Abnormal Activity Recognition Based on HDP-HMM Models

Derek Hao Hu[a], Xian-Xing Zhang[b], Jie Yin[c], Vincent Wenchen Zheng[a] and Qiang Yang[a]

[a] Department of Computer Science and Engineering, Hong Kong University of Science and Technology

{derekhh, vincentz, qyang}@cse.ust.hk

[b] State Key Laboratory for Novel Software Technology, Nanjing University, China

flyaway2009@gmail.com

[c] Information Engineering Laboratory, CSIRO ICT Centre, Australia

jie.yin@csiro.au

Abstract

Detecting abnormal activities from sensor readings is an important research problem in activity recognition. A number of different algorithms have been proposed in the past to tackle this problem. Many of the previous state-based approaches suffer from the problem of failing to decide the appropriate number of states, which are difficult to find through a trial-and-error approach, in real-world applications. In this paper, we propose an accurate and flexible framework for abnormal activity recognition from sensor readings that involves less human tuning of model parameters. Our approach first applies a Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM), which supports an infinite number of states, to automatically find an appropriate number of states. We incorporate a Fisher Kernel into the One-Class Support Vector Machine (OCSVM) model to filter out the activities that are likely to be normal. Finally, we derive an abnormal activity model from the normal activity models to reduce false positive rate in an unsupervised manner. Our main contribution is that our proposed HDP-HMM models can decide the appropriate number of states automatically, and that by incorporating a Fisher Kernel into the OCSVM model, we can combine the advantages from generative model and discriminative model. We demonstrate the effectiveness of our approach by using several real-world datasets to test our algorithm's performance.

1 Introduction

In recent years, activity recognition has been drawing growing interests from both artificial intelligence and pervasive computing researchers. Activity recognition aims to recognize the states and goals of one or more agents, given the observations of the agents' actions in some forms of input and probably the environmental conditions. Such a problem has important practical values and the research on activity recognition has witnessed a growing amount of research interest in past years. In the real world, activity recognition can be used in a variety of applications, including security monitoring to detect acts of terrorism [Jarvis *et al.*, 2004], where terrorist activities are defined as abnormal activities, and helping patients with cognitive disabilities [Pollack *et al.*, 2003].

In this paper, instead of considering how to perform accurate activity recognition, we consider the problem of detecting *abnormal activities*, where we follow the definition used in [Yin *et al.*, 2008] and define "abnormal activities" as "activities that occur rarely and have not been expected in advance". Such a problem may first appear to be very similar with the original activity recognition problem in principle. However, the problem of abnormal activity recognition is much harder than the original problem since such abnormal activities, by definition, rarely occur. This difficulty might become more significant during training phase since we lack such labeled sequences of abnormal activities. Up to now, most activity recognition algorithms [Lester *et al.*, 2005] are systems based on state space-based machine learning models, which require a significant amount of training data in order to perform accurate and successful parameter estimation. Nevertheless, in abnormal activity recognition, such requirements often cannot be satisfied. Most previous research tried to tackle the abnormal activity recognition problem by also using state-space models [Yin *et al.*, 2008], like Hidden Markov Models (HMMs) or Dynamic Bayesian Networks (DBNs).

There exists one serious problem with these state-space based models, especially HMMs, where one needs to define an appropriate number of states. Usually such a number is determined through a trial-and-error process. In practice, such a number is usually difficult to be known beforehand and the recognition accuracy is usually sensitive to the number of states chosen. However, in real-world applications, it is impossible to undergo this trial-and-error process when the recognizer is attached to humans already and when there is not enough data to validate the accuracy of the model under a particular number of states. Therefore, this drawback can become a major hurdle in real-world activity recognition systems.

