

HMM Based Semi-Supervised Learning for Activity Recognition

Marjan Ghazvininejad, Hamid R. Rabiee, Nima Pourdamghani, Parisa Khanipour

AICTC Research Center, Department of Computer Engineering
Sharif University of Technology, Tehran, Iran
rabiee@sharif.edu

ABSTRACT

In this paper, we introduce a novel method for human activity recognition that benefits from the structure and sequential properties of the test data as well as the training data. In the training phase, we obtain a fraction of data labels at constant time intervals and use them in a semi-supervised graph-based method for recognizing the user's activities. We use label propagation on a k -nearest neighbor graph to calculate the probability of association of the unlabeled data to each class in this phase. Then we use these probabilities to train an HMM in a way that each of its hidden states corresponds to one class of activity. These probabilities are used to learn the transition probabilities between hidden states of the HMM which is used to predict the classes of the test data. Experimental results show that the proposed method consistently outperforms the existing state of the art semi-supervised methods.

Author Keywords

Activity recognition, Semi-Supervised, Manifold, HMM.

ACM Classification Keywords

I.5 [Computing Methodologies]: Pattern Recognition

General Terms

experimentation, Algorithms

INTRODUCTION

Recognizing patterns of human activities is an essential building block for providing context-aware services in many applications, including intelligent environments [3], automated visual surveillance [5], human robot interaction [2], and assistive technology for the disabled [11]. Moreover, activity recognition has attained considerable interest in recent years due to its applications in elderly care. In health care, activity recognition can be used to automatically monitor the activities of daily living (ADLs) of old people, and offer just-in-time assistance. A wide range of sensors have been used for various applications of activity recognition. Using cameras

to capture videos and pictures as input data violates the sense of privacy for users and can not be used widely (specially in house settings). Therefore, the focus of our work is on using data from wearable sensors. However, the process of data annotation still requires the use of cameras and microphones which imposes excessive cost and effort. An alternative way to label the sensor data is to ask users to constantly report their activities in the training phase. This process can be exhaustive for the users, hence it can not be considered as a practical solution. One solution is achieved by using experience sampling [1], which is a method based on asking users to provide information about their activities in certain time intervals. This method is fast and easy to use, and does not bother users as far as the sampling rate is low enough. Consequently, in this setting only a fraction of data in the training phase is labeled. To decrease the level of experience sampling, the Multi-Graph based method introduced in [13] is used to estimate the underlying structure of data with the use of both labeled and unlabeled data in hand. In this paper, we propose an activity recognition method that utilizes the structure and sequential properties of the test data as well as the manifold structure of the training data. Our contribution is improving the previous Semi-Supervised Learning (SSL) graph-based activity recognition method [13] in two ways. First, by utilizing label propagation on the graph constructed on both labeled and unlabeled training data, we gain the association probabilities of data points to each class of activity. These probabilities are used to train an HMM in a way that each of its hidden states corresponds to one class of activity. We use these probabilities to learn the transition probabilities between hidden states of the HMM. Second, the manifold structure of the training data is used to estimate the probability distribution of test data in each of the states. Practically, using the manifold structure enables us to obtain observation probabilities of the HMM. Finally, with the aid of Viterbi algorithm we find the maximum likelihood sequence of hidden states or labels for test data.

The remaining of the paper is structured as follows. First, an overall discussion of the similar work is presented in the related work section, then basics and notations are introduced to provide the readers with the necessary concepts, followed by the proposed method section in which we explain our algorithm and justify the theory. Finally, the experimental results of applying our method on a well-known dataset are presented.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SAGAware'11, September 18, 2011, Beijing, China.

Copyright 2011 ACM 978-1-4503-0926-4/11/09...\$10.00.

RELATED WORK

For many years, supervised algorithms have been the main stream in activity recognition applications. Several probabilistic models have been proposed to model the sequence classification problem of activity recognition. These models include naive Bayes [15], decision trees [10], Support Vector Machines [8], boosting [12], Conditional Random Fields [16] and most common of all Hidden Markov Models [18] [4] [6].

In spite of their promising results in different applications of activity recognition, the usability of supervised algorithms in real-world cases remains in question due to the difficult and costly process of manually labeling the activity data.

