Improving kNN for Human Activity Recognition with Privileged Learning using Translation Models

Anjana Wijekoon[1][0000−0003−3848−3100] ✉, Nirmalie Wiratunga[1][0000−0003−4040−2496], Sadiq Sani[1][0000−0001−9784−8398], Stewart Massie[1][0000−0002−5278−4009], and Kay Cooper[2][0000−0001−9958−2511]

[1] School of Computing Science and Digital Media, Robert Gordon University, Aberdeen AB10 7GJ, Scotland, UK
[2] School of Health Sciences, Robert Gordon University, Aberdeen AB10 7GJ, Scotland, UK
{a.wijekoon, n.wiratunga, s.sani, s.massie, k.cooper}@rgu.ac.uk

Abstract. Multiple sensor modalities provide more accurate Human Activity Recognition (HAR) compared to using a single modality, yet the latter is preferred by consumers as it is more convenient and less intrusive. This presents a challenge to researchers, as a single modality is likely to pick up movement that is both relevant as well as extraneous to the human activity being tracked and lead to poorer performance. The goal of an optimal HAR solution is therefore to utilise the fewest sensors at deployment, while maintaining performance levels achievable using all available sensors. To this end, we introduce two translation approaches, capable of generating missing modalities from available modalities. These can be used to generate missing or "privileged" modalities at deployment to augment case representations and improve HAR. We evaluate the presented translators with k-NN classifiers on two HAR datasets and achieve up-to 5% performance improvements using representations augmented with privileged modalities. This suggests that non-intrusive modalities suited for deployment benefit from translation models that generates privileged modalities.

Keywords: Human Activity Recognition · Machine Learning · Case representation · Privileged Learning

1 Introduction

Human Activity Recognition (HAR) involves the computational analysis of human movement. The types of movement which are recognised are directly dependent on the application requirements. Typically these applications relate to tracking or monitoring movements such as ambulatory activities (i.e. walking or jogging) [9, 11], daily activities of living (i.e. gardening or cooking) [1] or exercises (i.e. muscle strength increasing exercises or stretching) [12]. In these situations we would expect to use sensing devices comprised of wearables (inertial sensors

such as an accelerometer or a gyroscope) and ambient sensors in the environment (such as movement sensors in a home) to track user activity.

Reasoning with multi-modal sensor data is an active area of AI research [17] with applications fielded in a range of domains, including health and well-being, smart cities, robotics and interactive natural interfaces. For HAR having different modalities for sensing is advantageous as it provides contextually richer representations. However access to all sensor modalities at deployment can be restricted due to a variety of reasons. Economics in some situations will limit the number of available sensors; erroneous behaviour may cause loss of data temporarily or ease of use may restrict the number of sensors one may be willing to use. In short, considerations such as usability, ease of deployment and cost all suggest that access to data from all modalities is likely to be a privilege to be had at training, and not necessarily at deployment (test time). This poses an interesting question of how representations learnt using all modalities at train time can also be exploited at test time. Here instead of simply ignoring missing modalities at test time we explore how performance gains can be achieved by learning to estimate them.

In this paper we focus on HAR in the context of Privileged Learning (PL) [14]. Specifically we show how PL can be used to estimate missing parts of a representation when one or more modalities are absent at test time. The key idea is to learn a generative model that can use existing modalities to estimate representations for any missing modalities. An initial study on linear correlation between modalities has as expected revealed that a simpler technique such as a linear regression is ineffective at estimating missing modalities. Our solution borrows ideas from computational language translation [13], but instead of translating between language pairs, we translate between data generated by sensor modalities - from present to missing modalities. The main assumption here is that there is a non-linear correlation between modalities and that we can discover them from a parallel corpus of modality pairs using translators. Accordingly we make the following three contributions:

- formalise PL in the context of HAR by recognising how different modalities contribute towards improved classification;
- introduce novel translation methods that can learn a mapping between sensors to estimate missing modalities at deployment; and
- conduct a comparative study of the proposed algorithms on the SelfBACK[3] and PAMAP2 [4] datasets to demonstrate their ability to achieve improved performance with fewer modalities at deployment.

This paper is organised as follows: in Section 2 we explore work related on HAR, PL and Sequence generation; in Section 3 we interpret Privileged Information (PI) in the domain of HAR and offer formalisations for our approaches.

[3] The SelfBACK project is funded by European Union's H2020 research and innovation programme under grant agreement No. 689043. More details available: http://www.selfback.eu. The SelfBACK dataset associated with this paper is publicly accessible from https://github.com/selfback/activity-recognition

[4] https://archive.ics.uci.edu/ml/datasets/PAMAP2+Physical+Activity+Monitoring

We detail the datasets, experiment design and evaluation techniques in Section 4; in Section 5 we present results and discuss outcomes; followed by conclusions and future improvements in Section 6.

