Optimal Recognition Method of Human Activities Using Artificial Neural Networks

Stefan Oniga[1,2], Sütő József[2]

[1]Department of Electrical, Electronic and Computer Engineering, Faculty of Engineering, Technical University of Cluj-Napoca, North University Center of Baia Mare, Dr. V. Babes Street, No. 62A, ZIP 430083, Baia Mare, Romania, stefan.oniga@cunbm.utcluj.ro
[2]Department of Informatics Systems and Networks, Faculty of Informatics, University of Debrecen, Egyetem ter, 1., 4032, Debrecen, Hungary, suto.jozsef@inf.unideb.hu, oniga.istvan@inf.unideb.hu

The aim of this research is an exhaustive analysis of the various factors that may influence the recognition rate of the human activity using wearable sensors data. We made a total of 1674 simulations on a publically released human activity database by a group of researcher from the University of California at Berkeley. In a previous research, we analyzed the influence of the number of sensors and their placement. In the present research we have examined the influence of the number of sensor nodes, the type of sensor node, preprocessing algorithms, type of classifier and its parameters. The final purpose is to find the optimal setup for best recognition rates with lowest hardware and software costs.

Keywords: Wearable sensor network, artificial neural networks, human activity recognition, pattern recognition.

1. Introduction

In the last two years there has been a spectacular growth in terms of research on human activity recognition. This trend can be explained by the importance of recognition of daily activities to assist the elderly or the sick persons, prevention of diseases, and alert in case of danger. Extraction of behavior patterns and detection of deviations from these patterns can provide early and useful information about health degradation.

The aim of this research is an exhaustive analysis of the different factors that may influence the recognition rate of the human activity using wearable sensors data. In a previous research, we analyzed the influence of the number of sensors and their placement. In the present research we have examined the influence of the number of sensor nodes, the type of sensor node, preprocessing algorithms, type of classifier and its parameters.

The final purpose is to find an optimal setup for best recognition rates with the lowest hardware and software costs. The hardware costs can be quantified by the number of required sensor nodes and the number of neurons of the neural network. The neural network classifier will be hardware implemented using reconfigurable logic. Lower software costs means less necessary preprocessing needs.

2. Ward database

For our research we have used the WARD (wearable action recognition database) database provided by NSF TRUST (Team for Research in Ubiquitous Secure Technology) Center at the University of California, Berkeley, University of Texas at Dallas, Tampere University of Technology, and Telecom Italia Laboratory [1], [2].

A. Database description

The database contains data supplied by wearable motion sensors network while performing a set of predetermined activities. The wireless sensor network is composed of sensor nodes equipped with one triaxial accelerometer and a biaxial gyroscope. The sensors are placed on both ankles, on the waist and on both forearms. The database contains 13 activities recorded from 20 persons, performed five times. These activities, as described in [1] and [2], are:
1. Standing: standing still more than 10 seconds.
2. Sitting: siting still more than 10 seconds.
3. Lying: lying still more than 10 seconds.
4. Walk forward: walking straight forward more than 10 s.
5. Walk forward left-circle: walking in counter-clockwise circle more than 10 seconds.
6. Walk forward right-circle: walking in clockwise circle more than 10 seconds.
7. Turn left: staying at the same position and turning left more than 10 seconds.
8. Turn right: staying at the same position and turning right more than 10 seconds.
9. Go upstairs: going up one flight.
10. Go downstairs: going down one flight.
11. Jog: jogging straight forward more than 10 seconds.
12. Jump: staying at the same position and jumping 5 times.
13. Push wheelchair: pushing a wheelchair more than 10 s.

The sampling rate was 20 samples/second. The intention of the authors was to create a database placed in the public domain with the aim to compare the results obtained in activity recognition using different recognition methods.

