Unsupervised Feature Transfer for Batch Process Based on Geodesic Flow Kernel

Zhang, Zheming; Wang, Fang; Pang, Yusong; Yan, Gaowei

DOI
10.1109/CCDC49329.2020.9164102

Publication date
2020

Document Version
Final published version

Published in
Proceedings of the 32nd Chinese Control and Decision Conference, CCDC 2020

Unsupervised Feature Transfer for Batch Process Based on Geodesic Flow Kernel

Zheming Zhang[1], Fang Wang[1], Yusong Pang[2], Gaowei Yan[1]

1. College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, Shanxi, China
E-mail: zhangzheming_tyut@163.com
E-mail: wangfang05@tyut.edu.cn
E-mail: yangaowei@tyut.edu.cn

2. College of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft 2628CD, Holland
E-mail: Y.Pang@tudelft.nl

Abstract: The problem of misalignment of the original measurement model is caused by nonlinear, time-varying characteristic of the batch process. In this paper, a method based on geodesic flow kernel (GFK) for feature transfer is proposed. By mapping data into the manifold space, the feature transfer from source domain to target domain is implemented. Distribution adaptation of real-time data and modeling data is performed to reduce the distribution difference between them. The historical data through distribution adaptation is used to establish a regression model to predict the real-time data, by which the unsupervised batch process soft sensor modeling is realized. The application of predicting the concentration of penicillin between different batches during the fermentation of penicillin demonstrated that the prediction accuracy of the model can be improved more effectively than the traditional soft sensor method.

Key words: Batch process, geodesic flow kernel, unsupervised, feature transfer, penicillin.

1 INTRODUCTION

At present, the difficulty in measuring key parameters during the batch process is mainly solved by multivariate statistical method based on measured data [1], such as Principal Component Analysis (PCA) and Partial Least Squares (PLS)-based algorithms [2,3]. However, when a large amount of process data is actually processed, there are often problems such as data drift [4], difficulties to obtain labels, and mismatch of the original model [5,6]. Multivariate statistical method is difficult to deal with such mixed dynamic characteristics.

Aiming at this problem, in [7], Artificial Neural Network (ANN) was applied to establish a soft sensor model of nonlinear process. However, its generalization ability cannot be guaranteed, so a well-trained model may lead to poor predictions of new observations. In [8], the author used Gaussian Mixture Regression (GMR) to establish multiple sub-models on historical data, and finally obtained the integrated regression model by weighting multiple sub-models according to the level of model output confidence. However, the output confidence of each sub-model is difficult to estimate for large structural risks. In [9,10], the author used the idea of Just-in-time learning (JITL), selecting the sample set that is most relevant to the current sample from the labeled historical data according to similarity metrics, and used machine learning methods to build a regression model to solve the problem of soft sensor for multiple working conditions. However, when the information of the current working condition is not included in historical data set, the established model cannot adapt the data of the current working condition, causing the model misalignment.

Transfer learning [11,12,13] used existing knowledge to solve the target domain problem by mining the shared features between domains, which introduces new ideas for the above-mentioned multimode soft sensor. In [14], the author introduced the semi-supervised domain adapted ELM algorithm to the soft sensor field of chemical processes. By using the source domain and a small number of labeled samples in the target domain, a mathematical model was constructed to realize the soft sensor of melt index in the process of industrial polyethylene under multi-working conditions. In fact, the problem of untagged samples in the target domain is common in the actual production process, and the semi-supervised algorithm is no longer applicable.