2 Related Work

Significant research has been carried out on reasoning with sensors for HAR using machine learning techniques. While early work was focused on using a single sensor to perform HAR with hand crafted features [6], more recent advancements are largely due to the successes of deep learning. Much of the latest research has focussed on exploiting multiple sensors for HAR with deep learning models to achieve state of the art performance [16, 8]. In [9] the authors explore the impact of different sensor placements on HAR performance and discuss the trade-off between convenience (wrist placement) versus accuracy (thigh placement). Ideally we want to optimise sensor placement convenience whilst minimising the negative impact this can have on accuracy.

Privileged Learning (PL) mimics how humans learn with a teacher. In a learning environment the teacher provides the student with explanations and additional information around the topic, but at test time the student must rely on what they have learned with no access to the teacher. This concept was introduced by [14] where they define an additional feature space, Privileged Information (PI) that guarantees 100% classification accuracy, but only available at training. We can draw parallels here in sensor placements; whereby sensors that lead to improved performance but not considered to be convenient placements are analogous to the teacher in PL. However unlike with PL, a privileged sensor placement can only promise positive improvements.

In this paper we explore how an additional PI space can be constructed for HAR. Typically PI can be viewed as an extra set of features describing the same problem. For example, using additional image masks to influence improved orthogonality in convolutional functions for image classification [2] or the use of skeleton information to improve depth sequence analysis [10]. In the latter paper, the authors demonstrate a system capable of learning to generate privileged (skeleton) information from training data (depth sequences) which can then be used to support classification at test time. Similarly, in our work we generate a PI feature space from existing sensors but use both feature spaces to improve HAR. However our translation model is a reusable standalone component which translates between sensor data compared to [10] where skeleton generation is continuously refined with classification.

Sequence to sequence (seq2seq), learning has been successfully applied in many domains, such as image/video captioning [18, 15], language translation [3] and time-series forecasting [4] with Recurrent Neural Networks (RNNs). We also see Sequence generation with Deep Belief Networks (DBNs) and Deep Autoencoders applied successfully in audio and video reconstruction [5]. Learning to reconstruct missing sensor data is similar to sequence generation, but as we focus on a small window of time, there are less temporal dependencies to be learnt.

The mapping between input and output data in an autoencoder is more relevant to our work. Unlike with autoencoders our input and output is not meant to be identical - instead they involve different sensor streams aligned in time. As such we focus more on learning a mapping between different sensor data to capitalise on their spatial dependencies.

3 HAR with Privileged Learning

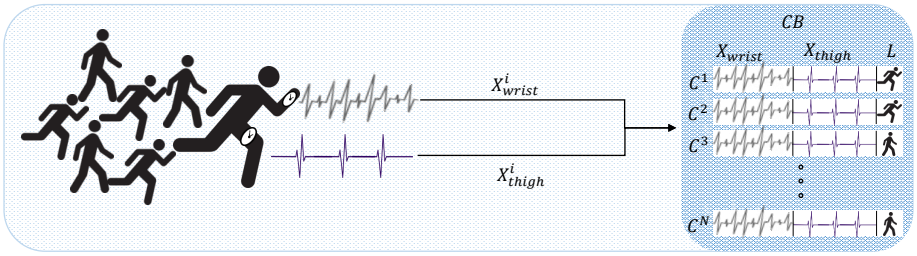

Fig. 1. HAR casebase creation

Figure 1 illustrates how we create the casebase for HAR task from a sample of people wearing two modalities (i.e. wrist and thigh). We use a sliding window approach, with a window size of w, to decompose data streams from each modality. Accordingly the case representation, $C = \{X_1, X_2, X_3, ..., X_n, L\}$, captures all n modalities together with the activity class label L at each time window, where X_i is the i_{th} sensor modality.

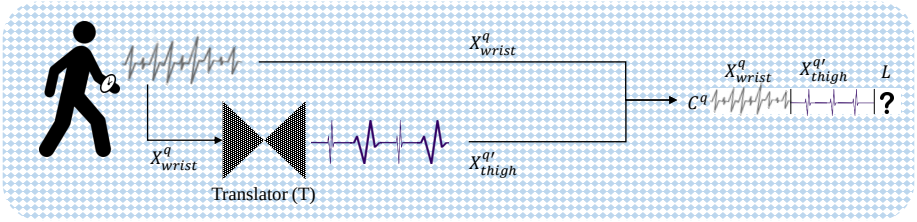

Fig. 2. Case creation at deployment

Unlike with Figure 1, in Figure 2 at deployment the user is wearing a single modality (a wrist sensor). We use a translator to estimate the missing modality, using the data that is present. Thereafter both the present modality X^q_i and the estimated modality $X^{q'}_j$ are used to form the query case C^q. Here $X^{q'}_j$ forms our privileged information and the translation model is simply a mapping between

the input and output modalities. In our example (Figure 2) the translation model can be learnt from a parallel corpus of wrist-to-thigh instances. We use cases from our casebase to learn the translation model, as each case contains all potential modalities. More generally, this mapping can be between any number of input and output modalities.

