

Sensor-based Abnormal Human-Activity Detection

Jie Yin, Qiang Yang, *Senior Member, IEEE* and Jeffrey Junfeng Pan

Abstract

With the availability of affordable sensors and sensor networks, sensor-based human activity recognition has attracted much attention in artificial intelligence and ubiquitous computing. In this paper, we present a novel two-phase approach for detecting abnormal activities based on wireless sensors attached to a human body. Detecting abnormal activities is a particular important task in security monitoring and healthcare applications of sensor networks, among many others. Traditional approaches to this problem suffer from a high false positive rate, particularly when the collected sensor data are biased towards normal data while the abnormal events are rare. Therefore, there is a lack of training data for many traditional data mining methods to be applied. To solve this problem, our approach first employs a one-class support vector machine (SVM) that is trained on commonly available normal activities, which filters out the activities that have a very high probability of being normal. We then derive abnormal activity models from a general normal model via a kernel nonlinear regression (KNLR) to reduce false positive rate in an unsupervised manner. We show that our approach provides a good tradeoff between abnormality detection rate and false alarm rate, and allows abnormal activity models to be automatically derived without the need to explicitly label the abnormal training data, which are scarce. We demonstrate the effectiveness of our approach using real data collected from a sensor network that is deployed in a realistic setting.

Keywords: Activity Recognition, Outlier Detection, Sensor Networks, Data Mining.

Jie Yin is with the Tasmanian Information Communication Technologies (ICT) Center, Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia. Email: Jie.Yin@csiro.au (<http://www.ict.csiro.au/staff/Jie.Yin/>)

Qiang Yang is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Email: qyang@cse.ust.hk (<http://www.cse.ust.hk/~qyang>)

Jeffrey Junfeng Pan is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Email: panjf@cse.ust.hk (<http://www.cse.ust.hk/~panjf>)

I. INTRODUCTION

With the increasingly accessible sensor technology, automatic activity recognition is becoming a reality. By attaching different types of sensors on various objects, locations and the human body, a user's activities can be tracked and monitored. An important application is security monitoring for identifying terrorist activities in high security areas [10], where each individual entering the security areas is issued a sensor-equipped security name-tag, through one or more sensors attached to the security tag, the activities of individuals can be continuously monitored to prevent acts of terrorism. When abnormal events are detected, the system will sound an alarm for immediate attention. Several methods have been proposed in the past, including rule-based approaches that have deep cognitive roots and template-based approaches that allow human to be tightly in the loop [31], [10]. Another potential application is helping patients who suffer from cognitive disorders. If a safe and smart house can be instrumented with a sensor network, the occupants would have a better chance to live safely and independently [22], especially when they suffer from life-threatening diseases such as Parkinson's Disease or Alzheimer's disease. Due to the importance and challenging nature of the problem, several approaches have been proposed to learn the daily activities of individuals based on sensor readings; examples include hidden Markov models (HMMs) [15] and dynamic Bayesian networks (DBNs) [16], [32]. These models can predict the type of a user's high-level activities based on collections of *normal* traces.

While detecting a user's normal activities is interesting and challenging, we consider the inverse problem – detecting abnormal activities of users from wearable sensors – to be just as challenging and interesting, if not more so. The abnormality detection problem is interesting because it has wide-ranging applications from security monitoring to healthcare for the elderly people. In a security monitoring application, suppose that our task is to monitor the movements of personnel in a high security area such as a government office building. One approach is to issue a sensor-equipped security name-tag to each person who enter the area. Then, abnormal activity detection algorithms can be applied to continuously track the movements of each individual to check whether the person's activities are out of the norm. In such cases, it would be beneficial to detect if a rare and unexpected event has occurred based on sensor readings, and to sound an alarm if necessary. Of particular interest to such applications are the objectives that the false positive rates, as well as false negative rates, must be kept at a minimum, in order for

the security warnings to be trustworthy. Abnormal event detection is also of great importance to other application domains. For example, consider a healthcare application, where instead of paying attention to normal activities, a doctor or nurse might be more interested in an elderly or a patient who show the first signs of a serious illness by wandering around aimlessly out of the normal patterns, or by simply falling down.

The problem is computationally challenging because, unlike normal activities, in abnormal activity detection, the data are extremely scarce. In a security application, for example, a monitoring system may be trained to recognize only the normal activities because of the availability of such data, while the abnormal activities are new to the system. This may happen, for example, after an activity profile is created, a user being monitored may change the patterns of activities without notifying the learning system. In this case, the abnormal activity detection system may wrongly consider these activities as abnormal and raise alarms. Therefore, it is a challenging task to design an abnormality detection system that has a low false positive rate. In this paper, we define abnormal activities as events with the following properties:

- 1) they occur rarely and
- 2) they have not been expected in advance.

In the past, several approaches have been proposed to tackle the abnormality detection problem. These approaches range from the computer vision area [28], [7], where variations of Markov models are designed to detect the out-of-norm behavior of users, to the data mining area [1], [13], where similarity or model based approaches are employed to detect outliers. However, in the sensor network area, there has been relatively few work on using wearable sensors to detect abnormal activities. As compared to vision data and computer network data, sensor data provide some unique advantages as well as challenges. For the advantages of sensors as compared to videos, sensors allow the movement of a person to be continuous monitored in an un-intrusive way. In contrast, videos can be sensitive to lighting and covering area, and images can introduce ambiguities when multiple objects appear in a field of view, in addition to being intrusive in privacy (for example, a patient may object to the use of videos in a home environment). However, sensors also introduce additional challenges to their deployment. For example, sensors are distributed in their computation, and they operate under such constraints as limited energy and communication channels. Depending on the type, a sensor may detect different types of signals,

including light, sound, temperature, acceleration and magnetic field. Because sensors are lightweight, they can be easily attached to a human body. In our work, the main computational challenges posed by sensors include uncertainty in the signals that are received by a sensor network, and the lack of negative examples (that is, the abnormal behaviors of a person) that can be used for training.

