Knowledge-Based Systems 90 (2015) 138–152

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Sensor-based human activity recognition system with a multilayered model using time series shapelets

Li Liu [a,b,*], Yuxin Peng [c], Ming Liu [d], Zigang Huang [e]

[a] School of Software Engineering, Chongqing University, Chongqing 400044, PR China
[b] School of Computing, National University of Singapore, Singapore 117417, Republic of Singapore
[c] School of Biomedical Engineering, National University of Singapore, Singapore 117575, Republic of Singapore
[d] Faculty of Computer and Information Science, Southwest University, Chongqing 400715, PR China
[e] School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, 85287, USA

ARTICLE INFO

Article history:
Received 27 April 2015
Revised 30 July 2015
Accepted 21 September 2015
Available online 3 October 2015

Keywords:
Human activity recognition
Complex activity
Multilayer
Time series
Sensor

ABSTRACT

Human activity recognition can be exploited to benefit ubiquitous applications using sensors. Current research on sensor-based activity recognition is mainly using data-driven or knowledge-driven approaches. In terms of complex activity recognition, most data-driven approaches suffer from portability, extensibility and interpretability problems, whilst knowledge-driven approaches are often weak in handling intricate temporal data. To address these issues, we exploit *time series shapelets* for complex human activity recognition. In this paper, we first describe the association between activity and time series transformed from sensor data. Then, we present a recursively defined multilayered activity model to represent four types of activities and employ a shapelet-based framework to recognize various activities represented in the model. A prototype system was implemented to evaluate our approach on two public datasets. We also conducted two real-world case studies for system evaluation: daily living activity recognition and basketball play activity recognition. The experimental results show that our approach is capable of handling complex activity effectively. The results are interpretable and accurate, and our approach is fast and energy-efficient in real-time.

1. Introduction

The increased accessibility of wearable and environmental sensors has raised interest in the development of human activity recognition techniques in ubiquitous computing [1]. The maturity of the miniaturized sensors and supporting technologies has pushed the research focus on context-aware triggered activity recognition and inference for a number of real-world applications, such as home monitoring and assisted living [2–4], smart hospitals [5,6], rehabilitation [7,8], physical and sport activities [9–11], terrorist detection [12], and so forth. Particularly, the prevalence of mobile devices, such as the smartphone, equipped with powerful sensors and high-speed processors, can offer advanced capabilities to recognize human activity for developing smartphone-based healthcare and wellbeing applications [13,14]. Consequently, a substantial number of projects and initiatives have been undertaken [15].

One goal of activity recognition is to uncover the knowledge of a user's behaviour that allows computing systems to proactively assist users with their tasks [16]. Computer vision-based activity recognition has been at the forefront in this research field, where a large number of researchers investigated machine recognition of gestures and activities from still images and video in well-controlled environments or constrained settings [17]. Advances in mobile devices motivated steps towards more challenging and real-world applications in unconstrained daily life settings or dedicated scenarios, where cameras cannot be deployed everywhere nor can be employed anytime.

Successful research so far has focused mainly on simplified scenarios involving single-user simple-activity recognition using sensors. The nature of human activities is complex and pose challenges that the majority of approaches and algorithms designed for simplified scenarios cannot handle in more complex scenarios [18].

Recent works have tried to address the modeling and recognition of complex activities. They generally can be categorized into two kinds of approaches, knowledge-driven and data-driven [15]. The data-driven approaches first collect sensed data, and then exploit the unseen correlations between activities and sensor data, and eventually establish a model to classify the activities. Some commonly used data-driven approaches are Hidden Markov Models [19–25],

* Corresponding author at: School of Software Engineering, Chongqing University, Chongqing 400044, PR China. Tel.: +86 13608337533.

E-mail addresses: dcsliuli@cqu.edu.cn, dcsliuli@gmail.com, dcsliuli@nus.edu.sg (L. Liu), biepeng@nus.edu.sg (Y. Peng), mingliu@swu.edu.cn (M. Liu), zigang.huang@asu.edu (Z. Huang).

http://dx.doi.org/10.1016/j.knosys.2015.09.024
0950-7051/

Conditional Random Fields [26–28], Bayesian Networks [29–33] and pattern mining techniques [34–37]. Despite the fact that most of these probabilistic and stochastic methods can handle sequential and concurrent activities with high classification accuracy, they suffer from inflexibility in recognizing different levels of activities [38,39]. They often need to construct a separate network structure for each class of activities at different levels. Unfortunately, for some applications like daily living behavior monitoring, relevant activities even cannot be clearly defined upfront [17]. On the contrary, knowledge-driven approaches start with an abstract model of common knowledge and then implement and evaluate the model through sensed data. One commonly used technique for building such a model is ontology modeling. The ontology-based approaches are semantically clear, logically elegant, and easy to interpret [15]. However, they often lack the expressive power to capture and propagate the temporal dependencies [33,40]. In addition, it would be rather difficult to handcraft each formula whose temporal relations among activities are intricate [41].