In this paper, we aim to solve the *abnormal activity recognition problem* via a three-phased approach. We first apply the Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM), which has an infinite number of states and can automatically decide the optimal number of states. Then, we incorporate a Fisher kernel into our model and apply a One-Class Support Vector Machine (OCSVM) to filter out the nor-

mal activities. Finally, we derive our abnormal activity model in an unsupervised manner. In this paper, two additional contributions, besides providing an effective and efficient algorithm for solving the *abnormal activity recognition* problem, are: (1) we provide an approach to automatically decide the optimal number of states in state-based methods; (2) combine the power of generative model (HDP-HMM) and discriminative power (OCSVM with Fisher Kernel). We demonstrated the effectiveness of our algorithm through extensive experiments.

One of our previous works [Zhang *et al.*, 2009] also aims to detect the abnormal events from video sequences using Hierarchical Dirichlet Processes. Our work differs from this previous work in several aspects. Firstly, our previous work, which detects abnormal events in video sequences, relies heavily upon feature vectors that we extract from video sequences. Such features normally contain more "representative knowledge" compared to sensor-based activity recognition, where sensor-readings can have both continuous and discrete attributes and understanding the role different sensor readings play in the feature vector is not direct. Secondly, instead of using an ensemble learning algorithm to extract the candidate abnormal events, which might be more heuristic and difficult to explain in principle, in this paper, we incorporate the Fisher Kernel into a One-Class Support Vector Machine model to bring benefit from both generative learning and discriminative learning. Thirdly, in this paper we perform more extensive experiments of our algorithm with different parameters and compare it to different baselines to show each of our components to be useful in our final abnormal activity recognition system.

The rest of the paper is organized as follows: In Section 2, we review some previous work related to our abnormal activity recognition problem. In Section 3, we describe our three-phase approach in detail. In Section 4, we present our experimental results using two real-world datasets compared to state-of-the-art abnormal activity recognition algorithms. Finally, we conclude our paper and discuss some possible future works in Section 5.

2 Related Work

There is much important previous research work done in trying to tackle the problem of abnormal activity recognition. Due to space constraints, we only review a few related papers which are most relevant to our approach.

With the recent development of sensor networks, activity recognition from sensor data becomes more and more attractive. Many real-world applications require accurate recognition results [Pollack *et al.*, 2003; Geib *et al.*, 2008]. So far, among the state-of-the-art learning-based activity recognition algorithms, state-space based models are quite representative. State-space based models usually treat the activities and goals as hidden states, and try to infer such hidden states from the low-level sensor readings by statistical learning. For example, [Bui, 2003] employed an Abstract Hidden Markov Memory Model to represent the probabilistic plans, and used an approximate inference method to uncover the plans. [Vail *et al.*, 2007] and [Liao *et al.*, 2007] focused on using Conditional

Random Fields and its variants to model the activity recognition problem. However, these algorithms are centered on the recognition of a set of predefined *normal activities*.

Previous approaches on abnormality detection problem range from the computer vision area [Duong *et al.*, 2005; Zhang *et al.*, 2009] to data mining areas of outlier detection [Lazarevic *et al.*, 2003]. Besides our previous work [Zhang *et al.*, 2009], the most relevant work to our approach is [Yin *et al.*, 2008], which also aims to detect a user's abnormal activities from body-worn sensors. We will describe their algorithm detail in brief since we will be using their algorithm as a baseline. They propose a two-phase abnormality detection algorithm where a One-Class SVM is built on normal activities which help filter out most of the normal activities. The suspicious traces are then passed on to a collection of abnormal activity models adapted via KNLR (Kernel Nonlinear Logistic Regression) for further detection. However, before training the One-Class SVM, they need to transform the training traces that are of variable lengths into a set of fixed-length feature vectors. To accomplish this task, they trained M HMMs where M is the number of normal activities, and then the likelihood between each sensor reading and normal activity is used as the feature vector. One major drawback of this model is we need to specify the state number of HMMs when training, and such a number will affect the overall algorithm performance a lot as we will show in our experiment section. Thus, their algorithm may not be easy to use in real-world situations since it is hard for users to tune this parameter easily.