Aiming at the problem of unlabeled target domain, manifold-based unsupervised transfer learning [15,16] has become a research hotspot. Manifold learning maps data to a reliable embedded projection to find the data representation in a low-dimensional subspace [17]. It can map different working condition data to different points on the potential continuous manifold space. Compared with Euclidean space, it can better reflect the inherent characteristics and rules between sample data of different working conditions. In [18], the author proposed an unsupervised transfer learning method based on geodesic flow for cross-domain image classification, mapping the target domain and source domain data to two points on the Grassmann manifold space [19]. In the direction of the geodesics at these two points, several intermediate points are selected and connected in order to realize the gradual domain transfer from the source domain to the target domain via the geodesic. In [20], the author further proposed a kernel method based on Grassmann manifold

This work is supported by the Major Science and Technology Projects of Shanxi Province 20181102017. Key R&D Projects of Shanxi Province 201903D121143. General Program National Nature Science Foundation under Grant 61973226

space, and realized the continuous transfer process from the source domain to the target domain by integrating an infinite number of subspaces. The results showed that compared with the Euclidean space, domain transfer in the manifold space can find the inherent rules of data between different domains.

In this paper, the feature transfer based on manifold space is introduced into the unsupervised soft sensor of batch process in the geodesic flow kernel method, and the model error caused by the large difference of data distribution in penicillin fermentation process is eliminated. The experimental results show that the soft sensor model has good availability and high measurement accuracy.

2 Proposed Method

2.1 Subspace dimension measure

Suppose the source domain data X_s and the target domain data X_t. In order to improve the effect of feature transfer, the dimensionality of the subspace d needs to be determined to reduce the dimensionality of the data D in order to extract the main features .The protagonist concept is used [21], it can be defined as [20]:

$$\mathcal{D}(d) = 0.5\left[\sin\alpha_d + \sin\beta_d\right] \tag{1}$$

where α_d denotes the d-th principal angle between the PCA_S and PCA_{S+T} and β_d between PCA_T and PCA_{S+T}. $\sin\alpha_d$ or $\sin\beta_d$ is called the minimum correlation distance[22]. The optimal dimension can be obtained by formula (2) [20]:

$$d^* = \min\left\{d \middle| \mathcal{D}(d) = 1\right\} \tag{2}$$

2.2 Construct geodesic flow

Let $P_S, P_T \in \mathbb{R}^{D*d}$ denote the two sets of basis of the subspaces for the source and target domains. Let $R_S \in \mathbb{R}^{D*(D-d)}$ denote the orthogonal complement to P_S, namely $P_S^T P_S = 0$. The geodesic flow is parameterized as $\Phi : t \in [0,1] \to \Phi : t \in G(d,D)$, under the constraints $\Phi[0] = P_S, \Phi[1] = P_T$. For other t [20]:

$$\Phi(t) = P_S U_1 \Gamma(t) - R_S U_2 \Sigma(t) \tag{3}$$

where $U_1 \in \mathbb{R}^{d\times d}$ and $U_2 \in \mathbb{R}^{(D\text{-}d)\times d}$ are orthonormal matrices. They are given by the following pair of SVDs [20]:

$$P_S^T P^T = U_1 \Gamma(t) V^T, R_S^T P^T = -U_2 \Sigma(t) V^T \tag{4}$$

Γ and Σ are $d \times d$ diagonal matrices. The diagonal elements are $\cos\theta_i$ and $\sin\theta_i$ for $i = 1,2,\ldots,d$. Particularly, θ_i are called the principal angles between P_S and P_T :

$$0 \le \theta_1 \le \theta_2 \le \cdots \le \theta_d \le \pi/2 \tag{5}$$

Moreover, $\Gamma(t)$ and $\Sigma(t)$ are diagonal matrices whose elements are $\cos(t\theta_i)$ and $\sin(t\theta_i)$ respectively.

