Engineering Applications of Artificial Intelligence 70 (2018) 109–122

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A wavelet tensor fuzzy clustering scheme for multi-sensor human activity recognition

Hong He [a,*], Yonghong Tan [a], Wuxiong Zhang [b]

[a] College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China
[b] Key Laboratory of Wireless Sensor Network and Communication, SIMIT, Chinese Academy of Sciences, Shanghai, China

ARTICLE INFO

Keywords:
Activity recognition
Dimensionality reduction
Fuzzy clustering
Wavelet packet transform
Tensor

ABSTRACT

With the increasing number of wearable sensors and mobile devices, human activity recognition (HAR) based on multiple sensors has attracted more and more attention in recent years. On account of the diversity of human actions, the analysis of multivariate signals of activities is still a challenging task. Clustering is an unsupervised classification technique which can directly work on unlabeled data and automatically identify unknown activities. Therefore, a new wavelet tensor fuzzy clustering scheme (WTFCS) for multi-sensor activity recognition is proposed in this paper. Firstly, feature tensors of multiple activity signals are constructed using the discrete wavelet packet transform (DWPT). Then Multilinear Principal Component Analysis (MPCA) is utilized to reduce the dimensionality of feature tensors so as to keep the inherent data structure. On the basis of the principal feature initialization and the tensor fuzzy membership, a new fuzzy clustering (PTFC) is developed to identify different activity feature tensor groups. Finally, the open HAR dataset (DSAD) is used to verify the efficiency of the WTFCS. Clustering results of seventeen activities of eight subjects show that potential useful features of human activities can be captured through combining DWPT-based feature extraction with MPCA-based dimensionality reduction. The PTFC is capable of discriminating various human activities effectively. Its correctness rate of activity recognition is higher than those of fuzzy c-means clustering and the fuzzy clustering based on the tensor distance.

1. Introduction

Activity recognition (AR) aims to provide accurate and appropriate information on people's activities and behavior by interpreting attributes derived from motion, location, physiological signals and environmental information. It can be used in healthcare recovery and wellbeing, safety, surveillance, smart home and also military operations (Bulling et al., 2014; Lara and Labrador, 2013; Pantelopoulos and Bourbakis, 2010; Gravina et al., 2017). According to the data type processed by the AR system, there are two main classes: vision-based AR and sensor-based AR. The former uses visual sensing devices such as camera-based surveillance systems, while the later measures actor's motions with various inertial sensors, such as accelerometers, gyroscopes, and magnetometers (Chen et al., 2015; Woznowski et al., 2016). Although vision-based AR continues to advance, it still suffers from some problems that cannot be neglected: personal privacy of user, pervasiveness of cameras and computation complexity (Lara and Labrador, 2013; Chen et al., 2015; Woznowski et al., 2016). Compared to vision-based AR, sensor-based AR is characterized by unobtrusive measurement, less memory, low computation and power requirement (Chen et al., 2015; Woznowski et al., 2016; Barshan and Yüksek, 2014). Hence, with the fast development of new low-power, low-cost, high-capacity and miniaturized wearable sensors and mobile devices, there is a shift towards using multiple sensors worn on the body to recognize people's behavior in recent years (Bulling et al., 2014).

A typical AR system with multiple wearable sensors comprises components for data acquisition and preprocessing, data segmentation, feature extraction and selection, clustering or classification, and performance evaluation. On account of the high-dimensional and multivariate properties of sensor time series, feature extraction and selection is a crucial step in the multiple sensor-based AR process. Signal characteristics in time domain and frequency domain are widely used for feature extraction of AR systems. Time-domain features often include mean, median, variance, energy, kurtosis, ranges, etc., while frequency-domain features involve peak frequency, spectral entropy and spectral centroid,

* Corresponding author.
E-mail address: heh@shnu.edu.cn (H. He).

https://doi.org/10.1016/j.engappai.2018.01.004
Received 30 April 2017; Received in revised form 26 November 2017; Accepted 8 January 2018
Available online 7 February 2018

etc. (Lara and Labrador, 2013; Gravina et al., 2017; Attal et al., 2015). Besides, extracted features from raw signals might contain redundant or irrelevant information that will affect the recognition accuracy. It is therefore necessary to select a minimum number of appropriate features to reduce computations. To find the optimal set of features, Minimum Redundancy and Maximum Relevance (Jatoba et al., 2008), correlation based feature selection (Capela et al., 2015; González et al., 2015), minimal redundancy maximal relevance heuristic (Liu et al., 2012a), principal component analysis (PCA) (Barshan and Yüksek, 2014; Altun et al., 2010) and other feature selection methods have been employed to multivariate sensor time series. Although the dimensionality of the raw data has been greatly reduced by these approaches, some techniques use specific classifiers as the evaluation tool to produce final selected features (Attal et al., 2015; Capela et al., 2015; González et al., 2015; Zhang and Sawchuk, 2011). It limits the application of feature extraction technique. Furthermore, nowadays widely used classification methods for sensor-based AR include Decision Trees (Barshan and Yüksek, 2014; Jatoba et al., 2008; Capela et al., 2015), Navie Bayes and Bayesian networks (Barshan and Yüksek, 2014; Jatoba et al., 2008; Capela et al., 2015; Altun and Barshan, 2010), k-Nearest Neighbor (Barshan and Yüksek, 2014), Neural Networks (Barshan and Yüksek, 2014; Altun and Barshan, 2010), Support Vector Machine (SVM) (Barshan and Yüksek, 2014; Liu et al., 2012a; Altun and Barshan, 2010). All these classification approaches are vector-based supervised methods that cannot be applied directly to the multivariate sensor time series matrix. The selected features of all sensor time series for one sample are often converted into a feature vector before pattern partition. It not only destroys the original data structure, but also increases the computation complexity of classification.

