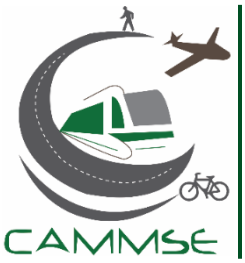

Center for Advanced Multimodal Mobility Solutions and Education

Project ID: 2021 Project 03

HAR-GCNN: DEEP GRAPH CNNS FOR HUMAN ACTIVITY RECOGNITION FROM HIGHLY UNLABELED MOBILE SENSOR DATA

Final Report

by

Abduallah Mohamed (ORCID ID: https://orcid.org/0000-0002-6074-6010)
PhD student, University of Texas Austin
Department of Civil, Architectural and Environmental Engineering
301E E Dean Keeton St c1700 Austin, Texas 78712

Christian Claudel. (ORCID ID: https://orcid.org/0000-0002-3783-4928)
Assistant Professor, University of Texas Austin
Department of Civil, Architectural and Environmental Engineering
301E E Dean Keeton St c1700 Austin, Texas 78712

for

Center for Advanced Multimodal Mobility Solutions and Education
(CAMMSE @ UNC Charlotte)
The University of North Carolina at Charlotte
9201 University City Blvd
Charlotte, NC 28223

September 2022

ACKNOWLEDGEMENTS

This project was funded by the Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE @ UNC Charlotte), one of the Tier I University Transportation Centers that were selected in this nationwide competition, by the Office of the Assistant Secretary for Research and Technology (OST-R), U.S. Department of Transportation (US DOT), under the FAST Act. The authors are also very grateful for all of the time and effort spent by DOT and industry professionals to provide project information that was critical for the successful completion of this study.

DISCLAIMER

Table of Contents

List of Figures

List of Tables

EXECUTIVE SUMMARY

The problem of human activity recognition from mobile sensor data applies to multiple domains, such as transportation engineering, but also in health monitoring, personal fitness, daily life logging, and senior care. In the context of transportation safety, a key objective of human activity recognition is to detect and classify current activities of a road user (for example a pedestrian), and use this information to forecast future actions. This forecasting of future actions is essential to solve the problems of latency, in autonomous vehicles or other advanced driver assist systems (ADAS).

A critical challenge for training human activity recognition models is data quality. Acquiring balanced datasets containing accurate activity labels requires humans to correctly annotate and potentially interfere with the subjects normal activities during data collection. Since the performance of learning and classification schemes improves with larger datasets, it is essential to generate accurate, extremely large datasets of human activities. While there exists a likelihood of incorrect annotation (or lack of annotation), there is often an inherent chronology to human behavior. For example, some activities tend to follow other precursor activities. This implicit chronology can be used to learn unknown labels in the training dataset and classify future activities.

The objective of this work is to propose a new method for predicting correct labels of unclassified or partially classified activities. To this end, we propose HAR-GCCN, a deep graph CNN model that leverages the correlation between chronologically adjacent sensor measurements to predict the correct labels for unclassified activities that have at least one activity label. We propose a new training strategy to ensure that the model predicts missing activity labels by leveraging the known ones. HAR-GCCN shows superior performance relative to previously used baseline methods, improving classification accuracy by about 25% and up to 68% on different standard datasets, including the PAMAP dataset.

Chapter 1. Introduction

1.1 Problem Statement

Human Activity Recognition (HAR) has been an active research field for the past two decades, covering a wide range of applications in health monitoring and fitness [1]–[4]. Technological advances regarding inertial sensors with longer battery life span and improved computing capabilities have enabled the gathering of larger volumes of continuous data that can be used for activity prediction [5]. Despite these recent advances, "ground truth annotation" which is the process of labeling sensor readings with the activity being performed by the user remains a critical challenge [5]. Such annotation tasks are typically performed manually and occur either in real-time or post hoc once the activity has been completed. In the majority of activity recognition studies, such annotation tasks can be expensive, onerous, prone to human error, and even condition the user interfering in the activity itself [6]. Thus, it is expected that collected data contains a significant amount of unreliable and missing labels, such as erroneously classified sensor readings. These incorrect or missing labels create gaps in the data which have a detrimental impact on model development and training. Thus, a paramount aspect of predictive models for HAR is learning in the presence of missing labels, while at the same time achieving high accuracy in classifying human activity. A variety of deep learning and machine learning methods have been previously proposed for single and multi-label classification of human activity [7]– [12]. The missing label(s) can be predicted based on the readings of multiple sensors such as accelerometer, gyroscope, location, etc [10]. However, fewer approaches leverage the context of "neighboring" activities for predicting a missing label which has been shown to improve activity recognition [11], [13].

