Clustering Mining of Urban Traffic Flow Based on CVAE

Min-Tong Su, Jin Zheng, and Zu-Ping Zhang
Department of Computer Science and Technology, Central South University, Changsha, Hunan, 410083, P.R. China
Email: {sumintong, zhengjin, zpzhang}@csu.edu.cn

***Abstract*—Understanding the urban traffic flow at intersections is helpful to formulate traffic control strategies, so as to ease traffic pressure and improve people's living standards. There are many related researches on traffic flow, and similarity research is one of them. Different from the traditional way, this paper studies the traffic flow from the perspective of image similarity. The Convolutional Variational Auto-Encoder (CVAE) is introduced to extract the low-dimensional features of traffic flow during a day, and Affinity Propagation (AP) clustering algorithm is used to cluster the features without real labels. Combining the clustering results with geographic coordinates reveals the distribution pattern of traffic flow. The experimental data includes about 10 million vehicle records at 650 intersections in Changsha on a certain day. The clustering results show that the traffic flow at the intersection of Changsha City can be divided into three categories according to the time-variant trends, and the distribution of each category basically conforms to the daily traffic laws of the city. Furthermore, the effectiveness of the clustering process is further verified by clustering the open source temporal data of different lengths.**

***Index Terms*—traffic flow, convolutional variational auto-encoder, feature extraction, affinity propagation clustering algorithm**

I. Introduction

In recent years, the advancement of technology has promoted the rapid development of society and transportation. But various traffic problems also arose at the same time, such as frequent traffic accidents and traffic jams. The traffic flow reflects the comprehensive state of urban traffic and is also an indispensable scientific basis for traffic management. Many researchers have analyzed it from different aspects [1]-[3]. Traffic flow as a kind of time series data can be studied from the perspective of similarity analysis. That is to find the similarity between the traffic flow at different intersections, different time periods or different areas.

The clustering analysis in data mining is one of the effective methods for mining traffic flow information and laws. It is also one of the most basic tasks in machine learning and artificial intelligence. Clustering urban traffic flow according to the temporal changes can not only dig out the traffic laws for urban road planning and traffic guidance, but also use the clustering results to build better traffic flow prediction model or missing traffic data estimation model. The clustering study of time series has also been receiving attention in many other research fields such as engineering, science, finance, economics, and government [4].

Clustering is the process of grouping similar objects into one class. Time series similarity analysis has always been a hot topic at home and abroad. To date, there are several methods for similarity measure in time series. One such method is the discrete Fourier transform (DFT). The main principle of this method is to use the equal Euclidean distance in the time domain and the frequency domain to calculate the similarity of the sequences [5]. The disadvantage is that some local extreme values will be ignored, resulting in loss of information. Compared with DFT, another method is the discrete wavelet transform (DWT) [6]. DWT can retain local details and is a better lossy compression, but the essence of the processing problem is the same as DFT, and there is not much advantage. Another method is the dynamic time warping (DTW) algorithm, which is used to calculate the similarity between two time series by stretching or shrinking to match sequences [7]-[9]. And the clustering research based on DTW also made a lot of progress [10], [11].

However, the calculation of similarity measure takes a long time, which is not conducive to the improvement of clustering efficiency and the further development of time series clustering. To solve this problem, the data dimension reduction or compression of the time series is an effective way. There are also well-known methods, such as Principal Component Analysis (PCA). PCA projects high-dimensional data into low-dimensional space with minimal loss of information to achieve the purpose of data dimension reduction. The second method is Linear Discriminant Analysis (LDA), which is a supervised linear dimension reduction algorithm. It differs from PCA in maintaining data information. LDA is to make it easier to distinguish data points after dimension reduction. The third method is Locally Linear Embedding (LLE). It is a non-linear dimension reduction algorithm, which can keep the original manifold structure of the data after dimension reduction. Although the above various methods have been widely used in similarity analysis, they are limited to specific application scenarios.

Manuscript received August 2, 2020; revised December 16, 2020.

doi: 10.18178/jtle.8.2.34-44

Therefore, time series similarity measurement and data dimension reduction are the two main problems to be solved in time series clustering. On the one hand, we need to ensure the comprehensiveness and accuracy of feature extraction. On the other hand, the data dimension should be as small as possible without affecting the main characteristics of the time series.

In this paper, we focus on cluster analysis of traffic flow at different intersections without real category labels. First, the one-dimensional traffic flow is converted into a grayscale image. Then, the model CVAE is used to mining the hidden features of traffic flow images. While extracting the main features of the traffic flow image, it compresses the traffic flow into a two-dimensional space. The advantage of CVAE is that it reduces the time cost of research on similarity measurement and obtains a good dimension reduction effect. Later, the low-dimensional features of traffic flow are clustered by AP algorithm without specifying the number of clusters.

The main contributions of this paper can be summarized as follows:

(1) Introduce CVAE to process the time series of traffic flow, extract the changing characteristics of temporal data from the perspective of images, and reduce the high-dimensional temporal data to two-dimensional space.

(2) The fully connected layers in CVAE are replaced with the convolutional layers. It reduces the spatial structure loss caused by the fully connected layer stretching the upper feature map to one-dimension.

(3) In the absence of real labels, the AP algorithm is used to cluster the low-dimensional feature vectors of the traffic flow, and then the clustering results are combined with the spatial information of the road intersections for visual analysis. Finally, a clustering distribution pattern with spatial-temporal characteristics was obtained, which revealed the daily traffic laws of the city.

This paper is organized as follows: Section 2 reviews related works. Section 3 describes the CVAE model and the AP algorithm in detail. The clustering experiment of traffic flow is shown in Section 4. Another verification experiment for the clustering process is shown in Section 5. Finally, Section 6 concludes this paper and proposes some suggestion for future work.

II. Related Work

At present, the research on traffic flow is mainly divided into two aspects. The first is the prediction of traffic flow, and the other is the similarity measure of traffic flow. Many of them are about traffic flow clustering.

Hu *et al.* [12] introduced both DTW and shortest path analysis methods to measure the similarity between traffic flow objects with spatial-temporal characteristics, and proposed a kind of clustering analysis method for road network traffic flow data, which aggregates traffic flow data objects with similar properties and space correlation into one class. They found the spatial distribution pattern of road traffic flow. Jiang *et al.* [13] discusses how to cluster traffic time series that have similar fluctuation pattern, using the simple average detrending method to make the time series smoother. Then they use principal component analysis (PCA) on raw data, and take the weight of the first d-components as the feature of the time series. Finally, they use the k-means algorithm to cluster the traffic time series. Yang *et al.* [14] uses non-negative tensor decomposition technique to extract the features of traffic flow in the within-one-day time, and then spectral clustering is performed in low-dimensional feature space. Song *et al.* [15] calculated the traffic flow of each road over a period of time based on the floating trajectory data set. They constructed a weighted directed graph of the road network. The adjacency, connectivity, and congestion between road segments and the congestion degree of road segments are made as attributes of the cluster similarity measurement.

In addition, some researches on traffic flow are based on clustering methods. In the paper [16], a multi-index fusion clustering strategy is proposed to improve the identification accuracy of traffic flow states. The model was solved by the method of Lagrange multipliers, producing the optimal weight combination. Then, the optimal weights were introduced to the fuzzy c-means (FCM) clustering, forming the multi-index fusion clustering method. Akın *et al.* [17] proposed a traffic flow forecasting approach based on OPTICS algorithm . OPTICS is a density-based clustering algorithm. Rao et al. [18] proposed an interval data-based K-means clustering method for traffic state identification to investigate the impact of traffic flow uncertainty on intersection.

