

Recognition Physical Activities with Optimal Number of Wearable Sensors using Data Mining Algorithms and Deep Belief Network

Ali H. Al-Fatlawi
Information Technology Research and
Development Center,
University of Kufa, Najaf, Iraq
Ali.H.Mahdi@uokufa.edu.iq

Hayder K. Fatlawi
Information Technology Research and
Development Center,
University of Kufa, Najaf, Iraq
hayder.fatlawi@uokufa.edu.iq

Sai Ho Ling, *SMIEEE*
Centre for Health Technologies,
University of Technology Sydney,
Sydney, Australia
Steve.Ling@uts.edu.au

Abstract— Daily physical activities monitoring is benefiting the health care field in several ways, in particular with the development of the wearable sensors. This paper adopts effective ways to calculate the optimal number of the necessary sensors and to build a reliable and a high accuracy monitoring system. Three data mining algorithms, namely Decision Tree, Random Forest and PART Algorithm, have been applied for the sensors selection process. Furthermore, the deep belief network (DBN) has been investigated to recognise 33 physical activities effectively. The results indicated that the proposed method is reliable with an overall accuracy of 96.52% and the number of sensors is minimised from nine to six sensors.

Keywords—Monitoring Physical activities, Deep Belief Network, Data mining, Features selection, Decision Tree, PART algorithm, Random Forest.

I. INTRODUCTION

Recently, several research studies highlighted the significance of monitoring human physical activities on improving the healthcare and treatment processes. Some diseases require continuous monitoring for the patient's activities to measure their reflections on the health status. For example, observing physical activities of the patients with the obstructive pulmonary disease has a benefit of knowing their lifestyle continuously and measuring the impact of each activity on their health [1]. Moreover, studies show that there is a significant relationship between the percentage of mortality and daily exercises [2],[3]. The death rate is lower in these people who are more active in their leisure time [4]. Dementia is another example of the diseases that are affected by the daily physical activities [5]. Thus, measuring these activities is a step in the treatment procedure [6].

Monitoring daily exercises of people and notifying them regarding their activities can save their life and improving their health care. The technology of wearable devices makes this idea more applicable and flexible. Apart from its benefits in health applications, recognition of the human physical activities by the wearable sensors can serve various kinds of sectors. It can support types of extreme sports and entertainment activities such as the Taekwondo [7], "Dance-Dance-Revolution" game [7] and gym exercises [8]. The human-computer-interface (HCI) is another application for the technology of the wearable sensors. For example, an independent Gestor controller for home appliances can be built based on that concept to make the control process easier and flexible [9]. Multiple wearable sensors are needed to detect the body's physical motions. These sensors should be carefully positions in order to get meaningful signals.

In this paper, data mining algorithms are applied to determine the impact of feature of sensors on the classification process. By examining this effect on identifying the activities, the number of used sensors can be reduced and classification performance is retained. Three data mining algorithms: Decision Tree, Random Forest and PART Algorithm, are applied to assess the significance of each input on the decision of the classifier. Furthermore, a deep belief network (DBN) is developed to classify each activity based on the reduced number of features. The DBN has been used because of its capability in classifying many classes (33 activities in the proposed system) with higher efficiency. The deep generative model of the DBN allows to process nonlinear dimensionality in two processing levels.

II. METHODOLOGY

In this paper, the main purpose is hybridising the data mining algorithms and DBN to monitor 33 daily human physical activities by using an optimised number of wearable sensors. Figure1 demonstrates the flow of the signals through the processing stages that are required to build an efficient system able to identify the activities with high accuracy rate.