In this paper, we propose a novel approach to abnormal activity detection based on signals from wearable sensors. Our approach is inspired by the observation that, although it is unrealistic to obtain a large amount of training data for abnormal activities, it is conversely possible to do so for normal activities. This enables the creation of well-estimated models for normal activities, which can be adapted for abnormal ones at a later stage. Therefore, we propose a two-phase approach to abnormal activity detection, which builds on the data available for normal activities. In the first phase, we build a one-class SVM solely based on normal activities, which can filter out the activities having a very high probability of being normal. Then, the suspicious traces are passed on to a secondary phase for further detection. In the second phase, we perform a kernel nonlinear regression (KNLR) analysis to derive abnormal activity models from a general normal activity model in an unsupervised manner. A major advantage of our approach is that it can achieve a good tradeoff between the abnormal activity detection rate and false alarm rate without the need to explicitly collect and label abnormal traces. We collect data from wearable sensors attached to a user and demonstrate the effectiveness of our approach using these data in a real-world environment.

II. RELATED WORK

In this section, we review the previous works in three parts. In the first part, we review research on activity recognition in artificial intelligence. In the second part, we review different approaches to outlier detection in data mining and machine learning areas. In the third part, we review previous work on cost-sensitive learning, which is closely related to our adopted approach.

A. Activity Recognition

Activity recognition has been a recent focus in artificial intelligence and ubiquitous computing, partly due to the increasing availability of wireless sensors. One approach is to exploit easily

understandable rules that can be used to describe the normal behavior of human activities [31]. This approach is particularly appealing because it can provide a mechanism for capturing *abnormal* rules as well, through the use of exceptional rules. Knowledge representation based on rules and exception rules has strong cognitive roots and be easily understandable by domain experts. This approach is complementary to our probabilistic-model based approach in that our approach (to be presented in the next section) provides an effective method for online tracking of potentially critical security violations when these violations *first appear* among a large collection of normal behaviors, thus our approach is applicable when not much data have been accumulated to train the rules and exceptional rules. These approaches are complementary because our method can benefit from rule-based explanation features through the techniques proposed in [31].

Another approach is to apply template-based plan recognition to security monitoring applications. For example, in [10], a plan-template based approach is proposed for recognizing and ranking the potential patterns of events that lead to a likely terrorist attack. The approach is to first compile a set of typical patterns using logical frameworks in artificial intelligence planning, and then match these patterns from observed actions and objectives. This approach will only be highly successful when the patterns and templates are defined A priori. Our problem in this article is complementary to [10] in that we detect and learn these templates when they appear for the first several times.

In the past decade, various model-based approaches such as hidden Markov models (HMMs) and dynamic Bayesian networks (DBNs) have been proposed to model and recognize users' activities. Lester *et al.* proposed a hybrid discriminative/generative approach to recognizing human activities in [15], in which useful features were first extracted to build an ensemble of static classifiers and HMMs were trained thereafter to recognize different activities. In [19], body-worn microphones and accelerometers were used to track users' daily activities. Most of these works employ supervised learning to recognize users' normal activities, which requires a large amount of labeled training data. Due to the lack of abnormal training data, these methods cannot be directly applied to abnormal activity detection.

Several approaches in the vision community have been proposed to detect unusual events in video sequences. The work of [28] adopted a dynamic Bayesian network to model each type of normal video patterns. An activity is identified as abnormal if the likelihood of being generated by normal models is less than a threshold. This approach is attractive due to its

simplicity; however, an appropriate threshold is hard to determine in practice. In [7], switching hidden semi-Markov models were applied to represent user's activities and perform abnormality detection. However, this work only focused on detecting a more subtle form of abnormality, which is the abnormalities only in the state duration, but not in the state order.

The Wireless local area network (Wireless LAN) provides an ideal ubiquitous and cheap network in which one can typically receive the radio signal strength (RSS) information from several beacon nodes that are called access points (AP). These information can be used to form a vector whose value is a function of the locations and time of the moving agent. Likewise, other context information can also be obtained by attaching to the agent more sensors. One important aspect of location computation is how to save the battery power while maintaining a high level of accuracy, a problem addressed in [5] for location estimation using access points in a Wireless LAN. When mapping the RSS values to the locations, it is sometimes possible to take advantage of the inherent correlation between the signal space and the location space, as is suggested in [21].

B. Outlier Detection

Similar problems are solved in the data mining and machine learning areas, where the focus has been on detecting outliers from a set of static data points. In general, these approaches can be classified into two categories: *similarity-based approaches* [1] and *model-based approaches* [13], [20]. For similarity-based approaches, the main task is to define pairwise distances between all the data points and identify outliers by examining the distance to an example's nearest neighbors. An example is the work of [3], who applied density-based clustering algorithm to efficiently detect local outliers. Based on the distance measure and user-defined density thresholds, these algorithms can efficiently detect the occurrence of outliers (or abnormal points) in a high-dimensional space. The basic principle is that if the neighboring points are relatively close, the example is considered normal; otherwise, the example is considered abnormal. The advantage of these approaches is that no explicit distribution needs to be defined to determine outliers and that the methods can be made efficiently for large datasets. However, a difficulty in these approaches lies in the question of how to define effective similarity measures when a large amount of uncertainty exists. For example, in the sensor-network area, the signals received from various sensory readings vary greatly from time to time, following a stochastic nature. Thus, it

would be very difficult to define a distance measure that is sufficiently robust in these settings, making it difficult also to define density measures. Another difficulty lies in the requirement in our problem that the algorithm must be online; that is, efficient models need to be trained ahead of time in order to *efficiently* detect abnormal events as they occur. Therefore, in summary, when the data do not provide clear-cut shapes, and when the data are stochastic in nature, as in the case of sensor signals, the similarity-based and distance based approaches cannot work well.

Model-based approaches in data mining usually characterize the normal data using predictive models. These models are used to detect outliers as deviations from the learned model. Among these models, one-class support-vector machines (SVMs) [24], [25] have been demonstrated empirically to be capable of detecting outliers in various domains. One-class SVMs fit a tight hyper-sphere boundary in the nonlinearly transformed feature space to include most of normal data, and then use the boundary to detect abnormal data. For example, the work of [20] first extracted useful features from time series data and then trained a one-class SVM for outlier detection. However, a significant drawback in using one-class SVMs is the difficulty in selecting a proper sensitivity parameter to achieve a good tradeoff between false negative rates and false positive rates.