3.1 A Privileged Classification Model with Translators

In this section we formalise classification with privileged learning for HAR. We consider privileged information in HAR as a set modalities that is present at casebase creation but missing at deployment. The HAR classifier receives n number of modalities as input to predict an activity class. Given a query case $C^q = \{X_1, X_2, ..., X_m\}$, where m is the number of modalities present at deployment, we determine the missing modalities as $n - m$. We then use one or more translators, T, to generate those missing modalities.

$$\chi' = T(\chi)$$

where $\chi \subset X$ and $\chi' \subset X'$. Here X denotes the set of modalities present at deployment and X' is the privileged information generated by translators for all missing modalities.

In this way, we augment the representation of the query case, using generated modalities to create the representation expected by the HAR classifier. Accordingly, the augmented query has the following representation:

$$C'^q = \{X_1, X_2, ..., X_m, X'_{m+1}, X'_{m+2}, ..., X'_n\}$$

$$C'^q = \{X, X'\}$$

In the rest of this section we describe two translation methods that can generate the missing modalities, $n - m$, from the m modalities.

3.2 k-Nearest Neighbour Translator

In this approach, the PI is generated for a query case by exploiting similarity based retrieval and solution reuse. Given the query case C^q and a case C from the casebase, we calculate their paired difference as follows:

$$Distance_{(C^q,C)} = \sum_{i=0}^{m} \delta(X_i^q, X_i)$$

where δ calculates the distance between a pair of modalities.

The top, k, cases are retrieved and their solutions (i.e. PI) is reused to estimate the missing modalities values in C^q. More specifically, we reuse the values from the privileged information attributes X_i as taken from the k nearest neighbours and average over k to estimate the privileged attribute X'_i.

$$X'_i = \frac{1}{k} \sum_{j=1}^{k} X_i^j$$

We iterate over all privileged modalities to form an augmented representation, C'^q, for the query case. This k-Nearest Neighbour Translator will be referred to as T^{kNN}.

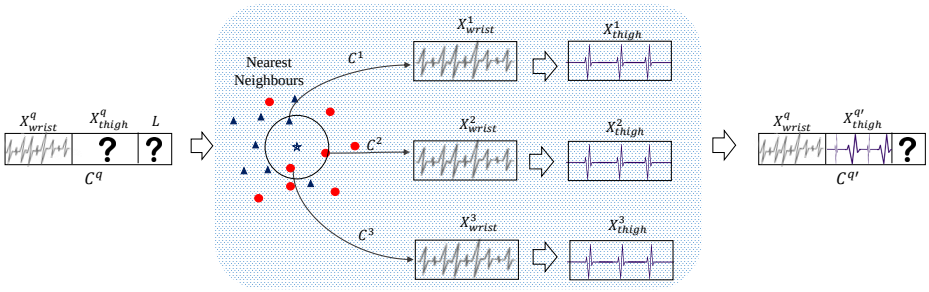

Fig. 3. An example wrist-to-thigh Nearest Neighbour Translator

Figure 3 illustrates an example T^{kNN} translator which retrieves the first 3 nearest neighbour cases from the casebase using the wrist modality attributes of the query case. The thigh attribute modalities of the retrieved cases are averaged to form an estimate thigh attribute for the query case. In this way an augmented representation is formed by combining the estimated thigh modality with the initial wrist modality.

3.3 Neural Translator

We use a fully connected neural network to generate privileged information; where it learns a neural mapping between its input and output layers. Here the input layer consists of features representing modalities that are present only at test time and the output estimates the missing modalities.

More specifically we have, $p * w$, input units where p is the number of input modalities and w is the window size and the output layer consists of units from a subset of missing modalities, $q * w$ where q is the number of output modalities. A single hidden layer is introduced to learn the feature mapping from input to the output units. We propose to use a narrow middle layer to force the network to learn the most significant features from the input when estimating its output. This also helps avoid learning arbitrary noisy features from the input.