3 Background and Our Proposed Approach

3.1 Overview

We first present an overview of our three-phase approach for abnormal activity recognition from sensor readings. In the first step, we extract the significant features from normal traces, where these features are then used to train an HDP-HMM-based classifier in a sequential manner. The classifier can then be used to decide on a suitable model for every feature automatically. In the second step, we learn a decision boundary around the normal data in the feature space and then use the boundary to classify activities as normal or abnormal via One-Class SVMs. We intentionally train the One-Class SVMs so that they can identify normal activities with a higher likelihood, under the assumption that everything else is abnormal with a lower likelihood. When choosing a threshold value for the general model, we tend to reduce the false positive rate. In the third phase, we perform model adaption to adapt the abnormal activity model to a new model, which gives each abnormal activity a "second chance" to be classified as normal activities[Zhang *et al.*, 2005].

In the remainder of this section, we first briefly review HDP and its Gibbs Sampling methods. We then describe how we incorporate HDP-HMM with OCSVM model. Finally, we describe how we build suitable model adaptation techniques.

3.2 HDP-HMM

Hierarchical Dirichlet Process hidden Markov Model

Consider groups of data, denoted as $\{\{y_{ij}\}_{i=1}^{n_j}\}_{j=1}^{J}$, where n_j denotes the number of data in group j, J denotes the total number of groups thought to be produced by related, yet unique, generative processes. Each group of data is modeled via a mixture model. A Dirichlet Process (DP) representation may be used separately for each of the data group. In an HDP, the base distribution of each of the DPs are drawn from a DP, which is discrete with probability 1, so each of the DPs can share the statistical strength, for instance, this encourages appropriate sharing of information between the data sets. An HDP formulation can decide the right number of states for the Hidden Markov Model (HMM) from its posterior density function on the appropriate number of mixture components, to some extent, the number of states in HMM can go to infinite if necessary. Besides, it learns the appropriate degree of sharing of data across data sets through the sharing of mixture components.

The HDP can be built as follows (Due to space constraint, we will omit the detailed explanation of HDP in this paper, interested readers please refer to [Teh *et al.*, 2006] for technical details.):

$$G_0(\theta) = \sum_{k=1}^{\infty} \beta_k \delta(\theta - \theta_k)$$

$$\beta \sim GEM(\gamma) \qquad \theta_k \sim H(\lambda) \qquad k = 1, 2, \ldots$$

$$G_j(\theta) = \sum_{t=1}^{\infty} \bar{\pi}_{jt} \delta(\theta - \tilde{\theta}_{jt})$$

$$\tilde{\pi}_j \sim GEM(\alpha) \qquad j = 1, \ldots, J \tilde{\theta}_{jt} \sim G_0 \qquad t = 1, 2, \ldots$$
$$\bar{\theta}_{ji} \sim G_j \qquad y_{ji} \sim F(\bar{\theta}_{ji})$$
$$j = 1, \ldots, J, i = 1, \ldots, N_j.$$

where $GEM(\cdot)$ stands for the stick-breaking process as follows:

$$\beta'_k \sim Beta(1, \gamma)$$

$$\beta_k = \beta'_k \prod_{l=1}^{k-1}(1 - \beta'_l), \qquad k = 1, 2, \ldots$$

HMM can be viewed as a doubly stochastic Markov chain and is essentially a dynamic variant of a finite mixture model. Therefore, by replacing the finite mixture with a Dirichlet process, we can complete the design of HDP-HMM (See Figure 1 for a graphical representation.)

To better illustrate the construction of HDP-HMM, we introduce another equivalent representation of the generative model using indicator random variables:

$$\beta \sim GEM(\gamma) \qquad \pi_j \sim DP(\alpha, \beta)$$

$$z_{ji} \sim Mult(\pi_j) \qquad \theta_k \sim H(\lambda) \qquad y_{ji} \sim F(\theta z_{ji})$$

