Reducing the Response Time for Activity Recognition Through use of Prototype Generation Algorithms

Macarena Espinilla[1(✉)], Francisco J. Quesada[1], Francisco Moya[1], Luis Martínez[1], and Chris D. Nugent[2]

[1] Computer Sciences Department, University of Jaén, Campus Las Lagunillas s/n, 23071 Jaén, Spain
{mestevez,fqreal,fpmoya,martin}@ujaen.es

[2] School of Computing and Mathematics, University of Ulster, Jordanstown, Coleraine BT37 0QB, UK
cd.nugent@ulster.ac.uk

Abstract. The nearest neighbor approach is one of the most successfully deployed techniques used for sensor-based activity recognition. Nevertheless, this approach presents some disadvantages in relation to response time, noise sensitivity and high storage requirements. The response time and storage requirements are closely related to the data size. This notion of data size is an important issue in sensor-based activity recognition given the vast amounts of data produced within smart environments. A wide range of prototype generation algorithms, which are designed for use with the nearest neighbor approach, have been proposed in the literature to reduce the size of the data set. These algorithms build new artificial prototypes, which represent the data, and subsequently lead to an increase in the accuracy of the nearest neighbor approach. In this work, we discuss the use of prototype generation algorithms and their effect on sensor-based activity recognition using the nearest neighbor approach to classify activities, reducing the response time. A range of prototype generation algorithms based on positioning adjustment, which reduce the data size, are evaluated in terms of accuracy and reduction. These approaches have been compared with the normal nearest neighbor approach, achieving similar accuracy and reducing the data size. Analysis of the results attained provide the basis for the use of prototype generation algorithms for sensor-based activity recognition to reduce the overall response time of the nearest neighbor approach.

Keywords: Activity recognition · Data-driven · Nearest Neighbor (NN) · Prototype Generation (PG) · Response time

1 Introduction

Sensor-based activity recognition is an important research topic that involves multiple fields of research including pervasive and mobile computing [1], context-aware computing [2] and ambient assistive living [3].

A. Geissbühler et al. (Eds.): ICOST 2015, LNCS 9102, pp. 313–318, 2015.
DOI: 10.1007/978-3-319-19312-0_28

The process of activity recognition aims to recognise the actions and goals of one or more people within the environment based on an observation series of actions and environmental conditions. Therefore, it can be deemed as a complex process that involves the following steps: i) to choose and deploy the appropriate sensors to objects within the environment in order to effectively monitor and capture a user's behavior along with the state change of the environment; ii) to collect, store and process information and, finally, iii) to infer/classify activities from sensor data through the use of computational activity models.

The k-nearest neighbour (k-NN) approach [4] is a Data-Driven Approach that is used for sensor-based activity recognition. It is considered to be one of the most popular algorithms among all machine learning techniques, mainly due to its simplicity and overall good levels of performance [5]. This approach is based on the concept of similarity due to the fact that patterns which are similar, usually, can be allocated to the same label class. K-NN have been used successfully in the past for the purpose of activity recognition [6].

The k-NN approach does, however, suffer from several shortcomings in response time, noise sensitivity and high storage requirements [7]. The response time and storage requirements are closely related to the size of the data. In the application domain of activity recognition, the size of the data is an issue given the vast amount of sensor data generated within smart environments. In order to take advantage of the main benefits provided by the k-NN approach and to avoid the drawback associated with the size of the datasets, this work proposes to use prototype generation (PG) algorithms to reduce the data size and, as result, reducing the response time in the k-NN approach.

PG algorithms [8] are focused on the identification of an optimal subset of representative samples from the original training data. This is achieved by removing noisy and redundant examples in order to generate and replace the original data with new artificial data [9]. The use of PG algorithms to improve the process of activity recognition can be viewed as an advance given that sensor data can be annotated and used to directly inform the training process. So, training data are increased and, proportionally, the response time. Therefore, a reduction in the number of stored instances, training samples, has the ability to reduce the response time of the k-NN.

Currently, there are a wide range of PG algorithms, which are categorized by the mechanism for prototyping [8]. It is therefore necessary to evaluate which of the available approaches are best suited for the problem domain of activity recognition considering the use of k-NN as the underlying classification model.

As a starting point, this work focuses on PG algorithms based on the mechanism of *positioning adjustment*. This type of PG algorithms is usually well adapted to numerical datasets and has been shown in the past to provide good levels of performance [8]. An evaluation is undertaken to consider of the effects of reduction rate and accuracy rate for activity recognition based on sensor data gleaned from smart environments.

The remainder of the paper is structured as follows: Section 2 reviews an overview of PG algorithms, focusing on the mechanism of positioning adjustment.

Section 3 presents an empirical study that analyzes PG algorithms. Finally, in Section 4, Conclusions and Future Work are discussed.

