Hierarchical classification method of electricity consumption industries through TNPE and Bayes

Zi-Wen Gu1, Peng Li1, Xun Lang1, Xin Shen2, Min Cao2 and Xiao-Hua Yang2

Abstract
As the multi-daily electricity consumption behaviors have the strong characteristics of dynamicity, nonlinearity and locality caused by temporal manifold structure, the existing methods are difficult to fine-grained and accurately classify it. To solve this problem, this paper proposes a hierarchical classification method based on the temporal extension of the neighborhood preserving embedding algorithm (TNPE) and Bayes. The input data are multi daily-load curves of a single consumer, including power-hour-day three dimensions, which contains the full information of the user’s consumption behaviors not only in hours, but also in days. Firstly, electricity consumption behaviors are divided into routine and non-routine types by k-means clustering algorithm. Secondly, the load feature mapping matrix of different industries is extracted through the TNPE, and each TNPE model can regard as one binary classifier, so the multi-classifier is constructed through multiple TNPE models. Finally, by converting the feature similarity between samples into probabilities, a Bayesian model is established to realize which the power consumption type belongs to. The case results show that this method can effectively recognize the local dynamic features in the temporal load data, and obtain a higher classification accuracy through a smaller number of training samples.

Keywords
Industry classification, electricity consumption behavior analysis, dynamic feature extraction, temporal extended manifold learning, data-driven

Date received: 4 January 2021; accepted: 28 January 2021

Introduction
With the development of smart grids within an active power distribution network, the analysis of user behavior from massive user data becomes crucial for realizing and optimizing a multi-energy complementary system. Electricity consumption behavior reflects the electricity consumption patterns in various fields relevant to the national economy and provides an important basis to the power grid for analyzing electricity consumers.1,2 However, while huge amounts of data have been accumulated in different information systems under the electric power sector, the value of such data has yet to be fully utilized.3,4 The categorization of power consumers effectively using the data-driven approach is an issue of wide concern for data mining in the current smart grid system.5

The characteristics of electricity consumption are most directly reflected by the load curve. Therefore, it has been the major research object for analyzing the electricity consumption behavior.6 The load curve can be analyzed using either clustering or classification.

Examples of typical clustering algorithms include the fuzzy C-means clustering,7 k-means clustering,8 density-based clustering,9 and consensus clustering.10 Yang et al.11 proposed the k-shape clustering algorithm based on load shape, which is to detect different levels of building energy consumption patterns, and further use the clustering results to improve the accuracy of the prediction model. Xiang et al.12 proposed a shape clustering method based on the segmented slope to solve the problem that the Euclidean distance as a measure of similarity is not enough to reflect the shape similarity of the load curve. In order to cluster the load curve data

1School of Information Science and Engineering, Yunnan University, Kunming, China
2Yunnan Power Grid Co., Ltd., Kunming, China

Corresponding author:
Peng Li, School of Information Science and Engineering, Yunnan University, East Outer Ring South Road, University Town, Chenggong Area, Kunming, Yunnan 650500, China.
Email: lipeng@ynu.edu.cn

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
more accurately, Iglesias and Kastner proposed a new method to calculate the Pearson distance, which helps to maintain the overall and local similarity of the load profile. The clustering algorithms are unsupervised learning methods for data processing; also, clustering result requires people to define the category. The classification algorithms offer a supervised learning approach to process the data. Besides, classification results do not require people to define the category. Examples of typical classification algorithms include multilayer feedforward network (MFN), artificial neural networks (ANN), support vector machine (SVM), and extreme learning machine (ELM). Kim and Lee adopt the Multi-Feature Combination that one of feature extraction techniques commonly used in audio signal processing to process power signals and select the Multi-Layer LSTM network as the classification model for further improvements. In Yang et al., a new semisupervised multilabel deep learning-based framework is proposed to mitigate the reliance on large labeled datasets. Varga et al. proposed a load profiles management software framework for encoding and classification in real-time, which tolerates defects and time shifts in input, so as to always provide accurate, fast, and reliable output.

Some researches analyze the consumption behavior based on the single daily-load curves in the power-hour dimensions. However, there is a stronger dynamic correlation between daily electricity consumption behaviors in the local time domain, and multi daily-load curves in the power-hour-day dimensions can more accurately reflect the user’s behavior information. Generally, multivariate statistical methods are used to reduce the redundancy and extract the true structure of the data. Compared with principal component analysis (PCA), independent component analysis (ICA), canonical variable analysis (CVA), etc., TNPE allows obtaining the global data feature from the local geometric structure and the temporal characteristics. Therefore, TNPE can more effectively extract the dynamic, non-linear, and local characteristics of the load data.

In this paper, a classifier based on TNPE and Bayes is proposed to identify the electricity consumer industries. Feature detection by TNPE algorithm can be regarded as a binary classifier, so the multi-classification can be constructed by multiple TNPE models. This paper mainly has the following three contributions. Firstly, the daily load data are divided into routine behaviors and non-routine behaviors by using the k-means algorithm based on Pearson correlation coefficient. Secondly, the load feature mapping matrix of different industries is extracted through the TNPE, and each TNPE model can regard as one binary classifier, so the multi-classifier is constructed through multiple TNPE models. Finally, by converting the feature similarity between samples into probabilities, a Bayesian model is established to realize which the power consumption type belongs to.

The organization of the rest of this paper is as follows. In section 2, the dynamic, nonlinear, and local characteristics of electricity consumption behaviors are analyzed. Section 3 briefly introduces the basic theory. Section 4 introduces the specific training and classification steps of the classification model proposed in this paper. OpenEI dataset is utilized for case studies in Section 5 to demonstrate the feasibility of the proposed approach. Finally, the conclusions are presented in Section 6.