C. Cost-Sensitive Learning and Unbalanced Data Problems

Our problem is clearly cost-sensitive in nature, because typically to classify an abnormal activity as a normal one is a much more serious problem than classifying a few normal activities as abnormal ones. Furthermore, our data are inherently unbalanced due to a much larger proportion of normal activity data. Therefore, our work is also closely related to cost-sensitive learning and learning for unbalanced data in data mining and machine learning.

Cost-sensitive learning addresses the issue of classification in the presence of different misclassification costs that are associated with the different types of errors[8], [26], and are useful for learning problems where the classes are extremely unbalanced. Problems having different misclassification costs are abundant in the real world, ranging from direct marketing, medical diagnosis, to intrusion detection. For example, in direct marketing, a task is to distinguish customers who respond positively to a marketing action from those who do not. This problem can be modeled as a two-class classification problem, where the positive class corresponds to the responsive customers. In this problem domain, the sizes of the positive data are usually tiny in

proportion as compared to that of the negative class. In addition, to classify a positive customer as a negative one will likely cost much more than the other way around. In such a case, a cost-sensitive learning method can be applied. In cases where unbalanced data are present, some researchers [2], [18] have argued for the use of ranking-based evaluation metrics such as area under the ROC curve (ROC) instead of using accuracy. Finally, Yang et al. [29], [17] presented an approach to integrate cost-sensitive learning with missing value handling, where additional test costs exist when acquiring the missing values for future data.

The unbalanced data problem is often solved using cost-sensitive learning techniques, by setting the false positive and false negative costs very differently, and incorporate the cost factors in a risk formula [8], [6]. Research in cost-sensitive learning falls into three main categories. The first category focuses on making particular classifier learners cost-sensitive, including methods specific for decision trees [11], neural networks [12] and support vector machines [9]. The second category designs wrappers for any classification algorithms, by applying the Bayesian risk theory and assigning to each example its lowest cost class [6]. The classifiers are required to output class membership probabilities, which is convenient for incorporating risk factors and ranking of the results. The third category consists of learning methods that modify the distribution of training examples before applying the classifier-learning algorithms, so that the classifiers learned from the modified distribution are cost-sensitive [4]. As a whole, the central objective of cost-sensitive learning is to balance out the false positive and false negative costs, in the two-class situations, by taking the total cost into consideration.

The abnormal-activity detection problem that we consider corresponds to the case of class-imbalanced data, where normal activities that correspond to the positive class are extremely small in proportion as compared to the normal activities that corresponds to the negative data. The training data are often grossly biased towards the negative classes. As a result, an accuracy based evaluation method would favor the larger-sized class. To solve the problem, we can exploit cost-sensitive learning algorithms, but we also observe that the false-positive and false-negative costs are usually unknown to us in our problems. Therefore, in our solution which we describe in the sequel, we will exploit a novel one-class classification method for abnormal-activity detection. In addition, we will use the ROC (Receiver Operating Characteristics) curve to evaluate the performance of algorithms in our experiments instead of using the accuracy measures.

III. OUR PROPOSED ALGORITHM

In this section, we present our approach to abnormal activity detection. Intuitively, activity-based abnormality detection techniques typically compare a profile of all normal activities to actual observations. Any deviation from the profile is flagged as a potential alarm. This positive activity model identifies normal activities and assumes that everything else is abnormal. Therefore, abnormal activity detection has the potential to detect even the unknown activities that have rarely happened. Based on this intuition, our proposed approach follows a two-phase pipelined process, which combines a one-class SVM for filtering out the normal activities while reducing false negative rate with a collection of secondary classifiers for abnormal activity detection which is aimed at reducing false positive rate.

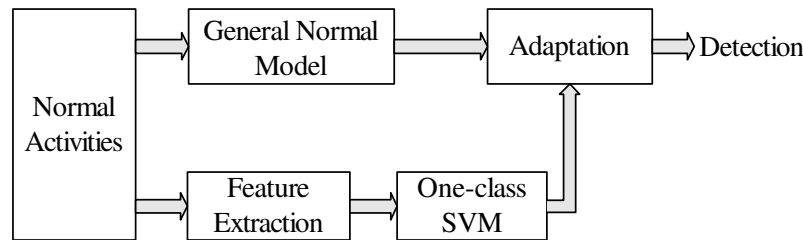


Fig. 1. Flow diagram of our proposed algorithm

Figure 1 shows the flow diagram of our proposed algorithm. Given a collection of normal traces $\{Y_1, Y_2, \dots, Y_N\}$, our proposed algorithm works in two phases. In the first phase, all the normal traces are directed via a module to compute significant features, which are then passed through a one-class SVM to eliminate the traces that have a high probability of being normal. The rest, which are suspicious traces, are then passed on to a secondary phase for further detection, in which abnormal activity models are derived from a general normal model in an unsupervised manner. These activity models serve as classifiers to further determine whether the input traces are normal or abnormal. In the following, we discuss the two major phases in detail.

A. Building One-Class SVM Classifier

Our objective is to automatically detect abnormal activities where we only have a collection of normal traces for training. Therefore, a direct approach is 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. For this purpose, we employ the one-class support vector machines (SVMs) [25].

1) *Feature Extraction*: To train a one-class SVM, we first need to transform the training traces which are of variable lengths into a set of fixed-length feature vectors. Our observation is that, since such traces are generated by a hidden mechanism associated with users' underlying activities, it is desirable to model such data using a generative model. In this paper, we adopt a set of HMMs with Gaussian observation density to model the normal traces.

Specifically, suppose that the training traces consist of M different types of normal activities. In order to capture temporal features involved in the traces, we train M Q -state HMMs with Gaussian observation density, one for each type of normal activities, using the Baum-Welch algorithm [23]. For each learned model with the corresponding parameters λ_i , $1 \leq i \leq M$, we calculate the log-likelihood of each of the N normal traces given the model parameters λ_i . The log-likelihood value for each pair consisting of trace and HMM is computed as follows:

$$L(Y_i; \lambda_j) = \log P(Y_i | \lambda_j), 1 \leq i \leq N, 1 \leq j \leq M. \quad (1)$$

This is calculated by applying the standard forward-backward algorithm [23]. In this way, for each training trace Y_i , $1 \leq i \leq N$, we can obtain an M -dimensional feature vector $\mathbf{x}_i = \langle L(Y_i; \lambda_1), \dots, L(Y_i; \lambda_M) \rangle$.