2.3 Compute geodesic flow kernel

Moving from the source domain to the target domain, the process of transfer from $\Phi(0)$ to $\Phi(1)$, the new feature can be expressed as:

$$z = g(x) = \Phi(t)^T x \tag{6}$$

The geodesic flow kernel is defined as [20]:

$$\left\langle z_i^\infty, z_j^\infty \right\rangle = \int_0^1 \left(\Phi(t)^T x_i\right)^T (\Phi(t)^T x_j) dt = x_i^T G x_j \tag{7}$$

where $G \in \mathbb{R}^{D\times D}$ is a positive semidefinite matrix, it can be calculated by equation [20]:

$$G = \left[P_S U_1 \ R_S U_2\right] \begin{bmatrix} \Lambda_1 & \Lambda_2 \\ \Lambda_2 & \Lambda_3 \end{bmatrix} \begin{bmatrix} U_1^T R_S^T \\ U_2^T R_S^T \end{bmatrix} \tag{8}$$

where $\Lambda_1, \Lambda_2, \Lambda_3$ are diagonal matrices, whose diagonal elements are:

$$\lambda_{1i} = 1 + \frac{\sin(2\theta_i)}{2\theta_i}, \lambda_{2i} = \frac{\cos(2\theta_i) - 1}{2\theta_i}, \lambda_{3i} = 1 - \frac{\sin(2\theta_i)}{2\theta_i} \tag{9}$$

A sample z that transfer the original sample features x along the geodesic direction can be obtained by:

$$z = \sqrt{G} x \tag{10}$$

Then the sample z_s mapped from x_s in the source domain can be obtained, so as z_t mapped from x_t in the target domain, and the existing labeled sample z_s can be modeled to predict the label of the sample z_t.

2.4 Unsupervised feature transfer based on GFK

This paper takes into account the difference in data distribution after batch changes and the potential associations between different batches [23], and a soft sensor modeling method based on transfer learning is introduced. The transfer from the source batch to the target batch is completed by GFK. The purpose of predicting the concentration of penicillin is achieved. Figure 1 is the schematic diagram of the method.

Multi-batch soft sensor based on geodesic flow kernel combines pre-processed known batches of labeled (source domain) samples X_s and unknown batches of unlabeled (target domain) samples X_t into overall data set $X = [X_s, X_t]$. The optimal dimension d^* is obtained by PCA_{X_s} and PCA_{X_t}, and source and target datasets are embed in a Grassmann manifold. PCA_{X_s} and PCA_{X_t} are used as the subspace P_S and P_T in the GFK framework. Considering P_S and P_T as two points in the higher dimensional space, a geodesic flow is constructed between the two points, and then integrate an infinite number of subspaces along the flow $\Phi(t)$. Concretely, raw features are projected into these subspaces to form an infinite-dimensional feature vector. Inner products between these feature vectors define a kernel function that can be computed over the original feature space in closed-form.

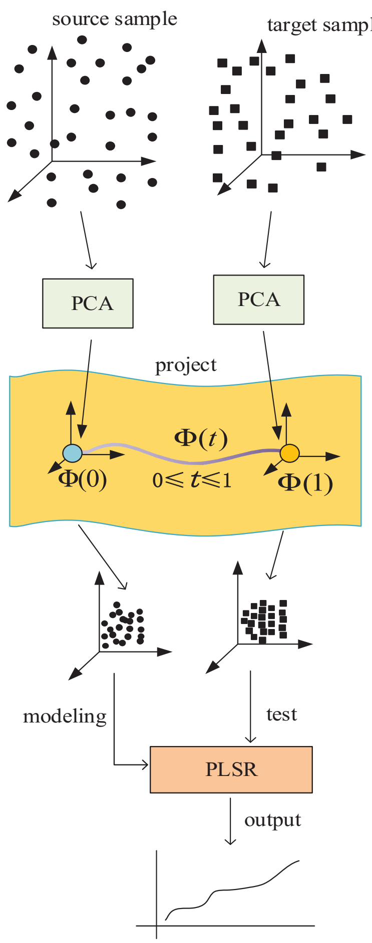

Figure 1. Schematic of the method

Thus, through equations (7), the geodesic flow kernel is formed. The task of feature transfer from source domain to target domain is completed by the geodesic flow.

Finally, a soft sensor model is established using the adapted source domain sample and the source domain label to achieve prediction of the target domain label. The algorithm of proposed method is shown in Table 1.

Table 1. The proposed method algorithm

Input: the source domain sample X_s , the target domain sample X_t , the source domain label Y_s .

