2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

A Hierarchical Classification Algorithm for Evaluating Energy Consumption Behaviors

Li Bu, Dongbin Zhao, *Senior Member*, *IEEE*, Yu Liu, Qiang Guan
The State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing, China
bulipolly@gmail.com, dongbin.zhao@ia.ac.cn, yu.liu@ia.ac.cn, qiang.guan@ia.ac.cn

***Abstract*—Researches on office building energy consumption have been hot in these years, but few researchers consider the classification of office energy consumption performance which can evaluate user behaviors in order to offer a clear analysis of energy consumption and improve their energy saving consciousness. In this paper, we propose a novel hierarchical classification algorithm for evaluating energy consumption behaviors at a real energy management system, which combines fuzzy c-means clustering with GA (genetic algorithm)-based SVM (support vector machine) to fully utilize collected samples. The experiment results with real energy consumption data show that the proposed algorithm works well to distinguish the abnormal behaviors and classify energy consumption behaviors accurately on normal offices.**

***Keywords*— building energy consumption; classification; hierarchical algorithm; support vector machine**

I. INTRODUCTION

With the continuous decrease of the earth's resources and the increase of carbon dioxide emissions due to the unreasonable consumption of energy, most of the countries, not only developed countries but also developing countries who are more dependent on energy, have realized the urgency of energy conservation. Almost every year, different conferences about energy or climate will be held, such as United Nations Climate Change Conference, World Energy Congress and World Future Energy Summit, and the primary task put forward by these conferences is to effectively make full use of existing energy and reduce emissions. Under this urgent strategic need, monitoring and analysis of energy consumption have attracted more researchers' attention. It's worth noting that in all categories of energy consumption, building consumption occupies a large proportion, even more than industrial consumption and transportation [1].

Nowadays, research directions of building energy mainly include: 1.using more renewable energy (such as wind and solar) to reduce the demand of fossil sources [2], and many of them are mainly aimed at residential buildings to lower user costs and the grid burden [3][4][5]; 2. Non-intrusive appliance load monitoring (NIALM), or disaggregating the total energy consumption into each of the electrical appliances [6][7][8], whose feedback can help energy saving on some appliances; 3. Building energy and comfort management systems to control systems by monitoring, data storage and communication [9][10]. Among these researches on intelligent buildings, office buildings occupy a large proportion [1], and many projects have been launched and funded by different institutes and universities [11][12]. The energy saving potential can be up to 58% followed by [13] in simulation, while in actual experiment [14] the saving potential has reached 25% much more than before.

Researches on office building energy consumption contain many topics [10][11][12][13][14], but few researchers consider the classification of energy consumption performance which can evaluate user behaviors in order to offer a clear analysis of energy consumption and improve their energy saving consciousness. Moreover, the energy consumption behavior of some office may be in an abnormal state because of decoration or function conversion. It poses challenges for evaluation of consumption behaviors accurately.

Since labeled examples are fairly expensive to obtain, semi-supervised learning which makes use of all collected data for training – typically a small amount of labeled data with a large amount of unlabeled data – has been a hot topic. The standard co-training algorithm [15] requires two sufficient and redundant views which are not always consistent with the actual situation. Zhou [16] proves the complete necessity theorem of co-training, and the experiment result shows that learning based on disagreement does not need multi-views, but only requires some appropriate disagreement between classifiers. In [17], Zhou provides some theoretical analysis and experimental results of co-training with insufficient views. The common feature of semi-supervised learning at present is that the labeled examples are given randomly.

In this paper, for evaluating building energy consumption behaviors accurately, we focus on a hierarchical classification algorithm to firstly distinguish the abnormal behaviors and then classify energy consumption behaviors accurately on normal offices which is very practical in real applications. The algorithm consists of three parts: unsupervised learning, labeling and classification with supervised learning, which draw lessons from the idea of semi-supervised learning. By using unsupervised learning, i.e. fuzzy c-means clustering in this paper, we can put the distribution information of all the

This work was supported in part by National Natural Science Foundation of China under Grant Nos. 61273136 and 61034002, and Beijing Natural Science Foundation under Grant No. 4122083, and Visiting Professorship of Chinese Academy of Sciences.

