

Outlier Detection in Energy Disaggregation Using Subspace Learning and Gaussian Mixture Model

Xiu-ming Tang¹, Rong-xiang Yuan¹, Jun Chen^{2,*}

¹ School of Electrical engineering, Wuhan University, Wuhan 430072, P.R. China

² Institute of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, P.R. China

tangxm2873@sina.com, rx yuan@whu.edu.cn, 97chenjun@gmail.com

Abstract

Special Complex non-Gaussian processes may have dynamic operation scenario shifts so that the traditional Outlier detection approaches become ill-suited. This paper proposes a new outlier detection approach based on using subspace learning and Gaussian mixture model(GMM) in energy disaggregation. Locality preserving projections(LPP) of subspace learning can optimally preserve the neighborhood structure, reveal the intrinsic manifold structure of the data and keep outliers far away from the normal sample compared with the principal component analysis (PCA). The results show proposed approach can significantly improve performance of outlier detection in energy disaggregation, increase the fraction true-positive from 93.8% to 97%, decrease the fraction false-positive from 35.48% to 25.8%.

Keywords: Energy Disaggregation, Gaussian Mixture Model, Locality Preserving Projections, Outlier Detection

1. Introduction

Energy disaggregation, aiming at determining the component appliance contributions from an aggregated electricity signal [1], uses machine learning and data mining techniques to study energy, energy consumption patterns and sustainability problems [2-4]. Outlier-free data is very important for building reliable systems models for energy disaggregation. Outliers are incorrect measurements unrelated to the actual output values, so they will affect the analysis of the data sets containing them [5]. In an automatic data acquisition environment, they arise from sporadic malfunctioning of the Reference Energy Disaggregation Data Set (REDD) collection sensors and equipment [1]. Most of times, these outliers affect accuracy in system modeling for energy disaggregation. It is important to identify and remove of these outliers [6-7].

Outlier detection refers to the problem of finding patterns in data that do not conform to expected behavior. There are many outlier detection methods cited in literature and all of these methods have shown to be able to detect obvious outliers [7-9], but these methods could not detect outliers in multi-Gaussianity, multiple states processes, so the traditional outlier detection approaches become inappropriate. As an alternative technique, Bayesian inference-based Gaussian mixture model (GMM) has been developed and applied to outlier detection in complex industrial applications, which consist of multiple operating modes and have significant multi-Gaussianity in normal operating data [10]. It provides the way of Outlier detection for REDD.

However, the distribution of REDD not only obey multi-Gaussianity and models with multiple states but also exhibit significant nonlinearity and multidimensional due to

* Corresponding Author: 97chenjun@gmail.com

various reasons such as time of day or weather conditions [2]. Thus the Gaussian mixture model can not efficiently capture the local nonlinear features and extract the sensitivity features of data [11]. Recently, a novel dimensionality reduction and subspace learning, called LPP was proposed by He and Niyogi [12] in differently with PCA based on preserving the global structure [13]. LPP is capable to recover important aspects of the intrinsic linear or nonlinear manifold structure by preserving local structure. Because of its ability to discriminate directions with the local largest variance in a given data set, now, LPP has been introduced for data preprocessing step in process monitoring and fault detection [14]. Thus, locality preserving projections can result in an appropriate representation of the data, which ensures outliers to stay away from the normal points.

In this paper, a novel outlier detection method was proposed by combining locality preserving projections with GMM and has greater robustness for outlier and noise compared to PCA in outlier detection problems. By using this approach, we conducted a thorough experimental investigation of our method on the REDD [2].

The remainder of this paper is organized as follows. In Section 2, briefly reviews the locality preserving projections methods. In Section 3, locality preserving projections GMM based outlier detection is proposed. Experimental results are reported in Section 4. Concluding remarks are presented in Section 5.

2. LPP of Subspace Learning

LPP aims at optimally preserving the neighborhood structure and revealing the intrinsic manifold structure of the data more effectively. In this study, LPP will be used to extract the effective information and then generate subspace dataset as the inputs of GMM.