Output: the target domain label Y_t .

1: Data preprocessing.

2: Calculating the optimal dimension d^* according to equations (1) and (2).

3: Constructing the geodesic flow $\Phi(t)$ by equation (3), getting the matrix G in the geodesic flow kernel according to equations (8) and (9), and obtaining the transferred data z_s and z_t with equation (10).

4: Using z_s and the source domain label Y_s to train a PLSR soft sensor regression model f .

5: Getting the target domain label Y_t based on f and z_t .

3 Experiments

Penicillin is the first large-scale clinically purified antibiotic used in humans. The penicillin fermentation process is a typical biochemical reaction process. It is a metabolic activity of penicillin-producing bacteria to grow and synthesize antibiotics under appropriate fermentation conditions [24,25].

In this paper, penicillin concentration, which was often analyzed offline during penicillin fermentation, was selected as the target variable. Table 2 lists process variables with high correlation as inputs to the soft sensor. Samples were collected every 0.5 hours during a 400-hour fermentation process, thus 800 samples were obtained, and the first five batches were selected as five different working conditions for transfer.

Table 2. Input variables for penicillin fermentation process

No.	Variable description	unit
1	Culture time	h
2	Aeration rate	L/h
3	Agitator power	W
4	Substrate feed rate	L/h
5	Substrate feed temperature	K
6	Substrate concentration	g/L
7	Dissolved oxygen concentration	g/L
8	Biomass concentration	g/L
9	Culture volume	L
10	Carbon dioxide concentration	g/L
11	pH	-
12	Fermenter temperature	K
13	Generated heat	kcal
14	Acid flow rate	L/h
15	Base flow rate	L/h
16	Cold water flow rate	L/h
17	Hot water flow rate	L/h

In order to quantify the prediction performance of various methods, root mean square error (RMSE) is used as the evaluation standard for measurement accuracy. The calculation formula is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{11}$$

where $\hat{y}_i$ and y_i represent the actual value and predicted value of the i-th sample, respectively, N is the number of test samples.

It is assumed that the known condition is the source domain and the condition to be measured is the target domain. In the experiment, PCA, PLS, ANN, GMR, JITL methods were used for comparison.

Figure 2 depicts the predicted result of penicillin concentration of each unsupervised method. It can be seen that under the same batch conditions, when the source batch and the target batch have a large difference in distribution, the accuracy of proposed method is higher compared with other methods. Figure 3 depicts the predicted result of GFK under the other batch conditions. The predicted value

