Contents

List of Figures

List of Tables

Chapter 1

Introduction

Due to the advances of sensor technology, smart power meters have been commonly deployed to collect electricity consumption of appliances in smart home environment. These usage data not only reveal the usage of appliances but also motivate us to extract useful knowledge by designing some mining techniques. Moreover, many applications also have been developed to utilize this knowledge for users to better understand how they use the household devices and to easily control their appliances.

Recently, with the concern of electricity conservation, one of the important applications is abnormal usage detection of appliance. Due to the significant efforts in reducing the emissions of CO2 and other GHGs (greenhouse gases), many researchers focus on the electricity conservation in the residential sector. Abnormal usage detection not only can help resident reduce electricity consumption, but also benefit our environment. Several previous researches [1-5] have focused on analysis of the usage behavior on single device and neglect the appliance correlation. However, in our daily life, we usually use different appliances simultaneously. For example, while the night, air conditioner and television in the living room may be turned on in the evening. Actually, the correlation among the usage of some appliances can provide valuable information to assist residents better detect abnormal usage of appliances.

Obviously, the consideration of correlation among appliance is a challenge issue for anomaly detection. Since the usage of a device has duration time, the correlations among appliances can be treated as an interval sequence. The abnormal usage based on the interval sequence is significantly different than that of the previous researches which only includes the information of a single appliance.

$$\begin{pmatrix} A^+ & B^+ & A^- & C^+ C^- & B^- \\ f_{A^+} & f_{B^+} & f_{A^-} & f_{C^+} f_{C^-} & f_{B^-} \end{pmatrix}$$

$$f(A^+) = \frac{1}{3h\sqrt{2\pi}}\left(e^{-\frac{1}{2}(\frac{x-2}{h})^2} + e^{-\frac{1}{2}(\frac{x-6}{h})^2} + e^{-\frac{1}{2}(\frac{x-13}{h})^2}\right)$$

$$h = \frac{range(\{2,6,13\})}{\sqrt{3}} = \frac{13-2}{\sqrt{3}} = 6.35.$$

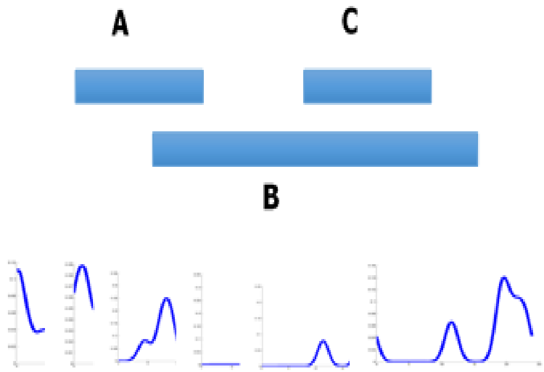

Fig. 1: An example of a correlation pattern.

Abnormal usage of the energy consumption for a particular period is significantly different than that of the previous time, during which some appliances are unexpectedly operating. In one of our companions' paper [2], an algorithm to transform the data log into correlation patterns was proposed. This study reported our findings of abnormal usage from a set of correlation patterns and a usage behavior pattern at a certain time. Fig. 1 shows an example of a correlation pattern and includes a frequent sequence (A+B+A–C+C–B–) and a corresponding probability density function set ($f_{A^+}f_{B^+}f_{A^-}f_{C^+}f_{C^-}f_{B^-}$). Appliances in a frequent sequence also show the correlation among the devices based on their locations in a house; for example, the usage of a television and a light is highly correlated when they are in the same room, but it is a coincidence when they are located in different rooms [2]. The correlation among the usage of some appliances can provide valuable information to assist residents better understand how they use their devices. The information of probability density function information is also very useful, which can be considered a normal distribution model for applying Extreme Value Theory (EVT). Fig. 2 illustrated the unknown behavior usage pattern at a certain time without information about the order of appliances turned on, which is useless. There follows in Fig. 2 a typical example of detecting abnormal usage, in which appliance A can be seen abnormal since Fig. 1 shows that A–often operated before C+ based on the correlation pattern. An example of the problem of anomaly usage detection in the smart home environment is sketched out in both Fig. 1

and Fig. 2 that demonstrated the habit of a family member. After having dinner his family, he usually goes to his bedroom and turns on light (A), light (B), and computer (C). He always turns off all appliances before sleeping. One day, if light A and light B are turned off while computer (C) is still working in the evening, it should be considered that he forgot to turn off the computer when he went to sleep. Hence, one of the applications of abnormal usage detection, Home Management System (HMS), would detect the on working computer and send a notification message to the user's smart phone.

Detection of abnormal usage is an important issue in smart home research. However, this is a challenging task when designing a remarkably effective and computationally reasonable solution. Our appliance behavior usage usually varies according to different periods of time and season, i.e. many behaviors of the same appliances in summer and in winter are totally different. For instance, a heater can be used daily in winter, but is seldom turned on in summer. In contrast, an air conditioner is usually turned on in summer, but is almost never used in winter. Appliances also have unique patterns such as seasonal types or daily types; for example, while the heater, a seasonal appliance, is frequently operated only in the summer; the light, a daily appliance, is usually turned on and off every day. Noticeably, this work determines tools aimed at designing solutions which can improve existing techniques for detecting anomaly usage in smart homes.

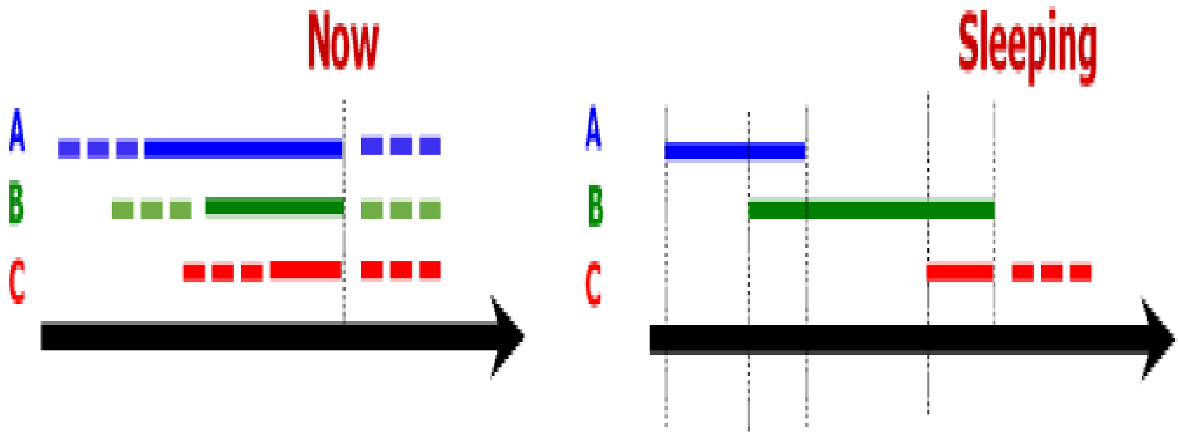

Fig. 2: An example of query patterns.

The proposed methods are advantageous. The first advantage is their simplicity in calculating anomaly score (these methods are easily comprehensible for all people, as well as the simplicity of its practical implementation). Another favor is these methods may be preferable when the observations are not exactly independent and identically distributed. For example, there may be a seasonal periodicity in terms of daily or yearly. Furthermore, it is the calculating speed which is very important when working in real time.

In this paper, several methods are proposed to detect anomalies. The contributions of our work are as follows. We develop an intelligent system, called Abnormal Detection System (ADS) to determine abnormal usage behavior, which includes three methods, namely Extreme Value for Measuring Anomaly Score (abbreviated as EVMAS), Sequence Patterns for Measuring Anomaly Score (abbreviated as SPMAS), and Time Intervals for Abnormal Detection (abbreviated as TIAD), to detect anomalousness in an appliance usage pattern. First, we used the probability density function as a model of normality for each appliance. Then, we redefined the definition of the “tails” to apply to our problem, and we set a threshold border in these tails of the model. It is necessary to redefine the tails because they are expected to determine normal distributions toward both positive infinity and negative infinity; however, the horizontal axis should be between 0 and 24, which is equivalent 00:00 AM to 23:59 PM. In other words, we find the minima distributions of the probability density functions that lie under a threshold. After that, we computed the anomaly score. An observed event is classified as abnormal when it exceeds another threshold. Moreover, it is known that the correlation among appliances is intrinsically complex. This pattern is really crucial for designing detection algorithm. Thus, we also propose an anomaly score for each appliance, and set some rules to determine when they are abnormal. Finally, we may know when appliances often or seldom operate in a day. To get this knowledge, we define the time intervals for each appliance by using the probability density function, based on which the status of appliance is determined as normal or abnormal. Experimental studies on real datasets indicate that the proposed algorithms are both efficient and practicability. Our experiments also show that the suggested approach consumes a much smaller memory space.

The rest of this paper is organized as follows. Section 2 provides a background, related work, and preliminaries. Section 3 introduces our abnormal usage behavior detection methods. Section 4 reports the experimental results in a performance study, and section 5 is our conclusion.

Chapter 2

Background

In this section, we review some related background knowledge which we apply to our approaches. First, we are summary the Extreme Value Theory (EVT), which is a branch of statistics which concerns the distributions of data of unusually low or high value. Second, we review several methods for estimation parameters which are used in EVT method. We also discuss some previous works utilized probability and sequence pattern for abnormal detection. Finally, we introduce some notions which will be used for our method.

2.1 Extreme Value Theory (EVT)

EVT is a tool used to consider probabilities associated with extreme and thus rare events, which has two major results. First, the asymptotic distribution of a series of maxima (minima) is modelled and under certain conditions the distribution of the standardized maximum of the series is shown to converge to the Gumbel, Fréchet, or Weibull distributions. The generalized extreme value (GEV) distribution is a standard form of these three distributions. All of these three distributions have one feature in common: If extreme value distributions are used to model empirical data, they will approximate the fat tail of the distribution with the highest precision. Second, the distribution of excess over a given threshold is modelled the behavior of the excess loss once a high threshold. The result is used to estimate the very high quantiles. The limit distribution is a generalized Pareto distribution (GPD). In this paper, we adapt the first result for our problem.

EVT forms representations for the tails of distributions. When discussing the properties of the tails of a distribution we will, for convenience, discuss the right-hand tail.

