Modelling low-level interactions on the Web: contributions to online learning, familiarity, and search behaviour evolution

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Science and Engineering

2022

He Yu
Department of Computer Science
Contents

Contents 2
List of figures 5
List of tables 7
Abstract 9
Declaration of originality 10
Copyright statement 11
Acknowledgements 12

1 Introduction 13

1.0.1 Research questions 17
1.0.2 Contributions 17
1.0.3 Thesis overview 18

2 Background and related Work 20

2.1 User Web interactions 20
2.2 Learning analytics 23
2.2.1 Data collection 26
2.2.2 Literature review 26
2.2.3 Properties of learning 30
2.3 Defining (Web) familiarity 32
2.4 Information retrieval and search skill 33

3 Methodology 37

3.1 Data pre-processing 37
3.2 Sequential pattern mining 39
3.3 Thematic analysis 41
3.4 User modelling 43

4 Modeling micro-interactions in self-regulated learning: a data-driven methodology 47

4.1 Chapter overview 47
4.1.1 Author’s contributions 48
4.1.2 Abstract 48
6.4 Study ................................................................. 92
  6.4.1 Platform and data collection .............................. 92
  6.4.2 Methodology .............................................. 95
6.5 Results .......................................................... 97
  6.5.1 Two periods ................................................. 97
  6.5.2 Four periods ............................................... 100
6.6 Discussion ................................................... 102
6.7 Conclusion .................................................... 105
  6.7.1 Limitations ............................................... 106

7 Discussion ....................................................... 107
  7.1 Findings and implications ................................. 109
  7.2 Theoretical and Practical Implications .................... 110
  7.3 Future research ............................................. 116
  7.4 Conclusion .................................................. 117

Appendices ......................................................... 155
  A Supplementary Material for Chapter 4 .................. 156
  B Supplementary Material for Chapter 5 .................. 160
  C Supplementary Material for Chapter 6 .................. 162
List of figures

2.1 Flow chart of pseudo systematic literature review ....................... 27

4.1 Graphical representation of the number of the active users for each session (brown line), the average of occurrences for each session among students who received a badge (blue line), the average of occurrences for each session among students who did not receive a badge (purple line) ....................... 63

5.1 Screenshot of the landing page of The University of Manchester School of Computer Science website ...................................................... 75

5.2 Distribution of responses for each survey question on each Likert items ... 82

6.1 A screenshot of the specialist search engine interface used in the study .... 93

6.2 User self-reported competency levels ........................................... 95

6.3 KMeans clustering with two periods (displayed sequentially from top to bottom), the figures on the left show the features Scroll and Click, the figures in the middle show Time and Session, and the figures on the right show Keyword with each axis presented on the same scale. Each user is represented as a data point and coloured based on their corresponding cluster. 98

6.4 KMeans clustering with four periods (displayed sequentially from top to bottom), the figures on the left show the features Scroll and Click, the figures in the middle show Time and Session, and the figures on the right show Keyword with each axis presented on the same scale. Each user is represented as a data point and coloured based on their corresponding cluster. 99

6.5 User transitions between clusters. Coloured nodes represent clusters in each period (left to right; period 1-4), with the users transitioning between clusters represented by grey streams. The number of users transitioning in each stream is listed sequentially in brackets. ......................... 102

A.1 Screenshot of the MOVING MOOC landing page. ......................... 158

B.1 URL domains with occurrences within each percentile and the overall user coverage of each group ................................................. 160

C.1 User self-reported competency levels ........................................... 162
C.2 KMeans clustering with three periods (displayed sequentially from top to bottom), the figures on the left show the features Scroll and Click, the figures in the middle show Time and Session, and the figures on the right show Keyword with each axis presented on the same scale. Each user is represented as a data point and coloured based on their corresponding cluster.

C.3 KMeans clustering with five periods (displayed sequentially from top to bottom), the figures on the left show the features Scroll and Click, the figures in the middle show Time and Session, and the figures on the right show Keyword with each axis presented on the same scale. Each user is represented as a data point and coloured based on their corresponding cluster.
List of tables

4.1 Example of captured student interface events. ........................................ 55
4.2 Example of theme generation process. The first column gives examples of
the initial codes (i.e. sequential patterns). The second column shows the
sub-themes generated from initial codes. The third column shows the final
themes (i.e. interactive behaviours) generated from the sub-themes. ......... 57
4.3 Description of the features of interactive behaviours. .......................... 58
4.4 Interactive behaviours generated from thematic analysis. ....................... 59
4.5 Behaviours and corresponding features correlated with achievement, where
correlation coefficients $| > 0.45|$ and $p < 0.001$. .......................... 59
4.6 Comparison between models including all the features (AllFeatures) and
models including engagement features only (EngFeatures). ................. 60
4.7 10-fold cross validation of the original dataset. .......................... 61
4.8 10-fold cross validation with oversampling. .......................... 62
5.1 Example of captured user interface events. ........................................ 76
5.2 Example of theme generation process. The first column gives examples of
the initial codes (i.e. sequential patterns). The second column shows the
sub-themes generated from initial codes. The third column shows the final
themes (i.e. activity patterns) generated from the sub-themes. ............. 78
5.3 Description of the features of activity patterns. .......................... 79
5.4 Activity patterns generated from thematic analysis. .......................... 81
5.5 Dependant variables constructed using survey responses ..................... 81
5.6 Number of responses for the survey questions at each point ................. 82
5.7 Activity patterns and corresponding features correlated with measured
variables ................................................................. 82
5.8 Classification results with different classifiers .......................... 83
6.1 Query and interaction features used to measure domain expertise, learning,
and search skill in previous research ........................................ 93
6.2 Example of captured user interface events. ....................................... 94
6.3 Cluster distribution and Silhouette coefficients for selected number of clusters
using KMeans ................................................................. 97
6.4 User transitions between clusters from period one to two .................... 99
A.1 Features of learning and their corresponding indicators identified from
previous works. ................................................................. 157
Abstract

Web technologies are constantly changing, and despite continuous advancements, the user still encounters difficulties interacting with Web platforms. Nowadays, researchers, Web designers, and website operators conduct extensive research on user Web interactions in order to better understand user behaviour, evaluate user experiences, and provide better support to users. Traditionally, user interaction research has relied heavily on controlled laboratory studies, surveys, and interviews. These techniques can potentially be intrusive, prone to bias, and lack ecological validity. While collecting and investigating users, Web interaction data provides an excellent alternative that is both unobtrusive, naturalistic, and ecologically valid. Web user interactions can be classified based on their level of abstraction, from the lowest level (e.g. finger pressing a key) to highest level (e.g. placing an order). We focus on low-level Web interactions, which are UI events generated by the user via user interfaces. There are several advantages of focusing on low-granularity data: it is easy to collect, contains a wealth of information, and potentially contains implicit user behavioural markers that are not observable in high-level data. Analysing fine-grained data, however, does present challenges; they can be noisy, less descriptive, and subject to high cardinality. There are also no well-established techniques for analysing low-level user interaction data. Thus, this thesis addresses this gap by proposing a data-driven methodology that combines data mining, qualitative analysis, and user modelling techniques in order to maximise the benefits of low-level data while mitigating its drawbacks. The proposed methodology and practical application of low-granularity user interaction data were demonstrated in this thesis through three studies, in the domains of online learning, Web familiarity, and search behaviour evolution. We identified user behaviours associated with the exploration of online learning forums, course materials, and Web navigation tools, and we were able to identify user groups with similar search behaviours and track changes in user behaviour over time. Additionally, our findings suggest that incorporating low-level user interaction into the analysis improves the explainability of user models. The significance of this work lies in the methodological contribution to the practical use of low-level user interactions in user modelling, as well as for the findings in the research areas of online learning, Web familiarity, and search behaviour evolution.
Declaration of originality

I hereby confirm that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright statement

i The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University’s policy on Presentation of Theses.
Acknowledgements

First and foremost, I wish to express my deepest gratitude to my supervisor, Dr Markel Vigo, for his invaluable advice, continuous support, and patience throughout my PhD. His guidance helped me all the time during my research, and I could not have imagined having a better supervisor. I also thank my co-supervisor, Prof Simon Harper, for all the insightful comments, suggestions, and the hard questions which incented me.

I would like to thank members of the IAM lab, Dr Aitor Apaolaza, Dr Ann Gledson, the people involved in the MOVING project, for your help and support.

I would also like to thank the University of Manchester School of Computer Science for the valuable opportunity to pursue my PhD here.

Lastly, I would like to thank my parents, my partner Dr Wolde-Michael, my new friends in the UK as well as old friends in China, and my cat, for your love and support.
Chapter 1

Introduction

With increasing access to the Internet and Web resources, digital skills required in contemporary society are often identified as part of the 21st-century skill set [Griffin and Care, 2014]. The connectivity to people and information is increasingly being supported and mediated by digital technology [Mangematin et al., 2014]. As a result, people are more familiar with various Web platforms and are better at navigating to find information to achieve their online browsing goals. Despite this, people can still find interacting with certain websites challenging [Eickhoff et al., 2013, White and Morris, 2007], which can result in, for example, e-commerce websites losing customers due to poor Web user experience or learners dropping out of online learning courses. As Web platforms and technologies continue to evolve, it is therefore important to continually monitor the quality of the user experience.

User Web behaviour is frequently studied to gain insights into the quality of the user experience, user characteristics, and user Web exploration outcome. It has been suggested that user behaviours can be recurring and periodic and that this periodicity may provide insights on user models [Aggarwal et al., 2020]. It has also been demonstrated that analysing these behaviours over time can be informative, and many works have focused on user behavioural patterns. Examples include; a time series analysis of educational log data which showed that the number of learning resources visited could be linked to learner achievement [Brooks et al., 2015a], pattern recognition techniques were used to identify excessive effort spent by users when completing tasks for identifying user interface deficiencies [Tamir et al., 2011]. Additionally, sequences of user behaviours investigated in information search sessions and results showing that user activity patterns could distinguish between task types and, to some degree, between task difficulty [Cole et al., 2015b], which can help better personalise user search experience.

Both qualitative and quantitative data analyses have been used in previous works regarding user Web behaviour. As a large amount of interaction data can be generated by users that is straightforward to collect, interaction data analysis is often at the centre of these works. When compared with traditional user experience studies that require self-reporting, survey participation, or lab monitoring, investigating user interaction offers a higher level of ecological validity for analysing user Web behaviours in the wild. While laboratory studies are more controlled and can generate more detailed interaction data using, for instance, eye-tracking technology, the reduced ecological validity of these studies may compromise their generalisability [Obendorf et al., 2007]. User studies that involve self-reporting and surveys can
be obtrusive and may be prone to, for example, sampling and confirmation bias. Investigating user interaction data offers an alternative, less obtrusive, and a highly ecologically valid method of analysing user Web behaviours. However, as interaction data is collected in the wild, without controlling for the tasks, topics, or user demographics, the ground truth of user behaviours may be subjective and open to interpretation.

User interactions can be characterised based on their level of abstraction, which varies from low-level physical events (e.g. keyboard presses) to high-level task-related events (e.g. completing a purchase) [Hilbert and Redmiles, 2000]. Many studies have been conducted at various levels of abstraction to investigate user interaction data; for instance, keystroke data was collected from a text editor to analyse keyboard function usage patterns, the results of which helped with the selection of the inverted T design of the arrow keys layout which is standard on keyboards today [Whiteside et al., 1982]. Compared with high-level user interaction data, investigating low-level user interactions has several advantages:

- **Low-level interface interactions are easy to collect.** Using simple scripts/tools, they can be collected efficiently, even automatically, without the need for annotation or human intervention.

- **Low-level interaction data is substantially larger in volume and more vigorous and can capture a variety of Web interactions.**

- **User interactions in low granularity may contain implicit behavioural markers - traces of unconscious or latent user behaviours, which are not observable in high-level interaction data.**

Investigating interaction data in low granularity is not without its challenges. Low-level interactions are less descriptive compared with high-level events, and therefore additional interaction context is often required to accurately interpret them. Low-level interaction data can also be noisy, as each high-level task comprises thousands or tens of thousands of low-level events. These low-level events may contain a considerable amount of repetitive or unintended user interactions. As there are many variations in Web interactions, especially for large-scale Web platforms with complex structures, building models and extracting insights from such interaction data is often more difficult due to the high cardinality [Apaolaza et al., 2015].

Currently, there are no widely accepted or well-established methods available for analysing low-level interaction data. This offers an opportunity to explore finer-grained users interactions and identify methods for learning and interpreting user behaviours. We propose a data-driven methodology to process and analyse low-level interaction data, which involves data mining techniques and qualitative analysis. A data-driven method, as opposed to hypothesis-driven approaches, has no clear direction at the beginning, and research is led by existing data. While a hypothesis driven approach only analyses a small number of isolated events based on pre-set hypotheses, a data-driven method allows us to investigate more complicated combinations of events and interaction patterns. Also, as a data-driven approach is less confined in the established scheme of thinking, it is less biased in the way that decision making...
can be more objective and the course of actions determined by data analysis. This therefore allows us to discover interesting, unknown behaviours and new insights. As mentioned above, concentrating exclusively on low-level interaction data makes it difficult to present a complete and clear picture of what the user is doing, even more challenging to analyse and investigate further. To mitigate this, additional contextual information can be collected and analysed in combination with the corresponding interactions [Apaolaza et al., 2015], for example, the triggering event, URL of the Web page, the specific Web elements triggered, and the timestamp of the event can be analysed together to enhance the semantics of the low-level interaction. It has also been suggested that both low-level and high-level interactions should be taken into consideration as information may be spread across multiple levels [Hilbert and Redmiles, 2000]. Thematic analysis was applied in our approach to systematically interpret the low-level interactions into high-level user behaviours where we consider user interaction patterns as codes and extract common themes among them. To manage noise and extract information from the interaction data, we apply a series of data pre-processing and pattern mining techniques to identify significant and meaningful user interaction patterns.

In this thesis, we begin by outlining the methodology used, followed by demonstrating the feasibility of our approach by applying it to an online learning platform - where we use low-level interaction data to model learners and their learning outcomes (Chapter 4). The rapid growth of online learning platforms (e.g. Massive Open Online Courses - MOOCs) has advanced the learning analytics research area as a significant amount of learner data can be collected and made available. The proposed methodology was applied on the MOVING MOOC\textsuperscript{1} which targets young researchers with an educational focus on academic information literacy, open science, and open research methods. The MOVING MOOC is a cMOOC platform (connectivist Massive Open Online Course), a type of MOOC that focuses on participatory learning with an emphasis on collaborative development. These differ from xMOOCs (eXtended Massive Open Online Courses), which are based on a more traditional classroom structure that involves pre-recorded video lectures and conventional assessments. One of the disadvantages of cMOOCs is the relative difficulty of accurately assessing knowledge acquisition, as participants are generally not focused solely on attaining a fixed set of knowledge or specific skills. In this study, we aim to provide an analysis of learner behaviours and indicators of learning outcomes for the purpose of user modelling and supporting platform design and quality learning. By investigating user interactions in the online learning domain, particularly within the cMOOC platform, we can potentially provide an alternative way to measure learning and ease the disadvantages of cMOOCs.

Our findings in Chapter 6 suggested that as learners engage with the platform, they become more familiar with it, and consequently, they become more proficient at navigating and using the platform. Therefore, we set our focus in this study to the Web familiarity domain. Previous research has investigated how user interactions evolve as they become more familiar with a Web page, and found that, for instance, as users become more familiar with the Web page, users’ interaction time with the mouse decreases while interaction session length

\textsuperscript{1}MOVING MOOC: Science 2.0 and open research methods, https://moving.mz.tu-dresden.de/mooc
within a Web page increases [Apaolaza et al., 2015]. However, this work was conducted via a hypothesis-driven approach that only focused on single events. In Chapter 5, we then conducted a secondary analysis aiming to measure user familiarity with low-level interaction data. The study is also motivated by the fact that as Web platforms differ, the user interface and functionality differs, users’ range of interactions and user models can vary, therefore, conclusions drawn from one may not be generalisable across different domains. To investigate the generalisability of our proposed methodology, we applied it to The University of Manchester School of Computer Science website (Chapter 5). Previous research has concluded that the sense of familiarity is often based on previous interactions, experiences, and learning from others [Luhmann and Morgner, 2019]. Research has shown that Web user experience is significantly impacted by familiarity, as users who are more familiar with a website are more efficient in browsing and navigation, and are less likely to be disorientated/lost [Chen et al., 2011, Galletta et al., 2006]. However, user Web familiarity as a strong indicator of user performance (e.g. navigation efficiency, information retention) and satisfaction has not been widely investigated. In this study, we aim to leverage low-level user interaction to indicate user level of familiarity with the website; our findings can be helpful for website designers to better support user browsing and navigation and improve user experience.

In Chapter 4, we briefly explored user behaviours’ occurrences over time and classified user behaviours based on the visualisations. Through the analysis in Chapters 4 and 5, we found that the features related to trends of user behaviour occurrences over time are mainly the features with stronger correlations with the predicted variables. Inspired by these findings, in Chapter 6, we demonstrate how low-level interaction data can be used to identify different user groups and investigate the evolution of user behaviours on a specialist search engine via unsupervised learning. The study was conducted on a specialist search engine (The MOVING search engine) which targets early career researchers. The cMOOC platform in Chapter 4 and the specialist search engine are part of the MOVING project\(^2\) which together allow users to improve their information literacy by training how to exploit data and text mining methods in their daily research tasks. Specialist search engines are information retrieval tools used by specialised audiences to search for information on specific domains. Despite the advanced support offered in modern search engines such as query formulation and document filtering, finding desired information can still be challenging, especially for specialist search engines [Eickhoff et al., 2013, White and Morris, 2007]. Monitoring user search behaviours has been shown to offer valuable information that can help search engines evaluate search interfaces and improve the quality of Web search experiences [Wang and Yen, 2007]; for instance, user search behaviour was used by Web search engines to analyse and predict user preference in order to present more relevant results and support user search experience [Teevan et al., 2005, Bogaard et al., 2019, Teevan et al., 2005]. Also, analysing low-level user interaction was shown to yield valuable information; for example, users struggling to locate desired information can be indicated. In this study, we aim to investigate user search behaviour and monitor search behaviour evolution using low-level user interactions.

\(^2\)The MOVING project (TraininG towards a society of data-saVvy inforMation prOfessionals to enable open leadership INnovation) http://moving-project.eu/
The implicit feedback from the user can help evaluate the search engine and enable targeted support.

Within this thesis, we use the term ‘micro-interaction’ and ‘low-level activity patterns’ interchangeably to match the terminology in different research domains. They refer to both user interactions patterns and user behaviours derived from low-level interactions.

1.0.1 Research questions

The thesis addresses the following research questions:

**RQ1. How can we use low-level interaction data for user modelling in each domain?**

Although low-level interaction data are easy to collect, users can quickly generate a significant amount of data, making data handling and analysis problematic. There is also a lack of an established methodology to interpret low-level user interactions. To mitigate these challenges, we propose a data-driven methodology involving sequential pattern mining and thematic analysis to extract and analyse low-level user interaction patterns. As a proof-of-concept, this methodology was tested on three separate platforms in different domains: an online learning platform with 193 students (Chapter 4), the University of Manchester School of Computer Science website with 35,819 users, including 268 revisiting users who reported their levels of familiarity (Chapter 5), and an academic search engine with 239 users (Chapter 6).

**RQ2. What insights can we learn about user behaviours in each domain?**

We performed three studies where we investigated user learning, browsing, and search behaviours. User behaviours were identified through thematic analysis of user interaction patterns in Chapters 4 and 5 and via cluster analysis in Chapter 6. Despite encouraging results from previous works, effectively utilising low-level user interactions to investigate user behaviours is still challenging, and any insights gleaned are uncertain. We sought for meanings in our findings after identifying the exhibited user behaviours, including what these behaviours imply about the users and how they connect to the user Web interaction outcome. User behaviours were included as features in user modelling to predict learning outcomes (Chapter 4) and user familiarity (Chapter 5). User behaviours over time were also investigated to investigate user search behaviour evolution (Chapter 6).

1.0.2 Contributions

This thesis contributes to the overall research domain of Human-Computer Interaction (HCI), particularly the user modelling and interactive data exploration domains. Our individual studies in Chapters 4 and 6 also contribute to the domain of learning analytics and information retrieval. The contributions are as follows:
- We investigated the application of low-level user interaction data in user modelling. We proposed a novel data-driven methodology and demonstrated the utility of this approach on three platforms, allowing for the generalisability to also be tested. We identified positive implications of the approach and noted opportunities for improvement that may be taken into account when applying to other Web platforms.

- We applied the methodology on an online learning platform where six interaction behaviours were identified that explained 82% of the variation in student achievement with a 91-95% accuracy in identifying low-achieving students. This information can be valuable for educators and online learning platform operators as it can be used to identify students who are struggling at an early stage. As mentioned before, one of the challenges of tracking learning progress for online learning platforms, particularly connectivist MOOCs (cMOOCs), is the lack of an assessment technique. Monitoring low-level user interactions may provide an assessment-free alternative. Moreover, when our models are compared to models with engagement-only features suggested in previous works, our approach offers 10% more explainability of student achievement variation based on Nagelkerke $R^2$ values, which is a measure of explained variation.

- The methodology was tested on the University of Manchester School of Computer Science website with user interaction data from 35,819 users over the course of 18 months. Within this, 268 revisiting users who reported their levels of familiarity with the platform were included in the final predictive models. Our model shows an accuracy of 82.7% when classifying users with higher levels of familiarity.

- On an academic search engine, we analysed the low-level interaction data of 239 users over the course of 20 months. User search behaviours were investigated using characteristics of users’ search queries as well as interface interactions. We were able to model user search behaviours, as well as evaluate how they developed over time. This allowed us to identify individuals whose search behaviours changed significantly and user groups whose behaviours changed in a similar fashion.

### 1.0.3 Thesis overview

This thesis is submitted in a journal/alternative format with permission from the supervisory team. Chapters 4, 5, and 6 are adapted from published or ready to submit journal/conference papers, written in the format appropriate for publication.

To provide a clear narrative and maintain coherency within the thesis, an overview section is included at the beginning of each of the three chapters. The overview sections explain the connections with previous chapters and the contribution to the overall aim of the thesis. The journal format was used to present this thesis as the three core chapters are connected and contribute to our overall goal, while each of these chapters also contributes to separate research domains and can be reviewed independently. The overall structure of the thesis and the content of each chapter are described below.
• Chapter 2 gives background on user Web interactions and the research area of learning analytics, familiarity, and search skill. In this chapter, previous works in modelling user interaction and interaction data exploration are discussed. A pseudo-systematic literature review was included that provides background in the area of online learning, assessments, and proprieties of learning.

• Chapter 3 describes the methodologies used in the subsequent studies, including their similarities and differences. It also presents the decisions made, along with our justifications when applying our methodology. Here, we review the overall data-driven methodology and the detailed differences applied to the studies. It also includes brief backgrounds on the techniques we adopted in the following chapters, such as sequential pattern mining and thematic analysis.

• Chapter 4 presents a study conducted in the self-regulated learning domain. It gives a background of user modelling techniques used in learning analytics and their challenges. We discuss the details of the study setup and methodology, present our results, and discuss our findings. In this chapter, we present the contributions of our finding to the self-regulated learning domain and demonstrate the usage of our data-driven methodology. The outcome of this initial study evaluated the feasibility of our methodology, its added value and limitations.

• Chapter 5 presents a study where we applied the methodology on the University of Manchester School of Computer Science website, where low-level interaction data was used to predict user level of familiarity. It gives a background on familiarity in general and in the context of Web exploration. In this chapter, we also discuss the findings in the user experience domain, implications of our findings, reflection on the proposed methodology.

• Chapter 6 describes a study on an academic search engine where low-level interaction data was used to model user search behaviour evolution. Here we used clustering to identify user groups and segmentation of user search sessions to evaluate user behaviours over time. We discuss the search behaviours their transitions and briefly discuss the impact of self-regulated online learning on the evolution of search behaviour.

• Chapter 7 provides a final concluding discussion where we revisit the research questions, evaluate the findings, and set out ideas for potential future work. Here we look at the overall findings and implications about our methodology derived from the three studies and discuss its advantages and challenges.
Chapter 2

Background and related Work

This chapter begins with an overview of user Web interaction, including the benefits and drawbacks of analysing low-level user interactions. In the following sections, we provide overviews of the topics and research domains relevant to our investigations - learning analytics, familiarity, and search skill - and proceed to discuss the research gaps identified within each of these domains.

2.1 User Web interactions

Over the past decade, the Web has become an interactive medium with an average of 9% growth in the world’s internet population every year [Kemp, 2021a]. Moreover, the Web has evolved significantly over the past decade - not only has the number of domains and users increased, but its character has also changed dramatically with the rise and ubiquity of social media platforms [Kemp, 2021a, Obendorf et al., 2007]. Advancements are constantly being made to aid Web users and improve user experiences through the introduction of new and improved Web innovations [Teevan et al., 2005, Bogaard et al., 2019]. User Web interactions are materialised by events generated by the user via user interfaces (UI) [Hilbert and Redmiles, 2000], which are now present on a wide range of devices such as laptops, mobile phones and tablets. These devices enable a wide range of Web-based interactions from users using keyboards, mice, touch displays, and other means [Albertos-Marco et al., 2016]. UI events generated by Web user interactions can provide in-depth information regarding user behaviour that can be captured, investigated, and analysed with the help of automated tools [Hilbert and Redmiles, 2000]. As UI events can be easily captured and used to indicate user behaviours, they can be considered a valuable source of information. Research has shown that by understanding the variability of user behaviours, interface improvements can be made to improve user information-seeking efficiency [White and Drucker, 2007], and that by identifying and modelling user behaviour, potential UI adaptations can be effected using predicted user behaviours [White and Drucker, 2007].

Analysing user Web behaviours in laboratory settings has been seen as disadvantageous by several studies [Apaolaza et al., 2013, Obendorf et al., 2007, Webb et al., 1999, Apaolaza et al., 2015]. Laboratory studies are well-established practices that provide a constrained environment in which confounding variables can be controlled. As such, these studies can
yield detailed data (e.g. eye movement tracking); however, they are also obtrusive and therefore possibly biased [Apaolaza et al., 2013, Obendorf et al., 2007, Webb et al., 1999]. Laboratory results are also heavily influenced by the assigned tasks and are unable to account for unanticipated circumstances in the real world [Apaolaza et al., 2013]. Their ability to accurately simulate the user’s daily work can be limited by the fact that users’ work environments often differ significantly from those observed in laboratories. One example of bias in Laboratory studies is the ‘Guinea Pig’ effect [Webb et al., 1999], which occurs when users become aware that they are being monitored and behave differently than they typically would. Moreover, it is not feasible to study individual user interactions over days or weeks, making it difficult to study distinct behaviours and recurring patterns [Apaolaza et al., 2015]. The potential distortion brought about by an artificial environment cannot be easily eliminated. Other data-collection methods such as questionnaires, interviews, ‘think alouds’, and focus groups are also considered intrusive and generally provide a less reliable view of users’ behaviours than participatory evaluation and direct observation [Novick et al., 2007, White and Drucker, 2007]. However, one method has been proposed as an unobtrusive alternative for data collection - automatically capturing user Web interactions. Among the advantages of analysing uncontrolled user interaction data are the availability of remote experimental settings and a broader target population, as well as the ability to obtain ecologically valid data [Apaolaza and Vigo, 2017]. For example, a long-term clickstream study was able to provide descriptive statistics on the behaviour of users on the Web and allow for long-term monitoring of user interaction and page re-visitation [Obendorf et al., 2007].

Interaction data has been used previously in several different contexts and application areas, such as predicting learner achievement [Brooks et al., 2015a], investigating Web search behavior [Kinley and Tjondronegoro, 2010], and inferring receptiveness to advertising [Guo et al., 2009]. For instance, learner log data was transformed into features to develop prediction models of learner success, and it was shown that the number of learning resources visited is associated with learners passing the course [Brooks et al., 2015a]. Web logs were used in conjunction with questionnaires and think-alouds to collect data on Web users. It was shown that users typically take a top-down approach to Web searching, beginning with a generic topic and then refining their queries to find specific information [Kinley and Tjondronegoro, 2010]. Low-level interaction data was captured via a modified toolbar to help develop a user interaction model for automatically inferring the interest of a user in search advertising [Guo et al., 2009].

Some data-collection techniques employ Web logs in which the data is too coarse to infer meaningful user interaction [Apaolaza et al., 2013]. User Web interactions with software applications can be analysed at multiple abstraction levels, where events at each level consist of events that occur at lower levels [Hilbert et al., 1997]. From the lowest level to highest level, the user Web interaction can be categorised as: physical events which involve a user interacting with physical devices (e.g. fingers pressing keys on a keyboard, hand moving the mouse), input device events generated by the device in response to the physical events (e.g. keys being triggered and mouse moving signals), UI events which manifest the input
device events on the user interface (e.g. keys entered in input fields, mouse movements on the screen), abstract interaction events (higher-level unit of user behaviours) indicated by UI events patterns (e.g. entering words in the input fields), domain/task-related events (unit of a user task, e.g. submitting a query), and goal/problem-related events directly related to overall interaction goals (e.g. document seeking and retrieval). The majority of modern UI systems can automatically generate user interface events in response to user interactions.

One of the advantages of investigating low-level user Web interactions is that they may contain traces of user unconscious behaviours. More detailed and fine-grained information than that provided by traditional log files is needed to obtain meaningful statements on how users interact with the Web [Apaolaza et al., 2013]. As low-level user Web interactions can capture users’ ‘implicit interaction’, they can provide additional information on how users interact with a Web application. Compared with explicit interaction, implicit interaction usually happens unconsciously [Atterer et al., 2006]. While users may be conscious as to their exploration goals, they may not be conscious of the interactions taken to achieve them. For example, rapid scrolling followed by reorientation indicates that the user is struggling to find the information they are looking for [Vigo and Harper, 2013], or time spent on a specific question in a questionnaire may indicate hesitation and doubtfulness [Atterer et al., 2006]. Although the notion of implicit human-computer interaction is well established [Schmidt, 2000], they have not been thoroughly investigated in modern Web platforms and UIs. In [Schmidt, 2000], researchers give an example of implicit interaction with the real world, defined as user interactions with the environment and artefacts to reach a goal. Implicit interaction, in the context of user web interactions, can be defined as the user’s observable unconscious behaviour while concentrating on achieving a goal [Schmidt, 2000]. Indeed, nowadays, people can generally complete a task or achieve a goal on the internet without focusing on the interaction with the computer or the user interface. For example, people place an order online without consciously considering where to click or fill in details. Implicit interaction traits such as the browsing speed and backtracking may be helpful to indicate user proficiency level [Galletta et al., 2006], and to monitor interface usability where user confusion and disorientation can be identified [Atterer et al., 2006]. Implicit information such as the above may only be provided by the user unconsciously, while low-level user Web interactions will spontaneously provide such variables [Schmidt, 2000]. While coarse interaction data allows Web designers to detect tasks that take an abnormally long time to complete, low-level interaction data provides deeper insights into the causes of these inefficiencies. For instance, Web designers can identify areas where users hover over elements for extended amounts of time, scroll excessively, or remain inactive for an extended length of time [Apaolaza and Vigo, 2017].

It has been said that high-level events do not directly reveal their composition from low-level events [Hilbert and Redmiles, 2000]. Thus, recording only high-level events will generally result in information loss. Moreover, low-level events do not carry enough information to reveal how they combine to form events at higher levels. Inferring higher-level events of interest from user interface events is often challenging, as low-level events typically do not contain enough information to firmly establish the application to which it should be related [Renaud
and Gray, 2004]. Therefore, methods of extracting information related to usability from UI events should consider that context may be spread across multiple events, and crucial contextual information may need to be recorded during data collection to allow for meaningful interpretation, and to describe how low-level events combine to form high-level events [Hilbert and Redmiles, 2000]. It has been shown that modelling the search context as well as behaviour can significantly improve the accuracy of ad click-through prediction compared to classification methods that do not model contextual and interaction information [Guo et al., 2009].

Another typical challenge of analysing low-level user interactions is that it requires a high degree of expertise and domain-specific knowledge [Apaolaza and Vigo, 2017, Apaolaza et al., 2015]. As user interface events typically contain substantial amounts of highly-detailed information, automated support is generally required to extract information at a level of abstraction [Hilbert and Redmiles, 2000]. We identify a methodological gap for utilising low-level interaction data that benefit from ecological validity and implicit interaction and address the contextual need and technical challenges.

### 2.2 Learning analytics

Technology Enhanced Learning (TEL), also known as ‘e-learning’, refers to ‘the use of technology to maximise the student learning experience’ [AdvanceHE, 2022], and often involves the application of modern technologies for teaching and learning. While TEL can refer to analogue and digital technologies, digital TEL, in particular, is expanding rapidly through the development of various educational software and online learning platforms. Learning Analytics tackles challenges in the Technology Enhanced Learning field. The widely agreed definition of Learning Analytics is ‘the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs’ [Siemens and Long, 2011]. This definition underlines the aim of learning analytics at transforming educational data into insights, decisions, and actions to improve learning and teaching. Measuring properties of learning (e.g. engagement, motivation, learning progress) [Brooks et al., 2015a, Eickhoff et al., 2014a, Prince, 2004, Zhang et al., 2006a], informing TEL decisions (e.g. curriculum design, platform improvements) [Xiao et al., 2015, Tsironis et al., 2016] and identifying and implementing adaptations [Pham and Wang, 2016, Rincón-Flores et al., 2019] at the forefront of research in Learning analytics.

Educational Data Mining (EDM) is a similar community to Learning Analytics. While the two communities share similar characteristics and goals, they hold distinct technological, ideological, and methodological positions [Liñán and Pérez, 2015, Lemay et al., 2021, Kavitha et al., 2017]. EDM and Learning Analytics both reflect the emergence of data-intensive approaches to education. They share the goals of improving teaching and learning by providing and enhancing the quality of analysis of large-scale educational data. However, the differences between EDM and Learning Analytics lie in the different focus on automated discov-
ery and leveraging human judgment where research in EDM utilises human judgement to assist automated data analysis, Learning Analytics utilises automated discovery to inform individuals [Siemens and Baker, 2012]. Another difference between the two communities is recognised - EDM models are more often used to help automated adaptation conducted by educational applications, while Learning Analytics models are more often used to inform and advise instructors and learners [Siemens and Baker, 2012]. Increased communication and collaboration between these communities has been called out and realised to share research, methodologies, and findings to advance both fields [Baker et al., 2021].