The optimization procedure of LPP includes two parts: firstly, an adjacency graph is built to describe the neighborhood relationship between the data points, and then LPP projection vectors are optimized so that the neighboring points in original space can remain close to each other in the low-dimensional embedding space.

The details of LPP algorithm is shown as follows.

$$\begin{aligned} Y &= A^T X, \\ A^T &= (a_1, a_2, \dots, a_d) \end{aligned} \quad (1)$$

Given the high-dimensional input data matrix $X = [x_1, x_2, \dots, x_n]^T \in R^{n \times m}$, which represents n samples of m measured variables, LPP seeks a transformation matrix A to project high-dimensional input data set X into a low-dimensional data set $Y = [y_1, y_2, \dots, y_n]^T \in R^d$ where $d (d \ll m)$ is the dimension of embedding subspace, so that the local structure between data points is preserved.

A reasonable criterion for choosing α is to minimize the following objective function[17],

$$\min_{\alpha} \sum_{i=1}^l \sum_{j=1}^l (y_i - y_j)^2 w_{ij} \quad (2)$$

Where the weight matrix $W = [w_{i1}, \dots, w_{il}]$ (called heat kernel) constructed through the nearest neighbor graph, is a symmetric weights $W_{ij} = W_{ji}$, it incurs a heavy penalty if neighboring points x_i and x_j are mapped far apart. Therefore, minimizing Equation (2), it is an attempt to ensure if x_i and x_j are “close” then y_i and y_j are

close as well. With this objection function, the local structure of the input data can be preserved.

For a matrix formulation, Equation (2) can be reformulated as

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (y_i - y_j)^2 w_{ij} \\ & = \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha^T x_i - \alpha^T x_j)^2 w_{ij} = \alpha^T X^T (D - W) X a = \alpha^T X^T (D - W) X a \end{aligned} \quad (3)$$

Where D is a diagonal matrix as $d_{ii} = \sum_i^n w_{ii}$, whose entries are column sum of W .

$L = D - W$ is usually called the Laplacian matrix. In order to remove the arbitrary scaling factor in the embedding, LPP imposes constrain as follows:

$$y^T D y = 1 \Rightarrow \alpha^T X D X^T a = 1 \quad (4)$$

Combining Eqs. (3) and (4), the optimization problem in LPP is expressed as

$$\begin{aligned} \min_{\alpha} \quad & \alpha^T X L X^T a \\ \text{st.} \quad & \alpha^T X D X^T a = 1 \end{aligned} \quad (5)$$

The above problem can be solved by the generalized eigenvalue decomposition as

$$X L X^T a = \lambda X D X^T a \quad (6)$$

The solutions of Equation (5) bring a series of eigenvectors a_0, a_1, \dots, a_{d-1} ordered by their corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_m$. The dimensionality of the problem can be reduced by only retaining the first $d(d \ll m)$ eigenvectors. For any test vector x_i , the project vectors can be computed by

$$y_i = A^T x_i, A^T = (a_1, a_2, \dots, a_d) \quad (7)$$

Where x_i is projected and then reduced to a d -dimensional vector y_i , which preserves the local relationship with its neighbors. Based on the projected data set Y , the high-dimensional data space is reduced into low-dimensional data space, but retaining majority local variation information in the projected data set. With the reduced dimension and preserved local variance information, the extracted features Y will be used as the new input features of the outlier detection model.

Therefore, LPP is capable of finding the optimal linear embedding that respects the structure presents in that matrix. The locality preserved character of the LPP algorithm makes it relatively insensitive to outliers and noise.