Consider a set Xn={X1, X2, …,Xn} of n i.i.d (independent and identically distributed) random variables, where Xi∈R is drawn from a distribution function F. The corresponding ordered sample in non-decreasing order is denoted by {X1:n, X2:n, …, Xn:n} where Xi:n, i = 1, …, n, stands for the i-th order statistic. In particular, X1:n and Xn:n represent the sample minimum and the sample maximum, respectively. In our paper, we focus only on the results about the sample minimum (the corresponding results for the sample maximum can be obtained from those of the sample minimum). We consider the sequence of minima M1=X1, Mn= Xn:n = min(X1, X2, …, Xn) for n≥ 2, obtained from the above sample. As mentioned, all the results for the sample maxima can be obtained from those of the sample minima, since mn = max(X1, X2, …,Xn) = –min(–X1, –X2, …, –Xn), then Mn= –mn. Hence, for large n,

$$\Pr\{M_n \le x\} = \Pr\{-M_n \ge x\} = 1 - \Pr\{M_n \le -x\}$$

$$\approx 1 - \exp\left\{-\left[1+\beta\left(\frac{-x-\alpha}{\eta}\right)\right]^{-\frac{1}{\beta}}\right\} = 1 - \exp\{-[1-\beta(\frac{x-\hat{\alpha}}{\eta})]^{-\frac{1}{\beta}}\}$$

where $1-\beta\left(\frac{x-\hat{\alpha}}{\eta}\right) > 0$ and $\hat{\alpha} = -\alpha$.

This distribution is the GEV distribution for minima.

Theorem If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr\left\{\frac{M_n - b_n}{a_n} \le z\right\} \to G(z)\ as\ n \to \infty$$

For a non-degenerate distribution function G, then G is a member of the GEV family of distribution for minima:

$$G(z) = 1 - \exp\left\{-\left[1-\beta\left(\frac{x-\hat{\alpha}}{\eta}\right)\right]^{-\frac{1}{\beta}}\right\}$$

on {z: $1-\beta\left(\frac{x-\hat{\alpha}}{\eta}\right)$> 0}, where $-\infty < \alpha < \infty, \eta > 0\ and -\infty < \beta < \infty$.

The parameters α, β, η correspond, respectively, to location, shape, and scale. The parameter β also called the tail index, indicates the thickness of the tail of the distribution. The larger the tail index, the thicker the tail. When the index is negative, it corresponds to a Weibull. When the index is equal to zero, the distribution H corresponds to a Gumbel. It corresponds to a Fréchet distribution when the index is greater than zero. The asymptotic distribution of the maximum which is estimated without making any assumptions about the original distribution, always belongs to one of these three distributions.

The GEV provides a model for the distribution of block minima (maxima). Its application consists of blocking the data into blocks of equal length, and fitting the GEV to the set of blocks. But in implementing this model for any particular dataset, the choice of block size can be critical. The choice amounts to a trade-off between bias and variance: blocks is likely to be poor if they are too small, leading to bias in estimation and extrapolation; large blocks generate few block extreme, leading to large estimation variance. In our paper, the choice of block size is not easy. For example, as shown in Fig. 3, the appliance 1 is often turned on in two periods: 10:00 AM to 11:00 PM and 18:00 PM to 04:00AM (of next day) and seldom turned-on from 05:00AM to 09:00AM and 12:00PM to 17:00PM. Therefore, we suggest a new method to choice minima values which will be introduced in next section.

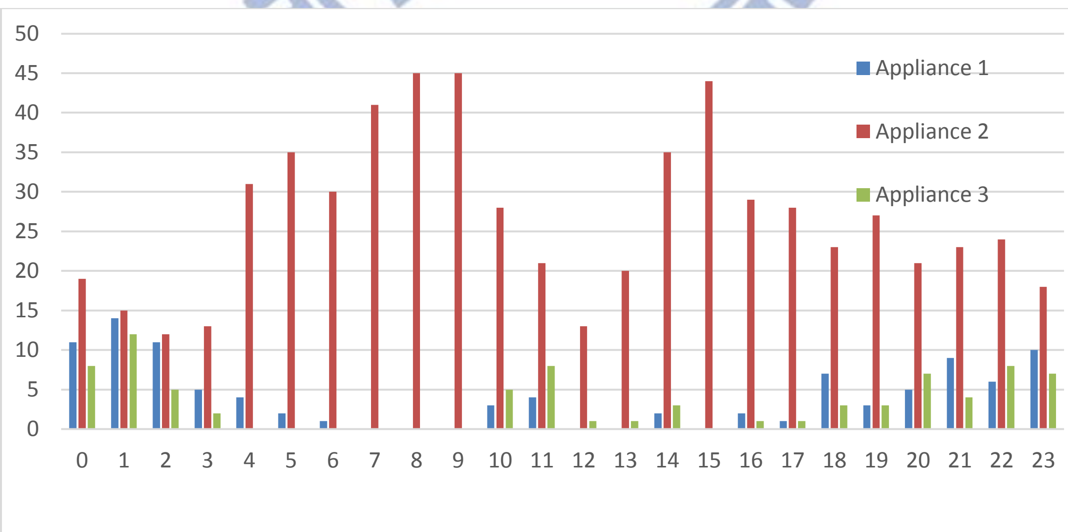

Fig. 3: An example of usage behavior log, turned on time information of three appliances for 5 weeks is illustrated.

2.2 Estimation of parameters in EVT

The sequential pattern mining originally focuses on the time point-based database [11, 26]. Han et al. [11] propose an efficient sequential pattern mining method, named FreeSpan. The general idea of FreeSpan is to integrate the mining of frequent sequences with that of frequent patterns and use projected sequence databases to confine the search and the growth of subsequence fragments. Pei et al. [26] propose an efficient sequential mining algorithm, named PrefixSpan, based on divide-and-conquer inspiration. PrefixSpan explores prefix-projection in sequential pattern mining, which substantially reduces the size of projected databases.

Estimating the GEV parameters (α, β, η) constitutes an important task in EVT approach, since it is a starting point for statistical inference about extreme values of a population. There are basically two approaches in order to obtain estimates for the GEV parameters, a parametric approach and a semi-parametric approach. In this paper, we follow a parametric approach that we can use the limiting distribution of the sample extremes as an exact distribution that can be fitted to data. The focus of this paper is on the sample minimum, Mn= min(X1, X2, …, Xn). This method is distinguished according to how many specific observations are picked up among the available sample data.

For this approach, the sample of size n is divided into m sub-samples size of k (or m blocks size of k), with $n = m \times k$ and k sufficiently large. In each block, the smallest observation is selected, so that we obtain a sample of sample minima.

$$Y = M_k = \min(X_1, \dots, X_k)$$

For considering m blocks, we get a collection of m sample minima, (Y1, Y2, …,Ym). The most popular estimation methods for the GEV are the Maximum Likelihood and the Method of Moments. These methods mainly focus on estimating Weibull parameters, namely, shape parameter (β) and scale parameter (η)[6, 7].

Maximum Likelihood Estimator (MLE)

The method of maximum likelihood [8-10]is a commonly used procedure since it has desirable properties. Let {x1, …,xn} be a random sample of size n drawn from a probability density function $f_x(x;\theta)$ where θ is an unknown parameter. The likelihood function of this random sample is

$$L = \prod_{i=1}^{n} f_{x_i}(x,\theta)$$

The maximum likelihood estimator of θ, say $\hat{\theta}$, is the value of θ that maximizes L or the logarithm of L. The MLE of θ may be a solution of

$$\frac{d\log L}{d\theta} = 0$$

We apply the MLE to estimate the Weibull parameters (the shape and the scale). The likelihood function will be

$$L(x_1,\dots,x_n;\beta,\eta) = \prod_{i=1}^{n}(\frac{\beta}{\eta})(\frac{x_i}{\eta})^{\beta-1}e^{-(\frac{x_i}{\eta})^{\beta}}$$

We different with respect to β and η and equate to zero

$$\frac{\partial\ln L}{\partial\beta} = \frac{n}{\beta} + \sum_{i=1}^{n}\ln x_i - \frac{1}{\eta}\sum_{i=1}^{n}x_i^{\beta}\ln x_i = 0$$

$$\frac{\partial\ln L}{\partial\eta} = -\frac{n}{\eta} + \frac{1}{\eta^2}\sum_{i=1}^{n}x_i^{\beta} = 0$$

On eliminating η between these two above equations, we get

$$\frac{\sum_{i=1}^{n}x_i^{\beta}\ln x_i}{\sum_{i=1}^{n}x_i^{\beta}} - \frac{1}{\beta} - \frac{1}{n}\sum_{i=1}^{n}x_i = 0$$

which can be solved to get the estimate of β. This can be accomplished by Newton – Raphson method. It will be given as

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

where

$$f(\beta) = \frac{\sum_{i=1}^{n}x_i^{\beta}\ln x_i}{\sum_{i=1}^{n}x_i^{\beta}} - \frac{1}{\beta} - \frac{1}{n}\sum_{i=1}^{n}x_i$$

and the computation of the derivative as [7] is

$$f'(\beta) = \sum_{i=1}^{n}x_i^{\beta}(\ln x_i)^2 - \frac{1}{\beta^2}\sum_{i=1}^{n}x_i^{\beta}(\beta\ln x_i - 1) - (\frac{1}{n}\sum_{i=1}^{n}\ln x_i)(\sum_{i=1}^{n}x_i^{\beta}\ln x_i)$$

Once β is determined, η can be estimated using the equation

$$\eta = \frac{\sum_{i=1}^{n} x_i^{\beta}}{n}$$

There are two things different between EVT and classical EVT. First, the probability density function decreases with increasing distance to the modes. Hence, extreme in magnitude are also minima in probability density values. Second, selecting the most improbable sample with respect to f is equivalent to selecting the sample of minimal magnitude with respect to the density function over f(X). Therefore, univariate EVT may be applied to samples drawn in the probability space.

Method of moments (MOM)

The MOM is another technique commonly used to estimate parameters. Given {x1, …,xn} represent a set o data, an unbiased estimator is given as

$$\widehat{m}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

where $\widehat{m}_k$stands for the estimate of mk. The kth moment is given as

$$\mu_k = \left(\frac{1}{\eta^{\beta}}\right)^{-\frac{k}{\beta}} \Gamma(1+\frac{k}{\beta})$$

where Γ signifies the gamma function

$$\Gamma(s) = \int_0^{\infty} x^{s-1}e^{-x}dx, (s > 0)$$

We can compute the first and the second moment as

$$m_1 = \hat{\mu}_k = \left(\frac{1}{\eta}\right)^{\frac{1}{\beta}} \Gamma(1+\frac{1}{\beta})$$

and

$$m_2 = \hat{\mu}_k^2 + \hat{\sigma}_k^2 = (\frac{1}{\eta})^{\frac{2}{\beta}}\{\Gamma\left(1+\frac{2}{\beta}\right) - [\Gamma(1+\frac{1}{\beta})]^2\}$$

We get a function of β by dividing m2 by the square of m1

$$\frac{\hat{\sigma}_k^2}{\hat{\mu}_k^2} = \frac{\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2(1+\frac{1}{\beta})}{\Gamma^2(1+\frac{1}{\beta})}$$

We have the coefficient of variation (CV) as

$$CV = \frac{\sqrt{\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2(1+\frac{1}{\beta})}}{\Gamma(1+\frac{1}{\beta})}$$

In order to estimate β and η, we need to calculate the coefficient of variation. The scale parameter (η) can be estimated as

$$\hat{\eta} = \{\bar{x}/\Gamma[(\frac{1}{\hat{\beta}}+1)]\}^{\hat{\beta}}$$

where $\bar{x}$ is the mean of the data.

