Accelerometer-based human activity recognition using 1D convolutional neural network

S Tsokov, M Lazarova and A Aleksieva-Petrova
Technical University of Sofia, Faculty of Computer Systems and Technologies, Department Computer Systems, 8 Kliment Ochridsky blvd., Sofia 1000, Bulgaria
e-mail: milaz@tu-sofia.bg

Abstract. Human activity recognition (HAR) is an important research field with a variety of applications in healthcare monitoring, fitness tracking and in user-adaptive systems in smart environments. The performance of the activity recognition system is highly dependent on the features extracted from the sensor data which makes the selection of appropriate features a very important part of HAR. A 1D CNN model trained on accelerometer data is suggested in the paper for automatic feature extraction in a HAR system. A semi-automatic approach is used that effectively and efficiently determines the number of convolutional layers in the network, the number of kernels and the size of the kernels. The experimental results show that the suggested model outperforms several existing recognition approaches that use the same data set.

1. Introduction

Human activity recognition (HAR) is a popular field for research. HAR systems usually consist of a set of sensors, the information from which is processed by machine learning algorithms that classify individual activities, such as walking, eating, running, etc. Popular traditional algorithms used to recognize activities are: Support Vector Machine (SVM), Hidden Markov Model, K-Nearest Neighbor (KNN), Decision Tree, Naïve Bayes classifier and Conditional Random Fields [1], [2].

The data used in activity recognition systems can be collected by different types of sensors [3]. One of them is video cameras. They usually provide high accuracy, but their performance depends on lighting, visual angle and other factors [4]. This type of sensors are not popular for use in smart homes because they are intrusive – there’s an invasion of the privacy of the occupants. A less intrusive type of sensors are those that are worn on the body (such as accelerometers, magnetometers, etc.). These sensors are particularly suitable for collecting information on specific movements, but they lack contextual information (environmental information), which can lead to difficulties in recognizing actions that are very similar [3]. Other disadvantages are that they can be uncomfortable to carry, that they have a short battery life, and that the occupants may forget to put them on. The third type of sensors are built into the environment and are sensors for movement, pressure, temperature, humidity, etc. They usually offer more limited information and a larger number of sensors are needed to recognize activities, but their advantages are that they are unobtrusive – they do not intrude on people's privacy and do not need to be worn. These sensors can provide contextual information, through the interaction of the occupant with the objects in the home, to help distinguish between similar activities [3].

Activity recognition has a number of applications in smart homes. One of them is related to the more efficient use of resources in the smart home (such as electricity, heating, lighting), by learning the behavior of the inhabitants. Another application is in helping the elderly to live independently in the comfort of their own home. Such systems can detect deviations from normal behavior in everyday life.
in a timely manner and inform the relevant persons. This is a particularly important problem, as the proportion of older people is expected to increase significantly over the next few decades [5], which could put a strain on the healthcare system.

In traditional methods, the raw signal from the sensors is processed before it is used for classification. This is usually done by segmenting it with a sliding window, after which certain features are derived from each segment, most often they are statistical features such as mean, variation, entropy and correlation coefficients [6]. Not all features are equally good for recognizing certain activities [6], i.e. some features are more appropriate in recognizing certain activities than others. For this reason, some authors use methods to find the most appropriate set of features for the activities that are to be classified [1], [7].

Recently the interest in deep neural networks has been growing. Deep neural networks are neural networks with more than one hidden layer. Unlike ordinary neural networks, deep neural networks can produce significantly more complex transformations of the input data, i.e. such networks can recognize more complex patterns in the data. A major disadvantage of deep neural networks compared to conventional ones is that training takes significantly longer [8].

