Computers and Electrical Engineering 68 (2018) 1–13

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Ontology-based feature generation to improve accuracy of activity recognition in smart environments☆

A.G. Salguero*,[a], M. Espinilla[b]

[a] Universidad de Cádiz, Cádiz, Spain
[b] Universidad de Jaén, Jaén, Spain

ARTICLE INFO

Keywords:
Activity recognition
Smart environments
Ambient assisted living (AAL)
Activities of Daily Living (ADL)
Ontology
Data-Driven approaches
Knowledge-Driven approaches

ABSTRACT

In recent years, many techniques have been proposed for automatic recognition of Activities of Daily Living from smart home sensor data. However, classifiers usually use features created ad hoc. In this work, the use of ontologies is proposed for the fully automatic generation of these features. The process consists of converting the original dataset into an ontology and then combine all the concepts and relations in that ontology to obtain relevant class expressions. The high formalization of ontologies allows us to reduce the search space by discarding many meaningless expressions, such as contradictory or unsatisfiable expressions. The relevant class expressions are then used as features by the classifiers to build the classification model. To validate our proposal, we have used as reference the results obtained by four different classification algorithms that use the most commonly used features.

1. Introduction

A very important process at the core of smart environments is the sensor-based activity recognition [1–3]. This kind of activity recognition is based on recognizing the actions of one or more persons within an intelligent environment by using a flow of observed events that depend only on the current activity. Common activities of interest are Activities of Daily Living (ADLs) such as “bathing”,“sleeping” or “dinning”. Objects or furniture can generate sensor events indicating, for example, the use of a faucet, the opening of a door, or the use of a light switch.

Approaches used for sensor-based activity recognition have been divided into two main kinds: Data-Driven (DDA) and Knowledge-Driven (KDA) approaches. DDA, are based on machine learning techniques in which a preexistent dataset of user behaviors is required. A training process is carried out, usually, to build an activity model which is followed by a testing process to evaluate the generalization of the model in classifying unseen activities [4]. The most remarkable features of the DDA are the capabilities of handling uncertainty and temporal information. DDA approaches need large annotated datasets for training and learning. In this context, it is interesting to mention the Open Data Initiative (ODI) [5] for Activity Recognition consortium that aims to create a structured approach to provide annotated datasets in an accessible format.

With KDA, an activity model is built through the incorporation of rich prior domain knowledge obtained from the application domain, using knowledge engineering and knowledge management techniques [6]. KDA has the advantages of being semantically clear, logically elegant, and easy to get started. In the context of KDA, ontologies for activity recognition have provided successful results [7]. In this kind of approach, interpretable activity models are built in order to match different object names with a term in an

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. A. Sangaiah.

* Corresponding author.
E-mail addresses: alberto.salguero@uca.es (A.G. Salguero), mestevez@ujaen.es (M. Espinilla).

https://doi.org/10.1016/j.compeleceng.2018.03.048
Received 31 May 2017; Received in revised form 26 March 2018; Accepted 27 March 2018

ontology that is related to a particular activity.

Ontologies can be seen as structured vocabularies that explain the relations among their terms (or classes). They are formed by concepts and relations that can be combined to form more complex class expressions. Because of the high rigidity of the logic behind ontologies, some hybrid approaches have been developed [8,9] that take advantage of the main benefits provided by DDA and the use of ontologies. Thereby, ontological ADL models capture and encode rich domain knowledge and heuristics in a machine understandable and processable way.

A hybrid approach for activity recognition is presented in this paper, where ontologies are used to automatically generate the features for the ADL classifiers. The features correspond to class expressions that have been created by combining the concepts and relations in the ontology, according to a given set of rules. Contradictory or unsatisfiable class expressions can be detected and discarded, greatly reducing the feature search space. Other similar proposals can be found in literature [10–12], but this is the only proposal that performs the process in a fully automatic way. In addition, and unlike the rest, the proposal presented in this paper does not incorporate external data to the knowledge base. It is based solely on the data available in the original dataset. The proposal presented in this paper could be considered as an approach for the problem of feature learning. However, the goal of feature learning is often to reduce the dimensionality of the dataset, selecting or aggregating features in order to produce low-dimensional versions of the original datasets [13], whereas our proposal expands the set of existing features, looking for new relevant features that help us to find hidden patterns in the dataset.