Outside the field of transportation, there are some researches on clustering of time series. Ruta *et al.* [19] used a dimensionality reduction technique, Symbolic Aggregate Approximation (SAX). With SAX, the time series data clusters efficiently and is quicker to query at scale. Li *et al.* [20] studied the bike usage clustering. Their results showed that DWT could effectively reduce dimensionality, filter out random errors and reveal the main characteristics of the raw time series. The clustering approach offers the ability to differentiate and discover bike usage patterns across different stations. Huang *et al.* [21] found the water quality trend patterns through using time series clustering. They used statistical methods to clean the data and extract the time-variant trends. Martins *et al.* [22] used clustering techniques to analyze time series of day-ahead electricity prices from European markets between 2015 and 2018 in order to identify different price patterns.

III. The Proposed Method

At present, there are many feature representation methods for time series, but the method using unsupervised learning is less applied. We have tried to reduce the dimension of the raw temporal data using ordinary auto-encoder, but the effect of dimension reduction is not ideal. On the one hand, the size of dimension reduction needs to research based on the difference in time series length or trend. On the other hand, low-dimensional feature vectors are less effective

for clustering. The solution to the above problem is Variational Auto-Encoder.

A. Variational Auto-Encoder

Variational Auto-Encoder (VAE) is an important type of generative model. It was proposed by Kingma and Welling [23]. The basic goal of VAE is to build a model that generates target data X from hidden variable Z. Specifically, giving a real sample X, we assume that there is a posterior distribution $p(Z|X)$, which is (independent, multivariate, standard) normal distribution. The normal distribution has two sets of parameters, the mean μ and variance σ^2 that can be fitted by a neural network.

$$p(Z)=\sum_X p(Z|X)p(X)=\sum_X N(0,1)p(X)=N(0,1)\sum_X p(X)=N(0,1)$$

$$\mu=f_1(X),\ \log\sigma^2=f_2(X)$$

Variance σ^2 is always non-negative and needs to be added to the activation function so that select to fitting $\log\sigma^2$. There is no need to add the activation function for $\log\sigma^2$, because it can be positive or negative. After getting the mean and variance of the X, we also get its normal distribution. Then sample a variable Z from this exclusive distribution, and get the $X=g(Z)$ through a generator. The model reconstruct X which is to minimize the distance between input and output, but this process is affected by noise. An additional loss that is KL divergence needs to be added to the reconstruction error, because variable Z is re-sampled and not directly calculated by the encoder. Generally KL divergence is used to measure the difference between two probability distributions $p(x)$ and $q(x)$, which defined as follow equation:

$$KL\left((p(x)\|q(x)\right)=\int p(x)\ln\frac{p(x)}{q(x)}dx=E_{X\sim P(x)}\left[\ln\frac{p(x)}{q(x)}\right]$$

The main property of KL divergence is non-negative. If fixed $p(x)$, there is formula below:

$$KL\left(p(x)\|q(x)\right)=0\Leftrightarrow p(x)=q(x)$$

If fixed $q(x)$, the above formula also holds. The result of minimizing the KL divergence is that both are as equal as possible. The situation of deriving a one-variable normal distribution is as follows:

$$\begin{aligned}
&KL\left(N(\mu,\sigma^2)\|N(0,1)\right)\\
&=\int\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}\left(\log\frac{e^{-(x-\mu)^2/2\sigma^2}/\sqrt{2\pi\sigma^2}}{e^{-x^2/2}/\sqrt{2\pi}}\right)dx\\
&=\int\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}\log\left\{\frac{1}{\sqrt{\sigma^2}}\exp\left\{\frac{1}{2}\left[x^2-(x-\mu)^2/\sigma^2\right]\right\}\right\}dx\\
&=\frac{1}{2}\int\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}\left[-\log\sigma^2+x^2-(x-\mu^2/\sigma^2)\right]dx
\end{aligned}$$

The final simplification result of the above formula is as below:

$$KL\left(N(\mu,\sigma^2)\|N(0,1)\right)=\frac{1}{2}\left(-\log\sigma^2+\mu^2+\sigma^2-1\right)$$

In addition, there is a heavy parameter trick in the VAE model. The model need to sample a variable Z from the normal distributed $p(Z|X)$, while the mean and variance are calculated by the model. This process is needed to optimize the model, but the "sampling" operation is non-differentiable and the sampling result is differentiable. Sampling a noise ε from $N(\mu,\sigma^2)$ so that $Z=\mu+\varepsilon\times\sigma$. In this way, the operation of "sampling" does not need to participate in gradient descent, and the result of sampling is involved instead, making the entire model trainable.

In general, the purpose of VAE is to learn the distribution transformation between the input and output, so that each data sample has its own mean and variance.

B. Affinity Propagation Clustering

Traffic flow data does not have any prior knowledge, so a clustering algorithm is needed without specifying the number of clusters. This paper chooses the clustering algorithm based on the actual application scenario and the distribution of low-dimensional features. It is called Affinity Propagation (AP) algorithm proposed by Frey [24].

The AP clustering algorithm is a clustering algorithm based on "information transfer" between data points. Unlike the k-means algorithm or the k-center point algorithm, the AP algorithm does not need to determine the number of clusters before running the algorithm. The "exemplars" that the AP algorithm looks for all actually exist in the data set, as the representative of each class.

Assumed data sample set $\{x_1,x_2,\cdots x_n\}$, there is no internal structure between the data. Characterizing the similarity between points by a matrix s. For example, $s(i,j)>s(i,k)$ if and only if x_i versus x_j are more similar than another.

The AP algorithm performs the steps of alternating two message passes to update the two matrices:

① Responsibility information matrix $R: r(i,k)$

It describes the degree which object k is suitable as the cluster center of object i, and represents the message from i to k.

② Availability information matrix $A: a(i,k)$

It describes the suitability which object i chooses object k as the cluster center, and represents the message from k to i.

Two matrices R, A are initialized to zero, which can be regarded as Log-probability table. This algorithm iterates through the following steps:

① Responsibility information $r_{t+1}(i,k)$ iterate according to equation (1):

$$r_{t+1}(i,k)=s(i,k)-\max_{k'\neq k}\left\{a_t(i,k')+s(i,k')\right\} \quad (1)$$

②Availability information $a_{t+1}(i,k)$ iterate according to equations (2) and (3):

$$a_{t+1}(i,k) = \min\left(0, r_t(k,k) + \sum_{i' \notin \{i,k\}} \max\{0, r_t(i',k)\}\right) \quad (2)$$

$$a_{t+1}(k,k) = \sum_{i' \neq k} \max\{0, r_t(i',k)\} \quad (3)$$

When these decisions remain unchanged after several iterations, more than the set number of iterations or a decision on sample point within a small area remains unchanged after several iterations, the algorithm ends.