2) *One-class SVM Training*: After transforming the n training traces into a set of feature vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$, we train a one-class SVM based upon the normal data. The basic idea is to find a *sphere* that contains most of the normal data, such that the corresponding radius R can be minimized:

$$\begin{aligned} \min \quad & R^2 + C \sum_{i=1}^n \xi_i, \\ \text{s.t.} \quad & \|\mathbf{c} - \mathbf{x}_i\|^2 \leq R^2 + \xi_i, \\ & \xi_i \geq 0. \end{aligned} \quad (2)$$

Here, the slack variables ξ_i are introduced to allow some data points to lie outside the sphere, and the parameter $C \geq 0$ controls the tradeoff between the volume of the sphere and the number of errors. Using the dual representation of the Lagrangian, the objective function is equivalent

to

$$\begin{aligned} \max \quad & \sum_{i=1}^n \alpha_i \langle \mathbf{x}_i \cdot \mathbf{x}_i \rangle - \sum_{i,j=1}^n \alpha_i \alpha_j \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^n \alpha_i = 1. \end{aligned} \quad (3)$$

This quadratic programming (QP) problem can be solved using standard optimization techniques [24]. To determine if a test data is within the sphere, the distance to the center of the sphere has to be calculated. If this distance is smaller than the radius R , the test trace is considered normal.

Typically, the training data are not spherically distributed in the input space. Therefore, the original data points are first mapped into a feature space so that a better data description can be obtained. Instead of requiring an explicit mapping function from the input space to the feature space, the solution can be obtained by simply replacing all the inner products $\langle \cdot, \cdot \rangle$ in Equation (3) by a kernel function $k(\cdot, \cdot)$:

$$\max \quad \sum_{i=1}^n \alpha_i k(\mathbf{x}_i, \mathbf{x}_i) - \sum_{i,j=1}^n \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j). \quad (4)$$

In our context, due to the nonlinear and noisy characteristics of sensor readings, the decision boundary of the one-class SVM classifier might be quite complex. Hence, we use a Gaussian Radial Basis Function (RBF) kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-w_1^2 \|\mathbf{x}_i - \mathbf{x}_j\|^2)$ for the one-class SVM. Here w_1 is a scaling factor that gives the width of the kernel function.

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. Figure 2 illustrates two decision boundaries of a one-class SVM built on data points with two features. In the figure, a wide decision boundary, indicated by the dashed curve, would result in too many false negatives, while a narrow decision boundary, indicated by the solid curve, would result in too many false positives.

To deal with this problem, we fit our one-class SVM by selecting parameters so that it is biased towards a low false negative rate. That is, our one-class SVM can identify, with high confidence, that a portion of data are normal. The rest of the data that are deemed suspicious are passed on to the second phase for further detection. Thus, our one-class SVM acts as a filter to a

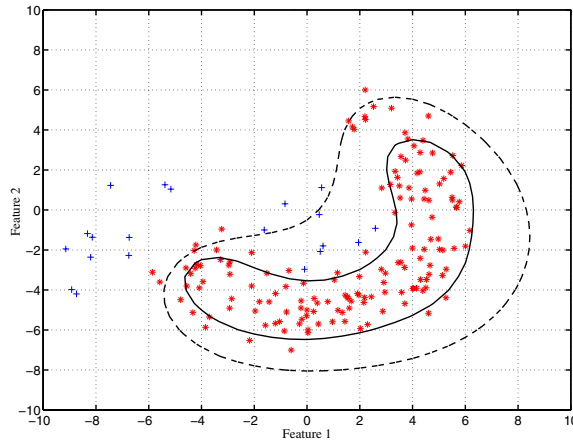


Fig. 2. Illustration about one-class SVMs

classifier by singling out the normal data without creating a model for abnormal characteristics.

B. Deriving Abnormal Activity Models

Taking the results of the one-class SVM as an input, the second phase of our proposed approach is aimed at deriving abnormal activity models from a general normal activity model. These models are used for abnormality detection.

1) *Iterative Adaptation Procedure*: We create models for abnormal activities in an iterative procedure. As shown in Figure 3(a), we start by having only a general model for normal activities. Specifically, we learn a general HMM with Gaussian observation density using all the normal traces. Given the well-estimated general normal model and a test trace, we first calculate the likelihood of this trace being generated by the general model. If the likelihood is less than a pre-specified threshold θ , we identify this trace as an outlier. The outlier is considered to be able to represent a particular type of abnormal activities, and thus it can be used to train an abnormal activity model. However, only one single outlier is clearly insufficient to generate a good estimate of the model parameters for an abnormal activity model. Therefore, we perform a kernel nonlinear regression (KNLR) analysis to adapt the general model to a particular abnormal activity model using the detected outlier (see Figure 3(b)). Later when another test trace arrives, we calculate the maximum likelihood of generating this trace by the existing models. If the maximum likelihood is given by the general model, we predict this trace to be of a normal

activity; otherwise, we identify it to be abnormal. In the latter case, we have to decide whether a new abnormal activity model is to be generated. If the maximum likelihood is more than the threshold θ , we consider this trace to belong to an existing abnormal model; otherwise, this trace is considered to be a new type of abnormal activity. Thus, we derive a new abnormal activity model from the general normal model (see Figure 3(c)).

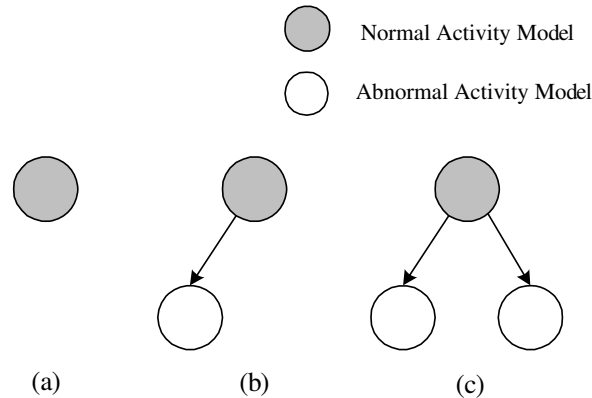


Fig. 3. Iterative adaptation procedure

We illustrate this iterative procedure in Figure 3. Initially, there is only one node in the tree, representing the general normal model. When an abnormal activity is detected, a new leaf node is split from the upper parent node, generating an abnormal activity model. When another abnormal trace is detected, if it can be well represented by one of the existing abnormal models, the tree structure remains the same; otherwise, a new abnormal activity model is derived from the parent node. This tree structure is modified in an online manner, allowing all the models to be created efficiently.