Identifying each $G(k)$ as describing both the transition probabilities $\pi_{kk'}$ from state k to k' and the emission distributions parameterized by $\phi_{k'}$, we can now formally define the HDP-HMM as follows:

$$\beta \sim GEM(\gamma), \qquad \pi_k \sim DP(\alpha, \beta), \qquad \phi_k \sim H, \quad (1)$$
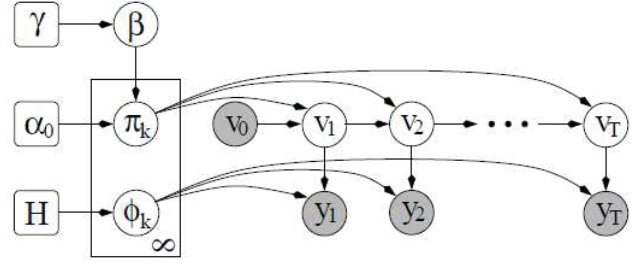$$s_t \sim Mult(\pi_{s_{t-1}}), \qquad y_t \sim F(\phi_{s_t}) \qquad (2)$$

Figure 1: A graphical representation of the HDP-HMM Model. [Teh et al. 2006]

The Gibbs Sampler

The Gibbs sampler was the first MCMC algorithm for the HDP-HMM that converges to the true posterior. [Teh *et al.*, 2006] proposed three sampling schemes, one of them that is heuristic to HDP-HMM builds on the direct assignment sampling scheme for the HDP, by marginalizing out the hidden variables π, ϕ from Equations 1 and 2 and ignoring the ordering of states implicit in β. Thus we only need to sample the hidden trajectory s, the base DP parameters β and the hyperparameters α, γ, for this sampler, a set of auxiliary variables m_{jk} is needed, we denote m_{jk} as the number of transitions from state i to state j, and $m_{j\cdot}, m_{\cdot j}$ denote the transitions out and in of state j, the sampling schemes are listed below:

Sampling β: According to [Teh *et al.*, 2006], the desired posterior distribution of β is:

$$p((\beta_1, \ldots, \beta_K, \beta_{\bar{k}}) \boldsymbol{t}, \boldsymbol{k}, y_{1:T}, \gamma) \propto Dir(m_{\cdot 1}, \ldots, m_{\cdot K}, \gamma).$$

Sampling s_t: We now determine the posterior distribution of s_t:

$$p(s_t = k \mid s_{\backslash t}, y_{1:T}, \beta, \alpha, \lambda) \propto$$

$$p(s_t = k \mid s_{\backslash t}, \beta, \alpha) p(y_t \mid y_{\backslash t}, s_t = k, s_{\backslash t}, \lambda)$$

According to the property of Dirichlet processes, we have

$$p(s_t = k \mid s_{\backslash t}, \beta, \alpha) \propto$$

$$\begin{cases} (\alpha\beta_k + m_{s_{t-1}k}^{-t})\left(\frac{\alpha\beta_k + n_{s_{t-1}k}^{-t} + \delta(s_{t-1},k)\delta(k,s_{t+1})}{\alpha + n_{k\cdot}^{-t} + \delta(s_{t-1},k)}\right) & k \in 1, \ldots, K \\ \alpha\beta_{\bar{k}}\beta_{s_{t+1}} & k = \tilde{k}. \end{cases}$$

The conditional distribution of the observation y_t given an assignment $s_t = k$ and given all other observations y_τ, having marginalized out θ_k, is derived as follows:

$$p(y_t \mid y_{\backslash t}, s_t = k, s_{\backslash t}, \lambda) \propto$$

$$\int_{\theta_k} p(y_t \mid \theta_k) p(\theta_k \mid \{y_\tau \mid s_\tau = k, \tau \neq t\}, \lambda) d\theta_k$$

Sampling m_{jk}:

$$p(m_{jk} = m \mid n_{jk}, \beta, \alpha) = \frac{\Gamma(\alpha\beta_k)}{\Gamma(\alpha\beta_k + n_{jk})} s(n_{jk}, m)(\alpha\beta_k)^m$$

where $s(n, m)$ are unsigned Stirling numbers of the first kind.