**Problem statement**

**Dynamic characteristic of electricity consumption**

The load curve has been considered as the main research object for analyzing the electricity consumption behavior. Generally, the consumption characteristics of users are analyzed through the single daily-load curve as shown in Figure 1(a), which including power-hour two dimensions. However, consumption behavior usually has dynamic daily characteristic, and a single daily-load curve cannot reflect the full characteristics of users. As shown by the multi daily-load curves in Figure 1.
The similarity between the morphologies become weaker with increasing temporal distance between them. The temporal correlation in days.

Figure 1(b), which including power-hour-day three dimensions. Within the time neighbor of \( x_t \), \( \{x_{t-2}, x_{t-1}, x_{t+1}, x_{t+2}, x_{t+m}\} \) is similar to \( x_t \), the daily load \( x_{t+3} \) and \( x_t \) have a big difference. However, the temporal correlation between \( x_{t+m} \) and \( x_t \) will become weaker with increasing temporal distance between them. The similarity between the morphologies of \( x_{t+m} \) and \( x_t \) is more likely to be a coincidence. Hence, the single daily-load curve only reflects the external shape characteristics of the load data in hours, but the multi daily-load curves can additionally reflect the temporal correlation in days.

Nonlinearity and local characteristics of electricity consumption
Due to the load data is one temporal series, the samples in the time neighbors have stronger correlation. This means that the user’s electricity consumption behaviors pay more attention to the local feature information of the load data. As shown by the daily load curve in Figure 2(a), Point A and point B are the power consumption values at 7:00 and 11:00 in a day, and there is a peak during these two moments. If the electricity consumption behavior analysis is applied by a global method, such as by the geodesic Path 1, it can only be concluded that the electricity consumption at 11:00 is higher than 7:00, but the fact that the electricity consumption peak exists is ignored. However, if the behavior analysis use a local method, such as by the manifold Path 2, the actual electricity consumption information can be fully reflected. Therefore, the load curve has characteristics of nonlinearity and locality. Similarly, when performing dynamic electricity consumption behavior analysis in the multi daily-load curves as shown in Figure 2(b), the real consumption behaviors between A and B is not represented by the geodesic path 1, but by the manifold path 2. So it is clear that nonlinearity and local characteristics of electricity consumption are caused by the manifold structure of it.

Basic theory and analysis

The TNPE algorithm
Through the analysis in section 2, the dynamic, non-linear, and locality exist in electricity consumption behavior, and the load data has a manifold structure. Hence, the classification problem of electricity consumption behaviors based on the multi daily-load curves should be solved by the manifold learning algorithm. Manifold learning, a branch of nonlinear dimensionality reduction, has become a hot topic in the field of information science since its proposition. 27 The TNPE algorithm has been proposed in a past study to enable the effective extraction of dynamic local features in multivariate temporal data through manifold learning. 28 TNPE allows the original data \( X = \{x_1, x_2, \ldots, x_N\} \in \mathbb{R}^d \) to be projected into the low-dimensional feature space through the mapping matrix \( A(a_1, a_2, \ldots, a_d) \). This procedure will generate a new sequence \( Y = \{y_1, y_2, \ldots, y_N\} \in \mathbb{R}^d \) satisfying \( Y = A^T X \) where \( d < D \). In the TNPE algorithm, the \( k \) nearest neighbors of each data point \( x_t \) will first be identified through the time window and the Euclidean distance, and then used to construct its temporal neighborhood \( P_t = \{p_{t1}, p_{t2}, \ldots, p_{tk}\} \) and spatial neighborhood \( S_t = \{x_{t1}, x_{t2}, \ldots, x_{tk}\} \). Here, the value of \( k \) is determined by the size of the time window.

Suppose data \( y_t \) is the mapping of data \( x_t \) in low-dimensional space. As shown in equation (1), \( \Phi(W) \) is used to obtain the local linear weight \( W_t \) between the original data \( x_t \) and its neighborhoods in the high-dimensional space. In the idea of the TNPE algorithm, the neighborhood weight \( W_t \) of the data \( x_t \) in the high-dimensional space can be preserved in the low-dimensional space. In equation (2), \( \Phi(y) \) reconstructs the data \( y_t \) in the low-dimensional space through neighborhood weight \( W_t \). Therefore, the reconstruction weights \( W_{tS} \) and \( W_{tP} \) of data \( x_t \) in the high-dimensional space are obtained by equation (1) firstly. The data \( y_{tS} \) and \( y_{tP} \) are then reconstructed by equation (2) with identical weights in the low-dimensional space.
The ultimate objective of the TNPE algorithm is to find \(d\) projection vectors \(\{a_1, \cdots, a_d\}\) to form the mapping matrix \(A\), which can minimize the information loss of the structural features during the mapping process of the data structure. The objective function is given as follows:

\[
J(a) = \min (\mu \Phi(y_S)) + ((1 - \mu) \Phi(y_P))
\]  

where \(\mu\) is the impact factor of the neighborhood, which measures the proportion of the neighborhoods \(S\) and \(P\) in the entire data during the mapping process.