3. Outlier Detection Using GMM Based On LPP

GMM, aiming at offering a comparatively richer model than the single Gaussian [15], can be viewed as a linear superposition of different Gaussian components in which each component is a basis function or a "hidden" unit. GMM can be expressed as

$$P(x_n | \theta) = \sum_{k=1}^k \pi_k N(x_n | \mu_k, \Sigma_k) \quad (8)$$

Where each Gaussian density $N(x_n | \mu_k, \Sigma_k)$ is called a component of the mixture and has its own mean μ_k and covariance Σ_k . Each component prior π_k can be viewed as positive weights and satisfying the requirements $0 \leq \pi_k \leq 1$ and $\sum_{k=0}^k \pi_k = 1$. The form of the Gaussian mixture distribution is governed by the set of parameters $\theta = (\pi_k, \mu_k, \Sigma_k)$. The optimal parameter θ is determined by Maximum likelihood principle. Given observations Y , Maximum likelihood tries to find θ such that $P(x_n | \theta)$ gets maximum value. For the sake of efficient optimization, it is typical to introduce the log likelihood function defined as follows:

$$\ln P(x_n | \theta) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^k \pi_k N(x_n | \mu_k, \Sigma_k) \right\} \quad (9)$$

The model parameters are optimized in Matlab using Expectation-maximisation (EM) so as to maximize the likelihood of the observed data. A good reference on the topic of expectation maximization is provided by Bishop [16]. The EM algorithm is initialized by centering the GMM components on randomly selected samples from the training set, it can not guarantee to find the global optimum, so that it is expected that the performance from different optimization runs will differ slightly. The EM optimization step is thus repeated a number of times, and the solution which offers the best results is subsequently selected.

Once a GMM is constructed by using the normal data set, it is then used to detect outlier. For each new input, GMM provides $P(x_n | \theta)$ (see Equation (8)), the unconditional probability density, which indicates how the input follows the probability distribution of the GMM trained by normal dataset. The GMM outputs corresponding to the outlier data should be enough smaller than outputs of the GMM for normal data, which is usually set to 5% in practice [18]. Thus, log likelihood can be a good quantification indication for evaluating. In general, since the log likelihood value is smaller than zero, to improve its intelligibility, negative log likelihood probability (NLLP) (Equation 10) is used as the quantification indication in this study.

$$\text{NLLP} = -\log P(x_n | \theta) \quad (10)$$

4. Experiment and Result Analysis

4.1. Data Description of Energy Disaggregation In House

In order to evaluate the proposed outlier detection approach in energy disaggregation, the actual electricity consumption measurements released by Kolter et al. [3] at <http://redd.csail.mit.edu>. are used here. The dataset contains both appliance and total consumption level data from six households in the USA collected in April and May 2011. The REDD consists of high frequency, 15kHz, whole home current and voltage data, low frequency power data of individual circuits and plugs sampled at an order around 1Hz. Some of the houses contain a single appliance (e.g., a dishwasher), other houses contain multiple appliances (e.g., lights, kitchen outlets), there are approximately 20 consecutive days of measurements available for each house.

4.2. Experiments and Result Analysis

In this study, we chose the energy consuming appliances in houses #5, and the 21 original features from the 26 multiple appliances (General information about experiments datasets which used is briefly introduced in Table 1). In order to investigate the effectiveness of the proposed approach, we extracted 1000 examples as normal data for train. And other 530 example are chose as the test data set which may include normal data and outlier data. All the samples (attribute) have been normalized into [0, 1].

Table 1. General Information of Chosen Variables of House # 5 In REDD

Variable no.	Used variables	Variable no.	Used variables	Variable no.	Used variables
1	microwave	8	subpanel	15	lighting
2	outlets_unknown	9	electric_heat	16	dishwasher
3	furnace	10	electric_heat	17	electronics
4	outlets_unknown	11	Lighting	18	lighting
5	washer_dryer	12	bathroom_gfi	19	kitchen_outlets
6	washer_dryer	13	Lighting	20	kitchen_outlets
7	subpanel	14	refrigerator	21	outdoor_outlets

4.2.1 The Effectiveness of Data Point Distribution Using Different Subspace Learning Between PCA and LPP

In this subsection, we visually compared effectiveness using different subspace learning between PCA and LPP. LPP and PCA of subspace learning techniques are implemented for finding the original manifold of the training data set in low dimensional space. PCA is based on the global geometric structure, LPP is based on preserving the neighborhood structure and revealing the intrinsic manifold structure of the data.