The structure of the data recorded in this database is presented below. The data belonging to a subject are recorded in individual directories. The directory contains five files for each activity, each file corresponds to a recording session. For each activity there were five recording sessions for each subject. Equation 1 shows a record structure from one sensor node (k) at time t having 5 components, three from the accelerometer (x, y, z) and two from the gyroscope (θ, ρ)

$$s_k(t) := (x_k(t), y_k(t), z_k(t), \theta_k(t), \rho_k(t)) \in \Re^5 \qquad (1)$$

Next equation represents the data vector of all L sensors at time t.

$$s(t) := (s_1(t), s_2(t), \dots s_L(t))^T \in \Re^{5xL} \qquad (2)$$

In equation 3, s_a collects all data corresponding to an activity of l samples.

$$s_a = (s(1), s(2), \dots s(l))^T \in \Re^{5xLxl} \qquad (3)$$

with a=1:13, and l being different for each activity.

B. Methods used for activity recognition

The authors of the WARD database have published their results obtained using nearest-neighbor (NN) and distributed sparsity classifier (DSC) algorithms [1].

We can find in the literature a huge variety of activity recognition methods. A survey of Human Activity Recognition using Wearable Sensors is presented in [3]. Generally, the methods used for activity recognition have two phases. First phase is the feature extraction and the second is the classification phase. For feature extraction, the most used methods may be on one hand statistical methods (max, media, variance, standard deviation, energy), called also time-domain features, and on the other hand frequency-domain features using FFT or time-frequency features using wavelet filter banks. A comparison between these methods is presented in [4]-[6].

Regarding the classification, in literature are presented many methods which include: data mining algorithms such as Naïve Bayes classifier [7], Bayesian Networks Classifiers [18]. Other classification methods encountered are Support Vector Machines (SVM), Decision Trees classifiers, k-nearest neighbor (kNN) classifier [4], hidden Markov models, and neural networks [6].

A comparison between classification methods is presented in [9] and [10].

3. THE METHOD PROPOSED BY US

A. Extracting data from the database and their reordering

For an easier access to these data we created a Matlab script which extracts data corresponding to all activities performed by a subject in a recording session (in total of 13), orders the data corresponding to the five sensor nodes, with five series of data each (acceleration 3 axes, gyroscope 2 axes) in 13 variables of l_a x 25 columns, where l_a is the number of samples corresponding to activity a, a=1:13. In this way we obtained 13 vectors as in (3). Finally, we created a 5996 x 25 matrix representing acquired data from a subject during a recording session.

The database contains a total of 20 x 5 matrices (number of subjects x number of recording sessions for each subject) available for simulations. During simulation, from these matrices, only the appropriate columns and sensor nodes that need to be considered can be automatically extracted. In this way all possible scenarios regarding the number of

sensors and all their possible combinations of x=1, 2, 3, 4 or 5 sensors can be simulated. For n=5 sensor nodes the total number of possible combinations is:

$$N = \sum_{x=1}^{5} C(n, x) = 31 \qquad (4)$$

B. Feature extraction

In our previous experiments presented in [11] we decided to use the sum of accelerations on the 3 axes and standard deviation of this sum as supplementary input data for the recognition system, because in this way we obtained better recognition rates. Standard deviation was calculated on non-overlapping windows of one second.

In the present work, these vectors could or could not be added to the data matrix, as new columns, according to the desired experimental configuration. Since the created experimental environment allows us to run all the scenarios, we simulated all configurations with and without sum of the accelerations on the 3 axes, respectively, with or without the standard deviation of this sum.

C. Preparing data for training and simulation

Next preparatory stage is the creation of the data for training the neural network and of the data for testing the network. These consist of two sets of data: input data and target data. The training input data consist of samples corresponding to the acceleration on the 3 axes from 1 up to 5 sensors to which one can add the data from the gyroscope, sum of the accelerations, and the standard deviation of the sum. In this way the simplest input data matrix has 5996 samples x 3 data, while the most complex contains 5996 samples x 35 data. Target data has a fixed size: 5996 x 13, each row contains a single 1 in the column representing the output corresponding to the activity that must be recognized.

The rows of the two matrices containing input and target data are mixed synchronously. Input data used for testing are in the same order as they were acquired to allow easier visual observation of simulation results.