2 Prototype Generation Algorithms

In this Section, we present an overview of the notion of PG algorithms with a particular emphasis on the mechanism of positioning adjustment. The PG algorithms are a kind of data reduction technique [10] that aim to identify an optimal subset from the original training set, by discarding noisy and redundant examples and modifying the value of some features of the samples to build new artificial samples that are known as prototypes [9].

PG algorithms are therefore designed to obtain a prototype generate set, which has a smaller size than the original training set. The cardinality of the prototype generate set is sufficiently small and has the subsequent effect to reduce both the storage and response time spent by the k-NN approach. A wide range of PG algorithms have been designed for the k-NN approach to reduce the data size which have been categorized into a taxonomy based on the following four mechanisms of prototyping [8]: *Positioning adjustment*, *Class relabeling*, *Centroid based* and *Space splitting*.

In this work, we focus on PG algorithms based on the mechanism of *positioning adjustment* to generate prototypes. The rationale for this choice of technique is due to the fact that the approach of positioning adjustment is usually well adapted to numerical datasets and has an accuracy rate close to k-NN. Based on the use of this technique an evaluation is conducted to consider the reduction rate and accuracy rate for activity recognition based on data gleaned from binary sensors in smart environments.

The mechanism of *positioning adjustment* is usually associated with two types of reduction: *fixed* or *mixed*. The *fixed* type of reduction establishes the final number of prototypes for the prototype generate set using a user's previously defined parameter related to the percentage of retention of original training set [8]. The *mixed* type of reduction begins with a preselected subset prototype generate set and then, additions, modifications, and removals of prototypes are undertaken in the prototype generate set.

3 Case Study

This Section details the evaluations under taken to investigate the effects of the performance of the PG algorithms based on positioning adjustment to enhance the response time when using the k-NN as a mean so of classification for activity recognition.

3.1 Activity Recognition Dataset

The case study presented in this contribution is based on a popular activity recognition dataset [11]. The dataset was collected in the house in which 14 state-change sensors were installed. Each sensor had the ability to provide two possible discrete values: 1 and 0, representing ON and OFF, respectively. Locations of sensors where on the doors of the apartment, cupboards, refrigerator and a toilet flush sensor. Seven different activities were annotated.

3.2 Evaluation of PG Algorithms

Eight PG algorithms based on positioning adjustment were considered. Six of them operated with fixed reduction and two of them with mixed reduction.

Table 1. PG algorithms based on positioning adjustment with fixed reduction

Acronym	Name
DSM	Decision Surface Mapping
HYB	Hybrid LVQ3 algorithm
LVQ3	Learning Vector Quantization 3
LVQPRU	Learning Vector Quantization with pruning
LVQTC	Learning Vector Quantization with Training Counter
VQ	Vector Quantization

Table 2. PG algorithms based on positioning adjustment with mixed reduction

Acronym	Name
MSE	Means of gradient descent and deterministic annealing
PSCSA	Prototype Selection Clonal Selection Algorithm

The name and acronym of each PG algorithm based on positioning adjustment is shown in Table 1 for fixed reduction and Table 2 for mixed reduction. In [8] is showed the paper in which each algorithm that is used in this paper was proposed and its optimal configuration.

3.3 Analysis and Empirical Results

Due to its simplicity and successful application, the classification rate is used as an accuracy rate. This is defined as the proportion of true results among the total number of class examined. To assess the performance of the PG algorithms, a 10-fold Cross-Validation was used to evaluate the accuracy rate of each PG algorithm. The main advantage of this validation is that all the samples in the dataset are eventually used for both training and testing. So, it matters less how the data gets divided.

The PG algorithms were implemented using Keel software [12]. Table 3 presents the average results obtained by the PG algorithms over the dataset with the 1-NN approach, indicating the reduction type in addition to the reduction rate. Furthermore, the accuracy rate is indicated together with the ranked order of approaches.

In view of the results, we can note that the k-NN approach achieves the maximum accuracy rate. Nevertheless, as already introduced, this approach presents problems of storage and, consequently, response time. The PG algorithms obtain acceptable results and dramatically reduce the training data size. As result, the response time to classify a new activity is reduced proportionally given a reduction in the number of stored instances, training samples, implies a reduction in

Table 3. Average results with 10-fold Cross-Validation obtained by the PG algorithms

Approach	Reduction Type	Reduction Rate	Accuracy Rate	Acc. Ranking
1-NN	-	1	0.963	1
PG algorithms	Reduction Type	Reduction Rate	Accuracy Rate	Acc. Ranking
DSM	Fixed	0.05	0.726	9
HYB			0.934	3
LVQ3			0.775	7
LVQPRU			0.942	2
LVQTC			0.893	5
VQ			0.743	8
PG algorithms	Reduction Type	Reduction Rate	Accuracy Rate	Acc. Ranking
MSE	Mixed	0.03	0.890	6
PSCSA		0.06	0.918	4

the response time, which is necessary to search through these training samples and classify a new activity.