In the following, we discuss the model adaptation technique in detail.

2) *Kernel Nonlinear Regression (KNLR) Adaptation*: In the field of speech recognition, several adaptation techniques have been proposed to generate speaker-dependent models from a speaker-independent model using very limited adaptation data. Among these, maximum likelihood linear regression (MLLR) [14] attempts to compute a set of transformations that can reduce the mismatch between an initial model and the adaptation data. Specifically, MLLR is a model adaptation technique that estimates linear transformations for the mean and/or variance parameters of an HMM with Gaussian observation density.

In our case, we choose to adapt the mean vectors of the general HMM model. Let $\boldsymbol{\mu}_i^{old}$, $1 \leq i \leq Q$, denote the mean vector of the i th state in the HMM. The adaptation is performed in two steps. First, given the new data, new estimates of the mean vector $\boldsymbol{\mu}_i^{new}$ are computed based on the general model. Second, the vector mean $\boldsymbol{\mu}_i$ is adapted using the following equation:

$$\boldsymbol{\mu}_i = \alpha \cdot \boldsymbol{\mu}_i^{old} + (1 - \alpha) \cdot \boldsymbol{\mu}_i^{new}, \quad (5)$$

where α is a weighting factor that controls the balance between the old model and new estimates. The smaller the value of α , the more contribution the new data makes to the adapted model.

A major limitation of MLLR is that it can only perform linear transformations between the general model and the adaptation data. Therefore, we perform the adaptation using a nonlinear generalization of MLLR, which is called kernel nonlinear regression (KNLR) [27]. The basic idea of KNLR is to map linear regression transformations to a high-dimensional feature space via a nonlinear kernel map. Let $\mathbf{A} = [\boldsymbol{\mu}_1^{old}, \dots, \boldsymbol{\mu}_Q^{old}]$ and $\mathbf{B} = [\boldsymbol{\mu}_1^{new}, \dots, \boldsymbol{\mu}_Q^{new}]$ denote the mean vectors corresponding to the old model and the new model, respectively. The mean vector $\boldsymbol{\mu}_i^*$ using KNLR can be computed as follows:

$$\boldsymbol{\mu}_i^* = (\mathbf{BK} + \beta\mathbf{AK}^{-1})(\mathbf{K}^2 + \epsilon\mathbf{I})^{-1}\mathbf{K}. \quad (6)$$

In this equation, similar to α in MLLR, β is a weighting factor which balances the old model and new estimates. \mathbf{I} is the identity matrix and ϵ is a user-defined regularization parameter. The matrix \mathbf{K} is a $Q \times Q$ kernel matrix:

$$\mathbf{K} = \begin{bmatrix} k(\boldsymbol{\mu}_1^{old}, \boldsymbol{\mu}_1^{old}) & \dots & k(\boldsymbol{\mu}_1^{old}, \boldsymbol{\mu}_Q^{old}) \\ \vdots & \vdots & \vdots \\ k(\boldsymbol{\mu}_Q^{old}, \boldsymbol{\mu}_1^{old}) & \dots & k(\boldsymbol{\mu}_Q^{old}, \boldsymbol{\mu}_Q^{old}) \end{bmatrix}, \quad (7)$$

where $k(\cdot, \cdot)$ is the kernel function. Here, in order to capture the nonlinear transformation between the general model and the adaptation data, we also use the RBF kernel $k(\boldsymbol{\mu}_i, \boldsymbol{\mu}_j) = \exp(-w_2^2 \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2)$ for model adaptation. By solving Equation (6), we can obtain a globally optimal solution for the final mean vectors $\boldsymbol{\mu}_i^*$, where $1 \leq i \leq Q$.

IV. EXPERIMENTAL EVALUATION

In order to evaluate the performance of our proposed algorithm, experiments were carried out on a real data set collected from a sensor network. For comparison, two other approaches were used as baselines. The first one is referred to as OneSVM which just builds a one-class SVM for abnormality detection. The second one first builds a one-class SVM and then uses the MLLR technique for adaptation, which is called SVM+MLLR. Our proposed algorithm is referred to as SVM+KNLR in the experiments.

A. Experimental Setup

We used three off-the-shelf Crossbow MICA2s (MPR400) with the sensor board MTS310CA for our experiments. Each sensor board consists of five different types of sensors including light, temperature, microphone, 2-axis accelerometer and 2-axis magnetometer. We programmed all the MICA2s so that the seven sensor readings can be collected at a sampling rate of 10Hz. The sensor readings are packaged to a sink at 915MHz radio frequency. After that, the data are sent to the computer via a serial port.

While at the moment we have to attach these sensors to different parts of a human body, we believe that with the rapid advancement in hardware sensor-network research, these sensors will become ever smaller in size and pervasive in our environment. Therefore, eventually they can be built into a single security card to be carried by people in a high security area, in order to facilitate security monitoring operations.

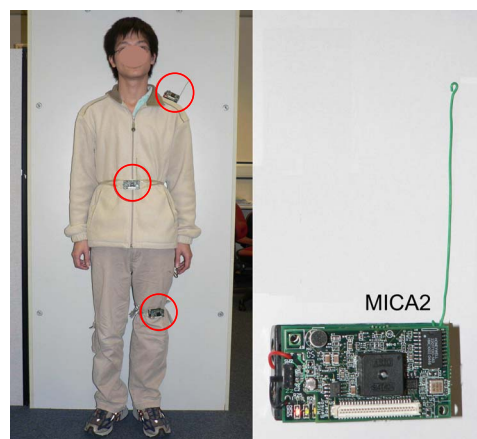


Fig. 4. The sensor type and placement

For the task of activity detection, we attached three sensors to a user’s body on positions that are circled in Figure 4. We collected 431 traces of a user’s normal daily activities in an indoor environment. To collect abnormal traces, the user was asked to simulate the effect of carrying out several abnormal activities, such as “slipping on a wet floor” and “falling down to a floor”. The total number of abnormal traces is 112. Table I lists several examples of normal and abnormal activities below.

Normal Activities	Abnormal Activities
sitting down	slipping on the ground
walking	falling down backwards
walking downstairs	falling down forwards
walking upstairs	...
running	...