To address the aforementioned issues, we believe that *time series shapelets* (or *shapelets* for short) is a promising technique basis for recognizing complex human activities. The shapelet has been shown efficiently in classification with extensive experiments [42]. It can inherently handle the temporal information for designing real-time recognition systems, and provide accurate and interpretable results. Most importantly, it is possible to combine shapelets with semantic-based techniques in order to recognize activities in different levels and granularity [43].

This paper extends the concept and usage of shapelets to the area of sensor-based complex human activity recognition. We present four contributions: a description of the transformation from sensor data to time series, a complex activity model, a shapelet-based recognition framework, and extensive experiments with public datasets and our dataset on daily living activities and sport activities. For the transformation approach, we represent a user's activity as a set of time series, each one from a particular measured sensor attribute. A shapelet is a representative of a class of time series [44]; that is to say, an activity can be represented by a shapelet. For instance, to represent activities in a game of basketball, "walking" can be represented as a set of time series collected from accelerometer sensors on a user's leg. For the activity representation and modeling approach, we present a recursive model of complex activities that are sequentially or concurrently composed of atomic activities. For instance, the activity "dribbling" is concurrently composed of overlapping atomic activities, like "running" and "bouncing ball"; the activity "slam dunk" is composed of a sequence of atomic activities, like "jumping", "throwing ball" and "hanging on hoop"; the higher level activity "offensive" is sequentially composed of some sequences of concurrent activities "dribbling" and one sequential activity "slam dunk". For shapelet-based recognition, we employ shapelets that represent atomic activities to recognize different types of activities whose time series are defined as a combination of shapelets through sequencing and overlapping.

Compared to other activity recognition approaches, time series and shapelets have a natural temporal ordering that make the recognition approach distinct from other data-driven or knowledge-driven approaches. A major limitation of ontology-based models concerns activities with temporal intra-relationships [39]. Although the time-domain or frequency-domain features are often extracted from sensor data in data-driven or knowledge-driven approaches, they are treated as independent points without taking into account these intra-relationship among data points over time [26]. In addition, the overlapping activities are hard to distinguish using such features [18]. In terms of temporal information, time series shapelet-based recognition can handle the intra-relationship. For instance,"set-shot" and "dribbling" are two concurrent activities, but the way that atomic activities concurrently compose them are different over time. For "set-shot", "throwing ball" is performed during the "jump" and has to occur after "jumping" starts, whereas "running" and "bouncing" are performed simultaneously from the beginning for "dribbling". Time series can record the occurrence time of each atomic activity, thereby recognizing the sequencing and overlapping activities.

To evaluate the time series shapelet-based approach for complex human activity recognition, we have implemented a prototype system comprising a server for training shapelets and a smartphone app for recognizing activities in real-time. We first evaluate our approach on two open datasets for atomic activity recognition, and then conduct two case studies for complex activity recognition, namely daily living activity recognition lasting for eight days, and basketball play activity recognition with 17 basketball play relevant activities of different levels. The experimental results are promising and show the capability of handling complex activity recognition effectively. We discuss the limitations and several directions to improve our approach in the conclusion.

2. Related work

Activity Recognition has been attracting growing attention in a number of application domains due to the prevalence of mobile devices, such as smartphones and tablets, which are often equipped with various sensors and powerful processors. As such, there are many works related to activity recognition. Aggarwal and Ryoo provided an extensive review on the vision-based methods in [45]. Chen et al. reviewed sensor-based activity recognition in [15], and wearable sensor-based activity recognition was further discussed by Lara and Labrador in [1]. In this section, we will highlight the works relevant to sensor-based activity representation and models and complex activity recognition.

2.1. Activity representation and models

In terms of recognition approaches, the activity models can be categorized as appearance-based representations and structure representations.

For computer vision-based methods, activities are determined using characteristic patterns of image appearance. Chowdhury and Chellappa et al. [46,47] incorporates the Kendall statistical shape theory to model the interactions of a group of people and the activities of individuals. In [48], activity is described as a sequence of flow vectors. 2D shapes are used to model activities in [49,50]. In [51], each human activity is represented by the 3D shapes. However, sensor-based activity recognition faces a number of unique characteristics.

In computer vision recognition, the problem definitions are often clear, such as "detect an object in image", and the recognition systems are also well-defined and fixed, e.g. a defined number and type of cameras. In contrast, the sensor-based recognition system suffers heterogeneous data sources – sensors that differ in their capabilities and characteristics [52]. In addition, for some sensor-based applications, there is no common definition of human activities, such as in long-term daily living activity monitoring, where relevant activities are highly diverse and even cannot be clearly defined upfront [17]. A multilayered structure is one way to describe such unclear, complex activities. As mentioned by Blanke and Schiele [38], the hierarchical structures are necessary for the system's recognition performance.

For knowledge-driven methods, activities are often modeled by structure representations. In [39,53], a multilevel structure was proposed to model activities that are represented in four different levels: atomic gestures that are in the lowest level and cannot be decomposed; manipulative gestures that are the execution of atomic gestures; atomic activities that a sequence of manipulative gestures; and the highest level, consisting of complex activities that are concurrent executions of atomic activities. The framework described the relationships between simple and complex activities. For the majority of data-driven methods, activities are modeled by logic or finite state