Study of feature extraction method of multi-information source for continuous casting process parameters

Xianqi Liang[1, a] , Zhufeng Lei[1] , Wenbin Su[1], Fali Zhong[1], Hang Zhao[1], Yanni Zheng[1]and Mingjie Ding[1]

Abstract. The control of the roll gap of the segment is one of the key links to ensure the quality of cast billet. In this paper, the big data in traditional continuous casting production operations is studied through in-depth experimental comparative analysis of linear and nonlinear dimension reduction method. The method is suitable for continuous casting to obtain the data of the dimension reduction. The method of principal component analysis is improved by using standardized data increment method. A faster and more efficient method of dimension reduction is obtained when the unrelated data, training time and reconstruction error are removed. Actual data simulation results show that this method is more efficient and suitable for continuous casting than any other dimension reduction method.

1. Introduction

In the modern steel industry, high-efficiency continuous casting technology, including key continuous casting equipment and new processes, has become the most internationally competitive core technology. The process of continuous casting is a complex continuous phase change process. There are many control aspects that affect the quality of slab. The key technology and core research of high quality steel continuous casting process are mainly focused on the precise control of mould, sector and composition of liquid steel [1]. Among them, the control of sector roll gap is one of the key links to ensure the quality of slab [2]. Not only the establishment of a dynamic and adaptive prediction model for the roll gap value of the sector continuous casting machine but also the real-time adaptive adjustment of the sector roll gap value according to the actual working conditions has important theoretical and practical values for improving the quality of slab.

1.1 motivation()

Due to the complexity of the continuous casting process as well as the non-intuitive and variability of the internal condition of the high-temperature slab, there are many factors affecting the roll gap parameter in the production process, resulting in many measurement points required for data collection. Therefore, the sampling frequency is higher and the sampling time is longer, so that massive data can be obtained [3.4.5]. For example, Handan Iron and Steel Group installed 6000 measuring points in a continuous casting production line to collect data every 3 seconds, so the amount of monitoring data is extremely large. Because of a large amount of data, how to refine the data related to the roll gap parameter and how to establish the roll gap prediction model service are the key to handle these data. Therefore, how to research and use the new method for data mining process parameters of continuous casting is a problem faced by this research.

Traditional data processing is usually based on the traditional method of dimension reduction, such as principal component analysis (PCA) and linear discriminant analysis (LDA). The process parameters of continuous casting have a lot of problems; for instance, multi-information source, real-time and large amount of data. Hence the traditional method of dimension reduction obviously cannot meet the requirements.

The third literature introduces an incremental principal component analysis（Candid Covariance-free Incremental PCA, CCIPCA）that does not compute covariance, which converges quickly and reduces computational cost when dealing with high-dimensional data. The fourth article introduces an incremental principal component analysis (Chunk IKPCA, CIKPCA), which decomposes bulk data. This method performs an eigenvalue calculation on a large data block and reduces the computational cost. MKaur[26] used LDA as the classification method from accuracy, sensitivity and prediction accuracy in the process of ECG signal analysis, and finally found that the isometric rotation method has the highest accuracy.

1.2Contribution

Many predecessors did a lot about the method of feature reduction which is better to solve the problem of real-time data and large amount of data，but some of them considered less about the accuracy and effectiveness of the reduction. The traditional linear dimension reduction method (PCA) and the emerging nonlinear dimension reduction methods (t-SNE, LLE, LTSA) were compared on the requirement of multi-information source, real-time and large amount of data of continuous casting production process parameters. Based on that, the dimension reduction methods which are suitable for predicting the roll gap parameters in the continuous casting process are analyzed and improved. Finally, a standardized robust incremental principal component analysis (Normalization Robust Incremental Principal Component Analysis, NRIPCA) was found. Before the reduction of the data of continuous casting, this method

firstly performs outlier monitoring and elimination and normalizes the processed data. This method considers not only the accuracy of feature reduction but also the validity of feature reduction.

2 Methodology(MATEC Web of Conferences)

Feature reduction removes irrelevant and redundant information from features based on guidelines. It maintains the important features and it can effectively reduce the data dimension for overcoming the 'dimension disaster' problem. This section compares the traditional linear dimension reduction method with the emerging nonlinear dimension reduction method. Thus, it finds out the dimension reduction method which is suitable for the prediction of the roll gap parameters of continuous casting.