MLE is more popular and attractive but this method its primacy in a small sample case, where it is outperformed by MOM [6]. However, MLE is the preferred parametric estimation method because of its flexibility of modification, to incorporate more complex problems. On the opposite, MOM has a serious difficulty in dealing with more complex structures. In this paper, we adapt MLE for estimation some parameters for EVT approach.

Chapter 3

Related Work

To the best of our knowledge, with the consideration of abnormal usage detection based on sequence patterns, there have been many proposed methods for behavior recognition [11-16], and some discussion of detection abnormal usages of appliances in a smart home environment [1, 17-19]. Some studies utilized the behavior patterns to detect anomaly user behavior that include a temporal-based approach where temporal relations were built to detect anomalies [2, 5, 14, 17, 18]. RFID-based algorithm was experimented for human behavior modeling and abnormal detection for elderly care by several authors[12, 13]. Neural network approach has been done to predict the future values which are used to inform the caregiver in case anomalous behavior is predicted [14-16]. Besides, Chen and Cook [1] proposed a framework to mine energy data and extend a suffix tree data structure and then use a clustering algorithm to detect energy patterns outliers which are far from their cluster centroids. Moreover, conceptual studies and used cases reported for abnormal events in the smart home context were proposed by some authors in [20, 21]. Chen et al. [3, 5] introduce some frameworks and algorithms to describe users' representative behaviors. Based on these patterns, we can be able to adapt our usage behavior for abnormal detection to conserve the energy easily. Jakkula et al. [18] propose an Apriori-based algorithm for activity prediction and anomaly detection from sensor data in a smart home.

Previous researches of abnormal detection mainly focused on sequence pattern [22-25] and probability density function based on EVT [26-31]. To the best of our knowledge, very few studies facilitate the detection of abnormal usage in smart home with the utilization of distribution based on EVT. While these papers mainly discussed the ways for detecting abnormal usage by using the sequence patterns, we intend to introduce the notion of the correlation pattern combined with probability density function into the detection of anomaly usage of appliances. Since we can see the probability function of happening time of each interval not only simply interprets the relations. The gap of happening time also make the patterns contain different meanings. With the probability function appended in the correlation patterns, we can employ it to our methods. There are several previous works using EVT for abnormal detection or novelty detection in some fields such as structural engineering, medical, finance, earth sciences, traffic prediction, and geological. Andreev et al. [27] use Extreme Value Theory for Stock Market in Russia. This paper utilized POT model and GPD distribution which give the description on tail distribution of financial returns/losses. They compute all necessary parameters, threshold and the value-at-risk for their method. However, they applied EVT and POT directly with no improvement or modification. Luca, S., et al. [26] Detect rare events using extreme value statistics applied to epileptic convulsions in children. They proposed an unsupervised method which uses EVT and seizure detection based on a model of normal behavior that is estimated using all recorded and unlabeled data. They also have several enhancements for EVT in their method but they lack of introducing how to determine threshold and parameters in detail. Clifton, D. A., et al. [32] developed a technique that is generally applicable and has the additional advantage of translating a multivariate model of behavior to an univariate model of minimal densities, using classical EVT. The approach uses Gaussian Mixture Model (GMM) fitting the data to create the probabilistic model and MEVS to define the threshold for outliers. The MEVS approach uses the model of normality to perform novelty detection in the probability space. Roberts, S. J.[29, 30] proposed an extension of classical EVT to mixtures of multivariate Gaussians, with three applications in biomedical engineering. This work recommends, for a given sample x, that only the extreme value distribution (EVD) associated with the kernel closest to x (in the Mahalanobis sense) is considered, which may be calculated using the known EVD of the single-sided univariate Gaussian distribution.

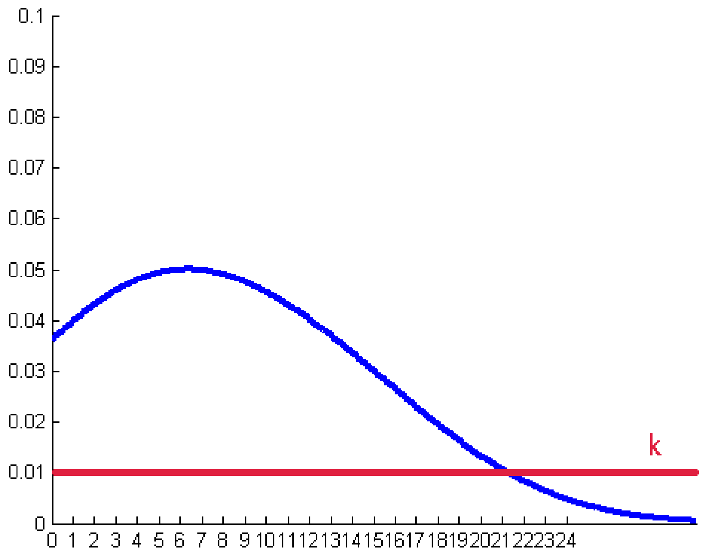

Fig. 4: The integration of bimodal probability distribution p(x) (blue line) and the contour p(x)=0.2 (dashed line).

In some previous studies [33-36], a heuristic novelty threshold has set on the pdf f(x)=k, such that x is abnormal when f(x)<k. f(x) is used simply as abnormal score, and the threshold is set such that separation between normal and abnormal data is maximized on the validation dataset. Some other approaches[33] use the cumulative probability Fn associated with fn. They compute the probability mass obtained by integrating fn over the region R where fn exceeds the novelty threshold. The region $R = \{x \in D | f_n(x) \geq k\}$:

$$F_n(k) = \int_R f_n(x)dx \qquad (1)$$

An example is shown in Fig. 4, in which the distribution fn is multimodal, and which has been approximated using GMM. A threshold is shown at f(x)=k=0.2 in the figure, in which the probability mass P enclosed by that threshold. Clifton et al. [28] have discussed some disadvantages of using Fn to identify the threshold. Setting a novelty threshold using Fn only has a valid when m=1 (with m is the number of observed data).

In this paper, the number of observed training data will be fix. Each appliance in our training dataset has a fix sample for pdf f. One thing that we need to define is a threshold for F(x). One suitable way is based on the size of training dataset. If appliance has m events then we set threshold at ln(m)≅ size. Hence, we will find the set of yk such that F(k)=size%. The area of F(k) is shaded area in Fig. 4 where yk<f(x).

Frequent sequences	
[A+|A-] : 65	[B+|B-|C+|C-] : 4
[A+|B+|A-|B-] : 12	[C+|C-] : 67
[A+|B+|C+|C-|A-|B-]:2	[C+|C-|C+|C-|C+|C-]:3
[A+|B+|D+|B-|A-|D-]:2	[C+|C-|C+|C-|D+|D-]:22
[B+|B-] : 66	[D+|D-]:55
[B+|B-|C+|A+|A-|C-]:5	[D+|D-|C+|C-]:25
[B+|B-|C+|C-] : 20	[D+|D-|C+|D+|D-|C-]:6
[B+|B-|C+|C-|C+|C-]:6	[D+|D-|D+|A+|A-|D-]:2
Probability density functions: $\boldsymbol{f}_{A^+}, \boldsymbol{f}_{A^-}, \boldsymbol{f}_{B^+}, \boldsymbol{f}_{B^-}, \boldsymbol{f}_{C^+}, \boldsymbol{f}_{C^-}, \boldsymbol{f}_{D^+}, \boldsymbol{f}_{D^-}$	

Table 1 – An example of a correlation pattern set, each frequent sequence has two part: the order of appliances' occurrence and number of occurrence's times.

Chapter 4

Preliminaries

Definition 1 (Sub pattern at a time period and sub-patterns set at a time period)

Given a correlation pattern P, a correlation sub-pattern at a time period of two appliances is a subset of correlation pattern P if end-time of the first appliance occurs after start-time of the second appliance. A set of sub-patterns at a time period is the collection of all sub-patterns at the time of all correlation patterns.

We take the database in Table 1 as an example. Let P1 be [A+B+C+C−A−B−], a sub-pattern is [A+], a set of sub-patterns at a time period is S1 = {[A+], [A+B+], [A+B+C+]}. Let P2 be [B+|B−|C+|A+|A−|C−], a set of sub-patterns at a time period is S2 = {[B+], [C+], [C+A+]}.

Definition 2 (Appliances' Combinations)

Given a pattern P, a combination is a way of selecting appliances from P, such that the order of selection does not matter. A subset S is a combination of appliances in P, denoted by S⊆P.

For example, P is [A+B+C+], there are seven subsets S= {[A+], [B+], [C+], [A+B+], [A+C+], [B+C+], [A+B+C+]}. A list of sub-patterns is a combination of all appliances with 2k-1 subsets. Notice that we do not use empty sub pattern.

Chapter 5

Anomaly detection methods

In this section, we propose three methods to detect anomalous usage behaviors by exploring the correlation patterns. The first method uses of Extreme Value Theory on the tails of the probability density function. The second method is that we compute the proportion of appliances' occurrences in data set for each appliance using frequent sequences in correlation patterns data set. Final, we determine time intervals for each appliance which can be used to determine anomalous.

Fig. 5 presents the system framework of Abnormal Detection. First, we collected the usage data of all appliances by smart meters and sent the data log to cloud server. After that we transform the data into correlation patterns using CoPMiner [2]. Then our Abnormal Detection System (ADS) uses this correlation patterns to detect abnormal usage behavior. Finally, we output all abnormal extraordinary behavior to users.

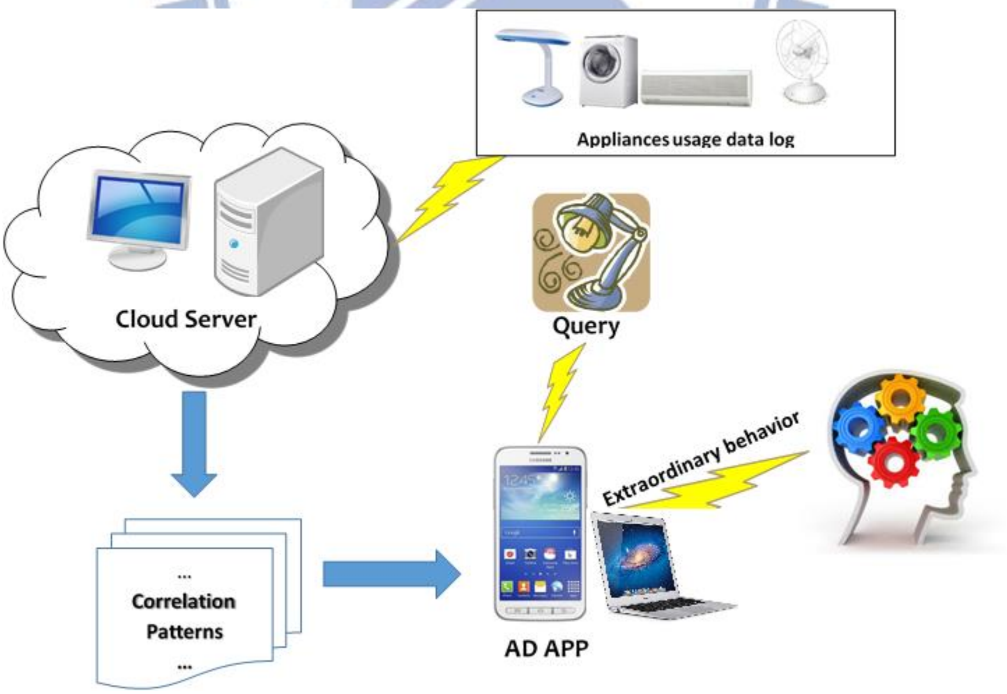

Fig. 5. The system framework of Abnormal Detection.