Human-centred learning analytics (HCLA) is a relatively new field that emphasises the human needs in learning analytics. HCLA involves the inclusion of users systematically throughout the design, implementation, deployment, and evaluation of the Learning analytics process to satisfy individual needs. The core of a human-centred approach is that aspects of the output system such as interaction designs, functionalities, and characteristics should be informed and defined by the intended users instead of platform designers or researchers. HCLA inspired design should therefore consider several aspects: the intended users, the intended usage with the system, the occasions of user interactions, the use of analytics, and the meaning of the analytics for the users [Giacomin, 2014].

Various process models have been proposed to conduct learning analytics studies. One commonality across these models is that they conceptualise learning analytics as a cycle. Learning analytics cycles often start with learners whose interactions generate data. The data are then processed and analysed, and the results finally inform interventions for learners [Clow, 2012]. Today, research in learning analytics focus more on collecting and analysing data about learners and their environments for understanding and improving learning outcomes. Moreover, increasing access to Massive open online courses (MOOCs) provides another way to collect learner data and a new particular field of analysis. MOOCs are the most recent and highly publicised entrant to a rapidly expanding universe of open educational resources, for example, organisations such as udacity, Coursera and edX. Different MOOCs exist, each with its structure, audience, and platform. The MOOC platforms described employ a typical classroom format that emphasises focused video material that mimics lectures—with short videos becoming more common nowadays—and frequently use automated testing to check learner knowledge as they progress through the course. These MOOCs have been dubbed ‘xMOOCs’. There is a distinct form of MOOCs that integrates learner-directed, open-ended learning, which is based on connectivist theories of learning, ‘cMOOCs’ [Mili ligan et al., 2013]. The connected aspect of learning is inherently personal and subjective, as participants create their own learning trajectory and build and navigate their own web of connections.

One concern for cMOOCs is that for novice learners, it has been argued that a minimal guidance connectivist approach is significantly less efficient than a worked example approach where instruction is linear and laid out [Kirschner et al., 2006]. However, for experts, a minimal guidance approach is more efficient than a worked example approach. It was argued based on cognitive load issue, and a connectivist MOOC is suitable for people with more de-
veloped learning strategies, greater motivation, and a capacity to remain motivated, whereas an xMOOC with ample guidance is one for more novice and intermediate learners [Kirschner et al., 2006]. Another challenge for cMOOCs is the lack of assessment. Whilst xMOOCs have discussion forums and allow people to communicate and discuss learning together, the core of the course is the guided lesson. Each learner’s journey/trajectory through the course is linear and largely depends on the absorption and understanding of a fixed set of knowledge. In xMOOCs, learning can be tested and certified. cMOOCs are non-restrictive, and participants set their own learning goals and engagement. They will not necessarily end with the intended set of specific skills or competencies or knowledge of a set body of content. cMOOCs are therefore tricky to grade, assess, or certify.

Another widely acknowledged criticism for MOOCs is the relatively low completion rates. Approximately only 3% to 13% of registered learners complete an online course [Lushnikova et al., 2012, Kizilcec et al., 2013a, Alonso-Mencía et al., 2020, Onah et al., 2014]. While this narrative has been used to argue and imply a binary categorisation of learners, this monolithic view of course completion does not consider the individual goals, background as the reason of discourse, learners who only wish to participate in certain aspects, or learners who do not require completion or achievement [Kizilcec et al., 2013a]. The sole focus on completion and binary categorisation of learners hinders the research on targeted intervention or personalised supportive adaptations. It has been shown that user goals and experience vary widely and have a substantial effect on user behaviour [White and Drucker, 2007]. Therefore, learner individuality should be considered without making assumptions about ‘correct’ or appropriate characteristics.

Research has shown that learning emerges from a dynamic process of interactions with and among learners, learning resources, and instructors [Kizilcec et al., 2013a]. Previous research on learner interactions has shown, for example, that high-level coarse-grained user interactions can be used to measure features of learning [Kizilcec et al., 2013b, Ferguson and Clow, 2015]. An overview of features of learning revealed that analysis primarily included interactions at the same abstraction level, typically coarse-grained events such as posting in forum [Crues et al., 2018]. Online learning platforms such as Coursera and Blackboard have been called to generate lower-level granularity data with rich semantics to enable the extraction of more interpretable information about learner behaviours [Maldonado-Mahauad et al., 2018, Fincham et al., 2019]. In the context of online learning, considering interaction patterns has also been suggested, as the order of interactions can encode valuable information about learning behaviour [Coleman et al., 2015].

Currently, the ability to use low-level fine-grained data to investigate learner interactions is distinctly absent. Our methodology utilises additional contextual information to add semantics and therefore achieve interpretability of learner behaviour while following a data-driven approach without presumptions of learner behaviours. Our study also addresses the assessment difficulty on cMOOC platforms, where we use low-level user interactions to analyse user behaviours and predict learning outcomes. We also provide additional background in the form of a pseudo systematic literature review of assessment technologies in learning de-
tailed in the following sections.

2.2.1 Data collection

We began the pseudo systematic literature review by searching for and examining relevant literature using Google Scholar, ACM Digital Library, IEEE Xplore, SpringerLink, and Wiley Online Library. The scope of the pseudo systematic literature review is assessments on MOOC platforms. We also included the ones focused on the properties of learning. Only the first 300 results for each search engine were considered to avoid publications with little relevance. There are mixed opinions for when MOOCs first emerged due to the ambiguity of the definition of terms such as ‘massive’, and ‘course’. In 2002, MIT established OpenCourseWare, followed by the release of OpenLearn by the Open University in 2006, representing the emergence of the ‘open education’ movement. However, the term ‘MOOC’ was first used by Dave Cormier in 2008 to describe a course called ‘Connectivism and Connective Knowledge’ (CCK08) offered by George Siemens of Athabasca University to over 2200 online students, and this is often quoted as being the first. The search query used was:

\[
\text{NOT instructor AND NOT teach AND NOT tutor} \\
\text{(AND (assess OR evaluate OR test OR formative OR summative) AND (student OR participant OR learner) AND NOT program AND NOT platform AND NOT environment AND NOT class AND NOT experience)} \\
\text{AND (MOOC or ‘massive open online course’)}
\]

The results of this search, after duplicates were removed, contains 1427 papers. The results were examined by the title and abstract, leaving 138 remaining. The excluded papers were primarily out of the scope of this pseudo-systematic literature review. We also excluded papers that were work-in-progress, posters, or works not on the topic of online learning. After the exclusion, 81 papers remained. The remaining papers were reviewed with their full-text documents. In the end, 62 papers were reviewed as the domain of some of the papers were irrelevant to computer science. Notably, these 62 papers included 11 papers that are on the topic of properties of learning.

2.2.2 Literature review

The most noticeable trend for assessment from the literature is the evolution from ‘assessment of learning’ to ‘assessment for learning’ [Stiggins and Chappuis, 2005]. It was suggested that assessment itself could be used as a learning strategy [Gálvez et al., 2009]. Research focused on different assessment methods in recent years has been considering assessment not only as a way to measure performance but also as an opportunity to help learners to improve learning outcomes.

It has been suggested that an e-learning system could be beneficial to students in their learning in three ways: enhanced engagement, self-directed learning opportunities, and faster
information dissemination [Leidner and Jarvenpaa, 1995, Liaw et al., 2007]. Assessment as a tool to help student learning was investigated by reviewing literature that claims to benefit students in these three ways.

The following sections describe the common themes discovered in the literature, including different assessment methods and common research directions in e-assessment.

**Comparison between e-assessment and traditional assessment**

Many research studies were focused on the impact of computer-based assessment and traditional ‘pen-and-paper’ assessment. Although there are mixed results and conclusions, with the development of e-assessment systems, the difference between traditional assessment and e-assessment is becoming less noticeable.

In 2002 researchers showed that a group of students who took an exam in the traditional manner performed better than the group of students that took the same exam on computers [Goldberg and Pedulla, 2002]. In the same year, it was also shown that tests taken on computer screens that require scrolling through the questions resulted in poorer performance than tests taken traditionally or on a computer screen without scrolling [Ricketts and Wilks, 2002]. Also, in 2002, It was concluded that students in higher educational institutions implementing e-learning in general performed better than those in face-to-face courses [Holley, 2002]. A decade later, in 2012, researchers observed the performance of students taking an online or offline assessment, and found that performance does not differ depending on whether the assessment format is according to their preferences [Hewson, 2012].

It is possible that, although whether a test is taken online or offline has minimal effect on student performance, the way online tests are designed and delivered, including the computer skill of the learner, affects their performance.

**Learning analytics**

The prime data source for most learning analytics applications is data generated by learner
activities. With the wide use of communication technologies in learning, the amount of data about students and their learning activities such as learner participation in formative assessments [Tempelaar et al., 2013] is accumulating rapidly [Aljohani and Davis, 2013]. The goal of learning analytics is to apply the outcomes of analysing the data gathered by monitoring and measuring the learning process to feedback to improve the same learning process, in this way, the assessment can also improve the learning outcome.

There are suggestions that learning analytics can assist student learning and teaching in different ways. Learning analytics focuses principally on data analysis at the level of the learning process [Siemens and Baker, 2012]. Six objectives were distinguished: predicting learner performance and modelling learners, suggesting relevant learning resources, increasing reflection and awareness, enhancing social learning environments, detecting undesirable learner behaviours, and detecting effects of learners [Verbert et al., 2012]. These objectives were commonly used in research studies for detecting disengagement and encouraging participation.

According to M.Booth, neither assessment nor learning analytics alone is enough to disrupt/transform education, adding that ‘assessment (when done well) is about the authentic and deep understanding and improvement of teaching and learning’, and that ‘analytics is about using the power of information technology to see patterns of success (or failure) in learning’. Only by combining the two, he claims, can we expect to produce something transformative, ‘a powerful inquiry into what supports authentic, deep, transformative learning for students’ [Booth, 2012].

**Predictive models to assess learning**

Analysing student performance indicators to assess learning has been a long-standing research direction [Giovannella et al., 2013]. While evaluating learner performance in an e-learning environment has been reported, it is still challenging to build predictive models for MOOCs [Qiu et al., 2016].

There have been conflicting conclusions drawn from different studies for performance indicators. For instance, it was reported that the use of discussion forums was a strong indicator of student performance [Alstete and Beutell, 2004], and online participation enhanced student engagement and thus improved learning effectiveness [Zhang et al., 2006a], while others found that students were able to learn equally well on online courses regardless of their level of online participation [Lu et al., 2003].

Another example is that researchers assumed learning styles to be a predictor of student success [Huang et al., 2012], while studies found that they had no significant effect on learning performance [Giovannella et al., 2013, Lu et al., 2003, Shih et al., 1998]. A summary of previous research on the relationship between learning style and learning performance showed that no unanimous conclusion could be drawn [Shaw, 2012].

Using predictive models to assess learning has been shown to be viable. Researchers investigated the possibility of predicting the learning outcome using machine learning tech-
niques [Al-Shabandar et al., 2017]. A strong correlation between clickstream actions and successful learner outcomes was identified. The results showed that machine learning is a viable approach to distinguish between success and failure outcomes. It was also demonstrated that the potential accuracy of forecasting techniques at identifying high-risk students early in the course term is possible [Wham, 2017]. The preliminary results suggested that 50 percent of students that earned a D or F could be identified prior to the start of the course.

A more general method is the learner model. Learner model records and diagnoses learning, including performance, knowledge level, misconceptions, etc. [Chrysafiadi and Virvou, 2013]. When the learner model is open to students, it is named as open learner model [Bull, 2004].

The research suggests that using a predictive model to measure performance is promising; however, the conclusions are mostly drawn without scalable implementation and evaluation, and the technology still needs support to be used on large-scale MOOCs or other platforms.

**MOOC assessment systems**

Assessment systems on MOOCs usually benefit from being interactive, allowing some level of personalisation. It was argued that MOOC design should favour a Learner-Centered Approach, providing strategies that are suitable for individual goals and a personal trajectory [Guàrdia Ortiz et al., 2013]. Therefore, it would be desirable if MOOCs allowed some degree of content customisation. State-of-the-art assessment systems are not only focused on evaluations but also on providing guidance and feedback to learners. In terms of the well-known MOOC platforms, a step unique to edX is Artificial Intelligence (AI) grading, which can machine-grade essays based on work conducted with natural language processing (NLP) and machine learning techniques. It was reported that its accuracy was close to instructor scoring. However, the AI grading system had shortcomings as it could not reliably grade answers which did not fit into the training examples, and it may be gameable for some students [Piotr F. Mitros et al., 2013].

It was stated that educational assessment has the purpose of improving a student’s understanding, measuring student performance, and evaluating the effectiveness of an educational program [Pellegrino et al., 2001]. Similarly, an interactive test system was proposed to enhance learning outcomes in computer programming courses [Yang et al., 2014]. After students answer the test item, the system gives feedback, including the correct answer, an explanation, and supplementary materials. Student feedback and the experimental results suggested that the proposed approach benefited students by correcting mistakes and guiding them to improve their learning process.

Assessment systems on MOOC platforms are being developed at a rapid pace. In the literature, we can see that many assessment systems have been tested and were shown to improve learning.

**Other approaches to MOOC assessments**
Novel approaches in MOOC assessment usually involve integrating AI technologies or learning theories such as the Bloom Taxonomy framework.

It was suggested that in many aspects, agents could be used to improve current automated tests [Daradoumis et al., 2013]. The benefits are, for instance, providing personalisation by modifying assignment questions according to the participants educational level, or altering the sequencing of the evaluation questions. It was also mentioned that agents could use the data produced from learning analytics of MOOCs to characterise the learning process and performance more efficiently.

Researchers have proposed a practical application of Shum and Crick’s theoretical framework [Shum and Crick, 2012] of a learning analytics infrastructure that combines learning dispositions data with data extracted from computer-based, formative assessments [Tempelaar et al., 2013]. The study showed that for students who had little experience in computer-based assessments, and who feel more comfortable with regulated learning, the digital test environments help in this transformation.

A recent paper published in 2018 discussed the use of adaptive learning technologies [Cai, 2018]. Benefits of adaptive technologies in education include enabling students to receive a real-time response to their work, providing teachers with insights into student needs, and increasing interaction productivity between faculty and online students [Kristen Hicks, 2015]. Adaptive learning technologies can customise learning based on the pre-determined knowledge state of the learner. An assessment-driven approach allows students to have their own learning path with individual learning stages, also providing different formative and summative assessments [Cai, 2018]. An adaptive learning system branded as Intellipath [Johnson, 2016] was used with detailed grading criteria based on student learning progress, mastery, and improvement. More and more adaptive assessment systems have been investigated in recent years, a trend that may be in coordination with personalisation. Adaptive assessment systems can take into consideration user progression and previous knowledge state. It has also been suggested that some courses may not be appropriate for implementing adaptive learning [Cai, 2018]. Courses with linear content, for example, can easily be adapted. Courses with units and lessons which are separate from one another do not allow for learning paths to be created.

Although these approaches are mostly still in the form of proposals, the results shown from the preliminary analysis appears promising.

2.2.3 Properties of learning

Engagement Student engagement has become the most popular aspect of learning to study in recent years. The motivation for its investigation derives from learning analytics and educational research, which established that students engagement with activities correlates well with performance [Kizilcec et al., 2013b].

Engagement was investigated mostly using learning analytics techniques. Research has
been conducted with user data to discover behaviours, patterns, and activities that can indicate engagement/disengagement. Many indicators were discovered for engagement, such as the curriculum design or individual prior experience. Interestingly, motivation as another property of learning has been shown to be a significant predictor of student course engagement [de Barba et al., 2016]. The relationship between motivation and engagement was also investigated, and it was found that motivation can predict behavioural, emotional and cognitive engagement [Sun and Rueda, 2012].

**Motivation** It is known that greater levels of self-regulation and motivation are required for online learning environments and self-regulated online learning respectively [de Barba et al., 2016]. However, motivation is a complex property and research on how it functions in MOOCs is still in its infancy. Research on student motivation is conducted chiefly through surveys and interviews. Motivation was shown to be connected with engagement and properties such as dropout and participation. It was shown that, although students are highly motivated throughout the first stages of MOOCs, only a tiny percentage stay to the course’s conclusion [Clow and Doug, 2013]. Previous studies have shown that a lack of persistence in online learning is associated with low levels of motivation [Hart, 2012, Vanthournout et al., 2012].

Lack of motivation is a primary reason for students dropping out of a MOOC [Xiong et al., 2015]. It has been shown that motivation influences and is influenced by students participation in an online course [Khalil and Ebner, 2014]. Five motivational components of motivation have been identified and are mentioned frequently in recent research: interest, achievement goals, value beliefs, self-efficacy, and control beliefs [Pintrich, 2003].

**Dropout and abandonment** A dropout in online courses refers to the persons who end their course participation, while abandonment refers to passive participants - students who remain registered without active collaboration in it [Rodriguez, 2012]. It is important to note that approximately only 10 percent of the registered students complete an online course [Lushnikova et al., 2012]. Although dropout rates and abandonment are easy to measure, the reasons these occur are the focus of relevant literature. Learning analytics is still the most commonly used technique to investigate dropout and abandonment. Many studies have investigated the reasons students dropout, revealing the most common causes of lack of intention to complete: personal circumstances, bad MOOC design, deficiency in digital skills, inaccurate expectations, and bad prior experience [Onah et al., 2014, Itani et al., Gerogina Gomez-Zermeno et al., 2016].

**Self-efficacy** Self-efficacy refers to an individuals confidence in his or her ability to effectively complete a specific task [Bandura, 1978]. It has been argued that self-efficacy influences an individuals behaviour in ways like performance in tasks and the amount of effort applied to tasks [Bandura, 1978, Willis III, 2013]. The instructional value of understanding self-efficacy lies in its ability to entail effort, participation, engagement, and performance. Self-efficacy is particularly critical for MOOCs, as students must have confidence in their ability to navigate and succeed in the course and the course’s ability to work for them.
Various studies have examined participants’ self-efficacy in relation to their learning achievement. A consensus opinion appears to have been reached; that self-efficacy significantly affects learning achievement, the motivation of individuals, and the final exam score [Zhang et al., 2001, Irizarry, 2001, Wang and Newlin, 2002].

### 2.3 Defining (Web) familiarity

The Oxford Languages dictionary defines familiarity as a ‘close acquaintance with or knowledge of something’. A sense of familiarity can be felt towards various things and processes. Research on familiarity is therefore spread across several domains and topics. Researchers have investigated topic familiarity, where a familiarity text classifier was developed aiming at improving re-ranking of Web query results [Kumaran et al., 2005]. Online information credibility and its relationship with familiarity were also looked at, showing, surprisingly, that familiar users have less trust in the information when they know its source [Lucassen and Schraagen, 2013], with research also showing that familiarity is a critical factor in resource selection [Quigley et al., 2002] and that users in general selected resources that they were familiar with [Xie and Joo, 2009, Bringman-Rodenbarger and Hortsch, 2020]. By exploring the familiarity of category labels in the technology industry, it was shown that familiarity influenced the sociocognitive dynamics of an emerging industry, revealing a bell-shaped relationship between both the familiarity of category labels and their adoption of early smartphone industry [Schneider et al., 2015]. It has also been shown that the experience of users with technology is strongly affected by their familiarity with it, in that novel (and unfamiliar) interactive artefacts require the development of new cognitive mechanisms [Tomasi et al., 2018].

Web familiarity research has mainly focused on aspects such as its effect on efficiency. Users have been shown to return to frequently visited websites twice as quickly as those visited occasionally [Bruce et al., 2004], and it has been shown that users who were familiar with a website developed more knowledge about the use of the system, which then made navigating the website easier compared to non-familiarised users [Veldof and Beavers, 2001]. In this case, users who were not familiar with the website had to re-read the task statements more often compared to the others [Chevalier et al., 2014], with the overall efficiency of users found to reduce when they first interact with an initially unfamiliar interface [Rosman et al., 2014]. Moreover, research on task performance in information retrieval tasks showed that familiarity has a positive effect on performance [Tomasi et al., 2018]. Information retention and recall was also shown to benefit from familiarity where web interface elements that are familiar to the users improved task performance [Wells et al., 2005]. Familiarity is also related to navigation, where research has shown that more familiar users are less likely to become confused/disoriented [Chen et al., 2011, Galletta et al., 2006], the level of trust or distrust [Zhang et al., 2006b, Luhmann and Morgner, 2019], and users’ perception of adverts’ entertainment value and importance, where familiar websites have been shown to have a positively influence [McCoy et al., 2013].
The ISO 9241-11 standard defines usability as "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use". Familiarity has also been researched together with usability, often focusing on the development of online consumer loyalty, and the role familiarity and usability have. At lower levels of familiarity, usability indirectly influences loyalty through consumer satisfaction. Alternatively, as familiarity increases, perceived usability now influences loyalty not through its effect on satisfaction as well as in a direct way [Casaló et al., 2008]. The results of this research confirmed that usability, familiarity, satisfaction and reputation have a direct and positive relationship with users’ loyalty as reflected in daily use of a news website [Hacek et al., 2017, Kaya et al., 2019]. Conversely, through a usability satisfaction assessment, participants non-familiarised with a website were found to be more satisfied compared to the familiarised participants, with the novelty of the website seeming to encourage participants non-familiarised with the website to be less demanding in their usability expectations, whereas familiarised participants less satisfied with usability [Chevalier et al., 2014]. This research suggested that the judgements of the familiarised participants may have been based both on their search activities during the experiment and their prior interactions with the website.

Familiarity can be viewed as being related to prior experience, repeated exposure, level of processing (e.g. processing of the meaning compared to the typeface of a word), and forgetting rate [Yonelinas, 2002, Moreland and Zajonc, 1982, Whittlesea, 1993, Weinreich et al., 2006]. Although the connection between familiarity and usability has not been explicitly made, familiarity is manifestly linked with usability metrics such as memorability. It states that the system should be memorable so that the casual user can return to the system after some time without re-learning [Hilbert and Redmiles, 2000]. Nevertheless, familiarity as a factor that can significantly impact user experience with websites [Chevalier et al., 2014, Chen et al., 2011, Ruddle, 2009] has not been widely investigated. User interaction analysis has long been focused on aspects of usability, and thus we identify this gap of analysing user interaction with a website to predict the level of familiarity. We also recognise similar problems identified previously, where measuring familiarity with interactive systems typically occurs in controlled settings or require users’ self-reporting and survey participation, which can be intrusive and prone to bias (e.g. selection and response bias). Our methodology provides an unobtrusive alternative.

2.4 Information retrieval and search skill

Finding information is the most common reason people go online, with nearly two-thirds of internet users stating that it is one of their top motivations [Kemp, 2021b]. Finding information, or Information retrieval (IR), is the ‘means by which users of an information system or service can find the documents, records, graphic images, or sound recordings that meet their needs or interests’ [Boyle et al., 2017]. IR is an active research domain in computing and information science that investigates the process of obtaining relevant information.
With an estimated 50 billion indexed pages available online [de Kunder, 2021], automated information retrieval systems are used to alleviate information overload. With 98% of people stating that they use a search engine at least once a month, web search engines are the most widely used automated IR applications [Kemp, 2021b]. Research has shown that continued use of a Web service is influenced both directly and indirectly by its perceived usability, trust in the medium, and the information quality it provides [Akram and Malik, 2012]. There is a critical need to satisfy search engine users in terms of the quality and relevance of search [Sameer and Balabantaray, 2015]. Modern commercial Web search engines have improved their prediction of user interests and intentions by monitoring user preferences and intentions [Teevan et al., 2005]. For instance, clickstream data has been used to determine user interests and re-rank Web search results [Bogaard et al., 2019, Teevan et al., 2005], and search engine logs of user behaviour have been used to address the task of learning rankings of documents [Radlinski and Joachims, 2007]. Significant improvements in rankings have also been demonstrated when users’ preferences for document complexity and readability are factored into Web search personalisation [Collins-Thompson et al., 2011].

Using search engines to navigate this space has evolved into a critical component of modern knowledge acquisition. It has been suggested that even the most ‘perfect’ search engine (i.e. one that returns precisely the desired results) may encounter problems; this occurs when users are unable to specify their information needs at a level sufficient for systems to function effectively [Teevan et al., 2004]. Additionally, it has been established that users engage in a search behaviour known as orienteering, in which they parachute into a relevant part of the information space and then rely on their recall and recognition abilities to guide them from there [O’Day and Jeffries, 1993]. Despite advancements in the relevancy and quality of search results, locating desired information can remain a difficult task, even more so for broad/exploratory research purposes [Eickhoff et al., 2013, White and Morris, 2007]. Casual users’ search behaviours are characterised by the use of fewer query terms, shorter search sessions, a focus on a single result page, and frequent errors with advanced query operators, all of which indicate difficulty satisfying information needs [Leroy et al., 2003, Hölscher and Strube, 2000, Jenkins et al., 2003]. While advanced tools have been developed to increase navigation efficiency, the quality of Web search experiences can still be improved by increasing search skills and utilising supportive search interfaces [Wang and Yen, 2007]. The majority of studies conclude that skilled searchers are more successful than unskilled searchers at completing search tasks. [Hölscher and Strube, 2000, Jenkins et al., 2003, Palmquist and Kim, 2000, Saito and Miwa, 2001] The term ‘search skill’ refers to the user’s ability to acquire desired information regardless of domain knowledge, such as through the use of advanced search expressions. Other research has demonstrated that individuals with a higher level of search ability integrate multiple search processes, including engagement with information content, interactions with content creators, and the ability to extract and manipulate information [Tucker, 2012]. Additionally, this term has been used to refer to users with more than five years of computer experience and more than a year of Internet experience [Jenkins et al., 2003]. Young people, in particular, rely heavily on search engines, but despite their familiarity with computers and ability to navigate the Web, they lack critical and analytical
skills when evaluating the information they find online [Rowlands et al., 2008].

It has been stated that most information retrieval systems take a ‘one-size-fits-all’ approach, displaying the same search interface to all users. However, one could argue that familiarity with the interface benefits users and reduces the cost to interface designers. However, it is recognised that users behave differently when confronted with difficulties with Web search [Joachims et al., 2005, Smith and Kantor, 2008], and that search outcome, to a great extent, is associated with behaviours such as the query reformulation strategies [Odijk et al., 2015]. Additionally, it has been demonstrated that individuals’ cognitive styles (i.e., their consistent tendency to use a particular type of information processing strategy), their background, knowledge of the Web, and search experience all have a significant impact on their search behaviour [Ford et al., 2002, Hölscher and Strube, 2000]. Therefore, as users perform more tasks using search engines, there is a growing need to understand more precisely what they are doing during the search process. Only with this understanding will we be able to create more effective interfaces that cater to a broader range of users’ queries and searching styles [White and Drucker, 2007]. Earlier research has proposed a variety of user behaviour models and applied them to a variety of information retrieval tasks, including deriving an optimal ranking principle [Fuhr, 2008], designing plausible evaluation metrics for IR systems [Moffat et al., 2013], and extracting relevant feedback from user behaviour logs [Joachims et al., 2017, Chuklin et al., 2015]. According to research, it is critical to understand the extent to which people’s search behaviours vary in terms of interaction flow and information sought when designing interfaces to assist World Wide Web users in searching more effectively [White and Drucker, 2007]. Understanding how these groups interact with the Web could enable the development of high-level improvement strategies targeted at each group [Tossell et al., 2012]. For instance, when we are able to model and identify consistent behaviour, we have the opportunity to adapt user interfaces to capitalise on predicted behaviour [White and Drucker, 2007].

Nowadays, more than 7 in 10 people report using at least one tool other than text-based search engines to locate information online on a monthly basis [Kemp, 2021b]. Numerous specialist search engines are being developed and used by an increasing number of people with tools that target specific domains and purposes. Specialist search engines are information retrieval tools that are targeted at particular audiences, domains, and data types. Among them are Wolfram|Alpha, a search engine for mathematical queries that require complex and dynamic computations, Google’s ‘Dataset search’, a search engine for datasets, and other search engines geared toward researchers and more advanced searchers [Fessl et al., 2019]. Several studies involving specialist search engines have been conducted, including an examination of how searchers interact with a web-based, faceted library catalogue for exploratory searches, utilising techniques such as eye-tracking and recall interviews to gain a better understanding of various aspects of search interface use [Kules and Capra, 2012]. Other investigations include gender-based inequalities in the context of résumé search engines, which are specialist tools for recruiters to search for candidates [Chen et al., 2018], and differences in search behaviours between a specialist search engine and ‘Google-like’ search interface for health
Navigating specialist search engines can be challenging: researchers searching through a large number of social sciences and humanities datasets struggled to gain a sufficient overview of each dataset and frequently chose datasets based on familiarity rather than relevance [Jay et al., 2016]. The ability to locate information quickly and efficiently is critical for researchers. To enhance the Web search experience and outcome, improvements can be made to either the search engine or the user's search efficiency [Moraveji et al., 2011], and while several methods of measuring the impact of search engine improvements have been proposed, with research often focusing on the algorithms, input modality, and visualisations, monitoring changes to search behaviour has proven to be more challenging [Moraveji et al., 2011, Teevan et al., 2004, Bhavnani, 2001].

Here we identify a gap for understanding user search behaviour variance by investigating user group behaviour characteristics. We aim to use low-level user data and cluster analysis to identify and investigate user group search behaviours. Analysing low-level data has been shown to yield valuable information in the past, such as how searchers examine and engage with web search results [Lagun et al., 2014, Yu et al., 2019], allowing the investigation of search tasks and information-seeking intentions [Liu et al., 2019], modelling the receptiveness of searchers in advertising [Guo et al., 2009], as well as the generation of effective personalised recommendations based on interest in retrieving specific information [Song et al., 2006]. We also intend to investigate search behaviour over time, as incorporating temporal factors enables us to monitor interaction characteristics such as session duration and the evolution of search behaviours. Additionally, user behaviours can be periodic and thus repeat over time, and the effect of periodicity on user models may provide insights on user models [Aggarwal et al., 2020].

When investigating search behaviours solely through low-level interactions without regard for the search topic or user objectives, interpreting users' explicit intentions and actions can be more challenging. Despite this, this work holds relevance for understanding real-world search behaviours as data collected by many search engine operators provide no explicit ground-truth information about user goals/desired query results [Scaria et al., 2014], and could help Web designers to better support specialist search engine users such as slow searchers and searchers who are struggling to satisfy their information needs.
Chapter 3

Methodology

In this chapter, we introduce the overall methodology used within the following studies in Chapters 4, 5, and 6. We present the decisions made at each step of the proposed methodology, as well as the criteria and justifications. We also discuss the similarities and differences of the steps applied in each study.

3.1 Data pre-processing

Although data pre-processing originated in data mining, it has evolved into a ubiquitous technique used in computing in general. Pre-processing data is frequently regarded as the first and, in some cases, the most critical stage of data analytics projects. Due to the fact that data collection is frequently loosely controlled, the resulting data quality may be low with, as evidenced by inconsistencies or missing values [Abbasian et al., 2018]. The data pre-processing phase frequently entails a number of data manipulation steps designed to improve the data’s quality and prepare it for further analysis. When there is a large amount of irrelevant and redundant data or noisy and unreliable data, knowledge discovery during the training phase becomes more difficult. A well-planned data pre-processing step is critical for avoiding potential problems during subsequent data analysis processes, as well as optimising model performance and simplifying model interpretation [Pyle, 1999].

User interactions can be classified depending on the level of abstraction, which ranges from low-level physical events (e.g. key press) to high-level task related events (e.g. completing an assignment) [Hilbert and Redmiles, 2000]. The platforms used in the following three studies were instrumented to generate low-level interface events, including a variety of Web interactions (e.g. mouse movements and keyboard presses). The example list of events that were captured can be viewed in Table 6.2 in Chapter 4 together with additional contextual information. Supplementary contextual information was extracted and used to process, reduce noise, and add semantics to the low-level interactions. The relevant additional information that was extracted in the studies are the following:

- User ID: user IDs were used to identify and tracks users. They were unique and anonymised codes (e.g. vYwkpQ3bXSm) stored in cookies assigned to the users the first time they accessed the website.
• URL: address of the Web page (e.g. cs.manchester.ac.uk) where the interactions occurred.

• Timestamp: time captured in milliseconds when the interface event was triggered (e.g. 1515026609).

• Node DOM and ID: the DOM element (e.g. BODY/DIV[1]/TABLE[1]/DIV) and ID (e.g. sidebar) of the Web page element where the interface event was triggered.

• Key: the corresponding key pressed when a keyboard event was triggered.

• Session start time: the timestamp when a Web page is loaded.

Similar data pre-processing steps were taken in Chapters 4 and 5 as the data was collected in the same low granularity and the subsequent data analysis steps are essentially the same. The steps were performed via Python scripts accompanied by the pandas library in Chapter 4, and via Javascript and PySpark in Chapter 5. Some of the interface events such as Copy and Paste are excluded as they were only partially supported by a few browsers. Events such as mobileTouchStart are excluded as they are only present for mobile users. Certain combinations of events were automatically triggered by the browser during user interactions; for example,mousedown and mouseup events typically follow each other, therefore to reduce noise and make it easier to process and interpret, these were combined and renamed as mousepress. To reduce noise, the same events which occur multiple times sequentially within the same Web page and node element were merged and renamed with the suffix ‘multi’ (e.g. multiple consecutive mousepress events were merged with the last occurring event renamed as mousepress+multi) while multiple sequential scrolling events were merged regardless of the triggering node element. Keydown events were renamed based on the function key pressed (e.g. commands key or alphanumeric key); for example, when the ‘Shift’ key was pressed, the event was renamed as keydown_command. Window events, which indicates when a Web page has gained or lost focus, combined with user-specific information, provide context for lower-level events, and was used to separate leavepage and switchtab events. leavepage events were created where both windowsblur and windowsfocus events are sequential and within the same tab (same page load timestamp), while windowfocus events with new page load times were marked as switchtab events.

