2014 IEEE International Conference on System Science ang Engineering (ICSSE)
July 11-13 2014,Shanghai, China

EM-LDA Model of User Behavior Detection for Energy Efficiency

Zhibin Zhao*, Weisheng Xu**, Dazhang Chen
* College of Electronics and Information Engineering, Tongji University, Shanghai, China
** Nanchang University, Jiang Xi, China
eric.zhao.zhibin@gmail.com

Abstract—**In energy efficient analysis, user behavior detection related to the dynamic demands of energy is a critical aspect to support the intelligent control schema of Building Management System. In this paper, anomalous occupancy of user behavior tends to be figured out from multiple time-series of occupancy record. The problems in this issue include the time-stamp detection and time-span identification of anomaly events. Most inference model based on Markov Chain can illustrate the time-stamp detection problem reasonably, but the time-span identification problem is just vaguely explained. Therefore, a Latent Dirichlet Allocation (LDA) model is declared to figure out those two problems efficiently. First, the discrete data of occupancy are expressed as mixture model of Poisson distribution, and are transformed to a dataset with several semantic concepts via Expectation-Maximization Algorithm. Then, the denotation of LDA components (including the words, the topic, the document, and the relevant parameters and hyper-parameters) are illustrated, according to the semantic dataset. Finally, particle filter algorithm is leveraged to sample latent variable of topic, according to the conditional posterior probability of word for specific topic. After iterations, the probability of samples is closely approximated the true marginal distribution of words with specific topic. Through the relation matrix of words and topic, the most possible topic can be explained for the specific document. If a document's topic is different with other document's topic, this document can be identified as a bias of point anomaly (noting generally the amount of topics setup to two). Due to a word can involve several time-stamps of the time-series in a time, other contextual anomalies nearby the point anomaly can be marked, and they are the notation of time-spans for anomalous events. With a step by step along the time-series, all time-stamps can be ergodic as the documents, then all the contextual anomalies can be explained as following the happening of point anomalous event.**

Keywords-Topic Model, Latent Dirichlet Allocation, Particle Filter, User Behavior Detection, Anomaly Detection, Energy Efficiency

I. INTRODUCTION

Nowadays, the main control schema of Intelligent Building Management System (IBMS) involving to the energy efficiency is to setup the time schedule of building subsystem, such as HVAC system, or to online-adjust the parameters of equipment dynamically with the variation of user's demands. However, the hybrid of static control schedule and dynamic self-adjustment is a popular and reasonable strategy for the balance between user's demands and energy efficiency. Dynamic control strategy, flexible management culture, and on-demand user response represent the personality feature of modern IBMS in each building. In the ubiquitous computing environment, the system is not just driven by the sensing value, but the deriving contextual information about reasons, intentions, desires and beliefs of the users.[1] Actually, user's demands are explicit representation of user behaviors involving to the energy requirement. Group behavior of users is semi-periodic in most public building, and is possible to predict the states of specific behavior. Modeling user behaviors is the attractive and popular field of building intelligence research, especially, the issues pertinent to the building energy efficiency.[2]–[5]

User Behavior Detection for energy efficiency of this paper is to detect the burst anomalies in the time-series of user behaviors, which induce the surge of requirement to bias control schedule of IBMS seriously. Thus, when the anomalies happened, the static control schema will either underestimate the demands of users or waste the energy for low requirement. However, the anomaly detection for user behavior is not just a point anomaly detection issue generally, but a hybrid of point anomaly and contextual anomaly.[6] According to the daily records of energy consumption and peak energy consumption in a building, Seem leverage the outlier detection methods to figure out abnormal energy, if the energy consumption for a particular day is significantly different than previous energy consumption.[7], [8] Ihler detected the contextual anomalies from a time series of arrival and departure of CalIt2 Building, according to MMNHPP approach.[9], [10]

In this paper, a Latent Dirichlet Allocation (LDA) model is declared to resolve the contextual anomaly detection from multi-dimensional dataset of user behaviors. First, all multidimensional discrete data points are clustered to five classes and give semantic definitions for these five classes, including Very high probability (V), High probability (H), Mild probability (M), Low probability (L), Tiny probability (T). Then, according to the density estimation of five classes, the discrete data time series will be sampled to construct a time series of characters. Based on multiple character time series, the elements of LDA, including words (2 characters), documents with each timestamp per week and corpus with all documents, can be organized. After that, LDA model with 2 topics will be leveraged to estimate the density of topics in each document. During the process of computation, approximate approach utilizes to generate the samples to approximate the density of posterior probability of topics in specific document.

