

Research on Online Multiple Model Soft-sensor

Shijie Wang^a, Zhenlei Wang^{a,*}, Xin Wang^b

^aKey Laboratory of Advanced Control and Optimization for Chemical Processes, East China University of Science and Technology, Shanghai 200237, China

^bCenter of Electrical & Electronic Technology, Shanghai Jiao Tong University, Shanghai 200240, China
wangzhen_l@ecust.edu.cn

Offline updating is a method that most of multiple model soft-sensors used to adapt the new operating conditions. Replacing online models with offline ones is bound to affect the efficiency of soft-sensors, and it costs manpower as well as time simultaneously. It takes maintenance staffs some time to re-train complete models, which requires a lot of historical data, and then the existing models will be changed with new ones. A soft-sensor that can be added or subtracted models online is proposed in this paper. Density-based spatial clustering of applications with noise (DBSCAN) is employed for clustering analysis. Compared with traditional kernel fuzzy clustering method (KFCM), DBSCAN improves the ability of filtering out noise and enhance the ability to decide whether there is a new working condition. However, the clustering results of DBSCAN are extremely sensitive to the input parameters. In this study, kernel density estimation (KDE) is applied to determine the number of subsets and a novel method is proposed to determine the parameters. The new sub-models can be directly added to the online models after trained. The results of soft-sensor achieved by a number of models according to the switching or weighted way. The method proposed in this paper is applied to the measurement of cracking depth of ethylene cracking furnace, which proves the practicability and effectiveness.

1. Introduction

It is essential for the process industry to achieve real-time estimation and control of key variables, which extremely relate to safety and efficiency. In actual plant, due to feed components, operating conditions and other changes, same process often contains multiple operating points. In order to predict key variables in these conditions, multiple model soft-sensor is proposed (Chu et al., 2009). It takes advantage of establishing multiple sub-models to cover the data characteristics of each operating point and the clustering algorithm is applied to divide the data subsets. Yang et al. (2012) proposed a novel soft sensor using multi-model neural network (MNN) based on modified kernel fuzzy clustering (KFCM) to release a single data-based soft sensor from suffering heavy burden calculation and poor accuracy. Tang et al. (2014) applied Dempster-Shafer rule (D-S) and least squares support vector machine (LSSVM) to improve the accuracy of weighted multiple soft-sensor models. For time-varying systems, a multi-mode moving-window Gaussian process regression (MWGPR) based approach for ARX modelling is used to capture process nonlinearity or switching dynamics (Xiong et al., 2016).

However, in the above methods, whether it is offline or online, the original model structure need be changed, rather than do some increase or decrease on the basis of the original part when model updates. Soft-sensors based on adaptive methods can update models with recently collected data, but the new models established are just local models, which only valid for current operating conditions. If a soft-sensor is asked for containing all the characteristics, old models deserve to be preserved. The re-clustering requires a lot of historical data, which increases the difficulty of on-site maintenance. It is also difficult to pick out the corresponding data in the database when a feature is not needed in current models.

To solve this problem, this study proposes a novel multiple soft-sensor method based on density clustering. With the help of kernel density estimation (KDE), new characteristics of inputs can be found. New sub-sets are collected by density-based spatial clustering of applications with noise (DBSCAN). New sub-models established by LSSVM are directly attached to the currently running soft-sensor structure without stopping the current models. Sub-models are also can be deleted according to the demand. This method avoids overturn the original

model, which greatly reducing the complexity of model training and extremely retaining effective historical characteristics.

2. Density-based spatial clustering of applications with noise and kernel density estimation

2.1 Kernel density estimation

Kernel density estimation is a nonparametric test method, commonly used to estimate the unknown density function. There is no need to make any assumptions about the data segment in advance. The kernel density estimation method is a method of studying the distribution characteristics of the data set entirely from the data samples. This method is applied in this study to analysis the distribution of unknown inputs and the results can facilitate the DBSCAN algorithm to determine the number of subsets.