TABLE I
 EXAMPLES OF NORMAL AND ABNORMAL ACTIVITIES

Each of these traces is a 7-dimensional time series. We collected these traces across multiple days in order to capture the day-to-day variations in the activities and sensor readings.

B. Evaluation Metrics

The performance of the abnormality detection algorithms can be evaluated in terms of two error rates: *detection rate* and *false alarm rate*. The detection rate is computed as the ratio of the number of correctly detected abnormal activities to the total number of abnormal activities. The false alarm rate is computed as the ratio of the number of normal activities that are incorrectly detected as abnormal activities to the total number of normal activities.

Based on the confusion matrix shown in Table II, the two metrics can be defined as follows:

$$\text{Detection Rate} = \frac{TN}{TN + FN}, \tag{8}$$

$$\text{False Alarm Rate} = \frac{FP}{FP + TP}. \tag{9}$$

The performance of an ideal abnormality detection algorithm should have a high detection rate and a low false alarm rate. Therefore, in the following, we evaluate the performance of the

		Actual Label	
		Normal	Abnormal
Predicted Label	Normal	True Positive (TP)	False Negative (FN)
	Abnormal	False Positive (FP)	True Negative (TN)

TABLE II
 CONFUSION MATRIX

algorithms using the ROC curve which plots the detection rate against the false alarm rate. In addition, we explicitly compute the area under the ROC curve (AUC) [2], [18] to compare the three algorithms. A desirable algorithm with a high detection rate and a low false alarm rate should have an AUC value closer to one.

C. Experimental Results

In our experiments, since the training traces are 7-dimensional time series with continuous values, we trained the general model and each normal activity model as two-state HMMs with Gaussian observation density. For SVM+MLLR and SVM+KNLR, the parameter C of one-class SVM was set to 0.05.

Experiments were first conducted to compare the performance of the three algorithms. We randomly selected a half subset of normal traces (216 traces) for training. The other 215 normal traces and all the 112 abnormal traces were randomly mixed together for testing. Figure 5 shows the ROC curve with respect to detection rate and false alarm rate. We can see from the figure that, OneSVM gives the poorest detection result because it achieves a high detection rate at the cost of incurring a high false alarm rate. By applying model adaptation techniques, SVM+KNLR and SVM+MLLR can improve the performance of OneSVM. This is because that the adapted abnormal activity models contributes to distinguishing normal and abnormal activities, which can significantly reduce false alarm rate. We can also observe that, by using nonlinear transformation for adaptation, SVM+KNLR consistently outperforms SVM+MLLR. For example, using the RBF kernel with $w_2 = 0.01$ and setting β to 10, SVM+KNLR can achieve a high detection rate of

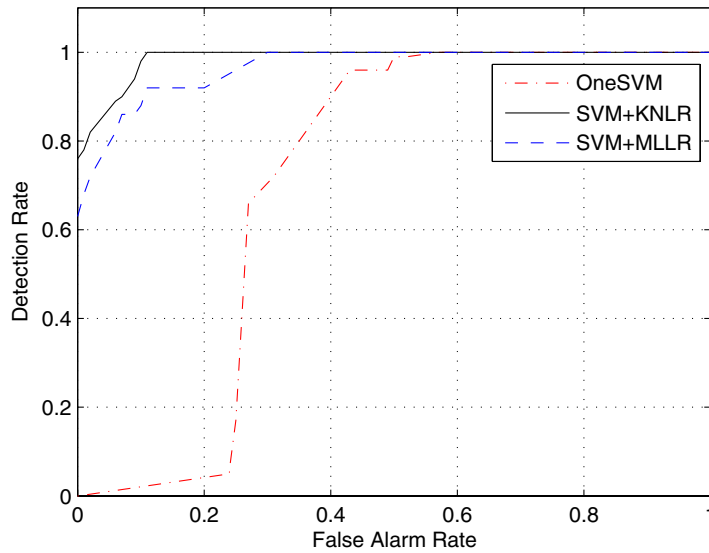


Fig. 5. Comparison of detection rate and false alarm rate vs. different numbers of training data: training on 216 normal traces

90% and a low false alarm rate of 7%. In this setting, the number of abnormal activity models generated by SVM+KNLR is 12. Therefore, we can conclude that, among the three algorithms, our proposed algorithm can achieve the best tradeoff between detection rate and false alarm rate.

We also performed experiments to investigate the effect of varying the number of training traces on the performance of the three algorithms. In this experiment, we kept the number of test data unchanged and reduced the number of training data. Figure 6 and 7 show the experimental results using 108 and 54 normal traces for training respectively.

We can see from the figure that, when the number of normal traces for training decreases, the performance of the three algorithms decreases accordingly. For OneSVM, when the training data are sparse, the calculated decision boundary may not exactly capture the data characteristics, which degrades the performance of abnormality detection. In addition, for SVM+MLLR and SVM+KNLR, sparse training data prevents us from having an accurate estimate of model parameters for the general normal activity model, which further results in an inaccurate abnormal activity model. Therefore, the two algorithms have degraded abilities in distinguishing normal and abnormal activities. As shown in Figure 7, when there are only 54 normal traces for training, the performance of SVM+KNLR and SVM+MLLR is comparable to that of OneSVM. Again,

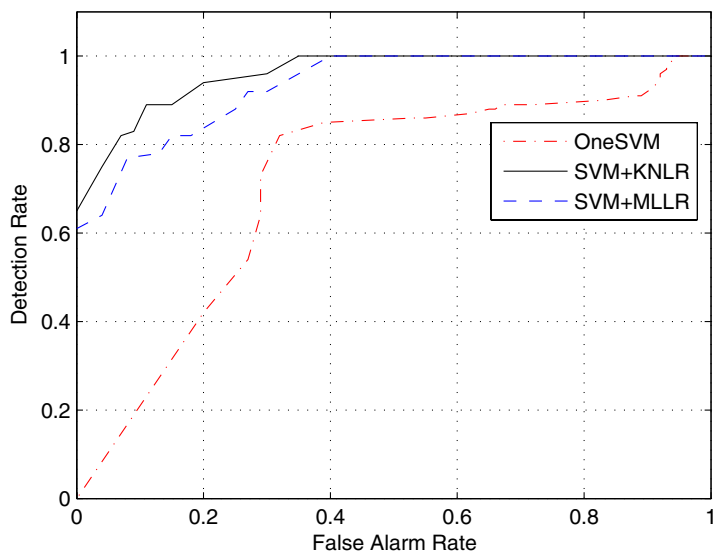


Fig. 6. Comparison of detection rate and false alarm rate vs. different numbers of training data: training on 108 normal traces