Considering that \(y = a^T X\), (3) can be transformed as follows:

\[
J(a) = \min (a^T XMX^T a)
\]  

s.t. \(y^T y = a^T XX^T a = 1\)

\[
M = \mu M_P + ((1 - \mu) M_S)
\]

\[
M_P = (I - W_P)^T (I - W_P)
\]

\[
M_S = (I - W_S)^T (I - W_S)
\]

Equation (4) can be transformed into a constrained optimization problem using the method of Lagrange multipliers. The solution can then be obtained using the generalized eigenvalue decomposition method, which yields the following:

\[
XMX^T a = \lambda XX^T a
\]  

To ensure that the information of the characteristic structure is minimized after data mapping, the eigenvectors associated with \(d\) minimum nonzero eigenvalues in equation (8) are used to form the mapping matrix \(A\).

The feature extraction capability of the TNPE algorithm is related to three parameters, namely, the size of the time window \(k\), the impact factor of the neighborhood \(\mu\), and the reduction in dimension \(d\). Currently, these three parameters are adjusted manually according to the requirements of problem solving and the characteristics of the data. The impact of parameter selection on the algorithm has been discussed in detail in a past study [21]. Further discussion is beyond the scope of this paper.

**Define of feature similarity**

According to the introduction in section 2.1, a single daily load curve cannot reflect the user’s dynamic behavior characteristics. For feature detection purpose between two periods of the multi-daily load curves, the \(T^2\) and \(SPE\) statistics were established as follows:

\[
\begin{align*}
\Phi(W) &= \sum_{i=1}^{n} \left\| x_i - \sum_{j=1}^{n} w_{ij} x_j \right\|^2 \\
\text{s.t.} &\sum_{j=1}^{n} w_{ij} = 1
\end{align*}
\]  

\[
\Phi(y) = \sum_{i=1}^{n} \left\| y_i - \sum_{j=1}^{n} w_{ij} y_j \right\|^2
\]

where \(\Phi(W)\) is the projection of the original data \(x\) on the low-dimensional space and \(A^{-1}\) is the sample covariance matrix of \(Y\).

\[
SPE = \left\| (I - AA^T) x \right\|^2
\]

Feature similarity coefficients are defined to describe the level of similarity. For the convenience of description, \(T^2\) and \(SPE\) are defined as the statistical parameters of the training sample, and \(t^2\) and \(spe\) are the statistical parameters of the test sample. They are all calculated by equations (9) and (10). If the \(t^2\) and \(spe\) statistics of the test sample are below the \(T^2\) and \(SPE\) statistics limits of the training sample, respectively, then the test sample conforms to the same structural features as the training sample. In this case, \(h(x_{new})\) is assigned with a value of 1. Otherwise, \(h(x_{new})\) equals 0:

\[
h(x_{new}) = \begin{cases} 
1 & \text{if } (t^2 \leq T^2) \text{ and } (spe \leq SPE) \\
0 & \text{if } (t^2 > T^2) \text{ or } (spe > SPE)
\end{cases}
\]

Hierarchical classification method based on TNPE and Bayes

Figure 3 briefly introduces the framework of hierarchical classification based on the TNPE and Bayes. According to section 3.2, feature detection can regard as a binary classifier. Then in order to achieve multiple classifications, the “one-against-the rest” strategy is adopted. Therefore, the \(K\) numbers of industries needs to design \(K\) feature detection models. Finally, the feature detection results are input to the Bayesian classifier to get the type of the user. The main blocks of this model are follows:

1. Separation model: electricity consumption behaviors are divided into routine and non-routine types by \(k\)-means clustering algorithm based on the Pearson correlation coefficient

2. Hierarchical training model: the binary classifiers of different electricity behaviors in each industry are trained through the TNPE algorithm.
Hierarchical classification model: the multi-classifier is combined by multiple binary classifiers, and a Bayesian model is established to realize the user’s consumption type.

A supervised dimensionality reduction model is shown in Figure 4, satisfying $Y = A^T X$ where $d < D$.

In the process of dimensionality reduction, supervised dimensionality reduction projects different types of data through different mapping relationships, while unsupervised dimensionality reduction projects all types of data through the same mapping relationship. The supervised dimensionality reduction is to make the data easier to distinguish, while the unsupervised dimensionality reduction is to keep data information as much as possible. Therefore, compared with the supervised dimensionality reduction, the unsupervised dimensionality reduction has little effect on data differentiation, and may make data points mixed together and indistinguishable.

By projecting different types of data into the low-dimensional spaces through different mapping relationships, the homogenous structural features are minimized between different types of data. However, the type of mapping relationship cannot be known in the supervised dimensionality reduction classification model. Figure 5 shows the basic principle of obtaining the type based on feature statistics in an ideal situation. According to the discussion in section 3.2, if the test sample has a similar feature structure as the training model, then the test sample can be considered the same type as the training model when the feature statistics ($t^2$ and SPE) of the test sample are less than the feature statistics ($T^2$ and $SPE$) of the training model.

Separation model
K-means is an unsupervised clustering algorithm. The objective of k-means algorithm is to form $k$ clusters $C = \{C_1, C_2, ..., C_k\}$ through the clustering of the data set $X \subseteq \mathbb{R}^D$. The specific sample is assigned to its closest cluster by calculating the center distance $\text{dist}(X, C)$ between the sample and all $k$
The objective is to minimize the intracluster error $E$:

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} \text{dist}(x, \mu_i)$$  \hspace{1cm} (13)$$

where $\mu_i$ is the center of cluster $C_i$ and $\text{dist}$ represents the distances between the data.

