November 7, 2014

Adaptive Activity Recognition with Dynamic Heterogeneous Sensor Fusion

Ming Zeng, *Carnegie Mellon University*
Xiao Wang, *Carnegie Mellon University*
Le T Nguyen, *Carnegie Mellon University*
Pang Wu, *Carnegie Mellon University*
Ole J Mengshoel, et al.

Adaptive Activity Recognition with Dynamic Heterogeneous Sensor Fusion

Ming Zeng, Xiao Wang, Le T. Nguyen, Pang Wu, Ole J. Mengshoel, Joy Zhang

Department of Electrical and Computer Engineering
Carnegie Mellon University
Moffett Field, CA, USA
{ ming.zeng, sean.wang, le.nguyen, pang.wu, ole.mengshoel, joy.zhang }@sv.cmu.edu

Abstract. In spite of extensive research in the last decade, activity recognition still faces many challenges for real-world applications. On one hand, when attempting to recognize various activities, different sensors play different on different activity classes. This heterogeneity raises the necessity of learning the optimal combination of sensor modalities for each activity. On the other hand, users may consistently or occasionally annotate activities. To boost recognition accuracy, we need to incorporate the user input and incrementally adjust the model. To tackle these challenges, we propose an adaptive activity recognition with dynamic heterogeneous sensor fusion framework. We dynamically fuse various modalities to characterize different activities. The model is consistently updated upon arrival of newly labeled data. To evaluate the effectiveness of the proposed framework, we incorporate it into popular feature transformation algorithms, e.g., Linear Discriminant Analysis, Marginal Fisher's Analysis, and Maximum Mutual Information in the proposed framework. Finally, we carry out experiments on a real-world dataset collected over two weeks. The result demonstrates the practical implication of our framework and its advantage over existing approaches.

Key words: Adaptive Activity Recognition, Sensor Fusion

1 Introduction

The rapid spread of wearable devices with sensing capabilities offers the opportunity for human activity recognition. Knowing a user's activity over a period of time enables applications such as continuous monitoring of user behavior, physical activity monitoring [18], abnormal activity detection [3], elderly care [23] and physical activity recognition [2].

The activity recognition is usually formulated as a classification problem [15]. Many classification methods have been leveraged in previous studies. The decision table, decision tree and naive Bayes classifier are experimented to recognize twenty predefined daily activities [2]. The support vector machine (SVM) and k-nearest neighbor (kNN) algorithm are used to perform fall detection [26]. The linear discriminant analysis and hidden Markov models are introduced to recognize predefined workshop activities [13].

However, most of the aforementioned activity recognition approaches frame activity recognition as a "static" machine learning problem, which assumes the types of activities to be recognized are predefined. This assumption does not hold for many real-life

applications such as Lifelogger [6], social activity pattern detection, etc. In these systems, the number of activities is not constant. Moreover, different users have their own definition of a "meaningful activity". It is infeasible to foresee activities that users may be interested in. So in the training phase, the systems are required to learn the most useful sensor modality combination according to different kinds of activity classes. We call these systems *Adaptive Activity Recognition Systems*

In order to recognize personal, unseen activities, some incremental methods [22, 1] are proposed. However, their results are similar to those of non-personalized models [12], indicating that the feature selection is crucial for activity recognition [12]. The semantic attribute sequence based models are also used for recognizing unseen new activities [5, 4], but still fail to consider the influence of different features. We have developed a *dynamic heterogeneous sensor fusion framework* for adaptive activity recognition. The key idea is to find the most discrimnative combination of sensor modalities (motion, sound, location, time of the day, WiFi environment, etc.) for each activity. For example, if all sensor modalities are leveraged, the system will not be able to recognize that the user is `walking` unless he walks with the same motion, at the same location and at the same time as the training `walking` examples. On the other hand, when the user annotates new types of activities, the system needs to adjust the model to use additional sensor modalities in order to discriminate a new activity from existing activities. Specifically, when the user labels an activity as `walking`, the system learns that motion feature is sufficient to recognize this activity. Several days later if the user labels a new type of activity: `grocery shopping`, which has very similar motion as `walking`, the system will need to incorporate location information to distinguish these two types of activities. Then the "motion" and "location" sensors play important roles in this case. The `sensor weight` is a value representing the importance of a sensor. To examine the effectiveness of the proposed framework, we integrate several feature transformation methods including Linear Discriminant analysis (LDA), Marginal Fisher's Analysis (MFA) and Maximum Mutual Information (MMI) algorithm.

To summarize, we develop a practical *dynamic heterogeneous sensor fusion framework*, which addresses the challenge of dynamic sensor fusion in adaptive activity recognition. The key contributions of the paper are highlighted as follows:

- We propose a sensor fusion framework to learn sensor weights for each activity class so that activities are easier to be discriminated in the new distance space. We implement several feature transformation algorithms including LDA, MFA, MMI based on our framework.
- To perform dynamic sensor fusion for each activity, we propose an adaptive activity recognition method based on the framework. In contrast to prior activity recognition methods, this framework learns the sensor weights without any prior knowledge about which sensor modalities are relatively more useful.
- Experimental results on our dataset are encouraging and confirm the effectiveness of the proposed framework on the activity recognition task.

