Disaggregating household loads via semi-supervised multi-label classification

Ding Li*, Kyle Sawyer* and Scott Dick*
*Department of Electrical and Computer Engineering
University of Alberta, Edmonton, AB, Canada, T6G 2V4
Email: {dli3, ksawyer, sdick}@ualberta.ca

***Abstract*—The essence of Non-Intrusive Load Monitoring (NILM) is to extract electricity consumption details of individual appliance from an aggregated house-level electrical measurement at the main panel without sub-metering each appliance. In this paper, an Expectation Maximization (EM) based semi-supervised multi-label classification technique is applied in NILM. It requires a one-time registration of individual appliance to obtain few samples during the training stage. After that, the total electricity is utilized to detect the states of each appliance and analyze electricity consumption information of individual appliance for each instance via the help of semi-supervised learning method. Experiments on house 1 and house 3 dataset of Reference Energy Disaggregation Dataset (REDD) verify the effectiveness of application of Semi-supervised learning techniques in NILM.**

***Keywords*—Non-intrusive load monitoring, semi-supervised learning, multi-label classification, EM algorithm, RAKEL**

I. Introduction

Energy conservation is a hot topic and challenging issue due to the increasing energy demand and the negative effects of energy generation on environment and climate change (CO_2 emissions, global warming). A significant reduction of energy wastage could be achieved in residential district via fine-grained energy consumption details, since it was said that 30% of the electricity usage is utilized in the residential sector in the European Union [1]. By providing customers with itemized appliance specific energy consumption information instead of the aggregated monthly bill of electricity. Customers could be aware of which appliances consume a lot of electricity and need updated or may be replaced for efficient ones. If peak timing and off-peak timing are applied in real life, itemized appliance specific electricity consumption details could make customers identify high-power appliances and run high-power appliances off-peak hours to reduce electricity bills and save energy. And energy management system could devise load scheduling schemes to optimize energy generation and utilization. Moreover, a recommender system could be produced to help customers modify their behaviors to cut down electricity bills and conserve energy. Thus a lot of energy could be saved from the residential area. One way to make it happen is appliance load monitoring which is able to obtain electricity consumption details of individual appliance. These years, a lot of smart meters are deployed on a large scale in the residential households by the government of UK and USA due to envision that appliance load monitoring will make it possible to charge electricity based on peak and off-peak timings [2]. And a lot of research work on appliance load monitoring has been carried out trying to realize it.

There exist two methods of appliance load monitoring. Namely intrusive load monitoring and non-intrusive load monitoring. Intrusive load monitoring provides fine-grained appliance energy consumption details since it installs sensors on every appliance of interest to monitor its state and energy consumption characteristic. And an automatic data collecting center gathers data to output the electricity consumption details for individual appliance. However, it requires significantly expensive and complicated hardware installation in order to obtain appliance-level electricity information. And it requires complex network for data storage and transmission. Moreover, appliances are replaced and purchased in a household from now and then, it is truly inconvenient and intrusive for experts to enter residential households to place sensors on newly bought appliances and uninstall them from old ones.

Non-intrusive load monitoring is developed one year later and is a good alternative since it requires only one single point measurement of house-level granularity at the main electricity panel outside the house. It is called non-intrusive since there is no need for experts to enter the household. Compared with intrusive load monitoring, few sensors are needed, few data needs collected, data collection, storage and transmission will be significantly simplified. Moreover, the equipment cost, equipment installation, maintenance, remove fees could be largely minimized. Nevertheless, there is one-time training data collection process of individual appliance. After that, the house-level electricity information is utilized to discern electricity details for individual appliance. It depends on machine learning and data mining techniques to extract appliance specific electricity consumption details from the aggregated electricity consumption data of the whole household.

In this paper, an EM-based semi-supervised learning technique is applied in NILM. It tries to reduce the amount of labeled data without sacrificing classification performance by combining a large amount of unlabeled data in training. The rest of the paper is organized as follows: Section 2 briefly introduces the state of NILM in literature. Section 3 presents data, data processing, methods and performance measurements. Section 4 reports on the experimental results. Finally, Section 5 concludes and raises several issues for the future work.