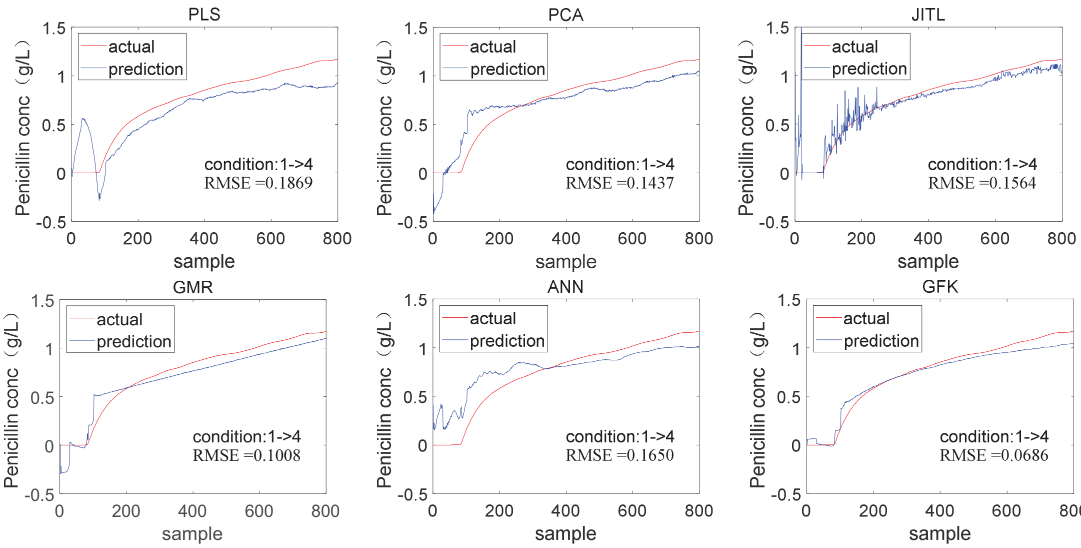

Figure 2. The prediction result of each algorithm under the conditions of batch 1-4

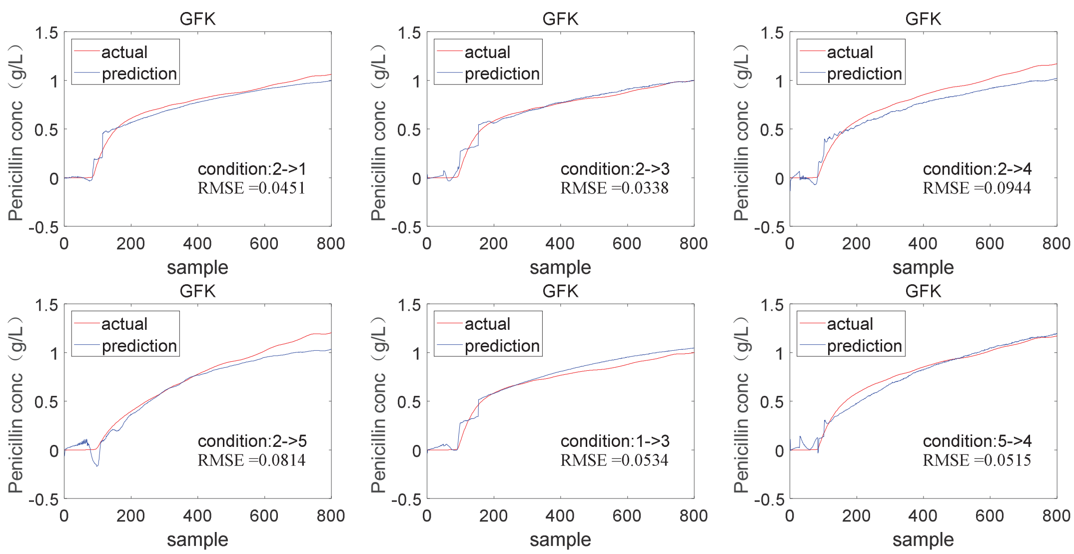

Figure 3. The predicted result of GFK under other batch conditions

can well track the measured value under all batch conditions, reflecting the advantages of proposed method. Table 3 describes the comparison result of penicillin concentrations predicted by different soft sensors under all batch conditions. The leftmost column “n→m” indicates the transfer from the nth batch to the mth batch. The last line represents the mean of RMSE of each algorithm. In order to compare the effects of different methods more intuitively, Figure 4 describes the evaluation accuracy of transfer to other batches when batches 1-5 are regarded as the source domain. Compared with the other prediction models, the proposed method achieved better predictive effect for the distribution adaptation of the source batch to the target batch by mapping the subspace to the manifold space for feature transfer.

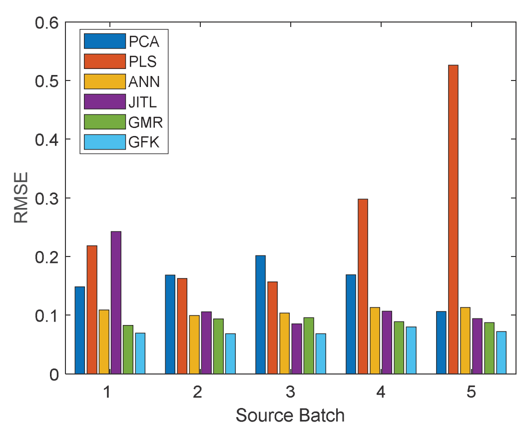

Figure 4. The average transfer result of source batches 1-5

Table 3. The comparison of root mean square error of different algorithms in each batch