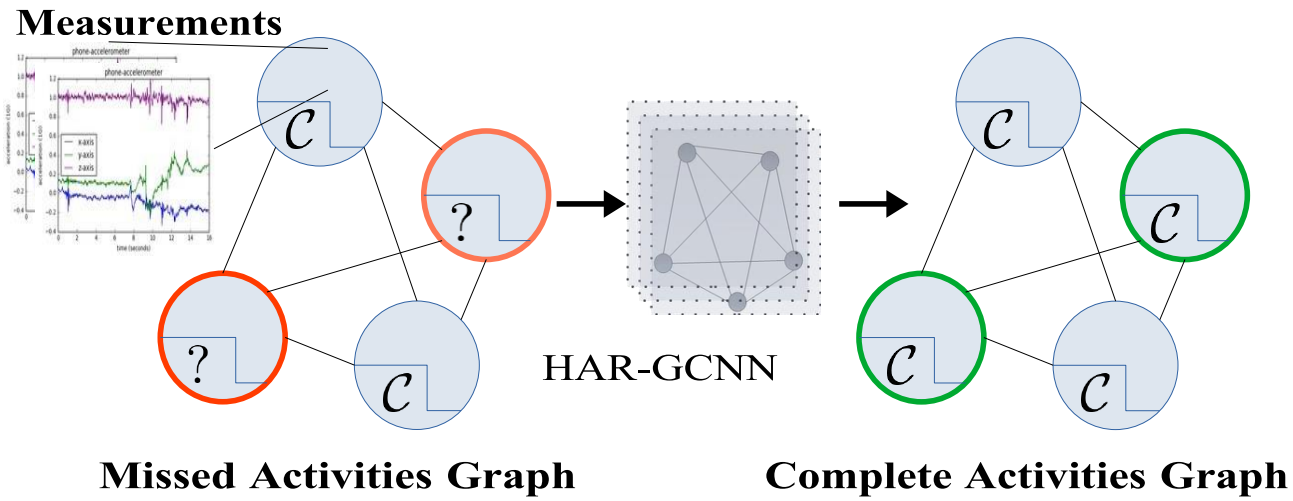

Figure 1 The input to HAR-GCNN is a graph consisting of partially labelled sensor measurements that is chronologically ordered activities. The model predicts the classes C of the unlabeled activities. Each input node contains a period of sensors measurements along side a class or missing class. The output is the activity class for each node.

1.2 Objectives

We hypothesize that human activities follow a chronological correlation which can provide informational context to improve HAR. The benefit of this hypothesis, if valid, is that one can leverage known or correctly labeled activities to predict the surrounding ones. In order to investigate the validity this assumption, we formulated the HAR problem comprising a sequence of chronologically ordered activities, some of which were correctly labeled while others were not labeled at all. We used deep learning as a data-driven approach to discover this chronological correlation. We first employed Recurrent Neural Nets (RNNs) which is the most straightforward deep model used to handle time series data. Nonetheless, our literature survey revealed that Convolutional Neural Networks (CNNs) can often be more powerful than RNNs in performing HAR [14]. Yet, both approaches CNNs and RNNS are structurally not geared towards directly leveraging the correlation between the sequential activities. For example, RNNs treat each time step separately by fusing it into the "neural memory", which is the only component that partially correlates the data. On a similar note, CNNs only make use of neighboring activity information based on the chosen kernel size. Based on the shortcomings of these architectures, we posited that graph-based structures more adequately capture the aforementioned features of the problem at hand. Graph representations in HAR allow for modeling each activity as a node, and the graph edges can directly model the relationship between these activities. A suitable tool for learninsuch graphs is deep Graph CNNs (GCNNs) [15]. Deep graph CNNs behave as ordinary CNNs but weigh the nodes based on the value of the edges. In other words, GCNNs can employ the sequence of information directly and exploit the correlation between all activities.

1.3 Expected Contributions

In this work we show that the chronological correlation indeed exists by experimenting with two commonly employed HAR datasets. One was collected in the wild, while the other was collected in a scripted manner. Our results show that the proposed models benefit from this correlation and use it to predict the neighbouring missing activities, improving performance relative to RNNs and CNNs benchmarks.

1.4 Report Overview

In the following sections, we discuss prior works, as well as the datasets utilized for training and validation. We provide details of our formulation including the graph structure used to model human activities. Then, we introduce our proposed HARGCNN, highlighting details of its architecture and implementation, and define the comparison baselines used to benchmark our approach. In the experiments section, we evaluate our model's performance relative to other previously reported deep neural network architectures employing widely used HAR datasets.

Chapter 2. Literature Review

A large number of traditional supervised machine learning techniques have been developed in the literature for HAR from mobile sensor data. Such models include, for example, logistic regression, k-nearest neighbors, decision trees, and multi-layer perceptrons (MLP) [7]–[9]. These models typically exhibit good performance when trained on controlled datasets that are fully and accurately labeled, which may not be the case for in-the-wild environments [16]. Furthermore, these models often require substantial feature extraction which can be time-consuming and rely heavily on domain knowledge [10].

