ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 15, No. 2, 2020, *pp*.146-153

Research on user behavior recognition based on 2D-CNN

Lei Zhao
School of Mathematics and Statistics, Nanjing University of Information Science & Technology,
Nanjing 210044, China
(Received July 10, 2020, accepted August 12, 2020)

Abstract: In this paper, we studied the application of two-dimensional convolutional neural networks to the classification of multivariate time series. Time series sample data is usually a set of measurement values of a single attribute or multiple attributes at continuous time points separated by uniform time intervals. It is a set of structured data, usually non-discrete, time-related between data Features such as sex, feature space, and large dimension. At present, most methods for time series classification problems need to go through an extremely complex data preprocessing process and related feature engineering and do not consider the long pattern information hidden in different time dimensions of time series data, and the different characteristics of multivariate time series data Relevant information in the space dimension between. By converting multivariate time series data into matrix form, this paper proposes an end-to-end deep learning model Pyramid-CNN based on two-dimensional convolutional neural networks, which uses two-dimensional convolution kernels to extract the spatial dimensions of multivariate time series data and the relevant information in the time dimension, and applied it to the user behavior recognition time series data set. The experimental results show that for this data set, compared with the existing methods, the model proposed in this paper has higher performance Accuracy and robustness, with a good classification effect.

Keywords: time series classification, deep learning, convolutional neural network

1 Introduction

Today's society has entered the information age of big data. With the rapid development of information technology, data has exploded in various fields and industries, such as stocks[1-3], currencies, precious metals, futures, and other trading quotations and buyers in the commercial field. And seller information feedback, etc.; in the field of science and technology robot detection records, aerospace information, mechanical control, etc[4][5].; in the field of medical information records[6], medical imaging records, disease monitoring, etc.; in the field of social media network chat content Records, digital images, video, and audio, etc. Every day, billions of searches are performed on the Internet, and hundreds of millions of megabytes of information are transmitted. It can be said that data is everywhere, and mankind has officially entered the era of big data. When faced with a huge amount of information, how to choose the available information for use has become an issue of widespread concern, and data mining research is extremely important.

Time-series mining has always been a hot issue in the field of data mining. In recent years, data mining summit KDD and neural information processing summit NIPS have held special seminars on time series related issues every year. Time series data refers to a kind of structured data formed by recording sample attribute values in chronological order. It often has the characteristics of high dimensionality, a large amount of data, noise, and different sample lengths. Time series analysis is to analyze the data change process and future trends based on historical record values. With the advent of the era of big data and the development of computer hardware technology, a large amount of diverse time series data has been generated in the fields of aviation, finance, medical treatment, and industrial production. For example, in motion recognition, the angle changes of the human skeleton joints are obtained as the characteristics of motion. The model measures the similarity of the change sequence for action recognition; in the financial market, it uses time series clustering to study the financial investment portfolio. Further exploration of the hidden information in the original time series will help discover knowledge, laws, and patterns. The problem of time series classification is one of its important research directions. It has obvious application value and theoretical guiding significance. It has been studied and explored for many years.

Time-series mining has always been a hot issue in the field of data mining. In recent years, data mining has topped the list of issues. Multivariate time series classification has been widely used and more complex than univariate time series data. It has always been a research focus on timing issues. This article mainly studies the problem of predicting user behavior based on the time-series data of the user's mobile phone sensor. This has a strong application background in preventing financial fraud and anti-fraud.

Without loss of generality,a labeled data set with N samples $\mathrm{X} = \{(\mathrm{x_i}, \mathrm{y_i})\}_{i=1}^{N} \in R^{N\times l}$. among them, $(\mathrm{x_i}, y_i)$ represents a sample pair, sample $\mathrm{x_i} = \{x_i^1, \mathrm{x}_i^2, \ldots, x_i^l\}$has l observations, for multivariate time series data, $\mathrm{x}_i^{\mathrm{j}}$ is a $\mathrm{p} \times \mathrm{l}$ vector, p is the number of sample variables, $\mathrm{y_l}$ is its corresponding label, there are usually C possible category values. The time series classification problem is to learn a mathematical model that can predict the corresponding label y for a given new input sample x.