To increase the interpretability and to reduce noise, we filtered and grouped the events by the URL domain topics (i.e. sub-domains such as the homepage, wiki page, and forum in Chapter 4), which can be extracted from the corresponding URL strings. In Chapter 5, as there are a large number of unique URLs (10,388) on the University of Manchester School of Computer science website, with some of them having expired or led to internal Web pages and documents, a few extra steps were involved. We first filtered the dataset by extracting the most commonly occurred and recognisable (i.e. each event triggers an occurrence) sub-domains: i.e. undergraduate, postgraduate research, staff, etc. To select the most representative URL domains included in the following analysis, we selected the set of URL domains by balancing the occurrences of additional sub-domains in the set and the added user coverage. We
extracted the occurrences of each of the URL domains in the interaction data. The URL domains with a larger number of occurrences were then added to the set where the overall user coverage (i.e., number of users who visited the domain) were calculated with each addition (see Figure B.1). By using this method and criteria, we are able to select a group of URL sub-domains that are widely used by the user. Chapter 5 also initially had 1780 unique node IDs, a large number of which were noisy and difficult to interpret. Similarly, we selected sets of node IDs to include for each URL sub-domains selected, again using occurrence and user coverage as the criterion to ensure that they were representative of user interactions.

Finally, to improve expressivity and ease the interpretation of the interface events, a event-node-URL triple representing each interaction was created for studies in Chapters 4 and 5. The triple consists of the interface event, the web element where it was triggered and the URL domain where it was triggered. For example, `mouseInorOut+main+wiki+multi` indicates multiple mouse movements in the main area of the wiki page. For the dataset in Chapter 4, we only used the node IDs to represent the web page element in the event-node-URL triple, while in Chapter 5, we used a combination of striped node DOMs and node IDs as the web elements are not all present as node IDs.

For Chapter 6, the data pre-processing steps are largely different from the that in Chapter 4 and 5 as we are conducting an exploratory data analysis in Chapter 6, and as we focus on a selected set of features that involves only a number of specific events (e.g., mouse clicks, scrolling) and that these events are collected and available independently within the dataset, events filtering and transforming was not needed. The interface events were, however, still categorised based on the URL domain, especially events performed on the search result page as the established features suggested by the previous research happens on the specific pages.

### 3.2 Sequential pattern mining

The goal of pattern mining is to uncover meaningful and hidden patterns in large datasets. The technique was widely used in retail settings to identify and analyse products that are frequently purchased together in order to better understand customer behaviour and make strategic sales decisions [Fournier-Viger et al., 2017, Aloysius and Binu, 2013, Borgelt, 2012]. In contrast to bag-of-words techniques, where user activities are investigated regardless of the order in which they were performed, sequential pattern mining (SPM) is a technique that identifies common patterns within sequences of data while taking into consideration the order in which sequences were performed [van Hoek and Carevic, 2017]. Preserving the temporal order in which user activity is conducted can be a significant advantage of SPM. These algorithms take as input a sequence database and a minimum support threshold and output sequential combinations in the data with occurrences above the set minimum support. This can then be used to find the most frequent sequences of events within a given dataset of interaction events [Mooney and Roddick, 2013].

The three types of patterns produced by SPM algorithms are frequent, closed, and max-
imal sequential patterns [Fournier Viger et al., 2017]. Patterns with occurrences surpassing a given minimum support are known as frequent sequential patterns (FS). Closed sequential patterns (CS) are frequent sequential patterns that are not sub-patterns of other frequent sequential patterns with the same minimum support, for instance we have sequence $S_a$ and $S_b$ with their support (i.e. the number of sequences that contain $S_i$ in the sequence database) being $sup(S_a)$ and $sup(S_b)$, then $CS = \{S_a | S_a \in FS \land \exists S_b \in FS \text{ such that } S_a \sqsubseteq S_b \land sup(S_a) = sup(S_b)\}$. Finally, maximal sequential patterns (MS) are frequent patterns that are not sub-patterns of other frequent sequential patterns (i.e. $MS = \{S_a | S_a \in FS \land \nexists S_b \in FS \text{ such that } S_a \sqsubseteq S_b\}$), and that $MS \subseteq CS \subseteq FS$. Research has shown that CM-SPADE generally outperforms other algorithms when mining frequent patterns [Fournier Viger et al., 2017], whereas CM-ClaSP and CloSPan are overall more suitable algorithms for mining closed sequential patterns, and VMSP performs better for maximal sequential pattern mining [Fournier Viger et al., 2014].

In Chapters 4 and 5, user interaction patterns were extracted using sequential pattern mining on the processed interaction data. To extract common patterns exhibited by groups of users, each of the input sequences were built using a single user’s interaction event across all active sessions. Here we first need to choose an appropriate sequential pattern mining algorithm, as well as a definition of user sessions and a suitable minimum support count.

In Chapter 4, we first benchmarked sequential pattern mining algorithms. We tested the execution time and memory use of a number of SPM algorithms, including CloSpan, CM-ClaSP, CM-SPADE, CM-SPAM, and VMSP, using the SPMF Java library [Fournier-Viger et al., 2016]. As all frequent patterns can be derived from maximal patterns, the set of maximal sequential patterns is often less than the set of (closed) sequential patterns. We recognised that it was more effective to pick a maximal sequential pattern mining approach to reduce the probable overlapping sequences in the following phase of the methodology in Chapters 4 and 5, where the occurrences of the detected sequences were sought in the original dataset.

Definition of user sessions is required to measure evolution and change in order to investigate interaction patterns and user behaviours over time. A user session can be defined in various ways, such as a continuous time of interaction (e.g. 10 minutes) or a session separated by specific interactions (e.g. leave the page). We used continuous inactivity as a criterion to separate the user sessions as it is unknown how long the user was spending on a task and accurately determine task completion vs abandonment. Sessions were divided into 5, 30, and 40 minutes of inactivity between two consecutive events, as recommended by the literature [Jones and Klinkner, 2008]. The selection criteria for the appropriate inactivity time are based on the generation of representative and informative sequential patterns, we are looking for session length that allows us to discover longer patterns exhibited by more users. The number of patterns in the search space depends on the set minimum support threshold and on how similar the sequences are in a sequence database. In general, as the minimum support is decreased, the number of sequential patterns found by SPM algorithms can increase exponentially. We maximised the number of patterns generated by a larger number of users to find representative patterns (i.e. the minimum support), and to find more descriptive patterns; we
focused on the sequence length as longer sequences contain more event-node-URL triples, which can be more informative regarding the user’s behaviour. We compared the datasets with different user session definitions with minimum support of 0.35-0.5 as values above 0.5 yielded a small number of patterns, while the number of patterns increases dramatically when the support decreases below 0.4. The minimum support and median were scaled to the same range. The minimum support, number of patterns, and their lengths were plotted for each dataset to identify intersections that maximised the criteria for representativeness and sequence length. This was measured as the point at which minimum support and sequence length are both maximised. This definition of user session was adopted in the latter two studies. The choice of using the maximal sequential pattern mining algorithms was also adopted in the study in Chapter 5. The only difference between the methods used in Chapters 4 and 5 is that for Chapter 4, the value space of minimum support we explored is much larger than in Chapter 5, as the number of patterns identified by the algorithm is much smaller in Chapter 5. SPM was not used in Chapter 6 as we only focus on a specific set of individual events.

Once we set the parameters, we then run the algorithms and extract the sequential patterns.

3.3 Thematic analysis

Thematic analysis is used to identify themes from the data that are important or interesting [Braun and Clarke, 2006]. There are a variety of approaches to conducting such analysis. The data suitable for this type of inquiry and research approach is qualitative, such as diaries, interviews, surveys, online forums, etc. In the domain of Human-computer Interaction, qualitative methods became common as researchers aimed to develop human-centred research practices that integrated the perspectives of the users [McDonald et al., 2019]. We employ thematic analysis to systematically interpret and group the extracted patterns produced by the SPM algorithms in the previous stage into higher-level interactive behaviours, being the first of its kind to adopt this method. Thematic analysis was used in studies in Chapters 4 and 5. There are various procedures to conducting thematic analysis, but the most common form follows a six-step process described in [Braun and Clarke, 2006]: familiarisation, coding, generating themes, reviewing themes, defining and naming themes, and writing up. The codes capture single ideas associated with a segment of data. They are conceptualised as the building blocks that combine to create themes. A theme captures a common, recurring pattern across a dataset. With small datasets, considerable overlaps between the coding stage and the stage of identifying preliminary themes may occur [Maguire and Delahunt, 2017]. The advantages of the approach include being more accessible, easier to implement, and theoretically flexible. It can be guided by concepts from a variety of fields, as well as research approaches (e.g. inductive, deductive, semantic, latent) [Braun and Clarke, 2019]. There are various theoretical approaches to thematic analysis, such as inductive/deductive and semantic/latent approaches. The distinction between inductive and deductive approaches is that an inductive approach involves allowing the data to determine the themes, and the themes are progressively refined, while a deductive approach involves some preconceived themes ex-
pected to find based on existing knowledge. There is also the distinction between a semantic and a latent approach: a semantic approach involves analysing the explicit content of the data, while a latent approach involves reading into the subtext and assumptions underlying the data.

We adopted the methodology with some adaptations: i) familiarising with the data, where we explore the variety of event-node-URL involved in the patterns and understand their meaning. ii) Using an inductive approach to determine themes due to the data-driven nature of this research. iii) Coding, where we considered the sequential patterns generated as initial codes and generate the themes. We first followed the semantic approach to transform each code into a sub-theme as a sentence describing the explicit interaction. After the generation of sub-themes, we followed the latent approach with assumptions about the underlying behaviours and determined the final themes. Each theme was created with a combination and transformation of the sub-themes. Examples of this process can be viewed in Table 4.2 (Chapter 4) and Table 5.2 (Chapter 5), where the codes are results from the sequential pattern mining (i.e. patterns detected in low-level events), and the final themes are the interactive behaviours/activity patterns users exhibit. For example, one of the patterns generated from sequential pattern mining is:

mousemove+sidebar+undergraduate
mouseinorout+MainBreadcrumbs+undergraduate+multi
mouseinorout+MainNavigation+undergraduate+multi
mousemove++undergraduate+multi

We first transform this into sub-themes: mouse movement at undergraduate page sidebar area-multiple mouse movement at undergraduate page breadcrumbs area-multiple mouse movement at undergraduate page navigation area, and finally we interpret this into a theme: Exploring the undergraduate page breadcrumbs, navigation, and sidebar area.

The thematic analysis process for Chapters 4 and 5 are essentially the same, while in Chapter 4, another coder was involved, and inter-rater reliability (IRR) was calculated. IRR is a statistical measurement designed to establish an agreement between two or more researchers coding qualitative data. We used Cohen’s kappa \(\kappa\) to measure this. However, with no one ‘accurate’ way to code data, the logic behind inter-rater reliability can be called into question. It has been argued argue that the IRR scores can show that two codes have been trained to code data in the same way, rather than that their coding is ‘accurate’ [Braun and Clarke, 2019]. While in the application of thematic analysis in low-level interaction patterns, the coding rules set out may influence the coding process a great deal, as the terminology of interpretation differs. Once the coding rules are set, however, the interpretation of interaction patterns are more uniform, leaving little room for ambiguity, which is the reason why we did not include a second coder in Chapter 5. In Chapter 5, as there were a more significant number of URL sub-domains and unique node IDs involved, there was a significant number of variations of the event-node-URL. This means there were a larger number of unique user interaction patterns, which in turn makes the number of themes more significant. A large number of unique interaction patterns and themes makes our dataset more sparse, as users tend to only exhibit a few of them. This can also be explained by the minimum support we
used for the study in Chapter 5, which is much lower than in Chapter 4, which means a smaller proportion of users tend to exhibit the behaviours. In the following user modelling process, to improve model performance, we returned to the thematic analysis process and combined certain themes that involve the same Web elements while in different URL domains; we also explain this in the next section.

3.4 User modelling

The user modelling in Chapters 4 and 5 shares similar processes with data preparation and feature engineering; both studies used supervised learning, while the user modelling technique used in Chapter 6 uses unsupervised learning.

For Chapters 4 and 5, the user behaviours (i.e. themes) generated from the thematic analysis are used to compute features. The behaviours identified through thematic analysis and their corresponding interaction patterns were sought in the original dataset for occurrences of each behaviour within each session for each user. These occurrences were represented in $m \times n$ matrices, where $m$ is the number of users, and $n$ is the number of sessions exhibited by the user, allowing us to investigate each user behaviour over time. In Chapter 4, we also created a boolean representation of the occurrences as an additional representation of data; this shows the active/inactive sessions for each of the behaviours.

For instance, the matrix for behaviour $B$ can be represented as follows:

$$B = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{bmatrix}$$

where $f_{ij}$ is the frequency or number of occurrences of behaviour $B$ for user $i$ in Session $j$. Since not all users have the same number of sessions, $f_{ij} = NA$ on those cells that do not have sessions.

Using the user behaviours matrix, we computed descriptive statistics of the frequencies, including the mean, median, and sum of occurrences. We also considered the ‘inactive’ sessions where the behaviours were not exhibited (i.e. $f_{ij} = 0$), such as the number of continuous ‘inactive’ sessions. Also, to show the evolution of interactive behaviours, we computed the frequency trend to indicate the general direction in which the number of occurrences is developing. To determine the trend, we calculated the mean, median, and sum of the occurrences across 3, 5, and 10 surrounding sessions. This is in consideration of the accuracy of the calculated trend for a relatively sparse dataset where long periods of ‘inactivity’ exist. The trend is calculated using two methods: the coefficient of its correlation with the session number variations and the slope in a polynomial equation with the horizontal line representing the increase in session number [Oliphant, 2006]. The trend, a combination of the aforementioned correlation and slope, was categorised into six levels of strength ranging from strong negative.
trend to strong positive trend. In total, we included 46 features for each behaviour as listed in Table 4.3 in Chapter 4.

In Chapter 6, as the number of user sessions and session duration varies for each user, to analyse user behaviour evolution across multiple search sessions, we used segmentation to group user sessions into periods and analysed the user behaviours in each period and behaviour change across periods. For each users interactions, we grouped the user sessions equally into \( N \) periods. For instance, when \( N = 2 \), we group the first half of the user sessions into one period and the remaining half into the other. When \( N = 5 \), each period contains 20% of the user sessions. For a user with a total of \( M \) sessions \( (S_M) \), each period \( P \) can be represented as:

\[
P_i = \{S_{1+\frac{(i-1)M}{N}}, \ldots, S_{\frac{iM}{N}}\}
\]

where \( P_i \) is the \( i \)th Period from a total of \( N \) periods. For the sessions that could not be equally divided into \( N \) periods, or when the total number of sessions is smaller than \( N \), these (additional) sessions are grouped into former periods.

In Chapter 6, After the grouping of user sessions, based on our analysis of low-level interactions in Section 6.3, we selected a set of established features that can be used to accurately identify search behaviours. These features were then computed for each user and averaged over all sessions for each period; they are used as the features in the following user modelling in Chapter 6.

In Chapter 4, We used indicators of learning as the target variable: the badge status (i.e. whether the user received a badge) and the number of learning units completed. Two surveys were involved in the studies in Chapter 5 and 6. For Chapter 5, the purpose of the survey was to compare users’ degrees of familiarity at different points in time. Responses to the questions were reported on a 5-point Likert scale. The survey results were used as the ground truth of user level of familiarity and were included as the target variable in the following user modelling. In Chapter 6, the search engine was hosted on the same platform as the cMOOC in Chapter 4. 62 users who enrolled in the MOOC also used the search engine. To establish individual learning paths for users within the MOOC curriculum, newly registered users were asked to fill in a short survey to assess their previous knowledge and digital competencies. In a three-point Likert scale, the survey assessed the perceived competency in the learning modules such as ‘Data and information literacy”; these survey results were used in the study in Chapter 6 to analyse the effect of self-regulated learning on user search skills evolution.

In Chapters 4 and 5, after the construction of features and dependant variables, we computed correlation tests between them. The correlation analysis is used as a feature selection step, where we select a subset of features to include in the following user model. Using the feature selection technique is an excellent way to deal with The curse of dimensionality, which, when increased, increases the amount of observation needed exponentially and makes it harder for the model to generalise the data, also increasing the risk of model overfitting. An example of the importance of feature engineering is when analysing DNA data, where there
can be thousands of features with much fewer samples.

We analysed the correlation results and selected the user behaviours with coefficients above a yield to include in the user models. To prevent overfitting, we only include the top two strongly correlated positive/negative features of each behaviour as long as they were not variations of the same feature. In Chapter 4, we used logistic regression analysis. We used the stepwise backward and forward techniques to build models and Nagelkerke $R^2$ to evaluate the models. For Chapter 6, we tested a variety of classification algorithms, Random Forest, Multi-layer Perceptron (MLP), K-Nearest Neighbour, and Support Vector Classification algorithms; The algorithms were evaluated based on the model performance (i.e. accuracy, precision, recall), we also performed a grid search for hyperparameter tuning for each classifier. We built models using the MLP algorithm and used precision, recall, accuracy in evaluating the models.

For Chapter 6, to analyse user search behaviour and user group characteristics, we use the clustering algorithm to model user groups. To select the clustering algorithm and the appropriate number of clusters. We performed Silhouette analysis on several clustering algorithms (i.e. KMeans, DBScan, Agglomerative clustering, and Meanshift). The Silhouette analysis is used for interpreting and validating the consistency within clusters [Arbelaitz et al., 2013]. It measures the degree of similarity of a sample to the assigned cluster compared to other clusters, and is often used to find the optimal number of clusters. Note that for the algorithms tested, only the Kmeans algorithm requires the pre-specification of the number of clusters. We computed and compared the average Silhouette scores for each algorithm and each number of clusters to select the clustering algorithms and the number of clusters. We then computed the clusters using the features mentioned before. We calculated the mean, median, and standard deviation of each feature for all users in each of the clusters, and statistics such as average time between clicks and average time spent on each keyword. The resulting clusters are analysed and labelled to represent the search behaviours and characteristics of its users using the descriptive statistics of the features of each cluster. The evolution of clusters and user behaviours over time was investigated by comparing the cluster characters and user transitions across different periods. We investigate the impact of MOOC participation on the evolution of search behaviours by comparing the characteristics and behavioural evolution of users who also participated in the MOOC, as well as with other users. We also compared the survey responses of the MOOC users on information literacy competence levels with their clustering results.

For Chapters 4 and 5, after the initial user modelling, we conducted 10-fold cross-validation. The k-fold cross-validation procedure is used to estimate the performance of a machine learning model on a limited data sample. As the sample distribution of the target variables is imbalanced in these two studies, we adopted resampling during cross-validation. To avoid losing information in the majority of examples to undersampling [Haixiang et al., 2016] and overly-optimistic estimates [Santos et al., 2018], we performed oversampling during the cross-validation procedure. Synthetic Minority Oversampling Technique (SMOTE) coupled with Tomek Links was performed as the combination of algorithms was suggested to prevent over-
fitting [Santos et al., 2018]. Stratified k-fold cross-validation was used to ensure that the proportion of positive to negative examples is kept in the folds. We report the results for the k-fold cross-validation and change in model performance (i.e. precision, recall, accuracy). In Chapter 4, we then conducted 10-fold cross-validation using various classifiers.

In Chapter 4, some of the features included in our previous regression analyses are closely related to engagement, such as the number of inactive sessions, as engagement is a property widely investigated and used to indicate learning outcomes in the online learning domain. We evaluate our methodology by examining the added value of including non-engagement metrics in the models. We computed a baseline model with those features that are known indicators of engagement and compared the $R^2$ values of the models, including all features the models including only engagement features. We computed the baseline model to compare our models against others consisting of common engagement features to further evaluate our methodology. We did not compute the baseline model in Chapter 5 as familiarity measuring features are less established.

At the end of Chapter 4, to investigate the occurrences of the behaviours over time, we analysed the occurrences of the identified user behaviours overtime, where we plotted the occurrences of the behaviours in each session for the users that eventually received a badge and users who did not alongside the average occurrences. We classified these user behaviours into clusters based on the observations.

In Chapter 5, the user matrices constructed were sparse as the URL domains were targeting specific user groups, and the URL domain and node IDs created a large variety of event-node-URL triples. As mentioned in the previous section, in pursuit of better classifying models, we combined certain themes so that the matrices were more populated. Themes that involve the same events and node IDs (despite the URL domains) were combined. We also sequestered nodes that were not navigational elements (e.g. main area), shortening the list as shown in Table B.2. Matrices were again calculated, along with correlations. We also adopted a separate method to re-select features - SelectKBest function from scikit-learn [Pedregosa et al., 2011], which select features according to the highest p-value and f-value scores; we then rerun the MLP algorithm with oversampling/added synthetic samples followed by cross-validation.
Chapter 4

Modeling micro-interactions in self-regulated learning: a data-driven methodology

4.1 Chapter overview

In Chapter 2, we introduced the concept of low-level user interactions, MOOCs and the online learning research domains, assessment methods on online learning platforms, and properties of learning. We discussed the advantages and disadvantages of low-level data analysis and the fact that, despite its potential, there are no established methods for it. It is unclear how they can be used or what practical utility and value they provide. We address this issue by introducing a data-driven methodology in the preceding section. In this chapter, we present a study in which we used the methodology to analyse user learning behaviours and predict learning outcomes using low-level user interaction data.

According to research, only about 10% of learners complete an online course, [Lushnikova et al., 2012], and user engagement typically decreases throughout the course. As user interaction data, such as clickstream data, is already being used to analyse learning properties such as user retention, engagement, platform efficiency, etc. [Park et al., 2017, Al-Shabandar et al., 2017], we believe that analysing low-level data can be advantageous because it may contain traces of user unconscious interactions and thus enable us to discover additional hidden user behaviours and contextualise them. Analysing user interaction data may be beneficial in increasing the understanding of user behaviours for course leaders or platform operators to provide better support and encouragement to users, allowing for possible personalisation and adaptations. Due to the ease with which low-level interaction data can be collected, we demonstrated how learning platforms can use low-level data for user modelling. This chapter focuses on the user learning outcomes, specifically user learning progress and achievement. The study was applied on a cMOOC platform, which often differs from a traditional classroom course structures found on xMOOCs. One of the distinctions or disadvantages is the absence of a method of assessment. Measuring knowledge acquisition in cMOOCs is challenging and relatively few studies have been conducted to address the lack of assessment on cMOOCs. Therefore, our work aims to address this gap by investigating whether low-level
user interactions can be used to effectively assess knowledge acquisition on cMOOCs.

The content of this chapter is adapted from He Yu, Simon Harper, and Markel Vigo. Modeling micro-interactions in self-regulated learning: a data-driven methodology. *International Journal of Human Computer Studies*, 151, July 2021. ISSN 1071-5819. doi: https://doi.org/10.1016/j.ijhcs.2021.102625. In this chapter, we use the term ‘student’ to refer to the previous chapters’ terms ‘learner’ and ‘user’. We use the term ‘micro-interaction’ to refer to both user interactions and user behaviours derived from low-level interactions. ‘Features of learning’ is used interchangeably with ‘properties of learning’ in this chapter.

Appendix A contains the chapter’s supplementary material. In Appendix A, we include an analysis of properties of learning and their corresponding indicators identified in previous research, a visual representation of the MOOC platform, and additional data for the results section.

4.1.1 Author’s contributions

He Yu carried out the literature review, data analysis, reporting and writing up. Markel Vigo provided continuous feedback and advice throughout all the stages of the study. Along with Simon Harper, they offered feedback and support that contributed to the work in this chapter.

4.1.2 Abstract

We explore whether interactive navigational behaviours can be used as a reliable and effective source to measure the progress, achievement, and engagement of a learning process. To do this, we propose a data-driven methodology involving sequential pattern mining and thematic analysis of the low-level navigational interactions. We applied the method on an online learning platform which involved 193 students resulting in six interactive behaviours that are significantly associated with learner achievement, including exploration of the first week’s materials and exploration of the forum. The value of including these behaviours in predictive models increased their explainability by 10% and accounted for an overall explainability of 82%. Performance evaluations of the models indicate 91-95% accuracy in identifying low-achieving students. Other relevant findings indicate a strong association between the reduction of the behaviours over time and student achievement. This suggests a relationship between student interface learnability and achievement: achievers become more efficient at using the functionalities of an online learning platform. These findings can provide context to learning progress and theoretical foundations of interventions against unhelpful learning behaviours.
4.2 Introduction

While online learning platforms provide students with the convenience to access learning materials at any time and location, concerns relating to the self-regulated nature of these platforms are growing. Tracking engagement, achievement, and abandonment features that occur during the learning process has become a rising research topic in the area of learning analytics. The reasons for this are twofold: first, we can provide context for assessment results and differences in student performance by understanding the differences in their level of engagement [Fredricks et al., 2004]. This information is valuable for educators as it can be used to identify students who are struggling at an early stage, with previous research suggesting, using performance predictors, that 50% of students who were close to fail or failed could have been identified before the start of the course [Wham, 2017]; second, for those online learning platforms such as connectivist MOOCs (cMOOCs), the absence of an assessment method is one of the challenges of monitoring learning progress. Monitoring these features may provide an assessment-free alternative.

Previous research has associated interactive behaviour with features of learning on Learning Management Systems (LMS) and online search [Yu et al., 2018a, Motz et al., 2019], where it was shown that some student interactions are indicators of engagement and can predict knowledge gain. For instance, the number of learning resources visited have been linked to learner achievement [Brooks et al., 2015a], the duration of learning sessions can be used to indicate knowledge status [Jansen et al., 2009], properties of search queries such as length or complexity have been linked to learning outcome [White et al., 2009], and the number of resources visited was shown to positively correlate with knowledge gain [Eickhoff et al., 2014a].

User interactions can be classified depending on the level of abstraction, which ranges from low-level physical events (e.g. key press) to high-level task related events (e.g. completing an assignment) [Hilbert and Redmiles, 2000]. Recent works demand that online learning platforms such as Coursera or Blackboard generate lower level granularity data with rich semantics to allow researchers to extract more interpretable information about what students do in order to create interventions [Maldonado-Mahauad et al., 2018, Fincham et al., 2019]. Established works use coarse-grained features that combine assessment data and user activity [Kizilcec et al., 2013b, Ferguson and Clow, 2015] suggesting the validity of behavioural data to measure features of learning. Crues et al. [2018] provide a comprehensive overview on these features, which mostly include interactions at the same abstraction level, typically coarse grained events such as the number of posts in the forum. Additionally, it has been suggested that both low-level and high-level interactions should be amalgamated as context may be spread across multiple levels and the composition of these events needs to be taken into consideration to give appropriate interpretations [Hilbert and Redmiles, 2000]. In the context of online learning, it has also been suggested considering interaction patterns, as the order of interactions can encode valuable information about learning behaviour [Coleman et al., 2015]. Low-level interaction patterns have been described elsewhere as micro-interactions [Breslav
et al., 2014], where repeated scrolling, re-visitation, and the number of times a response to an opinion question is changed were helpful in identifying issues such as disengagement, low self-efficacy, and confusion.

Inspired by the above works, we explore these connections in an online learning environment where the lack of formal assessment and the analysis of lower level interactions call for a data-driven approach. Specifically, we investigate whether micro-interactions (i.e. students’ interactive behaviours) can be used to measure learning progress, achievement, and engagement. In the context of this work, we define engagement as the student’s level of participation in course materials and activities on the learning platform [Fredricks et al., 2004], and the number of units completed and whether students received a badge are used to indicate learning progress and achievement respectively. We address the following research questions (RQ):

• **RQ1.** Can we identify micro-interactions that are relevant for online learning?

• **RQ2.** Which micro-interactions are associated with learning progress and achievement?

• **RQ3.** What is the added value of using micro-interactions?

To explore whether interface interactions are associated with learning progress and achievement, higher level task-related behaviours (e.g. watching a video) were generated from UI events (e.g. mouse clicks) by applying data mining techniques and qualitative analysis. Then, we explored the relationship between interactive behaviours and learning progress and achievement. Identifying these relationships provides both additional validation and insights on assessing learning. The contributions of this work are:

• We developed a data-driven methodology to isolate micro-interactions that are indicators of learning progress and achievement. We then showed the feasibility of our approach by applying this method on an online learning platform.

• Our approach identified six interactive behaviours which can explain 82% of the student achievement variation and provide a 91-95% accuracy in recognising low-achieving students. The added value of non-engagement metrics added a 10% of explainability of student achievement.

• In the learning platform that serves as a case study of the methodology, we discovered that interactive behaviours involving exploration of the forum, and exploration of the first week materials before leaving the page are significant predictors of student achievement.

• By discovering negative correlations between the frequency of interactive behaviours and achievement (i.e. the less frequent, the higher the achievement rate), we were able to associate the learnability of the online learning platform with the learning outcome.
4.3 Related work

This work builds on the metrics of online learning process (e.g. engagement, achievement), interactive behaviours to predict the features of learning, and sequential pattern mining to identify interactive behaviours.

4.3.1 Online learning metrics

In the context of self-regulated online learning, the definition and measurement of achievement can be a single (or a combination) of test scores and grades [Gong et al., 2018, Brooks et al., 2015a], course/personal learning goals [Farag, 2012, Wilkowski et al., 2014], and course completion [Robinson et al., 2016]. In a recent study, student achievement was measured as personalised learning objectives, and the rates of receiving a certificate significantly improved for students whose learning objectives include receiving a certificate when compared to the entire course population [Rohloff et al., 2020].

Engagement has been established to be closely related to learning [Kuh et al., 2008, Prince, 2004, Zhang et al., 2006a] in that the more students engage with activities the better they perform [Kizilcec et al., 2013b], while low-achieving students show lower level of engagement [Ding et al., 2018]. It has been suggested that low-achieving students receive fewer gamified badges and are less engaged than other students [Ding et al., 2018]. It has also been shown that evaluating engagement is key to assessing student retention, learning progress, and test performance [Baek and Shore, 2016, Crues et al., 2018, Ramesh et al., 2014, Singh et al., 2018]. Engagement is often measured by learner activities, and the amount of participation in online learning forums and the interaction with lecture videos are two of the most common measures of engagement. Motivation has shown to be a significant predictor of engagement [de Barba et al., 2016], and it was found that motivation can predict behavioural, emotional, and cognitive engagement [Sun and Rueda, 2012]. Previous work suggests that awarding badges has the potential to increase student motivation as it provides clarity of the learning goals and guidance on how to reach them [Hamari, 2017, Abramovich et al., 2013].

Approximately only 10% of registered students complete an online course [Lushnikova et al., 2012]. Hence, another frequently explored topic is student dropout/abandonment. Dropouts in online courses refer to the students who discontinue their participation in the course, while abandonment refers to passive participants —students who do not unregister and continue the course without being active participants [Rodriguez, 2012]. Many studies have investigated the reasons students drop-out, the most common reasons being lack of intention to complete, personal circumstances, bad course design, limited digital skills, unmet expectations, and negative prior experiences [Onah et al., 2014, Gerogina Gomez-Zermeno et al., 2016].
4.3.2 Modeling student interactions

Interactive behaviours and patterns refer to how the learners interact with an online learning platform and how these behaviours are exhibited over time. These factors have been the focus of previous works which involve assessment and prediction. Web navigation behaviours, for example, have been used to predict the next page students are likely to explore [Pardos et al., 2017], their level of engagement/participation [Ramesh et al., 2014], the likelihood to dropout [Whitehill et al., 2017], and their level of competence [DeBoer and Breslow, 2014, Käser et al., 2017, Falakmasir et al., 2016]. Similarly, student interactions have been used to evaluate participation behaviours in online collaborative learning [Daradoumis et al., 2006], as well as to predict gaming behaviours in an intelligent tutoring system [Muldner et al., 2011]. Further to this, informed by learning design, a recent study analysed interactive patterns to better understand student behaviours such as resource transition and review [Shen et al., 2020], and interaction patterns together with self-regulated learning strategies and self-reported variables were found to be accurate predictors of learner types [Maldonado et al., 2018].

It has also been suggested that learning is associated with interactive behaviours such as browsing patterns (e.g. how students transition through different tasks) [Geigle and Zhai, 2017], click patterns [Park et al., 2017], exploration strategies (including how students interact with videos) [Li et al., 2015a], exploration choices [Käser et al., 2017], and problem-solving strategies [Zhang et al., 2017, Guerra et al., 2014]. Research has shown that students’ grades can be accurately predicted using internal-course interactions combined with student activities and behaviours beyond the learning platform [Pérez-Sanagustín et al., 2019]. These works suggest that there are opportunities to relate online interactive behaviours with learning.

4.3.3 Sequential pattern mining

Sequential pattern mining (henceforth SPM) algorithms identify interaction patterns by extracting the sequential combination of interaction data that generates the most frequent patterns for a given support (i.e. the percentage of sequences that exhibit the pattern). SPMs have traditionally been used in retail where understanding shopping patterns can help increase profit by improving product allocation and display [Aloysius and Binu, 2013].

SPM algorithms generally produce three types of patterns: frequent, closed, and maximal sequential patterns. Frequent patterns are sub-sequences with frequencies exceeding a specified minimum support. Closed sequential patterns are frequent sequential patterns that are not strictly included in other sequential patterns having the same support. Finally, maximal sequential patterns are frequent sequential patterns that are not strictly included in other frequent sequential patterns. For instance, both pattern A \{i,j\} and pattern B \{i,j,k\} are frequent with their support counts being 0.3 and 0.2, respectively. We assume that for pattern A, its only frequent super pattern is B. In this case, pattern A is a closed pattern as its only super
pattern is less frequent than itself. However, pattern A is not a maximal pattern as its super pattern B is also frequent. The pattern B would be a maximal pattern if it has no super patterns which are frequent. For mining sequential patterns, CM-SPADE generally has the best performance [Fournier Viger et al., 2017], while CM-ClaSP and CloSPan are generally better suited for mining closed sequential patterns, and so is VMSP for maximal sequential pattern mining [Fournier Viger et al., 2014].

SPM is frequently used in the area of educational data mining; to identify sequences of events that can distinguish stronger/weaker groups in an online collaboration environment [Perera et al., 2008], and to extract learning features to classify learner groups [Wang et al., 2004]. It has been used in recent e-learning studies to improve online learning platforms, investigate online collaboration, and explore self-regulated learning [Dunaev and Zaytsev, 2017, Perera et al., 2008, Doko et al., 2018]. Previous work has also used SPM to mine student behaviours from length and correctness of steps taken in online programming courses [Hosseini et al., 2017]. Compared to other similar purpose data mining techniques such as process mining, SPM is more general as it can be applied to various types of sequences. Both techniques were evaluated qualitatively and quantitatively to predict dropouts in online learning, and it was shown that SPM is better suited to handle the data produced by learning processes for predictive purposes [Deeva et al., 2018].