Addressing these shortcomings, a variety of prior works have explored unsupervised and semi-supervised learning techniques on human activity data. For example, [17] developed a network of convolutional autoencoders on an unlabeled dataset to extract useful features, which were then used to complement their supervised learning mechanism. [17] proposed Deep Auto-Set, a deep learning classification model trained on raw multi-modal sensory segments. Furthermore, [4] implemented semi-supervised, active learning techniques operated both online and offline. [4] demonstrated superior performance relative to fully supervised approaches on activity recognition datasets where ground truth labels are scarce. More specifically, [4] evaluated the performance of different pool-based and stream-based active learning frameworks on various datasets relevant to human activity [1], [18]–[20] and contrasted classification accuracy against random forests, logistic regression, and k-nearest neighbors.

Besides these models, a great number of deep learning frameworks have been proposed for HAR [10]. For example, different CNN architectures [21]–[23] resulted in significantly higher accuracy than prior machine learning techniques applied on HAR. Further, [24] explored normalization and sensor data fusion using CNNs and demonstrated that these techniques can further improve the performance of deep learning models. [11] investigated deep convolutional recurrent models including deep feed-forward neural networks (DNN), CNNs, and different variants of Long Short-Term Memory (LSTM) cells. Their results show that recurrent networks significantly outperform convolutional networks on activities that are short in duration but follow a natural chronology. Such performance gains are likely because recurrent models can contextualize to improve recognition. [25] reported improved activity recognition using ensembles of deep LSTM learners relative to previously reported recurrent neural networks. [26] developed a deep RNN performing extensive hyper-parameters optimization and showed superior performance compared to other traditional machine learning models.

Several variants of graph-based models have been recently proposed leveraging spatial and temporal properties in the data for collective activity recognition (i.e., predicting the activity of an entire group as opposed to a single subject). Numerous of these models have been reported for computer vision applications to (deeply) learn the interactions between individual participants for a given sequence of video clips and to predict the overall activity or outcome of a group of people [27]–[31]. Similarly, [32] proposed a context-aware, semi-supervised graph propagation algorithm with a Support Vector Machine (SVM) classifier to address individual HAR. The results in [32] suggest that exploiting contextual data from neighboring activities (i.e., nodes) can result in improved performance even relative to fully supervised approaches, likely by discriminating outliers and erroneously labeled data.

GCNNs have recently gained significant popularity for various applications of semi-supervised learning [15], [33]. GCNNs leverage dependencies between the features and labels of nodes in a given graph resulting in improved predictive performance, particularly when a significant number of the training labels are missing. Considering these desirable characteristics, as well as the aforementioned advantages of models that can contextualize, GCNNs are an appealing modeling strategy that to the best of our knowledge has not been previously reported for HAR. In light of this, the main contributions of this work are:

- A formulation for HAR exploiting the chronological context of the activities embedded within a graph structure.
- A novel mechanism that trains HAR-GCNN to learn to predict the missing labels in the input graph with high accuracy.
- Extensive computational experiments evaluating the effect of the percentage of missing labels, as well as the effect of the number of activities on the prediction accuracy. Experiments performed on Extra-Sensory [18] and PAMAP [34] datasets, which are among the most commonly used in the activity recognition literature.
- Benchmarks of our proposed approach against previously reported deep architectures including CNN and LSTM models.

Chapter 3. Methods, Findings, & Discussion

3.1 Problem Definition

Given a set of time series sensory measurements $F = \{f_t | t \in T\}$ sampled over a window of time steps T, it is of interest to learn the classes of activities $C = \{c_t | t \in T\}$ associated with each of such measurements. The measurements are collected from a variety of sensors such as accelerometers, gyroscopes, etc. Each interval of measurements is associated with multiple labels corresponding to different activities (i.e., multi-label classification). Our problem formulation uses a set of prior and posterior measurements with known activities to predict the multi-label or single label classes of those unknown activities. In other words, using $(f_{t-m}, c_{t-m}),...,(f_{t-1}, c_{t-1}),(f_{t+1}, c_{t+1}),...,(f_{t+m}, c_{t+m})$ and f_t to predict c_t, where m is the number of neighbor activities to be used. In this way, we consider a sequence of activities represented by sensor measurements over a time horizon including the associated labels, whether they are known or not, and predict the class of the unknown labels exploiting the observed sequential order.

One important remark is that the datasets under consideration are recorded in-the-wild. That is, an in-the-wild dataset contains sensor readings from the users who were not previously instructed to perform a given set of activities (which is typically the case in more controlled environments). In-the-wild settings in turn reduce the bias within the collected data and results in a more natural chronology of activities from which our proposed model can learn. The ExtraSensory [18] dataset is an example of such in-the-wild measurements.

The dataset contains over 300,000 minutes of labeled sensor recordings from smartphones and smartwatches worn by a total of 60 study participants. The behavioral activities in this dataset can be categorized by 51 nonexclusive labels such as sitting, sleeping, strolling, cooking, etc. For this type of the data the HAR task at hand is a mutli-label classification problem. Further, the dataset contains 224 different raw features obtained from mobile sensor readings recorded over 20 seconds window every one minute.