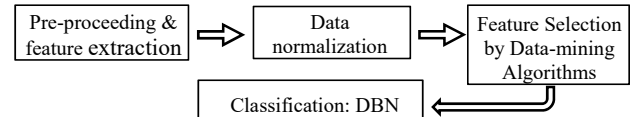


Figure 1. Processing Stages of the System

A. Data description and normalisation

The first process is to extract and present the features of the collected raw data. Then data normalisation is processed to reduce data redundancy and improve data integrity. For this normalisation operation, "mean zero standard derivation one" method has been applied and given below:

$$S_i = \frac{x_i - \bar{x}_i}{\sigma}, \quad i = 1, 2, \dots, n \quad (1)$$

where

$$\bar{x}_i = \frac{\sum_{j=1}^{i-1} x_j}{n}$$

$$\sigma = \sqrt{\frac{\sum_{j=1}^{i-1} x_j - \bar{x}_i}{n-1}}$$

x is the original inputs and S is the normalised input vectors respectively, and n is the number of the elements in the input vector.

B. Feature Selection process

For the feature selection task, three data-mining algorithms (Decision Tree, Random Forest and PART

Algorithm) are applied. The sensors generate signals in responding to each activity, and these signals will be transmitted to the processor which may be smartphones, tablets or computers.

Selection of an optimal subset of features is an elementary and efficient task which precedes classification. It aims to reduce the dimensionality (features) of data and then the required resources to perform the classification will be reduced.

I. Decision Tree:

A decision tree is a simple and powerful form of data analysis which allows us to predict, explain, describe, or classify a target [11]. A decision tree consists of two types of nodes; non-terminal nodes and terminal nodes. A non-terminal node denotes a condition based on a feature. This condition splits data records based on their characteristics. A terminal node, also it is called a leaf node, doesn't have a condition because it is used to make a decision. Each branch of the decision tree represents the result of the conditions, while the leaf node holds a class label. According to the number of branches, the Decision tree can be either a binary or multi-branch tree. In the binary decision tree; each non-terminal node branches to exactly only two other nodes [12]. However, in the multi-branch tree, more than two child nodes may be generated through the split operation.

The most critical issue during the building of the decision tree is evaluating of feature importance. In each non-terminal node, the most relevant feature with the target must be selected. Then, from all values of that feature, the best split point should be chosen to divide instances of data in that node. For this task, statistical measures could be used such as Information Gain and Gain Ratio. The feature which produces a maximum reduction in impurity represents the best feature. Impurity could be measured using statistical randomness measurement such as Entropy. Therefore, Information Gain is the gain of splitting operation indicates by the impurity of the class before and after splitting. Information gain [13] is governed by eq. (2),

$$IG = IG(P) - \sum_{i=1}^c \frac{NC(i)}{NC} IG(i) \quad (2)$$

where $IG(P)$ is the information gain of parent node before splitting, c is the number of feature's values, NC is the number of data instances in parent node, $NC(i)$ is the number of data instances in node i , $IG(i)$ is the information gain of node i . Entropy measure is utilised by Information Gain to measure the impurity. It tends to select features with more distinct values. A large number of values led to creating more child nodes in each iteration. Thereby; the number of instances would be decreased as the prediction reliability. To overcome this problem, Gain Ratio can be used. It weighted the information gain of each test condition by its number of child nodes as shown in the eq. (3) [12],

$$\text{Gain Ratio} = \frac{IG}{-\sum_{i=1}^c P(i) \log_2 P(i)} \quad (3)$$

where $P(i)$ is the fraction of data instances in the node i to the number of data instances in the parent node.

II. PART Rule Algorithm

PART is a rule based data mining technique which builds the classification model in the form of "if - then" statements connected by conjunctions [14]. There are two methods for building PART rules model as follow (i) the direct method which uses data directly to generate rules like RIPPER and OneR algorithms; (ii) the indirect method which generates rules depending on the output of other classification techniques such as C4.5 rules. PART merges the two methods by using a partial decision tree [13].