In order to avoid oscillation, an attenuation coefficient λ is introduced when the AP algorithm updates information. Each piece of information is set to λ times its updated value of the previous iteration plus $1-\lambda$ times its updates value. The attenuation coefficient λ is a real number between 0 and 1.So the number of $t+1$ times iteration is obtained form (4) and (5):

$$r_{t+1}(i,k) \leftarrow (1-\lambda) r_{t+1}(i,k) + \lambda r_t(i,k) \quad (4)$$

$$a_{t+1}(i,k) \leftarrow (1-\lambda) a_{t+1}(i,k) + \lambda a_t(i,k) \quad (5)$$

The AP clustering algorithm also needs to determine the similarity measurement method between data. This paper uses Euclidean distance, which measures the absolute distance between various points in a multidimensional space. The formula (6) is as follows:

$$dist_{ed}(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^2\right)^{\frac{1}{2}} \quad (6)$$

IV. Clustering Experiment of Urban Traffic Flow

The experiment is based on the traffic data in the intelligent transportation platform of Changsha City, and performs cluster mining when the label of traffic flow is unknown. The overall design of the method is as follows: (1) Extracting the vehicle records from the platform. (2) Preprocessing the vehicle records to obtain the traffic flow. (3) Transforming the sequential data into a grayscale image. (4) Building a CVAE model and applying it to traffic flow images. (5) Using AP algorithm for clustering and visual analysis with spatial information of urban intersection. (6) The verification experiment of clustering process by using open source temp oral data. The overall design flow of the experiment is shown in Fig. 1.

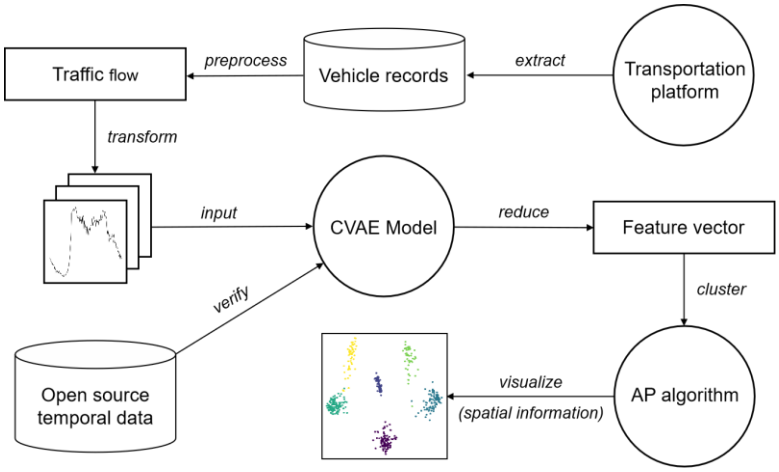


Figure 1. Flow chart of traffic temporal data clustering

A. Data Preprocessing

Through the Changsha Intelligent Transportation Platform, about 10 million vehicle records of about 650 intersections in the within-one-day time are extracted. These data include multiple labels such as the name of the intersection, the number of the intersection, the time of passing, the license plate number, and the vehicle model, but the experiment will not use each labels, so the passing record data needs to be filtered. In addition, due to the failure of the camera at the intersection or the bad weather, the data itself has problems such as missing and wrong. Therefore, the data must be cleaned first. For example, some data lack vehicle transit time, or the license plate number cannot be identified, which is not conducive to calculating traffic flow. So we remove this kind of data. Then the four labels of passing time, intersection number, license plate number and intersection name are retained in the data, as shown in Fig. 2 below.

Passing time	Crossing number	License number	Road name
201 16:08	730451300	Hu C61Z2	The intersection of Road
201 15:43	530011007	Hu G3U46	The intersection of Road and Wanbao Avenue
201 23:45	630011328	Hu A9551	The intersection of Road
201 8:29	730451304	Hu 1TK56	The intersection of Ring Road
201 6:51	530011007	Hu 6X28Z	The intersection of Road and Wanbao Avenue
201 18:28	730451304	Hu B555L	The intersection of Ring Road
201 20:28	730451304	Hu FG881	The intersection of Ring Road
201 20:14	530011007	Hu 671BW	The intersection of Road and Wanbao Avenue
201 15:15	730451304	Hu L4A05	The intersection of Ring Road
201 12:40	530011007	Hu 9Y09V	The intersection of Road and Wanbao Avenue
201 20:09	730451300	Hu K425P	The intersection of Road

Figure 2. Processed vehicle record data

Finally, we have divided 24 hours a day by 10 minutes to get 144 (24 hours * 6 time slices) time slices in the database. According to the intersection number, we counted the number of passing vehicles at each intersection in the time slice. For example, for intersection A, the number of passing vehicles between

time 0:00 to 0:10 is 50, and the number of passing vehicles between 0:10 to 0:20 is 49. And so on, temporal data of traffic flow at intersection A with a length of 144 is obtained. The main operations include: First, grouping by the intersection number and sorting the data in ascending order by the vehicle passing time. Second, counting the number of passing vehicles at all intersections every 10 minutes starting from 0 o'clock. Later, the sequences of traffic flow less than 144 are zero-filled. Two examples of traffic flow are shown in Fig. 3 below. Finally, the traffic flow data of about 650 road intersections in the city are set to different coordinate scales according to the corresponding flow size, and all temporal data are converted into 784-dimensions (28*28) grayscale image. Two examples of grayscale image data are shown in Fig. 4 below:

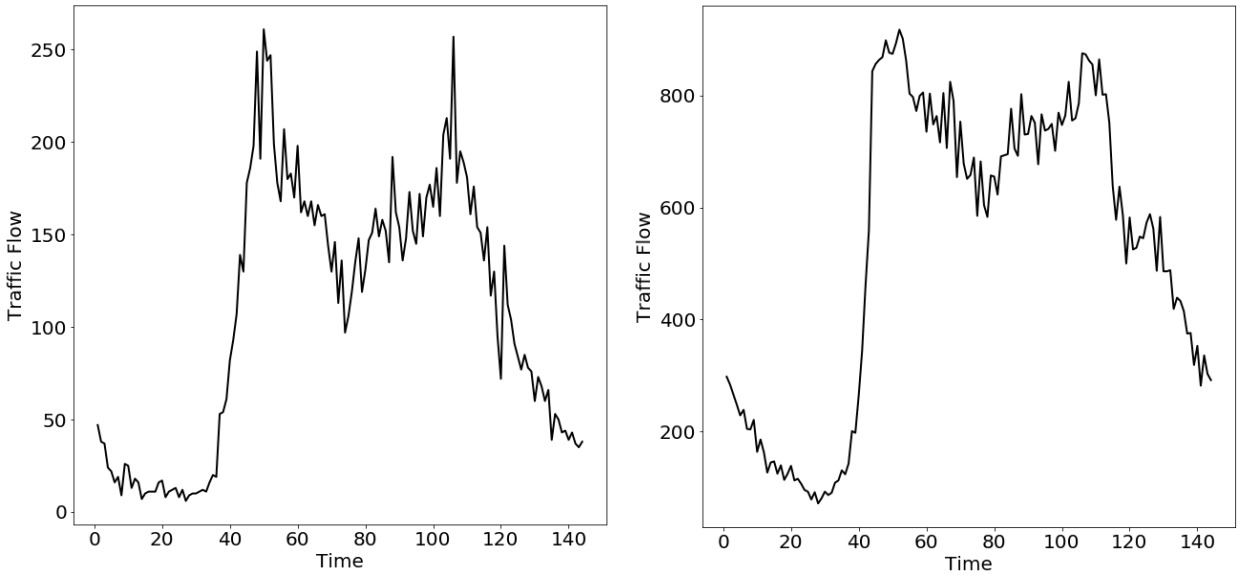


Figure 3. Examples of traffic flow image

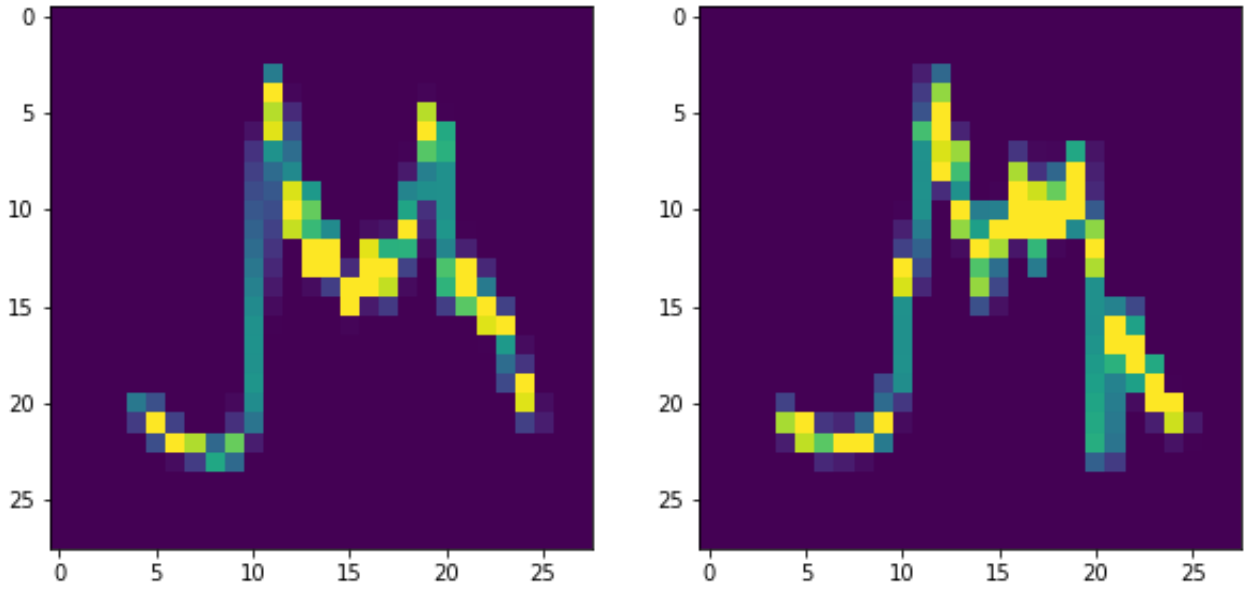


Figure 4. Example of grayscale image data of traffic flow

B. Implementation of Convolution Variational Auto-Encoder

The deep learning framework Keras was used to build a basic VAE. On this basis, modify and add new network layers including: three convolutional layers and three deconvolutional layers. The convolution kernel size of each layer is 3*3, and the step of the input layer and output layer is 1 when the other layers are 2. In addition, the number of convolution kernel is (32, 64, 64), and the batch normalization layer is added after the convolutional layer so that the input data of each layer of the neural network during the training process maintains the same distribution. The activation function uses "Relu". Then, the full connection layer of 256 dimensions is used to connect the convolutional layer after flattened. The CVAE model structure is shown in Fig. 5 below:

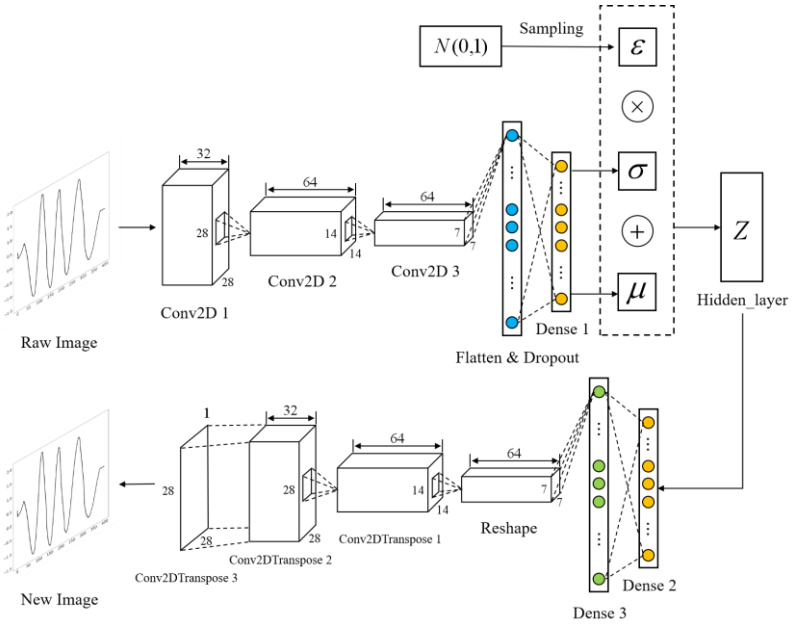


Figure 5. The CVAE model structure

In order to reduce the loss of spatial structure information when the feature image is flattened, the fully connected structure in the encoder is changed into the fully convolutional structure. As shown in the Fig. 6 below. The size of the convolution kernel of layer which name is "Conv2D_4" needs to be the same as the size of the feature image of the previous layer. Specifically, the size of the convolution kernel is 7*7, and the number of convolution kernels is set to 256, so the data structure is changed from (7, 7, 64) to (1, 1, 256).

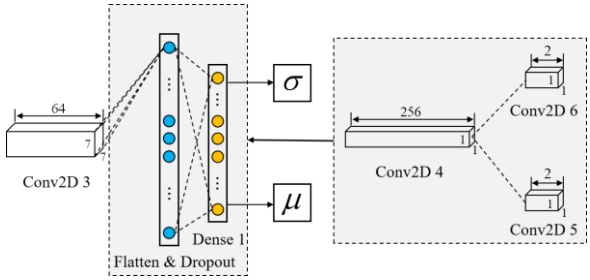


Figure 6. The replacement of the full connection layer

The model is trained using multiple loss functions, including cross entropy, relative entropy, and mean squared error. The relative entropy, also known as the KL divergence, is the basis functions of the VAE. The cross-entropy and mean square error are used to measure the similarity of the learning results from the perspective of probability and value. Finally, three functions are combined by weighted summation. The selection of weights is based on the distribution of low-dimensional features after experimental process. The initial weights are set to (0.01:1:1) corresponding to cross entropy, mean square error, and relative entropy respectively.

C. Experimental Environment

The experiment uses a server based on the Ubuntu 16.04.5 operating system, with a memory of 256G, and 72 Intel (R) Xeon (R) Gold 6140 CPU that clocked at 2.30GHz. The experimental programming language is Python and the Keras framework is used to build deep learning model. The clustering and evaluation methods are come from the machine learning library "Scikit-learn".