This paper is organized as follows. We begin with a discussion of the related work. Then we introduce our framework based on adaptive activity recognition with sensor fusion, which can learn sensor weights for each activity and adjust the weights according

to the newly arrived labeled data. Then we propose feature transformation algorithms by utilizing the proposed framework. After that, we describe the dataset that we use for evaluation and data preprocessing followed by details of the experiments and discussion of results. Finally, we conclude the paper with a summary and future work.

2 Related Works

2.1 Activity Recognition and Sensor Fusion

In the past decade, different methods have been applied to a variety of sensors to address the activity recognition problem. One influential work in this area is made by Bao et al. [2]. In their experiments, accelerometer data on different body positions are used to detect activities such as walking, sitting, standing still, watching TV, running, etc. The decision tree classifier shows the best performance among other classifiers, which achieves overall recognition accuracy of 84% on the predefined activity dataset. It is worth noting that Bao's results in terms of accuracy of "snippets" rather than accuracy of activity segments. CenceMe [14] is a personal sensing system that enables members of social networks to share their sensing presence with their friends in a secure manner. Relying on a two-tier split-level activity classification based on decision tree, it captures a small set of user status in terms of activity, disposition, habit and surrounding. However, its "static" characteristic hinders them to learn a more robust model using increasing labeled data. The sequential and temporal characteristic of activities attracts attention of applying dynamic models such as the variants of the Hidden Markov Model (HMM) [7, 11]. Although building HMM models does not require labeled training data, it requires prior knowledge to define the structure of the HMM model, which limits its feasibility.

In order to train an adaptive model, an incremental active learning method is applied for daily activity recognition [17]. It uses an unsupervised incremental learning algorithm (Growing Neural Gas) to select data points that the user should label, and then updates the supervised classier for activity recognition. Abdallah et al. describe a personal model that could be incrementally trained to adapt to changes in a user's activities [1]. Another incremental learning method is achieved using probabilistic neural networks and an adjustable fuzzy clustering algorithm [22]. Models based on semantic attribute sequence are used for recognizing unseen new activities [5, 4]. However, these methods do not consider the influence of different sensors.

Fusing information from different sensors to infer high level activities is also a hot topic. There are other approaches utilizing different sensors such as camera [7], acoustic sensor [25] and GPS to address the activity recognition problem. However, all these methods assume the targeted types of activities and the most useful sensor modalities can be known beforehand, which makes them not applicable under the adaptive activity recognition scenarios. Ad-hoc solution is presented to address the problem [9]. In their approach, the sensor fusion scheme is defined by human and is application specific. Stiefmeier et al. present a fusion technique based on classifier selection [19]: two classifiers work on the motion and ultrasonic sensor modalities and generate predictions respectively. By calculating the conditional probability of correctly recognizing one

class given another classifier's result, they are able to select the most reliable prediction from one of the classifiers. This classifier-wise approach can select the most robust sensor for recognizing certain activities, but cannot fuse information from a combination of different sensors.

Although above approaches of activity recognition work well in certain scenarios, their performance is unknown if used in the adaptive activity recognition system.

2.2 Feature Transformation

Feature transformation is an important task in pattern recognition systems. It can transform high-dimensional sensor features to a subspace, which makes classification much easier. An optimal feature combination can significantly improve the recognition accuracy. The traditional feature transformation can be embedded into our framework to find the optimal sensor combination. The well-known Linear Discriminant Analysis (LDA) [8] maximizes the mean value of Kullback-Leibler divergence between different classes when classes are sampled from Gaussian distribution with different means but an identical covariance. Another feature transformation method is Marginal Fisher's Analysis (MFA) [24]. MFA combines an intrinsic graph and a penalty graph to find a feature transformation that can well separate different classes. Besides, Maximizing the joint Mutual Information (MMI) between the features and the class labels can minimize the lower bound of the classification error [21] and has the similar discriminant characteristic as LDA.

3 Learning Activity-Specific Sensor Weights

In this section, we present adaptive activity recognition framework with sensor fusion. Since different sensors may play as "experts" on different activity classes, the sensor weight vector w is learned to maximize the discrepancy according to different objective functions in the sensor modality-distance space for each activity a_k (Figure 2). For a given activity class, the problem is framed as a binary classification problem. We separate the dataset into two partitions: in-class D_{a_k} and out-class $D_{\neg a_k}$. According to different objective function, it maximizes the projected distance between class D_{a_k} and class $D_{\neg a_k}$ in the modality distance space(Figure 1).