D. Creation of the network, training and simulation

The artificial neural network, we used for activity recognition was a feed forward back-propagation network with one hidden layer. The network is created automatically and simulated in Matlab with settable parameters. The number of inputs depends on the number of input data, the number is between 3 and 35, the number of outputs is 13 because this is the number of activities to be recognized. The number of neurons on the hidden layer is a parameter that can have three values 10, 20 or 30 for the networks with fewer inputs (simulations with 1:3 sensorial nodes), up to 50, 55, 60 in the case of networks for 5 sensorial nodes. We have chosen sigmoid as the activation function for the hidden layer and linear function for the output layer. Training is done using LM algorithm.

The trained network is simulated using three different output functions: round, competitive, and a threshold function with the threshold 0.5.

4. SIMULATION RESULTS

The purpose of the simulations was to observe the influence of various factors on the recognition rate and to identify the solutions that require fewer hardware or software resources. A total of 186 configurations were created and three different ANNs were trained for each configuration. For each possible combination of sensor nodes (a total of 31 combinations) we have imagined six configurations. For each configuration we created and trained three neural networks. Each network was simulated with three output functions giving a total of 1674 simulations results.

As was expected, best recognition rates can be obtained using all the 5 sensors and taking into account also the gyroscope data. Adding in this case the sumAcc data and std (sumAcc) does not rise significantly the recognition rates (only from 98.38 % to 99.8 %) and in this case, it is an disadvantageous solution because of its cost of hardware and additional software for the computation. Using preprocessed data together with raw data increases the recognition rates for the configurations with 3 and especially for those with 2 sensor nodes. In this case the calculation of extracted features is preferable, because the use of fewer sensor nodes provides an increased comfort to wear and the system is more reliable.

We present below some of the results obtained for configurations with two sensor nodes. It can be observed that for configurations with two sensor nodes we can obtain recognition rates over 95 %, e.g., for the combinations 13 and 35 if we use not only the acceleration data but also the gyroscope data. Also we must use the preprocessed data, i.e. the sum of the accelerations on three axes and its standard deviation, as additional input data for the neural networks.

A. Simulations based on the number of sensor nodes

Recognition rate depends on the number of used sensors, generally increasing with the number of sensors used. For some combinations of the sensor nodes together with data preprocessing, we can obtain recognition rates over 90 %, even using only 2 sensor nodes. Table 2. shows the recognition rates as a function of the number of sensors without preprocessed data and in two cases when we add two preprocessed features, i.e. Sum(Acc) and std(Sum), respectively.

B. Simulations based on sensor nodes composition

Fig.1. shows the dependence between recognition rates and type of sensors used in nodes. The blue line represents the results using only the acceleration sensors and the red line when we use both sensors. It is obvious that using both sensors we obtain better recognition rates. The results also depend on the location of the sensors. In this case the sensors on the lower part of the body were placed on the ankles and in this way they do not supply any information about upper-body movement.

In case of placement on a more suitable position, as we presented in our previous work, on the thigh just above the knee, the sensor nodes can provide also some information

about upper–body movement and in this way we obtained better results even without using a gyroscope.
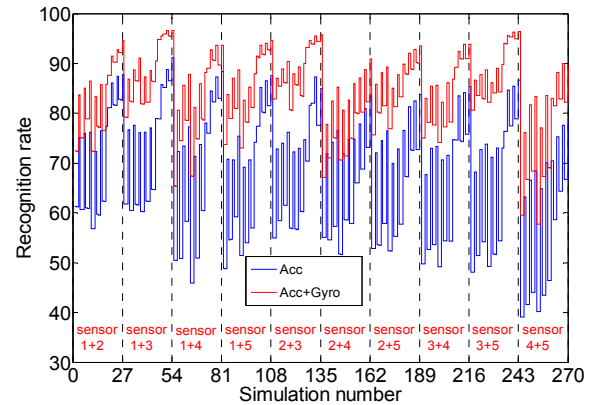

Fig.1. Comparison of recognition rates using acceleration vs. acceleration and gyroscope.

C. The effect of adding as input signal of the sum of accelerations on 3 axes

Using sum of the accelerations on the 3 axes, as a supplementary input for the neural network does not increase the recognition rates, as it can be observed in Table 1., where row 1 shows the results obtained without the sum of accelerations and row 2 the data with sum of the accelerations.