Evaluating the quality of the rules could be performed by the two criteria: coverage and accuracy. The coverage represents the fraction of instances that matches conditions of the rule to the number of data instances. The accuracy of the rule could be calculated by dividing the number of data instances that match both conditions and class of the rule, on the instances that match conditions of the rule [12]. From the current data, PART creates a partial decision tree that would be used to generate the rule. PART removes data instances that are covered by the rule and the partial decision tree. The whole procedure is repeated for each rule until stopping condition met [15] [16]. A rule in PART consists of one condition or more connected by "And", and each condition is related to one feature. Evaluating the rule could be measured by some metrics such as FOIL's Information gain, Laplace, and m-estimate. In this paper, FOIL's Information gain is used to evaluate the rule by calculating the difference of accuracies before and after adding a new test [12].

III. Random Forest

The methodology of an ensemble model is to combine a set of single models; each one tries to solve the same original task. It aims to obtain a better integrated global model [17] [18]. The size of the ensemble and the mechanism of combination among results of the trees should be taken into account when using ensembles [19]. Many techniques are developed for the ensemble model such as bagging, boosting, and stacking. The bagging combines the decisions of multiple trees by using the voting concept for both binary class and multi-class predictive tasks, and for the numerical predictive task, bagging calculates the average [19] [12]. Random Forest can be discriminated from other bagging techniques by choosing random subsets from features [13]. A subset would be used to select the best splitter feature for each node. Thereby, Random Forest uses randomness in the two steps: choose data instances for each single model and choose a subset of features for each node. The random forest has two parameters; the number of trees and the number of features in a subset [20].

The base model of Random Forest is mostly one of decision tree algorithms. Therefore, the best feature could be chosen from the random subset of features by using one of the metrics that explained in the previous section.

C. Classification: Deep Belief Network

For the classification process, a deep belief network (DBN) has been built to handling the nonlinearity of the selected features. DBN is a feed-forward neural network that

has a stack of Restricted Boltzmann machines [21]. The real values of this application require using Gaussian-Bernoulli Restricted Boltzmann machines (GBRBM). Therefore, the classifier combines two GBRBM, an input layer and an output layer. The model of the RBM is indicated in (4) where the energy function of the GBRBM [22] is defined as:

$$E(v, h | \theta) = \sum_{i=1}^{n_v} \frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{i=1}^{n_v} \sum_{j=1}^{n_h} w_{ij} h_j \frac{v_i}{\sigma_i} - \sum_{j=1}^{n_h} c_{h_j} \quad (4)$$

where b_i and c_{h_j} are biases corresponding to hidden and visible neurones, respectively, w_{ij} is the connection weight between visible and hidden neuron, and σ_i is the standard deviation of a Gaussian visible unit v_i .

Finally, for the fine tuning operation, the Backpropagation algorithm needs to be applied as a final training stage. The equation (5) and (6) identify the formula of the model [23],

$$y_j = f(b^1 + \sum_{i=1}^{10} (w_{ji}^1 \cdot x_i)), \quad i = 1, 2, \dots, 10, \quad j = 1, 2, \dots, n \quad (5)$$

$$z_k = f(b^2 + \sum_{j=1}^n (w_{kj}^2 \cdot y_j)), \quad k = 1, 2, \dots, p, \quad (6)$$

where $f(\bullet)$ refers to the activation function of the neurones, b^1 and b^2 are the weights of the bias in the visible and hidden layer respectively; x and z are the input and the output vectors respectively; vs , n and p are number of the elements in input, hidden and output layers respectively. By applying the feature selection algorithms, different numbers of features have been considered to feed the visible nodes of the network. The output layer represents the classes (the activities) that the system is designed to recognise. The binary coding representation [24] has been utilised to identify these classes in the output layer.

III. EXPERIMENT AND RESULTS

A. Experiment setup

This paper uses the Realistic Sensor Displacement (REALDISP) dataset [10]. The REALDISP includes data for 33 of daily human activities such as running, jump up, walking and cycling. Each sample contains 117 real-valued features extracted from nine wearable sensors. Table I specifies the positions of the wearable sensors in the human body. Each subject has three sets of data based on the displacement conditions as follows: ideal-placement, self-placement and mutual-displacement. In this paper, all of these subjects has been considered, and the dataset has been divided into training and testing sets.