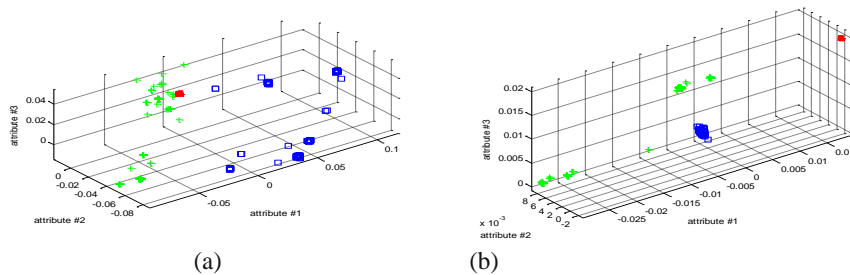


Figure 1. Data Point Distribution of GMM Based On the Subspace Learning By (A) PCA and (B) LPP

The LPP and PCA are implemented on the data set to obtain the three dimensional data vectors (in order to good data visualization). The projection of testing data on the selected three data vectors is performed by using Equation(7). Data point distributions based on the subspace learning in three dimensional data space by LPP and PCA are shown in Figure 1 (a), (b). From Fig.1 (a), it is found that method based on PCA is difficult to describe the data set when the distribution of data is non-Gaussian, multi-dimensional and multiple states processes. From Figure 1 (b), it is obvious that different Gaussian components are separated very effectively based on LPP. In comparison with PCA, LPP has a better separation of the four Gaussian components on low dimensional data space, and giving a better

description to reflect the intrinsic structure of the input data. The local structure is more important for outlier detection in energy disaggregation based on GMM.

4.2.2 Outlier Detection Results in Energy Disaggregation

The outlier detection results in energy disaggregation based on GMM with LPP or PCA compared on REDD are shown in figures.2-3, respectively.

During the training period, we use the normal dataset (i.e., 1000 samples) of choice to construct the corresponding GMM model and design two test scenarios are designed to examine the performance of Outlier detection in energy disaggregation based on GMM with LPP or PCA data preprocess.

LPP is first implemented based on the training data set according to Equation (7) to get subspace data projection. The first six eigenvectors corresponding to the larger six Eigenvalues are used to implement data projection. Then K-means clustering is implemented to identify the number of Gaussian components (i.e. the initialize parameters of GMM). (Experiments proved, GMM with 6 Gaussian components according to Equation (8). that is enough for modeling the train data set.)The trained GMM is used to detect outlier for the 530 test sample. The negative log likelihood probability of each test sample is shown in Figure 3. it is show that data points corresponding to the outlier state are well separated from the normal data points.

For the purpose of comparison, the PCA use the same dimensional data space for projection as LPP ways. The same test sample and initialize parameter of GMM are used. The results are shown in Figure 2. (The PCA and the original data set use the same initialize parameter of GMM and 530 test samples.)