One very popular deep network is the Convolutional Neural Network (CNN). The classical methods rely on hand-crafted features. The problem is that this process is time consuming and that it’s not always possible to select appropriate features. CNNs allow automatic extraction of relevant features by convolving the input signal with a set of filters. The result of which passes through a pooling layer, which reduces the size of the data and introduces a translational invariance. CNNs can consist of several consecutive convolutional and pooling layers, which leads to a hierarchical extraction of features – with each layer the data is presented in a more abstract way. Typically, after these layers, the convolution networks contain a fully connected multilayer perceptron, which performs the classification according to the extracted features. The ability of CNNs to take advantage of local dependencies in the data, as well as their translational invariance, makes them particularly suitable for recognizing activities [9]. One disadvantage of convolutional networks is that they require a large data set of labeled data [8].

This paper presents a 1D CNN for human activity recognition using accelerometric data. The rest of this paper is organized as follows: Section 2 presents related work. The proposed method and the data set used for the evaluation are described in Section 3. The experimental results are presented in Section 4 and the conclusion – in Section 5.

2. Related work
CNNs have been used to classify human activities using accelerometric data. In [6] the authors use a convolutional neural network with partial weight sharing to recognize activities from three public data sets. The system shows better results than other popular feature extraction methods. A convolutional neural network to recognize activities from multichannel time series data from body-worn inertial sensors is presented in [10]. The system is compared with SVM, KNN, means and variance with KNN and Deep Belief Network (DBN), showing better results. The authors also conclude that the CNN's performance is fast enough for online recognition. In [4] a convolutional neural network consisting of 3 convolutional layers and 3 pooling layers is suggested for activities recognition. To test the system the authors collect accelerometric data from 100 people performing 8 activities, such as falling, running, walking. The suggested CNN has an average accuracy of 93.8% and shows better results than a SVM classifier and an 8-layer DBN. In [11] the authors use a CNN to identify 20 activities related to the making of a cup of tea. The data used are collected from four people wearing an accelerometer on their wrist. The system is reported to show better results than methods based on k-means clustering, linear discriminant analysis and SVM. A convolutional networks for classification of multichannel time series is presented in [12]. The activities that are recognized are standing, walking, ascending stairs and descending stairs. Each activity is a three-channel time series. Features are extracted for each channel separately using the convolutional neural network. Then the extracted features from all channels are combined and fed into a multilayer perceptron for classification. The results show that the system performs better than other state-of-the-art methods. A convolutional network is described in [13] for extraction of features that are used by a SVM to recognize activities such as walking, ascending stairs, descending stairs, etc. To validate the system, the authors utilize inertial sensor data. The 1D signal from
the sensor is transformed into a 2D spectrogram and fed to the network. The authors use a pre-trained AlexNet convolutional network without additional training and the system is reported to show good recognition results.

3. Methods and materials
The CNN used for HAR in this paper consists of four 1D convolutional layers, each of which applies 16 kernels with size 2, and ReLu is used for the activation function. Each of these convolutional layers was followed by a max-pooling layer with pooling size 2. After the convolutional and max-pooling layers, a fully connected layer of 100 neurons with ReLu activation function was applied. To avoid overfitting, dropout regularization with $p = 0.5$ was applied to this layer. The last layer is the softmax layer. The hyperparameters were determined experimentally using 10-fold cross validation and the following procedure. The possible values that each of the hyperparameters can have are determined in advance. First, the number of layers that shows the best average $F_1$ score is selected, then for this number of layers the optimal number of kernels is selected (each layer will have the same number of kernels), and finally for the corresponding number of layers and number of kernels, the optimal kernel size is selected.

The neural network was implemented using the Keras neural network modeling library. The network parameters were optimized, by minimizing the cross-entropy loss function, using the Adam algorithm, with a maximum number of training epochs of 100. Training and classification were performed using NVIDIA GeForce RTX 2060.

The publicly available WISDM Actitracker data set [14] was used for the experimental evaluation of the suggested model. The set contains data collected from 36 people under controlled conditions. The data are recorded by a triaxial accelerometer with a sampling frequency of 20 Hz, and are labeled for 6 activities – walking, jogging, ascending stairs, descending stairs, sitting and standing. The sensor data was segmented using a 10s sliding window with 50% overlap and the accelerometer’s values were normalized using z-score normalization. 10-fold cross validation was used in the evaluation of the suggested 1D CNN model. The performance was evaluated using the following metrics: precision, recall, specificity, $F_1$ score and accuracy.