II. The state of the art of NILM

In the early 1990s, Hart proposed NILM to break down an aggregated energy consumption data into its constituents by examining the appliance-specific power consumption characteristics which he called appliance signatures [3]. The aggregated house-level energy consumption data is measured from the main panel outside the house, hence it is considered to be non-intrusive since it avoids installing measurements inside the customers' property. In contrast, the intrusive load monitoring, which needs experts entering the house to install distributed sensors on almost every appliance. The goal of NILM is to decompose the total load energy consumption data into its major components running during a period of time. The problem can be formulated as follows:

$$P(t) = \Sigma_{i=1}^{n} p_i(t) + \xi(t) \tag{1}$$

Where $P(t)$ is the total power consumption data, $p_i(t)$ is the power consumption data of individual appliance, n is the number of appliances running during a period of time, and $\xi(t)$ could be attributed to the contribution of measurements error or noise. It does not seem like a very difficult problem since optimization-based approaches could solve it, only if you know the power consumption data of individual appliance in advance. The truth is not so simple. Because appliances are purchased and replaced from now and then, it is difficult and expensive to track appliances for each household all the time. And the starting-up transients will make the solution complicated. Moreover, Hart categorized appliances into 4 classes, namely, ON/OFF state appliance like kettle, which has two states, either on or off; Finite State appliance which has finite states, like washing machine whose procedure is composed of water-fill, immerse, rinse, spin and dry operations; Continuously Variable appliance like refrigerator whose power consumption varies continuously and appliance Continuously Constant like alarm monitoring. Fig. 1 gives examples of the power draw characteristics of three classes of appliances. The above formulation is appropriate for disaggregating ON/OFF state appliances only.

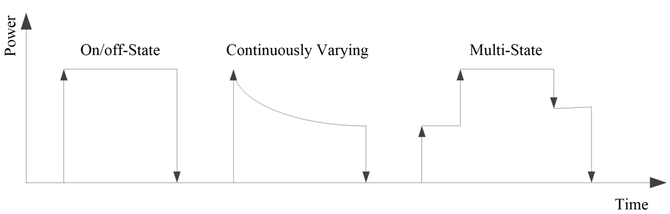

Fig. 1: Power curves of three types of appliances.

In [3], Hart presented different kinds of signatures which he considered as a set of measurable parameters of the total load that bearing information about the number and operating states of its constituents. Signatures are essential to NILM. Hart put them into two classes, one is passively observable and measurable that he called non-intrusive signatures, including steady-state signatures like power, and transient signatures like starting-up transients. Steady state signatures are derived from the difference between the measurements when appliance is in steady state operation. Transient signatures are obtained from the transient waveform like the size, duration, shape and harmonics. Transients provide less information and are expensive and difficult to detect. But they can be incorporated to provide additional information where appliances of the similar steady-state signatures can't be distinguished. Another he called intrusive signatures which either needs device installing on appliances to trigger signals to record signatures of individual appliance or injecting electrical signals like harmonics or transients to notice the behaviors and activities of appliances. Signatures could be transformed into a set of features to distinguish appliances with the help of algorithms.

In hart's paper, clustering algorithm was proposed to group and identify appliances based on real power and reactive power. The shortcoming is that the method can only be used to disaggregate ON/OFF state appliances and can't distinguish appliances with similar active power and real power. Also it failed to identify appliance whose power below 150w.

Since then, many research work has focused on NILM. New signatures have been presented. Including context-based information like the location of the house owner, temperature, duration and time of appliance usage [4]. In [5], [6], [7], peak and Root Mean Square (RMS) current and voltage were extracted as signatures to distinguish appliances. RMS was found to be more discriminative than peak values. In [7], [8], researchers have combined harmonics with real and reactive power to improve disaggregation performance. In [9], [10], V-I trajectory with normalized current and voltage values was creatively proposed to cluster appliances into 8 groups and then sub-partition within each group, this approach was proved to be more effective than existing approaches using power measurements. Transient signatures to date also develops fast. Researcher In [11] showed that energy calculated during the starting-up transient could be an effective signature to discriminate appliances. Power spikes and overshoots were tried as features in [12] to decompose loads. A very untraditional and novel method was presented in [13] to discriminate appliances without pre-training and supervision, the author used two basic units, namely rectangles and triangles, as features to identify appliances. Triangle unit was described by starting time, peak time, peak value and ending time while rectangle was described by starting time, peak time, peak value, steady time and steady power. The identification accuracy was 80%.