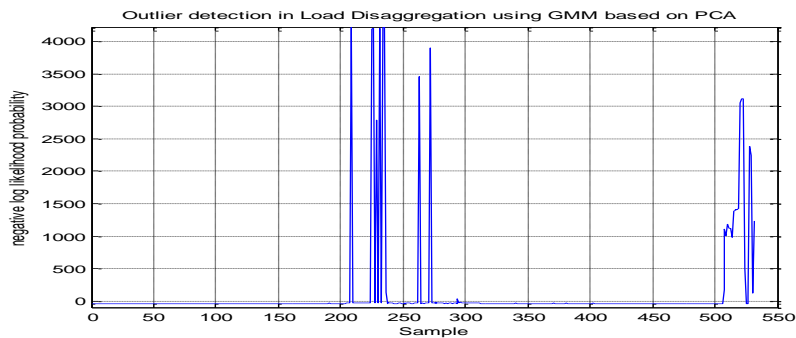


Figure 2. Outlier Detection in Energy Disaggregation using GMM based on PCA

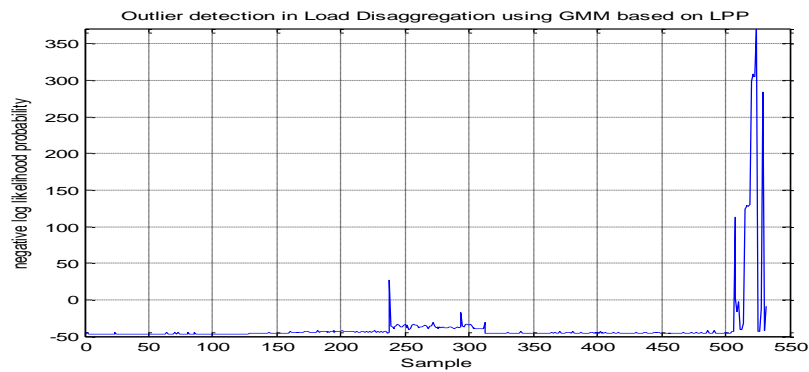


Figure 3: Outlier Detection in Energy Disaggregation using GMM based on PCA

As shown in Figures.2-3, we can observe the following: The proposed method achieves better performance compared on the method based on PCA method in outlier detection, From Figure 2, method based on PCA only identified part of the outliers and part of normal samples are classified incorrectly and its probability density is too small, especially the sample points from Num. 230 to 310. It is clear that GMM in REDD based on PCA can not correctly building the multiple states and multiple Gaussian model, and bad for our outlier detection. Figure 3 shows the final NLLP curve of outlier detection in energy disaggregation using GMM based on LPP. We can observe that a large proportion of the outliers is identified and at the same time is classified correctly normal samples with probability density is not too small. Data preprocess based on LPP can preserve local structure of the original normal data set, and better for energy disaggregation research.

In order to typically evaluate the performance of outlier detection in differently ways, two types of rates are used: namely the fraction true-positive (TP) and the fraction false-positive (FP) [7]. The TP defined in Equation (11) is computed as the fraction of the number of correctly detected normal samples to the total number of normal data. The FP defined in Equation (12) is computed as the fraction of the number of outlier examples that are incorrectly detected as normal data to the total number of outliers.

$$TP = \frac{\text{Normal samples correctly classified}}{\text{Total normal samples}} \quad (11)$$

$$FP = \frac{\text{Outliers incorrectly classified}}{\text{Total outliers samples}} \quad (12)$$

As shown in Tables 2, the outlier detection results based on GMM with LPP can significantly improve the performance of outlier detection in energy disaggregation, increase the fraction true-positive from 93.8% to 97%, decrease the fraction false-positive from 35.48% to 25.8%.

Table 2: Performance Compared between LPP and PCA

Data preprocess methods	FP	TP
LPP	25.8 %	97.0 %
PCA	35.4 8%	93.8 0%

5. Conclusion and Future Work

The use of outlier-free data is very important for building reliable system models with machine learning and data mining techniques for energy disaggregation. This paper has proposed an LPP-based GMM approach for outlier detection on REDD. The presented results clearly demonstrate that outlier detection using GMM based on LPP can obtain a fairly better and clearer separation outlier from normal data, at the same time maintain the original characteristics of the data. It showed stronger robustness, comparing with the PCA.

In the future, we would use outlier-free data of REDD to study energy, energy consumption patterns and sustainability problems.

Acknowledgements

We thank the anonymous reviewers for their very useful comments and suggestions. This work is supported by Science Fund for Creative Research Groups of the National Natural Science Foundation of China (61321003), Natural Science Foundation of China (61273169, 61134006, 61105080), the Scientific Research Fund of Hunan Provincial Education Department (13A016), the Science and Technology Planning Project of Xiangtan City Hunan Province (NY20141006), Hunan Provincial Natural Science Foundation of China(14JJ2099).

