# Feature Sub-set Selection for Activity Recognition

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

[1] Department of Computer Science, University of Jaén, Jaén, Spain
{fqreal,fpmoya,mestevez,martin}@ujaen.es
[2] School of Computing and Mathematics,University of Ulster, Jordanstown, Coleraine BT37 0QB, UK
cd.nugent@ulster.ac.uk

**Abstract.** The delivery of Ambient Assisted Living services, specifically relating to the smart-home paradigm, assumes that people can be provided with help, automatically and in real time, in their homes as and when required. Nevertheless, the deployment of a smart-home can lead to high levels of expense due to configuration requirements of multiple sensing and actuating technology. In addition, the vast amount of data produced leads to increased levels of computational complexity when trying to ascertain the underlying behavior of the inhabitant. This contribution presents a methodology based on feature selection which aims to reduce the number of sensors required whilst still maintaining acceptable levels of activity recognition performance. To do so, a smart-home dataset has been utilized, obtaining a configuration of sensors with the half sensors with respect to the original configuration.

**Keywords:** Activity recognition · Smart-homes · Feature selection

## 1 Introduction

The knock-on effects on ageing in society have now become widely appreciated. Health services, pension services and informal care provision are all experiencing increased levels of burden. A key focus, from a research perspective, has subsequently been identified in the area of healthy ageing and wellbeing with goal to deliver services which extend the period of time older persons can remain in their own homes.

One of the most common diseases within this cohort are cognitive related such as dementia. These illnesses are currently incurable, hence efforts are focused towards delaying their progression. In the early stages of dementia, it is useful to provide support in the form of prompting through the completion of activities of daily living (ADL) in addition to offering a series of reminders for tasks such as medication management, eating and grooming.

The advance in the miniaturization of electronic devices in addition to a reduction in their cost, has created an environment whereby we are surrounded


A. Geissbühler et al. (Eds.): ICOST 2015, LNCS 9102, pp. 307–312, 2015.
DOI: 10.1007/978-3-319-19312-0_27

by embedded sensing technology. Ambient Intelligence (AmI) characterizes a vision where humans are surrounded by computers [1]. The notion of a smart-home follows this vision with an environment of embedded technology and processing units with the ability to ascertain the behavior of its inhabitants.

At the core of this paradigm is the process of *Activity Recognition (AR)*, which gleans data from sensors embedded within the environment. Its main aim is to identify the different actions and/or activities which are taking place at that particular moment in time. Once the process recognizes the underlying activity automated assistance, in for example the form of a prompt or warning can be delivered through the smart environment itself.

Although there has been significant progress within the domain with promising results offer improvements in quality of life, it still remains expensive to deploy a full configuration of sensors within the home environment. For this reason, it is important to know which type of sensors and in which configuration are essential to detect the key ADLs. Thus, optimizing the configuration of sensors has the benefit to reduce costs from a technology perspective whilst having the additional benefit of reducing computational complexity.

In this work we focus on identifying an optimal set of sensors capable of detecting an inhabitants ADLs without a reduction in performance of the process of AR.

The remainder of the paper is structured as follows: in Section 2, the basic concepts of feature selection are reviewed. Section 3 presents the proposed methodology and a case study is presented in Section 4. The paper concludes with a Summary in Section 5.

## 2 Feature Selection

It is difficult to know, *a priori*, the relevant features which should be considered in a classification problem. It is therefore usual to gather information from multiple sources each having many features, in an effort to represent the domain as best as possible. Such an approach produces redundant or irrelevant information [2]. In addition it has the effect of increasing the size of the dataset to be processed hence increasing the computational complexity and potentially hindering the learning processes and generalization capabilities of the classifier.

Reducing the number of irrelevant or redundant features, clearly improves the time taken to deploy a learning algorithm and assists in obtaining a better insight into the concept of the underlying classification problem [3]. Thus, *Feature selection* methods aim to select a subset of relevant features to reduce the dimensionality of the classification problem without having a negative impact on classification accuracy. So, feature selection attempts to select the minimally sized subset of features according to the following criteria [4]: i) the classification accuracy does not significantly decrease; ii) the resulting class distribution, given only the values for the selected features, is as close as possible to the original class distribution, given all features.