3.3 Building One-Class SVM with Fisher kernel

Similar to [Yin *et al.*, 2008], we applied the One-Class SVMs to learn a decision boundary around the normal data in the feature space and then use the boundary to classify activities as normal or abnormal. However, in [Yin *et al.*, 2008] the author used Gaussian Radial Basis Function (RBF) kernel for the One-Class SVM, but we choose the Fisher Kernel to more effectively combine the strength from both the generative model (HDP-HMM) and the discriminative model (One-Class SVM). Such a combination is usually expected to obtain a robust classifier which has the strengths of each approach.

Fisher kernel

Fisher kernel is introduced in [Jaakkola and Haussler, 1998]. A kernel that is capable of mapping variable length sequences to fixed length vectors enables the use of discriminative classifiers for variable length examples. Fisher Kernel combines the advantages of generative statistical models (in our framework HDP-HMM) and those of discriminative methods (in our framework One-Class SVMs), where HDP-HMM can process data of variable length and automatically select the suitable model, while One-Class SVMs can have flexible criteria and yield better results. The gradient space of the generative model is used for this purpose since the gradient of the log likelihood with respect to a parameter of the model describes how that parameter contributes to the process of generating a particular example.

The Fisher Score is defined as the gradient of the log likelihood with respect to the parameters of the model:

$$U_X = \nabla_\theta \log P(X|\theta)$$

The Fisher kernel is defined as:

$$K(X_i, X_j) = U_{X_i}^T I^{-1} U_{X_j}$$

where I is the Fisher information matrix [Jaakkola and Haussler, 1998]and U_X is the Fisher score. In [Jaakkola and Haussler, 1998], the Fisher information matrix is proposed for normalization, while we also can use other measures to accomplish this task.

One-Class SVM Training

According to [Yin *et al.*, 2008], we first need to convert the training traces with variable lengths into a set of fixed length feature vectors, here we adopt a set of HDP-HMMs as described in the above section to model the normal traces, one for each type of M features, using beam sampling methods. And the feature vectors in our framework are just the log-likelihood value for each of the N normal traces computed as follows:

$$L_j(Y_i) = \log P_j(Y_i), 1 \le i \le N, 1 \le j \le M$$

where $\log P_j(Y_i)$ is the log-likelihood of the ith activities trained from the HDP-HMMs based on the jth feature. In this way, for each training trace Y_i, we can obtain an M-dimensional feature vector $X_i = \{L_1(Y_i), \cdots, L_M(Y_i)\}$ for One-Class SVM:

$$\max \sum_{i=1}^{n} \alpha_i K(X_i, X_i) - \sum_{i,j=1}^{n} \alpha_i \alpha_j K(X_i, X_j).$$

where $K(X_i, X_j)$ is the Fisher kernel described above.

As described in [Yin *et al.*, 2008], a major limitation of using a One-Class SVM for abnormality detection is the difficulty in selecting a sensitivity level that is sufficiently high to yield a low false negative rate and a low false positive rate. To deal with this problem, we also fit our One-Class SVM by selecting parameters so that it is biased toward a low false negative rate. That is, our One-Class SVM can identify, with high confidence, that a portion of data is normal. The rest of the data that are deemed suspicious are passed on to the third phase for further detection. Thus, our One-Class SVM acts as a filter to a classifier by singling out the normal data without creating a model for abnormal characteristics.

3.4 Model adaptation

In [Yin *et al.*, 2008], the abnormal events are derived from a general normal model in an unsupervised manner. The benefit of such an unsupervised manner is that this framework can address the unbalanced label problem due to the scarcity of training data and the difficulty in pre-defining abnormal activities. More specifically, after the second step we may get a high false negative rate, i.e., we may have many normal activities be incorrectly classified as abnormal activities, so it's necessary for us to apply a third phase, that is, to adapt models for the abnormal events, and use these abnormal classifiers to reduce the false negative rate. Besides, due to the lack of negative training data, we cannot directly build models for abnormal events. However, we can use adaptation techniques to get them during the test time or even in future use, that is, we can dynamically build the model for the abnormal event after the training phase. Here we briefly introduce the algorithm's framework first. The steps are listed as below:

Prerequisites: A well defined general HDP-HMM with Gaussian observation density trained by all normal training sequences.

