LSTM Networks for Mobile Human Activity Recognition

5 authors, including:

Kunhua Zhong
Chinese Academy of Sciences
11 PUBLICATIONS **99** CITATIONS

SEE PROFILE

Xueliang Zhao
Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sci…
8 PUBLICATIONS **65** CITATIONS

SEE PROFILE

LSTM Networks for Mobile Human Activity Recognition

Yuwen Chen[*], Kunhua Zhong, Ju Zhang, Qilong Sun and Xueliang Zhao

Chongqing institute of green and intelligent technology Chinese Academy of Sciences

[*]Corresponding author

Abstract—**A lot of real-life mobile sensing applications are becoming available. These applications use mobile sensors embedded in smart phones to recognize human activities in order to get a better understanding of human behavior. In this paper, we propose a LSTM-based feature extraction approach to recognize human activities using tri-axial accelerometers data. The experimental results on the (WISDM) Lab public datasets indicate that our LSTM-based approach is practical and achieves 92.1% accuracy.**

Keywords-Activity recognition, Deep learning, Long short memory network

I. INTRODUCTION

Although human activity recognition (HAR) has been studied extensively in the past decade, HAR on smartphones is a relatively new area. HAR is a classical multi-variate time series or sequence analysis problem, for which the task is to detect and classify those contiguous portions of sensor data streams that cover activities of interest for the target application. The predominant approach to HAR is based on a sliding window procedure, where a fixed length analysis window is shifted along the signal sequence for frame extraction. Preprocessing then transforms raw signal data into feature vectors, which are subjected to statistical classifiers that eventually provide activity hypotheses.

Activity recognition has a wide range of applications in mobile applications — from fitness and health tracking to context-based advertising and employee monitoring. Context-aware applications can customize their behaviour based on the current activity. For example [1, 2, 3, 4] have used the smartphone accelerometer to recognize movements, such as walking and running. Advances in the area of mobile sensing enable users to: quantify their sleep and exercise patterns [5], monitor personal commute behaviors [6], track their emotional state [7], or even measure how long they spend queuing in retail stores [8].

II. RELATED WORK

Extensive work has been done in the area of HAR using smartphone sensors, which has been summarized in [9]. Feature extraction for AR is an important task. Statistical features such as mean, standard deviation, entropy and correlation coefficients, etc. are the most widely used hand-crafted features in AR [10]. Fourier transform and wavelet transform [11] are another two commonly used hand-crafted features, while discrete cosine transform (DCT) have also been applied with promising results [12], as well as auto-regressive model coefficients [13]. Recently, time-delay embedding [14] have been applied for activity recognition. It adopts nonlinear time series analysis to extract features from time series and shows a significant improvement on periodic activities recognition .However, the features from time-delay embedding are less suitable for non-periodic activities. The authors in [15] firstly introduce feature learning methods to the area of activity recognition, they used Deep Belief Networks (DBN) and PCA to learn features for activity recognition in ubiquitous computing. The authors in [16] following the work of [15] applied shift-invariant sparse coding technique. The authors in [17] also used sparse coding to learn features.

The features used in most of researches on HAR are selected by hand. Designing hand-crafted features in a specific application requires domain knowledge [18], and maybe result in loss of information after extracting features. This problem is not unique to activity recognition. It has been well-studied in other research areas such as image recognition[19],where different types of features need to be extracted when trying to recognize a handwriting as opposed to recognizing faces. In recent years, due to advances of the processing capabilities, a large amount of Deep Learning (DL) techniques have been developed and successful applied in recognition tasks [20, 21]. These techniques allow an automatic extraction of features without any domain knowledge

In this work, we propose an approach based on Long - Short Term Memory(LSTM) [22]to recognize activities in various application domains.

