HAR-GCNN: Deep Graph CNNs for Human Activity Recognition From Highly Unlabeled Mobile Sensor Data

Abduallah Mohamed∗, Fernando Lejarza†, Stephanie Cahail‡, Christian Claudel§, Edison Thomaz¶
The University of Texas at Austin
{*abduallah.mohamed, †lejarza, ‡stephaniecahail, §christian.claudel, ¶ethomaz}@utexas.edu

Abstract—The problem of human activity recognition from mobile sensor data applies to multiple domains, such as health monitoring, personal fitness, daily life logging, and senior care. 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 in real-time. Despite the likelihood of incorrect annotation or lack thereof, there is often an inherent chronology to human behavior. For example, we take a shower after we exercise. This implicit chronology can be used to learn unknown labels and classify future activities. In this work, we propose HAR-GCNN, 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 enforcing that the model predicts the missing activity labels by leveraging the known ones. HAR-GCNN shows superior performance relative to previously used baseline methods, improving classification accuracy by about 25% and up to 68% on different datasets.

Index Terms—deep graph convolutional neural network, human activity recognition, deep learning

I. INTRODUCTION

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], [2]. Technological advances regarding inertial sensors with longer battery life spans and improved computing capabilities have enabled gathering larger volumes of continuous data to be used for activity prediction [3]. Despite these recent advances, “ground truth annotation” which refers to labeling sensor readings with the activity being performed by the user remains a critical challenge [3]. Such annotation tasks are typically performed manually and occur either in real-time or post hoc once the activity has been completed. Such annotation tasks can be expensive, onerous, prone to human error, and even condition the user interfering in the activity itself [4]. 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 their ability to learn in the presence of missing labels, while achieving high accuracy in classifying human activity. A variety of deep and machine learning methods have been previously proposed for single and multi-label classification of human activity [5]–[7]. The missing label(s) can be predicted based on the readings of multiple sensors such as accelerometer, gyroscope, location, etc [6]. However, fewer approaches leverage the context of “neighboring” activities for predicting a missing label which has been shown to improve activity recognition [7], [8]. 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 of 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 the value in 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 [9]. 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 neighbouring 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 learning such graphs is deep Graph CNNs
(GCNNs) [10]. 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.
In this work, we show that this chronological correlation indeed
exists by examining two 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. Related Work Several supervised machine learning
techniques have been developed for HAR from mobile
sensor data including, for example, logistic regression, k-
nearest neighbors, decision trees, and multi-layer perceptrons
(MLP) [5]. These models often require substantial feature extraction, which can be time-consuming and rely heavily on domain knowledge [6]. Addressing these shortcomings, a
variety of prior works have explored unsupervised and semi-
supervised learning techniques on human activity data [2],
[11]. Several variants of graph-based models (GCNNs) have
been recently proposed [10], [12], [13], leveraging spatial and temporal properties in the data for collective activity recognition. The results in [12] suggest that exploiting contextual data from neighboring activities (i.e., nodes) can cause improved performance even relative to fully supervised approaches. GCNNs are an appealing modeling strategy that, to the best of our knowledge, has not been previously reported for HAR. Considering this, the main contributions of this work are: 1) A formulation for HAR exploiting the chronological context of the activities embedded within a graph structure. 2) A novel mechanism for training HAR-GCNN to learn to predict the missing labels in the input graph with high accuracy.

II. PROBLEM FORMULATION & DATASETS

Given a set of time series sensory measurements \( \mathcal{F} = \{f_t| t \in T\} \) collected from a variety of sensors, and sampled over a window of time steps \( T \), it is of interest to learn the classes of activities \( \mathcal{C} = \{c_t| t \in T\} \) associated with each of such measurements window. Each interval of measurements is associated with multi or single labels corresponding to different activities. Our problem formulation uses a set of prior and posterior measurements with known activity labels to predict the classes of unlabeled activities. That is, using pairs of \((f_{t-m}, c_{t-m}), \ldots, (f_t, c_t), \ldots, (f_{t+m}, c_{t+m})\) and \(f_t\) to predict \(c_t\), where \(m\) is the number of neighbour activities to be used. This means we pick a sequence of activities (represented by the sensor measurements across a duration) and their labels whether they are known or not and predict the class of the un-
known labels exploiting their sequential order. One important
remark is that the datasets under consideration are recorded
\textit{in-the-wild}. The \textit{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). This in turn reduces the bias within the collected data and results in a more natural chronology of activities from which our proposed model can learn. The Extra-Sensory [14] dataset is an example of such \textit{in-the-wild} measurements and results in a \textit{multi-label problem}. Also, to further validate the advantages of the framework proposed we conduct experiments using the PAMAP dataset [15] which is a \textit{single label problem}. 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. PAMAP dataset is being used as an example of a scripted dataset. This means the chronological order of activities indeed contains a bias, which we highlight the experiments section.