During the $k$-means clustering process, the distance $\text{dist}$, which is used as a measure of similarity between data, can be expressed in several ways including the Euclidean distance, Minkowski distance, cosine correlation coefficient, and Pearson correlation coefficient. As the users will develop a fixed electricity consumption habit over time, $\text{dist}$ will exhibit a fixed form in the daily load curve. Therefore, the morphological similarity of the load curve should be emphasized when analyzing the electricity consumption behavior of users. The Pearson correlation coefficient focuses on capturing the direction of the morphological variation of the curve, and it does not require the specific normalization process of the data; thus, it is a good measure for expressing the similarity between the load curves. In this study, the Pearson similarity $\text{dist}_P$ is defined as follows:

$$\text{dist}_P = 1 - \frac{\sum x_i x_j - \sum x_i \sum x_j}{\sqrt{\left(\sum x_i^2 - \left(\sum x_i\right)^2\right) \left(\sum x_j^2 - \left(\sum x_j\right)^2\right)}}$$  \hspace{1cm} (14)$$

where $D$ is the dimension of the data and $\text{dist}_P \in [0, 1]$. The greater the similarity between $x_i$ and $x_j$, the smaller is the $\text{dist}_P$.

The original temporal load data are separated into multiple temporal load subsets by using the $k$-means algorithm based on the Pearson correlation coefficient:

$$U = C_1 \cup C_2 \cup \cdots \cup C_k$$  \hspace{1cm} (15)$$

where $U$ is the original load, and $C_i$ represents the load subsets.

In this paper, the cluster with the largest sample size is defined as the routine behaviors, and the remaining clusters are regarded as non-routine behaviors. Let $\text{num}_i$ represent the number of samples in cluster $C_i$, and the definitions of routine behaviors $r$ and non-routine behaviors $q$ are as follows:

$$\begin{cases} r = C_{\text{label}_r} \\ \text{s.t.} \text{num}(C_{\text{label}_r}) = \max(\text{num}_1, \text{num}_2, \cdots, \text{num}_k) \end{cases}$$  \hspace{1cm} (16)$$

$$\begin{cases} q = \bigcup_{i \in \text{label}_q} C_i \\ \text{s.t.} \text{num}(C_{\text{label}_q}) \neq \max(\text{num}_1, \text{num}_2, \cdots, \text{num}_k) \end{cases}$$  \hspace{1cm} (17)$$

Hierarchical training model

The objective of the hierarchical training model for electricity consumption behavior is to extract the characteristics of electricity consumption behavior through an optimized selection of training samples. The flowchart of the training model is shown in Figure 6. The model consists of two processes, namely, offline modeling and online training. The offline modeling will extract the electricity consumption characteristics of a user in the corresponding category. The online training will then update the electricity consumption characteristics of the sample library in the offline modeling process. Selecting preferential training samples through online training will prevent not only the overfitting caused by the incomplete extraction of structural features from the data, but also the extraction of redundant structural features due to the excessive size of the sample library.

The detailed procedures for offline modeling are described as follows:

1. Separation of electricity consumption behaviors. The electricity consumption behavior of the first user is separated in the training data according to section 3.1 for initializing the sample training libraries $R$ and $Q$ associated with routine and non-routine electricity consumption behaviors, respectively.

2. Extraction of feature-mapping matrix in the sample library. The feature-mapping matrices $A_R$ and $A_Q$ of the sample libraries associated with routine and non-routine electricity consumption behaviors, respectively, are extracted using the TNPE algorithm.

3. Calculate the feature statistics. The feature statistics ($T^2$ and SPE) for the routine and non-routine electricity consumption behaviors are calculated separately.
The detailed procedures for online training are described as follows:

(1) Separation of electricity consumption behaviors. The electricity consumption behaviors are separated based on the load data of the new user analyzed in the offline modeling process. The sample sets \( r \) and \( q \) associated with the routine and non-routine electricity consumption behaviors, respectively, for the new user are then extracted and stored in the data register.

(2) Calculation of feature statistics. The test data are projected onto the low-dimensional feature space according to the feature-mapping matrices \( A_R \) and \( A_Q \) obtained from offline modeling. The feature statistics \( T^2 \) and \( SPE \) of the routine and non-routine electricity consumption behaviors in the test data are then calculated using equation (9) and equation (10).

(3) Verification of the update of the training sample library. The level of feature similarity between the training model and the test data is calculated according to equation (12). If the feature similarity of the electricity consumption behavior from the new user is less than 90%, then this sample will be updated to the corresponding sample training library through the data register. Furthermore, the feature-mapping matrix \( A \) and the feature statistics \( T^2 \) and \( SPE \) will be updated in the sample library. Otherwise, the training process will be performed for the next round of users. The relationship between the update statuses of the training samples in each training round is given as follows:

\[
\text{Train}_i \supseteq \text{Train}_{i-1}
\]  

where \( \text{Train}_i \) represents the training sample in the \( i \)-th training round.

The structural feature of the user’s consumption behaviors mainly refers to the shape of the load curves, and it has scale changes, displacement changes, and noise changes.

**Scale changes.** There are differences in the scale of the load curve. For example, the load curve \( X_1 \) is transformed into \( X_2 = aX_1 + b \), where \( a \) and \( b \) are constants, and the similarity between \( X_1 \) and \( X_2 \) remains unchanged.

**Displacement changes.** The phase between the load curves has a certain deviation, such as \( X_2 = X_1(t - t_0) \), the similarity between \( X_1 \) and \( X_2 \) remains unchanged.

**Noise changes.** When two load curves have similar shapes, but are interfered with by different degrees of noise, the similarity between the two remains unchanged.