Traditional time series methods are mainly distance-based methods, such as DTW[6], ED[7] (Euler distance) feature-based methods, such as decision trees[8], Bayes[9], etc., and feature-based integrated learning methods, such as RF[10], XGBoost[11], LightGBM[12], etc. . However, traditional methods not only have higher data requirements, such as distance-based methods, the length of the sample must not be too short, and when the amount of data is large, the prediction time is longer and cannot be applied to real-time scenarios. Moreover, the requirements for engineers are also high. For example, for feature-based methods, it is necessary to extract features artificially. This requires engineers to have an understanding of specific business scenarios and the subject knowledge required by the business. This requires the knowledge of engineers. It must have a certain breadth and a certain depth, which requires a higher quality of engineers.

Compared with the current mainstream time series algorithms, deep learning models do not require heavy and complex data preprocessing processes and feature engineering, and the accuracy of classification has reached an advanced level. The time series algorithms based on deep learning have attracted the interest of researchers. In this paper, two-dimensional convolutional neural networks are applied to multivariate time series classification problems, and the application of two-dimensional convolution kernels enables the network model to extract local information and cross-combination information between different features. Sequence features have a new perspective, which improves the generalization ability of the model. This provides a new idea for the future application of deep learning technology to solve time series problems, and the development of a new deep learning model framework for time series related problems. It has a promoting effect.

2 Related work

In the research of time series classification, distance-based methods, feature-based methods, model-based methods, integrated learning-based methods, and deep learning-based methods are mainly used.

In the method based on distance measurement, the discriminative distance measurement plays a vital role in the performance of the classification model. First, define the distance function to calculate the similarity between two time-series, and then classify the sequence instances into the corresponding class according to the class of each time series instance and the closest instance in the training data. Commonly used distance measurement methods include Euclidean distance, dynamic time warping distance, longest common subsequence[13], and edit distance[14]. The earliest algorithm based on distance measurement is the method based on Euclidean distance. This method mainly measures the similarity between different time series samples by calculating the Euclidean distance between them, and then uses the KNN method to predict the label. But there is a problem with Euclidean distance. It requires the same length between different samples. In actual scenarios, the lengths between different time series samples are often different. For the case of the unequal length of time series, edit distance and dynamic time warping distance is usually used to solve the problem. There are two forms of edit distance: one is to measure the distance based on the number of conversions used to convert one time series to another; the other is to measure the distance based on the length of the longest common subsequence in two time-series data. Certainly, the edit distance is suitable for dealing with locally disturbing sequences, but not suitable for dealing with samples with severe phase distortion. The nearest neighbor algorithm based on DTW distance has been widely recognized and has been successfully applied to tasks such as time series classification and clustering.

The feature-based classification method usually includes two steps: defining time-series features and then training a classifier based on the defined time-series features for classification. The time series forest [15] algorithm calculates the mean, variance, and slope of the random sequence fragments, takes these statistical features as the characteristics of a fragment and then uses the random forest algorithm to search in a huge feature space, and the classification results are voted through all classification trees. The bag-of-words model is an extension of TSF. It uses the idea of the bag-of-words model method in the field of natural language processing to first obtain a new segment representation from the original time sequence segment, establishes the probability estimate of the sequence segment, and then use these probabilities to estimate the original. The time-series establishes the bag-of-words feature, and finally, similar to TSF uses the random forest to search and classify. Nanopoulos[16] et al. extracted statistical features, including features such as the mean and variance of the entire time series, and then used a multilayer perceptron neural network for classification. This method only extracts the global attributes of the time-series, but ignores the local features of the time series that contain important classification information. Geurts[17] extracts local time-series information features by discretizing time series so that traditional classification algorithms can classify time series data. The shapelet-based [18] time series classification algorithm searches for the best shapelet from the shapelet candidate space. When classifying, calculate the distance between the sample to be classified and the shaplet. When the distance is less than the distance threshold set by a certain shapelet, the category of the time series is consistent with the category of the current shapelet, and a good shapelet also reflects the time series data. Important mode information makes the results interpretable. The dictionary-based method can only find a single independent pattern for the shapelet method, and cannot take into account the frequency of the pattern. It is proposed to use the frequency information of the pattern to classify the time series.