In order to reduce the cost by precluding the need for labeled data, help in adaptation to non-stationary patterns, and for providing early exploratory tools, several unsupervised algorithms have been proposed [9]. Despite the aforementioned advantages of these methods, they fail to achieve acceptable accuracy levels. Fully unsupervised techniques lack a proper assumption over target activities in the system. Therefore, there is a risk of grouping data into clusters that do not correspond to the activity patterns required for the application.

The cost of data labeling and the amount of relatively inexpensive unlabeled data in hand, are two main reasons for the recent interest in semi-supervised learning (SSL) methods for activity recognition. SSL methods have eliminated the need for a large amount of labeled data in contrast to supervised methods. It has been shown that these methods can get acceptable accuracy compared to supervised methods which benefit from the labels of all the data. In this setting, it suffices to ask users to provide labels about their current activity on fixed time intervals and use these labels and other unlabeled sensor data to learn the activity pattern.

Some recent works use procedures such as self-training, co-Training [14], and En-Co-Training [7]. However they can not use unlabeled data in a proper manner and their results is not acceptable. Authors in [1] use multi-instance learning method for activity recognition and represent activity data as bags-of-activities. This representation enables them to require labels only on bag level. However this method does not use the information about underlying structure of data in a proper manner.

Although Graph-based techniques have been proven to be successful in different areas of machine learning, they have not yet been explored much in activity recognition applications. There are a few works that use this framework for activity recognition [13] [17]. However these works do not benefit from this structure in predicting the labels of test data. This structure is just used to estimate the labels of unlabeled data in the training phase. For instance, [13] constructs a graph on both labeled and unlabeled data that are provided in training phase. And then uses label propagation to estimate the labels for unlabeled data. Once all labels are

obtained, a Support Vector Machine (SVM) is trained to be used for classification.

Our proposed method uses the base idea introduced in [13].

However, we extend this work in two ways: first, as the structure of training and testing data is the same, and the unlabeled data in the training phase is adequate for capturing this structure properly, we utilize the structure obtained in the training phase for the test data as well. Second, our algorithm is aware of the sequential property of the test data, while the two previous methods allocate labels to each test data independently.

BASICS AND NOTATIONS

Manifold

Semi supervised methods surpass the supervised ones due to their stronger assumptions over the problem. In general, every learning algorithm needs some kind of prior knowledge (or assumption) over the problem to obtain generalization. However, if an assumption is not satisfied in a problem, it might mislead the algorithm and result in a low performance. Hence, there is a trade off between generality and strength of the assumption. One of the most common assumptions in SSL, which is both general and strong, is the manifold assumption. Manifold assumption is based on the intuition that label of the data in the dense areas of the feature space has to be similar. In an equivalent notation, manifold assumption assumes the input data to lie on a low dimensional manifold and requires the labels of the data to change smoothly over this manifold.

In activity recognition with the use of wearable sensor data, the first requirement is satisfied due to low degree of freedom in most of human activities. The second condition is also met because small variations in our inputs rarely affect the activity class.

As we do not know the exact structure of the manifold, we estimate it with a weighted graph. A manifold can be represented by sampling a finite number of points as the graph nodes, and drawing undirected edges between the nodes which are close on the manifold. In this paper we use k -nearest neighbors method, one of the classical methods in graph construction.

Label Propagation

One of the most common methods for employing manifold assumption in predicting the labels of the unlabeled data is label propagation. In label propagation, we employ some fixed sources of label (the vertices corresponding to the labeled data) along with some varying sources (the unlabeled data) which propagate their labels through the edges of the graph that models the structure of the data.

In what follows, we give a formal description of the label propagation algorithm used in this work. Let $X_L = \{x_1, \dots, x_l\}$ be the feature vectors of the labeled data and $Y_L = \{y_1, \dots, y_l\}$ represent their labels. Also, suppose that $X_U = \{x_{l+1}, \dots, x_{l+u}\}$ be the feature vectors of the unlabeled data. Our goal is to estimate the labels of X_U named $Y_U = \{y_{l+1},$

$\dots, y_{l+u}\}$ where $y_i \in \{1, \dots, C\}$ and C represents the number of activity classes. To this aim, we build a graph on the input data $X = X_L \cup X_U$ with adjacency matrix W where $w_{i,j} \in \{0, 1\}$. We define the transition probability matrix T as:

$$T_{i,j} = \frac{w_{i,j}}{\sum_{k=1}^{l+u} w_{i,k}}, i, j \in \{1, \dots, l+u\}$$

in this notation, $T_{i,j}$ represents the probability that a label is conducted from node i to node j .