Set x_1, x_2, \dots, x_n as independent random variables of the same distribution in data set D^m (m Dimensions), The relative density function $f(x)$ is defined as:

$$f_h(x) = \frac{1}{n} \sum_{i=1}^n K_{m,h}(x - x_i) \quad (1)$$

Where, $K_{m,h}(x) = \prod_{i=1}^m K(x_i / h_i) / h_i$ and $K(\cdot)$ is the kernel function, such as Gaussian function. h is called bandwidth (BW), which has great effect on the results of KDE. In other words, when h increases, some of the features be filtered out by the density estimation, so that the final result is a large cluster. In contrast, when it decreases, many small clusters will be highlighted, but they may not be some effective peaks that contain valid features.

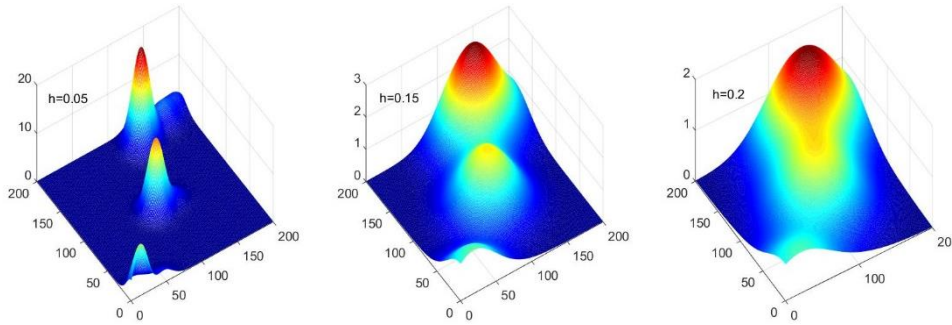


Figure 1: Density distribution based on KDE

Figure 1 shows the density estimation of a data set. When h is 0.15, a reasonable distribution occurs. If h comes to 0.2, some features are fused. In contrast, when h reduced to 0.05, some features become particularly prominent. Thus, as long as there are some appropriate peaks, the current distribution of the data set can be determined easily.

2.2 Density-based spatial clustering of applications with noise

DBSCAN is a density-based clustering method, the main idea is: In the neighbourhood Eps , the number of objects in each object cluster must equal to or more than the given value $MinPts$. Eps is a parameter similar to radius, mainly to limit the size of an object cluster. $MinPts$ is a parameter that represents the smallest number of objects in the spherical area which the object cluster as the centre point and Eps as the radius the distance formula is applied to calculate the number of objects including in the neighbourhood showed as follow:

$$D(p, q) = \sqrt{|p_1 - q_1|^2 + |p_2 - q_2|^2 + \dots + |p_m - q_m|^2} \quad (2)$$

According to different number of objects contained in the neighbourhood, the result is divided into three categories. One is the core point, the number of objects in the neighbourhood is greater than or equal to $MinPts$, which usually plays as an important role in the cluster. It will connect to other core points and boundary points in the surrounding density to constitute the same cluster. Other is the boundary point, the number of objects within the neighbourhood is less than $MinPts$, but there is at least a core point in it, which works as a relevant member attached to the core point around. If no core point can be found in the neighbourhood, it does not belong to any cluster and regarded as noise.

It is a long-running question to set appropriate *MinPts* and *Eps*. In particular, *Eps* will directly affect the distribution of the core points, so minor changes may lead to large different clustering results. Based on part 2.1, KDE is helpful to solve the problem. In this study, distribution of inputs need be acknowledged at first to determine the number of clusters, and then *Eps* can be approximated by dichotomy. The specific methods are as follows:

Step1: Set x_1, x_2, \dots, x_n as independent random variables of the same distribution in data set D^m (m Dimensions),

Calculate the centre point of the D^m by the distance formula, and then maximum distance in D^m called *DMax*, the minimum distance in D^m called *DMin* can be achieved. Set the number of clusters C according to KDE. Set an initial value to *MinPts* between 50 with 100. Next step.

Step2: Calculate $Eps = (DMax - DMin) / 2$. Then carry out $DBSCAN(D^m, Eps, MinPts)$ to get real number of clusters \hat{C} . Next step.

Step3: If $\hat{C} > C$, carry out $DMin = Eps$. If $\hat{C} < C$, carry out $DMax = Eps$. If $\hat{C} = C$, go to Step4, otherwise go back to Step2.