Figure 4 illustrates an example Neural Translator training using a single input modality (i.e. wrist) to generate another single modality (i.e. thigh). For training we use a parallel corpus of wrist-thigh pairs where wrist is input, and thigh is the solution that is being estimated by the network. The figure also indicates the node activation and loss functions expressions used for model training.

Let X_H denote the hidden layer representation of the input X_{wrist} and calculated with weights W_H and biases b_H on the hidden layer. The network derives X'_{thigh} using the hidden layer representation X_H and weights W_O and biases

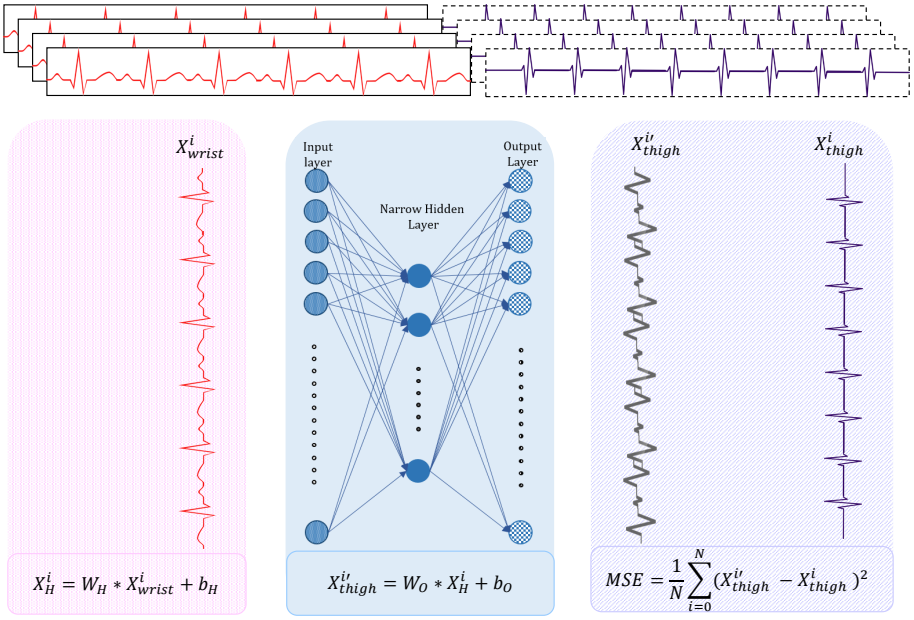

Fig. 4. An example wrist-to-thigh Neural Translator

b_O of the output layer. During training, given an input, the network learns to generate a representation of the output modalities that is as close to the actual values. This is enforced by using a loss function of Mean Squared Error (MSE) between predicted output and expected output, in Figure 4, it is the difference between predicted thigh X'_{thigh} and actual thigh data, X_{thigh}.

When a query case C^q is presented at deployment, we use one or more Neural Translators to generate all missing modalities required to re-construct C'^q for classification. We refer to this Neural Translator as T^N. Formally the privileged information generated by T^N is:

$$\chi' = \theta(\chi)$$

Here θ denotes parameters of the trained neural translator.

4 Evaluation

We conduct a comparative study to explore the utility of translation models to augment representations for HAR. Accordingly we include the following algorithms:

– T^{kNN} (Section 3.2) for several k values (1,3 and 5) with Euclidean Distance for δ; and

– T^N (Section 3.3) using the hyper parameters in Table 1 which were found to be empirically most effective.

Table 1. Hyper-parameters for Neural Translator

Hyper-parameter	Value
Number of Hidden Layers	1
Number of Hidden Units	96
Loss Function	Mean Squared Error
Optimizer/Learning Rate	Adam / 0.01
Number of Epochs	100

In the rest of this section we detail datasets, preprocessing and experiment designs.

4.1 Datasets

We use two HAR datasets in our experiments and their details appear in Table 2.

SelfBACK dataset was compiled with two tri-axial accelerometer data streams belonging to 6 activity classes performed by 34 individuals for approximately 3 minutes. Accelerometers were mounted on the right-hand wrist and thigh of the subject (thus forming 2 modalities). The data for three axes was recorded at $100Hz$ for each modality with time stamp. The dataset was recorded simultaneously on two sensors but dispersed as two separate datasets for each modality. For this study we merge the two datasets aligning them by timestamps to create a dataset with 8 columns as follows: 1 for the time stamp, 3 (x,y,z) columns each for wrist and thigh and the label.

PAMAP2 is a Physical Activity Monitoring dataset which contains data from 3 inertial measurement units (IMUs) located on wrist, chest and ankle. 18 different physical activities were recorded by 9 subjects following a pre-defined protocol [7]. Due to class imbalance within subjects in the dataset we filter out one subject and 9 activities with insufficient data. In addition we only selected accelerometer data from IMUs. The refined dataset contained 8 subjects and 9 activity classes. Previous literature of PAMAP2 dataset provides benchmark classification using all modalities [7]. But for the purpose of this research we created classification models using individual sensor modality.