The core idea of the model-based method is to use a probability generation model to fit the distribution of time series data, and to use the difference between functional models to measure the distance of the original time series data. The probabilistic generation model at the bottom layer can reflect the dynamic properties of the sequence, and it can also solve the problem of variable length. The upper classifier directly uses the label information to improve the classification accuracy of the algorithm. The hidden Markov model [19] is a commonly used time series classification generation model. The other is the current more mainstream time series classification algorithm based on the savings pool model. The echo state network model is its main representative model. Compared with the Markov model, it has a stronger fitting ability and weaker assumptions. The echo state network is a predictive model. Some researchers have proposed algorithms for time series classification based on the hidden state space of the ESN savings pool and the parameter space of the model. [20] proposed to use Martin distance in the hidden state space of the savings pool to measure the difference between function models. The dynamic state algorithm [21] uses DTW to measure the distance between the hidden state sequences based on the hidden state sequence in the savings pool, which improves the classification ability of the algorithm. [22] Use the read parameters of the ESN model to represent the original time series, and the distance between the model parameters to represent the original time series distance. Related research shows that learning a discriminative distance metric in the model space can help improve the performance of the classifier. The reason is that in the non-Euclidean model space, the traditional Euclidean distance metric can no longer be a good measure of the difference between the function models. distance. Therefore, exploring effective distance metrics in the model space of ESN has certain research significance. In addition, the classification performance of current model-based methods needs to be further improved. The reason is that a single model may be difficult to fit the complex data distribution structure in the model space.

The method based on ensemble learning is to combine different classifiers to achieve higher accuracy. Different ensemble models integrate different features and classifiers. Lines and Bagnall [23] will build a weighted ensemble model PROR based on 11 classifiers measured by elastic distance. SE generates a classifier through the combination of shapelet transformation and heterogeneous sets. The transform-based integration method COTE [24] integrates 35 different classifiers based on the features extracted in the time domain and frequency domain. The ensemble algorithm based on transformation is a powerful ensemble algorithm. It integrates representative algorithms from the time domain, frequency domain, change, and Shapelet transformation domain on a large scale, and finally integrates 35 classifiers. The type of voting sequence sample. In addition, the LightGBM model and the random forest model, which have performed well in major data mining competitions in the past two years, have also performed well in the field of processing time series.

All of the above methods require complex preprocessing of data, or complex and heavy feature engineering, especially feature-based methods. In recent years, with the rapid development of neural networks and deep learning, deep neural networks have made outstanding contributions to time series classification, especially convolutional neural networks for end-to-end time series classification. The multi-scale convolutional neural network introduces feature extraction and classification into a deep learning model and uses down-sampling, skip sampling, and sliding window techniques to automatically extract features of different scales and frequencies. The algorithm achieves competitive classification performance. Karim [25] et al. proposed LSTM-FCN and ALSTM-FCN models, combined with CNN and RNN to establish an end-to-end model, explored the performance of long- and short-term memory networks on UCR time series classification databases and proposed a fine-tuning approach. Wang [26] et al. proposed a well-recognized and relatively strong end-to-end full convolutional network model. Compared with the current mainstream ensemble model, this model not only achieves the classification performance of the ensemble model, but also does not require a lot of data preprocessing and feature engineering. , The three-layer convolution module can automatically extract effective local features.

Deep learning can fully extract various features of time series. In time-series data, the convolutional neural network, as a kind of time series convolution, extracts time-series features and primary features and advanced features under different time windows (as the network layer deepens). However, whether it is a convolutional neural network, a recurrent neural network, or a deep learning method such as LSTM, it is the extraction of sequence features. This has an obvious effect on one-dimensional time series data, but for multivariate time series data, the variables are ignored. The implicit relationship between. This paper proposes that the use of a two-dimensional convolutional neural network can not only extract the long pattern information hidden in the multi-dimensional time-series samples, but also mine the hidden relationships between different features, which greatly improves the robustness of the model.

