HETEROGENEOUS MULTIMODAL SENSORS BASED ACTIVITY RECOGNITION SYSTEM

Qiong Ning, Yiqiang Chen, Junfa Liu, Huiguo Zhang [*]

Pervasive Computing Research Center, Institute of Computing Technology, Chinese Academy of Sciences
No.6 Kexueyuan South Road Zhongguancun,Haidian District Beijing,China
{ningqiong, yqchen, liujunfa, zhanghuiguo}@ict.ac.cn

ABSTRACT

Activity recognition system is the key part in E-Health field. Traditional system needs more labeled training data to meet higher recognition accuracy. This means more calibration effort and time consumption. In this paper, with collaboration of heterogeneous multimodal sensors like a microphone, a camera and an accelerometer etc, we propose to design and implement a system to reduce the required amount of labeled data as well as achieve even better performance than traditional systems. The system consists of three phases: collaborative data collection, collaborative classifier training and collaborative classifier combination. The experimental results validate that with only 9% labeled data, our system can obtain as high accuracy as other systems which use 100% unimodal labeled data.

Index Terms— Activity recognition, Calibration effort, Heterogeneous multimodal sensors

1. INTRODUCTION

Activity recognition plays a very important role in E-Health field, such as its context-aware systems for fall detection and exercise persuasion . Apparently, higher recognition accuracy and lower calibration effort can make related systems more reliable and more popular.

However, when facing more complicated activities, existing systems just using homogeneous sensors cannot get satisfying recognition accuracy, because some activities with similar motion patterns may be confused. For example, with accelerators fixed on different body parts, it is still difficult to distinguish taking bus from taking taxi. Signals of all accelerators turn in a consistent performance. Nevertheless, differentiation degree can be improved with heterogeneous multimedia context such as image (bus and taxi) or sound.

But more heterogeneous information means more labeled data required. And labeling data is an expensive and time-consuming work. It will not be desirable to demand excessive calibration effort in practical system. So the accuracy and the calibration effort become a trade off to consider.

In this paper, we propose to design and implement a novel system to deal with this trade-off. The system includes three phases: collaborative data collection, collaborative classifier training and collaborative classifier combination. In the first phase, besides an accelerometer, we adopted a camera and a microphone to provide heterogeneous multimedia context. Enhanced co-training algorithm is proposed in the classifier training phase to reduce the amount of labeled data required while gaining better performance than standard supervised methods which use the same small labeled data set. Then product rule, one of classifiers combination rules, is adopted to further improve the recognition accuracy. Experimental results validate that our system can achieve high accuracy with low calibration effort.

The rest parts are organized as follows. In Section 2, we review the related work. And the system is detailed in Section 3. Then Section 4 presents all experimental results and our analysis. In the last Section, we conclude our work and present the future work.

2. RELATED WORK

More and more applications based on activity recognition appear in E-health field. F. Sposaro [1] presented a fall detection system using an integrated triaxial accelerometer and proposed an adaptive threshold to make the system more flexible. In order to get higher recognition accuracy for more complicated activities, Y.Q. Chen [2] fused motion information obtained from acceleration data and location information obtained from WIFI signal strength to infer activities and goals. A new sensor device [3] equipped with a camera, a microphone, an accelerometer and so on was designed to improve the recognition accuracy of IADLs (instrumental activities of daily living). However, little of these work paid special attention to the added calibration effort which hinders the extended applications of activity recognition. In order to reduce the amount of labeled data required, some researchers adopted semi-supervised learning method [4, 5]. But they often neglected the effect of classifier combination. Factually, the

[*] This work was supported in part by the National Basic Research Program of China (973 Program) under Grant No.2011CB302803, and Natural Science Foundation of China (61005045 , 90820303), and Beijing Municipal Natural Science Foundation (4112056).

