Expert Systems With Applications 74 (2017) 19–28

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Learning general model for activity recognition with limited labelled data

Jiahui Wen[a,b,*], Zhiying Wang[a,b]

[a] State Key Laboratory of High Performance Computing National University of Defense Technology, Changsha 410073, China
[b] School of Computer, National University of Defense Technology, Changsha, China

ARTICLE INFO

Article history:
Received 21 July 2015
Revised 2 January 2017
Accepted 3 January 2017
Available online 6 January 2017

Keywords:
Activity recognition
General model
Co-training

ABSTRACT

Activity recognition has been a hot topic for decades, from the scientific research to the development of off-the-shelf commercial products. Since people perform the activities differently, to avoid overfitting, building a general model with activity data of various users is required before the deployment for personal use. However, annotating a large amount of activity data is expensive and time-consuming. In this paper, we build a general model for activity recognition with a limited amount of labelled data. We combine Latent Dirichlet Allocation (LDA) and AdaBoost to jointly train a general activity model with partially labelled data. After that, when AdaBoost is used for online prediction, we combine it with graphical models (such as HMM and CRF) to exploit the temporal information in human activities to smooth out the accidental misclassifications. Experiments with publicly available datasets show that we are able to obtain the accuracy of more than 90% with 1% labelled data.

1. Introduction

Activity recognition (Banos, Damas, Pomares, Prieto, & Rojas, 2012; Chernbumroong, Cang, Atkins, & Yu, 2013; Huang, Lee, Kuo, & Lee, 2010; Ordóñez, Iglesias, De Toledo, Ledezma, & Sanchis, 2013; Wen, Loke, Indulska, & Zhong, 2015a) has gained much attention during the past ten years, both scientifically and commercially, because of the widely used applications in everyday life. For example, recognising human lifestyle can help to evaluate energy expenditure (Albinali, Intille, Haskell, & Rosenberger, 2010), and walking detecting and step counting can help to monitor elderly health (Brajdic & Harle, 2013). Recently, numbers of commercial products have been released for personal purpose, such as NIKE SPORTWATCH[1], MI miband[2] and APPLE iWatch[3]. In 2013, Google announced its Android activity recognition API[4], with which people can easily develop applications detecting activities such as *Stationary, On Foot, Cycling* and *In Vehicle*.

In order to provide a robust recognition system, it is necessary to train a general model with activity data from various users. The underlying reason is that people may perform the activities differently due to their age, gender and other physical characteristics. Therefore, activity models trained for a specific user may overfit the activity data of a group of users that present the similar patterns, and can not be scaled to others. A general model is extremely important for commercial products. Since they are users-oriented, low recognition accuracy can negatively affect the user experience and decrease the profit margin. However, annotating huge amount of activity data to build a general model is expensive, time-consuming and error-prone. Therefore, this paper aims to build a general activity model with limited labelled data and unlimited unlabelled data while maintaining a satisfactory accuracy.

Human activity recognition is a hot topic in pervasive computing community, and it has been addressed by numerous previous work with different sensing modalities and learning methods. For example, Zhan, Faux, and Ramos (2014) propose hierarchical classifiers to recognition daily activities with camera and accelerometer. In the low level, LogitBoost and Support Vector Machine (SVM) are used to recognise local vision and motion features. While in the high level, Conditional Random Fields (CRFs) are leveraged to exploit the temporal information in the human behaviours and smooth out the outliers. They find that video features are more accurate for stationary activities (e.g. Watching TV) and acceleration generally has good accuracy on locomotive activities (e.g. running). In Khalifa, Hassan, and Seneviratne (2015), the authors show that the motion (kinetic) energy harvested when the people is performing the activities can be used to classify human activities. The

* Corresponding author.
E-mail addresses: j.wen@uq.edu.au (J. Wen), zywang@nudt.edu.cn (Z. Wang).

1 https://secure-nikeplus.nike.com/plus/products/sport_watch/.
2 http://www.mi.com/tw/miband/.
3 http://www.apple.com/au/watch/.
4 http://developer.android.com/google/play-services/location.html.

http://dx.doi.org/10.1016/j.eswa.2017.01.002
0957-4174/

basic idea is that different activities produce kinetic energy in a different way leaving their signatures in the harvested power signal. Therefore, they are able to recognise activities without external sensing devices such as accelerometer. In Cheng, Griss, Davis, Li, and You (2013a); Cheng et al. (2013b), the authors propose zero-shot learning to recognise unseen activities. They introduce middle-level semantic attributes to relate low-level sensor readings and high-level activities. In this way, low-level observations can be classified into middle-level semantic attributes, and unseen activities can be recognised by exploiting the relationships between the semantic attributes and the unseen activities with commonsense knowledge and domain knowledge. Sundholm, Cheng, Zhou, Sethi, and Lukowicz (2014) propose to integrate a cheap, simple textile pressure sensor matrix into exercise mats to recognise and count exercise activities which are difficult to recognise with a single body worn motions sensors. Cao, Nguyen, Phua, Krishnaswamy, and Li (2012) develop a novel framework that contains simple pre- and post- classification strategies to improve the overall performance. They address the problem of class imbalance with structure preserving oversampling and exploit the sequential nature of sensor data with smoothing and classifier fusion. In Keally, Zhou, Xing, Wu, and Pyles (2011), the authors recognise activity with smartphone-based body sensor networks. They perform retraining detection by analysing the K-L divergence of the sensor data and sensor selection by identifying sensing redundancies with decision correlations among sensors. In this way, they are able to improve classification efficiency and reduce reliance on user annotated ground truth. There are also other works that address the energy efficiency by selecting a subset of the sensors dynamically (Gordon, Czerny, Miyaki, & Beigl, 2012; Keally et al., 2011; Zappi et al., 2008) or changing the sampling rate of the sensors adaptively (Yan, Subbaraju, Chakraborty, Misra, & Aberer, 2012). Other works (Huynh, Fritz, & Schiele, 2008; Seitr, Chiu, Fritz, Amft, & Troster, 2015; Sun, Yeh, Cheng, Kuo, & Griss, 2014) even use unsupervised methods to discover frequent patterns from human daily lives with accelerometers.