D. Clustering Results and Analysis

The traffic flow data set converted into a grayscale image is divided into a training set and a test set according to a ratio of 9:1, and training in CVAE model. The initial learning rate is set to 0.001 and the training epochs are 200. Finally, the trained CVAE model is split into two parts, an encoder and a decoder. And the low-dimensional feature vector distribution of the training set is obtained using the encoder, as shown in Fig. 7 below:

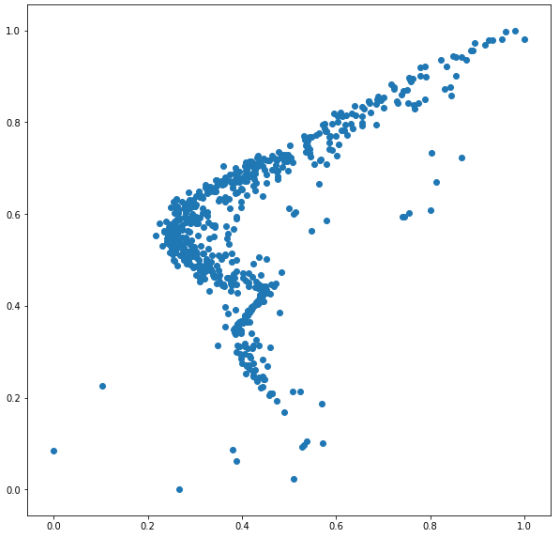


Figure 7. Two-dimensional feature vector distribution of traffic flow

As a generative model, the VAE can capture the structural changes of the image (tilt angle, circle position, shape change, or expression change, etc.), which is also a benefit of it. It has an explicit distribution and can be

easily visualize the distribution of the image. Therefore, the data manifold distribution graph of the traffic flow generated by the decoder is shown in Fig. 8 below:

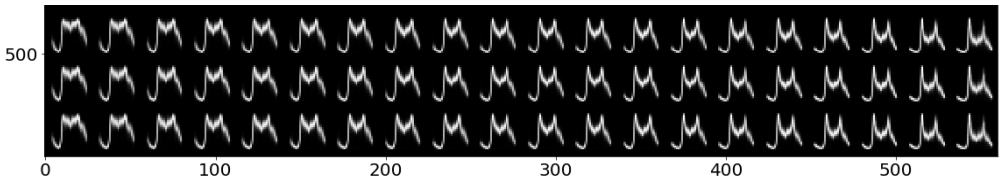


Figure 8. Manifold distribution of traffic flow data

From the data manifold graph, you can observe the change process of the traffic flow. Combined with the low-dimensional feature vector distribution, the number of target clusters is set to 3 types. Then the AP algorithm in the tool library scikit-learn is used for clustering. After experiments, the key parameters were adjusted as follows: the damping coefficient is 0.8, the preference degree is -3.5, and the similarity calculation method (or affinity) is "euclidean" which is the Euclidean distance. Finally, the experiment reached the result of target number, as shown in Fig. 9 below:

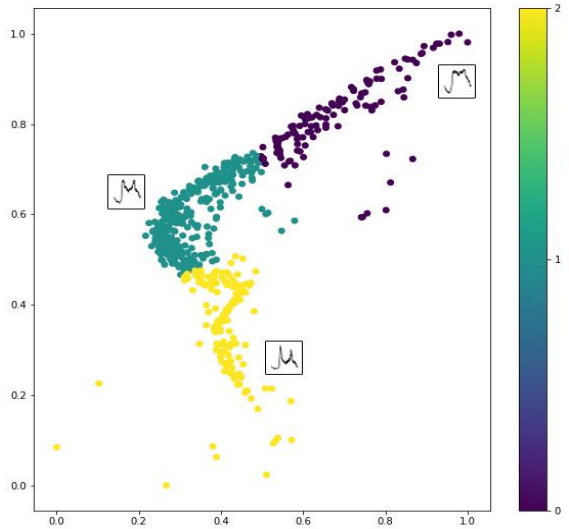


Figure 9. Traffic flow clustering results

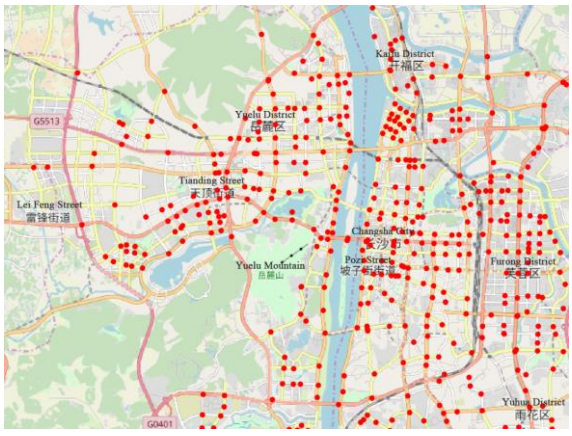


Figure 10. Geographical coordinates of road intersections

In order to further verify the clustering results, combined with the geographical coordinates of road intersections in Changsha City (as shown in Fig. 10 below), a traffic flow clustering distribution map containing spatial-temporal features was generated, as shown in Fig. 11 below:

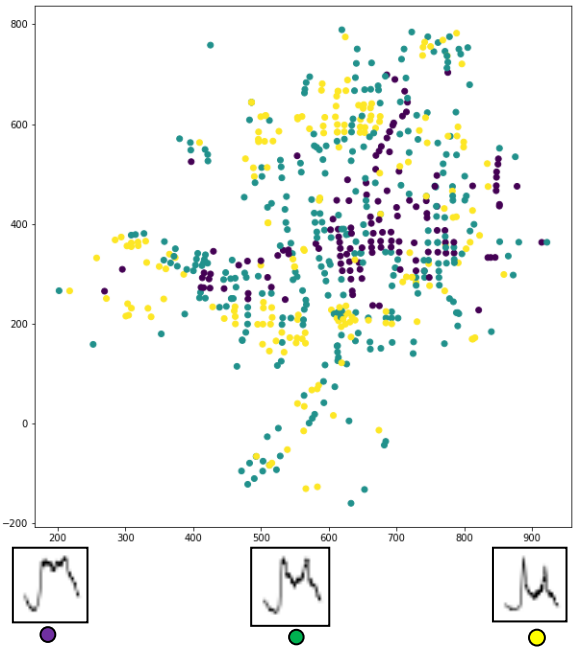


Figure 11. Clustering results of spatial-temporal characteristics of traffic flow

According to Fig. 10 and Fig. 11, the traffic flow clustering analysis is as follows:

(1) Intersections marked as "purple" are mainly concentrated in downtown sections and other main roads. There is a continuous wave of peaks. Except for the morning and evening peaks, the traffic volume decreases slightly in the afternoon.

(2) Intersections marked as "yellow" are mainly concentrated in the outer periphery of the urban area. There are two peaks and one trough. The traffic volume during the afternoon is the lowest compared to the other two categories.

(3) Intersections marked as "green" as the intermediate state of the other two categories. It also spatially distributed between the others. There are morning and evening peaks periods, and the traffic volume in the afternoon is significantly reduced compared to the morning and evening periods, but still higher than category "yellow".

In summary, the traffic flow of road intersections in Changsha can be divided into three types according to the changing pattern of temporal data, and the traffic flow

between each intersection has the characteristics of transition and gradual change. The morning and evening peaks occur of all intersections in the city within a day, but the reduction of traffic volume in the afternoon decreases from the city center and other main roads to the edge of the urban area. Therefore, the classification of categories is mainly based on the fluctuation of noon traffic, and its distribution basically conforms to the daily traffic laws of Changsha.

V. Validation Experiments of Clustering Methods

In the traffic flow clustering experiment, there is no category label as the experimental support data. In order to verify the effectiveness of the above clustering method, an open-source time-series data set with more data length is used for verification analysis.

The data set is selected from the University of California Riverside which name is "The UCR Time Series Classification Archive"[25]. This data set (http://www.cs.ucr.edu/~eamonn/time_series_data_2018/) was launched in 2002 and has become an important resource in a temporal data mining community. More than about one thousand published papers used at least one data set in the archive, which is called "Imagnet" in the time series field. This paper selects temporal data set of different lengths, as shown in Table I below:

TABLE I. Open Source Temporal Data Set

Data set name	Training set	Test set	Length	Category
Synthetic Control	300	300	**60**	6
CBF	30	900	**128**	3
Symbols	25	995	**398**	6
MALLAT	55	2345	**1024**	8
CinC_ECG_torso	40	1380	**1639**	4

A. Cluster Evaluation Index

The evaluation of clustering results is generally divided into internal indicators and external indicators. The external evaluation standard is based on classified labels, and the clustering results are evaluated according to the labels. The internal indicators are based on distance analysis between categories.