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

feature data of building energy consumption into consideration. Then, labeling samples which are cluster centers or along the boundary of their class region is carried out by experienced professionals. It is a critical step so that labeled data are no longer arbitrary but will affect class labels and margins which determine class performance. Training multi-class classifier with SVM whose training goal is to establish classification margins with support vectors is the last step to evaluate their consumption behaviors. Researchers can do a lot based on our work, such as taking corresponding control decision, establishing rewards and punishment mechanism for some companies, and improving energy saving consciousness.

The structure of the paper is as follows. At first, in section II, we briefly introduce Fuzzy c-means clustering and SVM. The choice of parameters and our hierarchical classification algorithm are given in section III. Features selection, definition of energy consumption behaviors and experimental results are given in section IV. Finally, conclusions are shown in section V.

II. BACKGROUND

In classification problems, samples close to class boundary are more vital than others, which affect generating a good classifier so as to affect classification performance. In Fig.1, all samples coming from two different types of distribution are separated by the blue dotted slash. However, it's hard to distinguish points between two classes, such as those hollow red and blue dots. When using an unsupervised learning to cluster these samples into two classes, the result is often as below that the red dotted slash is an actual classification surface which divides samples into red dots and blue dots, but some of them are wrongly classified. The solid dots belong to a certain class at a high possibility, while those hollow red and blue dots have low possibility belonging to any class, and they need further judgment.

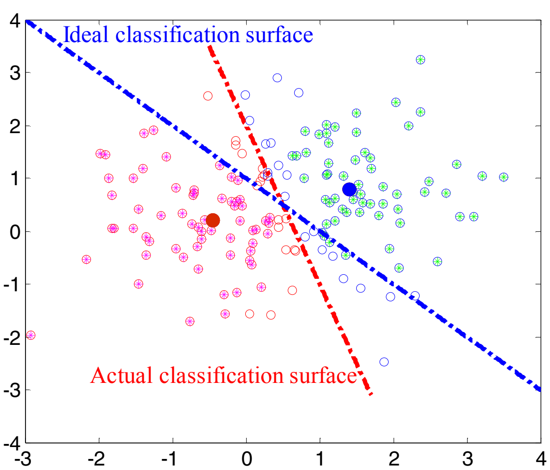

Fig. 1 Illustration with two types dataset from different gaussian distribution. The blue dotted slash is the ideal classification surface, and the red one is the actual surface generated by some learning method.

In this paper, we combine fuzzy c-means clustering with SVM together to fully exploit their advantages, where clustering can make the most of distribution information, and SVM is an effective classifier by establishing a maximum classification hyperplane.

A. Fuzzy c-means clustering

Fuzzy c-means clustering (FCM) is a combination of k-means cluster and fuzzy algorithm so that data can belong to more than one cluster to a certain degree [18], which is called soft clustering (relative to hard clustering in which each element belongs to exactly one cluster). Let $X=\{x_1,x_2,...,x_N\}^T$ be a sample of N observations in R^n; k is an integer, $2\le k<N$, which determines clusters number; $W=\{w_{i,j}, i=1,...,k, j=1,...,N\}$ is a matrix representation of the partition of N data with k clusters. Traditional hard clustering such as k-means can be described as [18]:

$$w_{i,j}=\begin{cases}1, & x_j\in cluster_i\\ 0, & otherwise\end{cases} \tag{1a}$$

$$\sum_{j=1}^{N} w_{i,j}>0,\quad for\ all\ i; \tag{1b}$$

$$\sum_{i=1}^{c} w_{i,j}=1,\quad for\ all\ j. \tag{1c}$$

In this case, one sample must belong to only one cluster. By introducing fuzzy logic, the elements of W are not just $\{0,1\}$, but in an interval $[0,1]$ which indicate the strength of the association between data and a particular cluster. If $k=2$, (1a) can be modified as:

$$w_{i,j}=\begin{cases}n, & x_j\in cluster_i\\ 1-n, & otherwise\end{cases} \tag{2}$$

where $n\in[0,1]$, and both (1b) and (1c) remain unchanged.