The update of the training sample set will only be triggered by the occurrence of new structural features in the electricity consumption behavior from a user in the corresponding industry. Therefore, the relationship
between the structural features of the training samples in each training round is given as follows:

\[ \text{Feature}_i \supseteq \text{Feature}_{i-1} \]  

(19)

where \( \text{Feature}_i \) represents the structural feature of the training sample in the \( i \)-th training round.

The original structural features of the data in the training samples are used as the local components of the updated training samples. The original samples can still be tested effectively using the feature detection statistics \( T^2 \) and \( SPE \).

Hierarchical classification model

The hierarchical classification model for electricity consumption behavior seeks to identify and classify the electricity consumption features. The flowchart of the model is shown in Figure 7. The feature similarity can be converted into probability based on the statistical characteristics of feature detection. The priori probability can be converted into a posterior probability through the Bayes classification model.30

The detailed procedures of hierarchical classification (H-TNPE-Bayes) model are described as follows:

1. Separation of electricity consumption behaviors. The routine electricity consumption behavior \( r \) and non-routine electricity consumption behavior \( q \) are first separated from the user’s load data.

2. Calculation of the priori feature probability. The feature similarity \( H_1^i(r) \) of daily electricity consumption behavior in each TNPE model of routine behavior is calculate by equation (12). The prior probability of testing a user’s electricity consumption behavior in the training model is given as follows:

\[ P_X(\omega_i) = \frac{H_1^i(r)}{\sum H_1^i(r)} \]  

(21)

where \( H_1^i(r) \) represents the feature similarity between the routine behavior \( r \) of the test user and the TNPE model \( i \).

3. Calculation of the conditional feature probability. Calculate the feature similarity \( H_2^i(q) \) of...
daily electricity consumption behavior in each TNPE model of non-routine behavior by equation (12). If all the conditional feature similarities are 0, then the feature of the test sample is too ambiguous, which causes the classification to fail. Otherwise, the conditional probability of the electricity consumption feature of the test user being included in the training model is given as follows:

\[
P_X(F_j|w) = \frac{H^2_j(q)}{\sum H^2_j(q)}
\]

where \(H^2_j(q)\) represents the feature similarity of the electricity consumption behavior \(q\) in the TNPE model \(j\) with respect to the test user.

(4) Bayes classification. As shown in equation (20), knowing the probability of the electricity consumption feature \(P(F_j)\) in the Bayes classifier in advance will not affect the classification results. Therefore, the prior probability \(P(w)\) is converted to the posterior probability \(P(w|F_j)\) through the conditional probability \(P(F_j|w)\). However, all the conditional probabilities \(P(F_j|w)\) will become 0 if the structural features of the test data are ambiguous. Therefore, the posterior probability \(P(w|F_j)\) is estimated using the largest prior probability \(P(w)\). The final Bayes classification conditions are given as follows:

\[
P(w|F_j) = \begin{cases} P(F_j|w)P(w) & P(F_j|w) > 0 \\ P(w) & P(F_j|w) = 0 \end{cases}
\]

(5) Category decision. The category with the highest probability is the user’s industry label.

\[C_j = \arg \max \{P_K(w_i|F_j)\}\]

where \(C_j\) is the category to which the electricity consumption behavior \(j\) belongs.

**Case analysis**

**Experimental environment**

The commercial user load data released publicly on the website of the US Department of Energy (OpenEI, https://openei.org/datasets/files/961/pub) are used to validate the effectiveness of the model proposed in this paper. Each set of data comprises 365 days of electricity load information for 16 industries collected at a sampling time of 1 h over 1 year. In this study, 584,000 sets of load data were considered as the training samples, whereas another 584,000 sets of load data were used as the test samples. The classification performance is evaluated based on three indicators, namely, the classification accuracy (CA), the classification average accuracy (CAA), and Macro F1-measure (Macro-F1):

\[
CA_i = \frac{TP_i + FN_i}{TP_i + FP_i + TN_i + FN_i} \times 100\% \tag{25}
\]

\[
CAA = \frac{1}{n} \sum_{i=1}^{n} CA_i \tag{26}
\]

\[
\begin{aligned}
Precision_i &= \frac{TP_i}{TP_i + FP_i} \\
Recall_i &= \frac{TP_i}{TP_i + FN_i} \\
F1_i &= 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \\
Macro-F1 &= \frac{1}{n} \sum_i F1_i
\end{aligned} \tag{27}
\]

where \(TP\) is to predict the positive class as a positive class; \(TN\) is to predict the negative class as a negative class; \(FP\) is to predict the negative class as a positive class; \(FN\) is to predict a positive class as a negative class. \(i\) represents the \(i-th\) category.

CA reflects the effectiveness of the classification of the proposed model for various industries. CAA provides a comprehensive measure for the classification performance of the proposed model for all industries. Macro-F1 can reduce the impact of categories balanced, and reflect the performance of the classifier with precision and recall.

**Comparison of performances between different classification models**

The TNPE algorithm is a dimensionality reduction algorithm based on supervised learning. To validate the effectiveness of the TNPE algorithm, we compared its performance with that of principal component analysis (PCA), which is an unsupervised dimensionality reduction algorithm. In addition, the effectiveness of the Bayes classification method based on feature detection was validated by comparing it with the ELM described in a past literature study. In this study, the industries associated with the electricity users are classified based on the daily load data of the users over one year. However, the original classification result obtained from the ELM is based on the daily load data for a single day. Therefore, the condition for realizing industry classification through the ELM is given as follows:

\[C = \arg \max \{\text{Num}_{C_1}, \text{Num}_{C_2}, \ldots, \text{Num}_{C_n}\}\]

where \(\text{Num}_{C_i}\) represents the number of daily load data being classified into category \(C_i\) among all the available data for the specific user.