3.1 Sensor Fusion Framework

To learn the weights of sensor for each activity, we need to build separate models for each activity. For each activity $a_k \in A$, where $|A| = K$ is a set of activity, we frame it as a binary classification task. The set of labeled data for training can be represented by $\{(x_i, c_i)\}$, $i = 1, ..., n$, $c_i \in \{0, 1\}$. $x_i \in R^M$ is the feature vector of M modalities which is a discreet/continuous random variable drawn from X. $c_i = 1$ denotes sample i belongs to activity a_k while $c_i = 0$ represents sample i is not activity a_k. For example, assume that the training set is $\{(x_1, walking), (x_2, running), (x_3, walking), (x_4, driving)\}$, to find the optimal transformation for activity "walking", the training set is converted

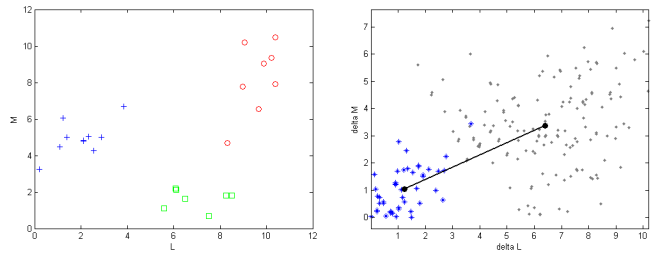

Fig. 1. Left) Original data modality space with three activity types. Right) The projected space: the blue points are all in-class modality distances, grey points are out-of-class, the line indicates the central distance between these two classes.

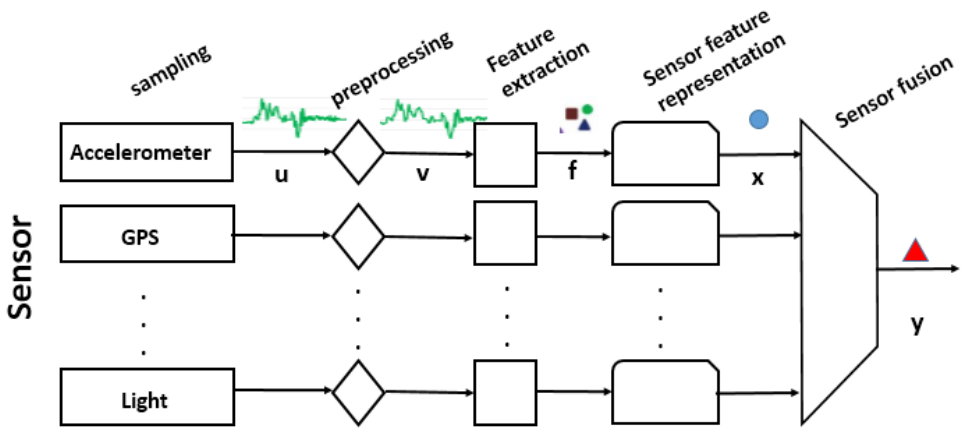

Fig. 2. The Heterogeneous Sensor Fusion Framework.

to $\{(x_1, 0), (x_2, 0), (x_3, 1), (x_4, 0)\}$. Then we try to find a feature combination in subspace R in which the discriminant of the data maximized according to different objective function. Whatever the objective is, to approximate the transformation from R^M to R, a linear/non-linear function f, wherein $w \in R^D$ can be obtained from

$$\begin{aligned} w &= \text{argmax}_w \quad J(c_i, y_i) \\ s.t., &\quad y_i = f(w, x_i) \end{aligned} \tag{1}$$

subject to specific constrains, e.g., $w^T w = 1$. The objective function $J(c_i, y_i)$ is designed for specific applications, e.g., it maximizes the discriminant of data in the selected subspace according to different assumptions or intuitions. For example, LDA selects a feature combination, where the ratio of the between-class scatter matrix and the within-class scatter matrix is maximized. $f(\cdot)$ can be a linear or non-linear transformation function upon the distribution of the data.

Once we define the objective function J and the function of transformation, we can perform gradient ascent to search for the optimal w for the framework as follows.

$$w_{t+1} = w_t + \eta \frac{\partial J}{\partial w} = w_t + \eta \frac{\partial J}{\partial y_i} \frac{\partial y_i}{\partial w}. \tag{2}$$

3.2 Incorporating Transformation Algorithms

Since we have a sensor fusion framework, any feature transformation algorithm can be embedded into according to the objective function. We provide three weighted transformation algorithms to show how to extend these classic algorithms under the proposed framework. These algorithms include the Linear Discriminant Analysis, the Marginal Fisher's Analysis, and the Maximum Mutual Information. According to Eq.(2), we only need to show $J(w)$ and $\partial J(w)/\partial w$ for adaptive sensor weight learning of conventional LDA, MFA, MMI.

Weighted Linear Discriminant Analysis Transformation Linear discriminant analysis [8] is a popular supervised feature transformation learning algorithm for learning discriminant feature transformation in the projected subspace, $y_i = w^T x_i$. It finds a feature transformation to preserve the class structure for classification defined in the high-dimensional space R^M. The class structure is described by the between-class scatter and the within-class scatter, which, respectively, measure the separation between different classes and the scatter of measurements around their corresponding class centers. The preservation is achieved by maximizing the ratio of the above two scatter matrices in R

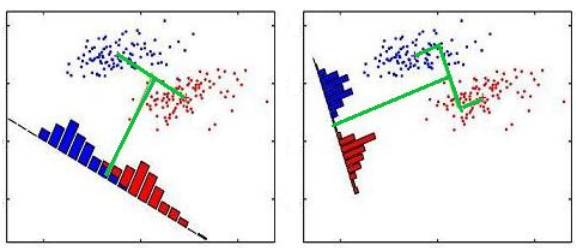

Fig. 3. The linear discriminant analysis. Left: A projection by a random vector; Right: A projection by a LDA vector.