Step4: Adjust *MinPts* appropriately obeys to actual results. If it is increased, the core points will be reduced, but more noise points can be kicked out.

Using the above method, subsets contain new characteristics are collected. This kind of screening method is only carried out in the prediction data with errors and retains the original models. So, there is no need to store data from original models. Compared with other clustering methods, it is possible to greatly reduce the number of training data when clustering, which eliminating unnecessary duplication and reducing time. As a result of applying KDE, new characteristics of current operating conditions are demonstrated in detail. However, conventional clustering methods lack approaches to determine the number of clustering centres, which will affect the actual results of clustering and soft-sensor.

3. Online multiple model soft-sensor

Support vector machine is a small sample machine learning method that can be used for classifier design and numerical regression. In this study, subsets representing different characteristics of the operating conditions, are selected by DBSCAN clustering. The sub-models are established by LSSVM. The predicting results of sub-models combined with switching or weighted way act as the final output of these sub-models, also as the output of soft-sensor.

After the sub-models are established, the soft-sensors output can be obtained by the following formula:

$$U_{ik} = \frac{[2 - 2K(x_k, v_i)]^{-1/(m-1)}}{\sum_{j=1}^c [2 - 2K(x_k, v_j)]^{-1/(m-1)}} \quad (3)$$

Where U_{ik} is the membership of the k -th input data for the i -th clustering centre and C is the number of clustering centres. If choosing switching way to combined multiple models, U_{ik} is be replaced by A_k :

$$A_k = i \quad \max(U_{ik}) \quad (4)$$

Therefore, the formula of output of multiple model soft-sensor showed as follow:

$$\begin{aligned} Output_k &= \sum_{i=1}^c U_{ik} Lssvm_i \quad \text{weighted} \\ Output_k &= Lssvm_{A_k} \quad \text{switching} \end{aligned} \quad (5)$$

The structure of online multiple model soft-sensor is demonstrated in Figure 2. The process in the red box shows the part of selection of subsets and establishment of sub-models which mentioned above. When new data come, their membership will be calculated based on existing sub-models. Results are the values mixed by all the sub-models. Those estimation results with bias are stored in the database and wait for new clustering analysis. The model in the dashed box represents the portion that can be added or deleted online. When number of training data in database reach the threshold or manual instruction has been received, new clustering subset is collected by DBSCAN and KDE. If the sub-models set are changed, the corresponding membership will change at the same time.

The model structure can be implemented by the following steps:

Step1: Determine whether to increase a sub-model. To add models, carry out Step 2. To delete models, carry out Step4. No action, re-execute Step1. Set a training data set to collect data whose prediction contains bias.

Step 2: Export data from training data set, remove the coarse error, calculate KDE, carry out DBSCAN clustering and train sub-models by LSSVM. If the accuracy of sub-models is satisfied, carry Step 3. Otherwise, re-execute Step 2 or back to Step 1.

Step 3: Adding or deleting some sub-models, recalculate the new sub-models' membership. The next predict data will using the updated models and membership. Back to Step 1.

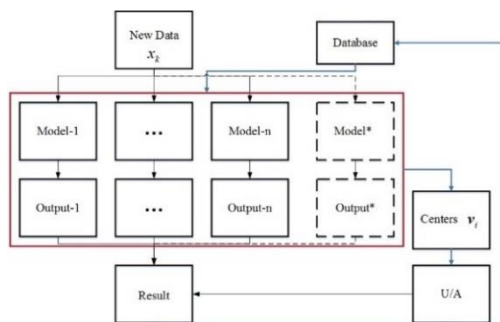


Figure 2: The structure of online multiple model soft-sensor

4. Case study

4.1 Description of depth estimation of ethylene cracking furnace

Ethylene cracking depth is an important indicator of the degree of ethylene cleavage, directly related to the yield of the product. Ethylene cracking process is the first step in the ethylene production process, which splitting raw materials into ethylene, propylene and other mixed products. The cracking reaction device is ethylene cracking furnace, as shown in Figure 3.