Many new algorithms in NILM have been developed and field tested. Either supervised or unsupervised. Broadly, the supervised algorithms could be partitioned into two classes, namely, pattern recognition based approach or optimization based approach. Different optimization methods have been tried in [14], [15], [16] to tackle NILM as an optimization problem. However, these methods can't detect unknown appliances and can't identify appliances with similar signature. Researchers have utilized committee decision mechanisms in [16], [17] to improve the overall decomposition accuracy with different algorithms and signatures. Support Vector Machines [8], Artificial Neural Network [6] and Hidden Markov Models [18] have also been utilized in NILM. Recently, many unsupervised methods [4], [19], [20], [21] have presented in NILM trying to achieve as less cost as possible. However, the accuracy is not high. Although NILM have been developed for so many years, there still exist many problems need to be solved before being widely accepted and applied.

Traditionally, supervised learning deals with data associated with single-label indicating its belongingness to a concept or not. However, in many real-world problems, objects may belong to multiple concepts simultaneously. For instance, in text categorization, a document may belong to several concepts; in bioinformatics, a protein may have multiple functions on cells; in movie categorization, a movie may belong to adventure, science fiction and drama simultaneously. Obviously, Binary classifier is unable to deal with these situations. Multi-label techniques are thus developed to handle such cases. Multi-label learning deals with data associated with multiple labels simultaneously. Under the framework of multi-label learning, each instance is associated with multi-labels, indicating a set of concepts it belongs to.

NILM is recently regarded as a multi-label classification problem in [22], [25]. Since the total electricity consumed of a house at each instance is associated with multiple appliances running simultaneously during each period of time. Thus the appliances can be taken as the set of labels for the total electricity at each instance.

III. METHODOLOGY

A. Data

The proposed EM-based semi-supervised multi-label classification method is evaluated on two real-world household dataset (house 1 and house 3) of REDD in this paper. The whole house-level and circuit/appliance-level real power consumption data is collected for both houses in the Boston area in the USA for a total of 119 days. The real power is plotted in Fig. 2. The circuit/appliance power information served as the ground truth for classification. The dataset of house 3 also consists of reactive power. It is calculated from currents and voltages measured from both phases of the main electricity panel. The data is publically available on the website: http://redd.csail.mit.edu.

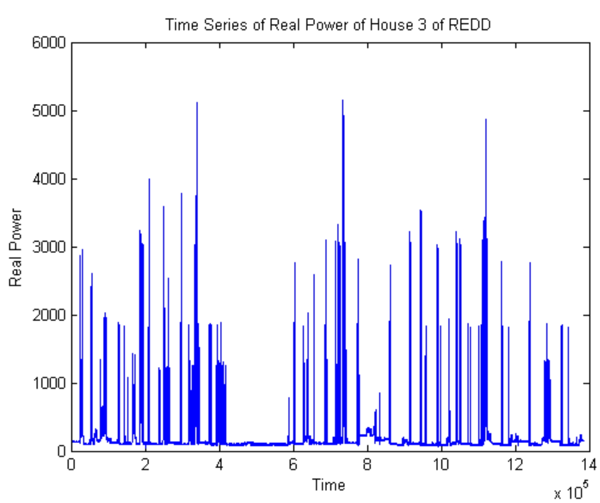

Fig. 2: Real power of house 3.

B. Data Processing via Delay Embedding

According to Takens' embedding theorem [23], the underlying phase space of a physical system can be reconstructed from an observable time series. The most important technique is the method of delays. Vectors in a newly embedded phase space are formed from scalar time measurements with delayed values. Given a single time series:

$$R = \{R_N, R_{N-1}, \cdots, R_2, R_1\} \tag{2}$$

The phase space can be represented by vectors:

$$\Phi_n = \{R_{N-(m-1)\tau}, R_{N-(m-2)\tau,\cdots,R_N}\} \tag{3}$$

where R is the observable time measurement. Φ_n is the reconstructed vector, the number of reconstructed vectors is $N-(m-1)\tau$. m is the embedding dimension, τ is generally referred to as the delay or lag. In this paper, mutual information and false nearest neighbor are respectively adopted to find the best embedding delay time τ and embedding dimension m.

C. Experimental Design

The experiments are carried out in MEKA which is an extension of WEKA Machine Learning Toolkit and it is specialized for multi-label classification. EM is a meta-classifier and it is currently the only algorithm in MEKA that could be adapted for semi-supervised multi-label classification. Some changes have been made to the JAVA code so that EM could incorporate unlabeled data into training. It needs two base classifiers. RAKEL is chosen as a MEKA base classifier and J48 is chosen as a WEKA base classifier. There are three parameters need to be explored. Namely, the number of iterations for EM, the confidence factor used for pruning and the minimum number of instances per leaf for J48.Ten-fold cross-validation is implemented to find the best set of parameters.