Batch	PCA	PLS	ANN	JITL	GMR	GFK
1→2	0.0520	0.1392	**0.0373**	0.0568	0.0486	0.0469
1→3	0.0806	0.1871	0.1219	0.1205	0.0684	**0.0552**
1→4	0.1437	0.1869	0.1425	0.1564	0.1008	**0.0686**
1→5	0.3179	0.3603	0.1347	0.6369	0.1122	**0.1078**
2→1	0.0541	0.0836	0.0449	0.0536	0.0594	**0.0441**
2→3	0.1444	0.1247	0.0566	0.0701	0.0574	**0.0347**
2→4	0.1635	0.1862	0.1335	0.1131	0.1256	**0.0949**
2→5	0.3116	0.2553	0.1633	0.1874	0.1327	**0.0997**
3→1	0.0662	0.0672	0.0798	0.0815	0.0734	**0.0526**
3→2	0.0679	0.0568	0.0743	0.0569	0.0531	**0.0335**
3→4	0.2098	0.1711	0.1249	0.1104	0.1260	**0.0992**
3→5	0.4631	0.3321	0.1360	0.0921	0.1301	**0.0893**
4→1	0.1223	0.2196	0.0669	0.0821	0.0781	**0.0664**
4→2	0.1464	0.2563	0.1001	0.1126	0.1003	**0.0971**
4→3	0.1579	0.2721	0.1494	0.1216	0.1082	**0.1060**
4→5	0.2485	0.4444	0.1365	0.1121	0.0692	**0.0506**
5→1	0.0824	0.5533	0.1184	0.1047	0.0804	**0.0755**
5→2	0.0921	0.5949	0.1126	0.0989	0.0923	**0.0551**
5→3	0.1730	0.5637	0.1297	**0.0963**	0.0985	0.1067
5→4	0.0787	0.3935	0.0924	0.0769	0.0783	**0.0515**
Average	0.1588	0.2724	0.1078	0.1270	0.0897	**0.0717**

4 Conclusion

In order to address the difficult problem of misalignment of the original measurement model is caused by complexity of the batch process. The unsupervised soft sensor for batch process based on geodesic flow kernel is used in this paper. According to utilizing common features between multiple batches and extracting similar knowledge structures to the target batch in the source batch, the performance of unsupervised soft sensor is improved. The concentration prediction during the multi-batch penicillin fermentation process is applied. The practicability and effectiveness of the model is illustrated by comparing experiments and cross experiments.

REFERENCES

[1] J. Yu, A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses, Computers and Chemical Engineering, 41:134–144, 2012.

[2] J. F. Cerrillo, J. F. MacGregor, Multivariate monitoring of batch processes using batch-to-batch information, AIChE Journal, 50(01):1219-1228, 2004.

[3] J. Tang, L. Zhao, W. Yu, H. Yue, T. Chai, Soft Sensor Modeling of Ball Mill Load via Principal Component Analysis and Support Vector Machines, Advances in Neural Network Research and Applications, Springer Berlin Heidelberg, 803-810, 2010.

[4] X. Feng, L. J. Zhao, H. Y. Zheng, Ball Mill Load Modeling Method Based on KPCA Spectrum Feature Extraction, Journal of Shenyang University of Chemical Technology, 28(2):175-180, 2014.

[5] J. G. Wang, Z. Xie, Y. Yao, B. H. Yang, S. W. Ma, L. L. Liu, Soft sensor development for improving economic efficiency of the coke dry quenching process, Journal of Process Control, 77:20-28, 2019.

[6] P. Cao, X. Luo, Modeling of soft sensor for chemical process, Ciesc Journal, 64(3):788-800, 2013.

