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
Accepted author manuscript

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

Unsupervised Soft Sensor 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 problems of nonlinear, time-varying, and multi-batch data distribution differences among batches during the batch process, label samples are difficult to obtain, and the original measurement model is inaccurate. In this paper, we use geodesic flow kernel (GFK) for feature transfer. By mapping data into the manifold space, the feature transfer from source domain to target domain is implemented. Perform distribution adaptation of real-time data and modeling data to reduce the distribution difference of data between them. Then use the historical data after the distribution adaptation to establish a regression model, and predict the real-time data after the distribution adaptation, to realize the unsupervised batch process soft sensor modeling, so as to improve the batch process soft sensor model accuracy. By predicting the concentration of penicillin between different batches during the fermentation of penicillin, it is verified that under the same conditions, the prediction accuracy of the model can be improved more effectively than the traditional soft sensor and deal with soft sensor problem in multi-batch process

Key words: Batch process, geodesic flow kernel, unsupervised, soft sensor, penicillin.

1 INTRODUCTION

At present, methods for batch processes are mainly multivariate statistical methods based on measured data [1], such as algorithms based on Principal Component Analysis (PCA) and Partial Least Squares (PLS) [2,3,4,5]. However, when a large amount of process data is actually processed, there are often problems such as data drift [6], difficult to obtain labels, and mismatch of the original model [7,8]. Multivariate statistical methods are difficult to deal with such mixed dynamic characteristics.

In response to this problem, in [9], Artificial Neural Network (ANN) is 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 [10], this paper uses Gaussian Mixture Regression (GMR) to establish multiple sub-models on historical data, evaluate the soft sensor results of each sub-model, weight multiple fusions based on the level of model output confidence, and finally obtain the integrated regression model. However, the output confidence of each sub-model is difficult to estimate, and there are large structural risks. In [11,12], this paper based on the idea of Just-in-time learning (JITL), select the sample set that is most relevant to the current sample from the labeled historical data according to similarity metrics, and use machine learning methods to build a regression model to handle multi-working conditions soft sensor. However, when the data of the current working conditions are lacking in the historical data set, the established model cannot be adapted to the data of the current working conditions, causing the model to be inaccurate.

Transfer learning [13,14,15] uses existing knowledge to solve the target domain problem by mining the shared features between domains, and introduces new ideas for the above-mentioned multi-modal soft sensor. In [16], this paper 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 is constructed to realize the soft sensor of melt index in the process of industrial polyethylene under multi-working conditions. But the semi-supervised learning method requires a small amount of labeled data in the target domain. However, in the actual production process, the problem of untagged samples in the target domain is common, and the semi-supervised algorithm is no longer applicable.

Aiming at the problem of unlabeled target domain, manifold-based unsupervised transfer learning [17,18] has become a research hotspot. Manifold learning maps data to a reliable embedded projection, that is, to find the data projected into a low-dimensional subspace representation [19]. Manifold learning 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 different working condition sample data. In [20], this paper 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 [21]. In the direction of the geodesic of these two

This work is supported by the General Program National Nature Science Foundation under Grant 61973226, Major Science and Technology Projects of Shanxi Province 20181102017.

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 [22], this paper introduced a kernel method on this basis, and realized the continuous transfer process from the source domain to the target domain by integrating an infinite number of subspaces. And achieved higher accuracy in cross-domain image classification, the results further show that compared with the Euclidean space, domain transfer in the manifold space can find the inherent rules of data between different domains.

To sum up, in this paper, the characteristic transformation 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 dealt with. The experimental results show that the soft sensor model has good adaptability 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 transformation, the dimensionality of the subspace needs to be determined to reduce the dimensionality of the data in order to extract the main features .We use the protagonist concept [23], it can be defined as [22]

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

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[24]. The optimal dimension can be obtained by formula (2) [22]

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

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. D is the dimensionality of the data. 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 Φ: $t \in [0,1] \to \Phi : t \in G(d, D)$, under the constraints $\Phi[0] = P_S, \Phi[1] = P_T$. For other t [22]

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

where $U_1 \in \mathbb{R}^{d\times d}$ and $U_2 \in \mathbb{R}^{(D-d)\times d}$ are orthonormal matrices. They are given by the following pair of SVDs [22]

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

$\Gamma \in d \times d$ and $\Gamma \in d \times d$ are 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 [22]

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

where $G \in \mathbb{R}^{D\times D}$ is a positive semidefinite matrix, it can be calculated by equation [22]

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

where Λ_1, Λ_2, Λ_3 are diagonal matrices, whose diagonal elements are

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

We can get a sample z that transfer the original sample features x along the geodesic direction

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

Then, the sample Z_s after X_s mapping and the sample Z_t after X_t mapping can be obtained, and the existing label sample Z_s in the source domain can be learned and modeled to realize the prediction of the sample Z_t label.

