Human Activity Recognition using Recurrent Neural Networks

Deepika Singh[1], Erinc Merdivan[1], Ismini Psychoula[2], Johannes Kropf[1], Sten Hanke[1], Matthieu Geist[3], and Andreas Holzinger[4]

[1] AIT Austrian Institute of Technology, Austria
deepika.singh@ait.ac.at, erinc.merdivan@ait.ac.at
[2] School of Computer Science and Informatics, De Montfort University, UK
[3] CentraleSupelec, France
[4] Holzinger Group, HCI-KDD, Institute for Medical Informatics/Statistics, Medical University Graz, Austria

Abstract. Human activity recognition using smart home sensors is one of the bases of ubiquitous computing in smart environments and a topic undergoing intense research in the field of ambient assisted living. The increasingly large amount of data sets calls for machine learning methods. In this paper, we introduce a deep learning model that learns to classify human activities without using any prior knowledge. For this purpose, a Long Short Term Memory (LSTM) Recurrent Neural Network was applied to three real world smart home datasets. The results of these experiments show that the proposed approach outperforms the existing ones in terms of accuracy and performance.

Keywords: machine learning, deep learning, human activity recognition, sensors, ambient assisted living, LSTM

1 Introduction

Human Activity recognition has been an active research area in the last decades due to its applicability in different domains and the increasing need for home automation and convenience services for the elderly [1]. Among them, activity recognition in Smart Homes with the use of simple and ubiquitous sensors, has gained a lot of attention in the field of ambient intelligence and assisted living technologies for enhancing the quality of life of the residents within the home environment [2].

The goal of activity recognition is to identify and detect simple and complex activities in real world settings using sensor data. It is a challenging task, as the data generated from the sensors are sometimes ambiguous with respect to the activity taking place. This causes ambiguity in the interpretation of activities. Sometimes the data obtained can be noisy as well. Noise in the data can be caused by humans or due to error in the network system which fails to give correct sensor readings. Such real-world settings are full of uncertainties and calls for methods to learn from data, to extract knowledge and helps in making

decisions. Moreover, the inverse probability allows to infer unknowns and to make predictions [3].

Consequently, many different probabilistic, but also non-probabilistic models, have been proposed for human activity recognition. Patterns corresponding to the activities are detected using sensors such as accelerometers, gyroscopes or passive infrared sensors, *etc.*, either using feature extraction on sliding window followed by classification [4] or with Hidden Markov Modeling (HMM) [5].

In recent years, there has been a growing interest in deep learning techniques. Deep learning is a general term for neural network methods which are based on learning representations from raw data and contain more than one hidden layer. The network learns many layers of non-linear information processing for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. Deep learning techniques have already outperformed other machine learning algorithms in applications such as computer vision [6], audio [7] and speech recognition [8].

In this paper, we introduce a recurrent neural network model for human activity recognition. The classification of the human activities such as cooking, bathing, and sleeping is performed using the Long Short-Term Memory classifier (LSTM) on publicly available Benchmark datasets [9]. An evaluation of the results has been performed by comparing with the standardized machine learning algorithms such as Naive Bayes, HMM, Hidden Semi-Markov Model (HSMM) and Conditional Random Fields (CRF).

The paper is organized as follows. Section 2 presents an overview of activity recognition models and related work in machine learning techniques. Section 3 introduces Long Short-Term Memory (LSTM) recurrent neural networks. Section 4 describes the datasets that were used and explains the results in comparison to different well-known algorithms. Finally, Section 5 discusses the outcomes of the experiments and suggestions for future work.

2 Related work

In previous research, activity recognition models have been classified into data-driven and knowledge-driven approaches. The data-driven approaches are capable of handling uncertainties and temporal information [10] but require large datasets for training and learning. Unfortunately, the availability of large real world datasets is a major challenge in the field of ambient assisted living. The knowledge-driven techniques are used in predictions and follow a description-based approach to model the relationships between sensor data and activities. These approaches are easy to understand and use but they cannot handle uncertainty and temporal information [11].

Various approaches have been explored for activity recognition, among them the majority of the techniques focuses on classification algorithms such as Naive Bayes (NB) [12], Decision Trees [13], HMM [5] , CRF [14], Nearest Neighbor (NN) [15], Support Vector Machines (SVM) [16] and different boosting techniques.