This paper is organized into 5 sections. After Introduction section, section 2 is to state the problems and

978-1-4799-4367-8/14/$31.00 ©2014 IEEE

claim the challenges of these problems. Section 3 illustrates the methodology of LDA algorithm, and represent the implementation of LDA. Section 4 describe a semantic process from multi-dimensional discrete data to semantic concepts, and explain the LDA modeling of user behavior detection. Section 5 make a conclusion and describe a perspective in the future.

II. PROBLEM STATEMENT

A. Contextual Anomaly Detection

Anomaly can be classified into three categories, including point anomaly, contextual anomaly, and collective anomaly. If an individual data instance can be considered as anomalous with respect to the rest of data, then the instance is termed a point anomaly. This is the simplest type of anomaly and is the focus of majority of research on anomaly detection. [6], [11] If a data instance is anomalous in a specific context, but not otherwise, then it is termed a contextual anomaly (also referred to as conditional anomaly). Contextual anomaly has to be identified from two attributes: behavioral attribute and contextual attribute.[12] In time-series data, time is a contextual attribute that determines the position of an instance on the entire sequence.[12] Behavioral attribute represents the non-contextual features of the instance, which focus on its own timestamp scenario and ignore the influence from other relevant timestamp. On the other side, contextual attribute mainly describes influence from other relevant time stamp.

During the process of anomaly detection, there are always some obstacles on the road. Firstly, the position of anomalous observation is close to the boundary of normal region, so that the explicit declaration of abnormal data is not coherent and natural. Secondly, the features of normal behavior are evolving continuously, and the conditions of normal identification need to adjust anyway. After that, the noises in the scenario of anomaly detection are difficult to distinguish to actual anomaly data. Finally, there are quite different definitions for same styles of anomaly for different application domains, so that a generalized approach of anomaly detection is not ease to develop for all scenarios.

B. Challenge of User Behavior Detection

Two challenges of user behavior detection for energy efficiency are discussed in this paper. First challenge is to figure out vague point anomalies from timeseries of user behavior. Generally, the obvious point anomalies, which obviously bias from the collection of normal data, are not complicated to discover precisely. However, some anomalies are not deviated from the normal region overtly, so that identification of this kind of anomalies is hard to declare explicitly. Moreover, the features of some observations are normal in one determinative phase, but identified as abnormal in other phase. Therefore, a self-adjusted approach of anomaly detection is needed to accommodate the transfer of different phases.

Second challenge is timespan recognition of anomaly, which describes the duration of anomaly behaviors happened. Generally, point anomaly detection of user behavior represents as interrupted results in time series, which only display the anomalies in its timestamps, can't display the contextual anomalies. Some anomalies need to be identified via contextual environment. The approach of contextual anomaly detection is rational to mark the timespan of anomaly, according to the feature of point anomaly in specific timestamp and the contextual attributes related to nearby timestamps.

III. LATENT DIRICHLET ALLOCATION

A. Description of LDA Model

Blei describes a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics.[13] Via the latent topic variables, the marginal probability of words per document in LDA model, which hard to resolve for a large scale of words and documents, can reduce the dimension of data, and decompose the marginal density estimation to ease the computation.

LDA model tend to resolve and evaluate three kinds of problem domains, including Document modeling, Document classification, collaborative filtering[13]. Document modeling problem tend to estimate the empirical marginal density of words sequence per document. Compared to probability, Latent Semantic Indexing, LDA model can avoid serious overfitting issues. Document classification problem tends to classify the document into two or more mutually exclusive class. Collaborative filtering problem is to make automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating). [13]

As the description of Figure 1, LDA models each of D documents as a mixture over K latent topics, each of which describe a multinomial distribution over a W word vocabulary, and each topic is a probability distribution over a finite vocabulary of words[13][14]. Topic z_i receives weight θ_j in document j, and word ω_i has probability with parameter ϕ_{zi} in topic z_i. The generative model can assume that documents are produced by independently sampling z_i for each word from θ_j and then independently sampling a topic z_i for each word from ϕ_{zi}.[15]