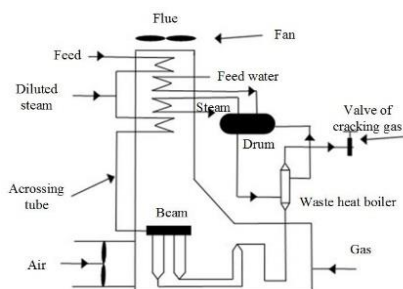


Figure 3: Structure diagram of tube cracking furnace

Cracking depth is an indicator, which usually expressed by the ratio of the yield of propylene and ethylene. Real-time estimation to the value of cracking depth is conducive to the operators to master the actual operation condition of cracking furnace, but also benefit the real-time control and process optimization. But the main problems faced by soft-sensor are changes in inputs, process parameters, etc. Some unpredictable factors, such as furnace tube coke, noise interference is also disadvantage for measurement. When faced with multiple operating conditions, the structure of traditional soft-sensor will become complex and its accuracy and generalization ability will be reduced. Problems, such as huge amounts of history data, complex calculation, prohibiting to shut down model when upload, etc., will increase the difficult for maintenance when updating models. The method proposed in this study is used to solve these problems as well as compare with the results of the single model LSSVM and KFCM-LSSVM.

4.2 The measurement of cracking depth of ethylene cracking furnace

The partial process data of an ethylene cracking furnace were selected and analysed. The feed of the cracking furnace is changeable, such as naphtha (NAP), liquefied petroleum gas (LPG) and mixed feed GAS+NAP, and the corresponding operating conditions are also changed. In this case, the single model LSSVM, KFCM-LSSVM and KDE-DBSCAN-LSSVM were used to estimate the depth of the crack respectively. The results are shown in Table 1 and Figure 4.

The first row in Table 1 is the results by using LSSVM. In the case of operating condition changes, the original model is poorly estimated for the current operating condition. So the models in the stage of condition change need to be re-trained. As each training is for a certain working conditions, so the corresponding MSE and COR are relatively high. The KFCM-LSSVM retrains the sub-models each time a new condition occurs. Compared to the single model LSSVM, it can effectively predict data which similar to data in subsets, that is, to record some of the historical data models. But updating needs to re-cluster all the data, and training sets become increasingly larger with the passage of time. It will cost more time and lager computer resource. Also, there are two different NAP feed conditions in actual process, it cannot be effectively identified by KFCM-LSSVM, resulting in the decline of model accuracy. In this study, the KDE-DBSCAN-LSSVM retains the original models and use the current data to train new sub-models when the working condition changes. As demonstrated in Figure 4, each of sub-models are stored by system, and a new sub-model is added in when inputs contain new characteristic occur. When new NAP data comes, the original NAP model can be deleted or invalidated (the membership maintains 0), then a new sub-model is added. Although the total number of support vectors is large, but each sub-model only need to contain part of them. This method can be regarded as this as a collection of knowledge, with the passage of time, the model structure has been improved and becoming integrated, nearing to the whole conditions of the real process.

Table 1: Results of soft-sensors

	NAP				LPG				GAS+NAP				NAP			
	MSE	COR	SVs	FL	MSE	COR	SVs	FL	MSE	COR	SVs	FL	MSE	COR	SVs	FL
Single	0.0010	97.03	392	Y	0.0164	99.61	249	Y	0.0255	96.90	179	Y	0.0005	89.94	505	Y
KFCM	0.0010	97.03	392	Y	0.0202	99.40	280	Y	0.0101	99.48	425	Y	0.0035	50.92	425	Y
KDE-DBSCAN	0.0010	97.03	392	Y	0.0217	99.29	669	P	0.0008	99.66	776	P	0.0005	90.70	1292	P

Where MSE is the mean square error, COR is correlation, SVs is the total number of support vectors contained in the model and FL is the flag represent the way model requires update, Y indicates that the model structure is completely updated, N is no update, and P is update partially.