5.1 Extreme Value for Measuring Anomaly Score (EVMAS)

Anomaly detection using EVT approach is based on a model of normal behavior which was presented by the probability distribution function. In this section we borrow the methodology which was proposed in [26].

Density Modelling using Gaussian Mixture Models

The first stage of the investigation is to construct a model of normality using normal training data. The model of normality is provided using two candidate techniques, Gaussian Mixture Model (GMM) and Parzen window estimation. These approaches estimate the unconditional probability density of training data, f(x). If we consider m training points from the input data x, the data density for the Parzen window estimator is defined as:

$$f_i(x) = \langle K(x), h, \{t_{i_1}, \dots, t_{i_m}\}\rangle = \frac{1}{mh}\sum_{j=1}^{m} K\left(\frac{x - t_{i_j}}{h}\right),$$

with $h = \frac{range(\{t_{i_1},\dots,t_{i_m}\}}{\sqrt{m}}$

and where K is Gaussian Normal Kernel,

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

The data density is defined as:

$$f(x) = \sum_{j=1}^{M} P(j)p(x|j)$$

Where p(x|j) is the probability of x w.r.t kernel j, and P(j) is the prior probability that x was generate by kernel j.

f(x) have to satisfy the following criteria:

The function should be non-negative throughout.

The function should integrate to 1.

EVT in Multimodal

Given a data set D (correlation pattern set), consisting of appliances, each appliance has the probability density function (pdf) y=f(x) which build from its training dataset. Anomaly detection address the question whether a query pattern Q={q1,…, qk} is drawn from f(x) or not. Each appliance is Q has the corresponding density values based on pdf f(x).

(y1, …,yk)=(f(q1), …, f(qk))

First, based on the equation (1), we compute a set of ymin in which the distribution is lower than the threshold. The threshold will be set based on the size of training dataset. If a training dataset has m events then setting threshold at $size = \ln(m)$ Then, set F(x)=size/100, the F(x) is given as:

$$F(x) = \int_R f_n(x)dx = size\%$$

The set $y_{min} = \{y|x \in D and F(x) = \mathrm{size}/100\}$: The distribution of ymin describes the distribution of minima of training dataset. An anomaly may be located in the tails of pdf f or between the modes of f. We find ymin which is the tails of f or the low probability between the modes of f.

The next step is that we apply the Weibull distribution for ymin. The form of the 3-parameter Weibull distribution is commonly used in practice

$$w(y_{min}) = \frac{\beta}{\eta}\left(\frac{y_{min} - \gamma}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{y_{min} - \gamma}{\eta}\right)^{\beta}\right)$$

Where parameters β >0, η and γ are shape, scale, and location respectively. The location parameter, γ, locates the distribution along the abscissa. The distribution moves to the right (if γ >0) or to the left (if γ<0). We set γ=0, the distribution starts at the origin. The parameters β and η can be found by using maximum likelihood estimates. The 2-parameter Weibull is obtained by setting γ=0, and is given by

$$w(y_{min}) = \frac{\beta}{\eta}\left(\frac{y}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{y}{\eta}\right)^{\beta}\right)$$

Since the probability of these are likely to very close to zero, the use of log helps emphasize their differences. The transformation is given as:

$$t = -\log(\mathrm{w})$$

Using this transformation, the short tail near zero of the Weibull distribution is then stretched out as the right tail of the Gumbel distribution for maxima. Hence, extreme values can be shown more clearly. The cumulative distribution function of the Gumbel distribution is

$$G(t) = \exp\left(-\,exp\left(-\frac{t-c}{d}\right)\right)$$

Where c=1/β and d= −ln(η).

In abnormal detection, extrema are regarded as potentially anomaly. The final step is that we define an anomaly score for each appliance. Hence, we can compute the anomaly score for each appliance is query pattern Q. Anomaly score, AS1, can be defined as:

$$AS_1(q_k) = G(\mathrm{t}) \quad (2)$$

Note that $AS_1(y_k)$ takes low values if x is close to the center of the distribution and increases as x becomes more abnormal.

Symbols	Meaning
A … E	Events or appliances
Q	Query pattern or unknown pattern
f(x)	Probability density function
F(x)	Cumulative probability distribution
AS1(x)	Anomaly score for appliance x.

Table 2 – Notations of EVT based method

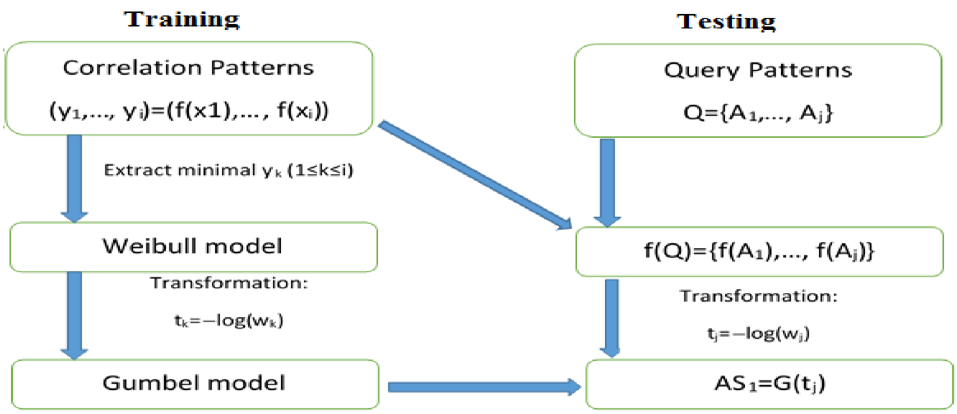

Fig. 6. Schematic of EVMAS algorithm.

We use database in Table 1 as an example, we want to check that appliance D turned–on at 23:30PM is normal or abnormal. First, we compute the size and collect the minima distribution from pdf $\boldsymbol{f}_{\boldsymbol{D}^+}$. We note that we do not use the pdf $\boldsymbol{f}_{\boldsymbol{D}^-}$ since we do not have turned-off information in query pattern, and we only check appliances which are turning on at a certain time. After collecting minima distribution, we fit parameters for Weibull distribution and Gumbel distribution, and then we compute the anomaly score for appliance D at 23:30PM. Appliance D in this dataset has ID 60 in real dataset, and the result is shown in Fig. 11. Appliance D is abnormal when turning on at 23:30PM. We can read more detail in section IV, and appliance 60 is described in more details. A schematic of the methodology is shown in Fig. 6.

5.2 Sequence Patterns for Measuring Anomaly Score (SPMAS)

We explore the frequent sequences of correlation patterns dataset to calculate the anomaly score for each appliance. Table 1 shows the set of frequent sequences which can be used to determine abnormal events. We assume that all sequence patterns in this dataset are normal patterns. All appliances occurred in normal scenarios. For example, a pattern, [A+|B+|D+|D-|B-|E+|E-|A-] describes a normal occurrence order of four appliances (A, B, D, and E). We can extract this sequence into four sub patterns as [A+], [A+B+], [A+B+D+], and [A+E+] as definition 1. This means that [A+] can turn on while [D+] turns on but [D+] and [E+] cannot turn on at the same time. This is necessary to decompose a correlation pattern into sub patterns because an appliance can occur many times and it is difficult to take out appliances which occur at a time period.

Symbols	Meaning
A … E	Events or appliances
L	Sub-pattern set
Q	Query pattern or unknown pattern
P(x)	The proportion of x in dataset. Ex: P(A+B+) is the proportion of occurrence of A+ and B+ in L.
AS2(x)	Anomaly score for appliance x.

Table 3 – Notations of sequential pattern

Algorithm 1: SPMAS(CP, Q)
Input: a correlation pattern dataset CP, a query pattern Q.

Output: all abnormal appliances A.

```
A←ϕ;
transform CP into sub pattern L by Definition 1;
transform Q into a combination list QL by Definition 2;
evaluate proportion P(x) for each element x in QL;
compute anomaly score AS2 for each appliance;
A← min(AS2);
output all appliances in A;
```

For unknown pattern, we actually do not know the order of appliances in this pattern. The order in query pattern is random, which depends on users' input. Therefore, it is not easy for us to compare the query pattern with existing correlation pattern set. Our method use probability theory to solve this problem.

Given a dataset D (correlation pattern set), consisting of appliances and their occurrences in classical sense, denoted by Ω. It is then assumed that for each element x∈Ω.

For the unknown pattern Q={X1, X1, X3} of space Ω, assume that all appliances in Q have occurred in dataset D. Anomaly score of Xi is defined as:

$$AS_2(X_1) = f(X_1) + f(X_1X_2) + f(X_1X_2) + f(X_1X_2X_3)$$

where $f(X_1) = \frac{number\ of\ occurrences}{total\ of\ sub\ patterns}$,

$f(X_1X_2) = \frac{number\ of\ occurrences\ of\ X_1\&X_2\ together}{total\ of\ sub\ patterns}$,

$f(X_1X_3) = \frac{number\ of\ occurrences\ of\ X_1\&X_3\ together}{total\ of\ sub\ patterns}$,

$$f(X_1X_2X_3) = \frac{number\ of\ occurrences\ of\ X_1, X_2\&X_3\ together}{total\ of\ sub\ patterns}$$

We also define some rules helps us identify the anomalies.

The rules are given as:

If P(Q) > 0 then the query pattern Q has no abnormal usage behavior.

If P(Q) = 0 then

Calculate AS2(Xi) for each appliance.

Min(AS2(Xi)) is abnormal.

Take the dataset in Table 1 as an example. First, we transform the correlation patterns into sub patterns by Definition 1. Table 4 shows the sub patterns corresponding to correlation patterns in Table 1.

Sub-patterns set (L)	
[A+] : 65	[B+], [C+]: 4
[A+],[A+B+]: 12	[C+] : 67
[A+],[A+B+],[A+B+C+]:2	[C+], [C+], [C+] : 3
[A+],[A+B+],[A+B+D+]:2	[C+], [C+], [D+] : 22
[B+]:66	[D+] : 55
[B+],[C+],[C+A+]:5	[D+], [C+] : 29
[B+],[C+]:20	[D+], [C+], [C+D+] : 6
[B+],[C+],[C+]:6	[D+], [D+], [D+A+]:2

Table 4 – A sub pattern set is generated from Correlation pattern set as definition 2 with 519 sub patterns.