TABLE I. THE POSITIONS OF THE WEARABLE SENSORS

| Symbol | Position | Symbol | Position | Symbol | Position |
|--------|-------------|--------|-----------------|--------|-----------------|
| S1 | left calf | S5 | back | S9 | right upper arm |
| S2 | left thigh | S6 | left lower arm | | |
| S3 | right calf | S7 | left upper arm | | |
| S4 | right thigh | S8 | right lower arm | | |

B. Features Selection results

Reducing the required sensors could be performing by selecting the most important features and exclude the non-significant features. In this paper, multi data mining techniques are used rather than typical feature selection methods. All features of the nine sensors have been tested

using three methods; Decision Tree, PART, and Random Forest. Evaluation the feature importance is performed by using two metrics: Information Gain and Gain ratio.

TABLE II. THE IMPORTANCE OF SENSORS ACCORDING TO TOTAL INFORMATION GAIN AND GAIN RATIO OF ITS FEATURES

| Sensor | Decision Tree | | PART | | Random Forest | |
|--------|---------------|---------------|--------------|---------------|---------------|---------------|
| | Info. Gain | Gain Ratio | Info. Gain | Gain Ratio | Info. Gain | Gain Ratio |
| S1 | 5.881 | 1.042 | 5.421 | 1.042 | 5.421 | 1.042 |
| S2 | 5.673 | 1.1214 | 5.742 | 1.1214 | 5.742 | 1.1214 |
| S3 | 6.002 | 1.1298 | 6.06 | 1.1298 | 6.06 | 1.1298 |
| S4 | 5.685 | 1.1065 | 5.617 | 1.1065 | 5.022 | 1.1065 |
| S5 | 5.587 | 1.0443 | 5.492 | 1.0443 | 5.587 | 1.0443 |
| S6 | 6.177 | 1.089 | 6.335 | 1.089 | 6.177 | 1.089 |
| S7 | 6.737 | 1.042 | 6.844 | 1.2463 | 6.737 | 1.2463 |
| S8 | 6.868 | 1.1214 | 6.836 | 1.2334 | 6.836 | 1.2334 |
| S9 | 7.329 | 1.1298 | 6.596 | 1.2024 | 6.596 | 1.2024 |

The result of the evaluation shows that PART and Random forest agrees that the sensors S1, S4 & S5 have the lowest importance based on their Gain Ratio and Information Gain metric. The results of the Decision tree partially agrees with this decision, as shown in Table II. Therefore, the selected sensors are S2, S3, S6, S7, S8 and S9.

C. Classification results

The feature selection process results in specifying the sensors that do not have a considerable consequence on the recognition rate. These sensors are S1, S4 and S5. Therefore, eight classifiers have been designed and optimised to examine the five selected cases that are shown in TABLE III. The table compares the performance of the clarified cases and the number of features in each case by using the Ideal-displacement data. First DBN is optimised to present a recognition system for the 33 activities by using the nine sensors (117 features). It achieves high overall recognition rate at 97%. Similar results have been achieved by eliminating the sensor S4 and keeping the other eight sensors. Removing the sensor S5 decrease the accuracy of the classification but with an acceptable range. This paper introduces an efficient classifier with an optimal number of sensors. The sufficient sensors that are required to monitor the 33 activities are six sensors (S2, S3, S6, S7, S8, and S9). By using the selected sensors only, the overall accuracy of the system is 96.52% with the ideal placement data.