Dash and Liu categorized the two major steps of feature selection as being the generation procedure and the evaluation function [4]:

1. **Generation Procedure** The total number of competing candidate subsets to be generated is $2^N$ if the original feature set contains $N$ number of features. There are different approaches for solving this problem:
   – *Complete* that carries out an exhaustive search for the optimal subset according to the evaluation function used.
   – *Heuristic* in which each iteration all remaining features yet to be selected (rejected) are considered for selection (rejection).
   – *Random* that sets a maximum number of possible iterations and usually search fewer number of subsets than $2^N$.
2. **Evaluation Functions** Normally, an evaluation function attempts to measure the discriminating ability of a feature or a subset to distinguish the different class labels. There are several types of evaluation functions:
   – *Distance Measures* have the idea that in a two-class problem, the most preferred features are those which induce a higher difference between the conditional probabilities of two classes. An example of this type of measure is the Euclidean distance measure.
   – *Information Measures* are based on the information gain. A feature is preferred to another if the information gain from the first feature is higher than the second.
   – *Dependence Measures* in which a feature is preferred to another if the correlation between the feature with a class is higher than the correlation between another feature and the same class.
   – *Consistency Measures* that deal with to find out the minimally sized subset that satisfies the tolerable inconsistency rate, that is normally set by the user.
   – *Classifier Error Rate Measures* that depend on the classifier itself in order to perform the feature selection.

## 3 Sensor-Based AR Optimisation

In this Section the method used in the current study to optimize the configuration of sensors within a smart environment to improve AR performance is presented. The motivation to reduce the number of sensors is two fold: firstly to reduce costs from a technology perspective and second to reduce the computational complexity of the AR process.

First of all, it is necessary to clarify that the AR method to be consider is sensor-based and data-driven [5]. Under the premise, the assumption is that a sensor network will generate an interpretable dataset, which is then used as the source to apply data mining and machine learning algorithms to recognise the activities that have been recorded in the dataset.

In this contribution it is proposed a method, which initially applies feature selection and subsequently the process of AR. Regarding the data-driven activity recognition, this proposal adds a pre-processing phase which applies feature selection to the original dataset. This pre-processing aims to remove all sensors that are irrelevant or redundant, hence avoiding unnecessary data. Once it is generated the sensor-reduced dataset, it is applied the activity recognition procedure as in the general scheme, generating the results of the process.

The application of feature selection techniques, has a knock-on effects with the process of AR. This is due to the fact that depending on the characteristics of the data there are some algorithms which are more appropriate than others.

The most popular classifiers for AR have been described by Wu et al. describe in [6] and include Naive Bayes (NB) [7], Support Vector Machines (SVMs) [8] and Nearest Neighbor (NN) [9].

## 4 Case Study

This Section details the dataset used in the current study and the effects of the feature selection on the overall accuracy of the AR process.

### 4.1 Activity Recognition Dataset

The case study presented in this contribution is based on a popular activity recognition dataset [10]. The dataset was collected in the house in which 14 state-change sensors were installed. Each sensors 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.

### 4.2 Applying Feature Selection

We have applied feature selection using Weka[11]. In practice, the way to apply feature selection is combining a generation procedure and an evaluated function as outlined in Section 2.

In the current work a complete generation method, specifically *Exhaustive Search*[12] has been used. This generation method produces an optimal result instead of the considerable computational cost.

Regarding the evaluation functions, *Dependence* and *Consistency Measures*, were used generating one dataset from each function. We have used the dependence function *CFsSubSetEval*[12], given that it produces a minimum subset of sensors and a high correlation with the class to be classified. Furthermore, the consistency function that we have used is *ConsistencySubSetEval*[12], because this evaluator seeks the smallest subset with a consistency that is equal or less as the consistency of the full attribute set.

Thus, applying the *Exhaustive Search* and the *ConsistencySubsetEval*, 10-sensors dataset was produced and, applying the *Exhaustive Search* and *CfsSubsetEval*, it is produced a 7-sensors dataset.

### 4.3 Classification Algorithms, Results and Discussion

A set of test were run using some popular classifiers used for AR. Among them, we used NB, SVMs and NN classifiers as reference methods (refer to Section 3). Apart from that, there are also used R+DRAH [13] and Decision Table (DT)[14] were also considered, given these algorithms have been previously used to perform a process which reduces the number of sensors[15].

