Downsampling wearable sensor data packets by measuring their information value

Document Version
Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Downsampling wearable sensor data packets by measuring their information value

Miguel A. G. Belmonte
and Alexander J. Casson
School of Electrical and Electronic Engineering,
University of Manchester, UK
Email: {miguel.belmonte, alex.casson}
@manchester.ac.uk

Niels Peek
School of Health Sciences,
University of Manchester, UK
Email: niels.peek@manchester.ac.uk

Abstract—**Long-Short Term Memory models (LSTMs) are data-driven routines that classify Human Activity Recognition (HAR) with minimum human input. The price to pay for analysing large sequences of on-body sensor measurements with LSTMs are high processing power and battery requirements. In this paper, we recognize that sensor data packets have differing information value to classify HAR and propose to quantify it with cross entropy (CrossEn), Kullback Leibler (KL) divergence and sample entropy (SampEn). Both, CrossEn and SampEn have the potential to guide dropping redundant data packets without compromising HAR. However, we do not find substantial improvements in dropping rates when downsampling by CrossEn and SampEn over computationally cheaper random and uniform alternatives. Our results show that the KL divergence, evaluated at training time is equivalent to the classification accuracy criteria that involves a testing set. The computational requirements to compute the KL in real-time could well guide sensor node design to downsample wearable measurements *near* the user.**

Index Terms—**Information value, deep learning, downsampling, classification accuracy, data packets.**

I. INTRODUCTION

Human activity recognition (HAR) has many applications in healthcare [1], [2]. Application of *deep learning* to the problem of HAR classification using wearable sensor measurements yields impressive recognition accuracies [1]–[4]. Deep learning models offer practitioners flexible data-driven routines that bypass manual feature selection procedures [5]. Long-Short Term memory models (LSTMs), an instance of deep learning algorithms are formulated around recurrent neural networks (RNNs) that incorporate memory cells capable of storing information over long time periods [6]. As such, LSTMs effectively capture long-term temporal dependencies within the movement data and have been applied to HAR classification with promising results in [4], [7]–[9].

The downside of training LSTMs is high processing and battery power. Fitting several layers of stacked RNNs [10] demands prohibitive memory and computational power. Transmission of large sequences of wearable sensor data, away from the device to a processing unit is the biggest source of energy consumption [11]. In [12] a design around online signal processing is suggested to optimise the operating lifetime of the device. Different approaches that aim to reduce the dimensionality of a fixed batch of sensor measurements have

been proposed. For example, [13]–[15] apply compression methods to electrocardiogram signals. Recent work investigated downsampling data packets as a method for reducing the transmission data rate without compromising machine learning classification accuracy. In [16], large amounts of redundant accelerometer data packets were found, which motivated the approach in [17] to uniformly downsampling accelerometer data in both sampling frequency and bit depth. A reduction of recorded measurements by three orders of magnitude improved battery life from weeks to years without compromising classification accuracy [17]. We investigate the impact on HAR classification performance of downsampling wearable sensor data before transmission. As an alternative to random and uniform downsampling, we propose reducing redundant sensor data packets based on quantifying their information value. High-frequency sensor measurements on human movement shows strong temporal dependence [9]. It is expected that conditional on most recent history, a participant will continue on the same activity for the next sensor measurement, except at *unexpected* switching times. We recognize that sensor data packets recorded during the same activity have less information value for HAR classification than those containing *surprisal* transitions between activities. We resort to cross entropy (CrossEn), the Kullback Leibler (KL) divergence and a multiscale sampling entropy (SampEn) estimate to evaluate the information content of sensor data packets. CrossEn is a measure of surprise and involves evaluating the cost function for every data packet that comprises the training set. The KL divergence quantifies the information loss incurred after dropping data packets in the training set. The information loss is measured by the dissimilarity between the discrete distribution characterised by the LSTM classifier applied first over a full training set and a successively downsampled training set. The SampEn quantifies the dynamic complexity of each data packet in the training set by evaluating a multiscale sample entropy estimate [18].

Both, random and uniform downsampling ignore dynamical properties of packet sensor measurements and are considered to benchmark the downsampling performance of the alternatives proposed. The comparison with CrossEn and SampEn is in terms of classification accuracy and KL evaluation. The aim
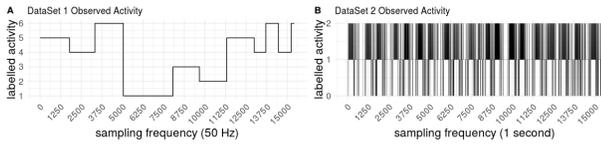
Figure 1: Paths of activity labels

Figure 2: Cross entropy estimates

of the comparison, is to find out which downsampling scheme achieves the best rate of downsampling while maintaining the classification performance of HAR. Complementary to comparing the four downsampling schemes, we investigate whether the KL divergence computed at training time is comparable to the classification accuracy criteria that requires a testing set for evaluation.