2.1 Principal Component Analysis (PCA)

As a mainstream method of linear dimension reduction, PCA has the most linear reconstruction error because its theory is complete. So its application is very extensive [6,7,8,17,18].PCA is mainly looking for a series of orthogonal basis vectors that maximize the variance of all data.

Suppose the sample set $X_i = \{x_1, x_2,...x_m\}$, in order to find a subspace to make the hyperplane projection of sample point xi is $W^T x_i$ in the subspace.

If the projection of all sample points is separated as much as possible, the variance of the projected sample point is the largest. The variance can be expressed as $\sum_i W^T x_i x_i^T W$; the optimization goal can be written as

$$\max_W tr(W^T XX^T W) s.t. W^T W = I$$
(1)

In equation (1) use the Lagrange multiplier method

$$XX^T W = \lambda W \qquad (2)$$

Among them, W is the best projection matrix, λ is the eigenvalue.

Therefore, we only need eigenvalue decomposition of the covariance matrix XX^T, then the eigenvalues will be sorted: $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ .Then, taking the first d eigenvalues corresponding to the eigenvectors $W = (w_1, w_2,...w_d)$ is the solution of principal component analysis.

2.2 T-distribution Stochastic Neighbor Embedding (t-SNE)

2.1.1 Formatting the title
The title is set in bold 14-point Arial, flush left, unjustified. The first letter of the title should be capitalised with the rest in lower case. You should leave 35 mm of space above the title and 6 mm after the title.
2.1.2 Formatting author names and author affiliations

Author names should be typed in 9-point Arial. The style for the names is First Names then Last Name, with a comma after all but the last two names, which are separated by "and". Do not use academic titles.

Affiliations of authors should be typed in italic 8-point Arial. They should be preceded by a numerical superscript corresponding to the same superscript after the name of the author concerned. Please ensure that affiliations are as full and complete as possible and include the country.

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)}$$
(3)

$$p_{ij} = \frac{p_{j|i} p_{i|j}}{2N}$$
(4)

Step2 Dimension reduction: learn probability distribution at low latitudes to define a d-dimensional map $y_i,....y_N, y_i \in R^d$ .Assuming that the similarity between y_i and y_j is q_{ij}.

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_{k \neq i}\left(1 + \|y_k - y_l\|^2\right)^{-1}}$$
(5)

The relative entropy (Relative Entropy) is introduced to measure the similarity between Q and P.

$$C = KL(P\|) = \sum_{i \neq j} P_{ij} \log \frac{p_{ij}}{q_{ij}}$$
(6)

Using the gradient descent method makes the relative entropy minimum.

$$\frac{dC}{dy_i} = 4\sum_j \left(p_{ij} - q_{ij}\right)\left(y_i - y_j\right)\left(1 + \|y_i - y_j\|^2\right)^{-1}$$
(7)

The advantage of t-SNE is that it takes a smaller distance to create a larger gradient that separates dissimilar points, and this gradient does not go too far away from dissimilar points. Although t-SNE is almost impossible to surpass the classic, there are some deficiencies. t-SNE is mainly used for visualization and cannot be used for other purposes. t-SNE is mainly used to preserve the local features because it is difficult to map the high intrinsic dimension data set to 2-3 Dimensional space very well. t-SNE is a probabilistic distribution problem and is not possible to consider reconstructing a model after dimension reduction.

2.3 Locally linear embedding (LLE)

Locally linear embedding (LLE) was proposed by S. T. Roweis in 2000. [22,23,24] It is a non-linear method of unsupervised dimension reduction. After dimension reduction of low-dimensional, the data can maintain the topology of the original data. This has been widely used in image classification and clustering, face recognition, natural

language processing, multidimensional data visualization and bioinformatics.

Step1 Establish the probability distribution of high dimensional objects: assume hypothetical high-dimensional observation dataset $X = \{x_1, x_2, ... x_n\}$, then search for k nearest neighbors $X_i = [x_{i1}, ..., x_{ik}], i = 1, ... N$.

Step2 Calculate the coefficients for each point which is reconstructed by the domain point: The coefficient of each sample point x_i linearly reconstructed by its neighborhood X_i can be obtained by minimizing the reconstruction error.

$$e(f_i) = \left\| x_i - \sum_j f_{ij} x_{ij} \right\|^2 \quad (8)$$

$$s.t. \sum_j f_{ij} = 1 \quad (9)$$

Among them, $\sum_j f_{ij} x_{ij}$ represents the linear reconstruction of point x_i, $f_{ij} = f(x_i, x_{ij})$ indicates that x_i is a coefficient reconstructed by x_{ij}. The solution

$$f_{ij} = \frac{\sum_h G_{jh}^{-1}}{\sum_{i,m} G_{im}^{-1}}$$

is .