4.4 Study: setting, apparatus and data collection

Interaction data was extracted from an online learning platform constructed as a cMOOC (connectivist Massive Open Online Course), a type of MOOC that focuses on participatory learning with an emphasis on collaboration and creation. The platform targeted a specific student group, early career researchers. The courses took place in three waves and ran for four weeks each time: 12th November–16th December 2018, 21st January–17th February 2019, and 17th June–14th July 2019. The learning topic was open science and open research methods. The learning materials were divided into four weeks and three learning modules including 63 micro-learning units. The three modules were:

- Data and information literacy: how to search, evaluate, and manage digital information.
- Communication and collaboration: how to use technologies to interact, share, communicate, and collaborate.
- Content creation: how to develop, integrate, and apply copyright to digital content.

The learning materials included text, videos, illustrations, links to online resources, etc. The platform provided a news page with updates about new modules, a wiki page where students share their reflections collaboratively, and a forum where students discuss the learning materials. Within each page there was a main area containing the learning materials and a left sidebar with links to other pages. For certain learning units the contents also included videos.
193 out of 382 students gave their consent to having their interactions collected. When each wave of the MOOC was finished, an online survey was sent to the students, which had a low response rate (10%). Hence, we should be careful making generalisations from the sample: the majority of students were based in the EU, the gender ratio was balanced and most of them were researchers or PhD students in natural sciences, engineering, and social sciences. We did not associate the demographic data to interaction data.

The platform was instrumented to generate low-level event interactive data, including browser window events such as page loads, and mouse and keyboard interactive data. Some of the events contain additional information, such as mouse coordinates for mouse events. We captured a total number of 281,087 interactive events of the types listed in Table 4.1 using WevQuery [Apaolaza and Vigo, 2017]. The low-level interface events generated directly from the platform captured student interactions in detail, which can lead to outputs that contain noise such as unintended/unnecessary student actions [Dev and Liu, 2017]. A series of pre-processing steps were conducted to combine or transform such events. For example, events which typically follow each other such as mouse-up and mouse-down, were combined and renamed as mouse-press. Key-down were renamed depending on whether they were commands (i.e. return key) or an alphanumeric key and same consecutive events were also merged. To ensure the interpretability of the event sequences, we grouped them by topic, i.e. the homepage, weekly material page, wiki page, forum, news page, and settings. Each recorded event had a corresponding timestamp, URL of the Web page where the event takes place, and the specific DOM element that triggers the event. Hence, a event-node-URL triple represented each event. For example, mouseinorout+main+wiki+multi indicates multiple mouse movements in the main area of the wiki page.

We computed two indicators of learning: the badge status and the number of units completed. The instructors of the course awarded an ‘Open Science Aficionado’ badge to those students that conducted weekly assignments. The specific requirements of the badge were to achieve at least 12 out of 24 attainable points from the following activities: vote on other student’s forum post (1 pt.); comment on other student’s forum post (2 pt); complete weekly assignment (3 pt.). On the weekly assignment of week 1, students had to read an article about the opportunities and challenges of Open Science. On week 2, students had to use social media to share ideas, ask for advice and find research partners to run a research project. Week 3 was about sharing research using workflows involving DOIs, ORCID, open data repositories such as Zenodo and use of Altmetrics. Finally, on week 4, students had to design a research plan using what they had learned in the previous weeks. While the data of all the participants was used for pattern analysis, we excluded 53 students from the predictive modeling analysis as the first wave of the course did not award any badge, leaving 140 students for the analysis involving badges and 193 for progress analysis.
<table>
<thead>
<tr>
<th>Type</th>
<th>Events</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>mousedown</td>
<td>Start of mouse click action</td>
</tr>
<tr>
<td></td>
<td>mouseup</td>
<td>End of mouse click action</td>
</tr>
<tr>
<td></td>
<td>mousemove</td>
<td>Mouse movement</td>
</tr>
<tr>
<td></td>
<td>mouseover</td>
<td>Hovering into target</td>
</tr>
<tr>
<td></td>
<td>mouseout</td>
<td>Hovering out from target</td>
</tr>
<tr>
<td></td>
<td>doubleclick</td>
<td>Double mouse click</td>
</tr>
<tr>
<td></td>
<td>mousewheel</td>
<td>Mouse wheel interaction</td>
</tr>
<tr>
<td>Selection</td>
<td>select</td>
<td>Selection of page content</td>
</tr>
<tr>
<td></td>
<td>cut</td>
<td>Content cut</td>
</tr>
<tr>
<td></td>
<td>copy</td>
<td>Content copy</td>
</tr>
<tr>
<td></td>
<td>paste</td>
<td>Content paste</td>
</tr>
<tr>
<td>Keyboard</td>
<td>keydown</td>
<td>Start of key press action</td>
</tr>
<tr>
<td></td>
<td>keyup</td>
<td>End of key press action</td>
</tr>
<tr>
<td></td>
<td>keypress</td>
<td>Key press action</td>
</tr>
<tr>
<td>Window</td>
<td>load</td>
<td>Page is loaded</td>
</tr>
<tr>
<td></td>
<td>resize</td>
<td>Browser window is resized</td>
</tr>
<tr>
<td></td>
<td>unload</td>
<td>Window is closed</td>
</tr>
<tr>
<td></td>
<td>windowfocus</td>
<td>Browser tab gains focus</td>
</tr>
<tr>
<td></td>
<td>windowblur</td>
<td>Browser tab loses focus</td>
</tr>
<tr>
<td></td>
<td>scroll</td>
<td>Change of scroll state</td>
</tr>
<tr>
<td>Other</td>
<td>change</td>
<td>Input element state change</td>
</tr>
<tr>
<td></td>
<td>contextmenu</td>
<td>Opening of context menu</td>
</tr>
</tbody>
</table>

4.5 Methodology

We propose a three-step methodology to mine interaction data, analyse sequential patterns, and generate interactive behaviours of interest. First, we benchmark sequential pattern mining algorithms and different versions of our dataset to identify the combination yielding the most representative and informative patterns. Once we set the parameters, we run the algorithms and extract the sequential patterns. Second, we apply thematic analysis to interpret and group the extracted patterns, transforming low-level interactions into higher-level interactive behaviours. Third, the interaction patterns that fall under the emerging themes are used to build user models.

4.5.1 Benchmarking algorithms and datasets

The success criteria for this stage aim at maximising the number of patterns generated and the sequence length. We benchmarked a number of SPM algorithms including CloSpan, CM-ClaSP, CM-SPADE, CM-SPAM, and VMSP, which were provided by the SPMF Java library [Fournier Viger et al., 2014].

To find representative patterns we maximised the number of patterns generated by a higher number of users. Each of the input sequences was constructed with a single student’s interaction events across all active sessions. To assess how representative a set of patterns are, we explored the value space of the minimum support parameter (i.e. the percentage of students that exhibit a given sequence) in the 0.35–0.8 range as values above 0.5 yielded a small number of patterns and the number of patterns increases dramatically when the support decreases below 0.4. To maximise the amount of information, we maximised the length of the pattern.
Longer sequences entail more events, and thus more information. Single-event sequences were therefore excluded.

To explore interaction patterns over time, we split our dataset into sessions. Sessions were split by 5, 30, and 40 minutes of inactivity in between two consecutive events which is in line with what is suggested in the literature [Jones and Klinkner, 2008]. This resulted in three different datasets: gap-5, gap-30, and gap-40 respectively. We compared the medians of sequence size of the three datasets with minimum support of 0.35–0.5. The number of patterns and their lengths were plotted for each dataset to identify intersections that maximised the criteria for representativeness and sequence length.

4.5.2 Thematic analysis

Thematic analysis is a widely used qualitative analytic method for identifying, analysing, and reporting patterns (common topics/themes) within, typically, qualitative data [Braun and Clarke, 2006]. While the use of thematic analysis has been used to evaluate online learning platforms [Meinert et al., 2018, Ossiannilsson et al., 2015, Alturkistani et al., 2019, Paphthoma et al., 2015], we employ thematic analysis to systematically find themes on the patterns produced by the SPM algorithms in the previous stage and reduce the high number of sequential patterns (and address RQ1), being the first of its kind to adopt this method. We adopted the most common form of thematic analysis, a six-step process described by Braun and Clarke [Braun and Clarke, 2006]: familiarising with the data, coding where we considered the sequential patterns generated as initial codes, and generating the themes. We conducted an inductive approach to determine themes as the data-driven nature of this research. We first followed the semantic approach to transform each code into a sub-theme as a sentence describing the explicit interaction. The sub-themes are essentially a more human-interpretable format of the initial pattern. After the generation of sub-themes, we followed the latent approach with assumptions about the underlying behaviours and determined the final themes. Each theme was created with a combination and transformation of the sub-themes. Examples of this process can be viewed in Table 4.2 where the codes are results from the sequential pattern mining (i.e. patterns detected in low-level events) and the final themes are the interactive behaviours student exhibit. The interpretation of the behaviours reveals the strategies used by students such as interacting with videos and forums. An independent coder was involved in the thematic analysis. The inter-rater reliability was particularly high, Cohen’s $\kappa = 90\%$. This may be due to the fact that sub-theme and theme analysis was conducted under a set of agreed rules: for example, load+windowfocus was defined as ‘explore’ rather than ‘load’, ‘focus’, or ‘view’, which reduced ambiguity.

4.5.3 Modeling student interaction

To investigate RQ2, the behaviours identified through thematic analysis and their corresponding interaction patterns were sought in the original dataset for occurrences of each behaviour
Table 4.2. Example of theme generation process. The first column gives examples of the initial codes (i.e. sequential patterns). The second column shows the sub-themes generated from initial codes. The third column shows the final themes (i.e. interactive behaviours) generated from the sub-themes.

<table>
<thead>
<tr>
<th>Codes</th>
<th>Sub-themes</th>
<th>Final themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>load+week_1</td>
<td>load and open the week 1 page</td>
<td>Explore the week 1 page main area, left sidebar, and interact with videos</td>
</tr>
<tr>
<td>windowfocus+week_1</td>
<td>multiple mouse movement at week 1 main area</td>
<td></td>
</tr>
<tr>
<td>mouseinorout+main+week_1+multi</td>
<td>multiple scrolling in week 1 page</td>
<td></td>
</tr>
<tr>
<td>scrollorwheel+week_1+multi</td>
<td>multiple mouse movement at week 1 left sidebar</td>
<td></td>
</tr>
<tr>
<td>mouseinorout+sidebar_left+week_1+multi</td>
<td>video activity in week 1 page</td>
<td></td>
</tr>
<tr>
<td>video+week_1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>windowfocus+week_1</td>
<td>open the week 1 page</td>
<td>Explore the homepage main area then leave page</td>
</tr>
<tr>
<td>mouseinorout+main+week_1+multi</td>
<td>multiple mouse movement at week 1 main area</td>
<td></td>
</tr>
<tr>
<td>mouseinorout+sidebar_left+week_1+multi</td>
<td>multiple mouse movement at week 1 left sidebar</td>
<td></td>
</tr>
<tr>
<td>video+week_1</td>
<td>video activity in week 1 page</td>
<td></td>
</tr>
<tr>
<td>load+home</td>
<td>load and open homepage</td>
<td></td>
</tr>
<tr>
<td>windowfocus+home</td>
<td>multiple mouse movement at homepage main area</td>
<td></td>
</tr>
<tr>
<td>mouseinorout+main+home+multi</td>
<td>leave home page</td>
<td></td>
</tr>
<tr>
<td>blurfocus_leavepage+home</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

within each session for each student. These occurrences were represented in \(m \times n\) matrices, where \(m\) is the number of students and \(n\) the number of sessions exhibited by the student. For instance, the matrix for behaviour \(B\) can be represented as follows:

\[
B = \begin{bmatrix}
    f_{11} & f_{12} & \cdots & f_{1n} \\
    f_{21} & f_{22} & \cdots & f_{2n} \\
    \cdots & \cdots & \cdots & \cdots \\
    f_{m1} & f_{m2} & \cdots & f_{mn}
\end{bmatrix}
\]

where \(f_{ij}\) is the frequency or number of occurrences of behaviour \(B\) for student \(i\) in Session \(j\). Since not all students have the same number of sessions, \(f_{ij} = NA\) on those cells that do not have sessions.

To identify which behaviours were associated with indicators of learning, we computed descriptive statistics of the frequencies including the mean, median, and sum of occurrences. We also considered the sessions where students were ‘inactive’ (i.e. \(f_{ij} = 0\)), as dwell time is indirectly related with achievement through engagement [Yi et al., 2014, Lu et al., 2019].

Also, to show the evolution of interactive behaviours, we computed the frequency trend to indicate the general direction in which the number of occurrences is developing. To calculate the trend, we calculated the mean, median, and sum of the occurrences across 3, 5, and 10 surrounding sessions. This is in consideration of the accuracy of calculated trend for a relatively sparse dataset where long periods of inactivity exist. The trend is calculated using two methods: the coefficient of its correlation with the session number variations, and the slope in a polynomial equation with the horizontal line representing the increase of session number [Oliphant, 2006]. The trend, a combination of the aforementioned correlation and slope, was categorised in six levels of strength ranging from strong negative trend to strong positive trend. In total, we included 46 features for each behaviour as listed in Table 4.3. We then conducted correlation and regression analysis between these features, and the number of units completed (indicator of progress) and the badge status (indicator of achievement).
4.6 Results

**Benchmarking** None of the five SPM algorithms was found to be problematic in terms of execution time and memory usage. As far as the number of patterns was concerned, for all of the datasets, CloSpan produced the fewest patterns while CM-SPADE produced the most. The ranges of sequence size under the minimum support of 0.5–0.8 was 2–4 for CloSpan and 2–8 for CM-SPADE, CM-ClaSP, and VMSP. In consideration of the next step of the methodology, it was more efficient to select a maximal sequential pattern mining algorithm to eliminate the possible overlapping sequences. We therefore performed our analysis using the maximal sequential pattern mining algorithm – VMSP on the three datasets.

The dataset with sessions separated by 40 minutes of inactivity (i.e. gap~40) is the one that generated the longer sequences exhibited by the largest number of students. This was measured as the point at which minimum support and sequence length are both maximised – which was 0.43 for minimum support. As a result, a total number of 110 patterns were generated with a sequence length range of 2–10.

**Thematic analysis** From the initial 110 patterns, 23 themes emerged from the thematic analysis, as shown in Table 5.4, including the exploration of the homepage, wiki, and the learning materials of week 1. Video and forum activities, which are known to be associated with learning [Cormier and Siemens, 2010, Atapattu and Falkner, 2017, Breslow et al., 2013] were involved in six and five of the 23 interactive behaviours, respectively. Exploration of the left bar menu, which contains the links to different resources, is also a recurrent element suggesting intent to explore pages. We also found that ‘leaving a page’ is present in four of the interactive behaviours, suggesting the action of clicking a link provided in learning materials or the student being distracted from the course (i.e. withdrawing from exploring its contents). Eight interactive behaviours were found to involve interactions on different pages including homepage, materials of week 1, and forum.
### Table 4.4. Interactive behaviours generated from thematic analysis.

<table>
<thead>
<tr>
<th>Index</th>
<th>Interactive behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Visit the wiki page</td>
</tr>
<tr>
<td>2</td>
<td>Explore the wiki page and left sidebar</td>
</tr>
<tr>
<td>3</td>
<td>Explore the week 1 page main area, left sidebar, and interact with videos</td>
</tr>
<tr>
<td>4</td>
<td>Explore the week 1 page main area, left sidebar, and forum main area</td>
</tr>
<tr>
<td>5</td>
<td>Explore the week 1 page main area then leave page</td>
</tr>
<tr>
<td>6</td>
<td>Explore the week 1 page main area and interacting with videos</td>
</tr>
<tr>
<td>7</td>
<td>Explore the week 1 page main area and left sidebar then leave page</td>
</tr>
<tr>
<td>8</td>
<td>Explore the week 1 page main area and left sidebar</td>
</tr>
<tr>
<td>9</td>
<td>Explore the week 1 main area and forum main area</td>
</tr>
<tr>
<td>10</td>
<td>Explore the week 1 page main area</td>
</tr>
<tr>
<td>11</td>
<td>Explore the homepage main area, left sidebar, and interact with videos</td>
</tr>
<tr>
<td>12</td>
<td>Explore the homepage main area and interact with videos</td>
</tr>
<tr>
<td>13</td>
<td>Explore the homepage main area and left sidebar</td>
</tr>
<tr>
<td>14</td>
<td>Explore the homepage main area then leave page</td>
</tr>
<tr>
<td>15</td>
<td>Explore the homepage main area and week 1 page main area, left sidebar, and interact with videos</td>
</tr>
<tr>
<td>16</td>
<td>Explore the homepage main area then week 1 page main area and interact with videos</td>
</tr>
<tr>
<td>17</td>
<td>Explore the homepage main area then week 1 page main area and left sidebar</td>
</tr>
<tr>
<td>18</td>
<td>Explore the homepage main area and left sidebar then week 1 page main area</td>
</tr>
<tr>
<td>19</td>
<td>Explore the homepage main area and left sidebar then forum main area</td>
</tr>
<tr>
<td>20</td>
<td>Explore the homepage main area</td>
</tr>
<tr>
<td>21</td>
<td>Explore the forum main area</td>
</tr>
<tr>
<td>22</td>
<td>Explore the forum main area and left sidebar</td>
</tr>
<tr>
<td>23</td>
<td>Explore the forum main area</td>
</tr>
</tbody>
</table>

### Table 4.5. Behaviours and corresponding features correlated with achievement, where correlation coefficients $| \rho, \tau > 0.45|$ and $p < 0.001$.  

<table>
<thead>
<tr>
<th>Id</th>
<th>Interactive behaviour</th>
<th>Correlated feature</th>
<th>Spearman's $\rho$</th>
<th>Kendall's $\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_1$</td>
<td>Explore the week 1 page main area, left sidebar, and forum main area</td>
<td>NumGap</td>
<td>0.55</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Epi_not0</td>
<td>0.47</td>
<td>0.45</td>
</tr>
<tr>
<td>$B_2$</td>
<td>Explore the week 1 page main area then leave page</td>
<td>$S_{10}$ mean</td>
<td>-0.48</td>
<td>-0.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$T_{10}$ mean</td>
<td>-0.50</td>
<td>-0.49</td>
</tr>
<tr>
<td>$B_3$</td>
<td>Explore the week 1 main area and forum main area</td>
<td>NumGap</td>
<td>0.48</td>
<td>0.46</td>
</tr>
<tr>
<td>$B_4$</td>
<td>Explore the homepage main area and interact with videos</td>
<td>NumGap</td>
<td>0.49</td>
<td>0.46</td>
</tr>
<tr>
<td>$B_5$</td>
<td>Explore the forum main area and left sidebar</td>
<td>NumGap</td>
<td>0.57</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sum</td>
<td>0.51</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$P_{10}$ sum</td>
<td>-0.49</td>
<td>-0.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$P_{5}$ sum</td>
<td>-0.49</td>
<td>-0.46</td>
</tr>
<tr>
<td>$B_6$</td>
<td>Explore the forum main area</td>
<td>NumGap</td>
<td>0.61</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Epi_not0</td>
<td>0.58</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$S_{3}$ sum</td>
<td>-0.72</td>
<td>-0.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$S_{5}$ sum</td>
<td>-0.64</td>
<td>-0.63</td>
</tr>
</tbody>
</table>

**Modeling student interaction and analysis** For each of the 23 emergent behaviours, we computed the Spearman and Kendall correlation tests between the features in Table 6.1 and the two indicators of learning we extracted: the badge status and the number of units completed. We found moderate to strong significant correlations $|\rho, \tau > 0.45|$ between the features of six of the behaviours and the badge status as shown in Table 5.7. We found no significant correlations between the behaviours and the number of completed units.

From the features in Table 6.1, Episodes$^1$ (i.e. the number of online sessions in total) correlates positively with whether students received a badge ($\rho = 0.42$, $p < 0.001$), which is in line with literature that suggests the more active a student is, the more likely a badge will

---

$^1$ All of the features are computed for each of the behaviours but Episodes, which is behaviour agnostic.
be awarded [Guo et al., 2014, Singh et al., 2018, Jansen et al., 2009]. The highest positive correlation was yielded by the NumGap feature of $B_6$ ($\rho = 0.61, p < 0.001$), which indicates a relationship between obtaining a badge and the number of continuous sessions where the forum’s main area was not explored. On the other hand, the highest negative correlation was $S3sum$ ($\rho = -0.72, p < 0.001$) for the same behaviour, suggesting a relationship between achievement and a decrease in the exploration of the forum over time.

Table 4.6. Comparison between models including all the features (AllFeatures) and models including engagement features only (EngFeatures)

<table>
<thead>
<tr>
<th></th>
<th>Cox &amp; Snell $R^2$</th>
<th>Nagelkerke $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AllFeatures</td>
<td>EngFeatures</td>
</tr>
<tr>
<td>Best</td>
<td>0.43</td>
<td>0.38</td>
</tr>
<tr>
<td>Average</td>
<td>0.42</td>
<td>0.35</td>
</tr>
<tr>
<td>Median</td>
<td>0.42</td>
<td>0.36</td>
</tr>
</tbody>
</table>

For the six interactive behaviours that yielded coefficients above 0.45 with badge status in Table 5.7, we included in a regression analysis the top two strongly correlated positive/negative features of each behaviour as long as they were not variations of the same feature (e.g. $S3sum$ and $S5sum$ for $B_6$) to prevent overfitting. These were the dependent variables of the regression analysis, while badge status was the dependent variable. As dependent variable is binary, we used logistic regression analysis. The logistic models with 1–6 predictors were computed following the stepwise backward and forward techniques, where Model 4.1 has the highest explainability (81.8%, Nagelkerke $R^2$) about whether a student received a badge:

$$-9.44 + 1.52B_1 - 1.08B_2 - 1.15B_3 - 0.04B_4 - 1.06B_5 + 2.40B_6 \quad (4.1)$$

The goodness-of-fit test (Hosmer–Lemeshow test) yielded a value of 0.246 and was insignificant ($p = 0.993$), suggesting that the model fits the data well. Two descriptive measures of goodness-of-fit presented are Cox and Snell $R^2 = 0.43$ as well as Nagelkerke. These indices are variations of the $R^2$ concept defined for Ordinary Least Squares regression (OLS) model.

According to Model 4.1, the log of the odds of receiving a badge is negatively associated with the exploration and withdrawal of the week 1 page main area ($B_2$, $p = 0.01$), the exploration of the week 1 main area followed by the exploration of the forum ($B_3$, $p = 0.37$), the exploration of the homepage followed by the interaction with videos ($B_4$, $p = 0.96$), and the exploration of the forum main area and left sidebar ($B_5$, $p = 0.36$). Those behaviours that were positively associated with receiving a badge include the exploration of the week 1 page main area, left sidebar, and the forum main area ($B_1$, $p = 0.30$), and the exploration of the forum main area ($B_6$, $p = 0.01$). The statistical significance of individual regression coefficients were tested using the Wald chi-square statistic. According to the results, only $B_2$’s $S10mean$ ($p = 0.01$) and $B_6$’s NumGap ($p = 0.01$) were significant predictors of badge status.
In other words, the more students revisited ($S10_{mean}$) week 1 materials to then leave the learning platform without interacting with any other section of the MOOC ($B_2$), the less likely they were to receive a badge. This could be because this behaviour models those who got stuck in week 1, try to catch up several times but make no further advances in the course. We confirm that the number of consecutive sessions in which the forum was not explored ($\text{NumGap}$ on $B_6$) is a strong predictor of badge status. The odds of students exhibiting $B_2$ ($S10_{mean}$) not receiving a badge were 3, whereas the odds for those exhibiting $B_6$ ($\text{NumGap}$) receiving a badge were 11.

Some of the features included in our previous regression analysis are closely related to engagement, including the number of inactive sessions. We know from previous work that dwell time is related to engagement [Yi et al., 2014, Lu et al., 2019, Guo et al., 2014, Singh et al., 2018]. To investigate the added value of including non-engagement metrics in the models and address RQ3, we computed a baseline model with those features in Table 6.1 that are known indicators of engagement. Similar to the previous analysis, the variable that contributed most was the number of consecutive sessions ($\text{NumGap}$) in which the forum was not explored ($B_6$). The model with the highest explanatory power (Nagelkerke $R^2 = 72.8\%$) for whether a given student receives a badge is:

$$-6.52 + 0.36B_1 - 0.001B_3 - 0.23B_4 - 0.62B_5 + 1.85B_6$$  \hspace{1cm} (4.2)

where only $B_6$ is a significant predictor of achievement ($p = 0.001$). We computed $R^2$ values of the top-20 performing models for both analyses: i.e. the models including all features (AllFeatures) and the models including only engagement features (EngFeatures) —see Table 4.6. Our results indicate that including non-engagement behaviours in the models adds a 9% of explainability in the best case and a 14% on average for the Nagelkerke $R^2$.

Table 4.7. 10-fold cross validation of the original dataset.

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
<th>Accuracy</th>
<th>Accuracy Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>K Nearest Neighbour</td>
<td>0.65</td>
<td>0.60</td>
<td>0.62</td>
<td>0.94</td>
<td>+0.06</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>0.72</td>
<td>0.80</td>
<td>0.73</td>
<td>0.92</td>
<td>+0.04</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.72</td>
<td>0.75</td>
<td>0.69</td>
<td>0.93</td>
<td>+0.05</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.85</td>
<td>0.75</td>
<td>0.78</td>
<td>0.94</td>
<td>+0.06</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.77</td>
<td>0.75</td>
<td>0.73</td>
<td>0.94</td>
<td>+0.06</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>0.90</td>
<td>0.55</td>
<td>0.67</td>
<td>0.94</td>
<td>+0.06</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>0.48</td>
<td>0.50</td>
<td>0.48</td>
<td>0.91</td>
<td>+0.03</td>
</tr>
<tr>
<td>Multi-layer Perception</td>
<td>0.57</td>
<td>0.60</td>
<td>0.56</td>
<td>0.93</td>
<td>+0.05</td>
</tr>
<tr>
<td>Bagging</td>
<td>0.90</td>
<td>0.60</td>
<td>0.70</td>
<td>0.95</td>
<td>+0.07</td>
</tr>
<tr>
<td>Gradient Tree Boosting</td>
<td>0.75</td>
<td>0.60</td>
<td>0.65</td>
<td>0.94</td>
<td>+0.06</td>
</tr>
</tbody>
</table>
We then conducted 10-fold cross validation using a variety of classifiers. As the distribution of students with and without a badge is imbalanced, only 12% (17 out of 140) of the participants received a badge with a 1:7.2 ratio, to avoid losing information in majority examples to undersampling [Haixiang et al., 2016] and overly-optimistic estimates [Santos et al., 2018], we performed oversampling during the cross validation procedure. Synthetic Minority Oversampling Technique (SMOTE) coupled with Tomek Links was performed as the combination of algorithms was suggested to prevent overfitting [Santos et al., 2018]. Stratified k-fold cross validation was used to ensure that the proportion of positive to negative examples is kept in the folds. The results for the 10-fold cross validation on the original dataset and with oversampling are shown in Tables 4.7 and 4.8, where the accuracy change (difference between the accuracy of the model and the default distribution percentage which is the students without a badge out of all the students), precision, recall, and f1 score are tabulated. We focus on the precision and accuracy as the importance of identifying students with tendencies to fail is greater than identifying students who are likely to succeed. For the original dataset (without oversampling), the accuracy of the models identifying failing students is 91-95%, 3-7% higher than using random selection (i.e. the percentage of students without a badge, 88%). The Bagging classifier yields the largest increase in accuracy across the two sets of results (7%), with precision being 0.90.

### Table 4.8. 10-fold cross validation with oversampling.

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
<th>Accuracy</th>
<th>Accuracy Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>K Nearest Neighbour</td>
<td>0.58</td>
<td>0.85</td>
<td>0.66</td>
<td>0.89</td>
<td>+0.01</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>0.57</td>
<td>0.95</td>
<td>0.69</td>
<td>0.89</td>
<td>+0.01</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.78</td>
<td>0.70</td>
<td>0.69</td>
<td>0.93</td>
<td>+0.05</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.75</td>
<td>0.75</td>
<td>0.71</td>
<td>0.92</td>
<td>+0.04</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.73</td>
<td>0.85</td>
<td>0.73</td>
<td>0.91</td>
<td>+0.03</td>
</tr>
<tr>
<td>Support Vector Machines</td>
<td>0.63</td>
<td>0.85</td>
<td>0.71</td>
<td>0.91</td>
<td>+0.03</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>0.72</td>
<td>0.80</td>
<td>0.73</td>
<td>0.94</td>
<td>+0.06</td>
</tr>
<tr>
<td>Multi-layer Perception</td>
<td>0.66</td>
<td>0.85</td>
<td>0.71</td>
<td>0.93</td>
<td>+0.05</td>
</tr>
<tr>
<td>Bagging</td>
<td>0.57</td>
<td>0.90</td>
<td>0.65</td>
<td>0.89</td>
<td>+0.01</td>
</tr>
<tr>
<td>Gradient Tree Boosting</td>
<td>0.62</td>
<td>0.65</td>
<td>0.59</td>
<td>0.92</td>
<td>+0.04</td>
</tr>
</tbody>
</table>
To investigate the occurrences of the behaviours over time, we analysed the occurrences of the 23 behaviours in each session for the students that eventually received a badge and students who did not. The number of the active students for each session was plotted alongside the average of occurrences for each session among the students who received a badge and the students who did not (data had been smoothed using LOESS Curve Fitting). Based on our observations of the visualisations as shown in Figure 4.1, we classified these behaviours into four clusters:

1. Behaviours of both groups evolve similarly over time (Figure 4.1a).
2. Behaviours were exhibited more by students with the badge (Figure 4.1b). Among the six behaviours included in our models, $B_1$, $B_3$, $B_5$, $B_6$ are in this cluster.
3. Behaviours of both groups evolve similarly for most sessions, with significant differences only at specific sessions (Figure 4.1c). Behaviours $B_2$ and $B_4$ are classified in this cluster.
4. Behaviours were exhibited more by students without the badge for most sessions, with significantly higher peak for badge holders in specific sessions (Figure 4.1d).
4.7 Discussion

We revisit the research questions formulated at the outset:

**RQ1. Can we identify micro-interactions that are relevant for online learning?**
We utilised sequential pattern mining and thematic analysis to find implicit regularities that are associated with learning including achievement and unit progress. Our methodology allowed us to investigate micro-interactions in different abstraction levels by grouping fine-grained interaction patterns into higher level behaviours. Our findings suggest that micro-interactions can potentially be used to predict achievement.

**RQ2. Which micro-interactions are associated with learning progress and achievement?**
We confirmed existing findings on an online learning platform and discovered that interactive behaviours involving exploration of forum, and leaving the online platform after the exploration of the week 1 materials are significant predictors of student achievement. Further to this, our findings provide more clarity on these behaviours: first, the less students abandon week 1 materials, the more likely they are to receive a badge; second, the number of consecutive sessions where the forum’s main area was not explored was strongly related to obtaining a badge. While the first behaviour confirms that withdrawal is directly related to the (lack of) achievement, it is unexpected that the less the forum is explored the more likely students are to achieve a badge [Anderson et al., 2014, Crues et al., 2018, Ramesh et al., 2014]. This may be due to the forum being perceived as a collaborative tool within the context of a cMOOC, which was used as a resource for students who were struggling. Hence, not required by advanced students.

The interactive behaviours include regularities such as exploration of the forum and interactions with videos. Forum participation and video interactions have been thoroughly investigated in previous research [Romero et al., 2013, Wang et al., 2015, Atapattu and Falkner, 2017, Kim et al., 2014], perhaps as they are the most representative of a classic classroom structure. The other frequent elements, for example exploration of homepage, week 1 materials, and leaving week 1 materials are within expectation: exploring the week 1 materials and leaving may be interpreted as distraction from the learning material and abandonment of the course, which resulted in low achievement. This is in line with the visualisations in Figure 4.1d illustrating the decrease of interactions over time and previous research indicating that participation tends to drop rapidly within the first week [Evans et al., 2016].

We found that a decrease in exploring the forum is a strong predictor of achievement while the visualisations show that successful students were more active in the forum. The findings may seem contradictory as they indicate that the levels of participation in forums decreased over time for students who received a badge while still remained higher compared to other students throughout the sessions. More forum activities suggests that the student is more likely to receive a badge, which is in line with related studies showing that forum participation
positively affects the learning outcome [Anderson et al., 2014, Crues et al., 2018, Ramesh et al., 2014]. Our findings can be explained in that, in sessions when students were less active in the forum, they were conducting other learning activities in the platform. It may also be the case that as students use the forum to make progress and learn, they depend on the forum less and become more autonomous learners.

We did not find strong relationships with the number of units completed. This may be because obtaining the badge involves a set of diverse activities and it is more informative than student progress. The models can be used to develop, for instance, browser extensions to capture interface interactions in real time and predict student learning trajectory to provide feedback and guidance to students and course leaders.

**RQ3. What is the added value of using micro-interactions?**
Features of engagement (e.g. the number of inactive sessions) are the strongest predictor of achievement according to our analysis. The explainability of models built with engagement features is as high as 72.8%. By comparing the models from two regression analysis containing different features, we discovered a 10% contribution to the explainability of student achievement from features of behaviours other than engagement, and perhaps learnability. When analysing the performance of the models, results indicate a 3-7% increase in accuracy in identifying students who will not receive a badge versus random selection (i.e. the percentage of students who did not receive a badge) suggesting that micro-interactions implicitly capture learning processes. This can be explained in that the relationship between cognitive engagement and learning outcomes is mediated by user activity [Fincham et al., 2019]. Consequently, it can be argued that the user activity in online learning platforms contains behavioural markers that are indicators of cognitive processes and learning outcome.