Given a query pattern Q= [A+B+C+D+]. As Definition 2, its subsets are QS={[A+], [B+], [C+], [D+], [A+B+], [A+C+], [A+D+], [B+C+], [B+D+], [C+D+], [A+B+C+], [A+B+D+], [A+C+D+], [B+C+D+], [A+B+C+D+]}.

Using information about occurrence of appliances in L, we compute the proportion of each subset of the query pattern. The proportion of each subset is calculated as following:

P(A+) = 108/519; P(A+B+) = 20/519; P(A+C+) = 7/519; P(A+D+) = 4/519; P(A+B+C+) = 2/519; P(A+B+D+) = 2/519; P(A+C+D+) = 0; P(A+B+C+D+) = 0;

➔AS2(A+)=(108+20+7+4+2+2+0+0)/519=0.2755;

AS2(B+)=0.2871; AS2(C+)=0.4277; AS2(D+)=0.2620;

D+ is abnormal in this query pattern Q because min (AS2(QS)) = AS2(D+).

Sub patterns as Definition 1 was shown in Table 4, which correspond to the correlation patterns set in Table 1. We need to decompose correlation patterns into sub patterns because of two reasons. First, we want to combine appliances which may turn on at a time period while an appliance can turn on and turn off many times in a correlation pattern. Each sub pattern indicates that appliances may turn on in the same period, which can be used for computing appliance anomaly scores. Second, we do not use turned off information in correlation patterns for our method because we have no turned off values of appliances in the query sequence. Therefore, we can eliminate turned off symbol (–) in sub patterns.

The first advantage of SPMAS is that we do not need time information of appliances to determine abnormal usage behavior since there is no time information in frequent sequences of correlation patterns. Another advantage of this method is that we do not pay attention to the order of appliance when compute anomaly scores because we have no order information in query pattern. However, we cannot identify exactly abnormal usage behaviors when there is only one appliance in the query pattern. If there are many appliances with low scores, we can use a threshold instead of using the minimum value.

5.3 Time Intervals for Abnormal Detection (TIAD)

First, we define time intervals of occurrence of each appliance in the correlation pattern that is to identify appropriate time periods for each appliance.

Symbols	Meaning
E	Events or appliances
T	Timestamps of E
P	The probability density function of T
M	The sorted list [00:00, 23:59]

Table 5 – Notations of time intervals

One possible approach is to generate the intervals while the correlation patterns discovery part. However, the time complexity increases, since all possible intervals have to be considered.

Algorithm 3: TIAD(pdf, Q)

Input: probability density function set pdf, a query pattern Q. Output: all abnormal appliances A. For each appliance: $\forall\, t_i, t_j \in M$: Let x_1 be $t_i \in M$ such that $\begin{cases} f(t_i) \geq k \\ f(t_{i-1}) < k\ \boldsymbol{or} t_{i-1} \notin M \end{cases}$ Let x_2 be $t_j \in M$ such that $\begin{cases} f(t_j) \geq k \\ f\left(t_{j+1}\right) < k\ \boldsymbol{or} t_{j+1} \notin M \end{cases}$ $interval_k \longleftarrow [x_1, x_2]$ A←appliance $\in interval_k$ Output all appliance in A.

Table 6 – Determining time intervals for each appliance

The potential drawback is that the quality of accuracy can be affected by how we define the intervals. However, we can minimize this possibility if we do not fixed-width time intervals. Instead determine the intervals based on the original dataset, we use the probability density function. The area under the curve from time t1 to time t2 gives us the probability that an appliance will turn on between t1 and t2.

We need to define the time intervals as intervals between local maxima of the probability density function. The main idea behind this approach is that a user often turn-on this appliance during a certain time period. For example, as illustrated in Fig. 4, a user may usually turn-on in two periods as between 03:00AM and 06:30AM; between 17:30PM and21:00PM. Since some appliances turn on more frequently than others, we define the time intervals by computing probability density functions for each appliance separately.

For anomaly detection, an appliance in the query pattern will be determined whether it is normal or abnormal. The appliance is normal when its time is in this appliance's time intervals. Each appliance will be determined by its time intervals. All appliances in query pattern must be existed in dataset.

Chapter 6

Experiment Evaluations and Results

For performance discussion, we compare our methods EVMAS, SPMAS, and TIAD together. All algorithms were implemented in Python language and tested on an Intel Core 2 Quad CPU Q9400 @2.66GHz with 8GB of main memory running Windows 7 system. This performance study has been conducted on both real world and synthetic dataset. First, we implement our three methods on data set in detail. Second, we compare the execution time using real world dataset at different threshold size. Finally, we compare the accuracy of the three methods on real world dataset.

6.1 Application

First Kind Synthetic Data Generator

Our aim is to illustrate the tail distribution estimation of a set of correlation patterns and use the results to quantify the anomaly score. Table 7 gives the list of the correlation patterns considered in our analysis. The illustration focuses mainly on the appliance ID 60, providing confidence intervals and graphical visualization of the estimates, whereas for the other appliances only point estimates are reported. The application has been executed in Python 3.3 programming environment.

Appliance ID	Appliance name	Observations
13	Kitchen outlet – 1500W	505
17	Kitchen outlet – 30W	624
24	Washer dryer 3W	497
29	Outlets	238
57	Furnace	258

60	Smoke alarm	682

Table 7 – Data analyzed – Correlation Pattern sets.

Fig. 7 shows the plot of the n=682 observed of the appliance ID 60. This is multimodal distributed such that anomalies possibly occur between the modes. Hence, we not only consider the left and the right tail of the distribution but also between the modes. We use maximum likelihood estimation, which is one of the most common estimation procedures used in practice. We also compute likelihood based interval estimates of the parameters and the quantities of interest which provide additional information related to the accuracy of the point estimates. These intervals, contrarily to those based on standard errors, do not rely on asymptotic theory results and restrictive assumptions. We expect them to be more accurate in the case of small sample size. Another advantage of the likelihood-based approach is the possibility to construct joint confidence intervals. The greater computational complexity of the likelihood-based approach is nowadays no longer an obstacle for its use.

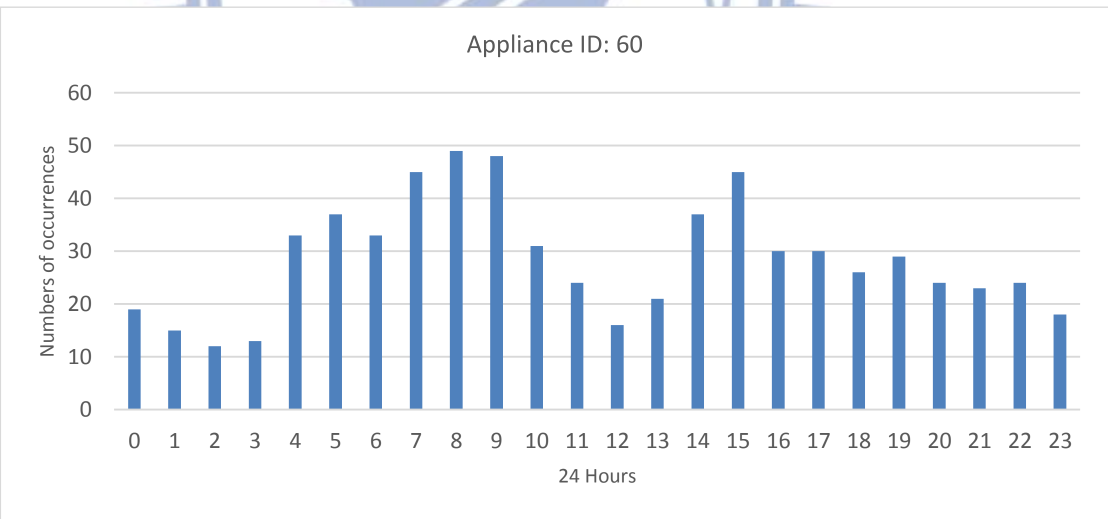

Fig. 7: Numbers of occurrences per hour for 45 days.

The implementation of the EVT method involves the following steps: select the threshold u, fit the Weibull and Gumbel distributions and then compute anomaly score.

Selection of the threshold u

We know that the higher the threshold the less observations are left for the estimation of the parameters of the tail distribution function. There is no automatic algorithm with satisfactory performance for the selection of the threshold u. In previous work [33-36] the threshold has been set on the pdf fn(x)=k. In this paper, we define the threshold for cdf F(x)=size. The size value is based on the size of samples dataset. Appliance 60 has 682 samples observations, we have ln(682)=6.525. Hence, we collect the set yk such that F(k)=6.525%. The number of observations exceeding the threshold is 53.

Maximum Likelihood Estimation

Given the theoretical results presented in the previous section, the distribution of the observations that we collect above should be drawn a Weibull distribution. We compute the value β and η that maximize the log-likelihood function for the sample yk.

Appliance ID	ln(n)	β	η
13	6.223	0.6124	3.8701
17	6.436	1.2495	16.6069
24	6.209	1.6596	134.082
29	5.472	1.3199	21.026
57	5.553	3.0131	284.27
60	6.525	0.6873	2.3496

Table 8 – Maximum Likelihood Estimation.

Weibull and Gumbel distributions

We obtain the estimates β=0.6873 and η=2.3496. Fig. 8 shows the pdf Weibull distribution for appliance 60.

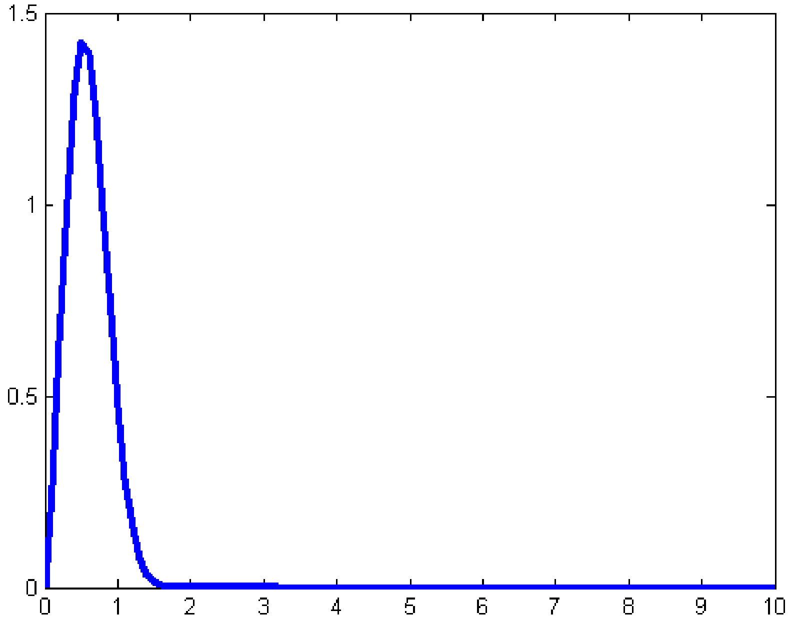

Fig. 8: Weibull 2-Parameter probability density function at β=0.6873 and η=2.3496.