III. LSTM-BASED ACTIVITY RECOGNITION

Long short-term memory (LSTM) is a recurrent neural network (RNN) architecture published[22] in 1997 by Sepp Hochreiter and Jürgen Schmidhuber. Unlike traditional RNNs, an LSTM network is well-suited to learn from experience to classify, process and predict time series when there are very long time lags of unknown size between important events. LSTM model which introduces a new structure called a memory cell (see Figure 1 below). A memory cell is composed of four main elements: an input gate, a neuron with a self-recurrent connection, a forget gate and an output gate. The self-recurrent connection has a weight of 1.0 and ensures that, barring any outside interference, the state of a memory cell can remain constant from one time step to another. The gates serve to modulate the interactions between the memory cell itself

and its environment. The input gate can allow incoming signal to alter the state of the memory cell or block it. On the other hand, the output gate can allow the state of the memory cell to have an effect on other neurons or prevent it. Finally, the forget gate can modulate the memory cell's self-recurrent connection, allowing the cell to remember or forget its previous state, as needed.
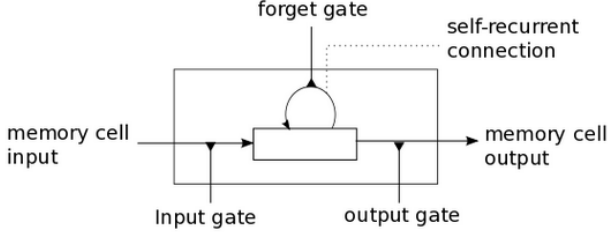
Following equations show how a layer of memory cells is updated at every timestep t.

x_t is the input to the memory cell layer at time t

W_i, W_f, W_c, W_o, U_i, U_f, U_c, U_o and V_o are weight matrices

b_i, b_f, b_c and b_o are bias vectors

i_t is the input gate, $\widetilde{C_t}$ the candidate value for the states of the memory cells at time t

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \qquad (1)$$

$$\widetilde{C_t} = tanh(W_c x_t + U_c h_{t-1} + b_c) \qquad (2)$$

f_t is the activation of the memory cells' forget gates at time t

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \qquad (3)$$

C_t is the memory cells' new state at time t

$$C_t = i_t * \widetilde{C_t} + f_t * C_{t-1} \qquad (4)$$

With the new state of the memory cells, compute the value of their output gates and their outputs :

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o C_t + b_o) \qquad (5)$$

$$h_t = o_t * tanh(C_t) \qquad (6)$$

HAR is a classical time series or sequence analysis problem, it is suitable to deal with the LSTM. In the next section, we will introduce our model LSTM-based Activity Recognition. Figure 2 show the structure of the model. Given a 3D acceleration time series we use a sliding window with a length of N values to extract input data for the LSTM. In order to form a richer representation of the data, our model has two LSTM layers.

X is the acceleration in the x direction as measured by phone's accelerometer. Y is the y direction, Z is the Z direction. In this paper, we combine the acceleration of the three directions into a three-dimensional vector with sliding window with a length of N value. So the input data for the LSTM is a time series of the NX3 matrix. Feature extraction of accelerometer raw data through long short memory network with N timesteps. We connected the output of the LSTM and form a new Feature vector. Finally, the feature vector is classified by a multi classifier.
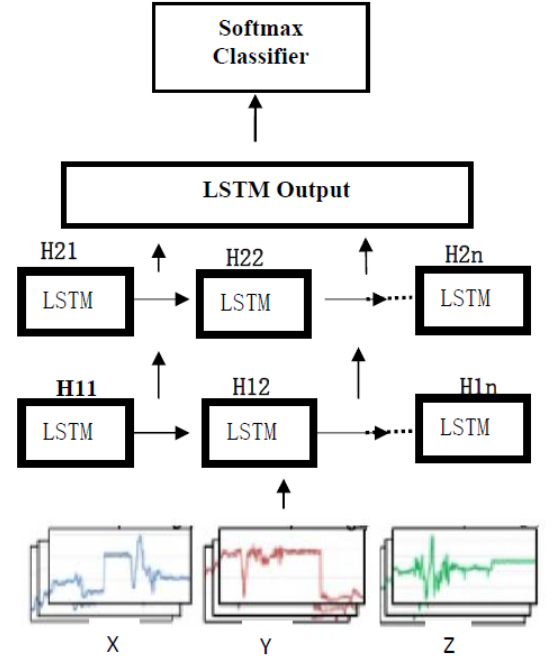

FIGURE II. STRUCTURE OF LSTM FOR HUMAN ACTIVITY RECOGNITION

IV. EXPERIMENTAL ANALYSIS

We selected one publically available datasets for our evaluation, which have been released by the Wireless Sensor Data Mining(WISDM) Lab[23] The Datasets related to human activities in different context sand have been recorded using tri-axial accelerometers. The datasets have 1,098,207 records and 6 Class attributes (class distribution as Figure 3 show.) In this paper, the sensor data was segmented using a sliding window with a size of 50 continuous samples. The acceleration values were normalized to have zero mean and unit standard variance for LSTM-based. We split the data into two parts.one is for training, another is for validation our model.