The PG algorithms with a fixed reduction of 5% significantly reduces the initial training data size which has 245 samples. So, the training data size is reduced to around 12 or 13 prototypes. Therefore, the response time will be clearly reduced in the same proportion, i.e., 5%, to classify a new activity.

Among PG algorithms with a fixed reduction, it is noteworthy that LVQPRU, HYB and LVQTC achieve a very good accuracy rate that is very close to the k-NN approach. In this group, LVQ3, VQ and DSM obtain acceptable results in the accuracy rate, though slightly far from the accuracy rate of k-NN approach due to the fact that these methods preserve the accuracy over the training set. However, the generalization accuracy over the test set can be negatively affected as in this case.

The PG algorithms with a mixed reduction obtain a very good performance in terms of accuracy rate. In this group of PG algorithms, the reduction rate is not fixed, however, they achieve a reduction rate very similar to 5%. On the one hand, the PSCSA algorithm obtains the greatest reduction ratio, 3%, which is translated into 8 prototypes, obtaining a high accuracy rate. On the other hand, the PSCSA algorithm obtains the lowest reduction ratio which is 6%, which is translates into 15 prototypes.

Analyzing the results, we can point out that PG algorithms based on positioning adjustment obtain acceptable results in terms of accuracy rate, reducing the number of instances to be stored. Therefore, the PG algorithms based on positioning adjustment for activity recognition can be deemed as being very useful given that a reduction in the number of stored instances corresponds to a reduction in the response time, which is necessary to search through these training samples and classify a new sample.

4 Conclusions and Future Works

This work has been focused on the use of prototype generation algorithms for the purpose of activity recognition through use of the k-NN approach in order to reduce the response time of classification, taking into appreciation that this is closely related with the size of the set of samples stored. Eight prototype generation algorithms have been reviewed with the mechanism of positioning adjustment; Six with a fixed reduction approach and two with a mixed reduction approach. Results from the evaluation demonstrated the ability of the approach to provide a good performance and a percentage reduction of approximately 5% of the instances stored which is directly proportional to the reduction in response time. Future work will be directed towards evaluating other kinds of mechanism to generate prototypes such as *Class relabeling*, *Centroid based* and *Space splitting*.

Acknowledgments. This contribution has been supported by research projects: TIN-2012-31263, CEATIC-2013-001, UJA2014/06/14. Invest Northern Ireland is acknowledged for partially supporting this project under the Competence Centre Program Grant RD0513853 - Connected Health Innovation Centre.

References

1. Satyanarayanan, M.: Pervasive computing: Vision and challenges. IEEE Personal Communications **8**(4), 10–17 (2001)
2. Emmanouilidis, C., Koutsiamanis, R.-A., Tasidou, A.: Mobile guides: Taxonomy of architectures, context awareness, technologies and applications. Journal of Network and Computer Applications **36**(1), 103–125 (2013)
3. Alam, M., Hamida, E.: Surveying wearable human assistive technology for life and safety critical applications: Standards, challenges and opportunities. Sensors (Switzerland) **14**(5), 9153–9209 (2014)
4. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Transactions on Information Theory **13**, 21–27 (1967)
5. Wu, X., Kumar, V.: The Top Ten Algorithms in Data Mining. Chapman & Hall/CRC, 1st ed. (2009)
6. Moayeri Pour, G., Troped, P., Evans, J.: Environment feature extraction and classification for context aware physical activity monitoring, pp. 123–128 (2013)
7. Kononenko, I., Kukar, M.: Machine Learning and Data Mining: Introduction to Principles and Algorithms. Horwood Publishing Limited (2007)
8. Garcia, S., Derrac, J., Cano, J., Herrera, F.: Prototype selection for nearest neighbor classification: Taxonomy and empirical study. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(3), 417–435 (2012)
9. Lozano, M., Sotoca, J., Sanchez, J., Pla, F., Pekalska, E., Duin, R.: Experimental study on prototype optimisation algorithms for prototype-based classification in vector spaces. Pattern Recognition **39**(10), 1827–1838 (2006)
10. Wilson, R.D., Martinez, T.: Reduction techniques for instance-based learning algorithms. Machine Learning **38**(3), 257–286 (2000)
11. Van Kasteren, T., Noulas, A., Englebienne, G., Krse, B.: Accurate activity recognition in a home setting, pp. 1–9 (2008)
12. Alcala Fdez, J., Fernandez, A., Luengo, J., Derrac, J., Garcia, S., Sanchez, L., Herrera, F.: Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. Journal of Multiple-Valued Logic and Soft Computing **17**(2–3), 255–287 (2011)