The complete probability model is:

$$\begin{aligned} &\omega_i \mid z_i, \phi_{z_i} \sim Multinomial\left(\phi_{z_i}\right), i = 1,\ldots, N, \\ &\phi_z \sim Dirichlet\left(\beta\right), z = 1,\ldots, K, \\ &z_i \mid \theta_j \sim Multinomial\left(\theta_j\right), i = 1,\ldots, N, \\ &\theta_j \sim Dirichlet\left(\alpha\right), j = 1,\ldots, D, \end{aligned} \tag{1}$$

where N is the total number of words in all documents, T is the number of topics, D is the number of documents, d_i are the document of the word w_i, and z_i are the topic of w_i.[13]–[15]

The goal of inference in this model is to identify the values of ϕ and θ, given a corpus with a sequence of words $w_N = \left(\omega_1, \ldots, \omega_N\right)$. In order to discover the set of topics used in a corpus w_N, where each word ω_i belong to some document d_i, an estimation of ϕ can represent a high probability to the words that appear in the corpus. The estimation problem of ϕ is to attempt to maximize $P\left(w \mid \phi, \alpha\right) = \int P\left(w \mid \phi, \theta\right) P\left(\theta \mid \alpha\right) d\theta$, where

$P(\theta)$ is a symmetric Dirichlet distribution with single parameter α. [16]–[19]

B. Collapsed Gibbs Sampling of LDA Model

Generally, the integral $P(w|\phi,\alpha)$ is intractable, so that some approaches of approximation are leveraged to estimate ϕ, including collapsed Gibbs sampling[16], [18], [19], expectation propagation, particle filter[18], [20], variational Bayes.[17] Griffiths advices the collapsed Gibbs sampling method to approximate the posterior probability $P(z|w)$, which is a classical and reasonable approach for LDA model.

Equation (2) illustrate the principal of Collapsed Gibbs Sampling approach of LDA model, and the samples can obtain from $P(z_i = j|z_{-i}, w)$ in equation (3).[16], [19]

$$
\begin{aligned}
&P(z|w) \propto P(w|z)P(z) \\
&P(\omega_i|z_i = j, z_{-i}, w_{-i}) = \int P(\omega_i|z_i = j, \phi^j)P(\phi^j|z_{-i}, w_{-i})d\phi^j \\
&P(\phi^j|z_{-i}, w_{-i}) \propto P(w_{-i}|\phi^j, z_{-i})P(\phi^j) \\
&P(\omega_i|z_i = j, z_{-i}, w_{-i}) = \frac{n^{\omega_i}_{-i,j} + \beta}{n^{(\cdot)}_{-i,j} + W\beta}
\end{aligned} \quad (2)
$$

$$
\begin{aligned}
&P(z_i = j|z_{-i}, w) \propto \frac{n^{\omega_i}_{-i,j} + \beta}{n^{(g)}_{-i,j} + W\beta} \frac{n^{d_i}_{-i,j} + \alpha}{n^{d_i}_{-i} + K\alpha} \\
&P(z_i = j|z_{-i}) = \int P(z_i = j|\theta^{d_i})P(\theta^{d_i}|z_{-i})d\theta^{d_i} \\
&P(z_i = j|z_{-i}) = \frac{n^{d_i}_{-i,j} + \alpha}{n^{d_i}_{-i,g} + K\alpha}
\end{aligned} \quad (3)
$$

According to the samples set of z, ϕ and θ can be estimated from equation (4).[16], [18], [19]

$$
\begin{aligned}
\hat{\phi}^{\omega}_j &= \frac{n^{\omega_i}_j + \beta}{n^{(g)}_j + W\beta} \\
\hat{\theta}^d_j &= \frac{n^{d_i}_j + \alpha}{n^{d_i}_g + K\alpha}
\end{aligned} \quad (4)
$$