The remainder of this paper is organised as follows. Section II overviews the three measures of evaluating information content in the sensor records and the implementation of downsampling schemes. Section III compares the dropping rates in terms of classification accuracy and KL after applying the downsampling methods to two datasets collected under different conditions. Section IV summarises findings and concludes.

II. METHODS

A. Data collection

We analyse two datasets, henceforth Dataset 1 and Dataset 2. Dataset 1 analysed in [19], contained 661056 observations from 20 participants on activity labels collected at 50 Hz sampling frequency. Six labelled activities were encoded: 0-walk on level surface; 1-walk upstairs; 2-walk downstairs; 3-sitting; 4-standing; 5-lying. Sensor measurements were recorded via a waist mounted inertial sensor that consisted on two sensing modalities, accelerometer and gyroscope. The six-dimensional sensor readings were reshaped into 5163 packets of 128 samples. Dataset 2 used in [20] contained 15421 observations from 10 participants on activity labels annotated by a team of observers every second. Three label activities were encoded: 0-ambulation activity; 1-static posture; 2-posture to posture transitions. Sensor measurements were recorded with a sampling frequency of 20 Hz via a wrist-worn tri-axial accelerometer. The three-dimensional sensor time series were reshaped into 481 packets of 32 measurements.

Dataset 1 and Dataset 2 were collected under different setups. Dataset 1 under controlled conditions for fixed movement in a protocol with balanced classes. Dataset 2 was obtained from participants under living conditions. A path of activity labels for both datasets visually illustrates in Fig. 1 their differences in activity persistence.

B. LSTM model classifier

The strength of LSTMs for HAR classification lies in an architecture capable of retaining information for hundreds of time periods [6]. Given differences in temporal persistence in activity labels between Dataset 1 and Dataset 2, 127 and 31 previous samples were arranged to work as memory. To avoid over-fitting, 20% of the training set was held out as a validation
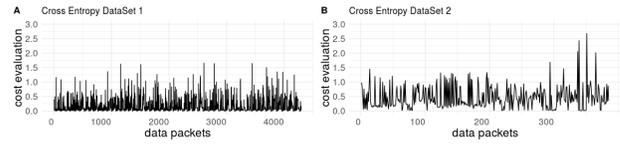
set. For Dataset 1, 4296 packets of 128 samples (LSTM 128) were considered for training and 867 for testing, whereas for Dataset 2 the split was 385 packets of 32 samples (LSTM 32) for training and 96 held out for testing. Further details on the computational implementation in Python [21] are detailed in [4].

The baseline classifying performance of the LSTM model applied on all available data before undertaking downsampling was assessed with a *10-fold* cross validation test harness [22]. Classification accuracy was on average $90.70\% \pm 1.53\%$ for Dataset 1 and $80.90\% \pm 3.39\%$ for Dataset 2.

C. Measurements of information value

The measurements of information value considered to evaluate the information content of sensor data packets were CrossEn, the KL divergence and multiscale SampEn.

The CrossEn is considered a measure of surprise between expected and observed activity [23]. Computing the *surprisal* effect involves evaluating the cost function for every data packet that comprises the training set, at linear computational cost. Cross entropy estimates for Dataset 1 and Dataset 2 are shown in Fig. 2. A mean value of 0.14 with associated standard deviation of 0.19 and mean of 0.46 with standard deviation of 0.40 for Dataset 2 indicate that the LSTM model finds the data packets in Dataset 1 less surprising than those of Dataset 2.

A related measure of information value is the KL divergence, introduced in [24]. We apply the KL to quantify, at linear computational cost, the information loss after each batch of 5% of data packets is dropped. The information loss is quantified by evaluating the dissimilarity between the discrete distribution characterised by the LSTM classifier applied on the full training set and the increasingly downsampled distribution.

