

Multiple Model Soft Sensor Based on Affinity Propagation, Gaussian Process and Bayesian Committee Machine^{*}

LI Xiuliang (李修亮), SU Hongye (苏宏业)** and CHU Jian (褚健)

National Key Laboratory of Industrial Control Technology, Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou 310027, China

Abstract Presented is a multiple model soft sensing method based on Affinity Propagation (AP), Gaussian process (GP) and Bayesian committee machine (BCM). AP clustering arithmetic is used to cluster training samples according to their operating points. Then, the sub-models are estimated by Gaussian Process Regression (GPR). Finally, in order to get a global probabilistic prediction, Bayesian committee machine is used to combine the outputs of the sub-estimators. The proposed method has been applied to predict the light naphtha end point in hydrocracker fractionators. Practical applications indicate that it is useful for the online prediction of quality monitoring in chemical processes.

Keywords multiple model, soft sensor, affinity propagation, Gaussian process, Bayesian committee machine

1 INTRODUCTION

Soft sensors are prerequisite to the immeasurable quality indicator monitoring and control in chemical processes. The operating variables are held at the set point nearby when chemical process operating is at normal production state. Occasionally, the production needs switch work mode (or named operating point). In such case, the historical samples are clustered around different operating points. The usual approach in order to get a clear description of the production process character is to collect historical samples as many as possible. But samples near different work modes display dissimilar properties in disturbance range and nonlinear properties. So the single model for complex system does not work any longer, because it caused lower estimation precision and over-fitting problems [1].

To overcome the aforementioned difficulties, multiple model (MM) has been advocated by some researches for complex processes. Multiple model involves several submodels and its global output is generated with the weighted sum of submodels outputs.

Intuitively, MM approach follows the principle of divide and conquer. The framework of MM soft sensor is illustrated in Fig. 1. The goal of a MM soft sensor is to estimate a mapping $f: x \rightarrow y$ in order to predict future samples. Sub-models are estimated with different nonlinear functions. Hence, MM consists of m sub-estimators, where m is usually determined a priori [2]. To improve the accuracy and robustness of MM, the global output of MM is a weighted sum of the sub-models outputs. Usually, the clustering method is carried out by k-means or fuzzy clustering method (FCM) [1, 3], and the modeling arithmetic is carried out through nonlinear regression method such as

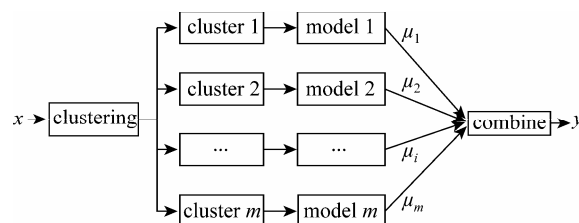


Figure 1 Framework of multiple model soft sensor

fuzzy Takagi-Sugeno (TS) modeling [2], artificial neural network (ANN) [3, 5], or support vector regression (SVR) [4, 6, 7].

However, the following drawback has appeared in practice. (1) The samples can not be effectively clustered to reflect the different work models. One of the main reason is that the second variables have the same weight factors in clustering model. The other reason is that the number of the sub-clusters is usually fixed by priori and the model can not track the movement of work points. (2) The combining method based on fuzzy membership in FCM is not suitable in practice, because the weight factor should be based on the degree of prediction confidence. (3) The output is not a probabilistic distribution prediction. Currently, the sub-estimators are modeled by point estimation technique, such as TS, ANN and SVR, so we can not assess the confidence and distribution of the prediction.

In this article, a multiple model soft sensor method is proposed. Like traditional MM soft sensor, it also consists of three distinct steps. In the first step, with affinity propagation (AP) clustering method [8], the training samples are clustered in according to the different work modes. In the second step, sub-models are estimated with Gaussian process regression [9, 10]

Received 2007-08-12, accepted 2008-10-14.

^{*} Supported by the National High Technology Research and Development Program of China (2006AA040309) and National Basic Research Program of China (2007CB714000).

^{**} To whom correspondence should be addressed. E-mail: hysu@iipc.zju.edu.cn

which provides a probabilistic non-parametric modeling approach. And in the final step, Bayesian committee machine is used to combine the estimators of sub-models and generate the global probabilistic distribution prediction.