2.4 Unsupervised soft sensor based on GFK

On the problem of multi-batch unsupervised soft sensor modeling, this paper takes into account the difference in data distribution after batch changes and the potential associations between different batches [25], and introduces a manifold-based transfer learning method. Using the characteristics of the GFK framework to continuously transfer along the geodesic in the manifold space, the transfer from the source batch to the target batch was completed, and the purpose of predicting the concentration of penicillin was achieved. Figure 1 is a schematic diagram of the method.

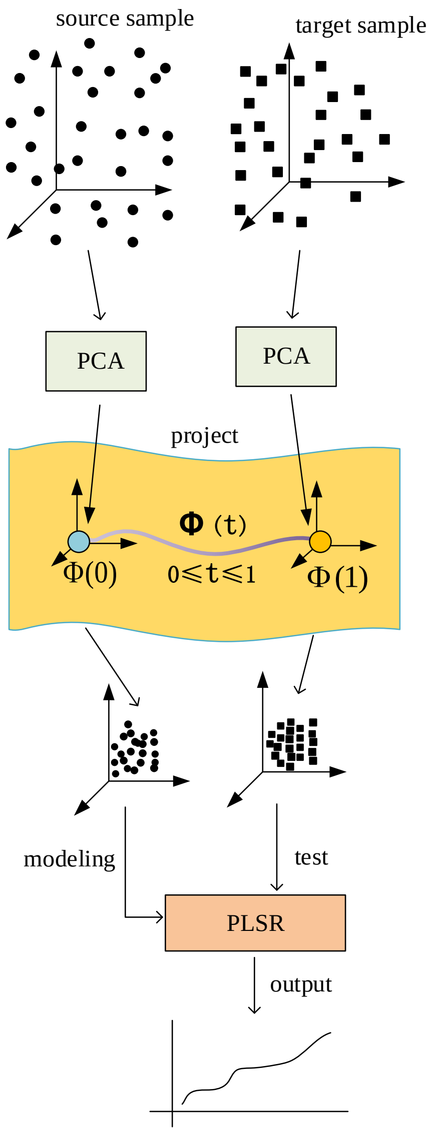

Figure 1 Schematic of the method

Combining Figure 1 and the description of related theories and algorithms, 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 from the angle between PCA_{X_s} and PCA_X and between PCA_{X_t} and PCA_X. Subsequently, map the reduced dimension PCA_{X_s} and PCA_{X_t} to the Grassmann manifold space, and it is used as a subspace P_S and P_T in the GFK framework, and combined with equation (3) to construct the geodesic equation, and then obtained from equations (8) and (9) the geodesic flow kernel is then used to obtain, the distribution-adaptive data through equation (10).

Finally, a soft sensor model is established using the adapted source domain samples and source domain labels to achieve prediction of the target domain labels. The algorithm flowchart of this method is shown in Table 1

Table 1 Algorithm flowchart

Input: source domain samples X_s, target domain sample X_t, label data Y_s for source domains.

Output: target domain label Y_t.

1: Data preprocessing.

2: Calculate the optimal dimension d according to equations (1) and (2).

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

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

5: Find 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 fermentation process is a typical biochemical reaction process. The penicillin fermentation process is a metabolic activity of penicillin-producing bacteria to grow and synthesize antibiotics under appropriate fermentation conditions [26,27].

In this paper, penicillin concentration that is often analyzed off-line during penicillin fermentation is selected as the target variable. Table 2 lists process variables with high correlation as inputs to the soft sensor. The data of the 400-hour fermentation process were selected, samples were collected every 0.5 hours, a total of 800 samples, 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