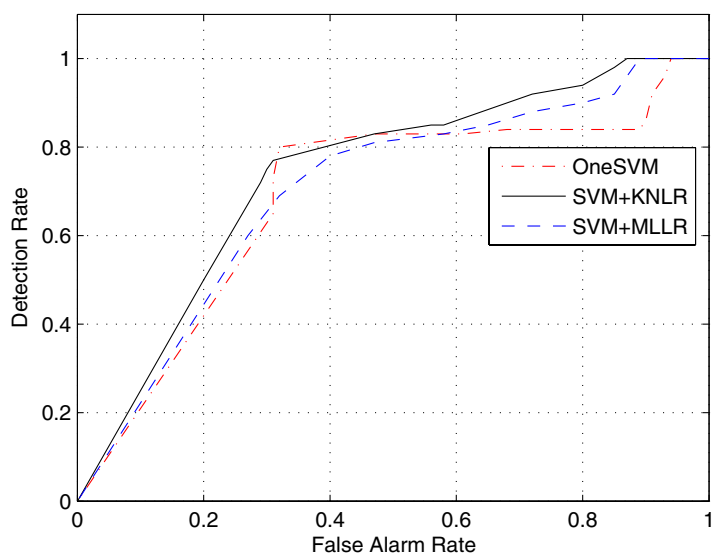


Fig. 7. Comparison of detection rate and false alarm rate vs. different numbers of training data: training on 54 normal traces

as shown in Figure 6, when we have 108 normal traces for training, SVM+KNLR still performs the best among the three algorithms.

Algorithms / # Training Traces	216 traces	108 traces	54 traces
OneSVM	0.685	0.720	0.689
SVM+MLLR	0.960	0.934	0.706
SVM+KNLR	0.985	0.961	0.739

TABLE III
 AUC VALUES WITH DIFFERENT ALGORITHMS AND DIFFERENT NUMBERS OF TRAINING TRACES

To explicitly compare the performance of the three algorithms, we also computed the AUC values by calculating the area under the ROC curves depicted in Figures 5, 6 and 7. The results are summarized in Table III. We can see from the second column of the table that, when 216 normal traces are used for training, the AUC values for OneSVM, SVM+MLLR and SVM+KNLR are 0.685, 0.960 and 0.985, respectively. Because SVM+KNLR can achieve an AUC value closer to one, we can conclude that SVM+KNLR performs better than the other two algorithms. Also, SVM+KNLR can be seen to perform the best among the three algorithms when we consider 108 and 54 normal traces for training as well.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel approach for detecting a user’s abnormal activities from body worn sensors. To deal with the scarcity of training data for abnormal activities, we propose a two-phase abnormality detection algorithm. In the first phase, a one-class SVM is built on normal activities, which helps to filter out most of the normal activities. The suspicious traces are then passed on to a collection of abnormal activity models adapted via KNLR for further detection. A major advantage of our approach is that it can achieve a better tradeoff between detection rate and false alarm rate. We demonstrate the effectiveness of our approach using real data collected from sensors attached to a human body.

A potential limitation of our approach is that there is a risk of generating a large number of abnormal models when abnormal activities suddenly becomes the norm. This may happen when a user being monitored repeats a certain behavior repeatedly after a certain time point. To solve this problem, we need to be able to detect when an abnormal activity turns normal, and vice

versa. Another direction is in detecting abnormal activities from continuous traces. This requires us to design an efficient segmentation and recognition algorithm which can precisely partition the whole trace into consecutive segments and detect abnormal activities from these segments in an online manner.

In the future, we wish to continue in the direction of detecting abnormal activities from continuous user traces. This requires us to design an efficient segmentation and detection algorithm which can precisely partition the whole trace into consecutive segments and detect abnormal activities from these segments in an online manner. In addition, we will also investigate how to best integrate intelligent planning for courses of actions when a abnormal situation is discovered, using statistical methods such as the actionable data mining algorithms in [30].

ACKNOWLEDGMENT

The work was supported by a grant from Hong Kong RGC 621606 and a grant from the National Basic Research Program of China (973 Program) under Grant No.2006CB303000. We thank Professor Jue Wang of Chinese Academy of Sciences and anonymous reviewers for valuable comments on this topic.

REFERENCES

- [1] S. D. Bay and M. Schwabacher. Mining distance-based outliers in near linear time with randomization and a simple pruning rule. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 29–38, Washington, DC, USA, August 2003.
- [2] A. P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30:1145–1159, 1997.
- [3] M. M. Breunig, H. P. Kriegel, R. Ng, and J. Sander. Identifying density-based local outliers. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages 93–104, Dallas, TX, USA, May 2000.
- [4] P. Chan and S. Stolfo. Toward scalable learning with non-uniform class and cost distributions. In *In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD98)*, pages 164–168, New York City, NY, USA, August 1998.
- [5] Yiqiang Chen, Qiang Yang, Jie Yin, and Xiaoyong Chai. Power-efficient access-point selection for indoor location estimation. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 18(7):877–888, 2006.
- [6] P. Domingos. Metacost: A general method for making classifiers cost-sensitive. In *In Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 155–164, San Diego, CA, USA, August 1999.
- [7] T. Duong, H. Bui, D. Phung, and S. Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-Markov model. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 838–845, San Diego, CA, USA, June 2005.