**4.7.1 The role of user interface learnability**

In general, beyond the decrease in the interactions of successful students with the forum, we discovered that students exhibited decreasing trends of participation throughout the online sessions. The decrease can be expected as engagement typically decreases over time on online learning platforms. The correlation results showed that most of the trends are negatively correlated with student achievement, which means that the less a student exhibits these behaviours, the more likely they are to receive a badge. As the trend features were calculated using windows of 3/5/10 sessions, the trends can be considered as fluctuations, which is confirmed by the visualisations of occurrences, as shown in Figure 4.1a. More fluctuations may mean that students exhibited different interactive behaviours at the same time. As students engage with the platform they become more familiar with the platform, and consequently they become more proficient at navigating and using it. This finding may further support our assumption that interface learnability has an effect on the learning outcome as measured by student achievement. Usability has been investigated in online learning settings to evaluate different platforms and to provide insights on the design of platforms and the learning re-
sources [Xiao et al., 2015, Tsironis et al., 2016]. Although it has been suggested that there is no widely-agreed definition for learnability, it can be described as ‘the ability to improve performance over entire usage history’ based on a taxonomy provided in [Grossman et al., 2009]. Learnability as an important and perhaps fundamental component of usability [Abran et al., 2003, Nielsen, 1994], has not been thoroughly discussed in online learning platforms.

The plausible association between becoming a more efficient user and student achievement may suggest that the easier a learning platform is to use, the better students learn. It may also suggest that students who are more familiar with the learning platform or skilled navigators are better at learning. This opens up new research avenues into analysing the learnability of platforms and student learning outcome.

4.7.2 Implications

**Methodological implications.** Previous research has identified the challenges of using sequential pattern mining algorithms in educational studies [Poon et al., 2017]: 1) the generation of excessive patterns with limited relevancy and value, and 2) the involvement of domain experts for filtering and labeling purposes. For the first challenge, we proposed a benchmarking process to select the most suitable algorithm, dataset and algorithm parametrisation so that we maximise how informative and representative the results are. The second challenge is relatively common in data mining and analysis scenarios. Through a number of established methodological steps, thematic analysis facilitates the work of domain experts.

**Theoretical implications.** We found that micro-interactions can be behavioural markers of online learning. Whether these micro-interactions are universal is unknown but they are probably not generalisable across online platforms. However, they provide further explainability in terms of interpretability and predictability of what students do.

**Implications for practice.** Despite the current demands on the use of rich interaction data [Maldonado-Mahauad et al., 2018], existing data analysis procedures in online learning do not contemplate low-level events. This may be due to the lack of infrastructure to do so although, actually, there is an availability of these tools [Apaolaza and Vigo, 2017] which facilitate low-level data collection. However, computing patterns and building student models require programming and data science expertise. While they are not widespread and their coverage is limited, tools such as WevQuery-PM [Apaolaza and Vigo, 2019] can lighten this burden.

4.7.3 Limitations

As each online learning platform is designed for different purposes and audiences, there may be limitations to the extent of the generalisability of our conclusions, which a common issue
in online learning [Alonso-Mencía et al., 2020]. It is well known that the models defined for MOOCs are typically valid within the scope of the MOOCs under scrutiny: early career researchers on a cMOOC. There are no universal models as students, learning contents, and course designs differ [Motz et al., 2019]. Hence, researchers emphasise on methodological approaches to build student models [Lehmann et al., 2012]. We contribute to such body of knowledge with this novel approach.

Also, compared to other online learning platforms, the sample size is relatively small due to the fact that the platform was targeting an specialised student group, early career researchers. We counterbalance this limitation with the large amount of data points collected, which give us an in-depth understanding of student behaviour.

While students interactions with the platform are used to predict whether they will receive a badge or not, achieving a badge is also rewarded partly for these same interactions. It could be argued, however, that low-level interactions used in analysis contain many more varieties of higher level interactions (such as viewing the video content and wiki page) than those used to give out the badge. Further to this, the criteria for receiving a badge includes activities and assignments which are not captured in these low-level interactions.

### 4.8 Conclusion

We propose a three-step methodology to determine if there are implicit micro-interactions that are associated with indicators of learning. To suggest its feasibility, we apply the methodology in an online learning platform. From the generated dataset, we extracted frequent student interaction patterns using sequential pattern mining algorithms. To handle the sheer numbers of patterns we applied thematic analysis to group the patterns into themes that represent interactive behaviours. Then we sought the presence of these representative behaviours in the original dataset to build student models that represent such behaviours. Finally, we conducted statistical analysis between features derived from the models and indicators of learning including achievement and unit progress.

We identified features from six interactive behaviours that strongly correlated with achievement that explain 72% of the student achievement variation, which was 10% higher when non-engagement features were included. The models with the interactive behaviours were 3-7% more accurate at identifying students who will not receive a badge than random selection. It would have not been possible to obtain the increases in explainability and accuracy without our data-driven approach. This increase involves deeper insights gained from the relationship between student achievement and learnability of the student interface. In summary, our data-driven approach provides self-regulated learning platforms with a fresh perspective on measuring indicators of learning, and has revealed connections between platform/student learnability and learning outcome.
Chapter 5

Low-level activity patterns as indicators of user familiarity with Websites

5.1 Chapter overview

In Chapter 4, we demonstrated our data-driven methodology on an online learning platform. We evaluated our methodology and found added value in investigating low-level user interaction data. There is a limitation for user modelling that is common in that user models generally can not be generalised across domains and platforms. Therefore, we are primarily interested in methodological contributions, we aim to further investigate whether our methodology can be applied to a different platform and domain and generate insights. In this chapter, we then apply the same methodology to the University of Manchester’s School of Computer Science website. In the previous chapter, we found that the interface learnability has an effect on the learning outcome as measured by student achievement, implying that students who are more familiar with the learning platform (or who are skilled navigators) are more adept at learning. Previous research has investigated how user interactions evolve with the increase in user familiar with a Web page [Apaolaza et al., 2015]. In light of this, we continue the investigation of familiarity by examining the use of low-level data to indicate user familiarity with the website. In this chapter we present a secondary analysis of the data used in [Apaolaza et al., 2015]. While this previous study was conducted using a hypothesis-driven approach, we perform a data-driven approach to analyse the same data. We also explored a variety of Web interactions while this previous work only focused on single events.

The term ‘(low-level) activity pattern’ is used in this chapter to refer to user behaviours that are derived from low-level interactions. Appendix B contains supplemental material for the chapter. We include additional information regarding survey results, URL domain selection, and additional thematic analysis results mentioned in Chapter 3.

5.1.1 Author’s contributions

He Yu carried out the literature review, data analysis, reporting and writing up. Markel Vigo provided continuous feedback and advice throughout all the stages of the study. Along with Simon Harper, they offered feedback and support that contributed to the work in this chapter.
5.1.2 Abstract

Familiarity is a quality of user experience that has traditionally been difficult to define, capture, and quantify. Existing works on measuring familiarity with interactive systems have relied on surveys and self-reporting, which is obtrusive and prone to biases. Here, we propose a data-driven methodology to associate low-level activity patterns with familiarity. As a proof-of-concept, this methodology was tested on a website with 35,819 users over the course of 18 months, including 268 revisiting users who had reported their levels of familiarity with the platform. By using activity patterns as features of predictive models, we were able to successfully classify users with higher levels of familiarity with an accuracy of 82.7%. These results suggest that there is a relationship between user familiarity and activity patterns involving the exploration and use of navigational artefacts including breadcrumbs, navigation bars, and sidebar areas. This research opens up further opportunities for unobtrusively analysing user experience on the Web.

5.2 Introduction

The experience of users with technology is strongly affected by their familiarity with it in that novel (and unfamiliar) interactive artefacts require the development of new cognitive mechanisms [Tomasi et al., 2018]. Familiarity can be defined as the level of knowledge or the sense of comfort and confidence with something. Familiarity can be viewed as a complex understanding which is often based on previous interactions, experiences, and learning from others [Luhmann and Morgner, 2019]. Various factors affect familiarity, such as prior experience, repeated exposure, level of processing (e.g. processing of the meaning compared to the typeface of a word), and forgetting rate [Yonelinas, 2002]. Previous work has shown that when using unfamiliar tools that require new skills and patterns of interaction, individuals will typically initially perform poorly [Dennis and Garfield, 2003]. It is expected that familiarity with a website can have significant impact on navigation, browsing, and the overall user experience, as users who are more familiar are less likely to become confused/disoriented while having to traverse through multiple levels of a deep website or resort to trial-and-error to achieve their goals [Chen et al., 2011]. Familiarity, therefore, reduces the cost of browsing, as time spent having to backtrack or seek new paths is minimised [Galletta et al., 2006]. By characterising and measuring user familiarity we can provide website designers with frameworks to optimise user experience and improve user navigation/browsing efficiency.

Measuring familiarity with interactive systems typically require users’ self-reporting and survey participation, which can be intrusive and may be prone to bias (e.g. selection and response bias). While more controlled settings such as laboratory studies can yield more detailed interaction data (e.g. eye movements) and control for confounding factors, the generalisability of these results can be called into question due to the reduced ecological validity of the settings [Obendorf et al., 2007]. Accessing server logs offers an alternative, less intrusive method of measuring user familiarity while sacrificing control over participants.
The use of user interactions to measure aspects of user experience is common practice and tool support has been provided for tracking real-time user experience using streams of user events [Kim et al., 2008]. User interactions and usage patterns have been used to identify navigation-related Web usability problems [Geng and Tian, 2015, Vigo and Harper, 2017]. Likewise, user activity patterns have been investigated to detect usability problems [Tamir et al., 2011]. A series of experiments have also identified those user browsing behaviours—such as time taken and amount of scrolling up a page—that predict navigation difficulty [Thomas, 2014]. Other features of user activity such as clickthrough, time spent on the search result page, and how a user exited a result or ended a search session (exit type/end action) can predict user satisfaction of search tasks [Fox et al., 2005].

Familiarity as a factor that can significantly impact user experience [Chen et al., 2011, Ruddle, 2009], has not been widely discussed. Also, despite the associations between previous interactions and familiarity, researchers have discovered that sense of familiarity can be evoked even without prior experience [Whittlesea, 1993]. Inspired by the above works involving activity patterns and aspects of user experience, we investigate whether low-level activity patterns can be used to indicate user familiarity. We collected the interaction data of 35,819 users over the course of 18 months on a university website. We apply a data-driven methodology that was recently introduced to associate activity patterns with qualities of the user experience [Yu et al., 2021]. We employ self-reported data from 268 revisiting users as the ground truth to measure familiarity.

In this work, we address the following research question (RQ): Are there any low-level activity patterns that are indicators of user familiarity on the Web? If so, which are these patterns? We adopted a data-driven approach to investigate the research question where low-level activity patterns were generated using pattern mining and qualitative analysis techniques. We then investigated the relationship between the activity patterns and the user’s level of familiarity. We applied this on The University of Manchester School of Computer Science website. As this study is conducted in the wild it does not control for browsing topics, tasks or user demographics. Hence the outcomes are relevant to understand if real-world user interactions are associated with the concept of familiarity, and may therefore inform Web designers to better support users and improve user experience and satisfaction.

Our analysis identified 15 activity patterns that were frequently exhibited by users found within multiple (sub)domains on a university website. Our findings are threefold: first, the activity patterns associated with familiarity include typically interactions on website navigation artefacts such as navigation and sidebar, suggesting that these elements are key to increase the familiarity of user with a website. Second, including discovered activity patterns in the classification analysis we can categorise user levels of familiarity with up to 82.7% accuracy. Third, correlations between certain activity patterns and variable representing user level of familiarity suggest that increased usage of the breadcrumb element over time is associated with users experiencing difficulty interacting with the website. Our findings are useful to inform Web designers on navigational elements and user experience monitoring to better support efficient Web navigation and to improve user experience.
5.3 Related work

In this study we used sequential pattern mining to identify interaction patterns that are frequently exhibited by the users. These interaction patterns were categorised into activity patterns later on and used to predict user levels of familiarity with the website. Here we gave brief introduction and review the literature on familiarity, activity pattern, and sequential pattern mining.

5.3.1 Defining familiarity. applications to Web familiarity

There have been several investigations relating to familiarity and its associated factors in the domain of psychology as well as on the Web. While exploring the relationship between familiarity, similarity, and attraction by performing experiments relating to facial recognition, researchers found that the familiarity experienced towards an object could be defined in terms of the frequency of exposure to that object, implying that repeated exposure can increase familiarity [Moreland and Zajonc, 1982]. Indeed, research has linked familiarity with the number of prior interactions, with feelings of familiarity shown to develop from repeated interactions [Whittlesea, 1993, Weinreich et al., 2006]. This is also well-established in information search, with familiarity affecting users’ decisions on whether to read a site’s content or not [Flavián et al., 2006]. For example, users have been shown to return to frequently visited websites twice as quickly as those visited occasionally [Bruce et al., 2004], and research showing that 50-80% of users website traffic involves previously visited pages [Herder, 2005, Tauscher and Greenberg, 1997].

By studying the differences in brain activation based on user familiarity of different websites, researchers were able to show that users’ cognition processed the websites differently depending on their previous exposure, with more familiar websites triggering brain activity in regions activated when users retrieve words or pictures from long-term memory [Neupane et al., 2017]. When using an unfamiliar interactive system (e.g. group support systems) for the first time, team performance (satisfaction, perceived effectiveness, and cohesiveness) can be low because using unfamiliar technology requires the development of new skills and patterns of interaction [Dennis and Garfield, 2003]. The performance (interface efficiency) of users also degrades when they move from an elementary to a more complex interface, however, they tend to perform better after an initial learning period. This varies from user to user, but research shows that several episodes tend to be required before users feel sufficiently comfortable with an initially unfamiliar interface to be able to confidently choose it and recognise the potential benefits [Rosman et al., 2014]. As users gain more experience they begin to gain knowledge about the system, and this familiarity makes the decision-making processes easier [Flavián et al., 2006]. This has a positive impact on user experience, with familiarity shown to have a significant positive moderating effect on the relationship between user satisfaction and loyalty [Kaya et al., 2019]. Although the link between previous interactions/experiences and familiarity has been demonstrated, researchers carrying out experi-
ments involving human memory have revealed that feelings of familiarity can also be aroused in the absence of prior experience [Whittlesea, 1993]. Notably, it has also been shown that the effect that prior experience has on performance (accuracy in information retrieval task) is related to the familiarity of an interface. That is, when an interface is familiar and users are given a close-ended task, experience has a positive effect on performance, and when the task structure is open-ended and the interface is unfamiliar, user experience has a negative effect [Tomasi et al., 2018].

The relationship between familiarity and trust is well established [Zhang et al., 2006b]. Research has shown that, although trust and familiarity are different qualities, trust is significantly affected by familiarity, being described as a building block and a precondition of trust [Gefen, 2000]. In the realm of e-commerce, for example, familiarity between agents in a multi-agent system is known to be an important factor in determining the level of trust [Zhang et al., 2006b]. Being familiar in this case creates a level of expectation based on previous interactions, providing a foundation for future expectations [Gulati, 1995]. Trust, therefore, has always been a focus in the financial field because greater trust is strongly related to better economic outcomes [Zhang et al., 2006b]. It is important to note that how a user’s prior experience was received plays a significant role in how familiarity is associated with trust. This is because prior experience can form the basis of trust, and when the experience was favourable this familiarity can be positive, but when the experience was unfavourable, familiarity can be associated with distrust [Luhmann and Morgner, 2019]. Despite this, researchers using surveys to investigate website continuance intention showed that in some cases familiarity (and intimacy) can be significant enough to overcome negative experiences, with users revisiting a website even following unsatisfactory experiences such as product delay in delivery [Lee and Kwon, 2011]. In choosing information sources, familiarity is likewise a key factor and is correlated with resource selection (e.g. read review journals, attend conferences). Users conducting a thorough search for a research topic cited ‘most familiar’ as the factor for selecting their resource [Quigley et al., 2002], and users rating their familiarity of resources on a five-point scale revealed that they in general selected resources that they were quite familiar with [Xie and Joo, 2009]. Research suggests that users are less motivated to seek out ‘high-tech’ e-learning resources, and instead select those with which they are familiar [Bringman-Rodenbarger and Hortsch, 2020]. Website design can improve familiarity and therefore user experience, for example by using a graphical web interface using a desktop metaphor that is familiar to the users (e.g. folder icons to assist the comprehension of an operating system), researchers noted that user task performance (information retention and recall) improved vs the standard, text-based design [Wells et al., 2005]. Familiar website design may have a significant effect, as navigation issues (e.g. fail to locate desired information) tend to relate to difficulty finding the general region on a site they know contains a piece of information they are searching for, rather than identifying the exact location once this general region has been identified [Ruddle, 2009]. Additionally, if a website is ‘deep’ (i.e. with a hierarchical structure and many levels of traverse), exploiting familiar product names, categories, and using multimedia images can help guide users [Chen et al., 2011]. This can be helpful because becoming disoriented and making several incorrect choices while exploring
levels of an unfamiliar site raises the effort and time of browsing substantially by requiring backtracking [Galletta et al., 2006].

5.3.2 User activity patterns on the Web

User interactions with the Web can be categorised based on the level of abstraction, ranging from low-level physical events (e.g. key press) to high-level task-related events (e.g. completing an assignment) [Hilbert and Redmiles, 2000]. Activity patterns refer to users’ interactions with a Web page and how these interactions are exhibited over time. Researchers have explored these patterns previously for various purposes in different domains. With the aim of improving personalisation of information systems, researchers showed that characterising low-level activity patterns consistently enabled prediction of task types, and suggests that differences in the patterns of user activity are able to indicate aspects of the higher levels of task in which the user is engaged [Cole et al., 2015a]. It has been suggested that both low-level and high-level interactions should be included for analysis, as context may be spread over multiple levels and the composition of these events needs to be taken into consideration to give appropriate interpretations [Hilbert and Redmiles, 2000]. In the context of online learning, when combined with complex data such as self-regulated learning strategies, activity sequence patterns can even be used as predictors for learner success [Maldonado et al., 2018]. When tackling identity theft and account sharing problems, researchers were able to exploit users’ game-play activity patterns such as idle time distribution to identify account hijacking – one of the most serious security problems in online gaming [Chen and Hong, 2007]. By comparing user activity patterns between mobile and non-mobile live streaming, researchers investigating viewing behaviours were able to identify high rates of abandoned viewing sessions for mobile users [Li et al., 2015b].

5.3.3 Sequential pattern mining

In contrast to bag-of-words techniques, where user activities are investigated regardless of the order in which they were performed, sequential pattern mining (SPM) is a data mining technique that identifies common patterns within sequences of data by/while taking into consideration the order in which sequences were performed [van Hoek and Carevic, 2017]. Preserving the temporal order in which user activity is conducted can be a significant advantage of SPM. These algorithms output sequential combinations in the data with occurrences above a set percentage (i.e. minimum support). This can then be used to find the most frequent sequences of events within a given dataset of interaction events [Mooney and Roddick, 2013]. Widely used in retail to help discover customer shopping patterns, SPM algorithms can be used to inform product planning and storage [Aloysius and Binu, 2013]. Under certain circumstances, the sequence of user interaction events can be used as indicators of usability problems [Vigo and Harper, 2017, de Santana and Baranauskas, 2015].

The three types of patterns produced by SPM algorithms are frequent, closed, and maxi-
mal sequential patterns. Patterns with occurrences surpassing a given minimum support are known as frequent patterns. Closed sequential patterns are frequent sequential patterns that are not sub-patterns of other frequent sequential patterns with the same minimum support. Finally, maximal sequential patterns are frequent patterns that are not sub-patterns of other frequent sequential patterns. For example, both pattern A \{a, b\} and pattern B \{a, b, c\} are frequent patterns with their minimum support being 0.6 and 0.5, respectively. Pattern B is the only frequent super-pattern of pattern A. We can conclude that pattern A is a closed pattern in this situation as its only super-pattern, pattern B, has a lower minimum support. However, pattern A is not a maximal pattern as its super-pattern, pattern B, is also a frequent sequential pattern. Pattern B, however, may be a maximal pattern if it has no super-patterns which are frequent. Research has shown that CM-SPADE generally outperforms other algorithms when mining frequent patterns [Fournier Viger et al., 2017], whereas CM-Clasp and CloSPan are overall more suitable algorithms for mining closed sequential patterns, and VMSP performs better for maximal sequential pattern mining [Fournier Viger et al., 2014].

5.4 Study

5.4.1 Data collection

Interaction data was extracted from The University of Manchester School of Computer Science website (see screenshot of the landing page at Figure 5.1). Information about the School of Computer Science can be found on the website and its sub-domains which are reachable via navigation tools from the homepage. The sub-domains include undergraduate pages where users can browse course information, research pages where research groups, facilities, and publications are listed, and ‘about us’ pages where the history of the school, news, and events are presented. Each sub-domain of the website contains a variety of UI elements including text, pictures, drop-down menus that link to other sub-domains, social media (e.g. Twitter) widgets, interactive tiles, etc. The platform was instrumented to generate low-level UI interactive data, including browser window events such as page loads, and mouse and keyboard interactive data. Some of the events contain additional information, such as mouse coordinates for mouse events. We captured a total number of 33 unique interactive events, such as those listed in Table 6.2 using W3vQuery, a tool to collect and query interaction data on the Web [Apaolaza and Vigo, 2017].

Users were identified and tracked via unique and anonymised codes stored in cookies which were deployed the first time they accessed the website. When users first visited the website, there was a 50% chance that a survey would be sent to them at the end of their interactions. If/when the remaining users returned to the site (identified by existing codes), a second survey was sent to them. In total, 392 answers were collected from first-time users, while 272 responses were received from returning visitors. The purpose of the survey was to compare users’ degrees of familiarity at different points in time. Responses to the above questions were reported on a 5-point Likert scale. The questions included in the first survey
Figure 5.1. Screenshot of the landing page of The University of Manchester School of Computer Science website.

The questions included in the second survey which was shown to users on their second visit to the website are:

- How easy is the site to interact with? (1-Not easy at all, 5-Really easy)
- How familiar do you currently feel with the site? (1-Not at all as easy as I expected, 5-Exceeded my expectations)
- Did you think consistency between Web pages across the Internet helped you feel more familiar with the site? (1-Not familiar at all, 5-I am entirely familiar)
- Was the Web site as easy to use as you expected? (1-Didn’t help at all, 5-Really helped)

The questions included in the second survey which was shown to users on their second visit to the website are:

- How easy is the site to interact with? (1-Not easy at all, 5-Really easy)
- How familiar do you currently feel with the site? (1-Not familiar at all, 5-I am entirely familiar)
- Do you think consistency across Web pages helped you feel more familiar with the site? (1-Didn’t help at all, 5-Really helped)
- Do you find the site easy to remember each time you come back? (1-Not easy at all, 5-Really easy)
• Did you find the site easier to interact each time you visited it? (1-Not easier at all, 5-Way easier over time)

A series of data pre-processing steps were required as the low-level interaction data was captured to a great level of detail which may contain noise such as unintended user interactions [Dev and Liu, 2017]. For example, key-up and key-down events were typically seen together, they were combined and renamed as key-press. Key-press events were then transformed based on the type of key pressed (i.e. commands, alphanumerical, other). Same events which occur multiple times sequentially were merged and renamed with the suffix ‘multi’ (e.g. multiple consecutive ‘mousepress’ events were merged with the last occurring event renamed as ‘mousepress+multi’). 8 out of 33 unique UI events remained after the pre-processing: mousepress, mouseinorout, scrollorwheel, blurfocus_leavepage, blurfocus_switchtab, keydown_write, keydown_command, keydown_other. Previous study conducted using similar methodology and low-level interaction data concluded that 40 min of inactivity generated longer sequences exhibited by the largest number of users, therefore, In line with established practices, user sessions were split when there were 40 minutes of inactivity in between two consecutive events [Jones and Klinkner, 2008, Yu et al., 2021].

Table 5.1. Example of captured user interface events.

<table>
<thead>
<tr>
<th>Type</th>
<th>Events</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>mousedown</td>
<td>Start of mouse click action</td>
</tr>
<tr>
<td></td>
<td>mouseup</td>
<td>End of mouse click action</td>
</tr>
<tr>
<td></td>
<td>mousemove</td>
<td>Mouse movement</td>
</tr>
<tr>
<td></td>
<td>mouseover</td>
<td>Hovering into target</td>
</tr>
<tr>
<td></td>
<td>mouseout</td>
<td>Hovering out from target</td>
</tr>
<tr>
<td></td>
<td>doubleclick</td>
<td>Double mouse click</td>
</tr>
<tr>
<td></td>
<td>mousewheel</td>
<td>Mouse wheel interaction</td>
</tr>
<tr>
<td>Selection</td>
<td>select</td>
<td>Selection of page content</td>
</tr>
<tr>
<td></td>
<td>cut</td>
<td>Content cut</td>
</tr>
<tr>
<td></td>
<td>copy</td>
<td>Content copy</td>
</tr>
<tr>
<td></td>
<td>paste</td>
<td>Content paste</td>
</tr>
<tr>
<td>Keyboard</td>
<td>keydown</td>
<td>Start of key press action</td>
</tr>
<tr>
<td></td>
<td>keyup</td>
<td>End of key press action</td>
</tr>
<tr>
<td></td>
<td>keypress</td>
<td>Key press action</td>
</tr>
<tr>
<td>Window</td>
<td>load</td>
<td>Page is loaded</td>
</tr>
<tr>
<td></td>
<td>resize</td>
<td>Browser window is resized</td>
</tr>
<tr>
<td></td>
<td>unload</td>
<td>Window is closed</td>
</tr>
<tr>
<td></td>
<td>windowfocus</td>
<td>Browser tab gains focus</td>
</tr>
<tr>
<td></td>
<td>windowblur</td>
<td>Browser tab loses focus</td>
</tr>
<tr>
<td></td>
<td>scroll</td>
<td>Change of scroll state</td>
</tr>
<tr>
<td>Other</td>
<td>change</td>
<td>Input element state change</td>
</tr>
<tr>
<td></td>
<td>contextmenu</td>
<td>Opening of context menu</td>
</tr>
</tbody>
</table>

5.5 Methodology

A three-step methodology was used to process interaction data, generate sequential patterns, and identify activity patterns of interest as demonstrated in previous work [Yu et al., 2021]. We implemented SPM algorithms and extracted the sequential patterns that satisfy a selected minimum frequency threshold. We then applied thematic analysis to interpret and group the extracted patterns, transforming low-level interactions into higher-level activities.
Interaction patterns that fall under the emerging themes are then used to build classification models.

5.5.1 Data pre-processing and setup

Each recorded event in the interaction data is associated with a timestamp, URL of the Web page where the event takes place, and the specific UI element (IDs) that triggers the event. As there are a large number of unique URLs (10,388) and some URLs expired or lead to internal Web pages and documents, to ensure the interpretability of the interaction sequences, we extracted the sub-domain’s topic: i.e. undergraduate, postgraduate research, staff, etc. The criteria to select the URL domains to be included in our analysis were as follows: we computed the occurrences of the URL domains in the interaction data within each percentile. The URL domains with larger number of occurrences (i.e. each event triggers an occurrence) were added to buckets where the user coverage (i.e. number of users who visited the domain) of each group were calculated. We selected the URL domains by balancing the occurrences of the domains in the group and the added user coverage. We had 1780 unique IDs (UI elements) initially, a large number which is noisy and may contain unintended interface events. We selected sets of IDs to include for each URL domain, again using user coverage as the criterion to ensure interpretability and that they are representative of user interactions. After the selection of URL domains and IDs, a event-ID-URL triple represented each event. For example, mouseinorout+main+undergraduate+multi indicates multiple mouse movements in the main area of the undergraduate page.

Sequential pattern mining was performed on the processed interaction data to extract user interaction patterns. To find representative patterns we maximised the number of patterns generated by a higher number of users. The inputs of SPM algorithms were sequences constructed with a single user’s interaction events across all active sessions. To assess how representative a set of patterns were, we explored the value space of the minimum support parameter (i.e. the percentage of users that exhibit a given sequence) in the 0.03–0.1 range, the range is selected based on the increase in the number of sequence along with the decrease of minimum support, as values above 0.1 yielded an extremely small number of patterns and the number of patterns increases dramatically when the support decreases below 0.03. To maximise the amount of information extracted, we sought longer patterns as they have more events and thus more information. Single-event sequences were therefore excluded.

5.5.2 Thematic analysis

Thematic analysis is a widely used analytical method for analysing and generating meaning across, typically, qualitative data [Braun and Clarke, 2006]. Although thematic analysis has been traditionally used to evaluate qualitative data such as survey/interview responses, here we employ thematic analysis to systematically extract themes from patterns produced by the SPM algorithms in the previous stage. We adopted the most common form of thematic
Table 5.2: Example of theme generation process. The first column gives examples of the initial codes (i.e. sequential patterns). The second column shows the sub-themes generated from initial codes. The third column shows the final themes (i.e. activity patterns) generated from the sub-themes.

<table>
<thead>
<tr>
<th>Codes</th>
<th>Sub-themes</th>
<th>Final themes</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>mousemove+courseprofile+undergraduate</code></td>
<td>multiple mouse movement at undergraduate page course profile area</td>
<td>Explore the undergraduate page course profile area with scrolling</td>
</tr>
<tr>
<td><code>mousemove</code>+<code>courseprofile</code>+<code>undergraduate</code></td>
<td>multiple mouse movement at undergraduate page course profile area</td>
<td>Explore the undergraduate page course profile area</td>
</tr>
<tr>
<td><code>scrollorwheel+courseprofile+undergraduate</code></td>
<td>multiple mouse movement at undergraduate page course profile area</td>
<td>Explore the undergraduate page course profile area</td>
</tr>
<tr>
<td><code>mouseinorout+MainBody+homepage</code>+<code>multi</code></td>
<td>multiple mouse movement at homepage main area</td>
<td>Explore the main home page main navigation area</td>
</tr>
<tr>
<td><code>mouseinorout+MainNavigation+homepage</code>+<code>multi</code></td>
<td>multiple mouse movement at homepage navigation area</td>
<td>Explore the homepage navigation area</td>
</tr>
<tr>
<td><code>mouseinorout+Sidebar+undergraduate</code>+<code>multi</code></td>
<td>mouse movement at undergraduate page sidebar area</td>
<td>Explore the undergraduate page sidebar area</td>
</tr>
<tr>
<td><code>mouseinorout+MainBreadcrumbs+undergraduate</code>+<code>multi</code></td>
<td>multiple mouse movement at undergraduate page breadcrumbs area</td>
<td>Explore the undergraduate page breadcrumbs, navigation, and sidebar area</td>
</tr>
</tbody>
</table>

Analysis: a six-step process described by Braun and Clarke [Braun and Clarke, 2006] which includes familiarising with the data, coding where we considered the sequential patterns generated as initial codes, and generating the themes. We conducted an inductive approach to determine themes as the data-driven nature of this research. We first followed the semantic approach to transform each code into a sub-theme as a sentence describing the explicit interaction. After the generation of sub-themes, we followed the latent approach with assumptions about the underlying activity patterns and determined the final themes. Each theme was created with a combination and transformation of the sub-themes. Examples of this process can be viewed in Table 5.2 where the codes are results from the sequential pattern mining (i.e. patterns formed by low-level interaction events) and the final themes are the activity patterns exhibited by the users. The interpretation of the patterns reveals the (navigation) activities conducted by users such as exploring the undergraduate page course profiles. A second coder was not required in the case of this study as sub-theme and theme analysis was conducted under a set of agreed rules: for example, `load+windowfocus` was defined as ‘explore’ rather than ‘load’, ‘focus’, or ‘view’, which reduced ambiguity and leaves little room for alternative interpretations.

5.5.3 Modeling user interaction

To investigate our research question, the activity patterns identified through thematic analysis and their corresponding interaction patterns were sought in the original dataset for occurrences of each pattern within each session for each user. These occurrences were represented in $m \times n$ matrices, where $m$ is the number of users and $n$ the number of sessions exhibited by the user.

For instance, the matrix for activity pattern $P$ can be represented as follows:

$$P = \begin{bmatrix}
    f_{11} & f_{12} & \ldots & f_{1n} \\
    f_{21} & f_{22} & \ldots & f_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    f_{m1} & f_{m2} & \ldots & f_{mn}
\end{bmatrix}$$

where $f_{ij}$ is the frequency or number of occurrences of activity pattern $P$ for user $i$ in Session $j$. Since not all users have the same number of sessions, this is not a square matrix and $f_{ij} = NA$.
We generated descriptive statistics of the frequencies, including the mean, median, and sum of occurrences, to determine which activity patterns were associated with familiarity. We also included the number of ‘inactive’ sessions (i.e. $f_{ij} = 0$) and the number of consecutive inactive sessions. We additionally computed the frequency trend to highlight the evolution of activity patterns and to demonstrate the overall direction in which the number of occurrences was evolving. The mean, median, and sum of the occurrences across 3, 5, and 10 neighbouring sessions were used to compute the trend. This takes into account the estimated trend’s reliability for a relatively sparse dataset with multiple periods of inactivity. The trend is calculated using two methods: the coefficient of its correlation with the growing session number, and the slope in a polynomial function with the horizontal line representing the increase of session number [Oliphant, 2006]. The trend was categorised into six degrees of intensity, ranging from a strong negative trend to a strong positive trend, based on a combination of the aforementioned coefficient and slope. We generated a total of 46 features for each activity pattern, which are listed in Table 6.1.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>The average of occurrences across all sessions</td>
</tr>
<tr>
<td>Sum</td>
<td>The sum of occurrences across all sessions</td>
</tr>
<tr>
<td>NumGap</td>
<td>Number of consecutive inactivity sessions</td>
</tr>
<tr>
<td>MaxGap</td>
<td>Maximum number of consecutive inactive sessions</td>
</tr>
<tr>
<td>Inactiv</td>
<td>Number of inactive sessions in total</td>
</tr>
<tr>
<td>AvInactiv</td>
<td>Proportion of inactive sessions in all sessions</td>
</tr>
<tr>
<td>Epi_not0</td>
<td>Number of active sessions</td>
</tr>
<tr>
<td>Episodes</td>
<td>Number of online sessions in total</td>
</tr>
<tr>
<td>Average_not0</td>
<td>Average of occurrences across all active sessions</td>
</tr>
<tr>
<td>Median</td>
<td>Median of occurrences across all sessions</td>
</tr>
<tr>
<td>T+3/5/10+mean/median/sum</td>
<td>Trend calculated as coefficient with the mean/median/sum of 3/5/10 surrounding sessions</td>
</tr>
<tr>
<td>P1+3/5/10+mean/median/sum</td>
<td>Trend calculated as slope with the mean/median/sum of 3/5/10 surrounding sessions</td>
</tr>
<tr>
<td>S+3/5/10+mean/median/sum</td>
<td>Trend strength of T3/5/10</td>
</tr>
<tr>
<td>Sp+3/5/10+mean/median/sum</td>
<td>Trend strength of p3/5/10</td>
</tr>
</tbody>
</table>

5.6 Results

URL domain and ID selection  Following sub-domain categorisation of the URLs, number of unique URLs were reduced from 10393 to 9750, the URL domains comprised the interactions of 96.6% of users. To determine the appropriate set of URL domains to include in subsequent analysis, the set was built by monitoring the change in overall user coverage following the addition of different URL domains. We noted that when adding URL domains that occur in less than 10% of data to the selection, the increase in user coverage began to stagnate. Following this process, we selected seven URL domains: ‘homepage’, ‘about us’, ‘professional development’, ‘postgraduate taught’, ‘postgraduate research’, ‘our research’, ‘undergraduate’, from which 83.7% of the interaction data remained, comprising the interactions 87% of users. The
same method was used to select sets of IDs to include for each URL domain. After filtering the IDs, 98.4% of the interaction data remained. In total, 18 unique IDs out of 1728 were kept across the seven URL domains, including ‘Twitter widget’, ‘graphic tiles’, ‘research staff profile’, ‘breadcrumbs’, ‘sidebar’, etc.