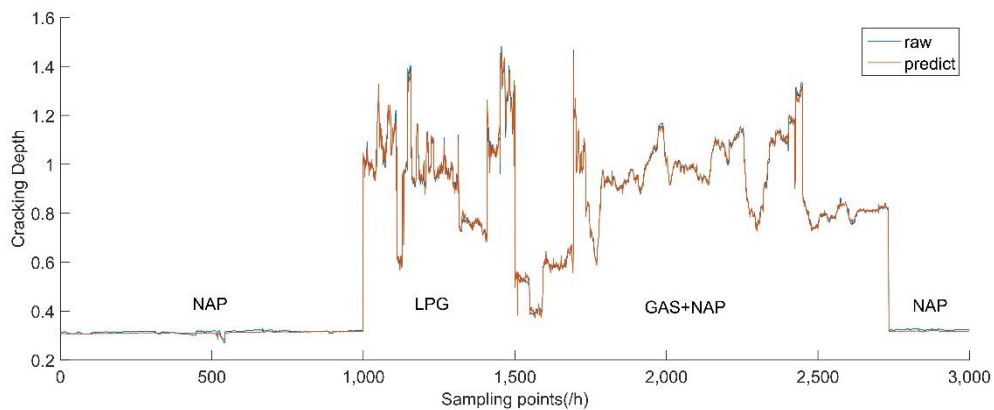


Figure 4: Result of KDE-DBSCAN-LSSVM

5. Conclusion

In this study, multiple model soft-sensor based on online increment and subtraction structure is proposed. Compared with the single model soft-sensor, it simplifies the structure of each sub-models, and only needs to change some sub-models, rather than completely change the original model structure when update. It is conducive to online maintenance and the impact of the current operating condition of the scene is also smaller. Compared with the traditional multiple model soft-sensor, DBSCAN plays a noise reduction and finds new characteristics from new process data. Traditional KFCM need to re-clustering all the data, but this method only need to cluster part of the data, which significantly simplifying computing complexity. Thus, this method can be applied in the process with multi-mode operating conditions. It is flexible for operators to make rapid and effective changes, which ensure the stability and accuracy of soft-sensors. Although industrial data show that this method is feasible, theoretical proof is still lacking and it will be discussed in future study.

Acknowledgments

This work was supported by National Key Technology Support Program (2015BAF22B02); the National Natural Science Foundation of China (61403141, 21406061). This work was supported by the Shanghai Natural Science Foundation of China (14ZR1421800) and State Key Laboratory of Synthetical Automation for Process Industries (PAL-N201404).

References

- Jin H.P., Chen X.G., Yang K., Wang L., 2015, Multi-model adaptive soft sensor modeling method using local learning and online support vector regression for nonlinear time-variant batch processes. *Chemical Engineering Science*, 131: 282-303.
- Li X.L., Su H.Y., Chu J., 2009, Multiple model soft sensor based on affinity propagation, Gaussian process and Bayesian committee machine. *Chinese Journal of Chemical Engineering*, 17(1): 95-99.
- Tang K., Wang X., Wang Z.L., 2014, Multi-model soft sensor based on Dempster-Shafer rule. *Control Theory & Applications*, 31(5): 632-637.
- Wang L., Chen X.G., Yang K., Jin H.P., 2017, Soft sensor modeling based on variable partition ensemble method for nonlinear batch processes. 7th International Conference on Electronics and Information Engineering. *International Society for Optics and Photonics*: 103222E-103222E-8.
- Wang Z.L., Tang K., Wang X., 2014, A multi-model soft sensing method based on DS and ARIMA model. *Control and Decision*, 29(7): 1160-1166.
- Xiong W.L., Zhang W., Xu B.G., Huang B., 2016, JITL based MWGPR soft sensor for multi-mode process with dual-updating strategy. *Computers & Chemical Engineering*, 90: 260-267.
- Zhang W.Q., Fu Y.J., Yang H.Z., 2012, Multi-model soft-sensor modeling based on improved clustering and weighted bagging. *CIESC Journal*, 9: 005.
- Zhang W., Xiong W.L., Xu B.G., 2015, Multi-model combination modeling based on just-in-time learning using Gaussian process regression. *Information and Control*, 44(4): 487-492.
- Zhou H.F., Wang P., Li H.Y., 2012, Research on adaptive parameters determination in DBSCAN algorithm. *Journal of Information & Computational Science*, 9(7): 1967-1973.