- [8] C. Elkan. The foundations of cost-sensitive learning. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 973–978, Seattle, WA, USA, August 2001.
- [9] G. Fumera and F. Roli. Cost-sensitive learning in support vector machines. In *In Proceedings of the Workshop on Machine Learning, Methods and Applications, held in the context of the Eighth Meeting of the Italian Association of Artificial Intelligence (AI*IA)*, Siena, Italy, September 2002.
- [10] P. Jarvis, T. F. Lunt, and K. L. Myers. Identifying terrorist activity with AI plan recognition technology. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI)*, pages 858–863, San Jose, CA, USA, July 2004.
- [11] U. Knoll, G. Nakhaeizadeh, and B. Tausend. Cost-sensitive pruning of decision trees. In *In Proceedings of the Eighteenth European Conference on Machine Learning (ECML)*, pages 383–386, Catania, Italy, April 1994.
- [12] M. Kukar and I. Kononenko. Cost-sensitive learning with neural networks. In *In Proceedings of the thirteenth European Conference on Artificial Intelligence (ECAI)*, pages 445–449, Brighton, UK, August 1998.
- [13] A. Lazarevic, L. Ertöz, A. Ozgur, J. Srivastava, and V. Kumar. A comparative study of anomaly detection schemes in network intrusion detection. In *Proceedings of the Third SIAM International Conference on Data Mining (SDM)*, pages 23–34, San Francisco, CA, USA, May 2003.
- [14] C. Leggetter and P. Woodland. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Computer Speech and Language*, 9:171–185, April 1995.
- [15] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford. A hybrid discriminative/generative approach for modeling human activities. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 766–772, Edinburgh, Scotland, July–August 2005.
- [16] L. Liao, D. Fox, and H. Kautz. Learning and inferring transportation routines. In *Proceedings of the Nineteenth National Conference in Artificial Intelligence (AAAI)*, pages 348–353, San Jose, CA, USA, July 2004.
- [17] Charles X. Ling, Victor S. Sheng, and Qiang Yang. Test strategies for cost-sensitive decision trees. *IEEE Transactions on Knowledge and Data Engineering*, 18(8):1055–1067, 2006.
- [18] C.X. Ling, J. Huang, and H. Zhang. Auc: a statistically consistent and more discriminating measure than accuracy. In *In Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 519–526, Acapulco, Mexico, August 2003.
- [19] P. Lukowicz, J. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, and T. Starner. Recognizing workshop activity using body worn microphones and accelerometers. In *Proceedings of the Second International Conference on Pervasive Computing (Pervasive)*, pages 18–32, Vienna, Austria, April 2004.
- [20] J. Ma and S. Perkins. Time-series novelty detection using one-class support vector machines. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pages 1741–1745, Portland, OR, USA, July 2003.
- [21] Jeffrey Junfeng Pan, James T. Kwok, Qiang Yang, and Yiqiang Chen. Multidimensional vector regression for accurate and low-cost location estimation in pervasive computing. *IEEE Transactions on Knowledge and Data Engineering*, 18(9):1181–1193, 2006.
- [22] M. E. Pollack. Intelligent technology for an aging population: The use of ai to assist elders with cognitive impairment. *AI Magazine*, 26(2):9–24, 2005.
- [23] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [24] B. Schölkopf, J. Platt, J. Shawe-Taylor, and A. Smola. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, July 2001.

- [25] D.M.J. Tax and R.P.W. Duin. Support vector domain description. *Pattern Recognition Letters*, 20(1113):1191–1199, 1999.
- [26] K. M. Ting. A comparative study of cost-sensitive boosting algorithms. In *Proceedings of the Seventeenth International Conference on Machine Learning (ICML)*, pages 983–990, Standord, CA, USA, June–July 2000.
- [27] I. W. Tsang, J. T. Kwok, B. Mak, K. Zhang, and J.J. Pan. Fast speaker adaptation via maximum penalized likelihood kernel regression. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. To appear, Toulouse, France, May 2006.
- [28] T. Xiang and S. Gong. Video behaviour profiling and abnormality detection without manual labeling. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1238–1245, Bei Jing, China, October 2005.
- [29] Qiang Yang, Charles Ling, Xiaoyong Chai, and Rong Pan. Test-cost sensitive classification on data with missing values. *IEEE Transactions on Knowledge and Data Engineering*, 18(5):626–638, 2006.
- [30] Qiang Yang, Jie Yin, Charles Ling, and Rong Pan. Extracting actionable knowledge from decision trees. *IEEE Transactions on Knowledge and Data Engineering*, 19(1):43–56, 2007.
- [31] Y. Yao, F. Wang, J. Wang, and D. D. Zeng. Rule + exception strategies for security information analysis. *IEEE Intelligent Systems*, 20(5):52–57, 2005.
- [32] J. Yin, X. Chai, and Q. Yang. High-level goal recognition in a wireless LAN. In *Proceedings of the Nineteenth National Conference in Artificial Intelligence (AAAI)*, pages 578–584, San Jose, CA, USA, July 2004.



Jie Yin is currently a researcher at the Tasmanian ICT Centre, where she is working on data management in wireless sensor networks, with a focus on developing techniques to recognize event patterns and identify data outliers. She received her PHD degree from the Department of Computer Science and Engineering, Hong Kong of Science and Technology, Hong Kong, in 2006 and her B.E. degree in Computer Science Department from the Xi'an Jiaotong University in China in 2001. Her research interests include data mining and machine learning, data integration in sensor networks, decision support systems and pervasive

computing . Contact her at Jie.Yin@csiro.au; <http://www.ict.csiro.au/staff/Jie.Yin/>



Qiang Yang is a faculty member in the Hong Kong University of Science and Technologys Department of Computer Science and Engineering. He is also an adjunct professor at Peking University, Beijing, and at Zhongshan University in Guangzhou, China. His research interests are AI planning and sensor based activity recognition, machine learning and case-based reasoning, and data mining. He received his PhD degree from the University of Maryland, College Park. He is a senior member of the IEEE, member of AAAI and ACM and an associate editor for the IEEE TKDE and IEEE Intelligent Systems,

as well as International Journal of Knowledge and Information Systems. Contact him at the Dept. of Computer Science and Engineering, Hong Kong Univ. of Science and Technology, Clearwater Bay, Kowloon, Hong Kong; qyang@cse.ust.hk; <http://www.cse.ust.hk/~qyang>.



Jeffrey Junfeng Pan is a Ph.D. candidate in the Department of Computer Science, Hong Kong University of Science and Technology. Before joining HKUST, he was an M.Phil. student in Shanghai Jiao Tong University from 2003 to 2004. He received the Bachelor's degree in the Department of Computer Science, Zhongshan (Sun Yat-sen) University in 2003. His research interests include data mining, machine learning, optimization and their applications in mobile computing and information retrieval. Contact him at panjf@cs.ust.hk; <http://www.cs.ust.hk/~panjf>.