We also obtain the estimates c = 1/β = 1.4545 and d = −ln(η)= −0.8542. Fig. 9 shows the pdf Gumbel distribution. The corresponding cdf of Gumbel is illustrated in Fig. 10.

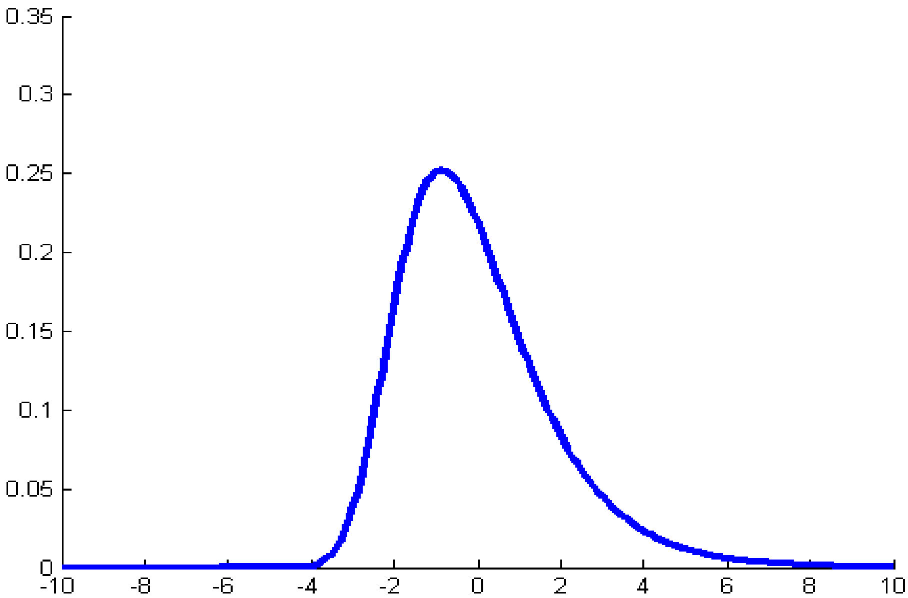

Fig. 9: Gumbel probability density function (pdf) at c = 1.4545 and d = −0.8542.

Anomaly score

An anomaly score of an appliance is defined as (2). An appliance which its anomaly score is high and between the Gumbel value of f(x) can be viewed as anomaly. In other words, an appliance is abnormal if this pdf value is lower than a threshold. Table 9 illustrates the pdf and Gumbel information of appliance 60 in a day. We can see that anomaly scores are higher than that of other scores when appliance turns on between 23:00 and 01:00. It means that the probability of appliance occurrence is lower than that of others in this period.

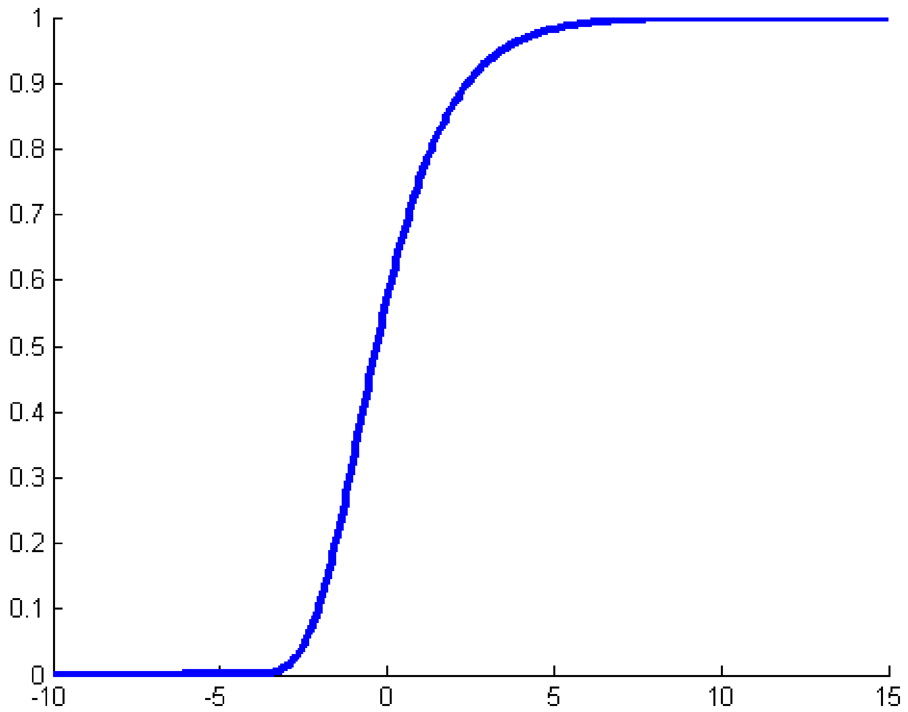

Fig. 10: Gumbel cumulative distribution function (cdf) at c =1.4545 and d = –0.8542.

Time	Pdf	AS1	Time	pdf	AS1
00:00	0.6412	4.0017	12:00	0.6463	3.0363
00:30	0.6426	3.6585	12:30	0.6460	3.0724
01:00	0.6437	3.4284	13:00	0.6458	3.1101
01:30	0.6445	3.2906	13:30	0.6455	3.1405

02:00	0.6450	3.2212	14:00	0.6454	3.1591
02:30	0.6451	3.1931	14:30	0.6453	3.1659
03:00	0.6452	3.1819	15:00	0.6453	3.1652
03:30	0.6453	3.1746	15:30	0.6454	3.1636
04:00	0.6453	3.1709	16:00	0.6453	3.1651
04:30	0.6453	3.1737	16:30	0.6453	3.1700
05:00	0.6452	3.1821	17:00	0.6452	3.1788
05:30	0.6451	3.1942	17:30	0.6451	3.1937
06:00	0.6450	3.2110	18:00	0.6450	3.2137
06:30	0.6449	3.2322	18:30	0.6449	3.2309
07:00	0.6447	3.2524	19:00	0.6449	3.2353
07:30	0.6447	3.2619	19:30	0.6449	3.2245
08:00	0.6447	3.2549	20:00	0.6451	3.2068
08:30	0.6449	3.2323	20:30	0.6451	3.1936
09:00	0.6451	3.1978	21:00	0.6452	3.1924
09:30	0.6454	3.1532	21:30	0.6451	3.2065
10:00	0.6458	3.1015	22:00	0.6448	3.2392
10:30	0.6462	3.0521	22:30	0.6444	3.2998
11:00	0.6465	3.0198	23:00	0.6438	3.4066
11:30	0.6465	3.0150	23:30	0.6429	3.5867

Table 9 – The pdf value and Gumbel value (Anomaly score) of appliance 60 from 00:00AM to 23:30PM

Fig. 11 shows the anomaly score for appliance 60 in a day for each 30 minutes and the Gumbel values for the set of f(x). We can set the threshold based on the pdf (0.644) or Gumbel value (3.4). Appliance 60 will be abnormal when its pdf smaller than 0.644 or anomaly score greater than 3.4. For this example, we just only compute anomaly score for appliance 60 with 48 times for each 30 minutes. We can identify that appliance 60 will be abnormal when it turns on between 23:00PM to 01:00 AM. For precision discussion, we will evaluate anomaly score for all appliances in a query pattern at a certain time (ex: current time)

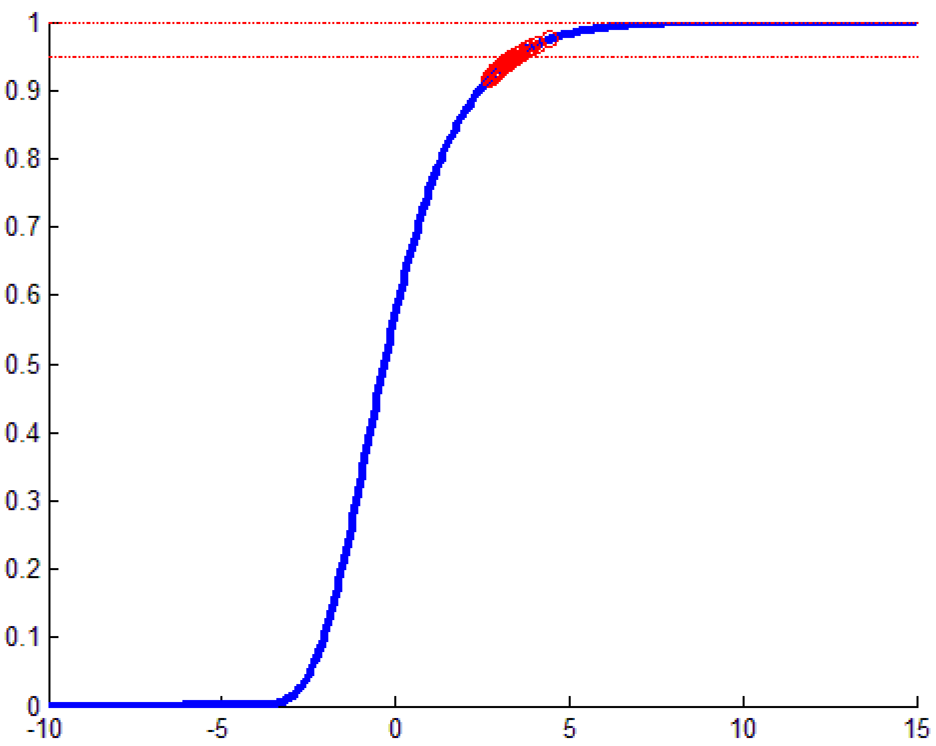

Fig. 11: Anomaly score and Gumbel values of f(x).

6.2 Performance on real world datasets

In this section, we describe our real-world dataset and show some correlation patterns that our algorithms discovered for each house. Although many smart home environment datasets are available, but little of them records the status along with space information for each appliance in smart home environment. Kolter et al. [18] collected the dataset REDD including detailed power readings of each appliance of six houses lasting for about five weeks. Therefore we can convert the raw data into suitable usage database. We use our sample house for the location information of each house. And we set the minimum space threshold fixed as 0.5 and mine the daily correlation patterns of each house.

In the following experiments, we compare the running time of EVMAS and TIAD method with threshold varied from 5% to 10% on A17-N624 dataset, while SPMAS test with query patterns varied length from 1 appliance to 6 appliances. A17–N624 dataset contains 624 events of appliance 17. Fig. 12 shows the running time of the three methods with different threshold and number of appliances in the query patterns. Obviously, when we continue to higher the threshold, the runtime for EVMAS and TIAD remain unchanged at around 56 (seconds). We can see that when the number of appliances increases, the processing time required for SPMAS increases. This is partly because EVMAS and TIAD use the probability density function information while SPMAS uses frequent sequences dataset. Many appliances in a query pattern lead to generate more number of combination sequences.