To evaluate the classifiers' performance in different situations, we have executed with a 10 fold Cross-Validation the original dataset (14 sensors) as well as the two datasets resulting from the two optimizations: the dataset with 10 sensors and the dataset with 7 sensors. The obtained results are presented in Table 1.

We can see how SVM has the highest level of accuracy with the full range of sensors. Nevertheless, in the case of the 7 sensor dataset, the highest level of classification was achieved for the R+DRAH. The approaches based on NN, DT and R+DRAH improved their rates regarding the 10 sensors configuration. This improvement is caused by cutting of the sensors that produce irrelevant or redundant information, which act as noise, confusing the classifier.

**Table 1.** Results following 10 fold Cross-validation

| Method | Original Dataset % | 10 Sensor Dataset % | 7 Sensor Dataset % |
|---|---|---|---|
| NB | 96.33 | 96.33 | 95.51 |
| NN(k=10) | 94.69 | 92.65 | 93.88 |
| DT | 95.51 | 94.69 | 95.92 |
| SVM | **96.73** | **97.14** | 95.1 |
| R+DRAH | 93.47 | 95.51 | **98.37** |

## 5 Concluding Remarks

This contribution presents a method to optimize the number of sensors required to inform the process of activity recognition in a smart environment. Applying this technique, the number of sensors was reduced, however the level of accuracy in the process of AR was maintained. This approach therefore provides a potential reduction in cost from a technology perspective and secondly reduces the computational complexity of the AR process. Regarding future works, we aim to focus on the classification with unbalanced datasets, in order to know how unbalanced data affects the process of AR.

**Acknowledgments.** This contribution was supported by Research Projects TIN-2012-31263, ERDF and CEATIC-2013-001. Invest Northern Ireland is acknowledge for partially supporting this project under the Competence Centre Program Grant RD0513853 - Connected Health Innovation Centre.

## References


1. Streitz, N., Nixon, P.: The disappearing computer. Communications of the ACM **48**(3), 32–35 (2005)
2. John, G.H., Kohavi, R., Pfleger, K., et al.: Irrelevant features and the subset selection problem. ICML **94**, 121–129 (1994)
3. Koller, D., Sahami, M.: Toward optimal feature selection. Stanford InfoLab (1996)
4. Dash, M., Liu, H.: Feature selection for classification. Intelligent data analysis **1**(3), 131–156 (1997)
5. Chen, L., Hoey, J., Nugent, C.D., Cook, D.J., Yu, Z.: Sensor-based activity recognition. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews **42**(6), 790–808 (2012)
6. Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Philip, S.Y., et al.: Top 10 algorithms in data mining. Knowledge and Information Systems **14**(1), 1–37 (2008)
7. van Kasteren, T., Krose, B.: Bayesian activity recognition in residence for elders. IET (2007)
8. Vapnik, V.: The nature of statistical learning theory. Springer Science & Business Media (2000)
9. Cherkassky, V., Mulier, F.M.: Learning from data: concepts, theory, and methods. John Wiley & Sons (2007)
10. Van Kasteren, T., Noulas, A., Englebienne, G., Kröse, B.: Accurate activity recognition in a home setting. In: Proceedings of the 10th International Conference on Ubiquitous Computing, pp. 1–9. ACM (2008)
11. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. ACM SIGKDD explorations newsletter **11**(1), 10–18 (2009)
12. Hall, M., Witten, I., Frank, E.: Data mining: Practical machine learning tools and techniques. Kaufmann, Burlington (2011)
13. Calzada, A., Liu, J., Wang, H., Kashyap, A.: Dynamic rule activation for extended belief rule bases. In: International Conference on Machine Learning and Cybernetics (ICMLC), vol. 4, pp. 1836–1841. IEEE (2013)
14. Kohavi, R., Sommerfield, D.: Feature subset selection using the wrapper method: Overfitting and dynamic search space topology. In: KDD, pp. 192–197 (1995)
15. Calzada, A., Liu, J., Nugent, C., Wang, H., Martinez, L.: Sensor-based activity recognition using extended belief rule-based inference methodology. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2694–2697. IEEE (2014)