The dataset is firstly divided into 80% for training and 20% for out-of-sample testing. Then the training data is further divided into labeled dataset, unlabeled dataset and testing dataset. Labeled dataset is obtained from the 80% training data by making sure each appliance contains 3 on events and three off events individually. After that, the Cross Validation (CV) partition is implemented on the remaining training data in WEKA to obtain unlabeled data and testing data for each fold via unsupervised filter, specifically, through the RemoveFolds method. When using ten-fold cross validation in training, the labeled data keeps the same for each fold, the unlabeled data and testing data used in each fold is obtained through the above CV partition. In out-of-sample testing with the best set of parameters, the labeled data and the whole unlabeled data are used in the learning process to classify the instances not used in training. The code for dividing dataset, parameters exploration and out-of-sample testing is written in ruby. Ruby code calls MEKA to perform ten-fold cross-validation to explore the best set of parameters and do out-of-sample testing.

D. Semi-Supervised Learning

In traditional machine learning techniques, there exists either supervised learning where training samples should all be labeled or unsupervised learning where training samples are unlabeled. When only a portion of the training data is labeled, an effective approach to enhancing the learning performance is to make use of the relatively small amount of labeled data in combination with a large amount of readily available unlabeled data. This exploiting unlabeled data to achieve stronger generalization method is nowadays a very hot topic and it is known as semi-supervised learning.

Many semi-supervised learning algorithms have been proposed in the literature. Among which a very famous type is the Support Vector Machine (SVM) based semi-supervised algorithms. Good examples like Semi Supervised Support Vector Machines (S3VMs) and Transductive SVM (TSVM). The two methods are both based on the assumption that decision boundary should pass through low-density regions and thus utilize unlabeled data to find the best decision boundary. Graph-based semi-supervised methods also develop fast. The basic idea is to construct a graph with all labeled data and unlabeled data and obtain weights between node pairs to quantify the degree of similarity between them. After that, labels propagate from labeled data to unlabeled data.

In this paper, Expectation-Maximization (EM) with multi-label classifier RAKEL is adopted to incorporate unlabeled data into supervised learning to enhance its learning performance. EM is an iterative process which is utilized to estimate the parameters of a generative model and the probability of unlabeled instances belong to each class. There are two stages in the iterative process to train a classifier. Firstly, use the current classifier built in the previous iteration to forecast labels for all of the unlabeled training data (In the first iteration, only labeled data is utilized to train a classifier); secondly, retrain the classifier with all of the labeled data and the most confidently predicted unlabeled data with the predicted labels. The process continues for a specified iteration or until some criteria is satisfied. Then the trained model is exploited to classify unforeseen dataset.

E. RAKEL

Generally, there exist two methods for treating multi-label classification problems. Namely, problem transformation and algorithm adaptation. Label Power set (LP) is a problem transformation approach which treating any combination of labels as a new class in the dataset, then training a single-label classifier on the transformed dataset. It actually transfers the multi-label classification problem into multi-class classification problem. The weakness of this approach is that the number of class it produced could be huge and some class may have only a few samples. Hence data imbalance could occur.

Random k-label set (RAKEL) [24] is thus proposed trying to solve the above problem of LP with an ensemble learning method and it takes label correlations into consideration. The RAKEL algorithm firstly partitions the total set of labels into multiple subsets of size k. And train a single-label classifier with the LP method for individual subset. Thus multiple classifiers are built in the training process. When it comes to predicting labels for future data, a majority voting method is adopted for each label by aggregating the predictions from the multiple classifiers.

F. Performance Measurements

The evaluation metrics used in multi-label classification are different from those used in single-label classification. Generally, the metrics can be categorized into instance-based, label-based and ranking based methods. Label-based metrics evaluate each label separately and then average the results over all labels to give a general evaluation result. The metrics used in this work are two popular label-based averaging method, namely, micro F_1 and macro F_1.