To evaluate the quality and efficiency of the methodology proposed in this work an experiment has been carried out, in which the datasets proposed in [14–16] have been used. The results obtained by using the classic approach for the recognition of ADL have been used as reference to measure the performance of our proposal. In this approach, the features are handcrafted, and usually represent the state of the sensors during the activity. Therefore, each sensor provides a single feature to the algorithm that generates the classification model. There is often more relevant information in the dataset, such as the order in which the sensors change. However, this kind of information is not usually taken into account because it requires the development of *ad hoc* applications [17]. The proposal presented in this paper automatically discovers relevant sequences of sensor changes as it generates more and more class expressions.

The remainder of the paper is structured as follows: Section 2 reviews the binary sensor data within the smart environment used in this proposal with the simple transformation into feature vectors. Furthermore, notions about ontologies are revised to understand our proposal as well as related works. Section 3 proposes the methodology to extend the set of feature vectors by means of an ontology. Section 4 presents an empirical study that analyzes our proposed methodology of extended feature vector in terms of accuracy based on three popular datasets by using the ontology. In Section 5, the results obtained are analyzed and discussed. Finally, in Section 6, conclusions and future works are presented.

2. Background

In this section, firstly, the process to transform a sensor data stream generated by a smart environment into classical feature vectors used by DDA to recognize activities is reviewed. Furthermore, the three datasets used to evaluate our proposal are described. Then, some relevant concepts related to ontologies are reviewed in order to understand our proposed methodology to extend the feature vectors with the inferred knowledge by the ontology. Finally, related works are also presented at the end of this section.

2.1. From sensor data stream to feature vectors. Smart environment datasets

Usually, feature vectors generated by a smart environment are computed from the temporal sensor data stream that is discretized into a set of time windows, denoting each time window by W^k, which is limited by each activity. The set of activities are denoted by $A = \{a_1, ..., a_i, ..., a_{AN}\}$, being AN the number of activities of the dataset.

Each feature vector is denoted by F^k and has $N_S + 1$ components, N_S being the number of sensors in the dataset denoted by $S = \{s_1, ..., s_j, ..., s_{N_S}\}$. Therefore, each computed feature vector is defined by the following expression:

$$F^k = \left\{f_1^k, ..., f_j^k, ..., f_{N_S}^k, f_{N_S+1}^k\right\}$$

being f_j^k; $j = \{1, ..., N_S\}$ a binary value that indicates if the sensor s_i was fired at least once, 1, or was not fired 0 in this time window W^k (see Fig. 1). The last component $f_{N_S+1}^k \in A$ indicates the activity carried out in the time window W^k.

In this paper, three popular activity recognition datasets of smart environments are used to evaluate our proposal, which are described below.

The first dataset was proposed in [14]. This dataset is composed by binary temporal data from a number of sensors, which monitored the ADLs carried out in a home setting by a single inhabitant. This dataset was collected in the house of a 26-year-old male who lived alone in a three-room apartment. This dataset contains 245 activities that are annotated in the stream of state-change sensors generated by 14 binary sensors. In this dataset, seven activities are classified: leave house, use toilet, take shower, go to bed, prepare breakfast, prepare dinner and, finally, getting a drink.

The second dataset was proposed in [15] that represents a sensor data stream in the Washington State University smart apartment. The data represents 20 participants performing eight ADL activities in the apartment. The activities were performed individually and sequentially. Each participant performed the same set of activities in any order. This dataset contains 178 activities that are annotated in the stream of state-change sensors generated by 45 sensors, three temperature sensor were omitted in the