Both the TNPE and PCA algorithms work by first projecting the original data to the low-dimensional space through dimensionality reduction, then converting the classification problem into a probability issue based on feature detection, and finally realizing the classification through the Bayes model. However, the mapping of data structural features and the output of classification results are achieved through neural networks in the ELM. Therefore, the detailed configuration of the comparative experiments conducted in this study is summarized in Table 1. During the comparative experiments, an
The S-type activation function is employed by the ELM algorithm along with 60 hidden nodes. The number of variables reduced by the dimensionality reduction is set as four in the PCA algorithm. The size of the time window, the magnitude of the neighborhood impact factor, and the number of variables reduced by the dimensionality reduction are set as 5, 0.4, and 4, respectively, in the TNPE algorithm.

While PCA is an unsupervised dimensionality reduction algorithm, the classification is still based on feature detection, which requires the extraction of the feature statistics from the training samples. Therefore, the five groups of classification models used in the comparative experiments are all supervised models. The other classification models except ELM use the online training method in the 4.2 section to reduce the training dataset. The numbers of training samples used in the experiments are shown in Figure 8. As PCA-Bayes, TNPE-Bayes, H-PCA-Bayes, and H-TNPE-Bayes can use the training structure described in section 3.3 to optimize the sample selection without affecting the feature detection of the previous samples, the sizes of the training samples are greatly reduced in these methods compared with that in the ELM. Furthermore, the selection of preferential training samples allows the detection of whether the structural features of the data have changed. Compared with the PCA-Bayes and TNPE-Bayes classification models, the H-PCA-Bayes, and H-TNPE-Bayes classification models require a much smaller training sample size due to the application of the divide-and-conquer algorithm. Thus, the strategy of separating electricity consumption behaviors can reduce the complexity of the structural features of the data effectively.

Table 1. Experimental control group.

| Experimental group | Description                                      |
|--------------------|--------------------------------------------------|
| ELM                | Based on neural network classification           |
| PCA-Bayes          | Unsupervised dimensionality reduction – classification based on feature detection |
| TNPE-Bayes         | Supervised dimensionality reduction – classification based on feature detection |
| H-PCA-Bayes        | Hierarchical – Unsupervised dimensionality reduction – classification based on feature detection |
| H-TNPE-Bayes       | Hierarchical – Supervised dimensionality reduction – classification based on feature detection |

Table 2. Classification results of classification models.

| User category       | ELM | PCA-Bayes | TNPE-Bayes | H-PCA-Bayes | H-TNPE-Bayes |
|---------------------|-----|-----------|------------|-------------|--------------|
| Full service restaurant | 95  | 100       | 100        | 100         | 100          |
| Hospital            | 98  | 99        | 99         | 98          | 99           |
| Large hotel         | 79  | 100       | 99         | 100         | 100          |
| Large office        | 96  | 97        | 99         | 96          | 100          |
| Medium office       | 81  | 100       | 100        | 98          | 99           |
| Midrise apartment   | 100 | 100       | 100        | 100         | 100          |
| Out patient         | 100 | 100       | 97         | 100         | 100          |
| Primary school      | 100 | 100       | 100        | 99          | 100          |
| Quick service restaurant | 94  | 100       | 100        | 100         | 100          |
| Secondary school    | 45  | 32        | 89         | 71          | 99           |
| Small hotel         | 93  | 100       | 100        | 100         | 100          |
| Small office        | 100 | 100       | 100        | 100         | 100          |
| Stand-alone retail  | 83  | 95        | 98         | 95          | 99           |
| Strip mall          | 88  | 95        | 97         | 99          | 99           |
| Super market        | 100 | 93        | 98         | 100         | 100          |
| Warehouse           | 100 | 100       | 100        | 100         | 100          |
| CAA                 | 90.75 | 94.44  | 98.5      | 97.25       | 99.81        |
| Macro-F1            | 91.69 | 95.18   | 98.06     | 97.47       | 98.06        |

Figure 8. Training sample size of each classification model.

Table 2 shows the classification results of the five classification models. By reflecting on the CA of each industry, ELM performs poorly in the classification of several industry types, such as Large Hotel, Medium
Table 3. Computational complexity of classification models.

| Experimental group | Computational complexity |
|--------------------|-------------------------|
| ELM                | $O(n^3mhn)$             |
| PCA                | $O(n^2m + n^3)$         |
| TNPE               | $O(m\log(m)n\log(k)) + O(mnk^3) + O(dm^2)$ |
| H-PCA              | $O(mnct) + O(m\log(m)n\log(k)) + O(mnk^3) + O(dm^2)$ |
| H-TNPE             | $O(mnct) + O(m\log(m)n\log(k)) + O(mnk^3) + O(dm^2)$ |

Table 4. Classification accuracy of Secondary School user.

| User category       | PCA-Bayes | TNPE-Bayes | H-PCA-Bayes | H-TNPE-Bayes |
|---------------------|-----------|------------|-------------|--------------|
| Secondary school    | 32        | 89         | 71          | 99           |

Office, Secondary School, Stand-alone Retail, and Strip Mall, compared to the four classification models. H-TNPE-Bayes can effectively and accurately distinguish all 16 industry users. Besides, the five classification models have the most significant performance on the difference in the recognition ability of Secondary School users. Reflected by CAA indicator, H-TNPE-Bayes performed the best, while ELM performed the worst. H-PCA-Bayes and H-TNPE-Bayes can effectively improve the classification ability of PCA-Bayes and TNPE-Bayes by adopting the strategy of separating electrical behavior. By observing the Macro-F1 indicator, the performance of H-TNPE-Bayes and TNPE-Bayes is the best, indicating that compared with traditional linear and static multivariate statistical algorithms, manifold learning can extract more data information.