The F_1 metric is the harmonic mean of precision and recall and it is widely adopted for single-label classification. Given the number of true positives (tp), true negatives (tn), false positives (fp) and false negatives (fn). F_1 is formulated as the following:

$$F_1 = \frac{2 \times tp}{2 \times tp + fp + fn} \tag{4}$$

Macro F_1 and micro F_1 are two averaging approaches to calculating the F_1 measure across labels. Given the number of true positives tp_λ, false positives fp_λ, true negatives tn_λ and false negatives fn_λ for label λ after being evaluated by F_1. F_{micro} and F_{macro} are respectively calculated as follows:

$$F_{micro} = F(\Sigma_{\lambda=1}^{M} tp_\lambda, \Sigma_{\lambda=1}^{M} tn_\lambda, \Sigma_{\lambda=1}^{M} fp_\lambda, \Sigma_{\lambda=1}^{M} fn_\lambda) \tag{5}$$

$$F_{macro} = \frac{1}{M} \Sigma_{\lambda=1}^{M} F(tp_\lambda, tn_\lambda, fp_\lambda, fn_\lambda) \tag{6}$$

IV. Experimental Results

The parameters of the data pre-processing are set as follows. For the real power time series of house 3 of REDD, the embedding delay time is 97 and the embedding dimension is 40. For the real power time series of house 1 of REDD, the embedding delay time is 32 and the embedding dimension is 8. The best set of parameters is achieved with the number of iterations for EM as 10, the confidence factor used for pruning as 0.3 and the minimum number of instances per leaf as 1. Next, the proposed Semi approach is applied to the two datasets. The experimental results are obtained and compared with two multi-label classification methods in [25]. The final out-of-sample testing results of the real power time series of house 1 and house 3 are demonstrated in Tables I and II, respectively.

TABLE I: Performance results for house 1.

	Semi	RAKEL	MLKNN
Micro F-measure	0.897	0.587	0.776
Macro F-measure	0.666	0.393	0.619

TABLE II: Performance results for house 3.

	Semi	RAKEL	MLKNN
Micro F-measure	0.660	0.923	0.921
Macro F-measure	0.297	0.492	0.471

The experimental results using Semi are much better than the other two approaches in [25] for house 1. But the results for house 3 are relatively bad. That might be because the dataset of house 1 is too large, the ratio of labeled data and unlabeled data is very low.

V. Conclusion

In this paper, an EM-based semi-supervised multi-label classification technique is applied to non-intrusive load monitoring. Experiments on REDD house 1 and house 3 data show promising perspective of this research topic. Nevertheless, the experimental results reported in the paper are rather preliminary. Thus, conducting more experiments on more dataset of NILM with many more semi-supervised algorithms will be an important issue to be explored in the near future. More specifically, many more algorithms will be explored, the perfect ratio between labeled data and unlabeled data will be studied.

References

[1] Uteley, J.; Shorrock, L. Domestic Energy Fact File 2008; *Technical Report for Building Research Establishment*: Garston, UK, 2008.

[2] Energy Consumption in United Kingdom; *Technical Report for Department of Energy & Climate Change*: London, UK, 2010.

[3] Hart, G.W. Nonintrusive appliance load monitoring. *IEEE Proc*. 1992, 80, 1870-1891.

[4] Kim, H.; Marwah, M.; Arlitt, M.; Lyon, G.; Han, J. Unsupervised Disaggregation of Low Frequency Power Measurements. *In Proceedings of the 11th SIAM International Conference on Data Mining*, Mesa, AZ, USA, 28-30 April 2011.

[5] Figueiredo, M.B.; Almeida, A.D.; Ribeiro, B. An experimental study on electrical signature identification of non-intrusive load monitoring (nilm) systems. *Adaptive and Natural Computing Algorithms*. Springer Berlin Heidelberg, 2011. 31-40.

[6] Ruzzelli, A.G.; Nicolas, C.; Schoofs, A.; O'Hare, G.M.P. Real-Time Recognition and Profiling of Appliances through a Single Electricity Sensor. *In Proceedings of the 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, Boston, MA, USA, 21-25 June 2010; pp. 1-9.

[7] Najmeddine, H.; El Khamlichi Drissi, K.; Pasquier, C.; Faure, C.; Kerroum, K.; Diop, A.; Jouannet, T.; Michou, M. State of Art on Load Monitoring Methods. *In Proceedings of the 2nd IEEE International Conference on Power and Energy Conference*, Johor Bahru, Malaysia, 1-3 December 2008; pp. 1256-1258.