Mathematically, centers of two classes are defined by

$$m_i = \frac{1}{n_i} \sum_{x_i \in D_i} x_i \tag{3}$$

The between-class scatter matrix is given by

$$S_b = \sum_{i=0}^{1} n_i \sum_{x_i \in D_i} (m_i - m)(m_i - m)^T \tag{4}$$

and the within-class scatter matrix is

$$S_w = \sum_{i=0}^{1} \sum_{j=1}^{n_i} (x_{ij} - m_i)(x_{ij} - m_i)^T \tag{5}$$

where n_i is the number of sample in class i, In our sensor fusion framework, the LDA objective function is given by

$$J(w) = \frac{|S_b|}{|S_w|} = \frac{w^T S_b w}{w^T S_w w} \tag{6}$$

with linear transformation $y_i = w^T x_i$. Actually, this optimization problem of (6) has a closed form solution,

$$w = S_w^{-1}(m_0 - m_1) \tag{7}$$

Weighted Marginal Fisher Analysis (MFA) Transformation LDA is motivated from the assumption that the data of each class is drawn from Gaussian distribution, which cannot always be satisfied in real world problems. Moreover, the within-class scatter cannot well characterize the separability of different classes of the data without the Gaussian distribution assumption. Algorithmically, MFA [24] constructs an intrinsic graph and a penalty graph. The intrinsic graph illustrates the inter-class compactness, and each sample is connected to its k-nearest neighbors of the same class. The within-class compactness is characterized by

$$\begin{aligned} S_w &= \sum_i \sum_{i \in N_k(j) or j \in N_k(i)} ||x_i - x_j||^2 \\ &= 2w^T X(D - E^w)X^T w \end{aligned} \tag{8}$$

where a possible way of defining E^w is as follows

$$E_{ij}^w = \begin{cases} 1 & if \quad j \in N_k(i) \quad or \quad i \in N_k(j) \\ 0 & otherwise \end{cases} \tag{9}$$

and $D_{ii}^w = \sum_{i \neq j} E_{ij}^w$, The between-class separability is characterized by

$$\begin{aligned} S_b &= \sum_i \sum_{(i,j) \in P_k(l_i) or (j,i) \in P_k(l_j)} ||x_i - x_j||^2 \\ &= 2w^T X(D - E^b)X^T w. \end{aligned} \tag{10}$$

where the definition of E^b is

$$E_{ij}^b = \begin{cases} 1 & if \quad j \in P_k(i) \quad or \quad i \in P_k(j) \\ 0 & otherwise. \end{cases} \tag{11}$$

and $D_{ii}^b = \sum_{i \neq j} E_{ij}^b$, $P_k(c)$ is a set of data pairs that are the k nearest pairs among $(i, j), i \in \pi_c, j \notin \pi_c$

The linear transformation in the sensor fusion framework is $Y = w^T X$, then the MFA feature transformation maximizes

$$J(w) = \frac{Y(D^b - E^b)Y^T}{Y(D^w - E^w)Y^T}. \tag{12}$$

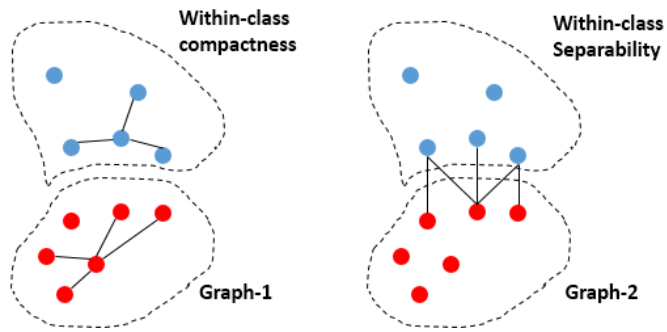

Fig. 4. Marginal Fisher Analysis. Left: The graph of intra-class with k-nearest; Right: The graph of inter-class with k-nearest

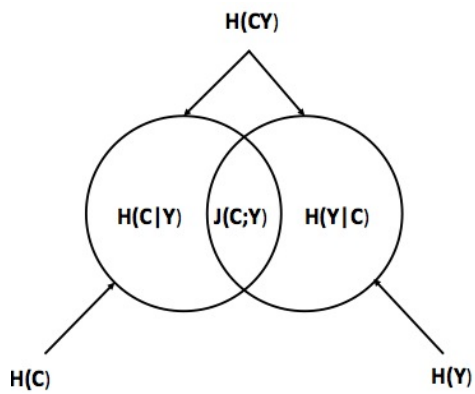

Fig. 5. The Mutual information J(C, Y).

Weighted Maximum Mutual Information (MMI) Transformation Since LDA only makes use of second-order statistical information, namely the covariance, it is optimal for data where each class has a unimodal Gaussian density with well-separated means. Under such scenarios maximum mutual information is proposed to learn a linear feature transformation.