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$\text{specificity} = \frac{TN}{TN + FP}$$

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

where TP are the true positives, FP are the false positives, TN are the true negatives, and FN are the false negatives.

As the data set is imbalanced (about 70% of the data belong to two classes – walking and running), accuracy is not an appropriate performance measure, but is included in the calculations in order to allow a comparison with other results from the literature. As suggested by [15], in order to avoid biased measurements in an imbalanced data set, the metrics were calculated from the total TP, FP, TN, FN of all 10 cross-validation runs.
4. Results
Table 1 shows the results of the 10-fold cross-validation. It can be seen that Walking, Jogging, Sitting and Standing are very well recognized with precision, recall, specificity and F1 score with values close to 1. On the other hand, the model doesn’t recognize Downstairs and Upstairs that well with F1 scores below 0.9. The mean F1 score is 0.9514 ± 0.0456.

Table 1: Results of the 10-fold cross-validation.

| Activity   | Precision | Recall  | Specificity | F1 score | Accuracy (%) |
|------------|-----------|---------|-------------|----------|--------------|
| Downstairs | 0.8970    | 0.8818  | 0.9898      | 0.8893   | 97.99        |
| Jogging    | 0.9877    | 0.9863  | 0.9944      | 0.9870   | 99.19        |
| Sitting    | 0.9882    | 0.9718  | 0.9993      | 0.9799   | 99.78        |
| Standing   | 0.9754    | 0.9855  | 0.9989      | 0.9804   | 99.83        |
| Upstairs   | 0.8755    | 0.8941  | 0.9840      | 0.8847   | 97.40        |
| Walking    | 0.9870    | 0.9873  | 0.9918      | 0.9872   | 99.01        |
| average    | 0.9518    | 0.9511  | 0.9930      | 0.9514   | 98.86        |
| std        | 0.0470    | 0.0451  | 0.0053      | 0.0456   | 0.90         |

Figure 1 shows the total confusion matrix of all 10 runs of the 10-fold cross-validation. Downstairs and Upstairs are the worst classified activities, and often one activity is misclassified as the other, which is most likely due to the fact that the two activities are quite similar. For the other activities, the misclassifications are significantly fewer.

Next, the effects of varying the hyperparameters on the performance of the system are evaluated.
Figure 2. Influence of the number of convolutional layers on the $F_1$ score.

Figure 2 shows that decreasing the number of convolutional layers leads to a deterioration in the performance of the system, with 1 convolution layer having a significantly lower average $F_1$ score.

Figure 3. Influence of the number of kernels on the $F_1$ score.

Increasing the number of kernels (Figure 3) leads to an improvement in performance, being most noticeable between 16 and 24, and almost insignificant between 32 and 64.
Figure 4. Influence of the kernel size on the F1 score.

It can be seen from Figure 4 that increasing the kernel size from 2 to 5 leads to an improvement in performance, but increasing the size above 5 does not lead to further improvement.

Table 2 shows a comparison of the proposed model with results of other deep models described in the literature that use the same data set. It can be seen that the proposed model shows better accuracy than the other models using this data set.

Table 2. Comparison of the proposed model against other HAR approaches.

| Neural Network Model                                    | Reference                  | Recognition accuracy (%) |
|----------------------------------------------------------|----------------------------|--------------------------|
| Deep Belief Network (DBN)                               | Alsheikh et.al. [16]       | 98.23                    |
| CNN with partial weight sharing                         | Zeng et. al. [6]           | 96.88                    |
| CNN with sums of temporal convolutions                  | Ravi et. al. [17]          | 98.20                    |
| 1D CNN                                                   | The suggested model         | 98.86                    |

5. Conclusion
A human activity recognition model based on 1D CNN and accelerometric data is suggested and experimentally evaluated. A semi-automatic approach is used that effectively and efficiently determines the number of convolutional layers in the network, the number of kernels and the size of the kernels. The experimental results show that the suggested model achieves higher recognition accuracy than existing approaches using the same data set. Future work will be aimed at utilization of the suggested HAR system for various smart home applications.