Sequential pattern mining  Following sequential pattern mining, we selected the support count by measuring the point at which minimum support and sequence length were both maximised – found to be 0.068 for minimum support, resulting in a total number of 130 patterns with a sequence length median of 6. The minimum support indicates that each of the 130 patterns was exhibited by at least 6.8% of the users (i.e. 2 436 users).

Although patterns could be identified in all seven URL domains, we found them most frequently in the undergraduate domain (75%), which suggest that the information about undergraduate courses was the most popular. Cross-domain patterns were not identified and the patterns only consisted of mouse-related events and scrolling. IDs found within the patterns included ‘sidebar’, ‘breadcrumbs’, ‘course profile’, ‘main area’, etc.

Thematic analysis of the patterns  Following thematic analysis, we initially discovered 24 themes, with 14 discovered on the undergraduate page, as shown in Table 5.4, including the exploration of breadcrumbs, navigation, sidebar, and course profile areas. Exploration of sidebar, breadcrumbs, and navigation area were involved in six, five, and seven of the 24 activity patterns, respectively. We also found that scrolling was present in three of the activity patterns, appearing in undergraduate and postgraduate research domains, and in main area and course profile area.

Modeling user interaction  The responses of 268 users to the survey on familiarity were used as the ground truth. These responses belonged to the second survey which was completed by those who had revisited the website as they had a chance to become familiar with it. When we compare the answers to the question ‘How familiar do you currently feel with the site?’ of the first and second survey, a Wilcoxon test shows that the familiarity of revisiting users is significantly higher than first-time visitors (W = 7621, Z = -2.06, p = 0.01954). Table 5.5 shows the features we derived from the survey in order to represent familiarity.

The responses for the five questions all average around 4, with medians at 4. The responses are heavily imbalanced as the majority of users reported higher levels of familiarity as can be seen in Figure 5.2 and Table B.1.

After thematic analysis, the matrices modelling user interaction were populated with the corresponding occurrences of each pattern within each session for each user. On average, users had 15 sessions (median: 4.0, SD: 54.8), and 85% of users had 15 sessions or fewer, revealing that there is a skewed distribution in that a few users revisited the site many more times than average users.
Table 5.4. Activity patterns generated from thematic analysis.

<table>
<thead>
<tr>
<th>Index</th>
<th>Activity patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Explore the about us page</td>
</tr>
<tr>
<td>2</td>
<td>Explore the homepage</td>
</tr>
<tr>
<td>3</td>
<td>Explore the homepage navigation area</td>
</tr>
<tr>
<td>4</td>
<td>Explore the our research page</td>
</tr>
<tr>
<td>5</td>
<td>Explore the postgraduate research page</td>
</tr>
<tr>
<td>6</td>
<td>Explore the postgraduate research page breadcrumbs and navigation area</td>
</tr>
<tr>
<td>7</td>
<td>Explore the postgraduate research page sidebar area</td>
</tr>
<tr>
<td>8</td>
<td>Explore the postgraduate research page with scrolling</td>
</tr>
<tr>
<td>9</td>
<td>Explore the postgraduate taught page</td>
</tr>
<tr>
<td>10</td>
<td>Explore the professional development page</td>
</tr>
<tr>
<td>11</td>
<td>Explore the undergraduate page</td>
</tr>
<tr>
<td>12</td>
<td>Explore the undergraduate page and sidebar area</td>
</tr>
<tr>
<td>13</td>
<td>Explore the undergraduate page and title area</td>
</tr>
<tr>
<td>14</td>
<td>Explore the undergraduate page breadcrumbs and navigation area</td>
</tr>
<tr>
<td>15</td>
<td>Explore the undergraduate page breadcrumbs and sidebar area</td>
</tr>
<tr>
<td>16</td>
<td>Explore the undergraduate page breadcrumbs area</td>
</tr>
<tr>
<td>17</td>
<td>Explore the undergraduate page breadcrumbs, navigation, and sidebar area</td>
</tr>
<tr>
<td>18</td>
<td>Explore the undergraduate page course profile area</td>
</tr>
<tr>
<td>19</td>
<td>Explore the undergraduate page course profile area with scrolling</td>
</tr>
<tr>
<td>20</td>
<td>Explore the undergraduate page navigation and sidebar area</td>
</tr>
<tr>
<td>21</td>
<td>Explore the undergraduate page navigation and title area</td>
</tr>
<tr>
<td>22</td>
<td>Explore the undergraduate page navigation area</td>
</tr>
<tr>
<td>23</td>
<td>Explore the undergraduate page sidebar area</td>
</tr>
<tr>
<td>24</td>
<td>Explore the undergraduate page with scrolling</td>
</tr>
</tbody>
</table>

Table 5.5. Dependant variables constructed using survey responses

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1/Q2/Q3/Q4/Q5</td>
<td>The responses for each question on a 5-point Likert scale (1-5 range)</td>
</tr>
<tr>
<td>averaged</td>
<td>The average of points for the five questions (5-25 range)</td>
</tr>
<tr>
<td>max</td>
<td>The maximum of points for the five questions (1-5 range)</td>
</tr>
<tr>
<td>median</td>
<td>The median of points for the five questions (1-5 range)</td>
</tr>
<tr>
<td>Above1/Below2/Above3</td>
<td>Number of responses with points above 1/2/3 (0-5 range)</td>
</tr>
<tr>
<td>Q1_0/Q2_0/Q3_0/Q4_0/Q5_0</td>
<td>Binary representation of each response with critical value at 2/3/4 (i.e. 0 if ≤ critical value, 1 if &gt; critical value), for instance, a response of 3 will be represented as 1 if critical value equals 2, or represented as 0 if critical value is set at 4</td>
</tr>
</tbody>
</table>

For each of the 24 emergent themes, we computed the Spearman correlation between the features in Table 6.1 and the indicators of familiarity we constructed in Table 5.5. We were able to identify weak correlations $|\rho > 0.20|$ between the features from four of the themes and the Q5 and Above2, as shown in Table 5.7.

To prevent overfitting, the above-mentioned features of each activity pattern were included in a classification analysis so long as they were not variations of the same feature (e.g. T10mean and T5mean for $P_1$). These were the independent variables of the classification analysis, while Q5 and Above2 were the dependent variables. We applied Random Forest, Multi-layer Perceptron (MLP), K-Nearest Neighbour, and Support Vector Classification algorithms, and performed a grid search for hyperparameter tuning for each classifier, as shown in Table 5.8. The model using MLP features T3mean, T3median, S3mean, p110mean for the four themes respectively, and dependant variable Above2 achieved an accuracy of 82.7%. The weighted precision, recall, and f1 scores were 0.76, 0.83, and 0.79. Due to the imbalanced nature of the survey responses, our model is better at classifying higher levels of familiarity than lower ones.

We then conducted 10-fold cross-validation using various classifiers. As the distribution
of high and low levels of familiarity was imbalanced (63% of users reported having high levels of familiarity), to avoid losing information to undersampling [Haixiang et al., 2016] and overly-optimistic estimates [Santos et al., 2018], we performed oversampling during the cross-validation procedure. Synthetic Minority Oversampling Technique (SMOTE) coupled with Tomek Links was performed as this combination of algorithms has been suggested previously for preventing overfitting [Santos et al., 2018]. Stratified k-fold cross-validation was used to ensure that the proportion of positive to negative examples was kept in the folds. The accuracy decreased significantly after resampling (from above 80% to 40%). Precision dropped from 0.76 to 0.62 and 0.65 while recall dropped from 0.83 to 0.40 and 0.40 respectively when using oversampling and SMOTE.

5.7 Discussion

To answer the main research question: ‘Are there any low-level activity patterns that are indicators of user familiarity on the Web?’ we implemented a methodology involving sequential pattern mining and thematic analysis to identify low-level activity patterns that may indicate user familiarity. This methodology enabled the investigation of user interactions in

---

**Table 5.6. Number of responses for the survey questions at each point**

<table>
<thead>
<tr>
<th></th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>89</td>
<td>52</td>
<td>105</td>
<td>88</td>
<td>73</td>
</tr>
<tr>
<td>4</td>
<td>114</td>
<td>117</td>
<td>99</td>
<td>108</td>
<td>102</td>
</tr>
<tr>
<td>3</td>
<td>46</td>
<td>70</td>
<td>51</td>
<td>51</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>20</td>
<td>11</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>9</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

---

**Table 5.7. Activity patterns and corresponding features correlated with measured variables**

<table>
<thead>
<tr>
<th>Id</th>
<th>Activity pattern</th>
<th>Correlated feature</th>
<th>Correlated variable</th>
<th>Spearman’s ( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Explore the undergraduate page</td>
<td>T3mean</td>
<td>Above2</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S3mean</td>
<td></td>
<td>0.21</td>
</tr>
<tr>
<td>2</td>
<td>Explore the homepage</td>
<td>S3mean</td>
<td>Above2</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3mean</td>
<td></td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>Explore the undergraduate page breadcrumbs, navigation, and sidebar area</td>
<td>T3median</td>
<td>Q5</td>
<td>-0.20</td>
</tr>
<tr>
<td>4</td>
<td>Explore the undergraduate page navigation and sidebar area</td>
<td>pl10mean</td>
<td>Q5</td>
<td>0.23</td>
</tr>
</tbody>
</table>
Table 5.8. Classification results with different classifiers

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
<th>Weighted precision</th>
<th>Weighted recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-layer Perceptron</td>
<td>82.7%</td>
<td>0.76</td>
<td>0.83</td>
<td>0.79</td>
</tr>
<tr>
<td>Random Forest</td>
<td>71.2%</td>
<td>0.69</td>
<td>0.73</td>
<td>0.71</td>
</tr>
<tr>
<td>K-Nearest Neighbour</td>
<td>63.5%</td>
<td>0.59</td>
<td>0.66</td>
<td>0.61</td>
</tr>
<tr>
<td>Support Vector Classification</td>
<td>65.2%</td>
<td>0.60</td>
<td>0.67</td>
<td>0.64</td>
</tr>
</tbody>
</table>

different abstraction levels by grouping fine-grained interaction events (i.e. mouse clicks, scrolling) into higher-level activity patterns (i.e. explore the undergraduate page and sidebar area). Through classification analysis, we discovered models which showed as high as 82.7% accuracy in predicting user levels of familiarity.

Previous research has linked familiarity with the number of prior interactions, with levels of familiarity shown to increase from repeated exposure [Whittlesea, 1993, Weinreich et al., 2006, Moreland and Zajonc, 1982, Yonelinas, 2002]. By analysing the survey responses, we similarly noted higher levels of familiarity for revisiting users compared to first-time visitors. When surveyed, 63% of revisiting users reported having high levels of familiarity when asked ‘How familiar do you currently feel with the site?’, compared to 49% of first-time users. The connection between higher levels of reported familiarity and revisitation may be explained by the concept of isomorphism. Users may identify different Web browsing tasks as isomorphic problems (i.e. different problems that share the same structure) and apply similar strategies to solve them, therefore, the sense of familiarity increases.

A set of URL domains and IDs were first selected to include in the subsequent analysis. The seven URL domains selected were ‘homepage’, ‘about us’, ‘professional development’, ‘postgraduate taught’, ‘postgraduate research’, ‘our research’, and ‘undergraduate’, and this set represented 87% of user activity, suggesting that although the targeted audiences are varied, most interactions fall within the same domains.

Following sequential pattern mining, the support count was then selected by measuring the point at which the minimum support and sequence length were both maximised –found to be 0.068– which means that at least 6.8% of the users exhibited these patterns at least once. Although this proportion is small, the sequential pattern mining was sufficient to capture at least the interactions of 2 436 users. We were also able to identify patterns in all seven domains, indicating at least 2 436 of these users exhibited the same interaction pattern in each of the domains, although 75% of the patterns were found in the undergraduate domain, suggesting that users who visit the undergraduate domains often display a variety of common patterns. The reason for this may be that users who visit the undergraduate domain (i.e. current or prospective students) share similar exploration goals (e.g. information on degree programs or courses). We did not identify any cross-domain patterns, which may be expected as the URL domains are specifically targeting different user groups (e.g. postgraduate, undergraduate students, research groups), and as the website is non-linear users are not required to follow set paths and can reach different domains when exploring. These patterns also only consist of mouse-related events and scrolling, which suggests a variety of navigation tools were available to explore the site, with users not required to perform keyboard events to search for
As with interaction patterns, more than half of the themes we discovered were in the undergraduate domain. The themes mostly involved the following elements: navigation, breadcrumbs, sidebar area, which all served as navigational elements, suggesting that navigational activity patterns are commonly exhibited by the users. These elements also appear in multiple themes in different combinations, which indicates that users specific navigational strategies are different. For example, some users may only use the sidebar to navigate while other users used a combination of all three.

Following thematic analysis, the matrices were constructed by the corresponding occurrences of each pattern within each session for each user. The user matrices constructed were sparse as the URL domains were targeting specific user groups, and combined with the URL domain and IDs creating a large variety of events, this may explain the weak correlation observed. Two activity patterns show weak correlations with the `Number of responses with points above 2' variable, and two activity patterns show weak correlations with Q5 - ‘Did you find the site easier to interact each time you visited it?’ These correlations with Q5 may corroborate existing findings that familiarity has a significant impact on user experience and that users are less likely to be disoriented and thus reduce the time spent looking for information [Galletta et al., 2006, Chen et al., 2011]. We note that the correlated features were all those that related to trend or trend strength which may suggest that including such features may compensate for having sparse datasets and, consequently, sparse matrices. We also note that the trend of activity pattern ‘Explore the undergraduate page breadcrumbs, navigation, and sidebar area’ has a weak negative correlation with Q5, as shown in Table 5.7, suggesting that the steeper the upward trend in exhibiting this pattern correlates to users finding the site, the less easy to interact it becomes at every revisit. More frequent use of specific navigation tools over time may suggest that the user is struggling to navigate the website, hence finding the website difficult to interact with. Alternatively, the trend/trend strength of activity pattern ‘Explore the undergraduate page navigation and sidebar area’ has a weak positive correlation with Q5, which suggests that the steeper the upward trend in exhibiting this pattern correlates to users feeling easier to interact with the site each time they visit. With the only difference between the two aforementioned patterns being the breadcrumb element which is an indicator of backtracking behaviour, the correlation suggests that users who interact with the breadcrumb element more frequently overtime may be disoriented/confused and thus may not be familiar with the site [Galletta et al., 2006, Chen et al., 2011].

Breadcrumbs are a textual depiction of a website’s hierarchical structure that have been widely used as a navigational aid in hypertext systems. This depiction of the website hierarchies allows users to access other levels within the website structure along a pathway in sequential order. In general, breadcrumbs are used to inform the user as to the location they are currently situated within the website, and to offer quick access for users to navigate directly to previous Web pages in the pathway without using other navigation tools. More importantly, breadcrumbs can assist users to backtrack to the appropriate level within the hierarchy by displaying their current location and the pathway they followed. This may be particularly helpful...
in error recovery. Research has suggested that breadcrumbs are most beneficial for large websites with many hierarchies [Teng, 2005]. If users become disoriented/lost, they will be able to use the breadcrumb to reach a ‘higher ground’, regroup, and resume their tasks [Mukherjea and Foley, 1995, Park and Kim, 2000, Blustein et al., 2005].

Previous studies have suggested that the availability of breadcrumb navigation results in more efficient site navigation and satisfaction [Maldonado and Jlesnick, 2002, Teng, 2005]. This, however, may contradict our finding that increased usage of the breadcrumb element over time is associated with users experiencing difficulty interacting with the website. Previous research has also reached the conclusion that regular breadcrumbs users were not more efficient than non-breadcrumbs users, and that overall breadcrumbs usage can be fairly low [Lida et al., 2003]. This may be explained as breadcrumbs being used as a navigation mechanisms when users felt lost, and that the increased usage of breadcrumb navigation may indicate poor usability of the site. Another study that investigated the impact of varying degrees of breadcrumb tutorial exposure on navigation efficiency found that brief breadcrumb training resulted in more efficient navigation, and that including breadcrumbs reduced the number of clicks required by users to complete tasks [Rogers and Chaparro, 2003]. Our discovery can perhaps also be explained by the fact that certain users are unfamiliar with breadcrumbs and their functionalities, therefore, the usage of breadcrumbs did not improve their navigation efficiency.

The activity patterns ‘Explore the undergraduate page’ and ‘Explore the homepage’ are both weakly correlated with the variable ‘Number of responses with points above 2’, which means that the steeper the upward trend in exhibiting these two activity patterns suggests that the user will have medium/high levels of familiarity. Since the activity patterns are not suggesting the use of any navigation functions, this may be explained as these users already being familiar with the website and therefore not needing much exploration to find the relevant information. Although, when we compare the activity pattern ‘Explore the undergraduate page’ and ‘Explore the undergraduate page navigation and sidebar area’, they both show weak positive correlations with measures of familiarity, which may suggest that (current or prospective) undergraduate students tend to be more familiar with the Web page than other user groups.

We then included features of each activity pattern with Q5 and Above2 as the dependent variables in classification analysis. Following attempts with multiple combinations of features and target variables, we reached an overall 82.7% accuracy using MLP algorithm with features T3mean, T3median,S3mean, p110mean for the four themes respectively and target variable ‘number of responses above 2 points’. As the survey results were imbalanced, our models are better at predicting high levels of accuracy than low levels. If we consider these survey results as a representation of the general users’ levels of familiarity, we can assume that a much larger proportion of users have a high level of familiarity, which may mean that there may be more effective techniques to capture users who have lower level of familiarity as sequential pattern mining (which aims to find commonly exhibited interaction patterns by large groups of users). This affects our model’s ability to identify users who are not fami-
iar with the website. Previous research has suggested that familiarity has significant impact on user navigation, browsing, and user performance [Dennis and Garfield, 2003, Chen et al., 2011]. Therefore, it is likely that users who possess high level and low level of familiarity may exhibit different interaction patterns, which further indicates that sequential pattern mining may not be effective as the algorithm does not capture uncommon patterns exhibited by users who are unfamiliar. Applying sequential pattern mining to each user group and analysing the common patterns can be a solution: patterns exhibited by users with high and low levels of familiarity can be identified separately and analysed together. Alternatively, anomaly detection techniques may be considered to identify distinct interaction patterns of minority user groups.

5.8 Conclusion

We investigated low-level activity patterns and their associations with user familiarity by applying a three-step methodology on a university website. 130 patterns each exhibited by at least 2,436 users were first identified using sequential pattern mining then we interpreted and grouped the patterns into 24 activity patterns using thematic analysis. A classification analysis was conducted between features of these activity patterns and measures of familiarity with survey responses.

We identified features from four activity patterns that demonstrate weak correlations with user-reported levels of familiarity. The classification model including the four activity patterns exhibited an accuracy of 82.7% at identifying levels of familiarity. Although our model was less effective at identifying users with low-level of familiarity, our analysis provides further insights into generalisability and highlights the limitations of the utilised data-driven methodology.

5.8.1 Limitations

As web interface design, user individuality, and domain vary, our studies have limitations in terms of the extent to which our conclusions are generalisable, which is a common issue in user modelling research [Alonso-Mencía et al., 2020] that conclusions drawn from one platform may not apply to others. Hence, researchers emphasise on methodological approaches to build student models [Lehmann et al., 2012]. We contribute to such body of knowledge with this novel approach.
Chapter 6

Modelling user search skill evolution on a specialist search engine

6.1 Chapter overview

In the two preceding chapters, we examined user interaction data in order to predict learning outcome and familiarity; we demonstrated how low-level user interaction data could be used in practice and discussed its benefits and drawbacks. In chapter 4, we conducted a brief exploration of user behaviour over time, we analysed them by visualisations, and we characterised the user behaviours based on them. The benefits of studying user behaviour over time has been demonstrated in chapters 4 and 5, where we also discovered that the features related to trends of user behaviour occurrences over time are mainly the features with stronger correlations with the predicted variables. This indicates that by taking into account the evolution of user behaviours, we can better associate these behaviours with online learning outcomes and user Web familiarity. In this chapter, we intend to continue our exploration of user behaviour over time, with a focus on search behaviours on an academic search engine. We are particularly interested in identifying users groups based on their search behaviours. In this chapter, we also aim to contribute to platforms, including most search engine providers, that are collecting search data but have no ground truth to indicate user search outcomes. Through our methodology, we can identify users who are more skilled at searching, and users who might be struggling to locate information, while benefiting from high ecological validity. We aim to provide web search engine providers information on how to use this type of data, be informed about platform design, and be able to better support the users.

Appendix C contains the chapter’s supplementary material. In Appendix C, we include additional information regarding survey and data analysis results unreported in the chapter.

6.1.1 Author’s contributions

He Yu carried out the literature review, data analysis, reporting and writing up. Markel Vigo provided continuous feedback and advice throughout all the stages of the study. Along with Simon Harper, they offered feedback and support that contributed to the work in this chapter.
6.1.2 Abstract

Despite the advances in online search tools and user interfaces, finding desired information can still be a complex task. This is particularly problematic for specialist search engines, which are non-generalist tools to search for information on domain-specific knowledge bases widely used by librarians, researchers, and life-scientists. To better understand the learning process on a new specialist search engine, we propose a methodology to monitor search behaviour evolution. As a case study, we analysed the low-level interaction data of 239 users on an academic search engine for 20 months. We modelled search behaviours using features derived from search queries as well as user interface interactions, and following clustering, we characterised users based on their search and exploration behaviours. We analysed the transitions between clusters over time to depict how search behaviour evolution manifests. Our method enabled us to identify individuals who exhibited significant changes in search behaviours throughout their search journeys. As the study was conducted in the wild, without controlling for the tasks, topics, or demographics, the methodology holds high ecological validity for search engines that have access to unconstrained user interaction data. Ultimately, our method informs user models to better support effective web-search interactions.

6.2 Introduction

With global Internet access at 4.66 billion, over half of the global population [Hootsuite and Social, 2020], and the number of indexed pages available online exceeding 50 billion [de Kunder, 2021], using search engines to navigate this space has become a central part of the modern-day knowledge acquisition process. Despite advances, finding desired information can still be a complex task, even for experts, becoming even more challenging for specialist search engines and more broad/exploratory research purposes [Eickhoff et al., 2013, White and Morris, 2007]. Research has shown that the continued use of a Web platform is related to the perceptions of its usability, trust in the medium, and quality of the information it provides [Akram and Malik, 2012]. Young people in particular have shown a heavy reliance on search engines, but despite their familiarity with computers and their ability to navigate the Web, they have demonstrated a lack of critical and analytical skills when assessing the information that they find online [Rowlands et al., 2008]. Moreover, while leading contemporary search engines do support more complex information needs, such as providing additional (only available on demand) search features, they have instead been developed and packaged for commercial purposes, and have been optimised for users seeking rapid answers to simple questions [Tucker and Edwards, 0]. While advanced tools have been developed to improve navigation efficiency, the quality of Web search experiences can still improve through higher levels of search skill and supportive search interfaces [Wang and Yen, 2007], with commercial Web search engines often tracking user preferences and behaviours to better predict user interests and intentions [Teevan et al., 2005]. Examples of this are the use of clickstream data to identify user interests and to re-rank Web search results [Bogaard et al., 2019, Teevan et al.,
Specialist search engines are information retrieval tools that focus on specific audiences, domains, and types of data. Examples include Wolfram|Alpha, a search engine for mathematical queries that involves complex and dynamic computations, ‘Dataset search’, a search engine for datasets launched by Google, and other search engines targeting researchers and more advanced searchers [Fessl et al., 2019]. There have been several investigations involving specialist search engines, such as the analysis of how searchers interact with a web-based, faceted library catalogue for exploratory searches, using methods such as eye-tracking and recall interviews to understand various aspects of search interface use [Kules and Capra, 2012]. Other investigations include gender-based inequalities in the context of résumé search engines, which are specialist tools for recruiters to search for candidates [Chen et al., 2018], and differences in search behaviours between a specialist search engine and ‘Google-like’ search interface for health data [Jay et al., 2016]. The navigation of specialist search engines can be difficult: researchers looking through a large number of social sciences and humanities datasets experienced difficulty getting a sufficient overview of each dataset and often selected datasets based on familiarity rather than relevance [Jay et al., 2016]. The ability to efficiently locate information is especially important for researchers. To improve the Web search experience and outcome, improvements can be made to either the search engine or the search efficiency of the user [Moraveji et al., 2011], and while several methods of measuring the impact of search engine improvements have been outlined, with research often focusing on the algorithms, input modality, and visualisations, monitoring changes to search behaviour has proven to be more difficult [Moraveji et al., 2011, Teevan et al., 2004, Bhavnani, 2001].

The influence of domain knowledge on query outcomes has been investigated previously, with some search behaviours shown to be predictors of domain knowledge (such as the number of documents saved, the average query length, and the average ranking position of documents opened) [Zhang et al., 2011], and others showing no significant differences between domain experts and non-experts when it comes to learning as an outcome of search activities [O’Brien et al., 2020]. It is understood that users tend to behave differently when they are having trouble with Web search [Joachims et al., 2005, Smith and Kantor, 2008] and that search outcome, to a great extent, is associated with behaviours such as the query reformulation strategies [Odijk et al., 2015]. In this paper, we are interested in those search behaviours associated with domain knowledge, knowledge gain and search skill as we expect they may vary over time.

Our main objective is, in spite of user variance, to investigate search behaviours using solely low-level interactions with the search engine. Analysing this particular granularity of data was shown to yield valuable information, such as how searchers examine and engage with the web search results [Lagun et al., 2014, Yu et al., 2019], allowing the investigation of search tasks and information-seeking intentions [Liu et al., 2019], as well as the generation of effective personalised recommendations based on interest in retrieving specific information [Song et al., 2006]. Incorporating temporal factors allows us to monitor interaction features such as session duration, as well as to track the evolution of search behaviours. User behaviours
can also be periodic and hence repeat over time, and the effect of periodicity may provide insights on user models [Aggarwal et al., 2020]. With both active periods and search duration varying from user to user, we use session segmentation to model behavioural evolution. The research question (RQ) that guides the overall direction and objectives of this research study is: What features of user interaction can be used to model search behaviour evolution? As the platform used in this study also includes a MOOC (Massive Open Online Course) aimed at improving users’ ability to search, evaluate, and manage digital information, we also aim to evaluate the impact of such self-regulated learning platforms on users’ search behaviours and search behaviour evolution.

As this study is based on user interaction data without controlling the search topic or user goals, interpreting users’ explicit intentions and actions can be more difficult. Despite this, this work holds relevance for understanding real-world search behaviours as data collected by many search engine operators provide no explicit ground-truth information about user goals/desired query results [Scaria et al., 2014], informing Web designers to better support specialist search engines users such as slow searchers and searchers who are struggling to satisfy their information needs.

The contributions of this work are:

- We propose a data-driven methodology to measure search behaviour evolution. As a proof-of-concept, this methodology was applied to an academic search engine.
- We were able to model search behaviours and identify search behaviour transitions, revealing how they change over time. This allowed us not only to investigate individuals whose search behaviours went through significant change, but also cohorts whose behaviour evolved similarly.
- Including users who also took part in the MOOC to learn advanced search skills enabled us to assess the impact of self-regulated learning on search behaviour evolution. We found no association between training and search skill development.

### 6.3 Related work

Research on search behaviours exhibited while performing learning-oriented search tasks indicates that user-perceived topic knowledge, perceived interest, and task difficulty increase with increasing cognitive levels of the task [Ghosh and Shah, 2017]. In task-oriented searching, it was found that the most important measures of search performance were task completion time and search outcome [Brand-Gruwel et al., 2005, Jenkins et al., 2003, Lazonder et al., 2000]. The outcome can refer to several variables, such as the number of relevant pages found or whether the correct answer was found. Search behaviours such as query complexity evolve over time, with particular features of the search sessions and page visits advancing the knowledge acquisition process, and metrics such as query complexity used to measure the gain of domain expertise [Eickhoff et al., 2014b]. The search behaviours of casual users
tend to use fewer query terms, have shorter search sessions, typically check just one result page, and often make mistakes with the advanced query operators, which denotes difficulties in satisfying information needs [Leroy et al., 2003, Hölscher and Strube, 2000, Jenkins et al., 2003].

Search skill is the user’s competence level of acquiring desired information independent of domain knowledge, such as the use of advanced search expressions. Domain expertise can be differentiated from search skill in that the former ‘concerns knowledge of the subject or topic of the information need, rather than knowledge of the search process’ [Wildemuth, 2004, White et al., 2009]. A number of user studies explored search skill, although there is little agreement on what defines a skilled searcher. For example, information brokers, reference librarians, and information architects were considered skilled [Hölscher and Strube, 2000], this has also been used to describe users with over five years of computer experience and over one year of Internet experience [Jenkins et al., 2003]. Most studies conclude that skilled searchers are more successful at executing search tasks than those considered unskilled [Hölscher and Strube, 2000, Jenkins et al., 2003, Palmquist and Kim, 2000, Saito and Miwa, 2001], however, some studies have identified only small differences between their search strategies and outcomes [Brand-Gruwel et al., 2005, Navarro-Prieto et al., 1999]. Other works have shown that those possessing a higher level of search skill integrate several search processes such as engagement with the information content, interactions with content creators, and the ability to extract and manipulate information [Tucker, 2012].

Knowledge gain is related to the ‘searching as learning’ concept, providing insights into search behaviours through the analysis of human learning and vice versa. Although few have studied the link between searching and learning, these are considered co-existing processes [Ghosh et al., 2018]. Knowledge gain is not to be confused with domain expertise, as mentioned above, which refers to user’s previous experience or inherent knowledge level about the search topic. There is an abundance of research on the domain expertise of searchers. The search behaviours of domain experts and non-experts were found to differ significantly [White et al., 2009]: factors associated with domain expertise (leading to higher search success for their topic of expertise) include performing more queries, a greater number of sites visited, more time spent reading, and using search strategies such as the use of filters and source selection, leading to higher search success for their topic of expertise. The comprehensibility of query results viewed was also found to be related to domain expertise, with research showing that users with higher domain expertise tend to prefer more technical results [Tan et al., 2012]. However, other works show that domain experts may be more sensitive to non-relevant search results, and are, therefore, more likely to feel unsatisfied [Vakkari, 2001]. Significant improvements to rankings have been achieved by factoring in users’ preferences in document complexity and readability into Web search personalisation [Collins-Thompson et al., 2011].

The relationship between domain expertise, knowledge gain, and search skill development can be observed during the search process. Users who are less familiar with search topics may achieve greater knowledge gains [Gadiraju et al., 2018]. With this increase in knowledge gain,
domain expertise can increase too, and by studying the evolution of knowledge across multiple sessions of the same task, researchers found a propensity for domain expertise to increase with each session [Liu et al., 2013, Zhang et al., 2011]. By investigating the connection between domain expertise and search strategy, it was found that, over time, those with lower levels of domain expertise tend to converge towards the same search patterns exhibited by experts [Wildemuth, 2004]. Previous studies have also shown that domain expertise influences users’ search behaviors [White et al., 2009, Duggan and Payne, 2008, Zhang et al., 2005] and effectiveness [Zhang et al., 2005]. Also, higher levels of domain expertise can be associated with better search performance [Duggan and Payne, 2008], and it has been shown that domain experts complete search tasks more efficiently than non-experts [Mao et al., 2018], and that domain expertise enables users to be better at conducting queries in exploratory search [Mao et al., 2018].

The above paragraphs suggest that there are interactive behaviours that can be measured to characterise the search process. While there has been significant interest in the effect of domain expertise on search behaviours, there is limited research on how search skill (independent from domain expertise) manifests in these search behaviours. Some of the works discussed in this section proposed features of domain expertise and learning which can be found in Table 6.1. Within this table, we took a sample of the query and interaction features that were previously used to measure search behaviours in the aforementioned topics: domain expertise, knowledge gain, and search skill. Utilising these state-of-the-art features, we aim to investigate search behaviours on a specialist search engine targeted at early career researchers, and to evaluate the impact of self-regulated learning programmes. Previous research on search behaviour generally aggregates their data across the search journey. We instead investigate the evolution of user behaviour to better understand the fluctuations in user search journeys.