Among them, G is the local Gram matrix: $G_{jh} = (x_i - x_{ij})^T (x_i - x_{ih})$. Reconstruction factor f_i describes the neighborhood geometry for rotation invariance and translation invariance for x_i and its neighbor point X_i.

Step3 Construct low dimensional embedding: Low-dimensional embedding is obtained by maintaining the weight of reconstructed data in each neighborhood in low-dimensional data; the objective function is

$$J(Y) = \sum_i \left\| yi - \sum_j f_{ij} y_j \right\|^2 \quad (10)$$

$$s.t. \frac{1}{N} \sum_i y_i y_i^T = I \, \& \, \sum_i y_i = 0 \quad (11)$$

So the low-dimensional data is still maintained at f_{ij}, the embedding result is completely characterized by the reconstruction weights that characterize the intrinsic geometry. The advantage of LLE algorithm is simpler calculation and less parameter selection. It can optimize the characteristic parameters. There is no local minimum problem. It can retain the essential characteristics of data

and express the internal structure of the data. However, LLE is sensitive to signal noise, because it is limited by the local neighborhood parameter k and the target dimension of the low-dimensional data, affecting the dimension reduction effect of high-dimensional data.

2.4 Local tangent space alignment (LTSA)

LTSA was proposed by Zhang Zhenyue and others from Zhejiang University in 2004 [25,26]. The method is mainly to gradually align all the tangent planes when a manifold is unfolded. The local geometry of the low-dimensional data is constructed by approximating the tangent space of each sample point. The local low-dimensional coordinates are obtained by the projection of the observed data points in the local tangent space. The local low-dimensional coordinates of the overlap are locally transformed to obtain the global low-dimensional embedding coordinates.

Step1 Suppose the high-dimensional observation dataset $X = \{x_1, x_2, ... x_n\}$, then search for its k nearest neighbors $X_i = [x_{i1}, ..., x_{ik}], i = 1, ... N$.

Step2 Calculate the base vector of the d-dimensional local tangent space of the neighbor dataset X_i and the local tangent coordinate Θ_i of the observed data.

Step3 According to the N local projection coordinates $\Theta_i = [\Theta_{i1}, ..., \Theta_{ik}], i = 1, ... N$, the global coordinates $\{y\}_{i=1}^N$ are obtained by the local affine transformation $Li \in R^{d \times d}$.

LTSA can reflect the local geometric characteristics of the data well, but it cannot handle large sample data, and it does not work well for popular learning with high curvature.

2.5 Comparison of four methods for dimension reduction

The dimension reduction methods will be tested with the actual production of continuous casting data in this section.
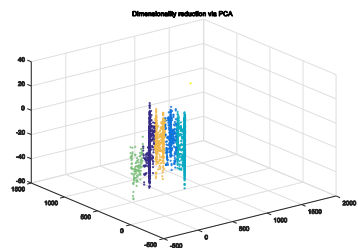

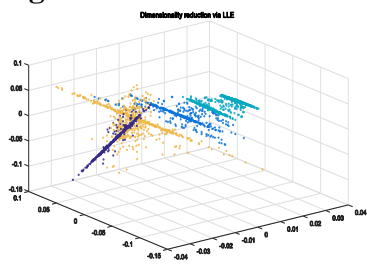**Figure. 1** Dimension reduction effect of PCA

Corresponding author: Xianqi Liang:1012363162@qq.com;Zhufeng Lei: 253574602@qq.com; Wenbin Su: 809355146@qq.com

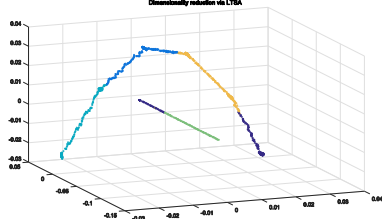**Figure. 2** Dimension reduction effect of LLE

Figure. 3 Dimension reduction effect of LTSA

As can be seen from the figures, PCA dimension reduction effect is normal and can reduce the dimension of the data, but the data classification is not obvious after the dimension reduction. In contrast, LTSA dimension reduction effect is better. Effective dimension reduction and data classification is obvious; different types of data aggregation are obvious. The reason why the two types of methods produce different results is that, in principle, linear dimension reduction mainly analyzes the importance of a certain dimension data, reduces the dimension according to its importance and deletes the data of lesser importance. In order to achieve dimension reduction, linear dimension reduction method is mainly to carry out the data clustering, as well as high similarity data aggregation and projection in low-dimensional space.