3 Two-dimensional convolutional neural network

In the past two years, more and more researchers have used various deep learning models to solve time series classification problems. Compared with time series ensemble models with high classification accuracy, deep learning models are an end-to-end learning framework, The model can automatically extract the effective features in the data, avoiding complex data preprocessing and feature engineering in the early stage, and the accuracy of some deep learning models is far ahead of most existing traditional models. Among them, the fully convolutional network is a recognized relatively strong time series classification model. The stacked three convolutional layers can well extract the local information hidden in the sequence. The application of the global average pooling layer greatly reduces the model parameters. First, better classification results were also obtained on the time series benchmark data set. However, the general convolutional neural network usually uses a one-dimensional convolution kernel to construct the network, which usually has a significant effect in processing univariate time variable data, but it often does not work well for multivariate and high-dimensional time series data. The dimensional convolution kernel is restricted by its own structure, and cannot extract relevant information between different variables, and cannot dig out the effective information hidden between sequence features. Aiming at this shortcoming, this paper proposes to use a two-dimensional convolution kernel to construct a convolutional neural network to process multivariate time series data, and apply it to user behavior recognition data sets, achieving better performance than traditional integrated learning and deep learning models. result.

A time-series data set is a set of series data, observations in a period of time at equal time intervals. In the time series classification problem, any real-valued ordered data can be regarded as a time series. For multivariate time-series data, each sample has multiple pieces of time series data. Suppose there is a set of time series $\mathbf{T} = \{\mathbf{T_1}, \boldsymbol{T_2}, \dots, \boldsymbol{T_n}\}$ with data size n. Each time series sample has $\mathbf{m_i}$ observations, that is,$\mathbf{T_i} = \{\boldsymbol{t_{i1}}, \boldsymbol{t_{i2}}, \dots, \boldsymbol{t_{im_i}}\}$, corresponding to the category labels $\mathbf{y_i}$, $\mathbf{y_i} \in \{\mathbf{1}, \mathbf{2}, \dots, \boldsymbol{C}\}$, and $\mathbf{C}$ represents the number of categories.

The structure of the two-dimensional convolutional neural network proposed in this paper is shown in Figure 1. Each convolution module first extracts relevant features through the convolution layer, and then processes the BN and activation layer, and finally extracts the high-level convolution module. The features are connected to the fully connected layer after global average pooling, and finally, enter the softmax classification layer for classification.

The two-dimensional convolutional neural network is usually used to deal with the problem of image classification and detection. It extracts the relevant features of the picture through the convolution operation of the two-dimensional convolution kernel, abstracts layer by layer, and combines low-level features into high-level features. The format of an image in computer storage is a two-dimensional matrix, and the convolution operation of the two-dimensional convolution kernel to extract image features is ultimately to extract relevant information in the matrix data. For multivariate time series data, we can also regard it as two-dimensional matrix data, use a two-dimensional convolution kernel to extract the relevant features of the time series data matrix, and combine high-order features of different attributes to improve the generalization ability of the model.

In a multi-layer two-dimensional convolutional neural network, the output of the previous layer is usually used as the input of this layer, the input data and the convolution kernel of this layer are convolved, and the output after the convolution operation is offset from the layer After the addition, it is activated by the activation function of this layer as the input data of the next layer. The formula is as follows:

$$z_j^l = \sum_i a_i^{l-1} * w_{ij}^l + b_j^l \tag{3.1}$$

$$a_j^l = f(z_j^l) \tag{3.2}$$

Among them, $*$ represents the convolution operation, a_i^{l-1} and a_i^l are the input and output of the convolution kernel, $f(.)$ is the activation function, which is the non-linear unit in the network, usually the ReLU activation function, The definition formula is as follows:

$$ReLU = \begin{cases} x, & x > 0 \\ 0, & x \le 0 \end{cases} \tag{3.3}$$

For the convolutional neural network, the convolution module is only a part of the network, and each convolution module is also connected with batch standardization technology and ReLU activation function. The calculation process of each convolution block is as follows:

$$y = W \otimes x + b \tag{3.4}$$

$$B = BN(y) \tag{3.5}$$

$$A = ReLU(B) \tag{3.6}$$

Among them, $\otimes$ represents the convolution operation, batch normalization is a commonly used data processing technique in the neural network training process, when the data is divided into small batches for gradient descent, the input of the previous layer of the network is first passed through the batch normalization process and then transmitted Entering the next layer of a network can speed up the training process of the network, accelerate the convergence and improve the generalization ability of the model, and prevent the gradient from disappearing and exploding.