Step 0 : Use the first outlier detected from the former phase - which is considered to be able to represent a particular type of abnormal activities - to train an abnormal event model by adaptation using beam sampler.

Step 1 : Slice the test sequence into fixed length segments, calculate the likelihood of these segments by the existing normal activity models, if the maximum likelihood is given by the general model, we predict this trace to be of a normal activity, then goto Step 4. Else goto Step 2;

Step 2 : If the maximum likelihood is larger than the threshold, we consider this trace to belong to an existing abnormal model; then we predict this trace to be possible abnormal events, go to Step 4, else go to Step 3;

Step 3 : Use adaptation methods to adapt the general model to a new abnormal activity model, then add this adapted abnormal model to the set of models and go to Step 4, here this outlier is regarded to represent one kind of the certain events.

Step 4 : Go to Step 1 if new outlier comes.

In this procedure, we provide the outlier with a second chance to be recognized as a normal event, so that normal events that tend to be unexpected or scarce in the training data

are not misclassified. Thanks to the effectiveness of beam sampler again, we can do the adaptation effectively without other special design. Suppose that we have the new parameters for the HDP-HMM λ, here we update the HDP parameters $\beta, \alpha_0, \gamma, K$ and HMM parameters π, μ. Notice that in Step 1, an abnormal activity sequence may be predicted as normal activities again, thereby decreases the false negative rate in this Step. And in Step 2, we classify such an abnormal activity sequence to one abnormal activity in the current activity set we are now holding. There may still be cases where we have not seen this abnormal activity before, and we perform Step 3 so that we can create a new abnormal activity set, and humans can be involved to analyze what this abnormal activity sequence actually means in real life. Such a framework is useful for real-world deployment of our abnormal activity recognition algorithm.

4 Experiments

In this section, we will study the effectiveness of our algorithm by validating it through several real-world abnormal activity recognition datasets and our algorithm is compared to the baseline algorithm described in [Yin *et al.*, 2008].

4.1 Datasets, Metrics and Baselines

We use two real-world activity recognition datasets. The first is the MIT PLIA 1 dataset [Intille *et al.*, 2006], which was recorded on Friday March 4, 2005 from 9AM to 1PM with a volunteer in the MIT PlaceLab. The dataset contains 89 different activities and was manually classified into several categories including: Cleaning, Yardwork, Laundry, Dishwashing, Meal Preparation, Hygiene, Grooming, Personal and Information/Leisure. Due to the fact that "abnormal activities" are usually hard to define and previous work including [Yin *et al.*, 2008] and [Zhang *et al.*, 2005] often manually defined some activities with low probabilities as abnormal activities, we manually selected some activities with low probabilities and consider such activities as abnormal activities we aim to detect from sensor readings. And the second dataset we are using, referred to as *Yin* in Table 2 is from [Yin *et al.*, 2008], where a number of traces of a user's normal daily activities in an indoor environment are recorded. In this dataset, the user was asked to simulate the effect of carrying out several abnormal activities.

The evaluation metric that we are using in this paper is the AUC (Area Under Curve) measurement [Bradley, 1997], since a good abnormal activity recognition system should have both high detection rate (defined as the ratio of the number of correctly detected abnormal activities to the total number of abnormal activities) and low false alarm rate (defined as the ratio of the number of normal activities that are incorrectly detected as normal activities to the total number of normal activities). The ROC curve plots the detection rate against the false alarm rate and therefore becomes our choice in such a problem.

The algorithms we plan to analyze in this paper are as follows: **HMM + RBF + KNLR**, which is the algorithm discussed in [Yin *et al.*, 2008]'s paper, **HDP + Fisher + Adaptation**, which is our proposed method by using HDP and

Support Vector Machine with Fisher Kernel, together with the model adaptation method we proposed, **HDP + RBF + KNLR**, which is exactly the original baselines except that we use a HDP-HMM in the first phase to automatically determine the optimal number of states in HMM. **HDP + RBF + Adaptation**, same as our algorithm but we use a traditional RBF kernel to train the OCSVM model. We design these baseline methods to demonstrate the effectiveness of our framework, and also show that our two main contributions, (1) using HDP-HMM to optimally decide the optimal number of states and (2) incorporating Fisher Kernel into the OCSVM model, are both effective in this problem.