Table 2. Summary - Datasets

	SelfBACK	PAMAP2
Number of Subjects	34	8
Number of Activity Classes	6	9
Accelerometer Calibration	± 8g, 100Hz	± 16g, 100Hz
Sensor Placements and Notation	Wrist (W) and Thigh (T)	Wrist (W), Chest (C) and Ankle (A)
Window Size	3s	3s
Number of Instances	9889	4833
Case Base	$\{C^1, C^2, C^3, ..., C^{9889}\}$	$\{C^1, C^2, C^3, ..., C^{4833}\}$
Case	$C^i = \{X^i_W, X^i_T, L^i\}$	$C^i = \{X^i_W, X^i_C, X^i_A, L^i\}$

4.2 Data Pre-processing

We perform three pre-processing steps on each dataset to create case bases for our translators and classification.

1. We use a sliding window size of 3 seconds with no overlap to create instances for each subject.
2. We convert the three-dimensional (x, y, z) raw data into a single dimension Discrete Cosine Transform (DCT) instance. First we convert each axis data instance of 300 timestamps into a DCT feature array and then select the first 60 DCT features. We append DCT features from all axes to form one array of length 180.
3. Finally we normalize all data instances.

We use DCT feature transformations as it has been proven to result in significant performance improvements over raw multi-dimensional features. DCTs extract generic and robust features compared to other statistically crafted features and was also shown to have slightly better or comparable results to deep feature embeddings[9]. Importantly for us, it simplifies the task of translators when the mapping can be carried over a proven feature representation for input and output data. Finally data normalisation ensures that the k-NN classifiers are unaffected by scalar differences between different modalities across all datasets.

4.3 Experiment Design

We employed Leave-One-Person-Out (LOPO) cross validation with all our experiments of HAR, with a k-NN classifier where $k = 3$. We use accuracy on classification to study the contribution of translators to performance gains in HAR and compare results for with and without privileged information. In order to establish which modalities are more likely to be considered as privileged in a given dataset, we also studied their individual performance and the contribution they each provide when combined with other modalities.

With the neural translator we perform several experiments to identify the most effective hyper-parameters. We experiment with different hidden units and hidden layers in the neural translator to understand the impact on learning the mapping between sensors. While maintaining the number of hidden layers to one, Figure 5 reports the results obtained for different hidden units for SelfBACK dataset. We can observe how performance increases with number of hidden units, but after 96 (which is closer to half of the size of input units) performance declines. This confirms claims we made in Section 3.3 on how a narrow hidden layer supports learning better mappings between sensors while discarding arbitrary noise.

Figure 6 presents performance results obtained with different hidden layers on the SelfBACK dataset. In the first four columns, we maintain a considerably narrow layer size compared to the input and output units, while increasing the number of layers. These four experiments do not show substantial performance gains from having additional layers. Later we increase number of layers and make them broader which saw a significant drop in performance. We can observe that, when the number of parameters of the network increases, the network tends to over-fit the training data and leads to poorer performance. Accordingly we use the best hyper parameters in Table 1 on all Neural Translator experiments.

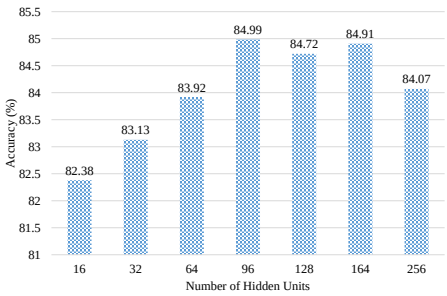

Fig. 5. T^N - Hidden Units

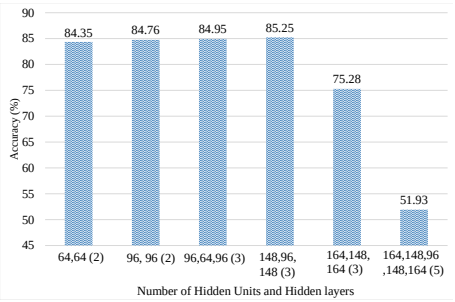

Fig. 6. T^N - Hidden Layers

We adopt the following naming convention, $f(X_i/X_j)$ to identify the different classifiers by the modalities that have been used for training, X_i, as well as to indicate which modalities (if any) are used as privileged information, X_j. Here $X_j = \emptyset$ indicates the absence of modalities for privileged information. For example, the $f(T/\emptyset)$ denotes a classification model trained and tested with the single modality thigh data using no privilege information; similarly $f(W, C/\emptyset)$ is a classification model trained on two modalities, W (wrist) and C (chest), again with no privileged information. In contrast $f(W, C, A/A)$ suggests the use of 3 modalities for training with modality A (ankle) forming the privileged information which will be estimated by a translator. With translators we adopt the following naming convention $T(X_i/X_j)$; For instance $T^N(W, C/A)$ indicates

a neural translator which generates A (ankle) as privilege information by translating from W & C (which are wrist and chest) data.