For continuous random variable $y \in Y$ from $y = w^T x$, the uncertainty of the class label, making use of Shannon's definition, is expressed in terms of class prior probabilities

$$H(C) = -\sum_c P(c) log P(c) \tag{13}$$

After having observed a feature y, the uncertainty of the class identity is now the conditional entropy

$$H(C|Y) = -\int_y p(y) \left(\sum_c p(c|y) log p(c|y) \right) dy \tag{14}$$

The amount by which the class uncertainty is reduced, after observing the feature y, is called the mutual information $J(C,Y) = H(C) - H(C|Y)$. Our target is to find an optimal y to maximize uncertainty reduction.

Actually, by $p(c,y) = p(c|y)p(y)$ and $P(c) = \int_y p(c,y)dy$, the MI in our sensor fusion framework can be written as

$$J(C,Y) = \sum_c \int_y p(c,y) log \frac{p(c,y)}{P(c)p(y)} dy \tag{15}$$

which can also be interpreted as the Kullback-Leibler (K-L) divergence [10] between the joint density and the product of the marginal densities of the variable: $KL(p(c,y), p(c)p(y))$.

In [21], the Renyi's quadratic [16] is used instead of Shannon's entropy for computational advantages.

$$H_R(C) = -log \sum_c p(c)^2 \tag{16}$$

$$H_R(Y) = -log \int_y p(y)^2 dy \tag{17}$$

Making use of quadratic functions, the mutual information between a discrete variable C and a continuous variable Y can be estimated as

$$J(C,Y) = V_{IN} + V_{ALL} - 2V_{BTW} \tag{18}$$

where V_{IN}, V_{ALL} and V_{BTW} are "information potentials"[16, 21]

$$\begin{aligned} V_{IN} &= \sum_{c=1}^{2} \int_y p(c,y)^2 dy \\ V_{ALL} &= \sum_{c=1}^{2} \int_y p(c)^2 p(y)^2 dy \\ V_{BTW} &= \sum_{c=1}^{2} \int_y p(c,y) p(c_p) p(y) dy \end{aligned} \tag{19}$$

Parzen estimation[21] and GMM estimation[20] are proposed to estimate the joint distribution and marginal distribution to get V_{IN}, V_{ALL} and V_{BTW}.

Mutual information $J(y_i, c_i)$ can now be interpreted as an information potential induced by samples of data in different classes. It is now straightforward to derive partial $\partial J / \partial y_i$. The three components of the sum give rise to following explanation[21]: 1) samples within the same class attract each other, 2) all samples regardless of class attract each other and 3) samples of different classes repel each other. This characteristic is similar as that of LDA.

4 Adaptive Activity Recognition with Sensor Fusion

In this section, we describe the adaptive activity recognition mechanism based on the sensor fusion framework and the incremental sensor weight update method.

4.1 Adaptive Activity Recognition

After we get the sensor weights for each activity, we can apply k-NN to calculate the distance between a testing activity segment with all labeled activity segments in the weighted distances in motion, location, speed, light modalities, which results in a modality vector $x_i = [x_i^{motion}, x_i^{location}, x_i^{speed}, x_i^{light}]$. For activity a_k, we calculate the optimal transformation w_k by discriminant analysis algorithm. The learned projection is actually a weight vector which indicates the ability of sensor modalities in separating a certain activity from other activity types. With the learned weighting vectors, we modify the standard k-NN algorithm for activity recognition accordingly. For an incoming activity segment, we calculate its relative distance $d_{a_i}^r$ with each of the labeled activity segments a_i in the database using the weighting vector w_i as

Fig. 6. Adaptive Activity Recognition with Different Sensor Weights for each Activity

$$d_{a_i}^r = \frac{d_{a_i}}{d_{a_i} + d_{\neg a_i}} \tag{20}$$

where d_{a_i} is distance between the newly arrived activity segment a^{new} and center of in-class a_i, and $d_{\neg a_i}$ is distance between a^{new} and center of out-of-class $d_{\neg a_i}$.

We then label the new activity segment according to the closest activity. The whole process is shown in Algorithm 1. For each activity, the sensor weights are calculated by closed form formula directly or iterative methods. Then the incoming activity can be classified to the nearest class according the weighted distance.

Algorithm 1: Adaptive Learning Activity-Specific Sensor Weights

Input: A labeled data set $D_{label} = \{x_i^l, a_i\}$, and an incoming activity unlabeled activity x^{un}

Output: weights of sensor for each activity w and labels C of hte unlabeled data X^{un}

foreach a_k *in the activity set* A **do**

- Initialize the sensor weight for each activity $w_k = [1/M, ..., 1/M]^T$, M is the number of modalities
- Separate the label dataset D_{label} into two class: $D_0 = \{x_i, c_0\}$: the label of x_i is activity a_k and $D_1 = \{x_i, c_1\}$: the label of x_i is is not activity a_k;

repeat

- Applying gradient descent update the sensor weight w_k by eq.(2) iteratively.

until w_k *convergences*;

- for the incoming activity x^{un}, calculate its distance of (d_{a_k}) and $(d_{\neg a_k})$ respectively, then use eq.(20) to get the relative distance $d^r_{a_k}$ between x^{un} and class a_k.

end

- The label for x^{un}, r^*, is calculated by $r^* = argmin_k\{d^r_{a_k}\}$, where $i = 1, ..., K$.