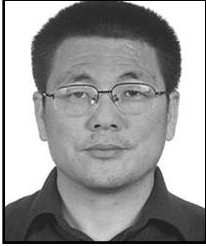
References

- [1] J. Z. Kolter and M. J. Johnson, "Redd: A public data set for energy disaggregation research." proceedings of the SustKDD workshop on Data Mining Applications in Sustainability.San Diego, CA,(2011)
- [2] J. Kelly and W. Knottenbelt, "Disaggregating multi-state appliances from smart meter data," SIGMETRICS'12 (2012).
- [3] M. Zeifman, "Disaggregation of home energy display data using probabilistic approach," Consumer Electronics, IEEE Transactions on, vol. 58, (2012), pp. 23-31.
- [4] H. Shao, M. Marwah, and N. Ramakrishnan, "A temporal motif mining approach to unsupervised energy disaggregation." Proceedings of the 1st International Workshop on Non-Intrusive Load Monitoring, Vol. 7, (2012),Pittsburgh, PA, USA.
- [5] J. C. Munoz and J. Chen, "Removal of the effects of outliers in batch process data through maximum correntropy estimator," Chemometrics and Intelligent Laboratory Systems, vol. 11, (2012), pp. 53-58.
- [6] V. J. Hodge and J. Austin,"A survey of outlier detection methodologies," Artificial Intelligence Review, vol. 22, (2004), pp. 85-126.
- [7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys (CSUR), vol.41, (2009), pp. 1-74.
- [8] S. Wang, J. Yu, E. Lapira, and J. Lee,"A modified support vector data description based novelty detection approach for machinery components," Applied Soft Computing, vol.13, (2013), pp. 1193-1205.
- [9] B. Liu, Y. Xiao, L. Cao, Z. Hao, and F. Deng, "SVDD-based outlier detection on uncertain data," Knowledge and information systems, vol.34, (2013), pp. 597-618.
- [10] S. Cateni, V. Colla, and M. Vannucci,"Outlier detection methods for industrial applications," Advances in Robotics, Automation and Control, (2008), pp. 265-282 .
- [11] J. Fan and R. Li,"Statistical challenges with high dimensionality: Feature selection in knowledge discovery," arXiv preprint math,(2006)
- [12] X. Niyogi, "Locality preserving projections", Adv. Neural Infor. Process. Systems,vol.16, (2003), pp. 153-160.
- [13] J. Kawahara, C. McIntosh, R. Tam, and G. Hamarneh, "Augmenting Auto-context with Global Geometric Features for Spinal Cord Segmentation", in Machine Learning in Medical Imaging Springer, (2013), pp. 211-218.
- [14] J.-B. Yu,"Bearing performance degradation assessment using locality preserving projections," Expert Systems with Applications, vol.38, (2011), pp.7440-7450.
- [15] C. M. Bishop, Pattern recognition and machine learning. springer New York, (2006).
- [16] N. Balakrishnan and M. H. Ling, "Expectation Maximization Algorithm for One Shot Device Accelerated Life Testing with Weibull Lifetimes, and Variable Parameters over Stress," Reliability, IEEE Transactions on , vol.62, (2013), pp.537-551.
- [17] X. He, D. Cai, and W. Min,"Statistical and computational analysis of locality preserving projection", In Proceedings of the 22th International Conference on Machine Learning,(2005),pp.281-288.Bonn, Germany
- [18] J. Yu, "A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes," Chemical Engineering Science,vol.68, (2012), pp.506-519.

Authors



Xiu-Ming Tang was born in YongZhou, Hunan, china in 1977. Correspondence author. He received his master's degree in the college of Electrical Engineering, Wuhan University (WHU), Wuhan, China, in 2008. She is currently working towards a Ph.D degree in Electrical Engineering at Wuhan University. She has been working on load model and data mining.



Rong-xiang was born in ShuangFeng, Hunan, China in 1965. He received the Ph.D. degree from Huazhong University of science and technology, Wuhan, China, in 1997, He is the professor with the college of electrical engineering, WHU, where he has been working on research of power electronic technology.



Jun Chen was born in XiangTan, Hunan, china in 1977. He is currently Lecturer with the Institute of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, Hunan, China. where he has been working on modeling and optimal control of complex industrial process and data mining.