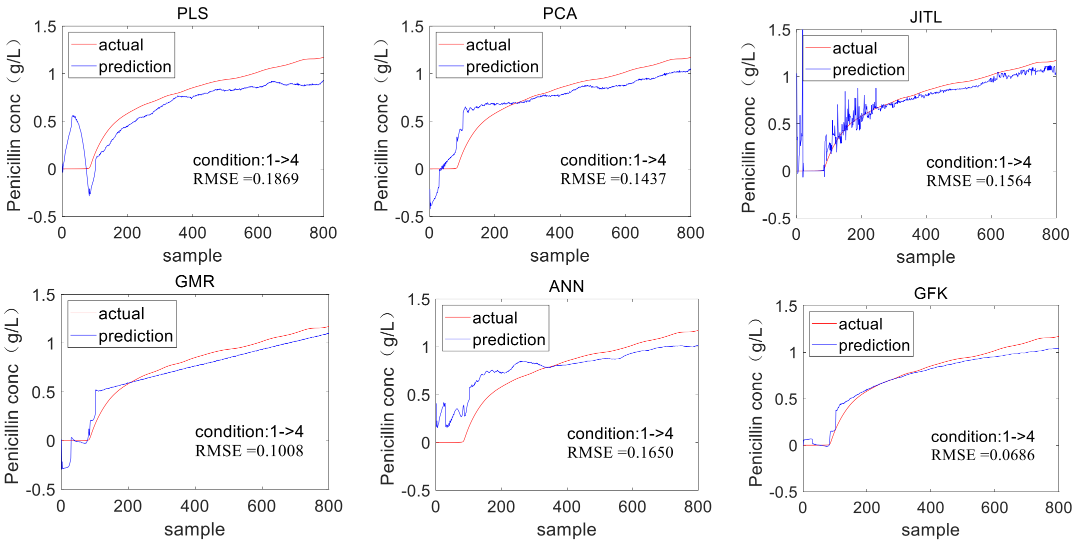

Figure 2 The prediction results of each algorithm under the conditions of batch 1-4

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, multiple algorithms were used to predict and compare the substrate concentration. This paper uses PCA, PLS, ANN, GMR, JITL methods for comparison.

Figure 2 depicts the comparison of the predicted results of penicillin concentration by each unsupervised method. It can be seen from the figure that under the same batch conditions, when the source batch and the target batch have a large difference in distribution, the accuracy of this method is improved to different degrees compared with other methods. It can be seen that the blue curve (predicted value) in the figure can better track the red curve (real value), which reflects the advantages of this method.

Table 3 describes the comparison results of penicillin concentrations predicted by different soft sensor under all batch conditions. The leftmost column “n→m” indicates transfer from batch n to batch m. The bottom line represents the mean of root mean square error of each algorithm. It can be seen that when using the PCA, PLS, ANN, GMR and

Table 3 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**

JITL, the prediction results are not ideal, and there are different degrees of accuracy degradation under different batches. Compared with other prediction models, GFK achieves the distribution adaptation of the source batch to the target batch by mapping the subspace to the manifold space for feature transformation, and taking into account the problem of feature differences, most prediction values have achieved better predictive effect.

4 Conclusion

This paper uses an unsupervised soft sensor for batch process based on geodesic flow kernel method to mine and utilize features common between multiple batches and extract knowledge structures similar to the target batch in the source batch to improve unsupervised soft sensor performance. In order to verify the validity of the method, it was applied to the soft sensor of concentration prediction during the multi-batch penicillin fermentation process. The multi-batch soft sensor modeling was completed. Experimental results show that the method used in this paper can effectively improve the prediction accuracy of the model.

REFERENCES

[1] J. Yu, A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses[J]. 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[J]. AIChE Journal, 50(01):1219-1228, 2004.

[3] C. Duchesne, J. F. MacGregor, Multivariate analysis and optimization of process variable trajectories for batch processes[J].Chemometrics and Intelligent Laboratory Systems，51：125-137, 2000.

[4] L. Zhao, J. Tang, W. Yu, Modeling of mill load for wet ball mill via GA and SVM based on spectral feature[C] IEEE Fifth International Conference on Bio-inspired Computing: Theories and Applications. IEEE,874-879, 2010.

[5] 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[M]. Advances in Neural Network Research and Applications. Springer Berlin Heidelberg, 2010.

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

[7] 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[J]. Journal of Process Control, 77: 20-28, 2019.

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

[9] 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[J]. Comput. Chem. Eng. 33, 43–49. 2009.

[10] 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[J]. Industrial and Engineering Chemistry Research, 54(30):7320-7345, 2015.

[11] 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[J]. Chemometrics and Intelligent Laboratory Systems, 184, 153–166, 2019.

[12] 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[J]. Chemometrics and Intelligent Laboratory Systems, 151, 228–244, 2016.

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

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

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

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

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

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

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

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

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

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

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

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

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

[26] Y. Liu, H. Q. Wang, Pensim Simulator and Its Application in Penicillin Fermentation Process[J]. Journal of System Simulation, 18.12, 2006.

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