To further validate the advantages of the framework proposed herein, we also conduct experiments using the PAMAP dataset [34]. The PAMAP dataset while collected in a more controlled environment provides a useful instance to validate our framework, particularly when labels are artificially hidden during training. The PAMAP dataset consists of data from 9 subjects, wearing 3 inertial measurement units and a heart rate monitor and performing 12 exclusive activities (i.e., the HAR task is a single-label classification problem) with 52 raw features per instance. The PAMAP dataset is being used herein as an example of a scripted dataset, which likely means that the chronological order of the recorded activities indeed contains some amount of bias and which we further emphasize in the numerical experiments section.

3.1.1 Training/ Testing details

Both datasets were split into training and test sets with a ratio of 2:1. The sequence of the recorded activities was kept and no randomization was used, which is crucial for learning from the natural sequential order of activities. The Extra-Sensory dataset serves to show that the

activities can be predicted with higher accuracy by exploiting such implicit order. Conversely, the PAMAP dataset is employed to show that if the sequence of activities is scripted a priori, HAR-GCNN will result in almost perfect classification performance as it can quickly learn the underlying script. We note that no test data was used to train the proposed model, and the data sets were kept separate to prevent any data leaking.

3.2 Description of the HAR-CGCNN

Constructing the activity graph: We model the set of activities as a graph, which is defined as G = (V,E) where V = $\{v|v \in V\}$ is a set of vertices. Each vertex v = [ft,ct] contains both measurements over a time window and the associated multi-label class, with the exception of the nodes whose activity label is missing and we thus want to predict, for which c is set to zero. The graph edges are fully connected, E = $\{e_{ij}|e \in E; i,j \in |V|\}$, where E is the set of all the graph edges. The weight of these edges is set to one, so that predicting the activity of all nodes is equally important. Our proposed model allows using any number of nodes during training. For example, the graph can consist of three nodes with one of them missing its label. In the experiments section we study the effect of the number of input labelled neighbour nodes on predicting a missing activity class. Further, we note that framework allows for predicting more than one missing label, which we explore in our numerical experiments.

3.2.1 Model Description:

First, the GCNN layer takes as input the aforementioned activity graph G. The GCNN model has a layer-wise structure defined as

$$\text{GCNN}(\mathcal{V}^{(l)}, A) = \sigma\left(A_{\text{norm}}\mathcal{V}^{(l)}W^{(l)}\right)$$

Where A is the adjacency matrix that defines the edges of the activity graph and $\sigma(\cdot)$ denotes an activation function. The GCNN operates as an ordinary CNN except that it weights the kernel by the value of the normalized adjacency matrix A_{norm} defined as:

$$A_{\text{norm}} = I - \hat{D}^{-\frac{1}{2}}(A + I)\hat{D}^{-\frac{1}{2}}$$

Where $\hat{D}$ is the degree node matrix of $A + I$, and I is the identity matrix. This normalization approach was introduced by [14]. More in-depth information about graph CNNs can be found in [15]. The output of the GCNN is a graph embedding that represents the whole information of the sensor measurements and their corresponding known labels.

The second step in HAR-GCNN is the CNN output layers. We use a sequence of CNNs (single layer convolutions with an activation function) to process the graph embedding to predict the activities classes C. The main reason for this architecture choice is that performance can deteriorate when the depth of the GCNNS increases as it was reported by [35]. In other terms, we could use a sequence of graph CNNs to predict each activity label but this would be inefficient. Thus, the usage of CNNs arises naturally as a necessity for to go deep in our model. The output of this CNN layers is directly regressed against the proper loss function (Binary Cross Entropy or

Cross-Entropy) to predict the labels C. Figure 2 describes the different steps included in our architecture.

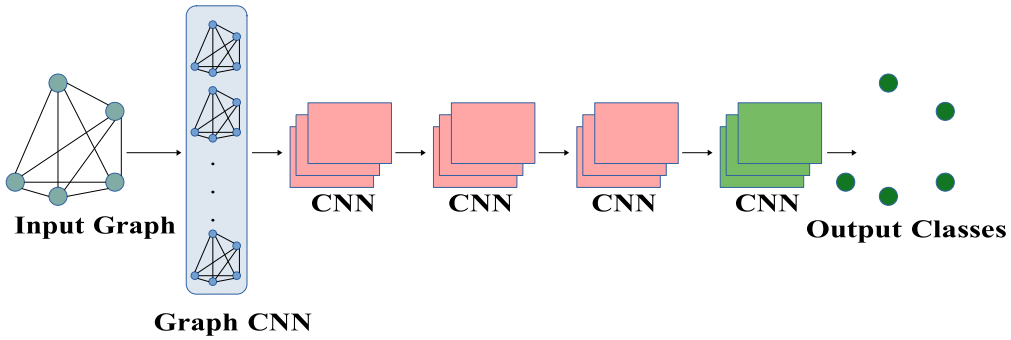

Figure 2 HAR-GCNN model description. *A* stands for adjacency matrix, V stands for the graph vertices. The CNN is just a single layer convolution with an activation function.