2 AFFINITY PROPAGATION

Clustering based on a measure of similarity is a crucial problem which appears throughout scientific data analysis. Recently, a powerful algorithm called affinity propagation (AP) based on message passing technique was published in Science by Frey and Dueck [8]. Affinity propagation has fast processing speed and can avoid many of the poor solutions which caused by unlucky initializations and hard decisions.

2.1 Principle and algorithm of affinity propagation

Affinity propagation is based on similarities $s(i, k)$ (or $n \times n$ similarity matrix S for n data points) between pairs of data points, and simultaneously considers all the data points as potential cluster centers (called exemplars). To find appropriate exemplars, AP accumulates evidence "responsibility" $r(i, k)$ from data point i for how well-suited point k is to serve as the exemplar for point i , and accumulates evidence "availability" $a(i, k)$ from candidate exemplar point k for how appropriate it would be for point i to choose point k as a final cluster center. Based on evidence accumulation, AP searches for clusters through an iterative process until a high-quality set of exemplars and corresponding clusters emerges. The algorithm is as follows:

(1) Input: S, p, λ

(2) Initialization: $r(i, k) = 0, a(i, k) = 0$ for all i, k

(3) Responsibility updates:

$$r^{\text{old}}(i, k) = r(i, k) \quad (1)$$

$$r(i, k) = s(i, k) - \max_{j: j \neq k} [a(j, i) + s(i, j)] \quad (2)$$

$$r^{\text{new}}(i, k) = (1 - \lambda)r(i, k) + \lambda r^{\text{old}}(i, k) \quad (3)$$

(4) Availability updates:

$$a^{\text{old}}(k, i) = a(k, i) \quad (4)$$

$$a(k, k) = \sum_{j: j \neq k} \max[0, r(j, k)] \quad (5)$$

$$a(k, i) = \min \left\{ 0, r(k, k) + \sum_{j: j \notin \{k, i\}} \max[0, r(j, k)] \right\} \quad (6)$$

$$a^{\text{new}}(k, i) = (1 - \lambda)a(k, i) + \lambda a^{\text{old}}(k, i) \quad (7)$$

(5) Iterative from step 3 and step 4 until a high-quality set of exemplars and corresponding clusters emerges. Then go to step 6.

(6) Making assignments:

$$c_i^* = \arg \max_k [r(i, k) + a(k, i)] \quad (8)$$

2.2 Parameter of affinity propagation

There are three important parameters in AP: S, p, λ .

Before the similarity S is calculated, the data x must be normalized with Eq. (9) to overcome the dimension effect of different second variables.

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sigma_j^2}, \quad j = 1, \dots, p \quad (9)$$

As the goal of soft sensor is to minimized squared error, each similarity $s(i, k)$ is set to a negative squared error (Euclidean distance):

$$s(i, k) = - \sum_{j=1}^p w_j (x_{ij} - x_{kj})^2 \quad (10)$$

w_j is the weight factor of second variables. It can be estimated according to the important degree of second variables to work modes by technologist and adjust in practice.

The preferences p (its initial value is negative) is used in diagonal of similarity matrix S and indicates the preference that data point i should be chosen as a cluster center, it influences the number of clusters. And usually a good choice is to set all the $p(i)$ to the median of S (exclude its diagonal). However this choice can not lead to an optimal clustering solution in the new multiple model soft sensor method, so we set

$$p = \beta \cdot \text{median}(S) \quad (11)$$

β can be adjusted in according to special applications.

In each iterative step, r and a are updated, and the damping factor $\lambda \in [0, 1]$ is set as default $\lambda = 0.5$.

3 GAUSSIAN PROCESS REGRESSION

The goal of soft sensor is to estimate a mapping $f: x \rightarrow y$ in order to predict future samples. Consider a input vectors $X = [x_1, \dots, x_n]^T$, $x_i \in \mathbf{R}^q$ and the vector of output data $y = [y_1, \dots, y_n]^T$, $y_i \in \mathbf{R}$. Based on the data $D = \{X, y\}$, for the future input vector x_* , we wish to find the prediction of the corresponding output y_* .