4.2 Sensor Weights Updating

Our assumption is that users will label the activity incrementally. To achieve dynamic adjustment of the sensor weight by the new arrival labeled data, we need to update the parameters for the two datasets for each activity, in-class D_k^0 and out-of-class, D_k^1. For example, if we use LDA, the within-class scatter S_w and between-class scatter S_b needs to be updated according the newly arrived labeled activity.

Another simple enhancement for iteratively updating the weights w_i is by using the previous sensor fusion iteration as the starting point for current weight updating. There are intuitively appealing reasons why the previous w_i might make a good initialization. Indeed, the new set $D_{a_i}^{new}$ added into the D_{a_i} should be "similar" to some extent, and thus the optimization problem solved by current gradient ascent is also similar to the previous one.

5 Experimental Analysis

We evaluate the performance of the proposed framework in terms of overall recognition performance and its robustness under the incremental learning scenario. We compared three activity recognition algorithms including the proposed adaptive recognition algorithm with dynamic sensor fusion, milti-class SVM and motion-based k-NN recognition.

Algorithm 2: Incremental Sensor Weight Updates

Input: Incoming labeled data $D^{new} = \{(x_{N+i}, a_{N+i})\}, i = 1, ...L$
Output: updated weights of sensor for each activity $w_i, i = 1, ...N$
foreach a_i *in the activity set* A **do**
  if $a_i \in D^{new}$*'s label set* **then**
    – Set the starting point of w_i as previous value
    repeat
      – Use eq.(2) to update the sensor weight w_i;
    until w_i *convergences*;
    – Remove the samples whose labels are a_i from D^{new}
  end
end
if D^{new} *is not empty* **then**
  – Create new sensor weight w for the new coming activity
end

5.1 Dataset

We use the annotated data collected by our mobile sensing platform (Fig 7). Sensor data is collected from two phones in two weeks. During the data collection period, a user is free to use the phone without any limitation or control. At the end of each day, the user annotates the ground truth by reviewing the trajectory and the audio recorded by the phone. Accelerometer data, GPS location, speed and ambient light data are used in this experiment. The sampling rate is shown in Table 1. The dataset contains 20 activity categories. The overall activity recognition performance is reported based on the result of 10-fold cross-validation.

Sensor	Sampling Rate
Accelerometer, Ambient Light	4Hz
Microphone	8KHz
GPS, Speed	every 2 minutes

Table 1. Sensors and their sampling rate used for the single-sensor classification experiment.

5.2 Evaluation Metric

To understand the performance of the proposed framework under real application scenarios, we design an iterative training and testing strategy for evaluation. The system is initialized with very few labeled data and then new labeled data is added incrementally.

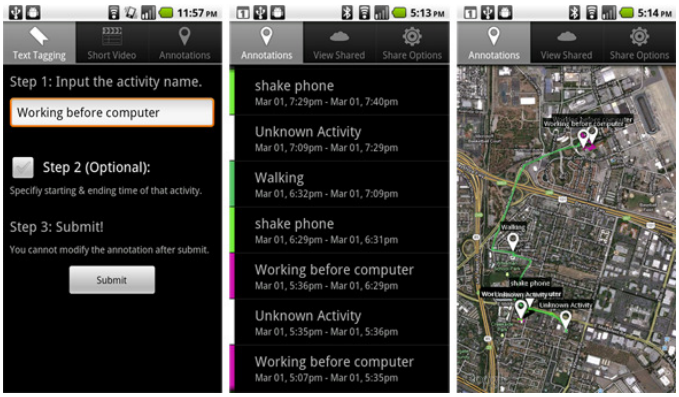

Fig. 7. We illustrate the interface of our mobile activity recognition application. Left: Naive self-report. Mid: Unsupervised activity segmentation. Right: Location trajectory review for activities.

From each of the 20 activity categories, 90% of the data will be selected randomly to the training pool and 10% left are removed the labels as test data. In the first iteration, the first 10% training data in the pool will be selected and used for training. In each of the following iterations, 10% more data is used for training. This procedure repeats 9 times until all labeled data is leveraged.

The reason of designing this iterative scheme is two-fold: we attempt to test the system's performance of classifying unknown activities as well as its capacity of correctly detecting known activities.

For the recognition, we evaluate the results using the F-measure. The F_β score can be calculated by the following formula:

$$F_\beta = (1+\beta^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}} \tag{21}$$

The performance of activity segmentation is measured by $F_{0.5}$ score because we believe recall is more important. We allow a ± 2 minutes' error-margin of activity boundary to be matched against the ground truth.