5 Results

In this section we will first identify PI for each dataset by comparing baseline results, next we present performance we obtained with k-Nearest Neighbour Translator, finally we present performance we obtained with our Neural Translator. We discuss results and their implications at each subsection.

5.1 Comparison of baselines to identify Privileged Information

For each dataset we create several baselines classifiers with no privileged information (see Figure 7). This allows us to analyse individual performance on the HAR classification task and identify modality placements that are ideal for activity recognition.

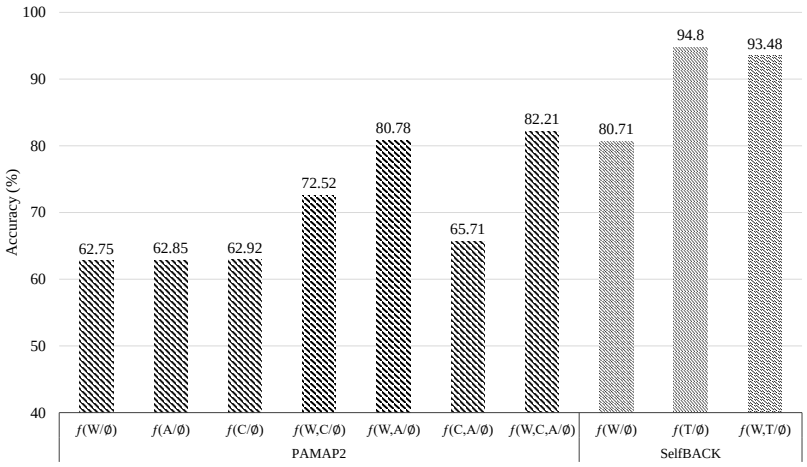

Fig. 7. Baseline classification results for SelfBACK and PAMAP2

In figure 7 there are results from 3 baselines created for the SelfBACK dataset. The best baselines are $f(W, T/\emptyset)$ and $f(T/\emptyset)$, with the inclusion of the wrist suggesting a slight decline in performance. With $f(W/\emptyset)$ using only the wrist, we see a considerable performance decline of almost 15% compared to the other two. These baseline results confirm that thigh is clearly a Privileged Information in the SelfBACK dataset.

The 7 baseline classifier accuracies for PAMAP2 dataset are also shown in Figure 7. Here we can see that each of the 3 single modality classifiers have

comparable performance but appear to have as much as a 10% performs degradation compared to the multi-modal baselines (e.g. $f(W, C, A/\emptyset)$). This might be explained by the similarities between some activity classes for example such as "Walking" and "Nordic Walking" which are harder to differentiate with a single sensor and would instead require multiple modalities.

Of the multi-modal classifiers on PAMPA2, $f(W, A/\emptyset)$ outperforms $f(W, C/\emptyset)$ and $f(C, A/\emptyset)$, furthermore, performance of $f(W, A/\emptyset)$ is notably close to the three-modality classifier $f(W, C, A/\emptyset)$. Surprisingly the two-modality classifiers $f(C, A/\emptyset)$ does not improve their single-modality performance substantially, but they both (Ankle and Chest) show improved performance when combined with wrist modality. Accordingly in this dataset we assess the use of both chest and ankle modalities as privileged information. We believe this is sensible especially when considering the intrusiveness of either of these wearables compared to an inertial sensor on the wrist.

5.2 Privilege Information generation with the k-NN Translator

In general k-NN as a translator failed to provide any significant improvements over classification without privileged information on the SelfBACK dataset (see Table 3). We studied two classifiers with both using a translator to generate thigh; where one used only thigh data (first column) and the other uses both wrist and thigh (second column). However at most, we only observed a classification performance improvement of only 1.32% over the baseline classifier $f(W/\emptyset)$. Increasing the number of neighbours (from k values 1, 3 to 5) also had no significant impact apart from a marginal improvement (as little as 1%).

Unlike with SelfBACK, in PAMPA2 we used only multi-modalities to train the HAR classifier following the poor results observed in Figure 7 with single modalities. However once again results here did not exhibit any substantial improvement or decline in performance compared to the baselines. In addition we observe no significant performance difference was to be had by increasing the neighbourhood sizes.