Label propagation works in an iterative manner. In each iteration, labels are transmitted according to the transition probability matrix. To describe the transition process formally, we define Z_k as a $(l+u) \times C$ matrix that represents the distribution of the estimated labels in iteration k . In other words, $Z_{k,t,j}$ represents the probability that x_t belongs to class j in iteration k . Z_1 is initialized as follows: $Z_{1,t,j} = 1$ if t is a labeled data with class j and $Z_{1,t,j} = 0$ otherwise.

The propagation process consists of three steps in each iteration.

First, the matrix Z is updated as below:

$$Z_{k+1} = T \times Z_k$$

Second, the rows of Z are normalized so that they represent an accurate distribution:

$$Z_{k+1,t,j} = \frac{Z_{k+1,t,j}}{\sum_{j=1}^C Z_{k+1,t,j}}, t \in \{1, \dots, l+u\}, j \in \{1, \dots, C\}$$

The labels for the training data should remain constant. Hence, the third step includes restoring these labels. For $t \in 1, \dots, l$ we set $Z_{k+1,t,j} = 1$ if $y_t = j$ and $Z_{k+1,t,j} = 0$ otherwise. The iteration stops when the class association probabilities converge. We refer to the final values of these probabilities as $z_{t,j}$.

PROPOSED METHOD

In this section, we present the proposed algorithm for recognizing human activities from labeled and unlabeled activity data. The approach consists of two phases. In the training phase we are provided with a training set which includes a smaller set of labeled data and a larger set of unlabeled data. In this phase, we build a neighborhood graph on both labeled and unlabeled inputs to estimate the manifold structure of the input data as explained in [13]. In this procedure, we consider each data point in the training set as a node and connect each node to its k -nearest neighbors in the input space and two nearest neighbors in the time space (i.e. its preceding and succeeding data). This way, we consider similarity in the feature space and sequence of data in the time space in constructing the graph. Next, we use label propagation on the constructed neighborhood graph to estimate the probability of association of each unlabeled data to each class of activity. After that, we train an HMM with hidden and observable states that represent activities and input features respectively. The probability of association of each unlabeled data to a special activity class is used to estimate the transition probabilities between the hidden states. In the test

phase, we calculate the probability distribution of test data in each of the states regarding the manifold structure. We utilize the HMM to classify test data and find the maximum likelihood sequence of hidden states for the test sequence.

Graph Construction and Label propagation

Several neighborhood graph construction methods are proposed to estimate the structure of the underlying manifold. In this paper, we employ a modification of the k -nearest neighbors (k -NN) method, one of the traditional methods for graph construction. In the proposed method, "feature similarity", defined as the Euclidean distance between input features, is used to connect input points to their k -nearest neighbors by undirected edges. In addition, as we expect conditional dependencies in sequences of data points, we connect each data point to its preceding and succeeding data in the sequence.

As mentioned above, the probability that the data point t belongs to the class j (i.e. $z_{t,j}$) can be calculated by label propagation on the constructed neighborhood graph. These probabilities will be used in the HMM.

training the HMM

As we aim to make use of the temporal sequence of the test data, we have to utilize a sequential probabilistic model. We selected HMM due to its well-developed and simple application. Each HMM has a set of hidden states $\{s_1, s_2, \dots, s_C\}$ each of which corresponds to one activity. Here, C is the number of activity classes in our problem.

To define the Markov model, the following probabilities have to be specified: transition probabilities between hidden states $a_{ij} = p(s_j | s_i)$, observation probabilities $b_i(y) = p(y | s_i)$, and the initial probabilities of hidden states $\pi_i = p(s_i)$.

In the training phase, we estimate the transition probabilities between hidden states and initial probabilities. The observation probabilities will be estimated in the test phase, as soon as the test data y arrives.

Traditionally, when we exactly know the hidden states corresponding to the training data, we can use the Baum-Welch algorithm to estimate the transition probabilities of the HMM. In such situation, maximum likelihood helps us to estimate the parameters. We set :

$$a_{ij} = \frac{\text{Expected number of transitions from state } s_i \text{ to } s_j}{\text{Expected number of transitions out of state } s_i}$$

In our case, the exact hidden states of the observations are not available. The proposed method utilizes the manifold structure of input data to calculate the probability of being in each hidden state. This way, we generalize the Baum-Welch algorithm to estimate the transition probabilities.