Implementation Details: We employ the same GCNN implementation as the one reported in the pioneering work of [15]. We believe that the chosen implementation is easier to interpret and deploy, relative to similar architectures available in PyTorch [36]. The model consists of a single GCNN layer that outputs a graph embedding. Then, three CNN layers are used to process the embedding and a final CNN layer is included with a sigmoid (softmax for PAMAP) activation function. We used PReLU [37] activation functions for intermediate layers. The total model parameters size is $\sim 15k$ for the Extra-Sensory dataset and $\sim 5k$ for the PAMAP dataset.

3.3 Baseline Methods

The CNN baseline The CNN baseline is a sequence of five CNN layers that have the same depth as HAR-GCNN model. The reason for choosing this architecture is that V represents a set of an arbitrary number of activities, and a CNN is capable of learning a kernel that is agnostic to the width and height of the input. The total parameters size is also $\sim 15k$ for the Extra-Sensory and $\sim 5k$ for the PAMAP datasets, thus allowing for a fair comparison with HAR-GCNN. The input to this base model is the V itself, which can be regarded as an image of width equal to the sensor readings, and of height equal to the number of nodes or activities in the graph. All layers used Parametric Rectified Linear Unit (PReLU) [37] activation function except the last layer with a sigmoid (softmax for PAMAP) function. The PReLU where used to allow a better gradients flow to the model. The intermediate three layers use a residual connection to improve performance during training.

The LSTM baseline The input to the LSTM are the graph nodes treated as a sequence of time steps. The model parameter size is also $\sim 15k$ and $\sim 5k$ for both Extra-Sensory and PAMAP respectively, same as HAR-GCNN model and CNN base model. The LSTM model can be seen in Figure 3. The 1D-CNN layers uses PReLU activation function except the last layer which uses a sigmoid (softmax for PAMAP) activation function.

3.4 Training

In addition to keeping the number of parameters between the baseline models and HARGCNN the same across the two datasets, all models were trained using the exact same training settings. We use Binary Cross Entropy (BCE) as a loss function in case of the Extra-Sensory dataset *because it is a mutli-label problem*, BCE is defined as:

$$\text{BCE loss}_j = -[\text{class}_j \cdot \log x_j + (1 - \text{class}_j) \cdot \log(1 - x_j)]$$

In case of the PAMAP dataset we used the Crossy-Entropy(CE) loss *because it is a single label problem*, the CE is defined as:

$$CE\ loss(x, class) = -\log(\frac{\exp(x[class])}{\sum_j \exp(x[j])})$$

Where *j* is the class number and *x* is the predicted class. We artificially impose missing labels in both of the dataset to simulate the case of missing labels. Our training framework randomly hides a percentage of the known activity labels C within the input graph G, which serve as the missing labels we want to predict. Moreover to make the model generalize better, we add random noise (Gaussian with 0 mean and 1 standard deviation) to disturb the measurements vector F, which helps generalize the performance of HAR-GCNN for unseen data while also accounting for possible measurement errors in the raw sensor data. In all of our experiments we used a 50% probability for either hiding a label or adding noise to the measurements. Further, we restricted the percentage of hidden nodes to be no greater than 66%, so that we are guaranteed that at least 33% of the nodes have their original labels. The disturbed data were generated once and were used across all of our experiments across different models in order to maintain a fair comparison. We use a test set that is about 33% of the data to computer performance metrics in reporting our results. We ran all of our experiments three times with a controlled random seed to ensure same settings for the training and test data. The reported results are the mean of these runs. Figure 3 illustrates the data-flow and the dimension of the data alongside the training mechanism.

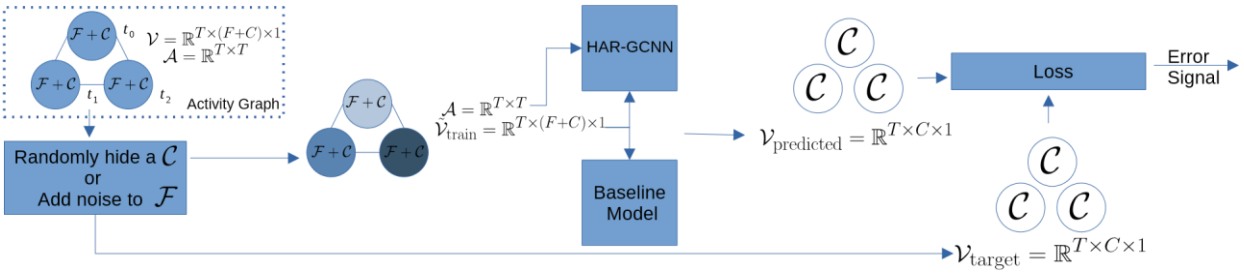

Figure 3 HAR-GCNN data flow diagram for model training as described in VI. F represents the sensor measurements with dimension F = 224,52 for the Extra-Sensory dataset and PAMAP dataset, respectively. C is the set of multi-label activity classes with dimensions C =51,12 for the Extra-Sensory dataset and PAMAP dataset, respectively. T is the number of time steps, where T corresponds to the number of the graph nodes. A stands for adjacency matrix, V stands for the graph vertices.