For a set of $m$ daily load curves $X$ with $n$ value records, the computational complexity of the five classification models is shown in Table 3. The computational complexity of ELM is $O(n^3mhn)$, where $h$ is the number of hidden neurons. In PCA, the computational complexity of the covariance matrix is $O(n^2m)$, and the eigenvalue decomposition is $O(n^3)$. Therefore, the complexity of PCA is $O(n^2m + n^3)$. The computational complexity of TNPE is mainly composed of three parts: the $k$-nearest neighbor is $O(m\log(m)n\log(k))$, the nearest neighbor reconstruction matrix $W$ is $O(mnk^3)$, and the low-dimensional space representation is $O(dm^2)$. Compared with PCA and TNPE, H-PCA, and H-TNPE increase the computational complexity of $k$-means. The computational complexity of $k$-means is $O(mnct)$, where $c$ is the number of clusters and $t$ is the number of iterations. ELM has the smallest computational complexity, but the worst classification performance. The computational complexity of PCA is less than that of TNPE, but the classification ability of TNPE is stronger than PCA. Compared with TNPE and PCA, H-TNPE, and H-PCA have higher computational complexity, but smaller samples of training required.

**Result analysis**

The method proposed in this paper based on feature detection, so whether the ability to extract the feature structure of the data effectively will affect the final classification result. Since the classification mechanism of ELM is different from the other model, this section mainly analyses the feature extraction capabilities of PCA, TNPE, H-PCA, and H-TNPE in load data. Table 4 shows the classification accuracy of Secondary School users, the four models have the most obvious differences in the performance of this type. Therefore, this paper selects Secondary School users to analyze the classification results.

As shown in Figure 9(a) and (b), the daily-load curves of Secondary school User 1 and User 2. Except for individual cases, most of them show daytime power consumption and the load characteristic is a single peak. In the summer electricity consumption, they both have a new load mode, which a higher electricity usage in the morning and evening. By observing the load heat map of the two users in Figure 9(c) and (d), the primary external performance of the difference in the electrical characteristics of User 1 and User 2 is the duration of different power consumption patterns in summer. Although the spatial characteristics of these two user loads are region consistent with the time-series changes. However, due to the influence of other factors such as region, climate, local policies, and so on, the characteristics of consumer electricity consumption in the same industry type have local dynamic changes.

In order to evaluate the performance of the multi-classifier specifically, the classification capabilities of the binary classifier should be discussed. Figure 10 shows the performance of the binary classifiers of PCA, TNPE, H-PCA, and H-TNPE, where the red line is the feature statistics of user 1, and the blue line is the feature statistics of user 2. If the blue line is below the red line, it means that User 2 and User 1 have similar electricity usage behaviors. Otherwise, the part of the blue line above the red line represents a difference in their electricity consumption behavior. Calculate the feature similarity of user 1 and user 2 according to equation (12) in section 3.2, the result is shown in Table 5. The higher value of the feature similarity, the better performance of the binary classifier. Figure 10(a) shows the feature detection result of PCA, there are differences in the summer consumption behaviors between user 1 and user 2. In fact, their summer electricity consumption behaviors have similar load curves, but the duration is different. This difference may be due to external factors such as geographic location, climate, weather, etc. Despite their electricity consumption behaviors in local differences, they still are secondary school users. Therefore, it is difficult for PCA to identify local dynamic changes in the consumption behaviors of the same industry. Figure 10(b) is the feature
detection result of TNPE. In the feature space detection, TNPE can identify the local dynamic changes of users’ electricity consumption behavior, but in the residual space detection, there are also problems similar to the PCA. Figure 10(c) and (d) are the feature detection results of H-PCA and H-TNPE respectively. By observing the statistics of feature space and residual space, H-PCA and H-TNPE are both better than PCA and TNPE. This means that the separation of electricity consumption behavior can improve the model’s ability to identify the industry to which electricity consumption belongs.

**Conclusions**

In view of the nonlinear and local characteristics in the load data, the single daily-load curves with power-hour cannot truly reflect the user’s dynamic consumption behaviors, and the multi daily-load curves with power-hour-day can reflect the truth. This paper developed a classification method through TNPE and Bayes, and it is composed of multiple binary classifiers based on feature detection. By separating the consumption behaviors, the relationship between behaviors can be simplified. Through hierarchical classification, it can effectively reduce the bad influence of external factors such as region and climate on the performance of the classifier. The results of the case study demonstrated that the model proposed in this paper could realize a refined classification of the electricity consumption industries using fewer training samples.

In order to simplify the processing of the proposed model, electricity consumption behaviors are divided into routine and non-routine types. However, the fact is that different users may have different types of electricity consumption behavior. Therefore, how to establish...
a variable multi-type electricity consumption behavior classification model to identify user industry types is the focus of this paper’s future work. In addition, the classification model proposed in this paper depends on the completeness and reliability of the data. Therefore, how to improve the robustness of the model is the focus in future work.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is supported by National Natural Science Foundation of China (No. 61763049); Science and Technology plan of Applied Basic Research Programs key Foundation of Yunnan province (No. 2018FA032).