Gaussian process (GP) is a flexible, probabilistic and non-parametric model [8, 9]. From function view, GP is a collection of random variables which have a joint multivariate Gaussian distribution [9]. Gaussian process has a property that examination of a larger set of variables does not change the distribution of the smaller one [9]. So through the clustering method and GPR, we can implement the incremental learning and get probabilistic prediction. With the presented MM soft sensor, we can gain knowledge of the different work modes with the assumptions of the dissimilar property prior in disturbance range and nonlinear character.

Assuming a relationship of the form $y = f(x) + \varepsilon$ between input x and output y , and assuming additive independent identically distributed Gaussian noise $\varepsilon \sim [0, \sigma_n^2]$, for a new test input x_* , the predictive

distribution is $f^q|(X, y, x_*)$ which is Gaussian, with mean Eq. (12), variance Eq. (13) and probability Eq. (14),

$$E(f^q|D) = k^T(X, x_*)[K(X, X) + \sigma_n^2 I]^{-1} y \quad (12)$$

$$\text{cov}(f^q|D) = k(x_*, x_*) - k^T(X, x_*).$$

$$[K(X, X) + \sigma_n^2 I]^{-1} k(X, x_*) \quad (13)$$

$$\ln p(y|X) = -\frac{1}{2} y^T [K(X, X) + \sigma_n^2 I]^{-1} y - \frac{1}{2} \ln |K(X, X) + \sigma_n^2 I| - \frac{n}{2} \ln 2\pi \quad (14)$$

where $k(X, x_*)$ is the covariance vector between the test and training cases, $K(X, X)$ is the covariance matrix for the training cases, and $k(x_*, x_*)$ is the covariance between the test itself. According to Mercer's theorem, the specification of the covariance function implies a distribution over functions. From the soft sensing view, we wish to specify the covariance that points with nearby inputs give rise to similar predictions, so the squared exponential covariance function Eq. (15) is recommended,

$$C(x_p, x_q) = \sigma_f^2 \exp\left[-\frac{1}{2l^2}(x_p - x_q)^2\right] + \sigma_n^2 \delta_{pq} \\ = k(x_p, x_q) + \sigma_n^2 \delta_{pq}. \quad (15)$$

The length-scale l , the signal variance σ_f^2 , and the noise variance σ_n^2 play the role of hyperparameters. With GP model, most of the effort consists of tuning the parameters of covariance function. This is carried out by maximizing the log-likelihood Eq. (14) of training data with respect to all hyperparameters using matrix operations.

4 BAYESIAN COMMITTEE MACHINE

With AP clustering method, the data set is divided into m sub-clusters $D = \{D^1, \dots, D^m\}$ and m models are developed on the individual data sets with GPR. Then the predictions of the individual models would be combined using a weighting scheme with Bayesian committee machine [11, 12].

When applied to a set of query points f^q , each of the m GPR estimator outputs a prediction $E(f^q|D^i)$ Eq. (12) together with covariance $\text{cov}(f^q|D^i)$ Eq. (13). BCM combines the m estimators and calculates an approximation to the expected values $\hat{E}(f^q|D)$ and covariance $\text{cov}(f^q|D)$ of the functional values at the query points:

$$\hat{E}(f^q|D) = C_{\text{BCM}}^{-1} \sum_{i=1}^m \text{cov}(f^q|D^i)^{-1} E(f^q|D^i) \quad (16)$$

$$C_{\text{BCM}} = \text{cov}(f^q|D)^{-1} = -\sum_{i=1}^m \Delta_i^{-1} + \sum_{i=1}^m \text{cov}(f^q|D^i)^{-1} \quad (17)$$

Δ_i is the priori predictive density for f^q in D^i . We recognize that the prediction of each module is weighted by the inverse covariance of its prediction. But note that we do not compute the covariance between the modules but the covariance between the f^q query points. This means that predictive densities at all query points contribute to the prediction of the BCM at a given query point. Modules that are uncertain about their prediction are automatically weighted less than modules that are certain about their prediction.