Table 3. T^{kNN} with SelfBACK and PAMAP2

	SelfBACK		PAMAP2		
	$f(T/T)$, $T^{kNN}(W/T)$	$f(W,T/T)$, $T^{kNN}(W/T)$	$f(W,A/A)$, $T^{kNN}(W/A)$	$f(W,C/C)$, $T^{kNN}(W/C)$	$f(W,C,A/A)$, $T^{kNN}(W,C/A)$
T^{1NN}	81.03	81.02	62.47	62.25	71.01
T^{3NN}	81.52	81.57	63.01	63.37	72.41
T^{5NN}	81.50	82.02	62.58	62.82	71.56

We believe the poor translation capability of the kNN method is primarily due to the inherent noise in some of the modalities. This is particularly the

case with SelfBACK (as observed in 7) and therefore is not surprising that the translation mapping was also not able to recover from this noise already captured in the case representation from wrist. However with PAMPA2 we did not see any significant difference between any of the single modalities (Figure 7) and as such do not believe that wrist is any nosier than, say chest for instance. Here we believe that the poor performance might be explained by the inability of the single modality to discriminate between the activity classes. These uncertainties are emphasised when selecting neighbours using single modality, thus end up not gaining any performance improvement from privileged information.

In general we expect that an incremental learner such as the neural translator will have a better opportunity to learn an improved mapping as it minimises the differences between estimated and actual privileged information during training. This alone helps to create an improved feature embedding compared to kNN.

5.3 Privilege Information generation with the Neural Translator

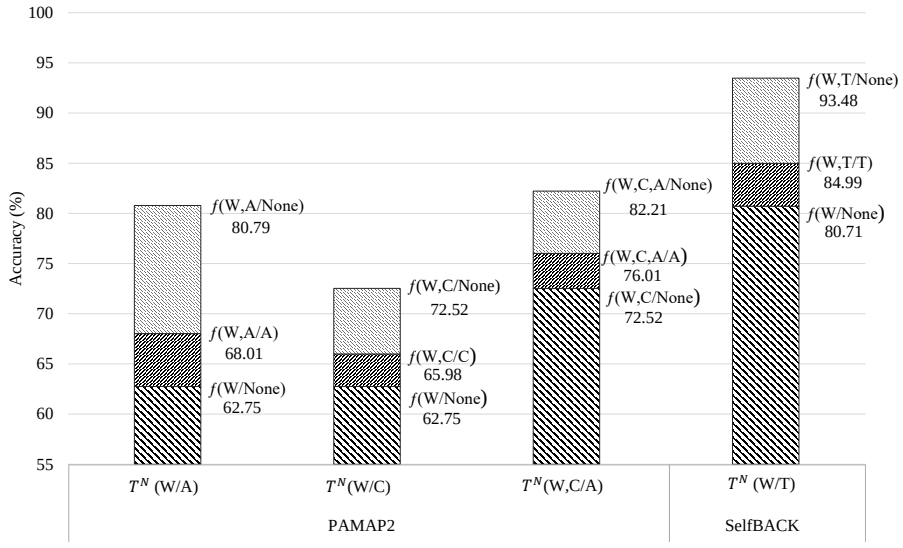

Fig. 8. T^N with PAMAP2 and SelfBACK

Figure 8 shows classification results for the Neural Translator for both SelfBACK and PAMAP2 datasets. Here each bar shows the lower and upper bounds set by the baselines. For instance the upper bound is simply the baseline that uses the actual data instead of the estimated generated by a translator; whilst the lower bound is when the privileged modality is not used for HAR. Ideally we want the translator to improve upon the lower bound to get closer to the upper.

On PAMAP2 we experimented with three multi-modal Neural Translators. Translator $T^N(W/A)$ learns from wrist case attribute to generate ankle case attribute from CB_{PAMAP2}. Results suggests a 5.26% increment in accuracy between $f(W, A/A)$ and the corresponding baseline $f(W/\emptyset)$. Similarly both Translators $T^N(W/C)$ and $T^N(W, C/A)$ improves accuracy of their corresponding baselines $f(W/\emptyset)$ by 3.23% and $f(W, C/\emptyset)$ by 3.49% respectively.

The SelfBACK results appear in the last column of Figure 8. Here we can see that the Neural Translator for SelfBACK has significantly improved the performance of the lower bound baselines $f(W/\emptyset)$ brining it closer to the upper bound set by $f(W, T/\emptyset)$ baseline (which is when all modalities are available without the need for translation). Specifically we observe that the $T^N(W/T)$ translator (wrist-to-thigh) achieves a 4.28% increment in accuracy using privileged information at deployment.