D. The effect of adding standard deviation of the sum of accelerations as input data

Adding standard deviation of sum(Acc) along with the raw data to the input of the ANN classifier significantly increases the recognition rate for the configurations with two sensor nodes. In this case the growth can be 15-20 %, as it can be seen in row 3 of Table 1. The growth is smaller for the configurations with a larger number of sensors.

E. Simulations depending on the number of neurons on the hidden layer

Increasing the number of neurons on the hidden layer generally results in increasing the rate of recognition, but the increase is not significant. Results are presented in Fig.2.
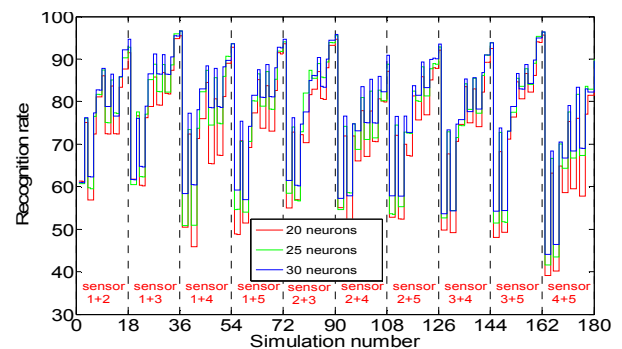

Fig.2. The effect of number of neurons on hidden layer on recognition rate.

Table 1. Overall recognition as a function of number of sensors and their configuration.

Preprocessing	Sensors configuration																														
	1	2	3	4	5	12	13	14	15	23	24	25	34	35	45	123	124	125	134	135	145	234	235	245	345	1234	1235	1245	1345	2345	12345
------	65.96	68.96	72.23	49.23	52.5	78.87	86.51	78.62	81.02	86.12	75.12	81.55	77.7	82.97	68.38	89.98	87.32	87.14	88.99	90.33	88.46	88.54	91.41	85.91	88.26	94.6	96.48	95.33	96.38	96.83	98.07
Sum	69.7	70.33	74.05	48.83	53.8	76.55	86.41	78.82	81.2	83.54	75.02	83.32	78.27	84.32	69.06	88.76	87.86	87.89	88.24	90.04	87.06	86.39	92.53	89.19	89.63	94.78	96.3	95.18	96.33	95.7	98.38
Sum+Std	83.64	74.38	87.88	60.41	60.64	92.21	95.51	89.71	92.73	94.43	82.35	90.23	90.96	94.95	82.22	97.93	95.38	96.25	97.33	97.45	95.55	95.66	98.17	93.58	97.08	98.82	99.3	98.73	98.88	99.07	99.8

Table 2. Comparison between results presented in [1] and results obtained by us.

Recognition method/ sensors configuration/ number of neurons	Recognition rate/ Activity (%)													
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	All
DSC	87.2	66.80	91.8	92.00	97.30	95.70	97.00	95.20	98.00	98.30	99.30	97.90	98.60	92.44
ANN-12345-50	100	100	100	91.88	97.63	95.64	99.10	98.70	95.67	95.02	97.85	97.31	100	97.77
ANN-12345-55	100	100	100	92.13	97.63	96.47	99.10	99.13	92.62	93.44	97.04	98.34	100	97.60
ANN-12345-60	100	100	100	94.92	98.49	97.10	99.10	99.13	93.13	94.57	98.39	97.93	100	98.07
ANN-13-25	100	100	100	82.23	92.90	92.32	99.55	97.18	88.80	95.93	96.24	98.55	98.25	95.93
ANN-35-25	98.9	99.35	100	89.85	93.98	85.68	99.10	98.48	87.28	87.78	98.12	99.38	98.25	95.36

5. CONCLUSION

The simulation results show that from a total of 1674 simulations performed if we consider only the cases when the recognition rates are over 90 %, still remain 467 cases. We are trying to identify those setups that provide better results than those presented in [1] using only 2 sensor nodes instead of 5, but adding sumAcc and std(sumAcc) as new preprocessed features to the input of the neural network.