In the roll gap prediction model of continuous casting, the characteristics of flow data from multi-information source determine that the feature reduction method should serve as the roll gap prediction model. When the nonlinear dimension reduction method is used, the original data is processed, resulting in better dimension reduction and improved operational efficiency. However, the data after dimension reduction cannot be evaluated; the prediction of the roll gap parameter cannot be realized. After repeated tests and comparison, the linear dimension reduction method is used to characterize the dimension of the continuous casting roll gap prediction model.

3 Experimental and Research on Improved PCA

In this section, PCA will be modified to make it suitable for feature reduction of continuous casting process parameters characterized by multi-information source data.

3.1 The theoretical basis of NRIPCA

The NRIPCA method updates the PCA model in real time based on sliding window thinking, replacing the oldest data in the window with the latest data.
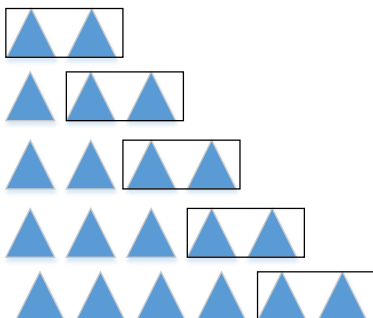

Figure. 4 Principle of sliding window

The main steps of this algorithm are:

(1) Data pre-process: use a certain amount of historical data for discrete point detection.

(2) The pre-process data are normalized to make the data mapped in a range, which makes the optimization process of the algorithm become more smooth and easier to correctly converge to the optimal solution.

(3) Principal component analysis: establish principal component analysis model and calculate the model parameters.

(4) Collect a new set of real-time data, and standardize the data with the current mean and variance.

(5) According to the current principal component analysis model, analyze whether the data is normal data, then go to the next step; otherwise, go back to the previous step.

(6) Add a new set of process parameters to the sliding window; meanwhile delete the oldest set of process parameters and update the principal component analysis model and related parameter values. Go to step 4.

3.2 data source

In order to verify the validity of the NRIPCA algorithm, a uniform parameter of a continuous casting production line of a steel mill was randomly collected in China for a week. These data include 32,154 dimensional data, of which 26,001 dimensional data are the switch data monitored by the control system, and 6153 dimensional data are valid process parameter monitoring data. The data in the data set cover important processes in the continuous casting process, including tundish, mould and sector segments. In order not to damage the sequence of data, Rada criterion is used to detect abnormal values of data, then the detected abnormal values are replaced by nearby values.

3.3 NRIPCA experiment on continuous casting flow data

Two important parameters of the NRIPCA algorithm are the size of both the sliding window and the sliding step. In order to study the size of both the sliding window and the step for the data set of this project, the total amount of data and the size of the step are changed respectively, and the number of stable principal components is obtained as the evaluation criterion. The result shows the ratio of sliding window size in the total amount of data.

Table 1 Sliding window size ratio of the total amount of data

The total amount of data	The ratio of sliding window size in each step to the total amount of data				
	30	60	120	240	480
One day （63700）	3/4	3/4	3/4	3/4	3/4
Two days （127400）	31/32	31/32	31/32	31/32	31/32
Three days （191100）	3/4	3/4	3/4	3/4	3/4

Four days（254800）	7/8	7/8	7/8	7/8	7/8
Five days（318500）	31/32	31/32	31/32	31/32	31/32
Six days（392200）	7/8	7/8	7/8	7/8	7/8
Seven days（445900）	3/4	3/4	3/4	3/4	3/4

From Table 1, it can be seen that the step size has little effect on the model within 480 (8 minutes), indicating that the change of continuous casting process parameters is a gradual process which can slow the model updating time in the subsequent online analysis. In other words, each group can reduce the computational cost without having to update the model parameters of the new process.

In this paper, a total of 446,000 samples are selected for sliding window PCA analysis in a week. The window size is 334500 and the step size is 60 (i.e. the principal component analysis model is updated once every minute). In order to calculate the reduction error, the latest model data is reconstructed with the current model parameters after each model update.

In order to quantitatively analyze the size of the reconstruction error, the average relative error between the reconstructed data of each parameter and the original data is calculated as shown in Table 2.