Fig. 13 shows the execution time of the three algorithms with threshold varied from 5% to 10% on A60–N682 dataset, while SPMAS test with query patterns varied length from 1 appliance to 6 appliances. A60–N682 dataset contains 682 events of appliance 60. From this figure, we can observe that SPMAS has the best running time performance. However, when the number of appliances in query patterns increases, number of combination sequences will be increased quickly. It leads to the runtime rising dramatically. On the contrary, EVMAS and TIAD stabilize at about 61 (seconds).

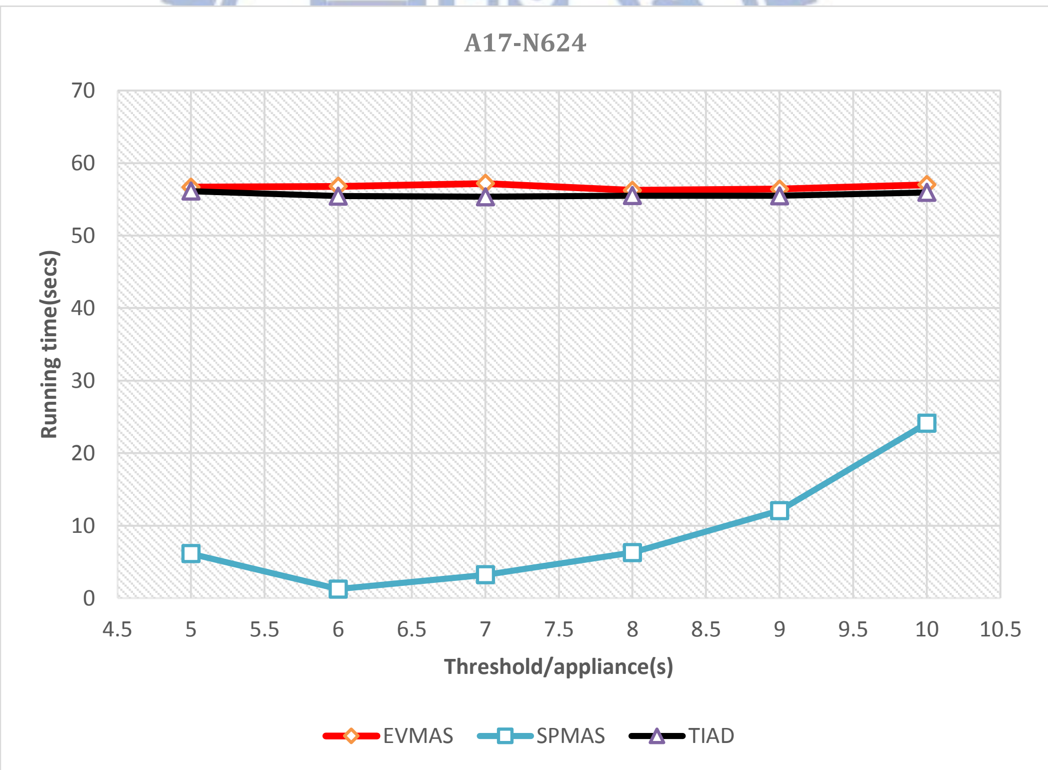

Fig. 12. Running time performance testing on A17-N624 dataset.

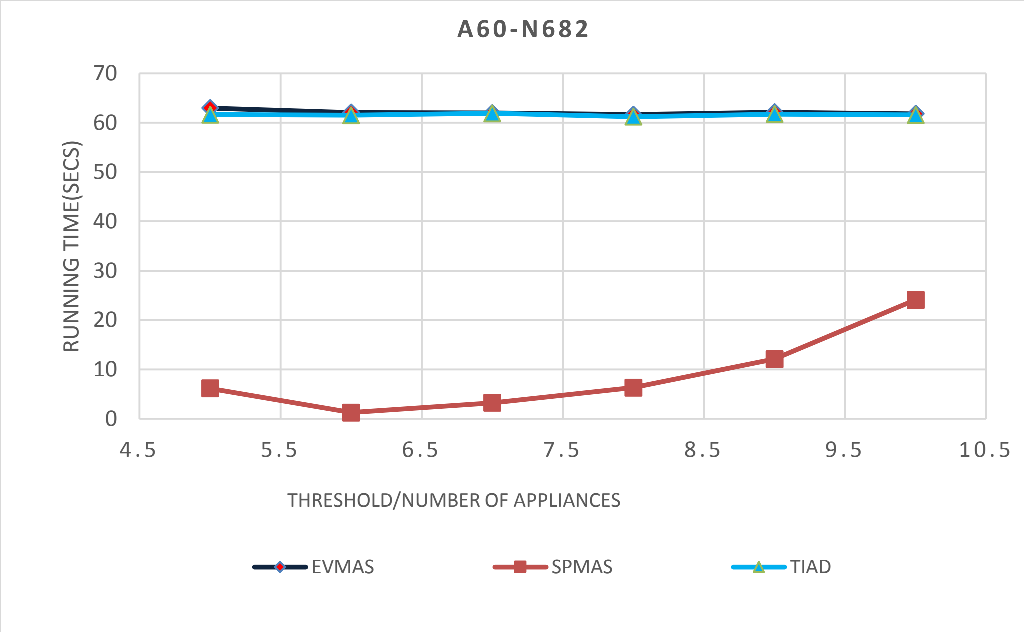

Fig. 13. Running time performance testing on A60-N682 dataset.

For testing the performances of anomaly detection, we have to generate the synthetic query because the original data does not label every day with normal or abnormal behavior. We take 1440 queries (1440 minutes per day), Q={t1, t2, …, ti|ti∊[0,24)}. As shows in Fig. 11, we can label normal or abnormal behavior for each minute. Fig. 14 shows the percentage of accuracy for each method. We can see that using probability density function information outperform taking frequent sequences. EVMAS and TIAD take more precise than that of SPMAS.

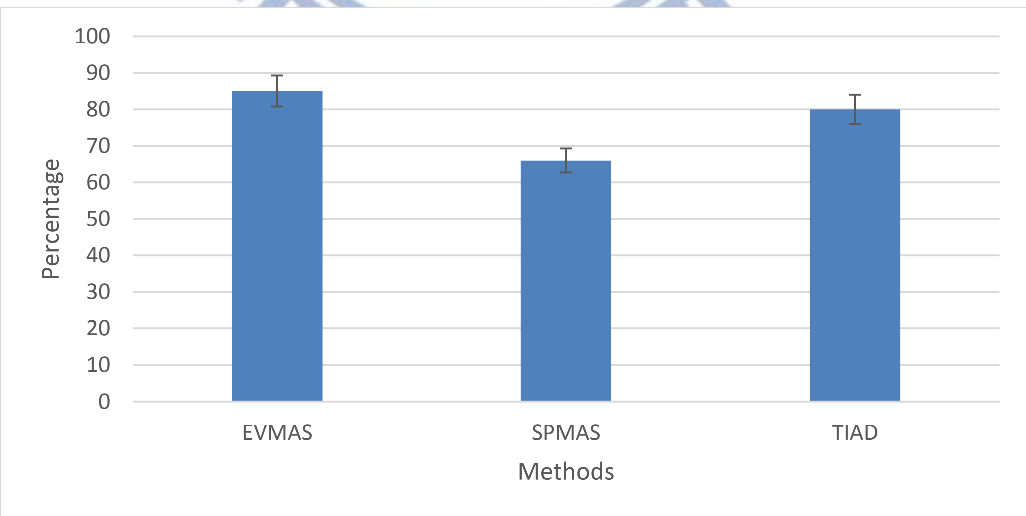

Fig. 14. Precision testing of the three methods on real world dataset.

6.3 Discussion on EVMAS

Anomaly detection using EVT approach is based on a model of normal behavior which was presented by the probability distribution function. There are five steps to execute the methods. First, we collect the information of appliances and then build the pdf for each appliance. This step will take a long time because it will read the original dataset and extract useful information for pdf. Second, we have to extract minimal value of y. We define where is the tails of the pdf and the collect the set of y that stay under the threshold. However, how to choose the best threshold is also a big problem for this method. We gradually increase value of threshold such that the area under the threshold equal to the size value (size=log(m)). This step takes the most running time of all steps in this method.

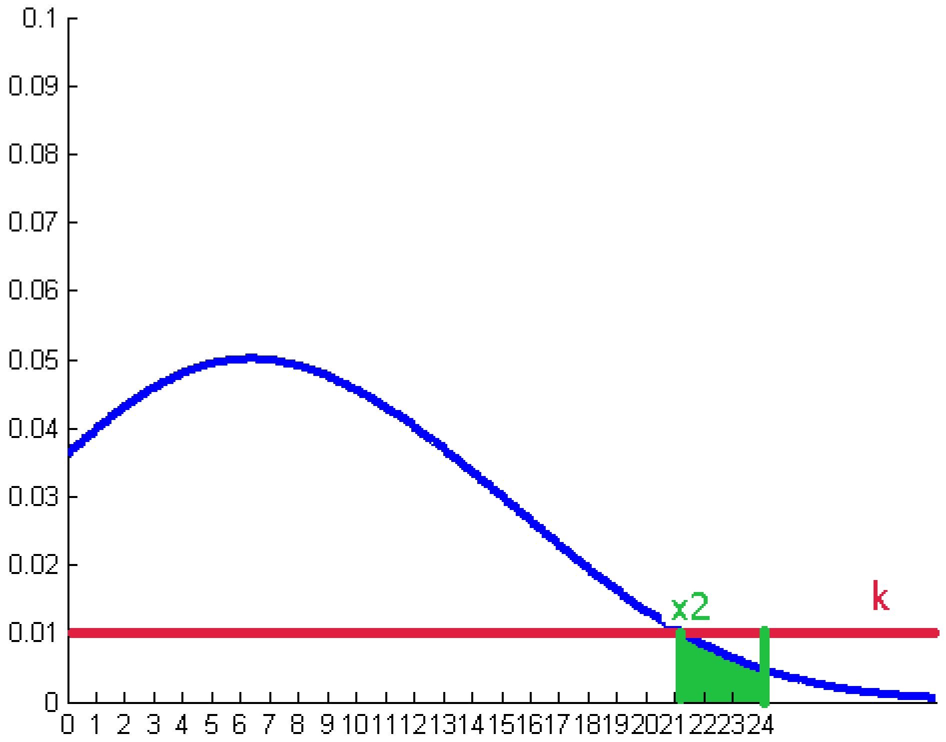

Fig. 15 Threshold and the area of threshold and axis.

However, it only take long time for training. When user appliances want to send a query pattern, it will respond in a short time because it just calculate the anomaly score for the query pattern.

About the performance of accuracy, we can see that EVMAS' accuracy is better than that of SPMAS and TIAD. It is really true because the anomaly score was calculated throughout the Weibull and Gumbel distribution. SPMAS just only use the pdf to determine the anomalies.