978-1-61284-350-6/11/$26.00 ©2011 IEEE

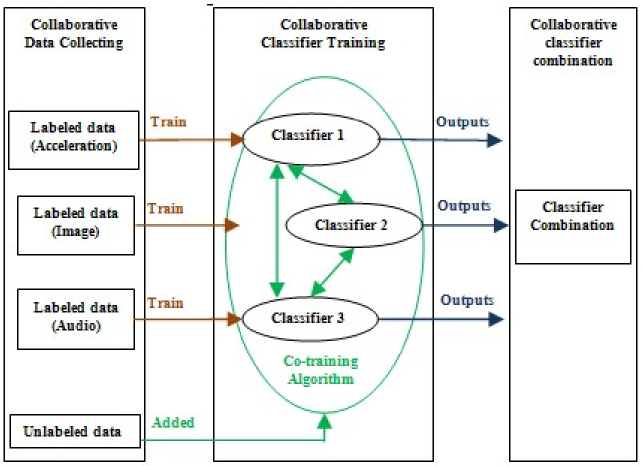


Fig. 1. Framework of activity recognition system

better the method of classifier combination used, the higher the accuracy obtained.

So with good performance and a small labeled data set required, our proposed activity recognition system is adapted to practical requirements.

3. SYSTEM IMPLEMENT

The novel activity recognition system aims at high accuracy with low calibration effort. The framework of the system is shown in Fig.1. Firstly in the phase of collaborative data collection, a small amount of labeled data together with a large amount of unlabeled data of the three heterogeneous sensors are obtained. Then in the phase of collaborative classifier training, after being trained respectively with the labeled data of three sensors, with the help of the unlabeled data, three original classifiers cooperate with each other through enhanced co-training algorithm and then get refined iteratively. Lastly in the phase of collaborative classifier combination, all of the outputs of the three classifiers are combined to further improve the accuracy of the system.

3.1. Collaborative data collection phase

Three heterogeneous sensors (a camera, a microphone and an accelerometer) are adopted to seek for more comprehensive heterogeneous information. Take into consideration that labeling data is an exhausting work while unlabeled data is easy to obtain, only a small part of data need to be labeled in our system. After raw data sets are prepared, features are extracted as follows.

- **Acceleration feature.** Features {mean, standard deviation and correlation} are extracted.
- **Image feature.** The numbers of pixels of representative colors are taken as image features [3].
- **Audio feature.** 13 order MFCCs (Mel-frequency Cepstral Coefficient) by Hamming window are calculated as audio features.

3.2. Collaborative classifier training phase

Here we take HMM (Hidden Markov Models) as the original classifier. Co-training [4, 5], in which classifiers can teach each other to augment their labeled data sets, is one of semi-supervised approaches. It can achieve good performance with small amounts of labeled training data.

However, there are two issues to be taken into account in standard co-training as follows. And some measures are taken to deal with them in our system.

- **Requirement on data set.** Data subsets for classifiers should be sufficient and independent to ensure the trustworthy of classification as well as equal distribution of the high confident samples labeled by different classifiers. As there are three heterogeneous sensors in our system, this requirement can be satisfied.
- **Computation efficiency.** Each classifier needs to pick out the most confident prediction to augment other classifiers' labeled data set. However, sometimes this procedure is time-consuming [6]. In fact, the confidence also depends on classifiers combination. We develops the procedure based on product rule [7], a simple but effective classifier combination method.

The enhanced co-training algorithm is shown in table 1.

3.3. Collaborative classifier combination phase

In this part, we explore six classifier combination rules [7] to fuse the outputs of classifiers. These rules include product rule, sum rule, max rule, median rule and majority vote rule.

There are two issues to be considered when using these rules. We take some measurements as well.

- **Requirement on combiner.** To avoid strong correlation in "misclassification", these rules demand 1) using different representation for the pattern (different features sets) or 2) using a different classification principle for each of the individual classifiers. With heterogeneous sensors, rule 1) is easily satisfied.

Table 1. Enhanced co-training algorithm

Given:

- a set L of labeled training data as three views L_1, L_2, L_3
- a set U of unlabeled data

Create a pool $U^{'}$ of examples by choosing u examples at random from U

Loop until U is empty:

- Use L_1, L_2, L_3 to train three HMM classifiers h_1, h_2, h_3 separately
- Allow h_1, h_2, h_3 to label data from U' separately
- Add p samples which have the maximal value in product rule (the most confident predictions)from $U^{'}$ to L.
- Randomly choose $2p$ examples from U to replenish $U^{'}$

- **Computational failure.** For product rule and min rule, once a classifier reports the correct class a probability as zero, the correct class cannot be identified. So we modify the probability as zero with value one to avoid the negative effect in our system.