Given the weight matrix W and data X, each center of one cluster can be calculated:

$$center_i=\frac{w_{i\cdot}^{\ m} X}{\sum_{j=1}^{N} w_{i,j}^{m}},\ i=1,...,k \tag{3}$$

where w_i is the i th row vector of W, and m, $1\le m<\infty$, is weighting exponent which is set to 2 in our experiment. The clustering criteria we chose is the generalized least-squared errors function which is the most popular and well-studied method [18]:

$$J_m(X,W)=\sum_{j=1}^{N}\sum_{i=1}^{k}(w_{i,j})^m \left\| x_j - center_i \right\|_2^2 \quad (4)$$

The iterative algorithm is to minimize (4) so as to identify (local) optimal fuzzy k -partitions in X .

B. Labeling samples

Labeling samples is a normal but critical step in every classification occasion. In our algorithm, we no longer label samples arbitrarily before experiments, but take actions after clustering which will offer us “important” samples.

After clustering by FCM, we can obtain the cluster centers and the weight matrix W which indicates the degree of any sample belongs to a certain class. Labeling the centers can tell what exactly the clusters mean, and labeling samples whose maximum weight is lower than a threshold, i.e. they are along the boundary of their class region, will determine class margins. For example, in Fig.1, the maximum weight of each hollow dot is smaller than 0.6 that it’s an “important” sample but hard to determine their classes.

C. Support vector machine

Support vector machine is a supervised learning model used for classification and regression analysis. In this paper, it is used as a classifier whose training purpose is to build an optimal hyperplane seperating the two different categories as farther as possible.

As described in [19][20], the optimization problem is cast in dual form:

$$\max\left(\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i,x_j)\right) \quad (5)$$

$$\text{subject to } \sum_{i=1}^{N} y_i\alpha_i = 0,\ C \geq \alpha_i \geq 0$$

where we choose the radial basis function (RBF) $K\left(x_i,x_j\right)=exp\left(-\frac{\left\|x_i,-x_j\right\|^2}{\sigma^2}\right)$ as kernel function, σ is the width of this kernel, $C>0$ is a cost parameter, α_i is an introduced Lagrange multipliers, and $y_i \in \{-1,1\}$ is the label of sample x_i . It is clear that parameters C and σ are very important to establish the optimal hyperplane. Here, we consider the GA-based algorithm [21] to select the optimal parameters; performances are evaluated with K-fold cross-validation to mitigate the fact that a limited labeled data set is available. Detailed algorithm process will be given in the next section.

III. A HIERARCHICAL ENERGY CONSUMPTION BEHAVIOR CLASSIFICATION ALGORITHM

A. GA-based SVM training algorithm

Different from exhaustive grid search which is a two-dimensional minimization procedure, genetic algorithm (GA) has more advantages that search for optimal parameters more effectively [21].

1) Chromosome design

The chromosome is composed of two parts, namely C and σ . The genotype of parameters C and σ should be transformed into phenotype after each iteration. Assume that the encoding of individual V is $V=v_1v_2...v_L$, where V is the representation of variables C and σ , $v_j \in \{0,1\}$ is the binary value. The corresponding decoding formula is:

$$V = a + \frac{b-a}{2^L-1}\left(\sum_{j=1}^{L} v_j 2^{L-j}\right) \quad (6)$$

where a is the minimum value of V , b is the maximum value of V , and L is the length of bit string. Moreover, $\delta = \frac{b-a}{2^L-1}$ is the coding accuracy.

2) Fitness function

Apparently, classification accuracy is an important standard of evaluating a classifier, so that fitness function is simply defined as: $Fit = accuracy(SVM)$.

3) Training GA-based SVM

The process of training GA-based SVM is shown in Fig.2. In the part of “training SVM”, K-fold cross-validation is taken to evaluate the performances.

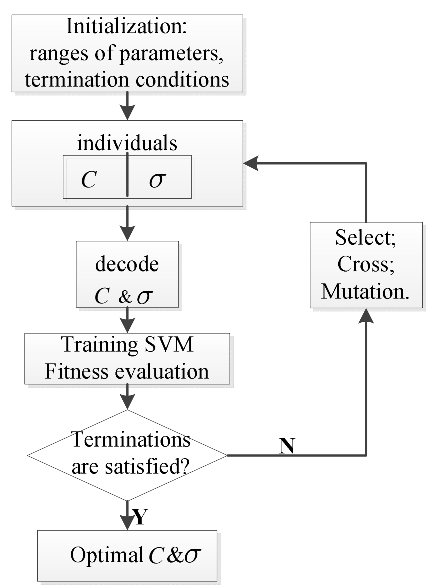

Fig. 2 Training process of GA-based SVM

B. Hierarchical classification algorithm

It's obvious that comparing normal energy consumption behaviors with abnormal behaviors does make no sense. We provide a hierarchical energy consumption behavior classification algorithm to firstly distinguish the abnormal behaviors and then classify energy consumption behaviors on normal offices. FCM and semi-supervised classification are used during the operation respectively, and the procedure is shown in Fig.3. Special experimental details and applications will be given in the next section.