For labeled open source data set, this paper chooses the FMI index (Fowlkes and Mallows Index) [26], which is a geometric mean of the recall and precision obtained between the training and validation data. As an external index for the evaluation of clustering results, the range of the performance metrics is between zero and one. The larger the value, the better the performance of the clustering. Its formula is as follow:

$$FMI = \sqrt{\frac{TP}{TP+FP} \cdot \frac{TP}{TP+FN}}$$

B. Experimental Results and Analysis

This experiment is also the verification of the above clustering process. Cluster analysis is performed under the same experimental environment and data processing. The FMI index was used to evaluate the clustering results, as shown in Table II below:

TABLE II. Clustering Results of Open Source Temporal Data

Data set name	Length	Real category	Cluster category	FMI
Synthetic Control	60	6	6	**0.9669**
CBF	128	3	3	**0.9198**
Symbols	398	6	5	**0.7079**
MALLAT	1024	8	8	**0.8507**
CinC_ECG_torso	1639	4	5	**0.6576**

The cluster results of open source temporal data are analyzed as follow:

First, from the perspective of the FMI index, the highest reached about 0.97 and the lowest was about 0.66. For time series data of different lengths, the clustering method proposed in this paper is effective.

Second, for the data set "Symbols", the distribution of low-dimensional features and the generated data manifold distribution map are visually analyzed, as shown in Fig. 12 below. We found that the distribution of over-similar time series data in the two-dimensional space and the data manifold is closer, which is not conducive to subsequent clustering. Over-similar refers to the small difference in the image structure of the time series.

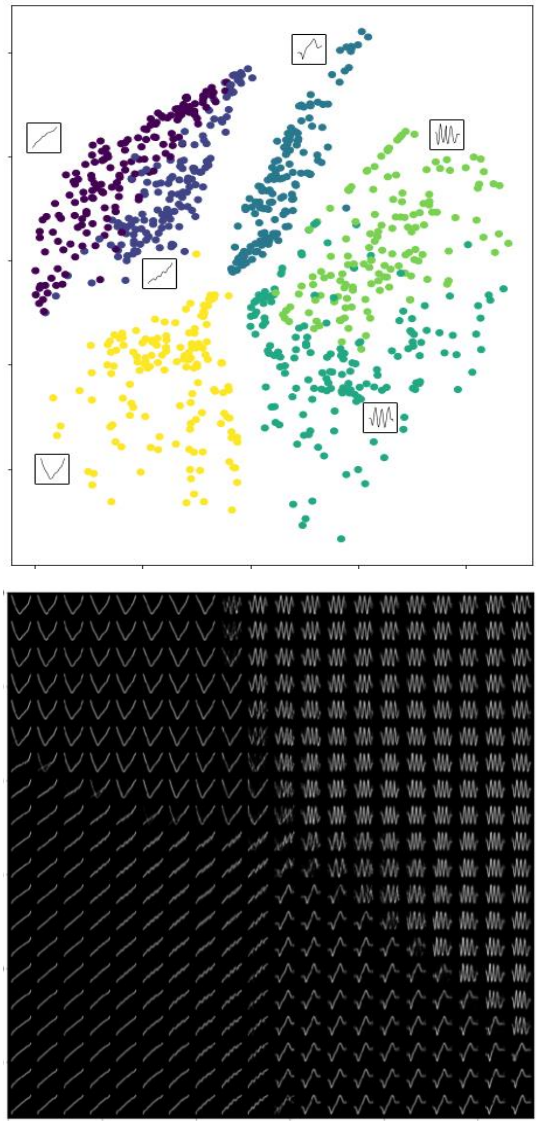
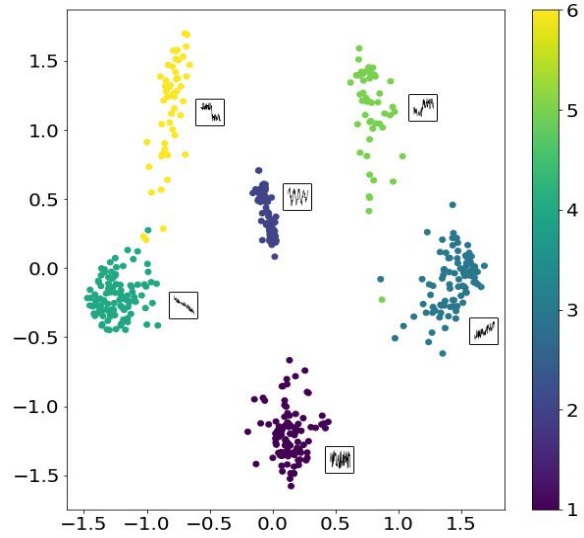


Figure 12. Data set symbols

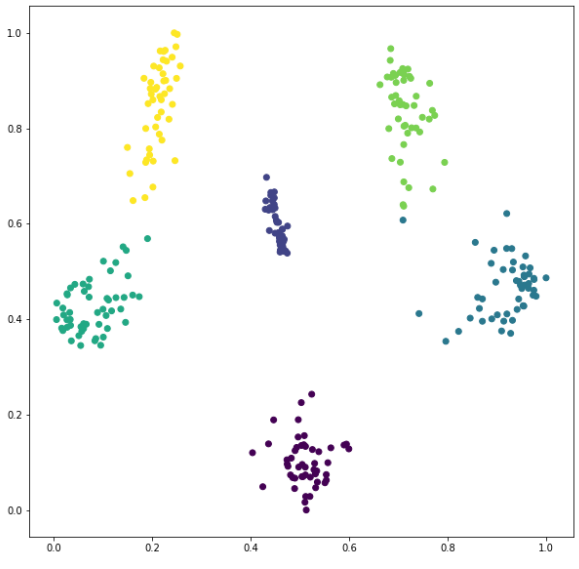
Third, for the data set "CinC_ECG_torso" with the longest length, its clustering accuracy is the lowest. Combined with the second point above, it can be inferred that the input image is not accurate enough. In the following research, we will try to increase the size of the input image to solve this problem.

In general, the clustering method in this paper is effective for temporal data. Data sets with obvious differences between each category of time image can achieve good clustering effect. However, for time images with small category differences, the clustering effect is poor. Of course, the size of the input image also has a great impact on the feature extraction, thus affecting the subsequent clustering.

The cluster visualization results are shown in Fig. 13 to Fig. 17. Among of them, the label (a) is the data distribution with real labels, and the label (b) is the result of AP clustering, and the data used for visualization only includes the train set of the original data.