After convolution by the multi-layer convolution module, the data will be passed to the global maximum pooling layer. The global maximum pooling layer is different from the maximum pooling layer. The maximum pooling layer is a subregion of the convolutional neural network feature map. The maximum pooling operation is performed, and the global maximum pooling is to perform the maximum pooling operation on the entire feature map. The application of the global maximum pooling layer can not only achieve the effect of regularization, but also greatly reduce the parameters of the model weight, and improve the generalization ability of the model. The following figure shows the structure of the convolutional neural network proposed in this article:

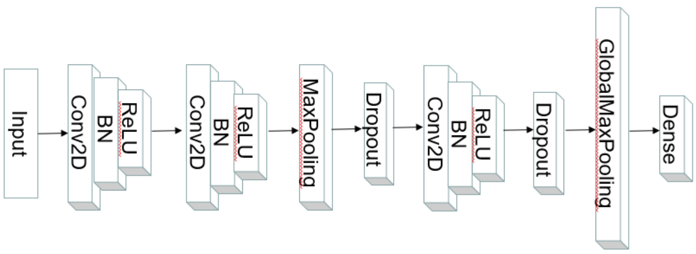

Fig.1 Network architecture of Pyramid-CNN

4 Experiment

4.1 Dataset

The data set used in this article is the 2020 Xinwang Bank Jiaozi Cup user behavior recognition data set. The data set is the offset in space of the mobile phone of various behaviors transmitted by the mobile phone sensor when the user uses the mobile phone. There are 7500 samples in total. Each sample is about 45~65 long and has 6 attributes: x-axis offset, y-axis offset, z-axis offset, x-axis drive offset, y-axis drive offset, and The offset of the z-axis driving amount. There are 19 categories of category tags, including 3 scenes: walking, standing, sitting, and lying; and 6 types of actions in the three scenes: playing games, using short videos such as vibrato or fast hands, watching long videos such as TV dramas or variety shows, Web browsing, text editing, other actions (such as taking photos, voice calls, video calls, voices, picture PS), and handing mobile phones are actions that do not distinguish between scenes. There are 19 categories (because privacy is involved, specific categories are not disclosed, Replace only with numbers). Some data of different categories are shown in the figure below:

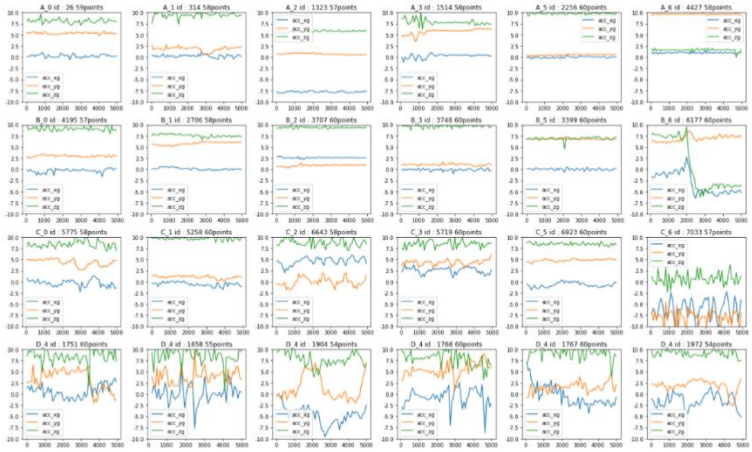

Fig.2 Schematic diagram of the offset of each category driving amount

From Figure 1, we can see the difference in the offset of different categories. According to the x, y, and z-axis offset data, we can naturally construct the character of the sample's modulus length in space, the formula is as follows:

$$m = \sqrt{x^2 + y^2 + z^2} \tag{4.1}$$

At this time, each sample in the data set has 8 attributes, but from Figure 1, we can see that the sequence length of each sample is different. This paper uses Fourier sampling to uniformly sample it to 60, and then The attributes of each sample are connected together to construct a single-channel matrix of $60 \times 8 \times 1$.

4.2 Experimental platform and related settings

The experimental environment is a laboratory personal computer, the system is Window10, and the GPU is GeForceGTX 1070, 6 GB. All models are trained in the Keras deep learning framework based on TensorFlow.

In the experiment, the convolutional neural network module consists of 4 convolutional blocks, two Dropout layers, a maximum pooling layer, a global maximum pooling layer, and a fully connected layer. After each convolution block, there is a BN layer and an activation layer, and the activation function is ReLU. The number of iterations is 500, the batch-size is 32, the optimizer is Adam, and the initial learning rate is 0.001. Set the callback function to adjust the learning rate during training. During 10 consecutive iterations, if the model accuracy is not improved, then learn The rate attenuation is 0.5. The classification process of the user behavior recognition data set in the two-dimensional convolutional neural network proposed in this article is as shown in Table 1.