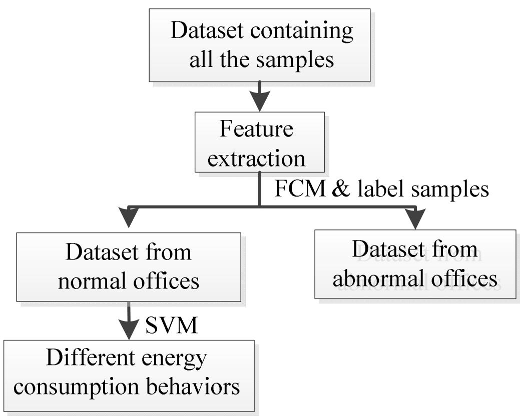

Fig.3 Hierarchical energy consumption behavior classification algorithm

C. Designing the novel semi-supervised classifier

As described in the introduction, the novel semi-supervised classification algorithm for classifying energy consumption behaviors on normal offices consists of three parts: fuzzy c-means learning shown in section II part A, labeling by human and GA-based SVM detailed above. The complete description of this algorithm is given below.

ALGORITHM: SEMI-SUPERVISED CLASSIFIER

---

1. 1. Dataset scale into [0,1]; define cluster number k and threshold ε;
2. Initialization of GA;
3. Clustering dataset with fuzzy c-means learning into k clusters with centers $c_i, i=1,\ldots,k$ respectively;
4. Labeled set = $\{x_j, \max_i w_{i,j} \geq \varepsilon\}$; unlabeled set = $\{x_j, \max_i w_{i,j} < \varepsilon\} \cup \{center_i, i=1,\ldots,k\}$;
5. Labeling the unlabeled set with interactive processing so that all the dataset, i.e. training set Ts, are labeled with relevant labels;
6. Training GA-based SVM with training set Ts to generate a classifier.

The significance of scaling (step 1) is to avoid imbalance in feature space between greater numeric ranges and smaller ones. Cluster number k is given according to the actual requirements of classification beforehand, and threshold ε is used to select data set which have fuzzy class attributes (step 4) but are extremely vital for the establishment of classification hyperplane. Initialization of GA (step 2) consists of the ranges of parameters, maximum iterative steps, the length of bit string, and the probabilities of cross and mutation. In general, these values are based on designers' experience as well as system requirements. For every real application, the labeled dataset are generated by referring to experienced professionals, but usually, labeled samples are arbitrary without pertinence of those "important" samples. In order to overcome this disadvantage effectively, we deliberately combine typical unsupervised learning (fuzzy c-means algorithm) and supervised learning(SVM) to make full use of the dataset distribution, and label samples which are fuzzy-class but important (step 4 & 5). After step 6, a well performing SVM classifier is obtained and used to classify coming samples.

IV. EXPERIMENT

A. Problem description

The office building energy dataset is collected with the Shugu Building in Qinhuangdao, Hebei Province, China. In every office, there are three electrical meters used to measure the energy of socket, air conditioning and lighting respectively, and then these electrical meters transmit data to the console every hour for data analysis and preservation. In this experiment, the chosen dataset covers six offices from the 4th floor without loss of generality, and have been collected during working days from June to the mid-September with the air conditioners working on cooling function. Fig.4 shows the socket energy consumption every day in the 4th floor, where the horizontal axis represents the serial number of hours, and the vertical axis represents the energy consumption every hour. In order to see the energy consumption in one day clearly, we pick out any day of office 1 from three meters, as shown in Fig.5, but the energy consumption behaviors do not keep the same every day or every office. That's why we need to distinguish abnormal consumption behaviors and further classify energy consumption behaviors on normal offices.