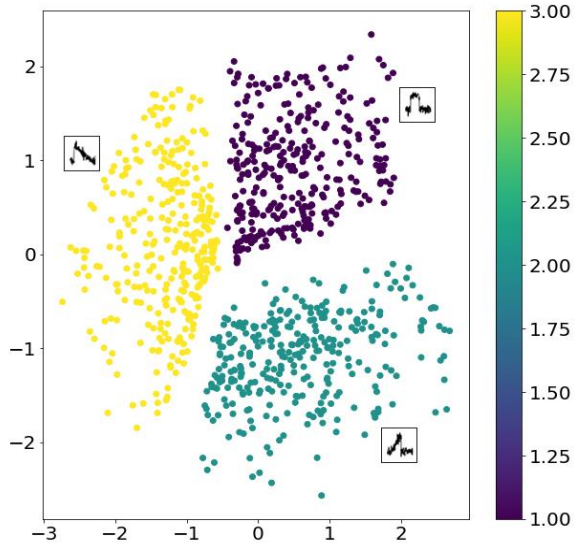


(a)

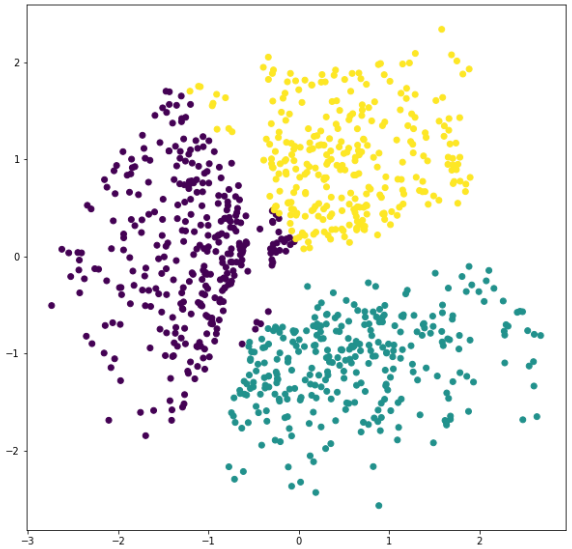


(b)

Figure 13. Data set synthetic control

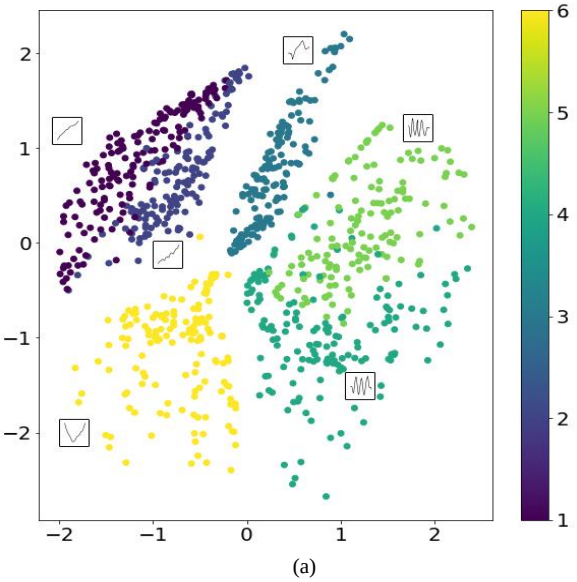


(a)

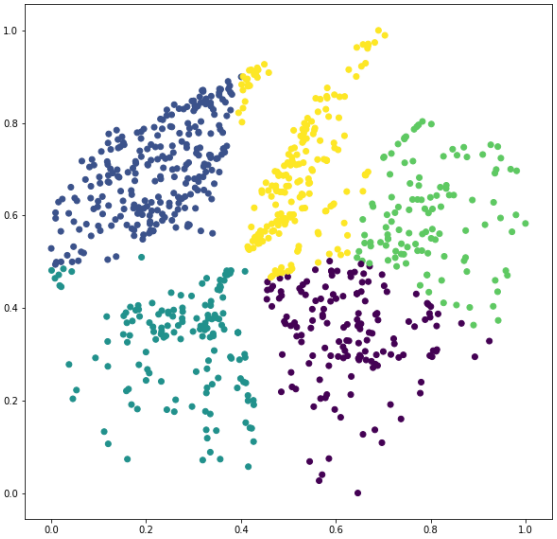


(b)

Figure 14. Data set CBF

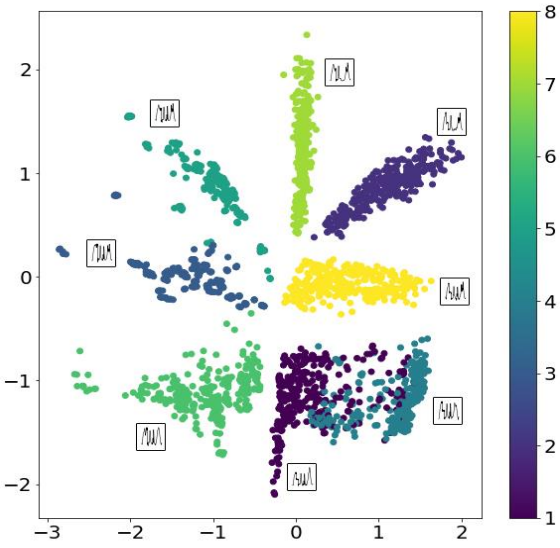


(a)

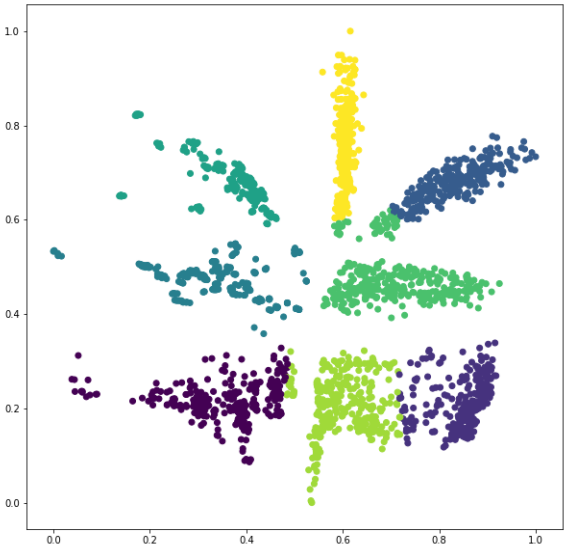


(b)

Figure 15. Data set Symbols

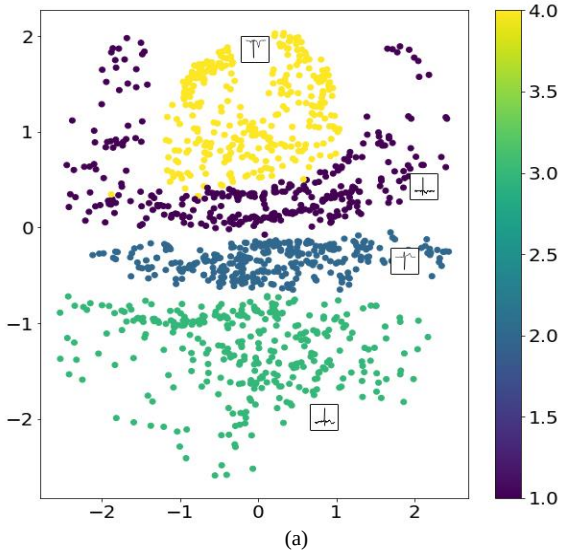


(a)