Let $z_{t,j} = p(s_j | x_t)$ denote the probability of association of data x_t to activity class j for all $1 \leq t \leq T$ and $1 \leq j \leq C$ where T is the number of training data. We have already calculated $z_{t,j}$ in the label propagation step for the training

data. Hence we get:

$$a_{ij} = \frac{\sum_{t=1}^T z_{t,i} z_{t+1,j}}{z_{t,i}}$$

To estimating the initial probabilities, we can use our prior knowledge about the problem or simply use $p(s_i) = \frac{1}{C}$ if we lack such an information.

In the next subsection, we will explain how to calculate observation probabilities for each test data (i.e. $p(y|s_i)$). Moreover, we will show how to convert the classification problem in activity recognition into the decoding problem in HMM.

Classifying Test Data

We use the Bayes rule to calculate the observation probability of test data y

$$p(y|s_i) \approx \frac{p(s_i|y)}{p(s_i)}$$

The value of $p(s_i)$ is computed in the previous section. To estimate $p(s_i|y)$ we benefit from our knowledge about the structure of the underlying manifold of training data to improve the classification performance. We estimate $p(s_i|y)$ in a manner similar to the label propagation used in the training phase. To this goal, we add the arriving test data to the graph and connect it to k of its nearest neighbors. Finally, $p(s_i|y)$ is estimated by performing one iteration of label propagation with the initial values equal to $z_{t,j}$ (calculated in the training phase) for every t and j .

Now that we have calculated all the parameters of HMM, we can utilize the Viterbi algorithm to find the maximum likelihood sequence of hidden states for the test sequence and classify the test data.

EXPERIMENTAL RESULTS

To evaluate the proposed method and compare the results with other activity recognition algorithms, we used the public dataset TU Darmstadt [9]. This dataset includes 164 hours of data recordings in 7 different days of a person's daily life in non-laboratory conditions over a sixteen day period. The subject wore two wearable accelerate sensors, that saved the mean and variance of accelerations within 50% overlapping, 30-second-long intervals, along with the time of the day. One of the sensors was placed at the right hip, and the other was on the right wrist.

The TU Darmstadt dataset includes a large set of activities: having lunch, having dinner, picking up cafeteria food, sitting, having a coffee, washing dishes, washing hands, personal hygiene, using the toilet, brushing teeth, standing, using the toilet, walking, walking freely, walking while carrying something, driving car, driving bike, sitting, desk activities, discussing at whiteboard, lying while reading, using computer, standing, talking on phone, queuing in line, and unlabeled.

To compare our work with the results reported in [13], we used the same set of features (mean and variance of acceleration In 30-second time intervals, along with its time stamp).

Like [13] we evaluated our algorithm for different sampling rates of every 10, 30, 60, 120 and 180 minuets . In constructing the features similarity graph , we set the number of nearest neighbors to 10.

The validity of manifold assumption in this dataset is evaluated by obtaining the labeling accuracy in the training phase. To do so, we assigned each unlabeled data (in the training set) to the class with the highest probability. The accuracy of this labeling is shown in Table 1

Table 1. Experimental Results: The accuracy of label propagation on the unlabeled training data

Time interval	Accuracy
10 min	92.6
30 min	90.2
60 min	87.5
90 min	84.9
120 min	81.8

To test the proposed algorithm, we conducted all experiments in a leave-one-day-out manner and performed 7 experiments. We compared the results of the proposed method with the algorithm in [13]. The results are presented in Table 2. The first column of this table shows the results for [13] and the second column presents our results.

Table 2. Experimental Results: The accuracy of proposed algorithm on the test data

Time interval	Accuracy of SVM+graph based	Accuracy of HMM+graph based
10 min	77.8	79.6
30 min	74.3	77.9
60 min	73.1	76.1
90 min	72.8	75.9
120 min	70.6	74.2

This table shows that the idea proposed in this paper results in a significant improvement in the accuracy of classification.