5.3 Result and Analysis

The overall results are shown in Table 2. The LDA and MFA dynamic sensor fusion algorithm outperform the motion-only k-NN (0.82 vs. 0.69, 0.73 vs. 0.69), which demonstrates that the effectiveness of the proposed framework. The multi-class SVM has the worst performance, which is 0.53. We find that the $F_{0.5}$ score of SVM can achieve 0.97 when the test is conducted on the training set. However, its performance drops significantly in the 10-folder cross-validation. One possible explanation is that SVM suffers from over-fitting when the training set is small. However, this is unavoidable in adaptive activity recognition scenario.

The simulation experiment (simulating how user provides annotations incrementally) (Figure 8) also shows that the performance of LDA in the proposed framework is

Algorithm	$F_{0.5}$ **Score**
LDA-Dynamic Sensor Fusion	0.82
MFA-Dynamic Sensor Fusion	0.73
MMI-Dynamic Sensor Fusion	0.41
Motion-only k-NN	0.69
Multi-class SVM	0.53

Table 2. The overall activity recognition performance.

robust under the incremental learning environment. Performance in the first two rounds is not stable because the number of labeled activity instances in each class is limited. The system performs much better when user annotates more instances and the performance is stable after the 4-th iteration. The $F_{0.5}$ scores remain about 0.75 after 9-th round of user annotation. The $F_0.5$ values of MFA and MMI are increasing through the increasing number of training samples. However, their $F_0.5$ values are lower than the simple LDA transformation because they are more complicated and prone to overfitting.

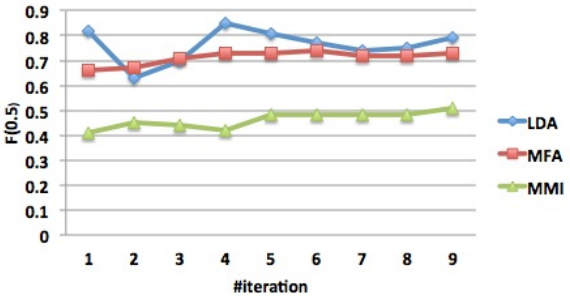

Fig. 8. We show the performance of the proposed framework in the simulated incremental activity annotation scenario. New labeled data is added in each iteration. Average $F_{0.5}$ score is 0.75.

Figure 9 shows the weights on motion (accelerometer) and location (GPS) sensor modalities of 12 activities classes. The weights for activity classes that have only one member are omitted because they are all set to the same initial value.

We observe that activities such as "Washing Dishes", "WaiteForTrain" "Paper Discussion" and "Meeting" have high weight on location, but very low reliance on motion (close to zero). This indicates that location is the most discriminative modality for these three activity types while motion is not. "Cooking", "WritingPaper", "WorkingBefore-Computer", "HavingMeal" need to incorporate both location and motion information to be distinguished from other activities, which results in more balanced weights. Notice that "Driving" also has a high weight on location. This is likely to be an artifact of our data collection as most users in this study only drive on regular routes.

Another interesting observation is that some weights are negative after the learning. For example, the motion weight of "Washing Dishes" and location weight of "Gaming" are all negative values. Since the weight vector w_k is calculated from the discriminant ratio between in-class D_{a_k} and out-of-class $D_{\neg a_k}$, a negative weight value means the

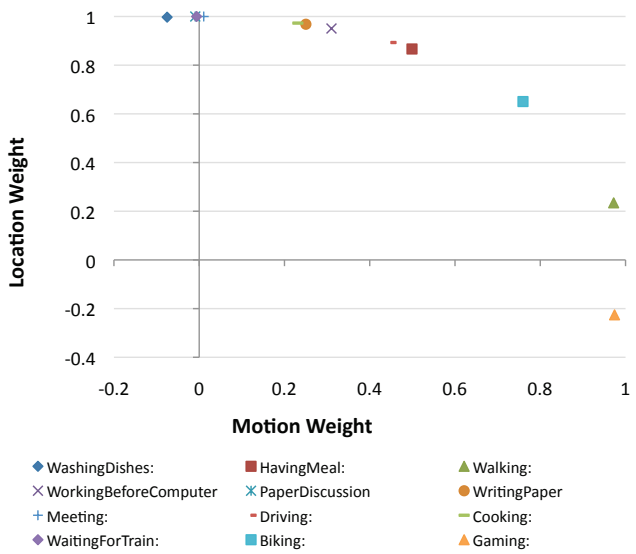

Fig. 9. The weight vector for motion and location of 12 activity classes with LDA.

diversity of activity class k on a certain data modality (e.g., motion) is larger than the average diversity of the remaining data.

6 Conclusion

In this paper, we developed a sensor fusion framework for adaptive activity recognition. The framework can evaluate the importance of each type of sensor data for each activity by feature transformation and adjust the sensor weights incrementally. Based on the proposed framework, we can apply various feature transformation algorithms including LDA, MFA and MMI. The experiments were carried out on a real world dataset and promising results were obtained. As future work, we will assess our framework on more datasets and examine its usability in more complicated activity recognition tasks.

7 Acknowledgement

This work is supported in part by the National Science Foundation through the Smart and Connected Health program under the award IIS1344768.