[7] J. C. B. Gonzaga, L.A.C. Meleiro, C. Kiang, R.M. Filho, ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process, Comput. Chem. Eng, 33:43–49, 2009.

[8] H. Jin, X. Chen, L. Wang, K. Yang, L. Wu, Adaptive Soft Sensor Development Based on Online Ensemble Gaussian Process Regression for Nonlinear Time-Varying Batch Processes, Industrial and Engineering Chemistry Research, 54(30):7320-7345, 2015.

[9] H. Jin, B. Pan, X. Chen, B. Qian, Ensemble just-in-time learning framework through evolutionary multi-objective optimization for soft sensor development of nonlinear industrial processes, Chemometrics and Intelligent Laboratory Systems, 184:153–166, 2019.

[10] H. Jin, X. Chen, L. Wang, K. Yang, L. Wu, Dual learning-based online ensemble regression approach for adaptive soft sensor modeling of nonlinear time-varying processes, Chemometrics and Intelligent Laboratory Systems, 151:228–244, 2016.

[11] S. J. Pan, Q. Yang, A Survey on Transfer Learning, IEEE, Transactions on Knowledge and Data Engineering, 22(10):1345-1359, 2010.

[12] M. He, J. Tang, X. Guo, G. Yan, Soft Sensor for Ball Mill Load Using RWNNRWNN Model, Acta Automatica Sinica, 45(2):398-406, 2019.

[13] K. Yan, D. Zhang, Calibration transfer and drift compensation of e-noses via coupled task learning, Sensors and Actuators B: Chemical, 225:288-297, 2016.

[14] Y. Liu, C. Yang, K. Liu, B. Chen, Y. Yao, Domain adaptation transfer learning soft sensor for product quality prediction, Chemometrics and Intelligent Laboratory Systems, 192:103813, 2019.

[15] J. D. Wang, W. Feng, Y. Chen, H. Yu, M. Huang, P. S. Yu, Visual Domain Adaptation with Manifold Embedded Distribution Alignment, Proceedings of the 26th ACM International Conference on Multimedia, 9:402-410, 2018.

[16] A. Shrivastava, S. Shekhar, V. M. Patel, Unsupervised domain adaptation using parallel transport on Grassmann manifold, Applications of Computer Vision, 277-284, 2014.

[17] C. R. Wang, Q. L. Zhang, X. D. Duan, Y. G. Wang, Z. D. Li. Research of Face Ethnic Features from Manifold Structure, Acta Automatica Sinica, 44(01):140-159, 2018.

[18] R. Gopalan, R. Li, R. Chellappa, Domain adaptation for object recognition: An unsupervised approach, IEEE, International Conference on Computer Vision, 999-1006, 2011.

[19] J. Zhou, Morse functions on Grassmann manifolds, Proceedings of the Royal Society of Edinburgh: Section A Mathematics, 135(01):209-221, 2005.

[20] B. Q. Gong, Y. Shi, F. Sha, K. Grauman, Geodesic flow kernel for unsupervised domain adaptation, Conference on Computer Vision and Pattern Recognition, IEEE, 2066–2073, 2012.

[21] J. Huang, Q. Qiu, R. Calderbank, The Role of Principal Angles in Subspace Classification, IEEE, Transactions on Signal Processing, 64(8):1933–1945, 2016.

[22] J. Hamm and D. Lee, Grassmann discriminant analysis: a unifying view on subspace-based learning, In Proc. of ICML, 376–383, 2008.

[23] S. Sun, H. Shi, Y. Wu, A survey of multi-source domain adaptation, Information Fusion, 24:84-92, 2015.

[24] Y. Liu, H. Q. Wang, Pensim Simulator and Its Application in Penicillin Fermentation Process, Journal of System Simulation, 18(12):3524-3527, 2006.

[25] C. Zhai, T. Qiu, A. Palazoglu, W. Sun, The emergence of feedforward periodicity for the fed-batch penicillin fermentation process, IFAC Papers OnLine, 51(32):130–135, 2018.