6.4 Discussion on SPMAS

We explore the frequent sequences of correlation patterns dataset to calculate the anomaly score for each appliance. The set of frequent sequences which can be used to determine abnormal events. We assume that all sequence patterns in this dataset are normal patterns. All appliances occurred in normal scenarios. The main disadvantage of this method is that the running time is fastest compare to EVMAS and TIAD. Thanks to this benefit, we can use this method for online system. It takes less time consumption than that of EVMAS and TIAD.

However, the first advantage of SPMAS is that we do not need time information of appliances to determine abnormal usage behavior since there is no time information in frequent sequences of correlation patterns. Another advantage of this method is that we do not pay attention to the order of appliance when compute anomaly scores because we have no order information in query pattern. However, we cannot identify exactly abnormal usage behaviors when there is only one appliance in the query pattern. If there are many appliances with low scores, we can use a threshold instead of using the minimum value. The accuracy of this method is also lower than that of EVMAS and TIAD.

6.5 Discussion on TIAD

Generating normal behavior time intervals using the pdf is easy way to define the normal time intervals. The advantage of this method is that we can use the threshold that we have done in EVMAS method. We save both running time and memory to find the threshold. The accuracy of this method is also better than that of SPMAS.

The potential drawback is that the quality of accuracy can be affected by how we define the intervals. However, we can minimize this possibility if we do not fixed-width time intervals. Instead determine the intervals based on the original dataset, we use the probability density function. The area under the curve from time t1 to time t2 gives us the probability that an appliance will turn on between t1 and t2. The running time of this method and EVMAS are nearly the same but the accuracy is a little bit lower than that of EVMAS.

Chapter 7

Conclusion

Recently, considerable concern has arisen over the electricity conservation due to the issue of greenhouse gas emissions. If abnormal behaviors of appliances usages are available, users can adapt abnormal behaviors information to conserve electricity effectively. In this paper, we explored the correlation patterns for abnormal detection and proposed three methods to detect abnormal usage behavior of appliances in a home. Our main method adapts the Extreme Value Theory on the tails of probability density function of appliances. Then, we can evaluate anomaly scores for appliances, and identify what the appliances usages are abnormal. One important thing is that the correlation patterns length need to be large enough for the Weibull and Gumbel distribution but too large values can lead to time-consuming. Other two methods use frequent sequences of correlation patterns to compute the proportion of occurrence of appliances in dataset and adapt probability density function to extract time intervals for normal periods, then we can determine what appliances usages are not normal at the time. The experimental studies indicate that our methods are efficient and precise. Moreover, EVMAS, SPMAS, and TIAD are applied on a real world dataset to show the practicability of abnormal detection.

Bibliography

[1] C. Chen and D. J. Cook, *Energy Outlier Detection in Smart Environments*, 2011.

[2] Y.-C. Chen, C.-C. Chen, W.-C. Peng, and W.-C. Lee, "Mining Correlation Patterns among Appliances in Smart Environment," 2013.

[3] Y.-C. Chen, Y.-L. Ko, W.-C. Peng, and W.-C. Lee, "Mining Appliance Usage Patterns in Smart Home Environment," in *Advances in Knowledge Discovery and Data Mining*. vol. 7818, J. Pei, V. Tseng, L. Cao, H. Motoda, and G. Xu, Eds., ed: Springer Berlin Heidelberg, 2013, pp. 99-110.

[4] A. Prudenzi, "A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recordings at meter panel," in *Power Engineering Society Winter Meeting, 2002. IEEE*, 2002, pp. 941-946 vol.2.

[5] C. Yi-Cheng, K. Yu-Lun, and P. Wen-Chih, "An Intelligent System for Mining Usage Patterns from Appliance Data in Smart Home Environment," in *Technologies and Applications of Artificial Intelligence (TAAI), 2012 Conference on*, 2012, pp. 319-322.

[6] M. A. Al-Fawzan, "Methods for Estimating the Parameters of the Weibull Distribution," October 2000.

[7] P. Bhattacharya, *Weibull Distribution for Estimating the Parameters*, 2011.

[8] A. C. Cohen, "Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples," *Technometrics*, vol. 7, pp. 579-588, 1965.

[9] H. L. Harter and A. H. Moore, "Point and Interval Estimators, Based on m Order Statistics, for the Scale Parameter of a Weibull Population with Known Shape Parameter," *Technometrics*, vol. 7, pp. 405-422, 1965.

[10] H. L. Harter and A. H. Moore, "Maximum-Likelihood Estimation of the Parameters of Gamma and Weibull Populations from Complete and from Censored Samples," *Technometrics*, vol. 7, pp. 639-643, 1965.

[11] J. Candamo, M. Shreve, D. B. Goldgof, D. B. Sapper, and R. Kasturi, "Understanding Transit Scenes: A Survey on Human Behavior-Recognition Algorithms," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, pp. 206-224, 2010.

[12] C. Chen, B. Das, and D. J. Cook, "A Data Mining Framework for Activity Recognition in Smart Environments," presented at the Proceedings of the 2010 Sixth International Conference on Intelligent Environments, 2010.

[13] H.-H. Hsu and C.-C. Chen, "RFID-based human behavior modeling and anomaly detection for elderly care," *Mob. Inf. Syst.*, vol. 6, pp. 341-354, 2010.

[14] S. C. Chin, A. Ray, and V. Rajagopalan, "Symbolic time series analysis for anomaly detection: a comparative evaluation," *Signal Process.*, vol. 85, pp. 1859-1868, 2005.

[15] A. Lotfi, C. Langensiepen, S. Mahmoud, and M. J. Akhlaghinia, "Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour," *Journal of Ambient Intelligence and Humanized Computing*, vol. 3, pp. 205-218, 2012/09/01 2012.

[16] K. Hara, T. Omori, and R. Ueno, "Detection of unusual human behavior in intelligent house," in *Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on*, 2002, pp. 697-706.

[17] V. J. a. D. J. Cook, "Anomaly Detection Using Temporal Data Mining in a Smart Home Environment," *Methods of Information in Medicine, Smart Homes and Embient Assisted Living special issue.*, 2008.

[18] V. Jakkula, D. J. Cook, and A. S. Crandall, "Temporal pattern discovery for anomaly detection in a smart home," in *Intelligent Environments, 2007. IE 07. 3rd IET International Conference on*, 2007, pp. 339-345.

[19] V. R. Jakkula and D. J. Cook, "Detecting Anomalous Sensor Events in Smart Home Data for Enhancing the Living Experience," *Artificial Intelligence and Smarter Living,* vol. 11, p. 07, 2011.

[20] W. Mao-Yung, W. Chao-Lin, L. Ching-Hu, Y. Hui-Wen, and F. Li-Chen, "Context-aware home energy saving based on Energy-Prone Context," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, 2012, pp. 5233-5238.

[21] A. C. Tran, S. Marsland, J. Dietrich, H. W. Guesgen, and P. Lyons, "Use cases for abnormal behaviour detection in smart homes," presented at the Proceedings of the Aging friendly technology for health and independence, and 8th international conference on Smart homes and health telematics, Seoul, Korea, 2010.

[22] L. Farinaccio and R. Zmeureanu, "Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses," *Energy and Buildings,* vol. 30, pp. 245-259, 8// 1999.

[23] A. O. a. M. B. H. Goncalves, "Unsupervised Disaggregation of Appliances using Aggregated Consumption Data," *KDD workshop on Data Mining Applications in Sustainability (SustKDD'11),* 2011.

[24] O. A. a. V. Negru, "A Methodology for Household Appliances Behavior Recognition in AmI Systems Integration," *Proceedings of 7th International Conference on Automatic and Autonomous Systems (ICAS'11),* pp. 175-178, 2011.

[25] T. Kato, H. Cho, D. Lee, T. Toyomura, and T. Yamazaki, "Appliance Recognition from Electric Current Signals for Information-Energy Integrated Network in Home Environments," in *Ambient Assistive Health and Wellness Management in the Heart of the City*. vol. 5597, M. Mokhtari, I. Khalil, J. Bauchet, D. Zhang, and C. Nugent, Eds., ed: Springer Berlin Heidelberg, 2009, pp. 150-157.

[26] S. Luca, P. Karsmakers, K. Cuppens, T. Croonenborghs, A. Van de Vel, B. Ceulemans, *et al.*, "Detecting rare events using extreme value statistics applied to epileptic convulsions in children," *Artificial Intelligence in Medicine,* vol. 60, pp. 89-96, 2// 2013.

[27] V. O. Andreev, Tinyakov, Sergey E.,Okunev, Oleg B., "Extreme Value Theory and Peaks over Threshold Model: An Application to the Russian Stock Market," *New York Science Journal,* 2012.

[28] D. Clifton, S. Hugueny, and L. Tarassenko, "Novelty Detection with Multivariate Extreme Value Statistics," *Journal of Signal Processing Systems,* vol. 65, pp. 371-389, 2011/12/01 2011.

[29] S. J. Roberts, "Novelty detection using extreme value statistics," *Vision, Image and Signal Processing, IEE Proceedings -*, vol. 146, pp. 124-129, 1999.

[30] S. J. Roberts, "Extreme value statistics for novelty detection in biomedical signal processing," in *Advances in Medical Signal and Information Processing, 2000. First International Conference on (IEE Conf. Publ. No. 476)*, 2000, pp. 166-172.

[31] M. Smith, S. Reece, S. Roberts, and I. Rezek, "Online Maritime Abnormality Detection Using Gaussian Processes and Extreme Value Theory," in *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, 2012, pp. 645-654.

[32] D. A. Clifton, S. Hugueny, and L. Tarassenko, "Novelty Detection with Multivariate Extreme Value Statistics," *J. Signal Process. Syst.*, vol. 65, pp. 371-389, 2011.

[33] L. Tarassenko, A. Hann, A. Patterson, E. Braithwaite, K. Davidson, V. Barber, *et al.*, "BIOSIGN™ : multi-parameter monitoring for early warning of patient deterioration," in *Medical Applications of Signal Processing, 2005. The 3rd IEE International Seminar on (Ref. No. 2005-1119)*, 2005, pp. 71-76.

[34] M. Lauer, "A Mixture Approach to Novelty Detection Using Training Data with Outliers," in *Machine Learning: ECML 2001*. vol. 2167, L. Raedt and P. Flach, Eds., ed: Springer Berlin Heidelberg, 2001, pp. 300-311.

[35] A. Nairac, T. A. Corbett-Clark, R. Ripley, N. W. Townsend, and L. Tarassenko, "Choosing an appropriate model for novelty detection," in *Artificial Neural Networks, Fifth International Conference on (Conf. Publ. No. 440)*, 1997, pp. 117-122.

[36] L. Tarassenko, P. Hayton, N. Cerneaz, and M. Brady, "Novelty detection for the identification of masses in mammograms," in *Artificial Neural Networks, 1995., Fourth International Conference on*, 1995, pp. 442-447.