TABLE III. CLASSIFICATION RESULTS FOR THE CONSIDERED CASES

| Case | # Considered Features | # Test Samples | Classification Results | | Accuracy (%) |
|-------------------------|-----------------------|----------------|------------------------|------|--------------|
| | | | Corrected | Fail | |
| Nine sensors (S1 to S9) | 117 | 32891 | 32042 | 849 | 97% |
| Without S1 | 104 | 32891 | 32072 | 819 | 97.51% |
| Without S4 | 104 | 32891 | 32003 | 888 | 97.30% |
| Without S5 | 104 | 32891 | 31753 | 1138 | 96.54 % |
| Without S1 & S4 & S5 | 78 | 32891 | 31904 | 987 | 96.52% |

D. Comparison with other works

Several researchers used this benchmark to evaluate their methods and approaches. [1], [2], [3] introduced their classifiers to recognize the activities with the following machine learning methods: (i) K-nearest neighbors (KNN), (ii) C4.5 decision trees (DT), (iii) nearest centroid classifier (NCC), (iv) Feature fusion, (v) Metal learner, (vi) hierarchical weighted classifier (HWC). Based on the results

of those papers, the KNN outperformed the two other methods in classifying the 33 activities for the three sets of the data: ideal-placement, self-placement and mutual-displacement. Table IV compares the classification accuracy of these works with the results of the proposed classifier that has been built by using the DBN based on features of nine sensors. The table shows that the DBN offers high-accuracy classifiers capable to recognise the data with different conditions of placements.

TABLE IV. COMPARISON OF THE ACCURACY WITH RELATED WORKS.

| Reference | Sensors Placement | Method | Accuracy (%) | |
|---|-------------------|----------------|-------------------|-----|
| [10] | Ideal | KNN | 96% | |
| | | NCC | 90% | |
| | | DT | 89% | |
| | Self-Displacement | KNN | 76% | |
| | | NCC | 72% | |
| | | DT | 32% | |
| | Mutual | KNN | 40% | |
| | | NCC | 38% | |
| | | DT | 17% | |
| [25] | Ideal | KNN | 95% | |
| | | NCC | 91% | |
| | | DT | 70% | |
| | Self-Displacement | KNN | 88% | |
| | | NCC | 77.5% | |
| | | DT | 64% | |
| | Mutual | KNN | 71% | |
| | | NCC | 61% | |
| | | DT | 60% | |
| [26] | Ideal | Metal learner | 93% | |
| | | Feature fusion | 89% | |
| | | HWC | 79% | |
| The proposed classifier with nine sensors | Ideal | DBN | 97% | |
| | | | Self-Displacement | 88% |
| | | | Mutual | 88% |

IV. CONCLUSION

This paper presents an efficient classifier to classify and monitor 33 daily physical activities by using optimized number of wearable sensors (six sensors). It aims to achieve two objectives. The first objective is to reduce the number of the required devices to recognise the activities. This can contribute to saving costs and make the system more applicable and flexible. Three data mining algorithms have been applied to calculate the optimal number of the necessary sensors and their positions by selecting only the features that have a significant impact on the recognition rate. DBN is used as a classifier to classify each activity. Various configurations have been applied to achieve the maximum possible accuracy. By reducing the three sensors using data mining algorithm, the overall accuracy of the proposed system is 96.52%. With this result, this paper presents a reliable system with a lowest possible number of sensors and higher accuracy rate.

REFERENCES

[1] R. Benzo, "Activity monitoring in chronic obstructive pulmonary disease," *Journal of cardiopulmonary rehabilitation and prevention*, vol. 29, p. 341, 2009.

[2] "World-Health-Organization". (30/12/2016). *Physical activity:Fact sheet*. Available: <http://www.who.int/mediacentre/factsheets/fs385/en/>

[3] V. H. Menec, "The relation between everyday activities and successful aging: A 6-year longitudinal study," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 58, pp. S74-S82, 2003.

[4] L. B. Andersen, P. Schnohr, M. Schroll, and H. O. Hein, "All-cause mortality associated with physical activity during leisure time, work, sports, and cycling to work," *Archives of internal medicine*, vol. 160, pp. 1621-1628, 2000.

[5] B. Chikhaoui, B. Ye, and A. Mihailidis, "Ensemble Learning-Based Algorithms for Aggressive and Agitated Behavior Recognition," in *Ubiquitous Computing and Ambient Intelligence: 10th International Conference, UCAmI 2016, San Bartolomé de Tirajana, Gran Canaria, Spain, November 29–December 2, 2016, Part II*, 2016, pp. 9-20.