6.4 Study

6.4.1 Platform and data collection

We extracted interaction data from a research platform containing an academic search engine targeting early career researchers (see screenshot of the interface at Figure 6.1). The data was collected over a period of 20 months. The platform integrated a variety of knowledge bases, databases, library catalogues, and information repositories including SpringerLink, Wiley, The German National Library and Flickr. Additional search features include advanced search, filters, results visualisation widgets, and a widget that makes recommendations based on the search history. A number of widgets for the visualisation of the results allow users to explore and discover the search results. In particular, users can explore the relationships between documents and their properties (e.g. author name, subject, and keywords), conduct an interest-driven exploration of search results, aggregate information about the properties of the retrieved documents, and examine keyword frequency in the retrieved documents.
Table 6.1. Query and interaction features used to measure domain expertise, learning, and search skill in previous research

<table>
<thead>
<tr>
<th>Features</th>
<th>Domain Expertise</th>
<th>Learning</th>
<th>Search skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration of the search session</td>
<td>White and Drucker [2007]</td>
<td>Eickhoff et al. [2014b]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jansen et al. [2009]</td>
<td>Yu et al. [2018a]</td>
</tr>
<tr>
<td>Number of queries</td>
<td>White and Drucker [2007]</td>
<td>Eickhoff et al. [2014b]</td>
<td>White and Morris [2007]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jansen et al. [2009]</td>
<td>Yu et al. [2018a]</td>
</tr>
<tr>
<td>Query length</td>
<td>Zhang et al. [2011]</td>
<td>Eickhoff et al. [2014b]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jansen et al. [2009]</td>
<td>Yu et al. [2019]</td>
</tr>
<tr>
<td>Number of unique search terms</td>
<td>Arguello [2014]</td>
<td>Jansen et al. [2009], Yu et al. [2018a]</td>
<td></td>
</tr>
<tr>
<td>Query complexity</td>
<td></td>
<td>Eickhoff et al. [2014b]</td>
<td></td>
</tr>
<tr>
<td>Number of clicks</td>
<td></td>
<td>Collins-Thompson et al. [2016]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Liu et al. [2019]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yu et al. [2018a]</td>
<td></td>
</tr>
<tr>
<td>Number of pages browsed</td>
<td>Arguello [2014]</td>
<td>Bhattacharya and Gwizdka</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gwizdka and Spence [2006]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jansen et al. [2009]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Liu et al. [2019]</td>
<td></td>
</tr>
<tr>
<td>Scroll distance</td>
<td></td>
<td>Yu et al. [2018a]</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.1. A screenshot of the specialist search engine interface used in the study

The platform was instrumented to generate low-level interaction data, including browser window events such as page loads and mouse/keyboard interactive events, with events containing additional information such as mouse event coordinates and the DOM element that triggered the event, as can be seen in Table 6.2. Using WevQuery [Apaolaza and Vigo, 2017], a tool to harvest and query interaction data on the Web, a total of 2,028,769 interactive events were collected alongside event timestamps and URLs. Following a data pre-processing step, events were filtered by URLs (e.g. search landing page, search results page, or advanced search page) to associate the events with the specific functionality/section of the search engine. In total, 239 users’ interactions were captured. In line with established practices, search sessions were then split when there were 40 minutes of inactivity in between two consecutive events [Jones and Klinkner, 2008].

The search engine also includes an online learning platform which also targets early ca-
Table 6.2. Example of captured user interface events.

<table>
<thead>
<tr>
<th>Type</th>
<th>Events</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>mousedown</td>
<td>Start of mouse click action</td>
</tr>
<tr>
<td></td>
<td>mouseup</td>
<td>End of mouse click action</td>
</tr>
<tr>
<td></td>
<td>mousemove</td>
<td>Mouse movement</td>
</tr>
<tr>
<td></td>
<td>mouseover</td>
<td>Hovering into target</td>
</tr>
<tr>
<td></td>
<td>mouseout</td>
<td>Hovering out from target</td>
</tr>
<tr>
<td></td>
<td>doubleclick</td>
<td>Double mouse click</td>
</tr>
<tr>
<td></td>
<td>mousewheel</td>
<td>Mouse wheel interaction</td>
</tr>
<tr>
<td>Selection</td>
<td>select</td>
<td>Selection of page content</td>
</tr>
<tr>
<td></td>
<td>cut</td>
<td>Content cut</td>
</tr>
<tr>
<td></td>
<td>copy</td>
<td>Content copy</td>
</tr>
<tr>
<td></td>
<td>paste</td>
<td>Content paste</td>
</tr>
<tr>
<td>Keyboard</td>
<td>keydown</td>
<td>Start of key press action</td>
</tr>
<tr>
<td></td>
<td>keyup</td>
<td>End of key press action</td>
</tr>
<tr>
<td></td>
<td>keypress</td>
<td>Key press action</td>
</tr>
<tr>
<td>Window</td>
<td>load</td>
<td>Page is loaded</td>
</tr>
<tr>
<td></td>
<td>resize</td>
<td>Browser window is resized</td>
</tr>
<tr>
<td></td>
<td>unload</td>
<td>Window is closed</td>
</tr>
<tr>
<td></td>
<td>windowfocus</td>
<td>Browser tab gains focus</td>
</tr>
<tr>
<td></td>
<td>windowblur</td>
<td>Browser tab loses focus</td>
</tr>
<tr>
<td></td>
<td>scroll</td>
<td>Change of scroll state</td>
</tr>
<tr>
<td>Other</td>
<td>change</td>
<td>Input element state change</td>
</tr>
<tr>
<td></td>
<td>contextmenu</td>
<td>Opening of context menu</td>
</tr>
</tbody>
</table>

reer researchers, a MOOC on open science and open research methods with the following modules:

- Data and information literacy: how to search, evaluate, and manage digital information.
- Communication and collaboration: how to use technologies to interact, share, communicate, and collaborate.
- Content creation: how to develop, integrate, and apply copyright to digital content.

Although the search engine assessed in this paper was introduced along with others in the above modules, it should be noted that the content primarily covered general search skills and did not feature the search engine heavily. The courses took place in three waves and ran for four weeks each time: 12th November–16th December 2018, 21st January–17th February 2019, and 17th June–14th July 2019. 62 users who enrolled in the MOOC also used the search engine.

To establish individual learning paths for users within the MOOC curriculum, newly registered users were asked to fill in a short survey to assess their previous knowledge and digital competencies. In a three-point Likert scale (1: basic, 2: intermediate, 3: advanced) the survey assessed the perceived competency in the learning modules described above. For instance, for Data and information literacy, the question asked was: When you think of your ability to search, evaluate and organise data and information on the internet, which of the following statements best describes your behaviour?, and the user is asked to select from the following three paragraphs that best describe their skills which indicates basic to advanced skill levels:

- I can look for information online using a search engine. I know not all online information is reliable. I can save or store files or content (e.g. text, pictures, music, videos, web
pages) and retrieve them once saved or stored. I keep information and files in a number of different physical supports (hard drive, USB stick, memory card).

- I can use different search engines to find information. I use some filters when searching (e.g. searching only images, videos, maps). I compare different sources to assess the reliability of the information I find. I classify the information in a methodical way using files and folders to locate these easier. I do backups of information or files I have stored.

- I can use advanced search strategies (e.g. using search operators) to find reliable information on the internet. I can use web feeds (like RSS) to be updated with content I am interested in. I can assess the validity and credibility of information using a range of criteria. I am aware of new advances in information search, storage and retrieval. I can save information found on the internet in different formats. I can use cloud information storage services.

In total, 76 users participated in the survey (73 to completion), 36 of whom used the search engine.

The reported competency levels can be seen in Figure C.1. The survey results show that nearly half of the participants believe they had advanced information literacy skills while only 24% and 35% of them believe they are advanced in content creation and communication, respectively.

Figure 6.2. User self-reported competency levels

6.4.2 Methodology

To analyse user behaviour evolution across multiple search sessions, with both active periods and session duration varying from user to user, we grouped the sessions equally into $N$ periods. The value of $N$ determines the percentage of sessions ($1/N$) that each period contains. For instance, when $N = 2$ the first 50% of sessions of each user are grouped into the first period and the remaining 50% into the second. When $N = 3$, each period contains 33% of the search sessions of each user, where the first bin contains the first 33% and so on. For a user with a total of $M$ Sessions ($S_M$), each period $P$ can be represented as:

$$P_i = \{S_{1+\frac{(i-1)M}{N}}, \ldots, S_iM\}$$
where $P_i$ is the $i$th Period from a total of $N$ periods. For the sessions that could not be equally divided into $N$ periods, or when the total number of sessions is smaller than $N$, these (additional) sessions are grouped into former periods.

Based on our analysis of low-level interactions in Section 5.3, we selected a set of established features that can be used to accurately identify search behaviours:

- **Click**: number of mouse clicks on search results page (positively correlated with domain expertise) [Collins-Thompson et al., 2016, Yu et al., 2018a, Liu et al., 2019].

- **Session**: session length calculated using timestamps of first and last events (positively correlated with domain expertise) [White and Drucker, 2007, Jansen et al., 2009, Yu et al., 2018a, Eickhoff et al., 2014b].

- **Time**: also known as dwell time is the aggregation of time spent on search results page (positively correlated with knowledge gain) [Arguello, 2014, White and Drucker, 2007, Yu et al., 2018a, Liu et al., 2019], calculated using timestamps of the first and last events on the search results page.

- **Scroll**: the scrolling distance calculated by vertical distance of mouse coordinates [Yu et al., 2018a], calculated using the absolute Y-coordinates of mouse scrolling events.

- **Keyword**: the total number of unique search terms (positively correlated with domain expertise) [Arguello, 2014, Yu et al., 2018a], calculated using the URLs which contains the search terms.

These features were then computed for each user and averaged over all sessions for each period. We then performed Silhouette analysis on several clustering algorithms (i.e. KMeans, DBScan, Agglomerative clustering, and Meanshift) to select the algorithm that generates the most consistent and informative clusters for our dataset, as well as to determine the appropriate number of clusters through Silhouette analysis, which is used for interpreting and validating the consistency within clusters [Arbelaitz et al., 2013]. The average Silhouette scores for each algorithm and each number of clusters was computed and compared, allowing us to determine the most appropriate number of clusters and algorithm —see Table 6.3. We then computed the clusters and descriptive statistics of the features in each cluster and period. Informed by these statistics and the meaning ascribed to the features by related work, each cluster was labelled to represent the search behaviours and characteristics of its users, which was used to analyse the evolution of clusters over time and the transition of users between clusters. By identifying users (n=62) who also participated in the MOOC and comparing with other users, we were able to investigate what impact participation in the MOOC had on the evolution of search behaviours and whether self-regulated learning support users in becoming more proficient searchers.
6.5 Results

While the average number of sessions was nine (median: 2, SD: 27.98) with an average session duration of 24m58s, over 80% of users had fewer than five sessions. Consequently, we explored the 2–5 space of \( N \) to minimise the number of empty periods. The Silhouette coefficients for the selected number of clusters are shown in Table 6.3, along with the distribution of users in each cluster. We noted that some Silhouette coefficients were high (over 90%) due to imbalanced clusters (with the majority of users assigned to one cluster, increasing the coefficient and therefore inflating the overall Silhouette score). We selected the KMeans clustering algorithm as it achieved higher Silhouette scores with more balanced clusters.

<table>
<thead>
<tr>
<th>Number of periods</th>
<th>Period</th>
<th>Clusters</th>
<th>Distribution</th>
<th>Silhouette Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>{75, 164}</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>{19, 220}</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>{237, 2}</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>{232, 7}</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>{231, 8}</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>{159, 8, 72}</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>{211, 28}</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>{229, 10}</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>{223, 16}</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>{80, 157, 2}</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>{210, 29}</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>{218, 21}</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>{232, 7}</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>{236, 3}</td>
<td>0.92</td>
<td></td>
</tr>
</tbody>
</table>

While we explored all the problem space, here we focus on exploring the problem space with two and four periods for the sake of brevity. Also, with three periods the clusters are highly imbalanced as can be seen in Table 6.3. The KMeans algorithm generated the cluster distributions in Figures 6.3 and 6.4, which show the distribution of users in each cluster (denoted by colour) for each period. We calculated the mean, median, and standard deviation of each feature for all users in each of the clusters and statistics such as average time between clicks (\( t_c \)) and average time spent on each keyword (\( t_k \)). Here we report the results of clustering, and how the above statistics characterise each cluster in terms of search behaviours.

6.5.1 Two periods

In the first period (see the first row of Figure 6.3), users of each of the two clusters can be characterised as *explorers* and *quick searcher and non-explorers*. The users in the explorers’ cluster (75 explorers, as depicted in blue dots) clicked (\( t_c : 26.3s \)) and scrolled (time between scrolls, \( t_s : 0.9s \)) less often in general, while clicking more (45.9) and spending more time on each keyword (23m54s) compared to users in the other cluster. They also had longer search sessions and spent a larger proportion of their time on the search results page (84.2%). Users in this cluster were characterised as *explorers* as their behaviours were more of an exploratory nature. Data for this cluster varies significantly for different users (standard deviation for the average number of clicks being 83.4 and time spent on search results page being 24m55s),
as shown by the distribution of blue dots, with a few users scoring high on all variables. 164 users in the *quick searcher and non-explorers’* cluster include users who clicked \( t_c: 19.8s \) and scrolled \( t_s: 0.5s \) more often overall while spending a third of the time of the *explorers’* cluster searching for each keyword (8m54s). Users in this cluster spend less time in the search results page and also interacted with the platform less than users in the *explorers’* cluster, which may be characteristic of navigational searchers who, having satisfied their information needs, swiftly leave the session (with the average length of sessions being around 10 minutes).

In the second period (see the second row of Figure 6.3), users in the two clusters are characterised as *explorer/passive searchers* and *passive searchers*. *Explorer/passive searchers* account for 19 users who clicked \( t_c: 47.1s \) and scrolled \( t_s: 1.7s \) less frequently, spent more time on each keyword (43m45s), and spent a large proportion of time on the search results page (81.4%) compared to the *passive searchers’* cluster. Compared to the *explorers’* cluster in the first period, the number of clicks per keyword (45.3), and the proportion of time spent on the search results page remained roughly equal, while the average number of clicks and scrolls decreased, and the session length increased. This may be interpreted as spending more time reading the search results, alternatively, it may also mean the users have left the platform idle after finding the desired results. The users in the *passive searchers’* cluster (220 users) did not click or scroll much on the search results page while spent more time on each keyword (9m30s) and a larger proportion of time on the search results page (72.7%) compared to users in the *quick searcher and non-explorers’* cluster in the first period. This cluster may be similar to the *explorer/passive searchers’* cluster in that they continued their searches from the last period. Some of these users, however, did not conduct further searches as the average number of keywords for this period is less than one —these are the *passive searchers*. The transition of users between clusters from period one to two can be seen in Table 6.4.
Figure 6.4. KMeans clustering with four periods (displayed sequentially from top to bottom), the figures on the left show the features Scroll and Click, the figures in the middle show Time and Session, and the figures on the right show Keyword with each axis presented on the same scale. Each user is represented as a data point and coloured based on their corresponding cluster.

Table 6.4. User transitions between clusters from period one to two

<table>
<thead>
<tr>
<th>No. of users</th>
<th>First period</th>
<th>Second period</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Explorers</td>
<td>Explorer/passive searchers</td>
</tr>
<tr>
<td>159</td>
<td>Quick searcher and non-explorers</td>
<td>Passive searchers</td>
</tr>
<tr>
<td>5</td>
<td>Quick searcher and non-explorers</td>
<td>Explorer/passive searchers</td>
</tr>
<tr>
<td>61</td>
<td>Explorers</td>
<td>Passive searchers</td>
</tr>
</tbody>
</table>

**Interpretation**

Moving from the first period to the second one, we observe a general decrease in the number of clicks, scroll distance, and the number of unique search terms, suggesting that users generally interacted less over time. We can assume that these users continued reading the results page content instead of conducting more searches. Passive searchers possess low values of features.
(Click, Scroll, and Keyword) that can be associated with discontinuing the search, which may suggest that following the last period, they found what they were looking for or had given up, or they may have returned after some time and did not conduct more searches. Most users of the quick searcher and non-explorers’ cluster in the first period remained and transitioned to the passive searchers’ cluster in the second period. This suggests that these users may be characterised as searchers who found desired information and continued exploring in the second period without conducting more searches as they spent a large proportion of time on the search results page (72.7%) and the average number of keywords for the second period is less than one. Five users in the quick searcher and non-explorers’ cluster (first period) moved to the explorer/passive searchers’ cluster in the second period by clicking and scrolling more often in the second period and spending more time on search results. These users appear to have conducted further reading as they spent more time on each keyword (43m45s), and spent a large proportion of time on the search results page (81.4%). Fourteen users from the explorers’ cluster remained in the same group and became explorer/passive searchers in the second period, suggesting that after explorations in the first period, these users may have satisfied their information needs and thus did not conduct further searches in the second period, or their search behaviours were consistent throughout their search journeys. 61 users moved from the explorers’ cluster to passive searchers, which suggests a decrease in the amount of interactions and perhaps abandonment in the second half of their journeys. This may correspond to users who did not find the desired information and had given up on doing more searches or they may have returned after some time and did not conduct more searches. Using two periods, we identified five users who exhibited changes in their search behaviours, moving from quick searcher and non-explorers to explorer/passive searchers. We also identified 61 users who may have abandoned the searches, moving from explorers to passive searchers.

Out of the 62 MOOC users, 60 remain in the same group (quick searcher and non-explorers – passive searchers) and two users switch groups from explorers to passive searchers. We therefore did not identify any users who participated in the MOOC whose search skill (or engagement) had improved. Of the 17 users who reported having advanced information literacy skills, 13 were grouped as quick searcher and non-explorers in the first period, and passive searchers in the second period, while four users moved from explorers to passive searchers. Of the 8 users who reported having basic information literacy skills, six were group as quick searcher and non-explorers and then passive searchers and two users moved from explorers to passive searchers. Of the 11 users who reported having intermediate levels of skill, nine were group as quick searcher and non-explorers and then passive searchers while two users moved from explorers to passive searchers. These results suggest that despite the differences in reported levels of information literacy, the participants performed most of the searching in the first half of their active periods.

6.5.2 Four periods

The clusters in the first period (see the first row of Figure 6.4) were characterised as quick searcher and non-explorers, explorers, and slow searcher and explorers. The quick searcher
and non-explorers’ cluster accounted for 159 users who clicked more frequently than others ($t_c: 18.4s$), users spent the least amount of time on each keyword (8m5s), clicked the least (17.5) per keyword and spent the lowest proportion of their session lengths on the search results page (65.2%). Users in this group seemed to have found desired information quickly without much exploring. Eight users in the explorers’ cluster did significantly more scrolling (9359.9) than others while spending more time (47m41s) and clicking more times for each keyword (45.5) compared to the quick searcher and non-explorers’ cluster. These users spent the largest proportion of their session lengths on the search results page (90.4%). Compared to their counterparts, the behaviours of users in this group are more of an exploratory nature. We therefore label this group as explorers. 72 users in the cluster slow searcher and explorers clicked ($t_c: 24.7s$) less frequently and scrolled ($t_s: 1.3s$) least frequently among the three clusters, while clicking more times (50.6) and spending the largest amount of time per keyword (24m44s).

In the second period (see the second row of Figure 6.4), the clusters (and their users) were characterised as quick searcher and explorers and slow searcher and explorers. 211 users in the cluster quick searcher and explorers clicked ($t_c: 33.4s$) and scrolled ($t_s: 0.7s$) much more frequently, spent less time on each keyword (9m22s), with a smaller proportion of time spent on the search results page (67.8%), compared to users in the slow searcher and explorers’ cluster. Compared to users in the quick searcher and non-explorers’ cluster in the previous period, they spent more time on each keyword but performed fewer clicks (11.4), with a greater proportion of time spent on the search results page. 28 users in cluster slow searcher and explorers resemble the slow searcher and explorers’ cluster in the previous period, with users clicking more (30.6) and spending more time (43m31s) per keyword.

We describe the clusters in the third period as explorers and quick and active searchers (see the third row of Figure 6.4). Compared to the cluster quick searcher and explorers in the previous period, the 229 users in the explorers’ cluster clicked more (12.5) and spent more time (10m6s) on each keyword, while clicking and scrolling less frequently overall. These users may be exploring search results. Compared to the cluster slow searcher and explorers in the previous period, the ten users of the cluster quick and active searchers clicked ($t_c: 52.3s$) and scrolled less frequently ($t_s: 1.8s$) overall. The users clicked more (25.5) and on average conducted the most searches (3.6) among all the clusters.

Finally, the users in the fourth period are characterised as inactive users and active searcher and explorers (see the last row of Figure 6.4). The 223 users in the inactive users’ cluster did not interact much with the platform, with only on average 2.5 minutes per search session. Compared to the cluster quick and active searchers in the previous period, the 16 users from the cluster active searcher and explorers clicked ($t_c: 33.6s$) and scrolled ($t_s: 1.3s$) more frequently overall, clicked more per keyword (36.6), and spent a larger proportion of their time on search result page (90.2%). The average number of unique search terms (2.0) for users in this cluster shows that the users are still actively conducting searches in this period.
Interpretation

Figure 6.5. User transitions between clusters. Coloured nodes represent clusters in each period (left to right; period 1-4), with the users transitioning between clusters represented by grey streams. The number of users transitioning in each stream is listed sequentially in brackets.

When we compare the transitions between clusters in Figure 6.5 with those of the two-period scenario (see Table 6.4), having four periods increased the resolution and enabled us to identify more changes in search behaviour and to acquire further insights into search behaviours evolution such as users who transitioned from slow searchers to quick searchers. Six slow searchers from period one transitioned to the cluster quick searcher and explorers, while only one user from period two transitioned from slow searcher and explorers to quick and active searchers. We can identify a tendency for user behaviours to converge over time, as shown in Figure 6.5. From the first period to the last, the user clusters became more unified and the majority of users moved to the dominant cluster.

Out of the 62 MOOC users, we identified three users whose search skill may have improved during the second period. Out of the 17 users who reported having advanced information literacy skills, we identified 13 and 16 users who were categorised as quick searchers in the first and second period, respectively. Of the eight users who reported to have basic information literacy skills, seven and six users were identified as quick searchers in the first two periods, respectively. Of the 11 users who reported having intermediate information literacy skills, ten and nine users were identified as quick searchers in the first two periods, respectively. The results suggest that despite the self-reported competency levels, most participants of the survey can be categorised as more efficient and thus more skilled searchers than they initially reported.

6.6 Discussion

We found that 80% of users had five search sessions or less, and we consider those who had more than five sessions to be frequent searchers who were struggling to satisfy their information needs, causing them to repeatedly discontinue and return later, or users who were performing exploratory searches and actively learning more about a topic. The latter two can
be characterised as users having a single goal performing search activities across multiple search sessions i.e. multi-session/cross-session searching [Li et al., 2020], which has been identified as a common approach in Web searching [10., Spink et al., 1999, 2002]. The literature categorises search needs as informational (e.g. searching for information), navigational (e.g. searching for the desired URL), or transactional (e.g. searching for sites which perform certain transactions) [Broder, 2002]. We expect users who are performing informational searches to spend more time exploring results, click on more results, and scroll more. Accordingly, users performing navigational queries are expected to spend less time on the result page and scroll less than for informational queries, but with more clicks and longer sessions as they repeat searches more often. As our specialist search engine targets researchers, it is likely that users are not carrying out transactional queries alone, but combine these with navigational tasks, such as finding a page to download a file. For instance, users in the quick searcher and non-explorers group spend significantly less time clicking and scrolling through results, and have shorter sessions overall —which indicates that these users are conducting navigational queries. Explorers can be characterised as users performing informational searches with long reading times as they spend more time on the results page but less time clicking and scrolling, compared to quick searchers. Due to the large amount of time spent clicking and scrolling, explorers struggling to find the desired information, as mouse movements and time spent on the page are known indicators of users not finding what they are looking for [Thomas, 2014]. This is not necessarily negative if the user is learning and consuming content [Raman et al., 2013], and searchers can still benefit from the information that they are exposed to as they search [White and Huang, 2010].

As this study is focused on search behaviour based on interaction data, we did not control for search topic and user goals. It is therefore difficult to know precisely whether users were performing an exploratory search or are struggling to find the desired information as the behavioural markers are similar [Marchionini, 2006, White and Chandrasekar, 2010]. However, with data collected by many search engine operators providing no ground-truth about user goals/desired query results [Scaria et al., 2014], our method offers a higher level of ecological validity for analysing search behaviours in the wild.

We labelled each cluster so it represents the search behaviours of its users at a given time, which we then used to analyse the evolution of clusters and the transition of users between clusters. We note that when plotting Time vs Session (see Figure 6.3 and 6.4) results in clusters which are easily distinguishable, whereas plotting Scroll vs Click and Keyword fails to show clear differences among the clusters. Figure 6.3 and 6.4 show that the Keyword, Click, and Scroll for users in most clusters varied significantly. Such variation suggests that differences between clusters are not distinctively manifested by these three features. For the session length and time spent on the search results page, there are clear differences between users in each cluster, other than a few outliers, these two features could be more significant for differentiating user behaviours. For certain periods the majority of users were assigned to a single cluster, indicating that there are significant behavioural overlaps between the users. This can be expected as it has been shown that behavioural differences between skilled and
unskilled searchers can be minimal [O’Brien et al., 2020, Brand-Gruwel et al., 2005, Navarro-Prieto et al., 1999]. Moreover, search behaviours influenced by domain expertise [White et al., 2009, Duggan and Payne, 2008, Zhang et al., 2005], and as we are operating on an academic search engine for researchers, we expect most users to have some baseline level of search proficiency and expertise regarding their own domains. Here, we attempted to uncover and interpret changes in search behaviours using the proposed data-driven methodology.

We revisit the research question formulated at the outset—what features of user interaction can be used to model search behaviour evolution? As the number of user sessions and session duration vary from user to user, we utilised segmentation of user interaction data to represent search behaviour changes occurring throughout the user search journey. We grouped user sessions into periods, with each period representing a time interval and holding an equal percentage of search sessions. This allowed us to track search behaviour evolution by examining user interactions over a set number of periods. We then conducted clustering to investigate search behaviours exhibited by groups of users. By analysing the clustering results for each period, we identified a trade-off between having more periods (allowing for user behaviour to be analysed more completely over time) and the number of sessions and interactions included in each period (with too few making it difficult to observe changes in behaviour).

We found that, for the current case study and dataset, three periods produce imbalanced clusters, and having two periods limited the resolution and prevented the extraction of details such as user transitions between clusters. Moreover, in the first period of the four-period scenario, different to that observed with the two-period scenario, users were clustered into three groups (i.e. quick searcher and non-explorers, explorers, slow searcher and explorers). The inactive searchers user groups became prominent only in the last period no matter the number of periods we analysed. This suggests that by increasing the resolution and separating the analysis into four periods we receive finer granularity of the findings while still ensure interpretability. However, a higher resolution implied that user behaviours showed less variability between different periods, with user groups labelled as explorers in multiple periods, suggesting that the number of sessions included in each period may not be enough to differentiate user groups between periods or that these users’ behaviours did not change. Although more periods (N=4–5) resulted in relatively limited variability between clusters across different periods, we were still able to differentiate the user groups and conduct more in-depth analysis of the user behaviour evolution. We therefore identified four as the most appropriate number of periods. This number is not generalisable as it is dependent on the number of search sessions: with our dataset, the average number of sessions was nine, but over 80% of users had less than five sessions. For example, in a dataset where users had conducted more search sessions on average, the ideal granularity would have been different. However, the trade-off we identified will likely still apply: having more periods allows for thorough analysis of user behaviour evolution while resulting in fewer number of sessions and interactions included in each period. In our dataset, we were able to detect differences in search behaviours across individuals and changes across periods, for example, we were able to identify seven users transitioning from slow searchers to quick searchers. As shown in Figure 6.5,
there is a tendency for user behaviours to converge over time. This confirms previous research suggesting that longer exposure to a specialist search engine causes user search patterns to resemble those of experts [Wildemuth, 2004]. The user clusters became increasingly cohesive from the first to the last period and the majority of users transitioned to the dominant cluster. While some studies identified only minor variations in search behaviours between skilled and unskilled searches [O’Brien et al., 2020, Brand-Gruwel et al., 2005, Navarro-Prieto et al., 1999], the proposed methodology allowed us to discover behavioural differences of users at the beginning of their search journeys and evolution of their search behaviours throughout.

By identifying those users who also participated in the MOOC, we could evaluate the impact of self-regulated learning on search skill evolution. This impact related to the participation on MOOC, the effectiveness of MOOC materials, and/or its compatibility with the platform. We found that most MOOC users remained in the majority cluster (i.e. quick searcher and non-explorers - quick searcher and explorers - explorers - inactive users) throughout their search journey, with no evidence that MOOC participation provided support to user search skill improvement. In our case, the MOOC focused on information literacy, open science, and research techniques, so it is possible that the MOOC content was too general for users to practically improve their search skills with this particular search platform. This could also be due to the fact that users degree of participation in the MOOC was insufficient to have a significant impact on user search skill. Another factor could be the time frame of user participation in the MOOC, its impact on user search skill may not be apparent at the time of this study and that user search skill improvement may take time to manifest.

6.7 Conclusion

We propose a data-driven methodology to measure search behaviour evolution using low-level interaction data. As a proof-of-concept, this was applied to an academic search engine. By segmenting user search sessions, we were able to monitor search behaviour throughout the user search journey and therefore monitor search behaviour evolution. Segmentation of user sessions allowed us to identify users who became more efficient at searching over time, demonstrating the utility of our methodology and potentially enabling targeted/personalised support. As our interaction data was collected from users performing searches in a setting with high ecological validity, with search topics and user goals uncontrolled, similar to the type of data extracted by many search engine operators, this work holds relevance for understanding real-world search behaviours. Notably, we discovered that most MOOC users remained in the majority cohort throughout their search journeys, and found no link between MOOC participation and search behaviour evolution, suggesting that self-regulated learning lead to improvements in search skill.
6.7.1 Limitations

As this study is focused on specialist search engines, there may be limitations to the generalisability of conclusions drawn. There are no universal models for specialist search engines as targeted users, domain areas, and interface designs differ. Hence, we contribute to the methodological space to build user models with the novel approach outlined in this paper. The applicative interest of our approach lies in providing designers for specialist search engines methods to monitor and analyse search behaviour evolution to improve platform learnability, usability, and potentially enable targeted/personalised support. To provide finer-grained diagnosis of search behaviours, a much greater level of detail is required, such as the user search goals and outcomes.
Chapter 7

Discussion

The aim of this project was to investigate the methodological and practical use of low-level user interaction data for user modelling in the topics of online learning, Web familiarity, and Web search behaviour. We identified a research gap in the lack of widely accepted methods for analysing low-level user interaction data, and our goal was to discover ways to make the most of low-level data while mitigating its drawbacks. While attempts have been made to use low-level user interaction, studies often solely focus on a few interface events, without fully exploiting the potential of analysing this type of data [Apaolaza et al., 2015]. In this thesis, we presented a novel methodology to utilise low-level data combined with data mining and thematic analysis techniques to extract and investigate user interaction patterns as well as high-level user behaviours. This work is the first to demonstrate the practical utility of low-level data in three distinct domains. Throughout the studies, we addressed the disadvantages of using low-level data identified in previous work [Renaud and Gray, 2004, Hilbert and Redmiles, 2000, Apaolaza et al., 2015]. We demonstrated the application of our methodology in order to hopefully establish a framework or foundation for future research examining low-level user interactions in greater detail.

Despite widespread access and increased interaction with Web technologies, people continue to have difficulty achieving their online objectives [Eickhoff et al., 2013, White and Morris, 2007]. Nowadays, in order to provide better support to users, websites, mobile applications, and web tools collect and track user data in order to analyse user behaviour, evaluate user experience, and make improvements [Teevan et al., 2005, Bogaard et al., 2019, Teevan et al., 2005, Radlinski and Joachims, 2007]. User modelling is a burgeoning area of research in a variety of domains. As long as there is human-computer interaction, end-user behaviour provides critical information about the user experience and can be used to inform interface design. Several studies have identified potentials for user modelling using low-level data [Apaolaza et al., 2015, Hilbert and Redmiles, 2000, Apaolaza et al., 2015, Guo et al., 2009]. We conducted a review of the literature on user Web interactions and discovered that the majority of studies focus on high-level interaction, possibly to ease analysis and interpretation. Hilbert and Redmiles [2000] also pointed out problems with analysing low-level data, which was echoed by [Renaud and Gray, 2004, Apaolaza et al., 2015]. This analysis informed and motivated our research goals and subsequent studies.

This section discusses the the first research questions posed in this thesis.
RQ1 How can we use low-level interaction data for user modelling in each domain?

There are challenges associated with investigating low-level interaction data. We conducted three studies that demonstrate the application of the proposed data-driven methodology we outlined in Chapter 3. We proposed a novel data-driven methodology for extracting and analysing user interaction patterns and behaviours through the use of sequential pattern mining and thematic analysis. To begin, we conducted background research on the domain of learning analytics, focusing on the MOOC domain in particular, where we identified a research gap related to the lack of assessment methods for cMOOCs. We then conducted a pseudo-systematic review of the literature on assessment methods for online learning and also on the properties of learning. The literature review informed the first study, which we conducted on a cMOOC platform and demonstrated our methodology by using low-level interaction to build user models that predicted user learning progress and achievement (Chapter 4). Using sequential pattern mining and thematic analysis, we were able to extract user interaction patterns and effectively interpret them to investigate user behaviours. We included features of user behaviours in machine learning models to seek behaviours indicative of properties of learning. We evaluated our approach by comparing user models, including only engagement features. Finally, we analysed user behaviours over time by plotting the occurrences of behaviours and classifying user behaviours based on our interpretations. In Chapter 4, we demonstrated the practical application of our methodology and its added value. To determine the generalisability of our methodology and to delve deeper into the usability topic discussed briefly in Chapter 4, we conducted a follow-up study presented in Chapter 5 in which we applied the methodology to the University of Manchester School of Computer Science website. The study was also motivated by Apaolaza et al. [2015] on the subject of familiarity. We presented a secondary analysis on user Web familiarity using low-level data while using a data-driven approach involving a large variety of events. We applied the same methodology, modifying it at times to accommodate the larger and more complex dataset in Chapter 5. In Chapters 4 and 5, we included features that represent trend of user behaviour occurrences and strength of the trend, where we discovered correlations between these features and the target variables, and incorporated these features into our user models. Additionally, in Chapter 4, we conducted a brief analysis of user behaviours over time, with this analysis and its implications supporting the argument that including temporal factors in such analyses can be useful. To further investigate these findings, we conducted a third study in which we examined the evolution of user search behaviours. Previous research showed that identifying differences in user groups can aid with personalisation. Therefore, we adopted unsupervised learning to identify clusters in users and analysed the characteristics of identified user groups. We also investigated the effect the online learning on search skill evolution by identifying the users who were active on both platforms.
7.1 Findings and implications

Here, we discuss the main findings from the studies, which addresses RQ2 *What insights can we learn about user behaviours in each domain?*

Our main findings in Chapter 4:

- We discovered that student achievement could be significantly predicted by interactive behaviours such as exploring a forum and exiting the online platform after exploring the week 1 materials. Exploring this further, we found that the less students abandon week 1 materials, the more likely they are to earn a badge, and the number of consecutive sessions in which the forum’s main area was not explored was significantly associated with earning a badge - a counter-intuitive correlation [Anderson et al., 2014, Crues et al., 2018, Ramesh et al., 2014]. We discuss that This may be due to the forum being used to learn and find support during cMOOCs, and advanced learners who become less reliant on this will then be more likely to receive a badge.