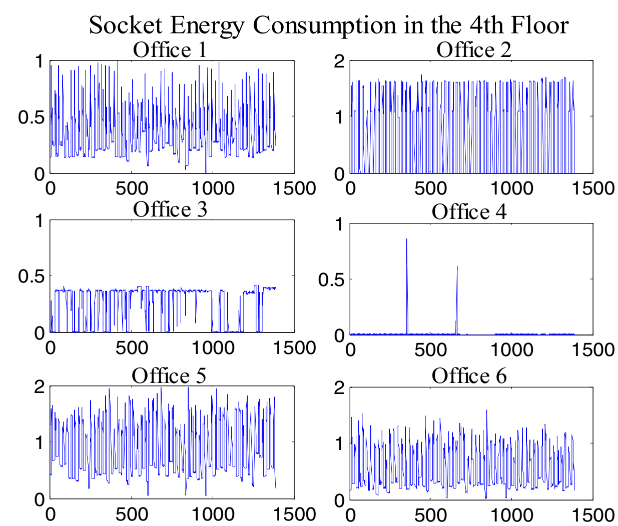

Fig. 4 Socket energy consumption in the 4th floor

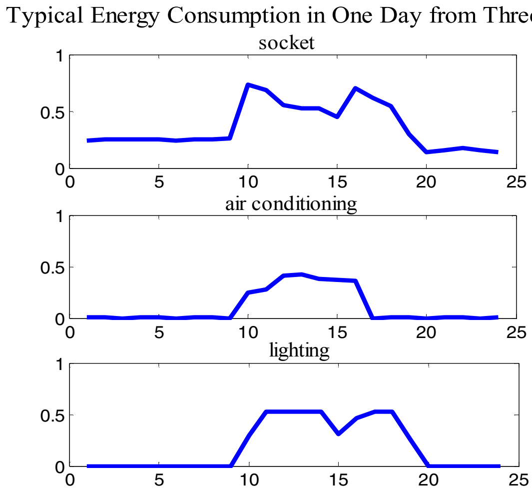

Fig.5 Typical energy consumption in one day from three meters

B. Definition of abnormal consumption behaviors and types of normal energy consumption behaviors

In all offices, it is possible that the energy consumption of some office is abnormal because of decoration, function conversion, etc. Distinguishing these abnormal consumption behaviors by FCM can better evaluate the normal energy consumption behaviors.

According to the actual requirements, we define three types of energy consumption behaviors: low energy consumption, high energy consumption, and uninterruptible consumption.

Low energy consumption (LC): energy consumption in one day is small or even close to zero;

High energy consumption (HC): consumption is low during non-working time, but very high during working time;

Uninterruptible consumption (UC): consumption may not be high, but almost continuous without power off.

C. Feature selection

In order to distinguish these three energy consumption behaviors on normal offices, we select eight kinds of features as the input of SVM: duration of consumption, average power during the morning, working time, night and the whole day, variance during working time and the whole day, the variance of changes of consumption.

D. Experiment setting

In the experiment, energy consumption datasets from three meters during mid-June to August containing 55 days are used to distinguish abnormal behaviors and then train semi-supervised classifier for each electrical type, and those containing 13 days during September are used to test the proposed algorithm.

In the first step to distinguish abnormal behaviors, the features for every office are their average consumption of three electrical meters every day, and cluster number for FCM is set to 2. In the second step to classify energy consumption behaviors on normal offices, cluster number is set to 3 for each electrical type, threshold ε is 0.5, ranges of C and σ are [0, 100], maximum iterative step is 200, the length of bit string L is 20, and the probabilities of cross and mutation are 0.4 and 0.2 respectively.

It's worth noting that for each dataset coming from different electrical meter that has different behavioral characteristic, we generate a different semi-supervised classifier respectively.

E. Experiment results

At first, we distinguish the "abnormal" offices from all the rooms with FCM (in Fig. 6). Except for Office 4, others are all "normal".

Then, we evaluate the energy consumption behaviors on normal office dataset. The classification results are shown in Fig.7. The horizontal axis represents the serial number of days for each office, and the vertical axis represents the classification results where "1" represents Low energy consumption (LC), "2" represents high energy consumption (HC), and "3" means uninterruptible consumption (UC). To compare with the classification results with real energy data, all the test data are shown in Fig.8, where the horizontal axis represents the serial number of hours, and the vertical axis represents the energy consumption every hour. It's obvious to see that consumption behaviors in line with LC characteristic are classified as LC accurately, and so as other behaviors.