A simple probabilistic classifier in machine learning is the Naive Bayes classifier which yields good accuracy with large amounts of sample data but does not model any temporal information. The HMM, HSMM, and CRF are the most popular approaches for including such temporal information. However, these approaches sometimes discard pattern sequences that convey information through the length of intervals between events. This motivates the study of recurrent neural networks (RNN) which promises the recognition of patterns that are defined by temporal distance [17].

LSTM is a recurrent neural network architecture that is designed to model temporal sequences and learn long-term dependency problems. The network is well suited for language modeling tasks; it has been shown that the network in combination with clustering techniques increases the training and testing time of the model [18] and outperforms the large scale acoustic model in speech recognition systems [19].

3 LSTM Model

LSTM is a recurrent neural network architecture that was proposed in [20]. Another version without a forget gate was later proposed in [21] and extended in [22]. LSTM has been developed in order to deal with gradient decay or gradient blow-up problems and can be seen as a deep neural network architecture when unrolled in time. The LSTM layer's main component is a unit called memory block. An LSTM block has three gates which are input, output and forget gates. These gates can be seen as write, read and reset operations for the cells. An LSTM cell state is the key component which carries the information between each LSTM block. Modifications to the cell state are controlled with the three gates described above. An LSTM single cell, as well as how each gate is connected to each other and the cell state itself, can be seen in Figure 1.

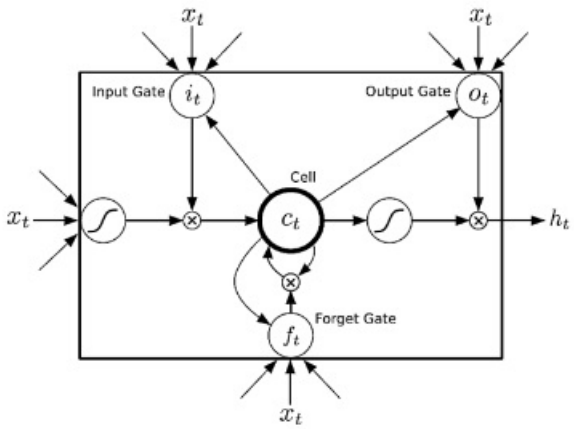

Fig. 1: LSTM single cell image [23].

Each gate and cell state are governed by multiplicative equations that are given by:

$$
\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i),\\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f),\\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o),\\
c_t &= f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c),\\
h_t &= o_t \tanh c_t,
\end{aligned}
$$

with W being the weight matrix and x is the input, σ being the sigmoid and tanh is the hyperbolic tangent activation function. The terms i, f and o are named after their corresponding gates and c represents the memory cell [23].

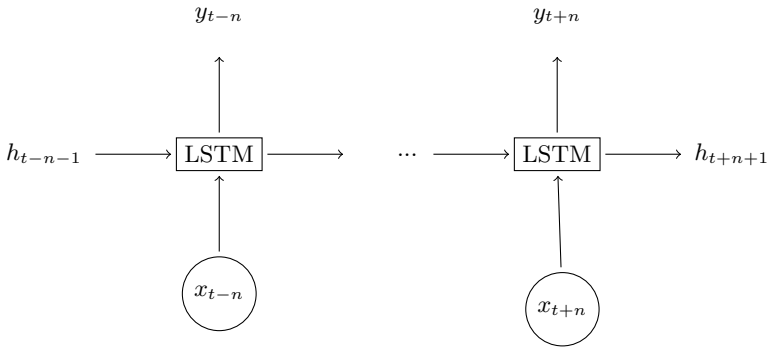

Fig. 2: Illustrations of an LSTM network with x being the binary vector for sensor input and y being the activity label prediction of the LSTM network.

By unrolling LSTM single cells in time we construct an LSTM layer where h_t is the hidden state and y_t is the output at time t as shown in Figure 2.

4 Experiments

4.1 Dataset

Publicly available and annotated sensor datasets have been used to evaluate the performance of the proposed approach [9]. In this dataset, there are three houses with different settings to collect sensory data. The three different houses were all occupied by a single user named A, B, and C respectively. Each user recorded and annotated their daily activities. Different number of binary sensors were deployed in each house such as passive infrared (PIR) motion detectors to detect motion in a specific area, pressure sensors on couches and beds to identify the user's presence, reed switches on cupboards and doors to measure open or

close status, and float sensors in the bathroom to measure toilet being flushed or not. The data were annotated using two approaches: (1) keeping a diary in which the activities were logged by hand and (2) with the use of a blue tooth headset along with a speech recognition software. A total of three datasets were collected from the three different houses. Details about the datasets are shown in Table 1 where each column shows the details of the house with the information of the user living in it, the sensors placed in the house and the number of activity labels that were used.