Table2 The average relative error of the process parameters

Process parameter name	Relative error（%）		Process parameter name	Relative error（%）	
	average value	Maximum		average value	Maximum
Tundish temperature of molten steel	0.11	0.65	Mold wide water flow	2.03	13.90
Overheating	4.89	22.32	Mold wide exit temperature	0.41	1.51
Pulling speed	1.51	6.71	Mold narrow surface pressure	2.94	21.28
Vibration frequency	1.39	9.03	Mold narrow surface water flow	0.23	2.52
Vibration amplitude	1.04	1.79	Narrow mold exit temperature	0.39	2.81
Drive roller pressure	4.24	22.22	Second cold water flow	4.89	42.69
Mold level	0.44	2.10	Second cold water pressure	7.73	54.24
Mold wide-face pressure	3.77	32.83			

As can be seen from Table 2, the average error between the original data obtained by PCA and sliding PCA is less than 5%, which is within the acceptable error range. Therefore, it is feasible to use the sliding window PCA to dynamically update the feature parameters of process flow data online.

4. Discussion

In this paper, fifteen kinds of process parameters that influence the roll gap value of continuous casting sector are reduced by one week's data. PCA and NRIPCA are used respectively. The former is to perform the feature reduction on the whole seven-day data; the latter uses part of the data as real-time data to simulate the process of online feature reduction. By observing the comparison between the reconstructed data and the original data, it can be found that both methods have a good fitting effect. In order to compare two methods more intuitively, the average reduction error of the reconstructed data and the original data of the two methods is shown in Table 3.

Table 3 The average relative error of reconstructed data of each parameter

Short for process parameters	PCA average relative error（%）	MWPCA average relative error（%）	Short for process parameters	PCA average relative error（%）	MWPCA average relative error（%）
GSWD	0.16	0.11	KMLL	1.26	2.03
GRD	4.25	4.89	KMWD	0.43	0.41
LS	1.53	1.51	ZMYL	4.35	2.94
ZDPL	1.12	1.39	ZMLL	0.68	0.23
ZDFZ	0.55	1.04	ZMWD	0.44	0.39
CDGYL	5.31	4.24	ELSLL	4.83	4.89
JJQYW	0.13	0.44	ELSYL	15.05	7.73
KMYL	2.58	3.77	Average	2.85	2.40

As can be seen from Table 3, NRIPCA reduction error is not small for each process parameter. And the reduction error of the two methods in the comparison of each parameter has its own advantages and disadvantages. But overall, the reduction error in NRIPCA is smaller than in PCA, meaning that NRIPCA has a small reduction error when feature reduction is performed on changing process parameters. From an algorithmic point of view, PCA calculates the global optimal reconstruction, while NRIPCA calculates the current local optimal reconstruction and has optimal reconstruction errors for each sliding window. Therefore, the overall reconstruction error in NRIPCA is smaller than in PCA.

5. Conclusion

In this paper, a standard robust incremental principal component analysis (NRIPCA) is proposed. The new method can reduce the robustness of real-time data by standardizing and pre-processing the data. Therefore, it is more suitable for the feature reduction of continuous casting production process parameters with multi-information source data characteristics. The experimental results prove the effectiveness of NRIPCA algorithm. The new method will provide a reliable source of data for the prediction of the roll gap parameters of the continuous casting and will be useful for establishing the prediction model for the parameters of the continuous casting. Further, we hope this feature reduction method can be applied to other industrial data area.

Acknowledgments

Corresponding author: Xianqi Liang:1012363162@qq.com;Zhufeng Lei: 253574602@qq.com; Wenbin Su: 809355146@qq.com

References

[1] ZHANG Guodong, SUN Jun. Application of Fuzzy Control on the Segment's Gap Control of Slab Continuous Casting Machine[J]. ELECTRIC DRIVE, 2009, 39, (05): 51-53.

[2]Fang Yiming, Hu Chunyang. Breakout Prediction Classifier for Continuous Casting Based on Active Learning GA-SVM[J]. China Mechanical.

[3]Carlos A Santos. A solidification heat transfer model and a neural network based algorithm applied to the continuous casting of steel billets and blooms[J]. Modelling and Simulation in Materials Science and Engineering. 2005, 10, (13): 1071-1087.

[4]Wang Nan, Dong Jianhong, Chen Min, YAO Yongkuan. Numerical Simulation on Thermal Stress Field in a Wide Slab Mould of Peritectic Steel Continuous Casting[J]. Journal of iron and steel research, 2012, (2): 909-912.