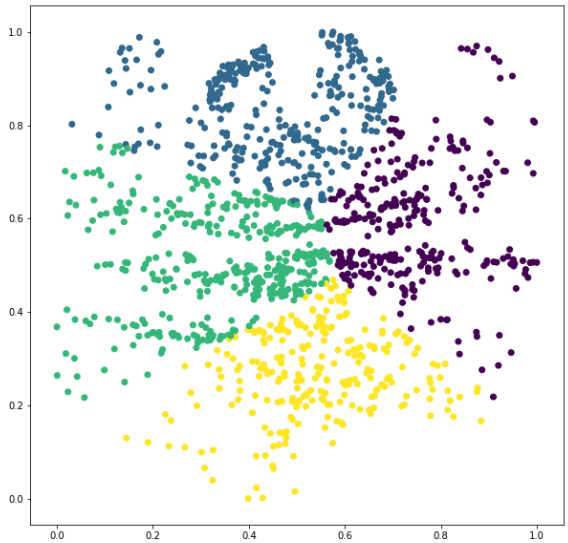


(b)

Figure 16. Data set MALLAT



(a)



(b)

Figure 17. Data set CinC_ECG_torso

VI. Conclusion and Outlook

Research on urban traffic flow is an important direction for the development of urban traffic. In this paper, the CVAE model is introduced to extract the low-dimensional features of traffic flow images, and then the high-dimensional time series image data is reduced to the low-dimensional feature space. Finally, AP algorithm is used to cluster the feature vectors to obtain the potential category relationship of traffic flow at intersections. The visual analysis of the clustering results combined with the spatial information of the intersection reveals the daily distribution pattern of traffic flow at the intersection in Changsha. The clustering process in this paper is verified in the open source temporal data set of the University of California Riverside, and experiments show that the clustering process is effective in dealing with different temporal data lengths and types.

The future research work focuses on two aspects:

(1) Considering the effect of input image size, such as trying to increase the image size or change the image into

a rectangle to obtain more image feature information of time series.

(2) New feature extraction methods will be explored, such as constructing new network structures by combining with the Long Short-Term Memory (LSTM) networks to obtain one-dimensional time series feature information.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Min-Tong Su conducted the research and wrote the paper; Jin Zheng and Zu-Ping Zhang provided the traffic data and assisted in its analysis; all authors had approved the final version.

Acknowledgement

This research is supported by The Hunan Key Laboratory for Internet of Things in Electricity No. 2019TP1016.

References

[1] S. Sun, C. Zhang, and G. Yu, "A bayesian network approach to traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 124-132, 2006.

[2] L. Qu, L. Li, Y. Zhang, and J. Hu, "PPCA-Based missing data imputation for traffic flow volume: A systematical approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 3, pp. 512-522, 2009.

[3] I. V. Eleni, G. K. Matthew, and C. G. John, "Short-term traffic forecasting: Where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, no. 1, pp. 3-19, 2014.

[4] T. W. Liao, "Clustering of time series data-a survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857-1874, 2005.

[5] C. Cerovecki and S. Hörmann, "On the CLT for discrete Fourier transforms of functional time series," *Journal of Multivariate Analysis*, vol. 154, pp. 282-295, 2017.

[6] X. Dai, L. Z. Cheng, J. C. Mareschal, D. Lemire, and C. Liu, "New method for denoising borehole transient electromagnetic data with discrete wavelet transform," *Journal of Applied Geophysics*, vol. 168, pp. 41-48, 2019.

[7] F. Petitjean, G. Forestier, G. I. Webb, A. E. Nicholson, Y. Chen, and E. Keogh, "Dynamic time warping averaging of time series allows faster and more accurate classification," in *Proc. IEEE International Conference on Data Mining*, 2014.

[8] H. Li, "On-line and dynamic time warping for time series data mining," *International Journal of Machine Learning and Cybernetics*, vol. 6, no. 1, pp. 145-153, 2015.

[9] P. E. Tsinaslanidis, "Subsequence dynamic time warping for charting: Bullish and bearish class predictions for NYSE stocks," *Expert Systems with Applications*, vol. 94, pp. 193-204, 2018.

[10] J. Paparrizos and L. Gravano, "Fast and accurate time-series clustering," *ACM Transactions on Database Systems*, vol. 42, no. 2, pp. 1-49, 2017.

[11] Z. G. Ives, "Technical perspective: K-Shape: Efficient and accurate clustering of time series," *ACM SIGMOD Record*, 2016.

[12] C. Hu and X. Yan, "Mining traffic flow data based on fuzzy clustering method," in *Proc. IEEE the Fourth International Workshop on Advanced Computational Intelligence*, 2011.

[13] S. Jiang, S. Wang, Z. Li, W. Guo, and X. Pei, "Fluctuation similarity modeling for traffic flow time series: A clustering approach," in *Proc. IEEE International Conference on Intelligent Transportation Systems*, 2015.

[14] S. Yang, J. Wu, Y. Xu, and T. Yang, "Revealing heterogeneous spatiotemporal traffic flow patterns of urban road network via tensor decomposition-based clustering approach," *Physica A: Statistical Mechanics and Its Applications*, vol. 526, p. 120688, 2019.

[15] Q. Song, J. Hu, R. Zhang, and Z. Zhang, "An urban topological map generation method for traffic flow prediction based on road segment clustering with floating vehicle trajectory dataset," in *Proc. 19th COTA International Conference of Transportation Professionals*, 2019.

[16] D. Bao, "A multi-index fusion clustering strategy for traffic flow state identification," *IEEE Access*, vol. 7, pp. 166404-166409, 2019.

[17] M. Akın, Ş. Sağıroğlu, and A. Değirmenci, "Traffic flow forecasting model with density based clustering algorithm," in *Proc. International Informatics and Software Engineering Conference*, 2019.

[18] W. Rao, J. Xia, W. Lyu, and Z. Lu, "Interval data-based k-means clustering method for traffic state identification at urban intersections," *IET Intelligent Transport Systems*, vol. 13, no. 7, pp. 1106-1115, 2019.

[19] N. Ruta, N. Sawada, K. McKeough, M. Behrisch, and J. Beyer, "SAX navigator: Time series exploration through hierarchical clustering," in *Proc. IEEE Visualization Conference*, 2019.

[20] D. Li, Y. Zhao, and Y. Li, "Time-Series representation and clustering approaches for sharing bike usage mining," *IEEE Access*, vol. 7, pp. 177856-177863, 2019.

[21] L. Huang, H. Feng, and Y. Le, "Finding water quality trend patterns using time series clustering: A case study," in *Proc. IEEE Fourth International Conference on Data Science in Cyberspace*, 2019.

[22] A. Martins, J. Lagarto, and M. G. M. S. Cardoso, "Electricity market price analysis using time series clustering," in *Proc. International Conference on the European Energy Market*, 2019.

[23] D. P. Kingma and M. Welling, "Auto-Encoding variational bayes," in *Proc. International Conference on Learning Representations*, 2014.

[24] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972-976, 2007.

[25] H. A. Dau, *et al*., "The UCR time series archive," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1293-1305, 2019.

[26] Y. Chu and B. Xu, "A RBF neural network classifier based on manifold analysis and AP algorithm," *Journal of Huazhong University of Science and Technology*, vol. 40, no. 8, pp. 23-25, 2012.

Min-Tong Su. Min-Tong Su was born in 1995 in Hunan province of China. He graduated in 2017 at the Harbin Engineering University. Since 2017, he continued his postgraduate studies at Central South University. Research area is related to the data analysis.

Jin Zheng. Jin Zheng is an associate professor at Central South University, mainly engaged in big data-related algorithm analysis and application, wireless sensor network distributed data management key technologies and other research.

Zu-Ping Zhang. Zu-Ping Zhang is a professor at Central South University and has been engaged in big data and knowledge engineering, information measurement and information fusion, software engineering and information system, parameter computing and biological computing for a long time.