ORCID iDs
Zi-Wen Gu https://orcid.org/0000-0003-1264-5730
Peng Li https://orcid.org/0000-0001-9432-8995

References
1. Al-Otaibi R, Jin N, Wilcox T, et al. Feature construction and calibration for clustering daily load curves from smart-meter data. IEEE Trans Industr Inform 2016; 12: 645–654.
2. Wang Y, Chen Q, Hong T, et al. Review of smart meter data analytics: applications, methodologies, and challenges. IEEE Trans Smart Grid 2019; 10: 3125–3148.
3. Borges FAS, Fernandes RAS, Silva IN, et al. Feature extraction and power quality disturbances classification using smart meters signals. IEEE Trans Industr Inform 2016; 12: 824–833.
4. Fang D-F, Su Z, Xu Q-C, et al. Multi-characteristics based data scheduling over the smart grid. Int J Autom Comput 2016; 13: 151–158.
5. Zhong S and Tam K-S. Hierarchical classification of load profiles based on their characteristic attributes in frequency domain. IEEE Trans Power Syst 2015; 30: 2434–2441.
6. Andersson M. Modeling electricity load curves with hidden Markov models for demand-side management status estimation. Int Trans Electr Energy Syst 2017; 27: e2265.
7. Raoofat M and Eghtedarpour N. A modified fuzzy clustering algorithm for market zonal partitioning in electricity markets. Int Trans Electr Energy Syst 2013; 23: 526–538.
8. Rhodes JD, Cole WJ, Upshaw CR, et al. Clustering analysis of residential electricity demand profiles. *Appl Energy* 2014; 135: 461–471.
9. Wang F, Li K, Dušić N, et al. Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns. *Energy Convers Manag* 2018; 171: 839–854.
10. Wang B and Mao Z. Detecting outliers in electric arc furnace under the condition of unlabeled, imbalanced, non-stationary and noisy data. *Meas Control* 2018; 51: 83–93.
11. Yang J, Ning C, Deb C, et al. K-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. *Energy Build* 2017; 146: 27–37.
12. Xiang Y, Hong J, Yang Z, et al. Slope-based shape cluster method for smart metering load profiles. *IEEE Trans Smart Grid* 2020; 11: 1809–1811.
13. Iglesias F and Kastner W. Analysis of similarity measures in times series clustering for the discovery of building energy patterns. *Energies* 2013; 6: 579–597.
14. Albu F, Mateescu A and Dumitriu N. Architecture Selection for a Multilayer Feedforward Network. In: *International Conference on Microelectronics and Computer Science*, Ottawa, ON, 3–6 June 1997, pp.131–134. New York: IEEE Communications Society.
15. Kirbas I and Kerem A. Short-term wind speed prediction based on artificial neural network models. *Meas Control* 2016; 49: 183–190.
16. Albu F and Martinez D. The application of support vector machines with Gaussian kernels for overcoming co-channel interference. In: *Proceedings of the 1999 IEEE signal processing society workshop: neural networks for signal processing IX*, Madison, WI, 25 August 1999, pp.49–57. New York: IEEE.
17. Dong S, Xu X, Liu J, et al. Rotating machine fault diagnosis based on locality preserving projection and back propagation neural network–support vector machine model. *Meas Control* 2015; 48: 211–216.
18. Albu F, Hagiescu D, Vladutu L, et al. Neural network approaches for children’s emotion recognition in intelligent learning applications. In: *Proceedings of the 7th international conference on education and new learning technologies*, Barcelona: IATED, 6–8 July 2015, pp.3229–3239.
19. Zhao L-J, Chai T-Y and Yuan D-C. Selective ensemble extreme learning machine modeling of effluent quality in wastewater treatment plants. *Int J Autom Comput* 2012; 9: 627–633.
20. Kim J-G and Lee B. Appliance classification by power signal analysis based on multi-feature combination multi-layer LSTM. *Energies* 2019; 12. DOI: 10.3390/en12142804.
21. Yang Y, Zhong J, Li W, et al. Semisupervised multilabel deep learning based noninvasive load monitoring in smart grids. *IEEE Trans Industr Inform* 2020; 16: 6892–6902.
22. Varga ED, Beretka SF, Noce C, et al. Robust real-time load profile encoding and classification framework for efficient power systems operation. *IEEE Trans Power Syst* 2015; 30: 1897–1904.
23. Balli S, Sağbış EA and Peker M. Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm. *Meas Control* 2018; 52: 37–45.
24. Ge Z, Xie L, Kruger U, et al. Local ICA for multivariate statistical fault diagnosis in systems with unknown signal and error distributions. *AIChE J* 2012; 58: 2357–2372.
25. Markazi AHD, Maadani M, Zabihišahr SH, et al. Adaptive fuzzy sliding mode control of under-actuated nonlinear systems. *Int J Autom Control* 2018; 13: 364–376.
26. Tan L, Li P, Miao A, et al. Online process monitoring and fault-detection approach based on adaptive neighborhood preserving embedding. *Meas Control* 2019; 52: 387–398.
27. Roweis ST and Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science* 2000; 290: 2323–2326.
28. Miao A, Ge Z, Song Z, et al. Time neighborhood preserving embedding model and its application for fault detection. *Ind Eng Chem Res* 2013; 52: 13717–13729.
29. He X, He Z, Song J, et al. NAIS: neural attentive item similarity model for recommendation. *IEEE Trans Knowl Data Eng* 2018; 30: 2354–2366.
30. Ahmadi H and Marti JR. Load decomposition at smart meters level using eigenloads approach. *IEEE Trans Power Syst* 2015; 30: 3425–3436.