Train the training data through the LSTM-model mentioned above. accuracy show Figure 4. the LSTM-based model achieves classification accuracy of 95.1%. Our training metrics are not smooth and fluctuations because we use small data sizes and the distribution of our training data is not uniform. If we used larger data we would get a smoother blue line. The results suggest that we need more data, stronger regularization, or fewer model parameters.

To analyze the results in more detail, we show the confusion matrix for the validation datasets, The confusion matrices indicate that many of the prediction error are due to confusion between these two activities: "Jogging", "Upstairs". This is because these two activities are relatively similar.
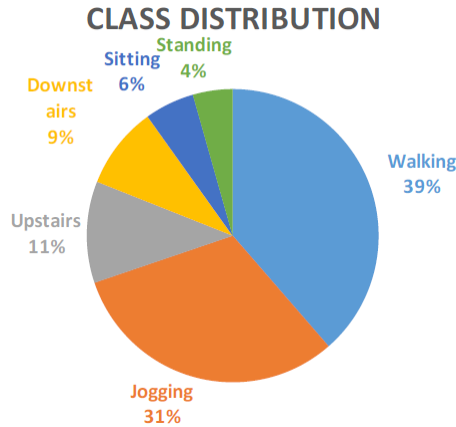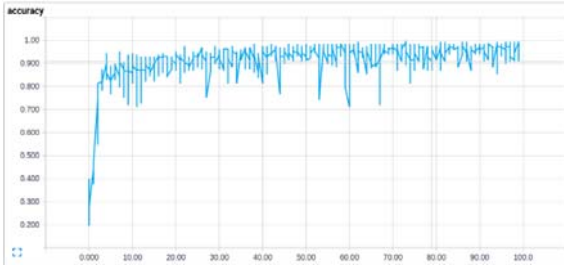

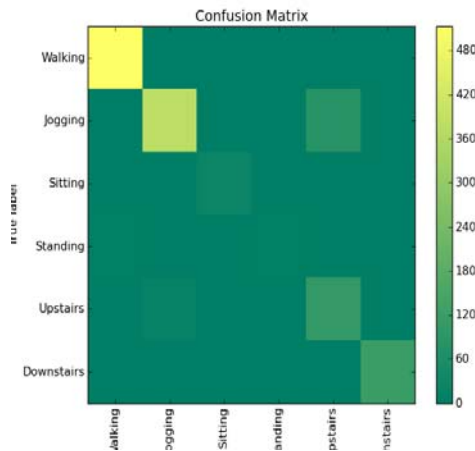FIGURE III. CLASS DISTRIBUTION FOR DATASETS

FIGURE IV. ACCURACY

FIGURE V. CONFUSION MATRIX

V. CONCLUSION

In this paper, we have proposed a LSTM-based feature extraction approach. The experimental results have shown that the LSTM-based approach is practical and achieves best 92.1% accuracy.

Experiments with larger datasets are needed to further study the robustness of the proposed technique. Further improvements maybe achieved by using more data and regularization.

REFERENCES

[1]]M. Fahim, I. Fatima, S. Lee, and Y. T. Park, "Efm: evolutionary fuzzy model for dynamic activities recognition using a smartphone accelerometer," Applied Intelligence, pp. 1–14, 2013

[2] O. D. Lara and M. A. Labrador, "A mobile platform for real-time human activity recognition," inProceedings of the IEEE Consumer Communications and Networking Conference (CCNC '12), pp. 667–671, IEEE, 2012.

[3] 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.