3.5 Experiments

To evaluate our model performance in predicting missing activity labels, we artificially hide a percentage of the labels and evaluate the model performance in classifying them. We considered 2 settings for the amount of missing classes C 33% and 66% in the input graph G. For example, in the case of 66% missing labels, it corresponds to the scenario in which $\frac{2}{3}$ of the measurements F are known but none are labeled, and we attempt to predict their labels C. This evaluation method gives us a perspective of the model's ability to generalize to unseen situations because the models were not trained on these settings as described in in the training mechanism section VI. We start by analyzing the performance of HAR-GCNN in comparison to the baseline methods. This will be followed by an analysis of the effect of the number of activities in the graph, which relates to how the length of the sequence of chronologically ordered observed activities influences the precision of the prediction. Because this results in a multi-label classification problem for the case of the Extra-Sensory dataset, using the macro F1 score and mean accuracy are suitable metrics to assess the models' performance. The macro F_1 score is defined as:

$$F_{1\ \text{macro}} = \frac{1}{N}\sum_{n \in N} F_1^n$$

Where N is the total number of classes. The mean accuracy is defined as:

$$\text{Acc}_{\text{mean}} = \frac{1}{N}\sum_{n \in N} \text{Acc}_n$$

Also the choice of the macro F_1 score is important to gain a better understanding of the performance of the models regarding instances of significant class imbalance. For example, some base models may result in a fairly high mean accuracy but their macro F_1 score is low compared to HAR-GCNN. Such results would suggest that HAR-GCNN is able to learn well in such class imbalance settings. For the PAMAP dataset we employed the ordinary F_1 and Acc metrics as the problem is a multi-class problem.

3.5.1 Performance of HAR-GCNN Against Baselines

Table I shows the main results of comparing the performance of HAR-GCNN versus the baseline models considered. We explain the main results by focusing on the 3 activity case (i.e., first row of Table I) but note that the observations generalize for all other instances in Table I. In the 3 activity case, 33% missing labels implies that only one out of the three activities does not have a label.

# of Activities	% missing labels	Extra-Sensory Dataset			PAMAP Dataset		
		HAR-GCNN	CNN	LSTM	**HAR-GCNN**	CNN	LSTM
3	33%	**0.781 / 99.52**	0.621 / 99.23	0.464 / 98.43	**0.999 / 99.92**	0.975 / 97.52	0.973 / 97.26
	66%	**0.792 / 99.52**	0.531 / 98.69	0.377 / 97.95	**0.999 / 99.94**	0.902 / 90.22	0.903 / 90.26
5	33%	**0.814 / 99.41**	0.715 / 99.41	0.504 / 98.67	0.998 / 99.76	**1.000 / 99.96**	0.990 / 99.05
	66%	**0.880 / 99.58**	0.659 / 99.25	0.459 / 98.37	**1.000 / 99.98**	0.996 / 99.57	0.950 / 95.05
10	33%	**0.813 / 99.14**	0.744 / 99.46	0.522 / 98.64	**0.999 / 99.91**	0.998 / 99.81	0.997 / 99.65
	66%	**0.868 / 99.52**	0.691 / 99.28	0.483 / 98.37	**1.000 / 99.98**	0.997 / 99.75	0.969 / 96.87
25	30%	0.748 / 98.76	**0.784 / 99.50**	0.532 / 98.73	0.998 / 99.79	**1.000 / 99.99**	0.998 / 99.75
	66%	**0.838 / 99.41**	0.724 / 99.23	0.494 / 98.46	**1.000 / 99.99**	0.997 / 99.68	0.981 / 98.10

Table 1 Comparing HAR-GCNN Performance to Baseline Models

An important result is that HAR-GCNN in case of 33% or 66% missing labels outperforms both CNN and LSTM baselines. The performance is about 25% more than CNN and 68% more than the LSTM model for the Extra-Sensory dataset. For the PAMAP dataset HAR-GCNN has a ~2% higher F-1 score than both CNN and LSTM, and almost reaching the upper performance bound of 1.00. We posit that the observed performance improvements stems from the nature of proposed graph formulation, in which the graph edges are able to more deeply capture the relationship between labels and their features. Therefore, if at least one of the activities is correctly classified then there is a high chance that the remaining nodes will be as well. Moreover, we note that the CNN baseline significantly outperforms the LSTM model across the two datasets considered. This connects with the results of HAR-GCNN in which we use a graph CNN, in other terms the CNN based approaches are more suitable to this problem. We due the good performance of CNNs that it does not accumulate errors while predicting the classes. In other terms, it is non sequential, with a global view of the whole state. Unlike in recurrent based methods like LSTM, the error in previous predictions propagates to the future ones. These findings aligns with the work of [14], [39].