4.3 Experimental results

The data set used in this article is user behavior recognition time series data. The category labels are divided into 3 kinds of scenes: walking, standing, sitting, and lying; and 6 kinds of actions in the three scenes: short videos such as playing games, brushing vibrato or fast hands, Watching long videos such as TV dramas or variety shows, browsing web pages, editing text, other actions (such as taking photos, voice calls, video calls, sending voices, picture PS), and handing mobile phones are actions that do not distinguish between scenes. There are 19 actions in total Category is a multivariate time series classification problem. This article uses classification accuracy as the evaluation index of the model, which is defined as follows:

$$acc = \frac{The\ sum\ of\ the\ scores\ on\ all\ test\ samples}{Number\ of\ test\ samples} \tag{4.2}$$

Among them, the prediction score rule is as follows: if it is the behavior of handing over the phone (the 19th category), the prediction is correct and the prediction is wrong 0 points; if it is not the behavior of handing the phone, then the scene and action are completely predicted correctly, 1 point only correctly predicts the scene, 1/7 points only correctly predict the action, 1/3 points; scene and action are all incorrect, 0 points.

Table 1. Two-dimensional convolutional network classification process

Layers		size
(1)	Input	(60,8,1)
Conv2D Batch normalization Activation	First layer convolution	(60,8,64) (60,8,64) (60,8,64)
Conv2D Batch normalization Activation	Second layer convolution	(60,8,128) (60,8,128) (60,8,128)
Maxpooling2D	Pooling	(30,8,128)
Dropout	0.2	(30,8,128)
Conv2D Batch normalization Activation	Third layer convolution	(30,8,256) (30,8,256) (30,8,256)
Dropout	0.3	(30,8,256)
GlobalMaxPooling2D	Pooling	(,256)
Dense		(,64)
softmax	output	(,19)

Because distance-based models are not easy to handle multivariate time series problems, this article selects the current integrated learning models with strong classification capabilities in the machine learning field: Random Forest and LightGBM, as well as one-dimensional full convolutional neural networks commonly used in the field of deep learning to deal with sequence problems As a comparison between the network and the RNN network based on LSTM units, the specific results are as follows:

Table 2. Comparison of classification accuracy of different models

Model	Acc
RF	0.675
LightGBM	0.690
FCN(Conv1D)	0.752
LSTM	0.640
Pyramid-CNN(Conv2D)	0.782

From the above table, we can see that the two-dimensional convolutional network is stronger than the representative model random forest and LightGBM in the traditional machine learning model, and it is also stronger than FCN and LSTM. Because deep learning models can extract features from themselves, their accuracy is generally higher than traditional machine learning models. Two-dimensional convolutional neural networks have more advantages than one-dimensional convolutional neural networks for multivariate time series data. It can not only mine sequence The information in the time dimension can also dig out the information in the spatial dimension between different characteristics of the sample, which improves the generalization ability of the model. Moreover, the two-dimensional convolutional neural network can be applied to multivariate time series problems, which means that such problems can use many techniques in the field of computer vision in deep learning, such as the Dropout layer and BN layer we have used, which greatly improves the model's performance The generalization ability improves the accuracy and robustness of the model. The application of a two-dimensional convolutional neural network also provides a new solution to the related problems in the field of time series with high-dimensional and multi-time nodes.

5 Conclusions

Time series analysis has always been a research hotspot in the field of data mining, and time classification is also an important research direction in time series analysis. Aiming at the shortcomings of traditional time series classification methods, an end-to-end Pyramid-CNN network is proposed through the study of deep learning models. In this paper, by using a two-dimensional convolution kernel to convolve on multivariate time series data, not only the information of different time scales but also the information of different feature combinations are obtained. In this paper, this model is applied to user behavior recognition and compared with some existing methods, a higher accuracy is achieved. We will further study the model and learn the ideas of various construction models in the field of deep learning so that the model can extract more comprehensive features and improve the accuracy and robustness of the model.