In addition, supervised and unsupervised learning are two principal approaches for human activity recognition. Since an AR system should return a label such as walking, sitting, running, etc., most of human activity recognition (HAR) systems so far work in a supervised fashion (Lara and Labrador, 2013). To obtain an accurate classification model of human activities, the first training phase of supervised learning often involves a given set of examples or observations labeled with specific classes of activities for discovering patterns from extracted sensor features. Nevertheless, in some cases, labeling data is not feasible because it may require an expert to manually examine the examples and assign a label based upon their experience. The collection of annotated or "ground truth labeled" training data is usually tedious, expensive, and time consuming (Bulling et al., 2014; Lara and Labrador, 2013). Besides, there is no common definition of human activities that would allow us to follow a clear and strict rule to perform. Human activities usually reveal a great diversity. Due to multiple factors from the environment or the subject, the variability often occurs in the same activity of different individuals, or even in the same activity of the same subject. Intraclass variability and interclass similarity of human activities bring more challenges to pattern recognition (Bulling et al., 2014). If only few labeled training samples are available, it is better to use semi-supervised or unsupervised approaches, i.e. clustering algorithm, for recognizing activities (Lara and Labrador, 2013). Moreover, unsupervised classification techniques directly work on unlabeled data by discovering groups with similar features. It can automatically identify unknown activities by clustering algorithm, which is helpful for simplifying the labeling process and extracting useful features of the complex AR system (Trabelsi et al., 2013). But up till now only few HAR works use clustering methods for multi-sensor activity recognition (Bulling et al., 2014; Lara and Labrador, 2013). The common unsupervised approach is k-means clustering which can be only used for feature vectors (Wawrzyniak and Niemiro, 2015).

Recently, tensor decompositions has attracted more attention for feature extraction due to its capability of capturing multi-linear and multi-aspect structures in datasets with large scale and higher order (Kolda and Bader, 2009). Massive amounts of multimodel data can be represented as high-order tensor data by tucker decomposition or canonical/parallel factors decomposition to accomplish feature extraction (Kolda and Bader, 2009; Cong et al., 2015). When tensor is applied to represent multivariate time series, the inherent structure of data can be preserved. Useful patterns can be recognized simultaneously from multiple orders. Nowadays tensor decomposition has been used in system prediction (Tan et al., 2013), detection of characteristic modules or points (Goovaerts et al., 2014; Kouchaki et al., 2015; Pester et al., 2015), and classification of time series (Zhang and He, 2015). In an activity recognition system, activity data from multiple sensors can be regarded as multivariate time series. Therefore, a wavelet tensor fuzzy clustering scheme (WTFCS) is developed in this paper for the wearable sensor-based activity recognition. The WTFCS integrates wavelet packet transform with tensor decomposition in the feature extraction of AR data. The general framework of WTFCS for wearable sensor AR is illustrated in Fig. 1. Firstly, the AR measuring data are converted into wavelet feature tensors using the wavelet packet transform to thoroughly investigate the sample in both time domain and frequency domain. Then Multilinear Principal Component Analysis (MPCA) is utilized to extract the optimal feature subsets of wavelet feature tensors. Moreover, a tensor fuzzy clustering(PTFC) based on the tensor membership and the first principal feature initialization is developed for the pattern partition of dimension-reduced feature tensors. Finally, the open AR dataset, i.e. Daily and Spots Activities (Anon, 2018), is used to verify the efficiency of the WTFCS.

The remainder of this paper is organized as follows: Tensor expression of wearable activity recognition data is described in Section 2; Then the dimensionality reduction of wavelet tensors by the MPCA is presented in Section 3. The novel tensor fuzzy clustering is described in Section 4. The experiment on the AR data is narrated in Section 5. Finally, conclusions and future work are given in Section 6.

2. Tensor expression of activity data by discrete wavelet packet transform

Wavelet transform has been regarded as a powerful time–frequency analysis and feature extraction tool for nonstationary signals. It has fine frequency resolution and coarse time resolution at lower frequency, and coarse frequency resolution and fine time resolution at higher frequency (Gao and Wavelets, 2011; He et al., 2015). Among various feature extraction methods, wavelet analysis is particularly useful in recognizing human activities, for it can identify activity transition points and generate time–frequency characteristics while enhancing the signal (Lai et al., 2013). Compared to discrete wavelet transform, the discrete wavelet packet transform (DWPT) is performed iteratively on every sub-band at a certain level to obtain the approximation and detailed coefficients, i.e. the low and high frequency coefficients at each stage. Since the DWPT offers a richer range of possibilities for signal analysis, it is more suitable than the wavelet in the feature extraction of activity signals (Wang et al., 2007; He et al., 2017). However, most of feature extraction techniques based on DWT or DWPT so far are only applied to signals from a triaxial accelerometer (Wang et al., 2007; He et al., 2017). It has been verified that the performance of a AR system with multiple sensors is superior than that with single sensor (Gravina et al., 2017; Altun et al., 2010). Currently, wearing same type of sensor units with various signals on different positions of the body has become a common method for the data acquisition of complex activity recognition. Therefore, it is necessary to further apply the DWPT to the feature extraction of multivariate time series from diverse sensors.

2.1. Wavelet packet transform

Suppose there are P same wearable sensor units used on different parts of the body for the measurement of human activities. Each unit simultaneously generates Q sensor signals. Since dividing the whole sequence into equal time windows further decreases the computational complexity (Incel et al., 2013), the preprocessed activity signals are