However, the aforementioned methods for activity recognition are based on personal activity data, disregarding the fact that different people may have different ways to perform the activities due to their age, gender and physical characteristic. Therefore, the activity model trained on one person cannot be scaled to others who present different activity patterns. In this paper, we will show the variations that people perform the activities with publicly available datasets, and build a general activity model with activity data of various people. As data labelling is expensive and time-consuming, we learn the general model with limited labelled data. Specifically, We build initial model with AdaBoost by using the limited labelled data, and combine LDA with AdaBoost to iteratively re-estimate the posterior distribution for each example. LDA is known to be effective in collaborative learning. In our case, each user has different mixtures of activity classes, while the parameters of each activity class is globally shared among the users. In this way, sensor data from all the users are collaboratively combined to overcome the problem of label sparsity. However, as LDA cannot be applied directly to the activity data, we combine it with AdaBoost to perform collaborative learning. AdaBoost resulted from the learning process is the general activity model. Finally, when AdaBoost is deployed for prediction, we combine it with graphical models (such as HMM and CRF) to smooth out the outliers.

To conclude, we propose an activity recognition method that is able to achieve high recognition accuracy with less annotated data. It is able to deal with variants of activity patterns because it trains the activity model with data of different users collaboratively, and it can exploit the temporal information to improve recognition accuracy. The propose method is related to expert and intelligent computing area since it requires less human effort for labelling the data while maintain high activity recognition accuracy. The contributions of this paper include:

1. We demonstrate that people perform the activities differently, and activity model built on one person cannot be scaled to others that have different activity patterns.
2. By combining AdaBoost and LDA, we propose a method to build a general activity model with limited labelled data and unlimited unlabelled data;
3. We propose a novel way of combining AdaBoost with HMM and CRF to exploit the temporal characteristics of human behaviour, so as to smooth out the outliers during the online prediction;
4. We demonstrate the effectiveness of the proposed methods with publicly available datasets, and analysis its effectiveness through comprehensive experimental and comparison studies.

2. Related work

Generally, the models for recognizing human activities can be classified into two categories: knowledge-driven models (Azkune, Almeida, López-de Ipiña, & Chen, 2015; Chen, Nugent, & Wang, 2012b; Fernández-Caballero, Castillo, & Rodríguez-Sánchez, 2012; Riboni, Pareschi, Radaelli, & Bettini, 2011; Wen, Indulska, & Zhong, 2016) and data-driven models (de la Concepción, Morillo, Gonzalez-Abril, & Ramírez, 2014; Kwon, Kang, & Bae, 2014; Wen & Indulska, 2015; Wen & Wang, 2016; Wen & Zhong, 2015; Wen, Zhong, & Wang, 2015c). In knowledge-driven models, the activities are usually represented in the form of rules specified with common sense, and the models have an advantage in being reused among different environments. However, the limitation of the statically and strictly defined rules makes the models being unable to deal with noises and uncertain information in sensor readings (Gu, Chen, Tao, & Lu, 2010). By contrast, data-driven models, which are trained with realistic data, are more powerful when facing the characteristics of randomness and erratic nature of human behaviours. To name a few, they include Naive Bayesian used in Bao and Intille (2004); Tapia, Intille, and Larson (2004), HMM used in Patterson, Fox, Kautz, and Philipose (2005); Van Kasteren, Noulas, Englebienne, and Kröse (2008), Support Vector Machine(SVM) in Brdiczka, Crowley, and Reignier (2009); Cook, Krishnan, and Rashidi (2013); Zhan et al. (2014), Decision Trees in Bao and Intille (2004); Hevesi, Wille, Pirkl, Wehn, and Lukowicz (2014), KNN in Hevesi et al. (2014); Sundholm et al. (2014) and Conditional Random Fields(CRF) in Vail, Veloso, and Lafferty (2007); Zhan et al. (2014). The reader is referred to survey (Chen, Hoey, Nugent, Cook, & Yu, 2012a; Ye, Dobson, & McKeever, 2012) for more details.

Since we need to train a general activity model with labelled and unlabelled data, our work is related to traditional semi-supervised activity recognition, in which the examples classified with high confidence are used to retrain and refine the model. For example, Stikic, Larlus, and Schiele (2009); Stikic and Schiele (2009) propose multi-graph based label propagation and multi-instance learning to iteratively model activities from both labelled and unlabelled data. Lee and Cho (2014) propose a mixture-of-experts co-trained model for activity recognition with label and unlabelled data. In their model, the global model and mixture-of-expert model iteratively select the instances that they are confident with and add them to each other's training data. However, these methods are proposed based on personalized sensor data, and are not applicable for multiple users who have significantly different activity patterns. Another potential problem with semi-supervised methods is that, for example, even though many labels have comparable likelihood in a step, it only considers the most confident one and ignore the others