Table 1: Details of the datasets.

	House A	House B	House C
Age	26	28	57
Gender	Male	Male	Male
Setting	Apartment	Apartment	House
Rooms	3	2	6
Duration	25days	14days	19days
Sensors	14	23	21
Activities	10	13	16
Annotation	Bluetooth	Diary	Bluetooth

The data used in the experiments have different representation forms. The first form is raw sensor data, which are the data received directly from the sensor. The second form is last-fired sensor data which are the data received from the sensor that was fired last. The last firing sensor gives continuously 1 and changes to 0 when another sensor changes its state. For each house, we left one day out of the data to be used later for the testing phase and used the rest of the data for training. We repeated this for every day and for each house. Separate models are trained for each house since the number of sensors varies, and a different user resides in each house. Sensors are recorded at one-minute intervals for 24 hours, which totals in 1440 length input for each day.

4.2 Results

The results presented in Table 2 show the performance of the LSTM model on raw sensor data in comparison with the results of NB, HMM, HSMM and CRF [9]. Table 3 shows the results of the LSTM model on last-fired sensor data again in comparison with the results of NB, HMM, HSMM and CRF. For the LSTM model, a time slice of (70) with hidden state size (300) are used. For the optimization of the network, Adam is used with a learning rate of 0.0004 [24] and Tensorflow was used to implement the LSTM network. The training took place on a Titan X GPU and the time required to train one day for one house is approximately 30 minutes, but training times differ amongst the houses. Since different houses have different days we calculated the average accuracy amongst

all days. The training is performed using a single GPU but the trained models can be used for inference without losing performance when there is no GPU.

Table 2: Results of raw sensor data

Model	House A	House B	House C
Naive Bayes	77.1 ± 20.8	80.4 ± 18.0	46.5 ± 22.6
HMM	59.1 ± 28.7	63.2 ± 24.7	26.5 ± 22.7
HSMM	59.5 ± 29.0	63.8 ± 24.2	31.2 ± 24.6
CRF	89.8 ± 8.5	78.0 ± 25.9	46.3 ± 25.5
LSTM(Ours)	$\mathbf{89.8 \pm 8.2}$	$\mathbf{85.7 \pm 14.3}$	$\mathbf{64.22 \pm 21.9}$

Table 2 shows the results of different models on raw data from three different houses. The LSTM model has the best performance for all three data sets. In House B and House C, LSTM improves the best result significantly especially on House C where the improvement is approximately 40%.

Table 3: Results of last-fired sensor data

Model	House A	House B	House C
Naive Bayes	95.3 ± 2.8	86.2 ± 13.8	87.0 ± 12.2
HMM	89.5 ± 8.4	48.4 ± 26.0	83.9 ± 13.9
HSMM	91.0 ± 7.2	67.1 ± 24.8	84.5 ± 13.2
CRF	$\mathbf{96.4 \pm 2.4}$	$\mathbf{89.2 \pm 13.9}$	$\mathbf{89.7 \pm 8.4}$
LSTM	95.3 ± 2.0	88.5 ± 12.6	85.9 ± 10.6

Table 3 shows the results on last fired data from three different houses using the same models as in Table 2. The LSTM model did not improve the results in this section but it matched the best performance for two data sets with a slight drop in House C.

5 Discussion

The results presented in this paper show that the deep learning based approaches for activity recognition from raw sensory inputs can lead to significant improvement in performance, increased accuracy, and better results. As shown in Section 4.2 our LSTM based activity predictor matched or outperformed existing probabilistic models such as Naive Bayes, HMM, HSMM and CRF on raw input and in one case improved the best result by 40%. Predicting on raw input also reduces the human efforts required on data preprocessing and handcrafting features which can be very time consuming when it comes to an AAL (Ambient Assisted Living) environment.

6 Future Work

Our future work will focus on reducing the variance on our predictions and early stopping criteria while training on different days. The LSTM model has different hyperparameters which affect the performance of the model significantly. Different optimization and hyperparameter search techniques could be investigated in the future. Since the LSTM model has proven to be superior on raw data it would be interesting to also apply other deep learning models. One problem is that deep learning badly captures model uncertainty. Bayesian models offer a framework to reason about model uncertainty. Recently, Yarin & Ghahramani (2016) [25] developed a theoretical framework casting dropout training in deep neural networks as approximate Bayesian inference in deep Gaussian processes. This mitigates the problem of representing uncertainty in deep learning without sacrificing either computational complexity or test accuracy.