III. HAR-GCNN METHOD DESCRIPTION

![Fig. 2. HAR-GCNN model description. The CNN is just a single layer convolution with an activation function.](image)

Constructing the activity graph: We model the set of activities as a graph, which is defined as \( \mathcal{G} = (\mathcal{V}, \mathcal{E}) \) where \( \mathcal{V} = \{v| v \in \mathcal{V}\} \) is a set of vertices. Each vertex \( v = (f_t, c_t)\) contains both measurements and the associated multi-label or single-label class, with the exception of the nodes whose activity labels are missing and are set \( c = 0 \). The graph edges are fully connected and the weights of these edges are set to one. Model Description: First, the GCNN layer takes as input the aforementioned activity graph \( \mathcal{G} \). The GCNN model has a layer-wise structure defined as GCNN\((\mathcal{V}^{(l)}, A) = \sigma (A_{\text{norm}} \mathcal{V}^{(l)} W^{(l)})\) 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 \( A_{\text{norm}} \) [10]. The output of the GCNN is a graph embedding that represents the complete 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 single layer CNNs to process the graph embedding to predict the activities classes \( \mathcal{C} \). The output of these CNN layers is directly regressed against the proper loss function (Binary Cross Entropy or Cross-Entropy) to predict the labels \( \mathcal{C} \). Figure 2 describes the distinct steps included in our architecture. The CNN baseline The CNN
**TABLE I**

| Dataset       | % missing | Model  | CNN    | LSTM    |
|---------------|-----------|--------|--------|---------|
| Extra-Sensory | 33%       | 0.761/0.99 52 | 0.621/0.99 23 | 0.464/0.98 43 |
|               | 66%       | 0.792/0.99 52 | 0.531/0.98 69 | 0.377/0.97 95 |
| PAMAP         | 33%       | 0.999/0.99 92 | 0.975/0.97 57 | 0.973/0.97 26 |
|               | 66%       | 0.999/0.99 94 | 0.902/0.90 22 | 0.903/0.90 26 |

We attempt to predict their labels. Evidently, HAR-GCNN outperforms both CNN and LSTM, and almost reaching the upper performance bound of 1.00. We posit that the observed performance improvements stem from the embedded graph structure, in which the graph edges can more deeply capture the relationship between labels and their features. Therefore, if at least one activity is correctly classified, then the remaining nodes will be as well. We note that the CNN baseline significantly outperforms the LSTM model across the two datasets considered. Thus CNN-based approaches are more suitable for the learning task at hand, aligning with previous results [9]. In this paper, we presented HAR-GCNN, a deep graph model 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. We introduced a new training mechanism that allows the model to learn from the chronological sequence to improve performance. We benchmarked our design against baselines showing that HAR-GCNN has superior relative performance in terms of F-1 score and accuracy. The results shown indicate that learning the chronology of activities to predict missing label(s) is a key driver for predictive performance improvements. Naturally, a limitation of HAR-GCNN is that if no neighbouring activity label is known its performance is expected to degrade, unlike CNNs and RNNs.

**References**

[1] 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 10th International Conference on Mobile Computing, Applications and Services. IEEE, 2014, pp. 197–205.

[2] R. Adami and E. Theizaz, “Leveraging machine 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]: https://doi.org/10.1145/3351228.3351232.

[3] 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.

[4] H. Kwon, G. D. Aboud, and T. Pietz, “Handling annotation uncertainty in human activity recognition,” in Proceedings of the 23rd International Symposium on Wearable Computers, 2019, pp. 109–117.

[5] J. Mantyjarvi, J. Himberg, and T. Seppanen, “Recognizing human motion with multiple acceleration sensors,” in 2001 IEEE International Conference on Systems, Man and Cybernetics. -Systems and E-Man for Cybernetics in Cyberspace (Cat. No. 01CH37236). vol. 2, IEEE, 2001, pp. 747–752.

[6] 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.

[7] N. Y. Hammerla, S. Halloran, and T. Pietz, “Deep, convolutional, and recurrent models for human activity recognition using wearables,” arXiv preprint arXiv:1604.08880, 2016.

[8] L. Cao, Y. Wang, B. Zhang, Q. Jin, and A. V. Vasilakos, “Ggchar: An efficient group-based context-aware human activity recognition on smartphones,” Journal of Parallel and Distributed Computing, vol. 118, pp. 67–80, 2018.

[9] 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.

[10] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” arXiv preprint arXiv:1609.02907, 2016.

[11] 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]: https://doi.org/10.1145/3286978.3287024.

[12] M. Stikic, D. Larlus, and B. Schiele, “Graph-based semi-supervised learning for activity recognition,” in 2009 international symposium on wearable computers. IEEE, 2009, pp. 85–92.

[13] L. Lu, Y. Lu, R. Yu, H. Di, L. Zhang, and S. Wang, “Gain: Graph attention interaction model for collective activity recognition,” IEEE Transactions on Multimedia, vol. 22, no. 2, pp. 324–339, 2019.

[14] Y. Vaizman, K. Ellis, and G. Lanckriet, “Recognizing detailed human context in the wild from smartphones and smartwatches.” IEEE Pervasive Computing, vol. 16, no. 4, pp. 62–74, 2017.

[15] 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.