4.2 Experimental Results

In this section we present our experimental results in Table 1. The AUC score of each algorithm is calculated and the training set is drawn at random ten times to calculate a variance of the AUC score. For the baseline methods, since the number of states in the HMM model Q needs to be manually defined, we tested the algorithm performance with varying numbers of Q from 2 to 8.

Algorithm	PLIA1 AUC (Variance)
HMM + RBF + KNLR (Q = 2)	0.683(0.025)
HMM + RBF + KNLR (Q = 3)	0.764(0.027)
HMM + RBF + KNLR (Q = 4)	0.793(0.025)
HMM + RBF + KNLR (Q = 5)	0.721(0.018)
HMM + RBF + KNLR (Q = 6)	0.657(0.030)
HMM + RBF + KNLR (Q = 7)	0.642(0.019)
HMM + RBF + KNLR (Q = 8)	0.631(0.016)
HDP + RBF + KNLR	0.811(0.032)
HDP + RBF + Adaptation	0.835(0.017)
HDP + Fisher + Adaptation	**0.857(0.028)**

Table 1: Performance Comparison of our algorithm and the baseline methods on the MIT PLIA Dataset

Algorithm	Yin's AUC (Variance)
HMM + RBF + KNLR (Q = 2)	0.713 (0.028)
HMM + RBF + KNLR (Q = 3)	0.725 (0.021)
HMM + RBF + KNLR (Q = 4)	0.748 (0.010)
HMM + RBF + KNLR (Q = 5)	0.785 (0.015)
HMM + RBF + KNLR (Q = 6)	0.732 (0.017)
HMM + RBF + KNLR (Q = 7)	0.718 (0.013)
HMM + RBF + KNLR (Q = 8)	0.707 (0.019)
HDP + RBF + KNLR	0.792 (0.018)
HDP + RBF + Adaptation	0.813 (0.021)
HDP + Fisher + Adaptation	**0.834((0.029)**

Table 2: Performance Comparison of our algorithm and the baseline methods on the dataset from [Yin *et al.*,2008]

From Table 1 and Table 2, we can see that our framework *HDP + Fisher + Adaption* outperforms the baseline algorithm and some other baselines that we have set. When we set Q from 2 to 8, we can see that the AUC score varies between 0.683 and 0.793 in PLIA1 dataset, and the AUC

score varies between 0.713 and 0.785 in [Yin *et al.*, 2008]'s dataset, which clearly indicates the difficulty of choosing an appropriate number of states and the impact of a non-optimal state on the final recognition accuracy cannot be neglected. When using *HDBP + RBF + KNLR*, we notice that its performance already outperforms that of HMM-based models. Therefore, adopting HDP-HMM in our model can automatically determine the appropriate number of states and algorithm performance will not be affected since we avoid a step of trial-and-error process. We can also see that using *HDP + RBF + Adaptation* is not as good as our proposed method which uses Fisher kernels on the two datasets we've tested, which suggests that our proposed approach for incorporating Fisher kernel into this framework will have stronger predictive strengths compared to incorporating the commonly-used RBF Kernels.

Therefore, in this section, by reporting the performance of our algorithm on two activity recognition datasets and by comparing the performance of our algorithm with the baseline algorithms, we have demonstrated empirically that our framework is useful at each step, and that introducing HDP and Fisher Kernel can improve the overall performance.

5 Conclusion and Future Work

In this paper, we have presented a novel framework for tackling the problem of abnormal activity recognition. Our method does not suffer the problem of hard to determine an optimal number of states as previous state-based approaches do. We applied an HDP-HMM model that can automatically select the suitable model with the optimal number of states. We analyzed the efficiency and effectiveness of introducing beam sampling in the HDP-HMM model. We also combined the powers of both generative models and discriminative models by using the Fisher Kernel in the One-Class SVM model in the second step. Finally, we described a model adaptation approach so that we can detect unseen abnormal activities. In the future, we wish to explore some effective online inference algorithms for us to tackle the abnormal activity recognition problem in a more natural way to meet the need of real-world applications.