Searching the trade-off between best recognition rates and less hardware resources needed and/or less preprocessing, we identified two setups with two sensor nodes, namely the configurations with the sensor pairs 13 and 35.

The method for designing such a multi-sensor system required consideration of the following: sensor selection, the placement of the sensors, selection of preprocessing algorithms, and classifier selection. The impact of each factor has been investigated and we selected the most appropriate approach in order to achieve a trade-off between recognition accuracy and fewer hardware resources and/or preprocessing requirements. The experimental results illustrate that the proposed multi-sensor system can achieve overall recognition accuracy over 97 % without any preprocessing using 5 sensors like those in [1]. We obtained a recognition rate over 95 % even for some of the setups with two sensors, adopting the sum of acceleration components and the standard deviation of the sum as preprocessed new features, when using an ANN classifier.

Table 2. presents comparison between the results presented in [1] using nearest-neighbor (NN) and distributed sparsity classifier (DSC) algorithms, for the configuration with 5 sensors (first row) and our results using an FFBP ANN classifier with one layer of 50, 55 and 60 neurons (rows 2-4). Columns A1-A13 present the results for the activities defined in chapter 2. Also we present in rows 5 and 6 results obtained for a 2 sensor configuration (1 and 3, respectively 3 and 5), using an ANN with 25 neurons and with the new features added as inputs. Even in this case the recognition rates are over 95 %.

Even better results can be obtained using a better placement of the sensors. With 2 sensors placed on the thigh and wrist and using only the accelerometer data and preprocessed data we obtained very good results as we presented in [12].

REFERENCES

[1] Yang, A.Y., Jafari, R., Sastry, S.S., Bajcsy, R. (2009). Distributed recognition of human actions using wearable motion sensor networks. *Journal of Ambient Intelligence and Smart Environments*, 1 (2), 103-115.

[2] Yang, A., Kuryloski, P., Bajcsy, R. (2009). WARD: A wearable action recognition database. In *27th Annual CHI Conference*, 4-9 April 2009, Boston, MA.

[3] Lara, O.D., Labrador, M.A. (2013). A survey on human activity recognition using wearable sensors. *Communications Surveys & Tutorials*, 15 (3), 1192-1209.

[4] Preece, S.J., Goulermas, J.Y., Kenney, L.P., Howard, D. (2009). A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering*, 56 (3), 871-879.

[5] Maurer, U., Smailagic, A., Siewiorek, D.P., Deisher, M. (2006). Activity recognition and monitoring using multiple sensors on different body positions. In *International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006)*, 3-5 April 2006, Cambridge, MA. IEEE, 113-119.

[6] Yang, J.Y., Wang, J.S., Chen, Y.P. (2008). Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers. *Pattern Recognition Letters*, 29 (16), 2213-2220.

[7] Gao, L., Bourke, A.K., Nelson, J. (2011). A system for activity recognition using multi-sensor fusion. In *Annual International Conference of the IEEE - Engineering in Medicine and Biology Society (EMBC 2011)*, Boston, MA. IEEE, 7869-7872.

[8]   Kouris, I., Koutsouris, D. (2013). Application of data mining techniques to efficiently monitor chronic diseases using wireless body area networks and smartphones. *Universal Journal of Biomedical Engineering*, 1 (2), 23-31.

[9]   Kouris, I., Koutsouris, D. (2011). A comparative study of pattern recognition classifiers to predict physical activities using smartphones and wearable body sensors. *Technology and Health Care*, 20 (4), 263-275.

[10] Gao, L., Bourke, A.K., Nelson, J. (2014). Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Medical Engineering & Physics*, 36 (6), 779-785.

[11] Oniga, S., Suto, J. (2014). Human activity recognition using neural networks. In *15th International Carpathian Control Conference (ICCC 2014)*, 28-30 May 2014, Czech Republic. IEEE, 403-406.

[12] Orha, I., Oniga, S. (2014). Study regarding the optimal sensors placement on the body for human activity recognition. In *IEEE 20th International Symposium for Design and Technology in Electronic Packaging (SIITME 2014)*, 23-26 October 2014. IEEE, 203-206.