4. EXPERIMENT

We validate our system in real environment. 13 activities {taking bus, taking subway, taking taxi, going up/down in an lift, going upstairs/down stairs, walking, running, stationary, falling, sitting down, standing up} are going to be recognized. 5-fold cross-validation is adopted in all experiments.

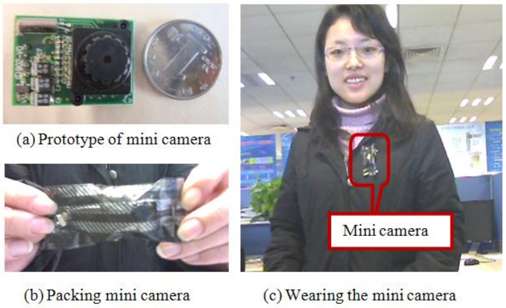
4.1. Collaborative data collection phase

As mentioned above, we select three heterogeneous sensors including an accelerometer, a microphone and a camera. We carried a Nokia N95 mobile phone equipped with an accelerometer and a microphone. And shown in figure 2, we developed a mini-camera as small as a coin. The mini-camera is attached to user's breast to collect environment information in front of user. In this phase, it took about three weeks to collect sensor data in daily life.

4.2. Collaborative classifier training phase

The performance of the enhanced co-training algorithm in this phase is evaluated.

Figure 3 shows the classification accuracies of supervised methods (blue bar) and enhanced co-training (red bar) under different percentage of labeled training data for acceleration (acc for short), image and audio, respectively. And supervised methods using 100% labeled data (blue line) are taken as the baseline for comparison. From the plot, we can see that enhanced co-training can improve the accuracy of the supervised method especially when the labeled data set is extremely small. And the increment of accuracy is highest for audio. From the plot, the superiority of enhanced co-training is obvious.



(a) Prototype of mini camera

(b) Packing mini camera

(c) Wearing the mini camera

Fig. 2. Mini camera

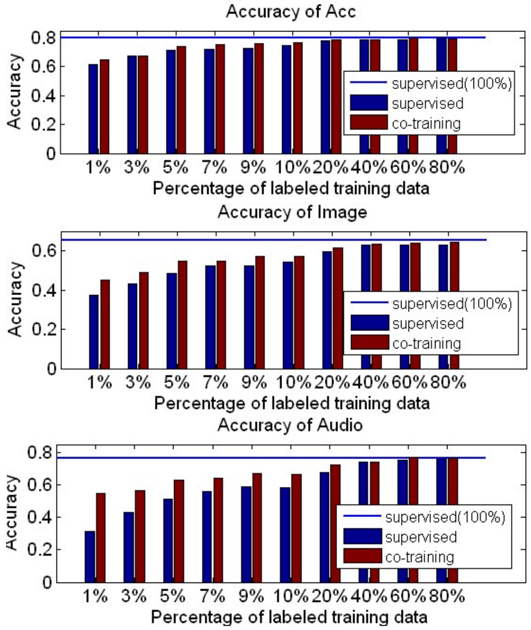


Fig. 3. Comparative performance of co-training and supervised learning method

4.3. Collaborative classifier combination phase

Based on enhanced co-training, the performance of six different classifier combination rules adopted in this phase is compared in figure 4. The dashed lines are the performance of enhanced co-training for acceleration, image and audio respectively. The solid lines represent the accuracy of six rules under different percentage of labeled training data. We can see that product rule (the magenta solid line) performs best in the six combination rules. It can increase accuracy of enhanced co-training on acceleration by 8% (from 80% to 88%), which proves the positive effect of product rule.

Table 2 shows the confusion matrices for enhanced co-training on acceleration and product rule. For convenience of comparison, we just focus on six confusing activities as shown in table. Every column represents that samples of an activity is assigned to different classes. In table 2(a) the num-

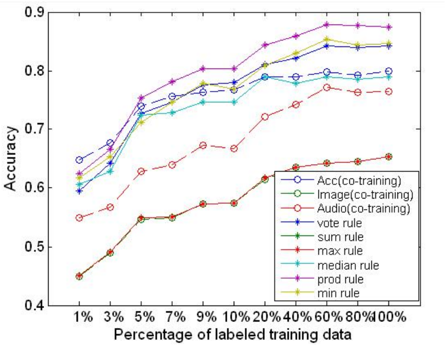


Fig. 4. Comparative performances of single classifier and classifiers combination

ber of correct predictions for each activity is colored. The red color emphasizes the number of correct predictions in product rule is more than that in enhanced co-training and the blue color for the contrary. A summary is made in Table 2(b). Although in product rule, there are 7 correct predictions less than in enhanced co-training for activities of taking taxi and going up in a lift, there are 37 more correct predictions for other activities, which means product rule is useful to distinguish those confusable activities.