References

[1] DELGADO M，CUELLAR M P，PEGALAJAR M C．Multiobjective hybrid optimization and training ofrecurrent neural networks．IEEE Transactions on Systems，Man，and Cybernetics，Part B(Cybernetics)，2008，38(2)：381-403．

[2] YE L，KEOGH E．Time series shapelets：a new primitive for data mining//Proeeedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.2009：947—956．

[3] CHENG CH，YANG J H.Fuzzy time-series model based on rough set rule induction for forecasting stock price．Neurocomputing，2018,302:33-45．

[4] Keogh E, Kasetty S. On the need for time series data mining benchmarks: a survey and empirical demonstration. Data Mining and Knowledge Discovery, 2003, 7(4): 349-371.

[5] LIU B，LI J，CHEN C，et a1．Efficient motifdiscovery for large-scale time series in healthcare i-．IEEE Transactions on Industrial Informatics，2015，1 1(3)：583-590.

[6] Faloutsos C, Ranganathan M, Manolopoulos Y. Fast subsequence matching in time-series databases//Proceedings of the ACM SIGMOD International Conference on Management of Data, Minneapolis, May 24-27, 1994. New York:ACM, 1994: 419-429.

[7] GHASSEMI M，PIMENTEL M A，NAUMANN T，et a1．A multivariate timeseries modeling approach to severity of illness assessment and forecasting in icu with sparse，heterogeneous clinical data//Twenty-Ninth AAAI Conference on Artificial Intelligence．2015．

[8] W. Buntine.Decision tree induction systems:a Bayesian analysis.Uncertainty in AI, 3:109–127,1989.

[9] Barber, D. (2012).Bayesian Reasoning and Machine Learning.Cambridge: Cambridge University Press

[10] Understanding Random Foresthttps://towardsdatascience.com/understanding-random-forest-58381e0602

[11] Chen, T. & Guestrin, C. 2016, in Proceedings of the 22Nd ACM SIGKDD In-ternational Conference on Knowledge Discovery and Data Mining, KDD ’16(New York, NY, USA: ACM), 785–794G

[12] Berndt D J, Clifford J. Using dynamic time warping to find patterns in time series//Proceedings of the AAAI Workshop on Knowledge Discovery in Databases, Seattle, Jan,1994. Menlo Park: AAAI, 1994: 359-370.

[13] Deng H T, Runger G C, Tuv E, et al. A time series forest for classification and feature extraction. Information Sciences,2013, 239: 142-153.

[14] Nanopoulos A, Alcock R, Manolopoulos Y. Feature-based classification of time-series data//Information Processing and Technology. Commack: Nova Science Publishers, Inc,2001.

[15] Geurts P. Pattern extraction for time series classification// LNCS 2168: Proceedings of the 5th European Conference on Principles and Practice of Knowledge Discovery in Databases, Freiburg, Sep 3-5, 2001. Berlin, Heidelberg: Springer,2001: 115-127.

[16] HILLS J，LINES J，BAR ANAUSKAS E，et a1．Classification of time series by shapelet transformation．Data Mining and Knowledge Discovery,2014，28(4)：851．881．

[17] RABINER L R A mtorial on hidden markov models and selected applications in speech recognition．Proceedings ofthe IEEE，1989，77(2)：257-286．

[18] ZHANG L，LI Y CHEN H．An effective martin kernel for time series classification//International Conference on Neural Information Processing．Springer,2017 384-393．

[19] BAGNALL A，DAVIS L，HILLS J，et a1．Transformation based ensembles for time series classification//Proccedings of the 2012 SIAM international conference on data mining．2012：307-3 18．

[20] GONG z'CHEN H．Sequential data classification by dynamic state warping．Knowledge and Information Systems,20 1 8，57(3)：545-570．

[21] Lines J, Bagnall A J. Time series classification with ensembles of elastic distance measures. Data Mining and Knowledge Discovery, 2015, 29(3): 565-592.

[22] Bagnall A J, Lines J, Hills J, et al. Time-series classification with COTE: the collective of transformation-based ensembles//Proceedings of the 32nd IEEE International Conference on Data Engineering, Helsinki, May 16-20, 2016.Washington: IEEE Computer Society, 2016: 1548-1549.

[23] Karim F, Majumdar S, Darabi H, et al. LSTM fully convolutional networks for time series classification. IEEE Access,2018, 6: 1662-1669.

[24] Wang Z G, Yan W Z, Oates T. Time series classification from scratch with deep neural networks: a strong baseline.