[4] L. Sun, D. Zhang, B. Li, B. Guo, and S. Li, "Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations," in Proceedings of the 7th international conference on Ubiquitous intelligence and computing (UIC '10), pp. 548–562, Springer, Berlin, Germany

[5] S. Consolvo, et al. Activity Sensing in the Wild: A Field Trial of UbiFit Garden. In CHI '08.

[6] S. Reddy, et al. Using mobile phones to determine transportation modes. ACM Trans. Sen. Netw., 6(2):13:1–13:27, Mar. 2010.

[7] K. K. Rachuri, et al. Emotionsense: A mobile phones based adaptive platform for experimental social psychology research. In UbiComp '10.

[8] Y. Wang, et al. Tracking human queues using single-point signal monitoring. In MobiSys '14.

[9] Incel, O.D., Kose, M., Ersoy, C.: A review and taxonomy of activity recognition on mobile phones. BioNanoScience 3(2), 145–171 (2013)

[10] D.Figo,P.C.Diniz,D.R.Ferreira,andJ.M.Cardoso.Preprocessingtechniques forcontext recognition from accelerometer data. Personal and Ubiquitous Computing, 14(7):645–662, 2010.

[11] Z.HeandL.Jin. Activity recognition from cceleration data based on discrete consine transform and svm. InSystems,ManandCybernetics.SMC2009.IEEEInternationalConference on, pages 5041–5044. IEEE, 2009.

[12] T. Tamura, M. Sekine, M. Ogawa, T. Togawa, and Y. Fukui. Classification of acceleration waveformsduringwalkingbywavelettransform. Methodsofinformationinmedicine,36(45):356–359, 1997

[13] Z.-Y.HeandL.-W.Jin. Activity recognition from acceleration data using a rmodel representation and svm. In Machine Learning and Cybernetics, 2008 International Conference on, volume 4, pages 2245–2250. IEEE, 2008.

[14] J. Frank, S. Mannor, and D. Precup. Activity and gait recognition with time-delay embeddings. In AAAI, 2010.

[15] Pl¨otz, T., Hammerla, N.Y., Olivier, P.: Feature learning for activity recognition in ubiquitous computing. In: Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence, vol. 2, pp. 1729–1734. AAAI Press (2011)

[16] Vollmer, C., Gross, H.-M., Eggert, J.P.: Learning features for activity recognition with shift-invariant sparse coding. In: Mladenov, V., Koprinkova-Hristova, P., Palm, G., Villa, A.E.P., Appollini, B., Kasabov, N. (eds.) ICANN 2013. LNCS, vol. 8131, pp. 367–374. Springer, Heidelberg (2013) 5. [17]Bhattacharya, S., Nurmi, P., Hammerla, N., Pl¨otz, T.: Using unlabeled data in a sparse-coding framework for human activity recognition. arXiv preprint arXiv:1312.6995 (2013)

[17] T. Pl¨otz, N. Y. Hammerla, and P. Olivier. Feature learning for activity recognition in ubiquitous computing. InProceedingsoftheTwenty-SecondIJCAIVolumeTwo,pages1729–1734. AAAI Press, 2011.

[18] D. G. Lowe. Object recognition from local scale-invariant features. In Computer vision, 1999. The proceedings of the seventh IEEE international conference on, volume 2, pages 1150–1157. Ieee, 1999.

[19] U.BagciandL.Bai. A comparison of daubechies and gabor wavelets for classification of mr images. In Signal Processing and Communications, 2007. ICSPC 2007. IEEE International Conference on, pages 676–679. IEEE, 2007.

[20] Y. Tang, R. Salakhutdinov, and G. Hinton. Robust boltzmann machines for recognition and denoising. InComputerVisionandPatternRecognition(CVPR),2012IEEEConference on, pages 2264–2271. IEEE, 2012.

[21] Sepp Hochreiter and Jürgen Schmidhuber (1997). "Long short-term memory" . Neural Computation 9 (8): 1735–1780.

[23] GM Weiss, JR Kwapisz,S Moore Activity recognition using cell phone accelerometers Proceedings of the Fourth International Works on konwledge Discovery from sensor Data., 2010, 12(12):74-82