4.4. Performance of the system

We compare our system with other systems which use different methods, as shown in fig 5. Here graph-based SSL [8] and LapRlsc [9] are two other well-known semi-supervised algorithms. Accuracies of single classifiers on different data sets (three straight lines) are taken as the baseline for comparison. We can see that the combination of co-training algorithm and product rule in our system outperforms other two combinations. And compared to single classifiers, only with 9% labeled data, our system can achieve as high accuracy as the system which use 100% of labeled acceleration data.

5. CONCLUSION AND FUTURE WORK

Balancing accuracy and calibration effort is an important issue in practical system of activity recognition. The contribution of this paper is that we propose to design and implement a novel system to achieve high accuracy with a small amount of labeled data collected from heterogeneous sensors. The experimental results validate the effectiveness of our system.

Table 2. Comparison of confusion matrix of the single classifier on acceleration with the one of our model

	Bus	Down-elevator	Stationary	Subway	Taxi	Up-elevator
Bus	43/47	0/0	1/1	4/5	3/7	1/0
Down-elevator	2/1	16/27	0/0	7/0	0/0	19/22
Stationary	2/2	0/0	66/73	9/1	0/0	0/0
Subway	1/5	1/0	7/0	38/53	0/0	1/0
Taxi	7/1	0/0	0/0	0/0	31/27	0/0
Up-elevator	1/0	50/39	0/0	3/3	0/0	53/50

(a)Confusion matrices (co-training on acceleration/product rule)

More correct predictions	Less correct predictions
37	7

(b)Comparison of the number of correct predictions in product rule with the one in co-training on acceleration

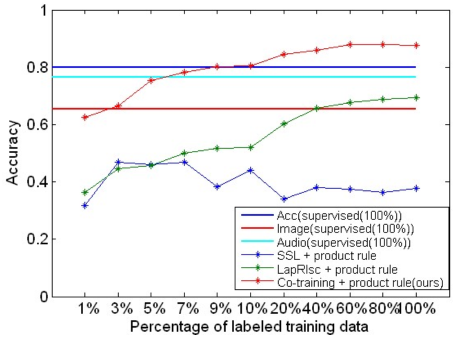


Fig. 5. Comparative performance of our system and other systems

In order to further improve our system, we plan to explore more effective methods such as active learning method to further decrease labeled data needed.

6. REFERENCES

[1] F. Sposaro and G. Tyson, "ifall: An android application for fall monitoring and response," *Conf Proc IEEE Eng Med Biol Soc.*, pp. 6119–6122, Nov. 2009.

[2] Y.Q. Chen, J. Qi, Z. Sun, and Q. Ning , "Mining user goals for indoor-location based services with low energy and high qos," *Computational Intelligence*, vol. 26, pp. 318–336, 2010.

[3] T. Maekawa and Y. Yanagisawa et al, "Object-based activity recognition with heterogeneous sensors on wrist," *Pervasive Computing*, vol. 6030, pp. 246–264, May 2010.

[4] D. Guan and W. Yuan et al, "Activity recognition based on semi-supervised learning," *Conf. RTCSA*, pp. 469–475, Aug. 2007.

[5] M. Stikic, K.V. Laerhoven, and B. Schiele, "Exploring semi-supervised and active learning for activity recognition," *IEEE ISWC*, pp. 469–475, Sep. 2008.

[6] S. Goldman and Y. Zhou, "Enhancing supervised learning with unlabeled data," *Conf. Machine Learning*, pp. 327–334, Jun. 2000.

[7] J. Kittler and M. Hatef et al, "On combining classifiers," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 20, pp. 226–239, Mar. 1998.

[8] D.Y. Zhou and O. Bousquet et al, "Learning with Local and Global Consistency," *Proc. NIPS*, pp. 9–18, 2003.

[9] M. Belkin and P. Niyogi et al, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *Machine Learning Research*, vol. 7, pp. 2399–2434, Dec. 2006.