The multivariate multiscale SampEn, introduced in [18] measures the dynamic complexity in physical systems that range between perfect regularity and complete randomness over multiple time scales. Small values of sample entropy are assigned to deterministic and uncorrelated signals, whereas large values are ascribed to correlated signals. To the best of our knowledge [18], [25]–[27] this is the first time that SampEn is considered to guide downsampling sensor packets of data. Results in [28] show the robustness of SampEn estimates for different sampling frequencies. For electrocardiogram readings over a range of 125 Hz - 2000 Hz frequencies, SampEn shows unaffected estimates. The computational cost of appending SampEn estimates to data packets is quadratic in the length of the sensor record [29].
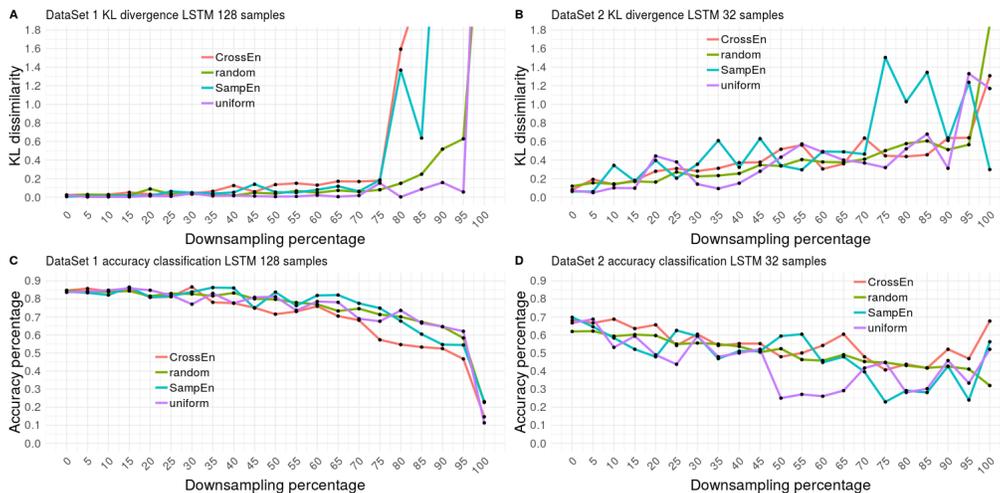
Figure 3: Comparison of KL and classification accuracy for all the downsampling schemes

D. Downsampling schemes

Four downsampling schemes were considered. Random downsampling indexed data packets with random numbers (between $1-4296$ for Dataset 1 and between $1-385$ for Dataset 2) from a uniform distribution. Uniform downsampling removed data packets starting from the beginning of the training set. CrossEn and SampEn downsampling removed data packets by quantiles that divide the range of CrossEn and SampEn estimates in epochs of 5%, starting from the lowest value. CrossEn, SampEn and the uniform scheme dropped data packets deterministically. Random downsampling required 50 runs to average the uncertainty out in the estimate of classification accuracy.

III. RESULTS

This section overviews the effect on classification accuracy and KL of downsampling data packets in epochs of 5% in the training set for Dataset 1 and Dataset 2.

Low rates of downsampling have negligible effects on classification accuracy and KL across all the schemes. The first big change in classification accuracy and KL marks the percentage of redundant packet readings; 75% for Dataset 1 and 70% for Dataset 2. For Dataset 1, homogeneity in persistence between data packets in the training and testing set imply that further downsampling does not compromise HAR classification.

The baseline classifying performance of the LSTM model was preferably established with a 10-fold cross validation test harness (section II-B). A single training and testing data split is shown in Fig. 3 for a 0% downsampling rate. Differences in classifying performance between datasets are due to their differing temporal persistence in activity labels [9]. Data packets in Dataset 1, collected under sustained participant activity show homogeneous persistence over long time periods and contain the same number of rare switching in both, the training and testing set. In contrast, data packets in Dataset 2 collected under living conditions contained differing number of transitions, short durations in activity and the temporal persistence across the sample is vastly different. These dif-

ferences in dependence across data packets between the two datasets help explaining the distinct pattern that successive downsampling introduces in classification accuracy and KL. Smooth and steady for Dataset 1, jagged and irregular with respect to previously downsampled batch for Dataset 2.

The effect of dropping the last 5% batch on classification accuracy also differs between datasets. For Dataset 1, preventing any training of the LSTM, reduces the classification accuracy to less than chance at 10%. For Dataset 2, training and testing data packets are so dissimilar, that lack of training of the LSTM actually increases classification accuracy to 50%.

Computing estimates of CrossEn and SampEn showed that every data packet had different information content. However, we did not find evidence that pre-processing data packets with computation of CrossEn and SampEn clearly outperformed random and uniform schemes. Visual comparison of KL and classification accuracy in Fig. 3 showed an inverse relationship between the two. This suggests evaluating the KL near the user to guide the dynamic decision of whether to keep a data packet.

IV. CONCLUSION

On-body sensor datasets collected under different protocols are difficult to compare in classification accuracy. Large percentage of accelerometer and gyroscope readings are redundant. There is ample scope to improving battery performance at the wearable itself. Pre-processing sensor measurements with CrossEn and SampEn does not offer an advantage in terms of classification accuracy and KL over computationally cheaper downsampling schemes. The KL divergence has shown a comparable measure of information loss with the classification accuracy. The computational requirements to processing sensor measurements and computing the KL in real-time could well guide sensor node design to downsample data packets on the sensor device, *near* the user.