[5]Zhang Qunliang, Guo Zhaohui, Model control technologies of dynamic secondary cooling and soft reduction for slab continuous casting[J]. Baosteel technology research, 2012, 06, (3):61-64.

[6]WENG J Y, ZHANG Y L, HWANG W S. Candid Covariance — free Incremental Principal Component Analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(8): 1034-1040.

[7]JOSEPH A, TOKUMTO T, AZAWA S. Online Feature Extraction Based on Accelerated Kernel Principal Component Analysis for Data Stream[J]. Evolving Systems, 2015:1-13.

[8]Martins M, Santos C, Costa L, et al. Feature reduction with PCA/KPCA for gait classification with different assistive devices[J]. International Journal of Intelligent Computing and Cybernetics, 2015,8(4):363-382.

[9]Kaur M, Arora AS. Classification of ECG signals using LDA with factor analysis method as feature reduction technique[J]. Journal of Medical Engineering & Technology, 2012,3 6(8): 411-420.

[10]Wise B M, Gallagher N B. The Process Chemometrics Approach to Process Monitoring and Fault Detection[J]. Journal of Process Control (S0959—1524), 1996, 6(6): 329-348.

[11]Dung D, McAvoy T J. Nonlinear Principal Component Analysis Based on Principal Curves and Neural Networks[J]. Computers and Chemical Engineering(S0098—1354), 1996, 20(1): 65-78.

[12]You C K, Vanrolleghem P A, Lee I-B, Nonlinear Modeling and Adaptive Monitoring with Fuzzy and Multivariate Statistical Methods iu Biological Wastewater Trcatment Plants[J]. Journal of Biotechnology(S0168—1656), 2003, 105:135—163.

[13]Wang Xun, Kruger Uwe, Lennox Barry. Recursive Partial Least Squares Algorithms for Monitoring Complex Industrial Processes[J]. Control Engineering Practice(S0967-0661),2003，1l：613—632．

[14]Zhang Y，Dudzic M，Vaculik V. Integrated Monitoring Solution to Start — up and Run — time Operations for Continuous Casting[J]. Annual Review of Control(S1367—5788), 2003, 27: 141—149.

[15]Ivan Milefic, Shannon Quinn, Michael Dudzic, et al. An Industrial Perspective on Implementing On-line Applications of Multivariate Statistics[J]. Journal of Process Control(S0959-1524), 2004, 14: 821-836.

[16]L.J.P. van der Maaten, E.O. Postma, and H.J. van den Herik. Dimensionality Reduction: A Comparative Review. Tilburg University Technical Report, TiCC-TR 2009-005, 2009.

[17]Li W, Yue H, Valle — Cervantes S, Qin S. Recursive PCA for Adaptive Process Monitoring[J]. Journal of Process Control(S0959—1524), 2000, 10(5): 471—486.

[18]Wold S. Exponentially Weighted Moving Principal Components Analysis and Projections to Latent Structures[J]. Chemometrics and Intelligent Laboratory Systems(S0169—7439), 1994, 23: 149—161．

[19]L.J.P. van der Maaten and G.E. Hinton. Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research 9(Nov):2579-2605, 2008.

[20]N. Pezzotti, B.P.F. Lelieveldt, L.J.P. van der Maaten, T. Hollt, E. Eisemann, and A. Vilanova. Approximated and User Steerable t-SNE for Progressive Visual Analytics. IEEE Transactions on Visualization and Computer Graphics 23(7), 2017.

[21]L.J.P. van der Maaten. Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine Learning Research 15(Oct):3221-3245, 2014.

[22]Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. SCIENCE.5500(290): 2323-+,2000.

[23]Donoho DL, Grimes C. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data.PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA. 100(10): 5591-5596. 2003.

[24]Yao Chao, Liu Ya-Feng, Jiang Bo. LLE Score: A New Filter-Based Unsupervised Feature Selection Method Based on Nonlinear Manifold Embedding and Its Application to Image Recognition. IEEE TRANSACTIONS ON IMAGE PROCESSING.26(11):5257-5269.2017.

[25]Zhang ZY, Zha HY. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. SIAM JOURNAL ON SCIENTIFIC COMPUTING. 26(1): 313-338.2004.

[26]Zhang Tianhao, Yang Jie, Zhao Deli. Linear local tangent space alignment and application to face recognition.NEUROCOMPUTING. 70(7-9): 1547-1553. 2007.