Acknowledgement

This work has been funded by the European Union Horizon2020 MSCA ITN ACROSSING project (GA no. 616757). The authors would like to thank the members of the project's consortium for their valuable inputs.

References

1. Roecker, C., Ziefle, M., Holzinger, A.: Social inclusion in ambient assisted living environments: Home automation and convenience services for elderly users. In: Proceedings of the International Conference on Artificial Intelligence (ICAI 2011). CSERA Press, New York (2011) 55–59
2. Chen, L., Hoey, J., Nugent, C.D., Cook, D.J., Yu, Z.: Sensor-based activity recognition. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **42** (2012) 790–808
3. Holzinger, A.: Introduction to machine learning and knowledge extraction (MAKE). Machine Learning and Knowledge Extraction **1** (2017) 1–20
4. Roggen, D., Cuspinera, L.P., Pombo, G., Ali, F., Nguyen-Dinh, L.V.: Limited-memory warping lcss for real-time low-power pattern recognition in wireless nodes. In: European Conference on Wireless Sensor Networks, Springer (2015) 151–167
5. Duong, T.V., Bui, H.H., Phung, D.Q., Venkatesh, S.: Activity recognition and abnormality detection with the switching hidden semi-markov model. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Volume 1., IEEE (2005) 838–845
6. Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th annual international conference on machine learning, ACM (2009) 609–616
7. Lee, H., Pham, P., Largman, Y., Ng, A.Y.: Unsupervised feature learning for audio classification using convolutional deep belief networks. In: Advances in neural information processing systems. (2009) 1096–1104

8. Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., et al.: Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine **29** (2012) 82–97
9. Kasteren, T.L., Englebienne, G., Kröse, B.J.: Human activity recognition from wireless sensor network data: Benchmark and software. Activity recognition in pervasive intelligent environments (2011) 165–186
10. Yuen, J., Torralba, A.: A data-driven approach for event prediction. Computer Vision–ECCV 2010 (2010) 707–720
11. Ye, J., Stevenson, G., Dobson, S.: Kcar: A knowledge-driven approach for concurrent activity recognition. Pervasive and Mobile Computing **19** (2015) 47–70
12. Tapia, E.M., Intille, S.S., Larson, K.: Activity recognition in the home using simple and ubiquitous sensors. In: International Conference on Pervasive Computing, Springer (2004) 158–175
13. Bao, L., Intille, S.S.: Activity recognition from user-annotated acceleration data. In: International Conference on Pervasive Computing, Springer (2004) 1–17
14. Van Kasteren, T., Noulas, A., Englebienne, G., Kröse, B.: Accurate activity recognition in a home setting. In: Proceedings of the 10th international conference on Ubiquitous computing, ACM (2008) 1–9
15. Wu, W., Dasgupta, S., Ramirez, E.E., Peterson, C., Norman, G.J.: Classification accuracies of physical activities using smartphone motion sensors. Journal of medical Internet research **14** (2012) e130
16. Zhu, Y., Nayak, N.M., Roy-Chowdhury, A.K.: Context-aware activity recognition and anomaly detection in video. IEEE Journal of Selected Topics in Signal Processing **7** (2013) 91–101
17. Ribbe, M.W., Ljunggren, G., Steel, K., Topinkova, E., Hawes, C., Ikegami, N., Henrard, J.C., JÓNnson, P.V.: Nursing homes in 10 nations: a comparison between countries and settings. Age and ageing **26** (1997) 3–12
18. Sundermeyer, M., Schlüter, R., Ney, H.: Lstm neural networks for language modeling. In: Interspeech. (2012) 194–197
19. Sak, H., Senior, A., Beaufays, F.: Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: Fifteenth Annual Conference of the International Speech Communication Association. (2014)
20. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9** (1997) 1735–1780
21. Gers, F.A., Schmidhuber, J., Cummins, F.: Learning to forget: Continual prediction with lstm. Neural computation **12** (2000) 2451–2471
22. Gers, F.A., Schraudolph, N.N., Schmidhuber, J.: Learning precise timing with lstm recurrent networks. Journal of machine learning research **3** (2002) 115–143
23. Zhang, S., Zheng, D., Hu, X., Yang, M.: Bidirectional long short-term memory networks for relation classification. In: PACLIC. (2015)
24. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
25. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Balcan, M.F., Weinberger, K.Q., eds.: Proceedings of The 33rd International Conference on Machine Learning (ICML). Volume 48., PMLR (2016) 1050–1059