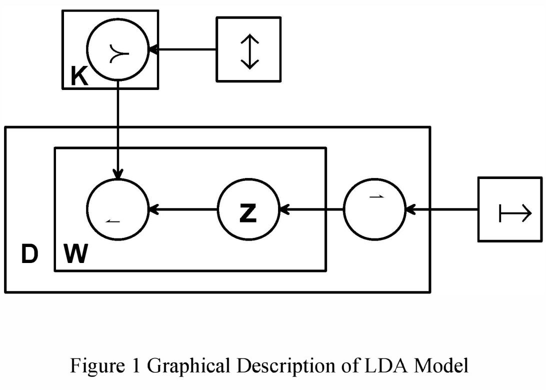

Figure 1 Graphical Description of LDA Model

IV. EM-LDA MODEL OF USER BEHAVIOR DETECTION

The LDA model of user behavior detection tends to figure out the topic estimation of in each timestamp collection per week, and it can be considered as document classification. Generally, the user behavior for energy efficiency in the building consists of two categories, user behavior of existing, and user behavior of reaction. In this paper, user behavior of existing is discussed via empirical dataset, CalIt2 Building People Count Dataset.

A. CalIt2 Dataset

Calit2 Building People Count Data Set is a public data set from UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/), which record the people count at the main door of Calit2 Building, UCI. The whole observation contains 2 data streams, 10,080 instances, people's arrival or departure count of the building, over 15 weeks, 48 times slices per day. There is also a basic truth of events in the data set, which records the anomalous events such as conferences in the building relevant to the unusually high people counts for that day/time period, to validate the efficiency of prediction. The data of observations are generated by a pair of battery powered optical detectors that measure the presence and direction of objects as they pass through the building's main set of doors. The number of count in each direction is then communicated via a wireless link to a base station with internet access, at which they are stored. In the observation data set, the record of people count is organized into 4 attributes, including the Flow ID (in or out), Date, Time, and Number of counts for previous half hour. In the events record, there are 4 attributes, including Date, Time of event begin, Time of event end, and Name of event. However, the non-periodic activities or anomalous events obviously increase the count before the events and leaving after the events. The events with conspicuous abnormal count are easily specified, but the other events with implicit features of anomaly need to mark out from the relevant contextual environment.

B. Mixture of Poisson and EM

Known as a classic algorithm, Expectation-Maximization (EM) is developed by Dempster, to estimate the maximum likelihood estimation (MLE) and Maximum a posterior estimation (MAP) [21]. EM algorithm always utilizes to resolve the cluster problem of non-parametric density estimation with fixed number of components[11], [22], [23]. For the user behavior of existing, its non-parametric density estimation always chooses Gaussian Mixture or Poisson Mixture. Due to the arrival and departure data, Poisson kernel will be better to fit the density probability of count data, and sparse matrix with anomaly is ill-conditioned for Gaussian kernel. Therefore Poisson Mixture is the first choice for the scenario of user behavior detection. Equation (5) shows the form of Poisson Mixture:

$$
P(x) = \sum_{k=1}^{K} \omega_k Poi(x|\lambda_k) \quad (5)
$$

Where K is the number of kernel in the mixture, λ is the parameter of Poisson distribution, and ω is the scale parameter of mixture. In this paper, the mixture is setup as $k \in \{0, K, K\}$, and means different semantic concepts.

EM approach estimates the parameters of Poisson mixture via construct a latent variable z, and based on density of z, compute the expectation and maximize the expectation via new parameters recursively until the estimation converged. The Algorithm describes as following:

Problem: figure out Maximum Likelihood Estimation

$\arg\max P(X|\theta)$ or $\arg\max \log P(x|\theta)$

Analysis:

$$\ln P(X|\theta) = \ln\left(\sum_Z P(Z|\theta)P(X|Z,\theta)\right)$$

$$\ln\left(\sum_Z P(Z|\theta)P(X|Z,\theta)\right) \geq \sum_Z P(Z|\theta)\ln P(X|Z,\theta)$$

$$\sum_Z P(Z|\theta)\ln P(X|Z,\theta)$$

$$= \sum_Z P(Z|\theta)\ln\left[\frac{P(X,Z|\theta)}{P(Z|\theta)}\right] + \sum_Z P(Z|\theta)\ln\left[\frac{P(Z|\theta)}{P(Z|X,\theta)}\right]$$

$$KL = \sum_Z P(Z|\theta)\ln\left[\frac{P(Z|\theta)}{P(Z|X,\theta)}\right] \geq 0$$