[6] A. Mihailidis, J. N. Boger, T. Craig, and J. Hoey, "The COACH prompting system to assist older adults with dementia through handwashing: An efficacy study," *BMC geriatrics*, vol. 8, p. 1, 2008.

[7] E. H. Chi, J. Song, and G. Corbin, "Killer App of wearable computing: wireless force sensing body protectors for martial arts," in *Proceedings of the 17th annual ACM symposium on User interface software and technology*, 2004, pp. 277-285.

[8] K. Kunze and P. Lukowicz, "Dealing with sensor displacement in motion-based onbody activity recognition systems," in *Proceedings of the 10th international conference on Ubiquitous computing*, 2008, pp. 20-29.

[9] S. Kallio, J. Kela, P. Korpiää, and J. Mäntyjärvi, "User independent gesture interaction for small handheld devices," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 20, pp. 505-524, 2006.

[10] O. Baños, M. Damas, H. Pomares, I. Rojas, M. A. Tóth, and O. Amft, "A benchmark dataset to evaluate sensor displacement in activity recognition," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 2012, pp. 1026-1035.

[11] B. De Ville, *Decision trees for business intelligence and data mining*: SAS Institute, 2006.

[12] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*: Elsevier, 2011.

[13] T. Pang-Ning, M. Steinbach, and V. Kumar, "Introduction to data mining," in *Library of congress*, 2006.

[14] E. Frank and I. H. Witten, "Generating accurate rule sets without global optimization," 1998.

[15] S. Vijayarani and M. Divya, "An Efficient Algorithm for Classification Rule Hiding," *International Journal of Computer Applications*, vol. 33, pp. 39-45, 2011.

[16] R. Lahiri, "Comparison of data mining and statistical techniques for classification model," Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Master of Science in The Department of Information Systems & Decision Sciences by Rochana Lahiri BE, Jadavpur University, India, 2006.

[17] L. Rokach and O. Maimon, *Data mining with decision trees: theory and applications*: World scientific, 2014.

[18] G. Seni and J. F. Elder, "Ensemble methods in data mining: improving accuracy through combining predictions," *Synthesis Lectures on Data Mining and Knowledge Discovery*, vol. 2, pp. 1-126, 2010.

[19] Z.-H. Zhou, *Ensemble methods: foundations and algorithms*: CRC press, 2012.

[20] F. Livingston, "Implementation of Breiman's random forest machine learning algorithm," *ECE591Q Machine Learning Journal Paper*, 2005.

[21] A. H. Al-Fatlawi, M. H. Jabardi, and S. H. Ling, "Efficient diagnosis system for Parkinson's disease using deep belief network," in *2016 IEEE Congress on Evolutionary Computation (CEC)*, 2016, pp. 1324-1330.

[22] K. Cho, A. Ilin, and T. Raiko, "Improved Learning of Gaussian-Bernoulli Restricted Boltzmann Machines," in *Artificial Neural Networks and Machine Learning – ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I*, T. Honkela, W. Duch, M. Girolami, and S. Kaski, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 10-17.

[23] A. Al-Fatlawi, S. Ling, and H. Lam, "A Comparison of Neural Classifiers for Graffiti Recognition," *Journal of Intelligent Learning Systems and Applications*, vol. 2014, 2014.

[24] M. H. N. Jabardi and H. Kaur, "Artificial Neural Network Classification for Handwritten Digits Recognition," *International Journal of Advanced Research in Computer Science*, vol. 5, 2014.

[25] O. Banos, M. A. Toth, M. Damas, H. Pomares, and I. Rojas, "Dealing with the effects of sensor displacement in wearable activity recognition," *Sensors*, vol. 14, pp. 9995-10023, 2014.

[26] O. Baños, M. Damas, H. Pomares, and I. Rojas, "Activity recognition based on a multi-sensor meta-classifier," in *International Work-Conference on Artificial Neural Networks*, 2013, pp. 208-215.