3.5.2 Effect of The Cardinality of The Graph Nodes

We also explored the effect of considering an increasing number of chronologically ordered activities embedded within the graph structure on the classification accuracy when predicting the missing label(s). The cost of adding more activities leads to a larger training time and slower inference, thus answering the last question is crucial. From Table I we notice that the results discussed in the previous section hold consistently in general when there is a larger number of activities considered in the input graph. From the results obtained using the Extra-Sensory dataset we notice that the performance of all base models improves as more graph nodes are considered. Nonetheless, the the performance of HAR-GCNN is still significantly better than the base models and does not change significantly when the number of activities is increased from 5 to 25. These results imply that HAR-GCNN learns a proper representation that have a

constant performance irrelevant from the number of activities. While the CNN and LSTM baselines has an increase in the performance with more activities. We note that the performance of HAR-GCNN is still superior for increasing node cardinality than the one of the baseline models, suggesting that HAR-GCNN is economic in terms of training and inference time. The results obtained using the PAMAP dataset in Table I show that the performance across all models (HAR-GCNN and baseline models) does not change significantly with the graph cardinality. This constant performance can be attributed to the nature of the PAMAP datasets, which was collected from users being asked to perform tasks in pre-scripted manner. Models trained on this dataset might readily infer such pre-scripted sequence of activities and thus resulting in very high F-1 score and accuracy.

Chapter 4. Summary and Conclusions

4.1 Introduction

In this report, we presented HAR-GCNN, a deep learning model based on graph CNNs to predict missing activity labels leveraging the context of chronological sequences of these activities. We proposed an approach for modelling the activities as nodes in a fully connected graph, and leveraged the context of these activities in terms of the connection between graph nodes. We introduced a new training mechanism that forces the model to learn the chronological sequence without memorizing it, this was supported by the introduced experiments. To gain a quantitative understanding of our proposed strategy, we bench-marked our design against other commonly encountered deep network architectures having the same number of parameters and design structure as HAR-GCNN. Our results indicate that HAR-GCNN has superior performance in terms of F-1 score and accuracy in comparison to the baseline models. Further, our results suggest that learning a chronology of activities to predict the missing label is the key driver for performance improvements. When at least one of the activities in the graph is correctly labeled, HAR-GCNN performs better at classifying unlabeled sensor measurements than the other two baselines. Furthermore, our experimental results reveals that HAR-GCNN has a very stable performance, independently from the number of chronologically ordered activities considered within the input graph.

4.2 Summary and Conclusions

Our results indicate that HAR-GCNN has superior performance in terms of F-1 score and accuracy in comparison to the baseline models. Further, our results suggest that learning a chronology of activities to predict the missing label is the key driver for performance improvements. When at least one of the activities in the graph is correctly labeled, HAR-GCNN performs better at classifying unlabeled sensor measurements than the other two baselines. Furthermore, our experimental results reveals that HAR-GCNN has a very stable performance, independently from the number of chronologically ordered activities considered within the input graph.

References

1. O. Amft, D. Bannach, G. Pirkl, M. Kreil, and P. Lukowicz, “Towards wearable sensing-based assessment of fluid intake,” in
2. 2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). IEEE, 2010, pp. 298–303.
3. J. W. Lockhart, T. Pulickal, and G. M. Weiss, “Applications of mobile activity recognition,” in Proceedings of the 2012 ACM Conference on Ubiquitous Computing, 2012, pp. 1054–1058.
4. M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, “Convolutional neural networks for human activity recognition using mobile sensors,” in 6th International Conference on Mobile Computing, Applications and Services. IEEE, 2014, pp. 197–205.
5. R. Adaimi and E. Thomaz, “Leveraging active learning and conditional mutual information to minimize data annotation in human activity recognition,” Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 3, no. 3, Sep. 2019. [Online]. Available: https://doi-org.ezproxy.lib.utexas.edu/10.1145/3351228
6. A. Bulling, U. Blanke, and B. Schiele, “A tutorial on human activity recognition using body-worn inertial sensors,” ACM Computing Surveys (CSUR), vol. 46, no. 3, pp. 1–33, 2014.
7. H. Kwon, G. D. Abowd, and T. Plotz, “Handling annotation uncertainty in human activity recognition,” in¨ Proceedings of the 23rd International Symposium on Wearable Computers, 2019, pp. 109–117.
8. J. Mantyjarvi, J. Himberg, and T. Seppanen, “Recognizing human motion with multiple acceleration sensors,” in 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat. No. 01CH37236), vol. 2. IEEE, 2001, pp. 747–752.
9. J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” ACM SigKDD Explorations Newsletter, vol. 12, no. 2, pp. 74–82, 2011.
10. S. Pirttikangas, K. Fujinami, and T. Nakajima, “Feature selection and activity recognition from wearable sensors,” in International symposium on ubiquitious computing systems. Springer, 2006, pp. 516–527.
11. J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, “Deep learning for sensor-based activity recognition: A survey,” Pattern Recognition Letters, vol. 119, pp. 3–11, 2019.
12. N. Y. Hammerla, S. Halloran, and T. Plotz, “Deep, convolutional, and recurrent models for human activity recognition¨ using wearables,” arXiv preprint arXiv:1604.08880, 2016.
13. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.
14. L. Cao, Y. Wang, B. Zhang, Q. Jin, and A. V. Vasilakos, “Gchar: An efficient group-based context—aware human activity recognition on smartphone,” Journal of Parallel and Distributed Computing, vol. 118, pp. 67–80, 2018.
15. A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, “Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 14 424–14 432.
16. T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” arXiv preprint arXiv:1609.02907, 2016.