[8] Srinivasan, D.; Ng, W.; Liew, A. Neural-network-based signature recognition for harmonic source identification. *IEEE Trans. Power Del.*, 2006, 21, 398-405.

[9] Lee, W.K.; Fung, G.S.K.; Lam, H.Y.; Chan, F.H.Y.; Lucente, M. Exploration on Load Signatures. *In Proceedings of International Conference on Electrical Engineering (ICEE)*, Sapporo, Japan, 4-6 July 2004; pp. 1-5.

[10] Lam, H.Y.; Fung, G.S.K.; Lee, W.K. A Novel method to construct taxonomy electrical appliances based on load signaturesof. *IEEE Trans. Consum. Electron.*, 2007, 53, 653-660.

[11] Chang, H.; Yang, H.; Lin, C. Load identification in neural networks for a non-intrusive monitoring of industrial electrical loads. *Computer Supported Cooperative Work in Design IV*. Springer Berlin Heidelberg, 2008. 664-674.

[12] Cole, A.I.; Albicki, A. Data Extraction for Effective Non-Intrusive Identification of Residential Power Loads. *In Proceedings of Instrumentation and Measurement Technology Conference*, St. Paul, MN, USA, 18-21 May 1998; Volume 2, pp. 812-815.

[13] Wang, Z.; Zheng, G. Residential appliances identification and monitoring by a nonintrusive method. *IEEE Trans. Smart Grid*, 2012, 3, 80-92.

[14] Baranski, M.; Voss, J. Non-Intrusive Appliance Load Monitoring Based on an Optical Sensor. *In Proceedings of IEEE Power Tech Conference*, Bolgona, Italy, 23-26 June 2003; pp. 8-16.

[15] Suzuki, K.; Inagaki, S.; Suzuki, T.; Nakamura, H.; Ito, K. Nonintrusive Appliance Load Monitoring Based on Integer Programming. *In Proceedings of SICE Annual Conference*, Tokyo, Japan, 20-22 August 2008; Volume 174, pp. 2742-2747.

[16] Liang, J.; Ng, S.K.K.; Kendall, G.; Cheng, J.W.M. Load signature study Part I: Basic concept, structure, and methodology. *IEEE Trans. Power Del.*, 2010, 25, 551-560.

[17] Liang, J.; Ng, S.; Kendall, G.; Cheng, J. Load signature study-part II: Disaggregation framework, simulation and applications. *IEEE Trans. Power Del.*, 2010, 25, 561-569.

[18] Zia, T.; Bruckner, D.; Zaidi, A. A Hidden Markov Model Based Procedure for Identifying Household Electric Loads. *In Proceedings of the 37th Annual Conference on IEEE Industrial Electronics Society (IECON 2011)*, Melbourne, Australia, 7-10 November 2011; pp. 3218-3223.

[19] Johnson, M.J.; Willsky, A.S. Bayesian nonparametric hidden semi-markov models. *The Journal of Machine Learning Research*, 14.1 (2013): 673-701.

[20] Shao, H.; Marwah, M.; Ramakrishnan, N. A Temporal Motif Mining Approach to Unsupervised Energy Disaggregation. *In Proceedings of the 1st International Workshop on Non-Intrusive Load Monitoring*, Pittsburgh, PA, USA, 7 May 2012.

[21] Goncalves, H.; Ocneanu, A.; Berg?es, M.; Fan, R.H. Unsupervised Disaggregation of Appliances Using Aggregated Consumption Data. *In Proceedings of KDD 2011 Workshop on Data Mining Applications for Sustainability*, San Diego, CA, USA, 21-24 August 2011.

[22] Basu, K.; Debusschere, V.; Bacha, S.; Maulik, U.; Bondyopadhyay, S., Nonintrusive Load Monitoring: A Temporal Multilabel Classification Approach, *Industrial Informatics, IEEE Transactions on*, vol.11, no.1, pp.262,270, Feb. 2015

[23] Takens, F. Detecting strange attractors in turbulence, *in Lecture Notes in Mathematics*, No.898 (Springer-Verlag,1981).

[24] Tsoumakas, G.; Katakis, I.; Vlahavas, I. Random k-labelsets for multilabel classification. *IEEE Transactions on Knowledge and Data Engineering*, 23.7 (2011): 1079-1089.

[25] Mostafa Tabatabaei; Scott Dick; Wilsun Xu. Non-Intrusive Load Monitoring via Multi-Label Classification. submitted to Information Sciences.