REFERENCES

[1] A. Murad and J. Y. Pyun, "Deep recurrent neural networks for human activity recognition", *Sensors*, vol. 17, no. 11, p. 2556, 2017.

[2] J. Wang, Y. Chen, S. Hao, *et al.*, "Deep learning for sensor-based activity recognition: A survey", *Pattern Recognition Letters*, Jul. 2017.

[3] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors", *ACM Comput. Surv.*, vol. 46, 33:1–33:33, 2014.

[4] T. Zebin, N. Peek, and A. Casson, "Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks", in *IEEE EMBC*, Apr. 2018.

[5] T. Plötz, N. Y. Hammerla, and P. Olivier, *Feature learning for activity recognition in ubiquitous computing*, 2011.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural computation*, vol. 9, pp. 1735–80, Dec. 1997.

[7] F. Li, K. Shirahama, M. A. Nisar, *et al.*, "Comparison of feature learning methods for human activity recognition using wearable sensors", *Sensors*, vol. 18, 2018.

[8] F. J. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition", *Sensors*, vol. 16, no. 1, 2016.

[9] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables", in *ACM IJCAI*, New York, Jul. 2016.

[10] C. Laurent, G. Pereyra, P. Brakel, *et al.*, "Batch normalized recurrent neural networks", in *IEEE ICASSP*, Shanghai, Mar. 2016.

[11] A. Godfrey, V. Hetherington, H. Shum, *et al.*, "From A to Z: Wearable technology explained", *Maturitas*, vol. 113, pp. 40–47, 2018.

[12] G. Chen, E. Rodriguez-Villegas, and A. J. Casson, "Wearable algorithms: An overview of a truly multi-disciplinary problem", in *Wearable Sensors*, E. Sazonov and M. R. Neuman, Eds., Oxford: Academic Press, 2014, pp. 353–382.

[13] T. Yousefi Rezaii, S. Beheshti, M. Shamsi, *et al.*, "ECG signal compression and denoising via optimum sparsity order selection in compressed sensing framework", *Biomed. Signal Process. Control.*, vol. 41, pp. 161–171, Mar. 2018.

[14] R. Kumar, A. Kumar, and G. Singh, "Electrocardiogram signal compression based on singular value decomposition (SVD) and adaptive scanning wavelet difference reduction (ASWDR) technique", *AEU - International Journal of Electronics and Communications*, vol. 69, no. 12, pp. 1810–1822, 2015.

[15] M. S. Manikandan and S. Dandapat, "Wavelet-based electrocardiogram signal compression methods and their performances: A prospective review", *Biomed. Signal Process. Control.*, vol. 14, pp. 73–107, 2014.

[16] A. Khan, N. Hammerla, S. Mellor, *et al.*, "Optimising sampling rates for accelerometer-based human activity recognition", *Pattern Recognition Letters*, vol. 73, pp. 33–40, 2016.

[17] X. Fafoutis, L. Marchegiani, A. Elsts, *et al.*, "Extending the battery lifetime of wearable sensors with embedded machine learning", in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, Feb. 2018, pp. 269–274.

[18] M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy: A tool for complexity analysis of multichannel data", *Phys. Rev. E*, vol. 84, p. 061 918, 6 Dec. 2011.

[19] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Evaluation of supervised classification algorithms for human activity recognition with inertial sensors", in *IEEE Sensors conf.*, Glasgow, Nov. 2017.

[20] N. Twomey, T. Diethe, M. Kull, *et al.*, "The SPHERE challenge: Activity recognition with multimodal sensor data", *CoRR*, vol. abs/1603.00797, 2016.

[21] F. Chollet. (2013). Keras: The python deep learning library.

[22] Y. Bengio and Y. Grandvalet, "No unbiased estimator of the variance of K-fold cross-validation", *J. Mach. Learn. Res.*, vol. 5, pp. 1089–1105, 2004.

[23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[24] S. Kullback and R. A. Leibler, "On information and sufficiency", *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[25] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series", *Phys. Rev. Lett.*, vol. 89, p. 068 102, 6 2002.

[26] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy", *Am J Physiol Heart Circ Physiol*, vol. 278, no. 6, H2039–H2049, 2000.

[27] D. E. Lake, J. S. Richman, M. P. Griffin, *et al.*, "Sample entropy analysis of neonatal heart rate variability", *Am J Physiol Regul Integr Comp Physiol*, vol. 283, no. 3, R789–R797, 2002.

[28] B. Singh, M. Singh, and V. K. Banga, "Sample entropy based HRV: Effect of ECG sampling frequency", *Biomed. Sci. and Eng.*, vol. 2, no. 3, pp. 68–72, 2014.

[29] M. Aktaruzzaman, "Low computational cost for sample entropy", *Entropy*, vol. 20, Jan. 2018.