17. Y. Vaizman, N. Weibel, and G. Lanckriet, “Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 1, no. 4, pp. 1–22, 2018.
18. A. A. Varamin, E. Abbasnejad, Q. Shi, D. C. Ranasinghe, and H. Rezatofighi, “Deep auto-set: A deep auto-encoder-set network for activity recognition using wearables,” in Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, ser. MobiQuitous ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 246–253. [Online]. Available: https://doi.org/10.1145/3286978.3287024
19. Y. Vaizman, K. Ellis, and G. Lanckriet, “Recognizing detailed human contextin the wild from smartphones and smartwatches,” IEEE Pervasive Computing, vol. 16, no. 4, pp. 62–74, 2017.
20. R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Troster, J. d. R. Mill¨ an, and D. Roggen, “The opportunity´ challenge: A benchmark database for on-body sensor-based activity recognition,” Pattern Recognition Letters, vol. 34, no. 15, pp. 2033–2042, 2013.
21. A. Reiss and D. Stricker, “Creating and benchmarking a new dataset for physical activity monitoring,” in Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments, 2012, pp. 1–8.
22. R. Grzeszick, J. M. Lenk, F. M. Rueda, G. A. Fink, S. Feldhorst, and M. ten Hompel, “Deep neural network based human activity recognition for the order picking process,” in Proceedings of the 4th international Workshop on Sensor-based Activity Recognition and Interaction, 2017, pp. 1–6.
23. F. Moya Rueda, R. Grzeszick, G. A. Fink, S. Feldhorst, and M. Ten Hompel, “Convolutional neural networks for human activity recognition using body-worn sensors,” in Informatics, vol. 5, no. 2. Multidisciplinary Digital Publishing Institute, 2018, p. 26.
24. F. Cruciani, A. Vafeiadis, C. Nugent, I. Cleland, P. McCullagh, K. Votis, D. Giakoumis, D. Tzovaras, L. Chen, and R. Hamzaoui, “Feature learning for human activity recognition using convolutional neural networks,” CCF Transactions on Pervasive Computing and Interaction, vol. 2, no. 1, pp. 18–32, 2020.
25. S. Munzner, P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen, and R. D¨urichen, “Cnn-based sensor fusion techniques¨ for multimodal human activity recognition,” in Proceedings of the 2017 ACM International Symposium on Wearable Computers, 2017, pp. 158–165.
26. Y. Guan and T. Plotz, “Ensembles of deep lstm learners for activity recognition using wearables,”¨ Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 1, no. 2, pp. 1–28, 2017.
27. M. Inoue, S. Inoue, and T. Nishida, “Deep recurrent neural network for mobile human activity recognition with high throughput,” Artificial Life and Robotics, vol. 23, no. 2, pp. 173–185, 2018.
28. L. Lu, Y. Lu, R. Yu, H. Di, L. Zhang, and S. Wang, “Gaim: Graph attention interaction model for collective activity recognition,” IEEE Transactions on Multimedia, vol. 22, no. 2, pp. 524–539, 2019.
29. J. Zhang, F. Shen, X. Xu, and H. T. Shen, “Temporal reasoning graph for activity recognition,” IEEE Transactions on Image Processing, vol. 29, pp. 5491–5506, 2020.
30. J. Wu, L. Wang, L. Wang, J. Guo, and G. Wu, “Learning actor relation graphs for group activity recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 9964–9974.

31. D. Singh and C. K. Mohan, “Graph formulation of video activities for abnormal activity recognition,” Pattern Recognition, vol. 65, pp. 265–272, 2017.
32. J. Tang, X. Shu, R. Yan, and L. Zhang, “Coherence constrained graph lstm for group activity recognition,” IEEE transactions on pattern analysis and machine intelligence, 2019.
33. M. Stikic, D. Larlus, and B. Schiele, “Multi-graph based semi-supervised learning for activity recognition,” in 2009 international symposium on wearable computers. IEEE, 2009, pp. 85–92.
34. J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, “Graph neural networks: A review of methods and applications,” arXiv preprint arXiv:1812.08434, 2018.
35. A. Reiss and D. Stricker, “Introducing a new benchmarked dataset for activity monitoring,” in 2012 16th International Symposium on Wearable Computers. IEEE, 2012, pp. 108–109.
36. G. Li, M. Muller, A. Thabet, and B. Ghanem, “Deepgcns: Can gcns go as deep as cnns?” in¨ 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 9266–9275.
37. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in Advances in Neural Information Processing Systems 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alche-Buc, E. Fox, and R. Garnett, Eds.´ Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperativestyle-high-performance-deep-learning-library.pdf
38. K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” CoRR, vol. abs/1502.01852, 2015. [Online]. Available: http://arxiv.org/abs/1502.01852
39. T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, “Convolutional, long short-term memory, fully connected deep neural networks,” in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015, pp. 4580–4584.
40. S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” arXiv preprint arXiv:1803.01271, 2018.