CONCLUSION

In this paper, we introduced a novel approach to human activity recognition. The problem was divided into training and test phases. In the training phase, we periodically obtained labels of the data in constant time intervals. Due to the availability of cheap unlabeled data, we utilized a semi-supervised graph-based method to benefit from the underlying structure of the data. We used label propagation on a k -nearest neighbor graph to calculate the probability of association of the unlabeled data to each class in the training phase. These probabilities were used to learn the transition probabilities of an HMM in order to predict the classes of the test data. In the test phase, we had to predict the classes of the test data in an online manner. To achieve this goal, we used one iteration of label propagation on the graph that was constructed in the training phase to obtain a primary estimate of the probabilities of association of the test data

to each class. Then, by using these probabilities, we utilized the trained HMM to predict the class of the test data. Experimental results showed that the proposed method significantly outperforms the state of the art semi-supervised method presented in [13].

ACKNOWLEDGEMENTS

We would like to thank anonymous reviewers for their helpful comments. This work was supported by National Elite Foundation of Iran.

REFERENCES

1. Stikic, M. and Schiele, B. Activity recognition from sparsely labeled data using multi-instance learning. In *International Symposium on Location and Context Awareness*, pages 156–173. Springer, 2009.
2. Bennewitz, M. , Burgard, W. , Cielniak, G. , and Thrun, S. Learning motion patterns of people for compliant robot motion. *International Journal of Robotics Research (IJRR)*, 10(1):31, 2005.
3. Brumitt, B, Meyers, B. , Krumm, J. , Kern, A. , and Shafer, S. Easyliving: Technologies for intelligent environments. In *Handheld and Ubiquitous Computing*, pages 97–119. Springer, 2000.
4. Bui, H.H. , Phung, D.Q. and Venkatesh, S. Hierarchical hidden Markov models with general state hierarchy. In *Proceedings of the National Conference on Artificial Intelligence*, pages 324–329. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2004.
5. Dee, H. and Hogg, D. On the feasibility of using a cognitive model to filter surveillance data. In *IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 34–39. IEEE, 2005.
6. Duong, T.V. , Bui, H.H. , Phung, D.Q. and Venkatesh, S. Activity recognition and abnormality detection with the switching hidden semi-markov model. In *Computer Vision and Pattern Recognition*. IEEE Computer Society, 2005.
7. Guan, D., Yuan, W., Lee, Y.K., Gavrilov, A. and Lee, S. Activity recognition based on semi-supervised learning. 2007.
8. Huynh, T. , Blanke, U. and Schiele, B. Scalable recognition of daily activities with wearable sensors. *Location-and context-awareness*, pages 50–67, 2007.
9. Huynh, T. , Fritz, M. and Schiele, B. Discovery of activity patterns using topic models. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 10–19. ACM, 2008.
10. Logan, B. , Healey, J. , Philipose, M. , Tapia, E.M. and Intille, S. A long-term evaluation of sensing modalities for activity recognition. In *Proceedings of the 9th international conference on Ubiquitous computing*, pages 483–500. Springer-Verlag, 2007.
11. Patterson, D. , Etzioni, O. , Fox, D. and Kautz H. Intelligent ubiquitous computing to support Alzheimers patients: Enabling the cognitively disabled. In *First International Workshop on Ubiquitous Computing for Cognitive Aids*, page 2. Citeseer, 2002.
12. Ravi, N. , Dandekar, N. , Mysore, P. and Littman, M.L. Activity recognition from accelerometer data. In *Proceedings of the National Conference on Artificial Intelligence*, volume 20, page 1541. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.
13. Stikic, M. , Larlus, D. and Schiele, B. Multi-graph Based Semi-supervised Learning for Activity Recognition. In *International Symposium on Wearable Computers*, pages 85–92. IEEE, 2009.
14. Stikic, M. , Van Laerhoven, K. and Schiele, B. Exploring semi-supervised and active learning for activity recognition. 2008.
15. Tapia, E.M. , Intille, S.S. and Larson, K. Activity recognition in the home using simple and ubiquitous sensors. *Pervasive Computing*, pages 158–175, 2004.
16. Vail, D.L. , Veloso, M.M. and Lafferty, J.D. Conditional random fields for activity recognition. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, pages 1–8. ACM, 2007.
17. Wang, L. and Suter, D. Learning and matching of dynamic shape manifolds for human action recognition. *Image Processing, IEEE Transactions on*, 16(6):1646–1661, 2007.
18. Wilson, D.H. and Atkeson, C. Simultaneous tracking and activity recognition (STAR) using many anonymous, binary sensors. *Pervasive computing*, pages 62–79, 2005.