These results suggest that using a classifier trained with multiple modalities, with a single or subset of modalities in deployment, is not only possible but improves performance significantly. Unlike the k-NN Translator, the Neural Translator is less affected by the ambiguities of the source modalities. Instead, it learns relationships that help to map between source and target modalities. As a result the generated modalities improve performance of the HAR classifiers at deployment using the estimated knowledge.

6 Conclusions

We introduced two Translator approaches for privileged learning with HAR. Our results showed the Neural Translator to have significant performance improvements over the baselines which have no privilege learning. kNN translators were less effective in this domain, and we concluded that this was due to the inherent noise and class ambiguities in HAR which requires effective case representations. But unlike the neural translator, the kNN translator had no mechanisms to iteratively refine its representations.

Overall the neural translator had significantly outperformed the lower bounds set by the baseline classifiers on both datasets. However we believe there is further opportunity to improve on the translator generated representations allowing us to move closer to the upper bound or optimal performance observed when actual privileged information is used.

Accordingly in future work we will explore a number of directions in which to improve our Neural Translator, for instance exploring other network optimisation techniques, different data representations and also considering how ideas from case adaptation might be employed here in a neural setting. Another direction involves the creation of personalised translators that are better able to capture personal traits and individual differences when estimating missing modalities.

Finally this research has demonstrated that translation methods can help to minimise the number of sensors needed at deployment; which we argue is one of the key components of an optimal HAR solution.

References

1. Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S.T., Tröster, G., Millán, J.d.R., Roggen, D.: The opportunity challenge: A benchmark database for on-body sensor-based activity recognition. Pattern Recognition Letters **34**(15), 2033–2042 (2013)
2. Chen, Y., Jin, X., Feng, J., Yan, S.: Training group orthogonal neural networks with privileged information. arXiv preprint arXiv:1701.06772 (2017)
3. Luong, M.T., Le, Q.V., Sutskever, I., Vinyals, O., Kaiser, L.: Multi-task sequence to sequence learning. arXiv preprint arXiv:1511.06114 (2015)
4. Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y.: Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies **54**, 187–197 (2015)
5. Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y.: Multimodal deep learning. In: Proceedings of the 28th international conference on machine learning (ICML-11). pp. 689–696 (2011)
6. Preece, S.J., Goulermas, J.Y., Kenney, L.P., Howard, D.: A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. IEEE Transactions on Biomedical Engineering **56**(3), 871–879 (2009)
7. Reiss, A., Stricker, D.: Introducing a new benchmarked dataset for activity monitoring. In: Wearable Computers (ISWC), 2012 16th International Symposium on. pp. 108–109. IEEE (2012)
8. Ronao, C.A., Cho, S.B.: Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications **59**, 235–244 (2016)
9. Sani, S., Massie, S., Wiratunga, N., Cooper, K.: Learning deep and shallow features for human activity recognition. In: International Conference on Knowledge Science, Engineering and Management. pp. 469–482. Springer (2017)
10. Shi, Z., Kim, T.K.: Learning and refining of privileged information-based rnns for action recognition from depth sequences. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
11. Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T.S., Kjærgaard, M.B., Dey, A., Sonne, T., Jensen, M.M.: Smart devices are different: Assessing and mitigating-mobile sensing heterogeneities for activity recognition. In: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems. pp. 127–140. ACM (2015)
12. Sundholm, M., Cheng, J., Zhou, B., Sethi, A., Lukowicz, P.: Smart-mat: Recognizing and counting gym exercises with low-cost resistive pressure sensing matrix. In: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing. pp. 373–382. ACM (2014)
13. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. pp. 3104–3112 (2014)
14. Vapnik, V., Vashist, A.: A new learning paradigm: Learning using privileged information. Neural networks **22**(5), 544–557 (2009)
15. Vinyals, O., Toshev, A., Bengio, S., Erhan, D.: Show and tell: A neural image caption generator. In: Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on. pp. 3156–3164. IEEE (2015)
16. Yao, S., Hu, S., Zhao, Y., Zhang, A., Abdelzaher, T.: Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In: Proceedings

of the 26th International Conference on World Wide Web. pp. 351–360. International World Wide Web Conferences Steering Committee (2017)
17. Yin, W., Schütze, H., Xiang, B., Zhou, B.: Abcnn: Attention-based convolutional neural network for modeling sentence pairs. arXiv preprint arXiv:1512.05193 (2015)
18. Yu, H., Wang, J., Huang, Z., Yang, Y., Xu, W.: Video paragraph captioning using hierarchical recurrent neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4584–4593 (2016)