- We developed models based on the interaction patterns associated with user achievement, and they achieved an accuracy of 91-95% in identifying low-achieving students. By comparing the models from two regression analyses containing different features, we discovered a 10% contribution to the explainability of student achievement from features of behaviours other than engagement. This finding implies that low-level user interactions can capture implicit user behaviours during the learning process, which can contribute to user modelling of the learning outcome.

- We discovered an association between *becoming a more efficient user* and student achievement, which suggests that the learnability of the interface has an effect on the learning outcome as measured by student achievement. The association may imply that the more user-friendly a learning platform is, the more efficient students can navigate and use the platform, the more effectively students learn. Additionally, it may imply that students who are more familiar with the learning platform or who are skilled navigators are more adept at learning. This creates new research opportunities for analysing platform learnability and student learning outcomes.

Our main findings in Chapter 5:

- By identifying patterns that were correlated with user survey responses, we discovered the difference element –breadcrumb, that explains the negative correlation with the user reported ease when interacting with the website. The correlation suggests that users who interact more frequently with the breadcrumb element over time have a more difficult time interacting with the website. Breadcrumbs have been widely used as a navigational aid in hypertext systems used to inform users of their current location within a website and to provide quick access to navigate directly to previous Web pages in the pathway. This finding contradicts previous research indicating that breadcrumbs benefit users for
navigation efficiency and satisfaction [Teng, 2005, Mukherjea and Foley, 1995, Park and Kim, 2000, Blustein et al., 2005, Maldonado and Jlesnick, 2002] However, it can be explained that users who are unfamiliar with breadcrumbs and their functionality may be impacted, and that using breadcrumbs did not improve their navigation efficiency, which is supported by research that brief breadcrumb training results in more efficient navigation [Rogers and Chaparro, 2003].

- We included features of the identified user interaction pattern with variations of survey responses as the dependent variables in classification analysis, reaching an overall 82.7% accuracy when classifying users with higher levels of familiarity.

Our main findings in Chapter 6 are:

- We were able to discern differences in search behaviours between individuals and changes over time; We identified seven users transitioning from slow searchers to quick searchers. By analysing user group transitions over time, we found that the user clusters became increasingly cohesive, and the majority of users transitioned to the dominant cluster. We also discovered the trend toward user behaviour convergence over time, which corroborates previous research [Wildemuth, 2004].

- By identifying those users who also participated in the MOOC, we evaluated the impact of self-regulated learning on search skill evolution. We found that most MOOC users remained in the majority cluster (i.e. quick searcher and non-explorers - quick searcher and explorers - explorers - inactive users) throughout their search journey, with no evidence that MOOC participation aided in user search skill improvement.

7.2 Theoretical and Practical Implications

In this section, we discuss the advantages and drawbacks of our proposed methodology and identify the studies’ limitations.

There is a lack of established methods of interpreting low-level user interactions. To address this, we proposed a data-driven methodology involving sequential pattern mining and thematic analysis to extract and analyse low-level user interaction patterns. As a proof-of-concept, this methodology was applied to three distinct platforms in different domains; an online learning platform with 193 students (Chapter 4), the University of Manchester School of Computer Science website with 35,819 users, including 268 revisiting users who reported their levels of familiarity (Chapter 5), and an academic search engine with 239 users (Chapter 6).

We began with the methodology where we first conducted data pre-processing; as we discussed in Chapter 3, data pre-processing is a critical step in any type of data analysis. It serves as the foundation for the subsequent steps in our case. While low-level interaction
data is relatively easy to collect, users can quickly generate a large amount of data, complicating data management and analysis. Fine-grained interaction data may contain numerous repetitions and predefined patterns, such as mouse movements. By combining and replacing them, interpretability and scalability can be improved. We can extract this information by representing these interactions with a $event-node-URL$. The data pre-processing assisted us in mitigating and reducing the volume of data and noise. Reflecting on the steps taken: we first filtered the events as some are only available on mobile platforms. Users’ interaction patterns and behaviours may differ based on the devices they use; for example, people may need to scroll more when visiting a website using mobile devices due to their smaller screens. These differences may affect the following steps and eventually the user models. It is possible to filter user interactions by the device type, which may be useful for web platforms that cater to specific user types, such as mobile users, website users, or app users. Following this, we conducted a series of transformations to the UI events to add semantics to them. From our process and findings, these steps were proved to be critical; for example, we were able to extract user behaviours involving leave page actions. This additional discovery was only possible after we transformed the windows focus/blur events. We then filtered the events by URL domain; we applied additional selection as the number of unique URLs, and the domains they belong to was too large for the study in Chapter 5. We ended up with 7 URL domains, in retrospective from our findings, because the website has a diverse range of users with varying backgrounds and browsing goals (current UG students looking for course information, prospective PG students looking for school information), their behaviours may vary significantly as a result. Although we enter a dilemma here, focusing on a small set of URL domains will likely reduce our dataset, the number of users, and the number of interactions, but the added focus would likely lead to more patterns as the users are more similar in the sense that they visit Web pages within the same small set of domains, and that their behaviours and which elements they interact with will also be similar. This can be beneficial as, for example, we found that increasing the use of breadcrumbs makes users feel less comfortable interacting with the website, website navigation issues in specific domains may be visible in a narrow focus; however, when the focus is expanded, user behaviours become more representative (exhibited by a larger number of users), but also more generic; we will be unable to detect distinct behaviours in a single or a small set of domains. However, focusing exclusively on a small set of domains limits the potential to detect cross-domain behaviours; for example, in Chapter 4, we discovered patterns that spread across the homepage, week 1 page, and the forum. However, we did not detect cross domains behaviours in Chapter 5. This could also be because user behaviour on MOOC platforms is more constrained to a small set of domains, whereas on the university website, users can navigate to a variety of pages, making it more difficult to track behaviours across domains. Similarly, we filtered the node IDs based on user coverage for each of the seven URL domains. Concentrating on a large number of web elements makes it difficult for SPM algorithms to identify representative patterns. Likewise, focusing on a small set of node IDs makes the patterns generic and difficult to identify distinct behaviours across a variety of web elements. There must be a compromise.

We can say that for future investigation of low-level interaction data, if a hypothesis-driven
approach is used, the focus can be on particular URL domains and web elements of interest, but when using a data-driven approach such as ours, apart from using user coverage to identify representative patterns, depending on the structure of the web platforms, focusing on a larger set of URL domains may be more beneficial for platforms that are more nested and linear with a limited number of interactive web elements, than Web platforms which are more spread out and much more interactive with a more diverse set of web elements than for which reducing the focus would potential be better.

We then added semantics to low-level events by creating the event-node-URL triple; this proved to be effective in improving the interpretability and descriptiveness of the events. Without the additional contextual information contained in the event-node-URL triples for each event, we would be unable to interpret and identify user behaviours during thematic analysis. Additionally, including contextual information strengthened the validity of the thematic analysis, as there is little room for alternative interpretation when these factors are included. On reflection, there may be instances where we can augment the triple with additional information. For example, we combined multiple recurring sequential events occurring within the same user session and added the suffix ‘multi’, this reduced noise and simplified interpretation. However, one could argue that the frequency of recurring events and the time intervals between them contain information. For instance, there are behavioural differences between users who click excessively and quickly. This may be an area for future applications to improve.

We then moved on to sequential pattern mining; this step was taken to identify user interaction patterns. Our intention was to begin by identifying the algorithm to use using a standard method for evaluating algorithms’ time and space efficiency. However, while the data size and complexity of the study on the university website are significantly larger and more complex due to the study’s much larger user base (more sequences), we benchmarked the algorithms using the MOOC dataset. We encountered no inefficiency issues with any of the datasets. Previous studies have noted that the performance of pattern mining algorithms varies significantly depending on the compiler used and environmental setup used to conduct performance comparisons [Goethals, 2003]. This benchmarking procedure could be applied to the selection of appropriate SPM algorithms for other datasets. We selected the VMSP algorithm because it is a maximal SPM algorithm that will be efficient in the subsequent process of looking for patterns of occurrences. Other algorithms, we discovered, may also be used, although steps to eliminate duplicate results will be required, as patterns may be subsets of others. Previous research has identified several difficulties associated with the use of sequential pattern mining algorithms [Poon et al., 2017]: 1) the generation of excessive patterns with limited relevance and utility, and 2) the use of domain experts to assist in filtering and labelling. For the first challenge, We developed criteria for dataset selection and algorithm parametrisation in order to maximise the informative and representative nature of the results, which will increase their relevance.

To address the second challenge, we proposed incorporating thematic analysis to systematically interpret the extracted user interaction patterns. By utilising this technique, we were
able to investigate higher-level user behaviour composed of low-level events. Another limitation of traditional sequential pattern mining algorithms is that their assumption that sequence databases are static. Indeed, conventional sequential pattern mining algorithms are referred to as batch algorithms because they are intended to be applied once to a sequence database in order to extract patterns. The algorithms must then be rerun if the database is updated to obtain the updated patterns. This approach is inefficient because minor changes to a database may occur that do not require re-running the pattern search. To address this issue, several incremental sequential pattern mining algorithms have been developed [Lin et al., 2015, Nguyen et al., 2005]. Another potentially beneficial extension of sequential pattern mining is top-k sequential pattern mining [Fournier-Viger et al., 2013]. It computes the k most frequently occurring sequential patterns. Using traditional sequential pattern mining algorithms, it is frequently difficult for users with no prior knowledge of a database to set the minimum support threshold. If the minimum support threshold is set too low, too many patterns may be discovered and the algorithms may become extremely slow, while setting it too high results in too few patterns being discovered. Top-k sequential pattern mining algorithms address this issue by allowing users to directly specify the number of patterns to discover, rather than through the minimum support parameter.

Thematic analysis is a widely used technique for analysing qualitative data. Throughout the two studies, we treated the extracted user interaction patterns as initial codes; however, in common qualitative analyses, the number of initial codes is likely to be smaller than in our study, where we had over 100 patterns. We reflect on the process; as discussed in Chapter 4, the patterns have little room for alternative interpretation; compared to common thematic analysis, our themes are more specific and less dependent on human judgement. This may contradict the theory of thematic analysis, but this then allows for scalability. If the number of the extracted interaction patterns is larger (e.g. 1000+), analysing them one by one can be time-consuming, and due to the fact that the interpretation of the patterns is more defined, it is possible that a set of interpreting rules can be developed in the initial stage, after familiarising and exploring the patterns (codes), with iterations of analysis and adding/updating when the new pattern (outside the rules) appears, supporting the scalability of this approach. Also, although the room for interpretation is limited, the level of abstraction and detail required for interpretation vary; for instance, at the end of the study in Chapter 5, we regrouped some of the themes from thematic analysis in order to deal with a sparse dataset. We excluded some of the elements from the themes, and certain URL domains were combined. Due to the level of detail that thematic analysis is focused on, the resulting themes can vary.

We created user behaviour matrices that contain the occurrences of user behaviours for each user in each session. On the basis of these occurrences, we computed features. We then use correlation for feature selection. This is a critical step as it reduces the dimensionality of the feature space; we used correlation in Chapters 4 and 5, and also used selectKbest in Chapter 5. Correlation was shown to be an effective method for selecting features to include in models, as the models we computed were acceptable. However, more advanced dimensionality reduction techniques exist and are widely used (e.g. Principal Component Analysis); these
techniques are capable of projecting data to lower-dimensional space (new data), making them more effective and contributing to more robust models; however, the interpretability is likely to suffer. As a result, simple techniques such as correlation may be more advantageous in preserving the inverse of interpretability from low-level user interactions. Additionally, we recognised the benefits of parameter selection in Chapter 5. In this chapter, we used GridSearchCV to perform hyperparameter selection which allows the algorithm to optimally solve the machine learning problem. This was also recognised in practice as we built and tested the models, where the models performed better with more appropriate hyperparameters. We also note that the features involving trends have a higher correlation and are included in many models, indicating that considering the temporal factor and devotion of user behaviours is important.

In Chapter 5, we note that our models are more accurate at predicting high levels of familiarity than they are at predicting low levels of familiarity. This may be due to the fact that the survey results were imbalanced; more data points belong to users with a high level of familiarity. However, another possibility worth noting is that given that user behaviours are influenced by familiarity [Dennis and Garfield, 2003, Chen et al., 2011], it is possible that users who reported a high level of familiarity behave differently from users who reported low-level of familiarity. There may be distinct behaviours observable within users who were not familiar with the website. However, through sequential pattern mining, we extract the patterns that are exhibited by the largest proportion of users. Users who are not as familiar with the website, as a minority group, may be overlooked in this process. As a result, the interaction patterns captured do not accurately represent the less familiar users. Nonetheless, the badge status in Chapter 4 was also imbalanced, while the models performed relatively well. This may be explained by the fact that the users in Chapter 4 represent a distinct demographic (early career researchers), and research has demonstrated that differences in background have an effect on user behaviour. User behaviours for users in Chapter 4 may be more similar than those in Chapter 5, as the School of Computer Science website is used by a variety of people. Additionally, as the MOOC platform is significantly less complex than the university website, the URL domains web elements have fewer variations, implying that user behaviour variations are more constrained, which may explain the models’ performance in Chapter 4. For Chapter 5, applying sequential pattern mining to each user group and analysing the common patterns may be a solution. Patterns exhibited by users with high and low levels of familiarity can be identified separately and analysed together. Alternatively, anomaly detection techniques may be considered to identify distinct interaction patterns of minority user groups.

In Chapter 6, we utilised segmentation of user interaction data to represent changes in search behaviour that occur throughout the user search journey. We grouped user sessions into periods, with each period representing a time interval and holding an equal proportion of search sessions. By analysing the clustering results for each period, we identified a trade-off between having more periods (allowing for user behaviour to be analysed more completely over time) and the number of sessions and interactions included in each period (with too few making it difficult to observe changes in behaviour). The more appropriate balance of this
may depend on the specific study. This number is not generalisable as it is dependent on the number of search sessions. We also note that when plotting Time vs Session (see Figure 6.3 and 6.4), the result is clusters which are easily distinguishable, whereas plotting Scroll vs Click and Keyword fails to show clear differences between the clusters. The figures show that the Keyword, Click, and Scroll for users in most clusters varied significantly. Such variation suggests that differences between clusters are not distinctively manifested by these three features. This presented potential problems for our analysis where we relied on statistics of these features to label each cluster in each period and did not control for search topic and user goal. It, therefore, becomes difficult to interpret user behaviours precisely, thus we think to suggest the inclusion of more features in the clustering model to increase the interpretability and accuracy of interpretation of the clusters. With too many features, however, this becomes a problem as observations become more difficult to cluster because too many dimensions cause every observation in the dataset to appear equidistant from all others, and because certain clustering algorithms, such as Kmeans, use a distance measure such as Euclidean distance to quantify the similarity between observations. If all of the distances are roughly equal, all of the observations appear to be equally similar (and equally different), and no meaningful clusters can be formed. By adding a pre-clustering step to the algorithm, spectral clustering avoids the curse of dimensionality: in this case, reducing dimensionality either by using PCA on the feature data or by modifying the clustering algorithm with 'spectral clustering.'

In conclusion, our proposed methodology addressed the issue of low descriptiveness and interpretability of low-level interaction data while demonstrating that, in practice, we can contribute to user models by identifying implicit user behaviour traces that are otherwise unobservable. We have shown that the methodology can be adapted and applied in different domains. Our method has a higher level of ecological validity when it comes to analysing wild search behaviours. By applying our methodology, for instance, website designers or online learning organisers will be able to take advantage of user low-level interaction data that is easily collectable, and generate insights into user behaviours, characteristics, and interaction outcomes. Additionally, we identified areas for future consideration and enhancements to specific steps in our methodology.

Our studies have limitations in terms of the extent to which our conclusions are generalisable, which is a common issue in user modelling research [Alonso-Mencía et al., 2020]. Due to the fact that web interface design, user individuality, and domain vary, conclusions drawn from one platform may not apply to others. Nonetheless, our methodology is applicable. Additionally, we note that the data size during the user modelling stage is small due to the fact that the cMOOC and search engine were targeting a specific audience, and the users who responded to the survey in Chapter 5 were small. This is mitigated by the fact that we collected large amounts of low-level data with numerous events.
7.3 Future research

We revisit the main recommendations and design implications of the proposed methodology for future research:

- Identify and independently analyse the events performed on different platforms as user behaviours can be affected by the devices/tools used.
- Include additional context information to the **event-node-URL** triple, such as the number of sequential re-occurring events and the time gaps between them.
- Employ data mining techniques to extract distinct user behaviours or conduct separate sequential pattern mining for each users group.
- Include more features in the clustering analysis to increase the interpretability of the cluster user behaviours.

We also note the future work recommendations for the three studies:

In Chapter 4, we discovered that user behaviour is associated with forum activity, that the less the forum is explored, the more likely students are to earn a badge, and that successful students were also more active in the forum. We hypothesised that when students were less active in the forum, they were engaged in other learning activities on the platform, and that as learners gain autonomy, their reliance on the forum decreases. This necessitates additional research into user forum activities and their relationship to learning properties. Additionally, we discovered that the learnability of the interface has an effect on learner achievement. The plausible relationship between increasing efficiency as a user and student achievement may imply that the more accessible a learning platform is to use, the more effectively students learn. It may also imply that students who are more familiar with the learning platform or who are skilled navigators are more adept at learning. Learnability is critical and perhaps a fundamental component of usability, but this has not been thoroughly discussed in online learning platforms. This creates new research opportunities for analysing platform learnability and student learning outcomes. We evaluated our user models by comparing them to models that included only engagement-related characteristics; in future research, it would be interesting to compare our methods to other assessment metrics; we can also learn whether our methods can provide appropriate context for assessment results.

By investigating low-level interactions, we were able to build user models that could indicate learner achievement. Online learning platforms, particularly cMOOCs, could use these models to develop browser extensions to capture interface interactions in real-time and predict student learning trajectories, providing feedback and guidance to students and course leaders. It can also be used for personalisation, to improve learner experience by analysing user behaviours to evaluate platform design, course material, and other properties of learning. The failure, however, to transition from exciting concept demonstrators to embedded practical
tools has long dogged educational technology [Scanlon et al., 2013]. Although the conclusions of our studies may lack generalisability, we believe that by demonstrating the proposed methodology, more platforms may recognise the value of low-level user interactions and the potential of their analysis.

In Chapter 5, we found that the frequent use of breadcrumbs over time indicated that users have difficulty interacting with the website. We surmised that the reason could be that users are using the breadcrumbs to support their navigation within the webpage, and that the increased usage of breadcrumbs may indicate usability issues of the site. This highlights the importance of future research into breadcrumbs, which are frequently used as navigational aids. As with Chapter 4, we believe that by implementing our methodology, websites can adapt to the unique needs of users who are more or less familiar with the website. Active adaptations to the user interface may be available to users who are most familiar with the website, allowing them to adjust their browsing needs independently and precisely. Other platforms, such as e-commerce platforms, may also benefit from investigating familiarity through low-level interactions in order to analyse user trust, loyalty, and perceptions.

The literature categorises search requirements as informational (e.g. searching for information), navigational (e.g. searching for the desired URL), or transactional (e.g. searching for sites that perform transactions) [Broder, 2002]. In Chapter 6, we briefly characterise users based on statistics of the features associated with these search needs categories. Additional research can be conducted to determine which features can more accurately identify users’ search needs. As search engines often do not collect ground truth about user goals [Scaria et al., 2014], search engine operators may find it beneficial to identify search needs based on user interactions in order to personalise search results and satisfy user information needs. We evaluated the impact of self-regulated learning on search skill evolution and found no evidence that MOOC participation provided support to user search skill improvement. We suspected that the MOOC martial relevancy and quality, as well as the level of participation, might have impacted this. This calls for further research into the effect of online learning on search skill evolution, as search skill is critical in modern times and can primarily benefit user information goal-achieving. Analysing the effect of online learning on search skills can help educators develop more effective strategies for increasing information literacy and research query techniques.

7.4 Conclusion

This thesis contributes towards the methodological and practical use of low-level user interactions in user modelling. We proposed a novel data-driven methodology that capitalises on the benefits of using low-level interaction while addressing its challenges. We demonstrated the methodology and its practical use of low-level data in three studies. Our work contributes to the research domain of learning analytics, Web familiarity and Search skill. Through the investigation of user behaviours, we discussed the findings and implications in these three domains. Additionally, we reflected on the proposed methodology and identified
potential challenges for future adoption considerations, as well as recommendations for future work in these three domains.
Bibliography


Naif Radi Aljohani and Hugh C. Davis. Learning analytics and formative assessment to provide immediate detailed feedback using a student centered mobile dashboard. IEEE, 2013. URL https://doi.org/10.1109%2Fngmast.2013.54.


Miguel L. Bote-Lorenzo and Eduardo Gómez-Sánchez. Predicting the decrease of engagement indicators in a MOOC. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK ’17*, pages 143–147, New York, New York,


Susan Bull. Supporting learning with open learner models. 01 2004.


Carlo Giovannella, Federico Scaccia, and Elvira Popescu. A pca study of student performance


Patrick Griffin and Esther Care. Assessment and teaching of 21st century skills: Methods and approach, 2014.


Lourdes Guàrdia Ortíz, Albert Sangrà Morer, and Marcelo Fabián Maina. Mooc design principles. a pedagogical approach from the learner”s perspective. 2013.

Julio Guerra, Shaghayegh Sahebi, Yu-Ru Lin, and Peter Brusilovsky. The problem solving genome: Analyzing sequential patterns of student work with parameterized exercises. 2014.


Jorge Maldonado, Mar Pérez-Sanagustín, Pedro Moreno-Marcos, Carlos Alario-Hoyos, Pedro Merino, and Carlos Delgado-Kloos. Predicting Learners’ Success in a Self-paced


Vivek Singh, Balaji Padmanabhan, Triparna de Vreede, Gert-Jan de Vreede, Stephanie Andel, Paul E. Spector, Steve Benfield, and Ahmad Aslami. A content engagement score for


Chenhao Tan, Evgeniy Gabrilovich, and Bo Pang. To each his own: Personalized content selection based on text comprehensibility. In Proceedings of the Fifth ACM International


Katrien Verbert, Nikos Manouselis, Hendrik Drachsler, and Erik Duval. Dataset-driven research to support learning and knowledge analytics. 2012.


Yao Xiong, Hongli Li, Mindy L Kornhaber, Hoi K Suen, Barton Pursel, and Deborah D Goins. Examining the relations among student motivation, engagement, and retention in a MOOC: A structural equation modeling approach., volume 2. ERIC, 2015.


Xiangmin Zhang, Michael Cole, and Nicholas Belkin. Predicting users’ domain knowledge from search behaviors. In *Proceedings of the 34th International ACM SIGIR Conference*
Appendices
Appendix A

Supplementary Material for Chapter 4

A.1 Features of learning and their corresponding indicators identified from previous works
Table A.1. Features of learning and their corresponding indicators identified from previous works.

<table>
<thead>
<tr>
<th>Features of learning</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement</td>
<td>✓</td>
</tr>
<tr>
<td>Engagement</td>
<td>✓</td>
</tr>
<tr>
<td>Performance</td>
<td>✓</td>
</tr>
<tr>
<td>Knowledge gain</td>
<td>✓</td>
</tr>
</tbody>
</table>

- Type of resource visited (e.g. forum, quiz) [Brooks et al., 2015b]
- Number of days active [Crues et al., 2018, Reeve and Tseng, 2011, Brooks et al., 2015b]
- The time duration for which student viewed the content [Guo et al., 2014, Singh et al., 2018, Cicchinelli et al., 2018, Motz et al., 2019, Yu et al., 2018b]
- Browser interactions (page loads, tab opens/closes switches, key presses, mouse moves, mouse clicks and drags) [Arapakis et al., 2014, Thomas et al., 2016, Yu et al., 2018b, Cicchinelli et al., 2018]
- Number/percentage of videos watched [Crues et al., 2018, Boté-Lorenzo and Gómez-Sánchez, 2017]
- Average time spent on assignments [Feild et al., 2018, Motz et al., 2019]
- Number of forum views [Anderson et al., 2014, Crues et al., 2018, Ramesh et al., 2014]
A.2 The MOVING MOOC platform

![Screenshot of the MOVING MOOC landing page.](image)

Figure A.1. Screenshot of the MOVING MOOC landing page.

A.3 Execution times of five SPM algorithms with different minimum supports

Table A.2. Execution times of five SPM algorithms with different minimum supports.

<table>
<thead>
<tr>
<th>Minimum support</th>
<th>CM-SPADE</th>
<th>CM-SPAM</th>
<th>CM-ClaSP</th>
<th>CloSpan</th>
<th>VMSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>612 ms</td>
<td>1225 ms</td>
<td>2288 ms</td>
<td>194 ms</td>
<td>1401 ms</td>
</tr>
<tr>
<td>0.5</td>
<td>211 ms</td>
<td>571 ms</td>
<td>423 ms</td>
<td>134 ms</td>
<td>571 ms</td>
</tr>
<tr>
<td>0.6</td>
<td>292 ms</td>
<td>534 ms</td>
<td>222 ms</td>
<td>98 ms</td>
<td>426 ms</td>
</tr>
<tr>
<td>0.7</td>
<td>67 ms</td>
<td>322 ms</td>
<td>72 ms</td>
<td>63 ms</td>
<td>306 ms</td>
</tr>
<tr>
<td>0.8</td>
<td>31 ms</td>
<td>379 ms</td>
<td>36 ms</td>
<td>50 ms</td>
<td>228 ms</td>
</tr>
<tr>
<td>0.9</td>
<td>22 ms</td>
<td>212 ms</td>
<td>25 ms</td>
<td>32 ms</td>
<td>207 ms</td>
</tr>
</tbody>
</table>

The CM-SPADE algorithm performed quicker than CM-SPAM for mining frequent patterns while CloSpan was fastest for mining closed patterns as shown in Table A.2. The execution times for each algorithm decreased as minimum support (and therefore the search space) was reduced. Overall, among the five algorithms, CloSpan was the quickest. Memory use varied a great deal depending on the values of minimum support and results regarding memory use were inconclusive.
A.4 Themes’ categories

Table A.3. Categories of themes

<table>
<thead>
<tr>
<th>Categories</th>
<th>(Sites/UI elements involved)</th>
<th>Number of themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homepage</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Week_1 page</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Forum</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Video</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Left sidebar</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Leave page</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Wiki</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Multiple pages</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

A.5 Logistic regression analysis results

Table A.4. Logistic regression analysis results

<table>
<thead>
<tr>
<th>Variables in the Equation</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$ (NumGap)</td>
<td>1.515</td>
<td>1.452</td>
<td>1.088</td>
<td>1</td>
<td>0.297</td>
<td>4.547</td>
</tr>
<tr>
<td>$X_2$ (S10mean)</td>
<td>-1.083</td>
<td>0.432</td>
<td>6.288</td>
<td>1</td>
<td>0.012</td>
<td>0.338</td>
</tr>
<tr>
<td>$X_3$ (NumGap)</td>
<td>-1.145</td>
<td>1.283</td>
<td>0.796</td>
<td>1</td>
<td>0.372</td>
<td>0.318</td>
</tr>
<tr>
<td>$X_4$ (NumGap)</td>
<td>-0.042</td>
<td>0.836</td>
<td>0.002</td>
<td>1</td>
<td>0.960</td>
<td>0.318</td>
</tr>
<tr>
<td>$X_5$ (NumGap)</td>
<td>-1.063</td>
<td>1.155</td>
<td>0.847</td>
<td>1</td>
<td>0.357</td>
<td>0.345</td>
</tr>
<tr>
<td>$X_6$ (NumGap)</td>
<td>2.404</td>
<td>0.882</td>
<td>7.433</td>
<td>1</td>
<td>0.006</td>
<td>11.070</td>
</tr>
<tr>
<td>Constant</td>
<td>-9.442</td>
<td>2.833</td>
<td>11.111</td>
<td>1</td>
<td>0.001</td>
<td>0.000</td>
</tr>
</tbody>
</table>

A.6 Second logistic regression analysis results

Table A.5. Second logistic regression analysis results

<table>
<thead>
<tr>
<th>Variables in the Equation</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$ (NumGap)</td>
<td>0.359</td>
<td>0.992</td>
<td>0.131</td>
<td>1</td>
<td>0.717</td>
<td>1.432</td>
</tr>
<tr>
<td>$X_3$ (NumGap)</td>
<td>-0.001</td>
<td>0.822</td>
<td>0.000</td>
<td>1</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>$X_4$ (NumGap)</td>
<td>-0.230</td>
<td>0.792</td>
<td>0.085</td>
<td>1</td>
<td>0.771</td>
<td>0.794</td>
</tr>
<tr>
<td>$X_5$ (NumGap)</td>
<td>-0.616</td>
<td>1.026</td>
<td>0.361</td>
<td>1</td>
<td>0.548</td>
<td>0.540</td>
</tr>
<tr>
<td>$X_6$ (NumGap)</td>
<td>1.845</td>
<td>0.546</td>
<td>11.424</td>
<td>1</td>
<td>0.001</td>
<td>6.327</td>
</tr>
<tr>
<td>Constant</td>
<td>-6.520</td>
<td>1.390</td>
<td>22.018</td>
<td>1</td>
<td>0.000</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Appendix B

Supplementary Material for Chapter 5

B.1 Number of responses for the survey questions at each point

Table B.1. Number of responses for the survey questions at each point

<table>
<thead>
<tr>
<th></th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>89</td>
<td>52</td>
<td>105</td>
<td>88</td>
<td>73</td>
</tr>
<tr>
<td>4</td>
<td>114</td>
<td>117</td>
<td>99</td>
<td>108</td>
<td>102</td>
</tr>
<tr>
<td>3</td>
<td>46</td>
<td>70</td>
<td>51</td>
<td>51</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>20</td>
<td>11</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>9</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

B.2 Selection of URL domains

Figure B.1. URL domains with occurrences within each percentile and the overall user coverage of each group
Table B.2. Activity patterns after merging.

<table>
<thead>
<tr>
<th>Index</th>
<th>Activity patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Explore the about us page</td>
</tr>
<tr>
<td>2</td>
<td>Explore the our research page</td>
</tr>
<tr>
<td>3</td>
<td>Explore the postgraduate research/taught page</td>
</tr>
<tr>
<td>4</td>
<td>Explore the postgraduate research/taught page with scrolling</td>
</tr>
<tr>
<td>5</td>
<td>Explore the professional development page</td>
</tr>
<tr>
<td>6</td>
<td>Explore the undergraduate page</td>
</tr>
<tr>
<td>7</td>
<td>Explore the undergraduate page breadcrumbs and sidebar area</td>
</tr>
<tr>
<td>8</td>
<td>Explore the undergraduate page breadcrumbs area</td>
</tr>
<tr>
<td>9</td>
<td>Explore the undergraduate page breadcrumbs, navigation, and sidebar area</td>
</tr>
<tr>
<td>10</td>
<td>Explore the undergraduate page navigation and sidebar area</td>
</tr>
<tr>
<td>11</td>
<td>Explore the undergraduate page with scrolling</td>
</tr>
<tr>
<td>12</td>
<td>Explore the Webpage breadcrumbs and navigation area</td>
</tr>
<tr>
<td>13</td>
<td>Explore the Webpage navigation area</td>
</tr>
<tr>
<td>14</td>
<td>Explore the Webpage sidebar area</td>
</tr>
<tr>
<td>15</td>
<td>Explore the homepage</td>
</tr>
</tbody>
</table>

B.3 Activity patterns after merging
mouseinorout++postgraduate_research+multi
mousemove++postgraduate_research
mousemove+sidebar+postgraduate_research
30
mousemove+MainBody+homepage
mouseinorout+MainBody+homepage+multi
mousemove+MainBody+homepage+multi
mouseinorout+MainNavigation+homepage+multi
mousemove+MainNavigation+homepage
31
mouseinorout++postgraduate_research
mouseinorout++postgraduate_research+multi
mouseinorout+MainBody+postgraduate_research
mousemove++postgraduate_research
mousemove++postgraduate_research+multi
mousemove+MainBody+postgraduate_research
32
mouseinorout+MainBody+undergraduate
mouseinorout+MainBreadcrumbs+undergraduate+multi
mousemove+MainBody+undergraduate
mousemove+MainBreadcrumbs+undergraduate+multi
mouseinorout+MainNavigation+undergraduate+multi
mousemove++undergraduate+multi
mousemove++undergraduate
mousemove++undergraduate+multi
33
mouseinorout+MainBody+postgraduate_taught+multi
mousemove+MainBody+postgraduate_taught
mousemove+MainBody+postgraduate_taught+multi
34
mousemove+MainBody+professional_development+multi
mousemove+MainBody+professional_development
Appendix C

Supplementary Material for Chapter 6

C.1 User self-reported competency levels

Figure C.1. User self-reported competency levels
C.2 KMeans clustering with two periods

Figure C.2. KMeans clustering with three periods (displayed sequentially from top to bottom), the figures on the left show the features Scroll and Click, the figures in the middle show Time and Session, and the figures on the right show Keyword with each axis presented on the same scale. Each user is represented as a data point and coloured based on their corresponding cluster.
Figure C.3. KMeans clustering with five periods (displayed sequentially from top to bottom), the figures on the left show the features Scroll and Click, the figures in the middle show Time and Session, and the figures on the right show Keyword with each axis presented on the same scale. Each user is represented as a data point and coloured based on their corresponding cluster.