Tend to maximize $L = \sum_Z P(Z|\theta)\ln\left[\frac{P(X,Z|\theta)}{P(Z|\theta)}\right]$,

then $KL = 0$, Given $Q(\theta) = \sum_Z P(Z|X,\theta^{old})\ln P(X,Z|\theta)$

Resolution:

1. Initialize θ^{0} , and compute $Q(\theta^{0})$
2. Step E: $P(Z|X,\theta^{old})$
3. Step M: $\theta^{new} = \arg\max Q(\theta|\theta^{old})$
4. Recursion: Step E and Step M

Until $Q(\theta^{K+1}|\theta^{K}) - Q(\theta^{K}|\theta^{K-1})$ converged, θ^{K} is the answer

According to the description of EM General Process, the EM for Poisson Mixture (EMPM) is described as following:

Problem: Given $X = \{x_1,\ldots,x_n\}$, find ω_k and λ_k

From $P(X|\theta) = \prod_{i=1}^{N}\sum_{k=0}^{K}\omega_k Poi(x_i|\lambda_k)$

Resolution:

1. Initialize ω_k^{0} and λ_k^{0} , and compute

   $\ln P(X|\omega^0,\lambda^0) = \sum_{i=1}^{N}\ln\sum_{k=0}^{K}\omega_k^0 Poi(x_i|\lambda_k^0)$

2. Step E: According to ω_k and λ_k

   calculate $P(z_k = 1|x_i)$

3. Step M: $\omega_k^{new} = \dfrac{\sum_{i=1}^{N} P(z_k = 1|x_i)}{N}$

   $\lambda_{km}^{new} = \dfrac{\sum_{i=1}^{N} P(z_k = 1|x_i)x_{im}}{\sum_{i=1}^{N} P(z_k = 1|x_i)}$

4. Recursion: Step E and Step M

Until $\ln P(X|\omega_k^{new},\lambda_k^{new}) - \ln P(X|\omega_k,\lambda_k)$ converged,

ω_k and λ_k are the answer

C. Sementic Resampling

Considering the Callt2 Dataset with 2 dimensional time series, the EMPM is utilized to classify the data in each day per week to 5 categories, such as the data in all Monday are clustered into 5 categories, or 5 semantic concepts, including Very high probability (V), High probability (H), Mild probability (M), Low probability (L), and Tiny probability (T), showing as figure 2.

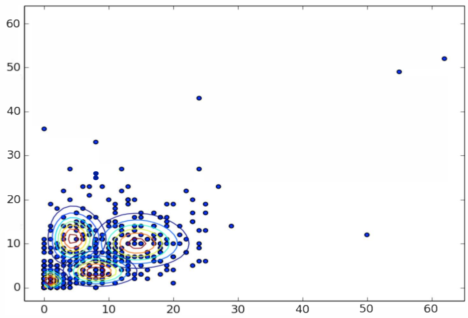

Figure 2 Categories from cluster of Monday dataset from Callt2 via EMPM

Due to the Poisson Mixture with 5 components, and all parameters already get via EMPM process, the probability of count at specific timestamp can be calculated. Given the proportion of 5 components, Monte Carlo approach is used to resample character at specific timestamp[24], [25]. The process describes as following:

Poisson Mixture (PM) of each day per week with 5 components:

$$P(X|\theta) = \prod_{i=1}^{N}\sum_{k=0}^{4}\omega_k Poi(x_i|\lambda_k)$$

Given $X = \{x_1,\ldots,x_n\}$, ω_k and λ_k , the proportion $c_k = (c_1,\ldots,c_5)$ of character, and $\sum_k c_k = 1$

(V, H, M, L, T) can be calculated from PM.

Resolution:

1. Sampling $u_m \sim Uniform(0,1)$ for M samples.
2. According to u_m and c_k^i , resampling the character (V,H,M,L,T) at specific timestamp i.

D. LDA Model and Collapse Gibbs Sampling

According to the character time series with 5 semantic concepts (V, H, M, L, and T), the elements of LDA model attempt to be constructed, as following figure 2. In this model, words are organized by characters at 2 timestamps $t_i \in \{t_1, K, t_n\}$, and document $W_j = \{\omega_1, K, \omega_v\}$ is along the timestamps per week in time series, corpus D_m of document collections consists of the documents of specific timestamp from all weeks. In CalIt2 dataset, $n = 5040$, $j = 15$, $m = 335$, v is determined by the times of recursive resampling.