References

1. Z. S. Abdallah, M. M. Gaber, B. Srinivasan, and S. Krishnaswamy. Streamar: incremental and active learning with evolving sensory data for activity recognition. In *ICTAI 2012*, volume 1, pages 1163–1170. IEEE, 2012.
2. L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive computing*, pages 1–17. Springer, 2004.
3. A. Bourke, P. Van de Ven, M. Gamble, R. OConnor, K. Murphy, E. Bogan, E. McQuade, P. Finucane, G. Olaighin, and J. Nelson. Evaluation of waist-mounted tri-axial accelerometer based fall-detection algorithms during scripted and continuous unscripted activities. *Journal of biomechanics*, 43(15):3051–3057, 2010.
4. H.-T. Cheng, M. Griss, P. Davis, J. Li, and D. You. Towards zero-shot learning for human activity recognition using semantic attribute sequence model. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 355–358. ACM, 2013.
5. H.-T. Cheng, F.-T. Sun, M. Griss, P. Davis, J. Li, and D. You. Nuactiv: Recognizing unseen new activities using semantic attribute-based learning. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, pages 361–374. ACM, 2013.
6. S. Chennuru, P.-W. Chen, J. Zhu, and J. Y. Zhang. Mobile lifelogger–recording, indexing, and understanding a mobile users life. In *Mobile Computing, Applications, and Services*, pages 263–281. Springer, 2012.
7. T. V. Duong, H. H. Bui, D. Q. Phung, and S. Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-markov model. CVPR '05, pages 838–845, Washington, DC, USA, 2005. IEEE Computer Society.
8. J. H. Friedman. Regularized discriminant analysis. *Journal of the American statistical association*, 84(405):165–175, 1989.
9. I.-J. Kim, S. C. Ahn, H. Ko, and H. G. Kim. Automatic lifelog media annotation based on heterogeneous sensor fusion. In *Proceedings of IEEE International Conference on Multi Sensor fusion and Integration for Intelligent systems*, Seoul, Korea, August 20-22 2008.
10. S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, pages 79–86, 1951.
11. X. Lexing, C. Shih-Fu, D. Ajay, and S. Huifang. Unsupervised discovery of multi-level statistical video structures using hierarchical hidden markov models. *International Conference on Multimedia and Expo (ICME)*, 2003.
12. J. W. Lockhart and G. M. Weiss. The benefits of personalized smartphone-based activity recognition models. *SIAM*, 2014.
13. P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, and T. Starner. Recognizing workshop activity using body worn microphones and accelerometers. In *Pervasive Computing*, pages 18–32. Springer, 2004.
14. E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell. Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 337–350. ACM, 2008.
15. S. J. Preece, J. Y. Goulermas, L. P. Kenney, D. Howard, K. Meijer, and R. Crompton. Activity identification using body-mounted sensorsa review of classification techniques. *Physiological measurement*, 30(4):R1, 2009.
16. J. C. Principe, D. Xu, and J. Fisher. Information theoretic learning. *Unsupervised adaptive filtering*, 1:265–319, 2000.

17. J. Rebetez, H. F. Satizábal, and A. Perez-Uribe. Reducing user intervention in incremental activityrecognition for assistive technologies. In *ISWC 2013*, pages 29–32. ACM, 2013.
18. A. Reiss, D. Stricker, and G. Hendeby. Confidence-based multiclass adaboost for physical activity monitoring. In *ISWC 2013*, pages 13–20. ACM, 2013.
19. T. Stiefmeier, G. Ogris, H. Junker, P. Lukowicz, and G. Troster. Combining motion sensors and ultrasonic hands tracking for continuous activity recognition in a maintenance scenario. In *Wearable Computers, 2006 10th IEEE International Symposium on*, pages 97 –104, oct. 2006.
20. K. Torkkola. Learning discriminative feature transforms to low dimensions in low dimensions. In *NIPS*, pages 969–976, 2001.
21. K. Torkkola and W. M. Campbell. Mutual information in learning feature transformations. In *ICML*, pages 1015–1022. Citeseer, 2000.
22. Z. Wang, M. Jiang, Y. Hu, and H. Li. An incremental learning method based on probabilistic neural networks and adjustable fuzzy clustering for human activity recognition by using wearable sensors. *Information Technology in Biomedicine, IEEE Transactions on*, 16(4):691–699, 2012.
23. J. Witlox, L. S. Eurelings, J. F. de Jonghe, K. J. Kalisvaart, P. Eikelenboom, and W. A. Van Gool. Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. *Jama*, 304(4):443–451, 2010.
24. S. Yan, D. Xu, B. Zhang, and H.-J. Zhang. Graph embedding: A general framework for dimensionality reduction. In *CVPR 2005*, volume 2, pages 830–837. IEEE, 2005.
25. K. Yatani and K. N. Truong. Bodyscope: a wearable acoustic sensor for activity recognition. In *Ubicomp 2012*, pages 341–350. ACM, 2012.
26. T. Zhang, J. Wang, P. Liu, and J. Hou. Fall detection by embedding an accelerometer in cellphone and using kfd algorithm. *International Journal of Computer Science and Network Security*, 6(10):277–284, 2006.