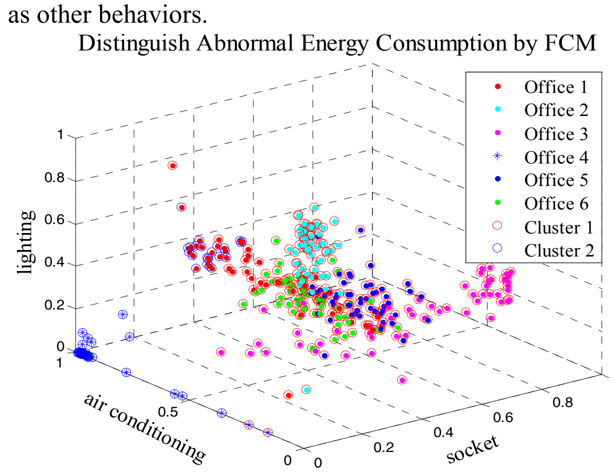

Fig.6 Distinguish abnormal energy consumption by FCM

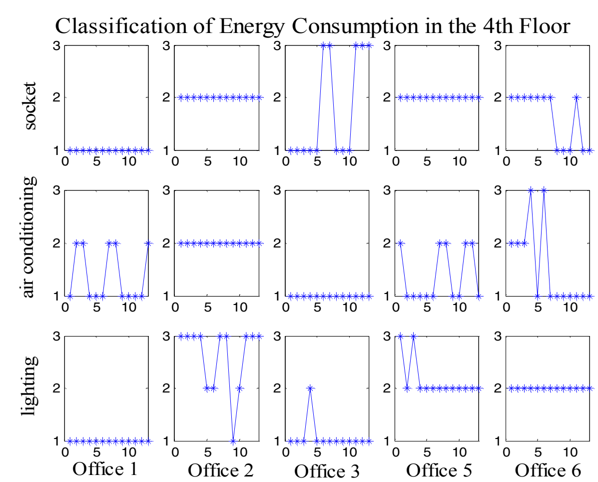

Fig7. Classification results of office building energy consumption in the 4th floor

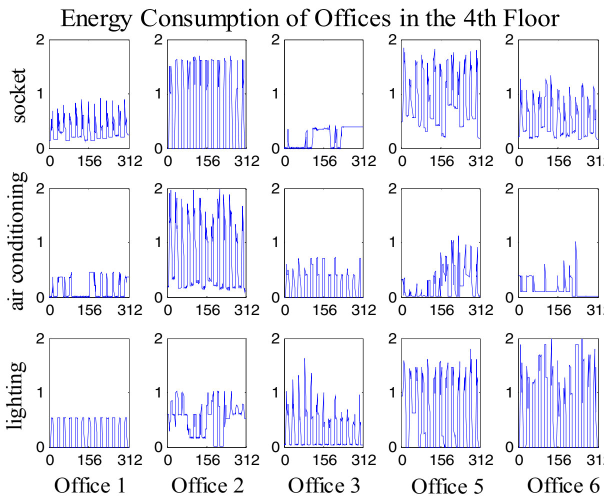

Fig.8 Energy consumption dataset of normal offices in the 4th floor

V. CONCLUSION

To analyze the performance of office building energy consumption behaviors, we present a novel hierarchical classification algorithm to distinguish abnormal behaviors and classify energy consumption behaviors on normal offices based on a novel semi-supervised classifier. This classifier can make full use of these two learning algorithms where Fuzzy c-means clustering can fully utilize the distribution information of all the features of building energy dataset, and then give a weight matrix as output which indicates the strength of the association between that data and a particular cluster. Furthermore, we label samples which belong to certain classes at low probabilities but vital for generating an optimal class margin. This is the most important step in our algorithm that makes labeled samples more targeted. In the final step, GA-based SVM with optimal parameters established well-performed classifiers to evaluate energy consumption behaviors on normal office dataset. The test results show that this method can effectively distinguish abnormal consumption behaviors and classify energy consumption behaviors on normal offices. These classification results can be used in subsequent work, such as taking corresponding control decision.