Acknowledgments
This work is supported by the European Regional Development Fund and the Operational Program "Science and Education for Smart Growth" under contract UNITe № BG05M2OP001-1.001-0004 (2018-2023).
References

[1] Fang H, He L, Si H, Liu P and Xie X 2014 Human activity recognition based on feature selection in smart home using back-propagation algorithm ISA transactions 53 (5) 1629-1638

[2] Nweke H F, Teh Y W, Al-Garadi M A and Alo U R 2018 Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges Expert Systems with Applications 105 pp 233–261

[3] Guan D, Ma T, Yuan W, Lee Y K and Jehad Sarkar A M 2011 Review of sensor-based activity recognition systems IETE Technical Review 28 (5) pp 418-433

[4] Chen Y and Xue Y 2015 A deep learning approach to human activity recognition based on single accelerometer 2015 IEEE international conference on systems, man, and cybernetics pp 1488-1492

[5] Liu Z, Song Y, Shang Y and Wang J 2015 Posture recognition algorithm for the elderly based on BP neural networks The 27th Chinese Control and Decision Conference (2015 CCDC) pp 1446-1449

[6] Zeng M, Nguyen L T, Yu B, Mengshoel O J, Zhu J, Wu P and Zhang J 2014 Convolutional neural networks for human activity recognition using mobile sensors Proc. of 6th IEEE International Conference on Mobile Computing, Applications and Services pp 197–205

[7] Oukrich N, Maach A, Sabri E, Mabrouk E and Bouchard K 2016 Activity recognition using back-propagation algorithm and minimum redundancy feature selection method 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt) pp 818-823

[8] Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B and Yang G Z 2016 Deep learning for health informatics IEEE journal of biomedical and health informatics 21 (1) pp 4-21

[9] Ronao C A and Cho S B 2015 Deep convolutional neural networks for human activity recognition with smartphone sensors Neural Information Processing ICONIP 2015 Lecture Notes in Computer Science 9492 (Springer, Cham) pp 46–53

[10] Yang J, Nguyen M N, San P P, Li X L and Krishnaswamy S 2015 Deep convolutional neural networks on multichannel time series for human activity recognition Twenty-Fourth International Joint Conference on Artificial Intelligence

[11] Panwar M, Dyuthi S R, Prakash K C, Biswas D, Acharya A, Maharatna, K, Gautam A and Naik G R 2017 CNN based approach for activity recognition using a wrist-worn accelerometer 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) pp 2438-2441

[12] Zheng Y, Liu Q, Chen E, Ge Y and Zhao J L 2014 Time series classification using multi-channels deep convolutional neural networks International Conference on Web-Age Information Management (Springer, Cham) pp 298-310

[13] Wagner D, Kalischewski K, Velten J and Kummert A 2017 Activity recognition using inertial sensors and a 2-D convolutional neural network 2017 10th International Workshop on Multidimensional (nD) Systems (nDS) pp 1-6

[14] Kwapisz J R, Weiss G M and Moore S A 2011 Activity recognition using cell phone accelerometers ACM SigKDD Explorations Newsletter 12 (2) pp 74–82

[15] Forman G and Scholz M 2010 Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement Acm Sigkdd Explorations Newsletter 12 (1) pp 49-57

[16] Alsheikh M A, Selim A, Niyato D, Doyle L, Lin S and Tan H P 2016 Deep activity recognition models with triaxial accelerometers Proc. of 30th AAAI Conference on Artificial Intelligence

[17] Ravi D, Wong C, Lo B and Yang G Z 2016 Deep learning for human activity recognition: A resource efficient implementation on low-power devices 2016 Proc. of 13th IEEE International Conference on Wearable and Implantable Body Sensor Networks (BSN) pp 71–76