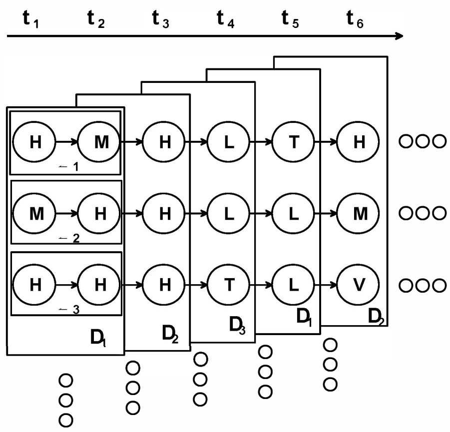

Figure 2 Structure of LDA model for energy efficiency

Along the timestamps, corpus D_m has the collection with documents $W_{mj} \in \{W_{m1}, K, W_{m15}\}$, the posterior of topic $P(z_{mj} | W_{mj})$ tend to figure out to explain the proportion of topic in the document W_{mj}. If the topic of one document is quite different with other documents, this document will be identified as abnormal document. The relevant timestamps will be marked to be anomaly. The length of time span of event happened is the length of words in abnormal document.

The inference process of LDA model describes as following[14]–[16]:

Condition: Given word $X = \{x_1, K, x_n\}$, hyperparameters α, β topic number $K = 2$, document number in corpus m, and word number v; count variables n_m^k and n_k^v represent the number of topic k in document m, and the number of words v involving topic k, respectively.

$\sum_k n_m^k = n_m$ and $\sum_v n_k^v = n_k$

$P(z_{m,v} | w, z_{-(m,v)}, \alpha, \beta)$

Problem: obtain topic z, parameters ϕ, θ

Resolution:

Initialize all count variables to 0:
for all documents :
  for all words ω_v in document d_m :
    sample $z_{mv}^k \sim Multinomial(0.5)$;
    $n_m^k = n_m^k + 1$;
    $n_m = n_m + 1$
    $n_k^v = n_k^v + 1$
    $n_k = n_k + 1$

While not finished:
  For all documents :
    For all words ω_v in documents d_m :
      $n_m^k = n_m^k - 1$; $n_m = n_m - 1$;
      $n_k^{w_{m,v}} = n_k^{w_{m,v}} - 1$; $n_k = n_k - 1$
      sample topic index $\bar{k} \sim P(z_{m,v} | w, z_{-(m,v)}, \alpha, \beta)$
      $n_m^{\bar{k}} = n_m^{\bar{k}} + 1$; $n_m = n_m + 1$;
      $n_{\bar{k}}^{w_{m,v}} = n_{\bar{k}}^{w_{m,v}} + 1$; $n_{\bar{k}} = n_{\bar{k}} + 1$
  if converged and iterations L times:
    set ϕ, θ as equation (4)
  evaluate topic samples z of document d to choose the normal feature.
  set document d_m to anomaly or not, according to the z

V. CONCLUSIONS

In this paper, the problems involving anomaly detection of user behavior detection are stated, and the challenges are analyzed in details. An EM-LDA model was leveraged to inference contextual anomaly of user behavior for energy efficiency from a discrete dataset of CalIt2. Firstly, the time series of discrete dataset was transformed to a time series of character with specific semantic concepts via EM approach. After that, the elements of LDA were defined, including the word, document, and corpus, based on the time series of character. Finally, the timespan of anomaly happened also marked via the word structure of LDA, and can adjust the word structure to intensify or diminish the relationship of timestamp in contextual environment. As a generative model, this LDA use collapsed Gibbs Sampling approach to sample the latent variable z to estimate the model parameters ϕ and θ. According to the density estimation of topic for specific document, the main topic of corpus can be identified, and it is the normal feature of anomaly detection. If the irrelevant topic was figure out from specific document, this document can be identified as abnormal document, and all timestamps in the document are considered as anomaly. In the perspective of future work, a self-adjusted word structure will be implemented to LDA model, in order to resolve the evolving timespan of events for changeable contextual environment, and the detail of experimental results will be illustrated.