Acknowledgment

References

[Bradley, 1997] Andrew P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997.

[Bui, 2003] Hung Hai Bui. A general model for online probabilistic plan recognition. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI 2003)*, pages 1309–1318, 2003.

[Duong *et al.*, 2005] Thi V. Duong, Hung Hai Bui, Dinh Q. Phung, and Svetha Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-markov model. In *CVPR (1)*, pages 838–845, 2005.

[Geib *et al.*, 2008] Christopher W. Geib, John Maraist, and Robert P. Goldman. A new probabilistic plan recognition algorithm based on string rewriting. In *ICAPS*, pages 91–98, 2008.

[Intille *et al.*, 2006] Stephen S. Intille, Kent Larson, Emmanuel Munguia Tapia, Jennifer Beaudin, Pallavi Kaushik, Jason Nawyn, and Randy Rockinson. Using a live-in laboratory for ubiquitous computing research. In *Proceedings of the Fourth International Conference on Pervasive Computing (Pervasive 2006)*, pages 349–365, 2006.

[Jaakkola and Haussler, 1998] Tommi Jaakkola and David Haussler. Exploiting generative models in discriminative classifiers. In *NIPS*, pages 487–493, 1998.

[Jarvis *et al.*, 2004] Peter Jarvis, Teresa F. Lunt, and Karen L. Myers. Identifying terrorist activity with ai plan recognition technology. In *AAAI*, pages 858–863, 2004.

[Lazarevic *et al.*, 2003] Aleksandar Lazarevic, Levent Ertöz, Vipin Kumar, Aysel Ozgur, and Jaideep Srivastava. A comparative study of anomaly detection schemes in network intrusion detection. In *SDM*, 2003.

[Lester *et al.*, 2005] Jonathan Lester, Tanzeem Choudhury, Nicky Kern, Gaetano Borriello, and Blake Hannaford. A hybrid discriminative/generative approach for modeling human activities. In *IJCAI*, pages 766–772, 2005.

[Liao *et al.*, 2007] Lin Liao, Dieter Fox, and Henry A. Kautz. Extracting places and activities from gps traces using hierarchical conditional random fields. *International Journal of Robotics Research (IJRR)*, 26(1):119–134, 2007.

[Pollack *et al.*, 2003] Martha E. Pollack, Laura E. Brown, Dirk Colbry, Colleen E. McCarthy, Cheryl Orosz, Bart Peintner, Sailesh Ramakrishnan, and Ioannis Tsamardinos. Autominder: an intelligent cognitive orthotic system for people with memory impairment. *Robotics and Autonomous Systems (RAS)*, 44(3-4):273–282, 2003.

[Teh *et al.*, 2006] Yee Whye Teh, Michael I. Jordan, Matthew J. Beal, and David M. Blei. Hierarchical dirichlet processes. *Journal of The American Statistical Association*, 101(476):1566–1581, 2006.

[Vail *et al.*, 2007] Douglas L. Vail, Manuela M. Veloso, and John D. Lafferty. Conditional random fields for activity recognition. In *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2007)*, page 235, 2007.

[Yin *et al.*, 2008] Jie Yin, Qiang Yang, and Junfeng Pan. Sensor-based abnormal human-activity detection. *IEEE Trans. Knowl. Data Eng.*, 20(8):1082–1090, 2008.

[Zhang *et al.*, 2005] Dong Zhang, Daniel Gatica-Perez, Samy Bengio, and Iain McCowan. Semi-supervised adapted hmms for unusual event detection. In *CVPR (1)*, pages 611–618, 2005.

[Zhang *et al.*, 2009] Xianxing Zhang, Hua Liu, Yang Gao, and Derek Hao Hu. Detecting abnormal events from hierarchical dirichlet processes. In *PAKDD*, 2009.