REFERENCES

[1] T. A. Nguyen and M. Aiello, "Energy intelligent buildings based on user activity: A survey," *Energy Buildings*, 56 (2013), pp. 244–257.

[2] C. W. Potter, A. Archambault and K. Westrick, "Building a Smarter Smart Grid Through Better Renewable Energy Information," *Proc. Power Syst. Conf. Expo*., pp.1 -5 2009.

[3] B. B. Alagoz, A. Kaygusuz, M. Akcin, and S. Alagoz, " A closed-loop energy price controlling method for real-time energy balancing in a smart grid energy market," *Energy*, vol. 59, pp. 95-104, Sep. 2013.

[4] M. Severini, S. Squartini, and F. Piazza, "Hybrid soft computing algorithmic frame-work for smart home energy management," *Soft Computing,* vol.17, pp. 1983-2005, Nov. 2013.

[5] S. Squartini, M. Boaro, F. De Angelis, D.Fuselli and F. Piazza, "Optimization algorithm for home energy resource scheduling in presence of data uncertainty," *In IEEE Proceeding of ICICIP,* 2013.

[6] G. W. Hart, "*Nonintrusive appliance load monitoring*," *Proc. IEEE*, vol. 80, no.12, pp.1870–1891, 1992.

[7] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Using Hidden Markov Models for Iterative Non-intrusive Appliance Monitoring," *in Neural Information Processing System, Workshop on Machine Learning for Sustainability*, 2011.

[8] H. Gonçalves, A. Ocneanu, and M. Bergés, "Unsupervised disaggregation of appliances using aggregated consumption data," *In 1st KDD Workshop on Data Mining Applications in Sustainability*, Aug.2011.

[9] G. J. Levermore, Building Energy Management Systems: Applications to Low-Energy HVAC and Natural Ventilation Control, 2nd Edition, E & FN Spon, London, 2000.

[10] B. Dong and B. Andrew, "Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings," *In Proceedings of the 11th international IBPSA Conference,* Glasgow, U.K, 2009.

[11] G.R. Newsham and B.J. Birt, "Building-level occupancy data to improve ARIMA-based electricity use forecasts, " *In Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, ACM, New York, NY, USA, 2010, pp. 13–18.

[12] Z. N. Zhen, Q.S. Jia, C. Song, and X. Guan, "An indoor localization algorithm for lighting control using RFID," *in Energy 2030 Conference*, pp. 1–6, Atlanta, Georgia, USA, 2008.

[13] D.T. Delaney, G.M.P. O'Hare, A.G. Ruzzelli, "Evaluation of energy-efficiency in lighting Systems using sensor networks", *In Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, ACM, New York, NY, USA, 2009, pp. 61–66.

[14] V. Garg and N. Bansal, "Smart Occupancy sensors to reduce energy consumption," *Energy and Buildings ,* vol.32, pp.81–87, 2000.

[15] A. Blum and T. Mitchell, "Combining Labeled and Unlabeled Data with Co-Training," *Proc. 1th Ann. Conf. Computational Learning Theory*, pp. 92-100, 1998.

[16] W. Wang and Z. H. Zhou, "A new analysis of co-training.," *Proc. ICML*, pp.1135–1142, 2010.

[17] W. Wang and Z. H. Zhou, "Co-training with insufficient views," *In: Proceeding of the 5th Asian Conference on Machine Learning*, Canberra, Australia: ACML, 2013.

[18] J. C. Bezdek, R. Ehrlich, W. Full. "Fcm: the fuzzy c-means clustering algorithm," *Computers & Geosciences,* vol. 10, no. 2-3, pp. 191-203, 1984.

[19] C. Alippi, L. Bu and D. B. Zhao. "SVM-based just-in-time adaptive classifiers," *in Proc. The Int. Conf. Neural Information Processing Part II*, LNCS 7664, pp.664-672, 2012.

[20] C. Alippi, D.R. Liu, D.B. Zhao and L. Bu. "Detecting and reacting to changes in sensing units: the active classifier case," *IEEE Transactions on System, Man and Cybernetics Part A – Systems and Humans*, DOI 10.1109/TSMC.2013.2252895.

[21] C. L. Huang and C. J. Wang. "A GA-based feature selection and parameters optimization for support vector machines," *Expert Syst. Appl.*, vol. 31, no.2, pp. 231−240, 2006.