ACKNOWLEDGMENT

I really appreciate for the support and encouragement from my supervisor Prof. Xu Weisheng, Tongji Univeristy firstly. The guidance from Prof. Chen Dazhang, Tongji University, is also indispensable for my research process. Then, I have to mention Prof. David A. Bradley, Univeristy of Abertay Dundee, to thank the advice involving to anomaly detection. All helps from colleague and classmate, Qu Haini, Lu Qianli, Chen Yan etc., give me the reason to advance my research to new level.

Finally, I would like to thank for the support from my wife, Wang Zhili.

REFERENCES

[1] D. Bruckner and R. Velik, "Behavior Learning in Dwelling Environments With Hidden Markov Models," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 11, pp. 3653–3660, Nov. 2010.

[2] R. Bleischwitz, N. Bader, J. Stephenson, B. Barton, G. Carrington, D. Gnoth, R. Lawson, and P. Thorsnes, "Energy cultures: A framework for understanding energy behaviours," *Energy Policy*, vol. 38, no. 10, pp. 6120–6129, 2010.

[3] C. Harris and V. Cahill, "Exploiting user behaviour for context-aware power management," in *WiMob'2005), IEEE International Conference on Wireless And Mobile Computing, Networking And Communications, 2005.*, 2005, vol. 4, pp. 122–130.

[4] T. Yu, "Modeling Occupancy Behavior for Energy Efficiency and Occupants Comfort Management in Intelligent Buildings," in *2010 Ninth International Conference on Machine Learning and Applications*, 2010, pp. 726–731.

[5] Z. B. Zhao, W. S. Xu, and D. Z. Cheng, "User behavior detection framework based on NBP for energy efficiency," *Automation in Construction*, vol. 26, pp. 69–76, 2012.

[6] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, Jul. 2009.

[7] J. Seem, "Pattern recognition algorithm for determining days of the week with similar energy consumption profiles," *Energy and buildings*, vol. 37, no. 2, pp. 127–139, Feb. 2005.

[8] J. E. Seem, "Using intelligent data analysis to detect abnormal energy consumption in buildings," *Energy and Buildings*, vol. 39, no. 1, pp. 52–58, 2007.

[9] A. Ihler, J. Hutchins, and P. Smyth, "Learning to detect events with Markov-modulated poisson processes," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 3, p. 13–es, Dec. 2007.

[10] A. Ihler, J. Hutchins, and P. Smyth, "Adaptive event detection with time-varying poisson processes," *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*, p. 207, 2006.

[11] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques, Third Edition*, Third Edit. Waltham, USA: Morgan Kaufmann, 2011.

[12] X. Song, M. Wu, C. Jermaine, and S. Ranka, "Conditional Anomaly Detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 5, pp. 631–645, May 2007.

[13] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, no. 4–5, pp. 993–1022, 2003.

[14] I. Porteous, D. Newman, A. Asuncion, and M. Welling, "Fast Collapsed Gibbs Sampling For Latent Dirichlet Allocation Categories and Subject Descriptors."

[15] K. Canini, L. Shi, and T. Griffiths, "Online inference of topics with latent Dirichlet allocation," *12th International Conference on Artificial Intelligence and Statistics 2009*, no. 1999, pp. 65–72, 2009.

[16] T. L. Griffiths and M. Steyvers, "Finding scientific topics.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101 Suppl, no. January, pp. 5228–35, Apr. 2004.

[17] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, no. 4–5, pp. 993–1022, 2003.

[18] K. Canini, L. Shi, and T. Griffiths, "Online inference of topics with latent Dirichlet allocation," *12th International Conference on Artificial Intelligence and Statistics 2009*, no. 1999, pp. 65–72, 2009.

[19] I. Porteous, D. Newman, A. Asuncion, and M. Welling, "Fast Collapsed Gibbs Sampling For Latent Dirichlet Allocation Categories and Subject Descriptors."

[20] D. Sontag and D. Roy, "Complexity of Inference in Latent Dirichlet Allocation.," *NIPS*, pp. 1–14, 2011.

[21] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[22] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 4. Springer, 2006.

[23] B. W. Silverman, *Density Estimation for Statistics and Data Analysis (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)*. Chapman and Hall/CRC, 1986.

[24] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 4. Springer, 2006.

[25] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan, "An Introduction to MCMC for Machine Learning," *Machine Learning*, vol. 50, no. 1–2, pp. 5–43, Jan. 2003.