5 PERFORMANCE EVALUATION INDICATOR

Generally, we use root mean square error [RMSE, Eq. (18)] as the indicator to evaluate the error of model. But they are impacted by the dimension value of the output y , so they are not suitable to estimate the tracking precision between the prediction and the measurement value.

$$R_{\text{MSE}} = \sqrt{\sum_{i=1}^n [y_i - f(x_i)]^2 / n} \quad (18)$$

In order to overcome the effect of dimension, we introduce a new indicators to evaluate tracking precision: relative variance tracking precision [13] (RVTP) Eq. (19).

$$R_{\text{VTP}} = 1 - \sigma_{\text{error}}^2 / \sigma_{\text{property}}^2 \quad (19)$$

where σ_{error}^2 is the variance between the model prediction and the measurement value, and $\sigma_{\text{property}}^2$ is the variance of the measurement value. RVTP indicator can reflect whether the soft sensor can also remain with enough precision when the value of the measurement changes. When the R_{VTP} value is less than zero, the precision of soft sensor is very low and the use of this soft sensor will bring interference to the process. The more R_{VTP} approach one, the more accurate soft sensor tracks the real process.

6 CASE STUDIES

6.1 Simulation and analysis

The case study considered is the real-time prediction of the butane concentration (C_4) in the bottom flow of a debutanizer column, on the basis of a set of available measurements ($D = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, y\}$). It was first introduced in [14] and the data is available at the book website [15]. To simulate the offline measurement by artificial laboratory, the training data (598 samples) is sampled from the whole dataset (2394 samples) with an interval of four samples, and the left of the data is collected as test data (1796 samples). The simulation result can be seen in Table 1.

In AP+GP+BCM method, the AP clustering weight factor in Eq. (10) is set to [1.59, 1.52, 0.98, 0.33, 0.63, 0.78, 0.31] and weight factor β in Eq. (11) is set to 65.

Table 1 Simulation result

Method	Sub-model number	R_{MSE}	R_{VTP}
$y_* = y$	1	0	1
$y_* = \bar{y}$	1	0.159	0
PLS	1	0.149	0.0266
k-means+PLS	2	0.135	0.151
	3	0.148	0.0806
	4	0.144	0.109
	5	0.184	-0.0633
GP	1	0.139	0.134
k-means+GP	2	0.105	0.343
	3	0.104	0.358
	4	0.111	0.321
	5	0.105	0.257
AP+GP+BCM	3	0.0935	0.414

Partial least squares (PLS) and k-means PLS method are observed. They have lower accuracy than those of nonlinear models, implying that nonlinear methods are better than linear one for debutanizer column.

Noting the k-means method, most of them have better performance than the global model. It means that multi-model methods outperform single global model. The performance of k-means methods are determined by the number of clusters which is set a priori.

See the AP+GP+BCM method proposed in this article. It gives a Gaussian distribution prediction and shows the best performance. The clustering distance weighted factor is determined by their relativity to the work modes and the number of sub-models is determined automatically.

6.2 Application and results

In hydrocracker fractionators process, the immeasurable quality indicator is the light naphtha end

point (95% distillation range) and should be strictly monitored and controlled. The end point can be offered by artificial laboratory per eight hours. Its prediction is one of the most difficult tasks in industrial process field. We apply the proposed MM soft sensor to estimate it.

According to mechanism analysis and experience in practice, we choose ten second variables: feed temperature, feed flow rate, temperature of the fractionator top, reflux ratio of the fractionator top, temperature of the reflux jar on the fractionator top, temperature of cut point of heavy naphtha, temperature of cut point of diesel, temperature of the middle reflux, flow rate of the middle reflux and temperature of the fractionator bottom. The output variable is the end point of light naphtha.

Data collected in plant database is generally corrupted by the presence of outliers and noise, and the data inconsistent with the majority of recorded data can greatly affect the performance of data driven soft sensor design. Here, we used three steps to prepare the data for soft sensor. First, the strategy of 3σ edit rule [14] is used to detect outliers. In 3σ edit rule, the normalized distance d_i is computed:

$$d_i = \frac{x_i - \text{median}(\mathbf{x})}{\sigma_x} \quad (20)$$

and data are assumed to follow a normal distribution, so that the probability that the absolute value of d_i is greater than 3 is approximately 0.27% and an observation x_i is considered an outlier when $|d_i|$ is greater than this threshold. Second, data reconciliation [16] is used to the input and output flow variables. And third, a five minute smooth filter is used to smooth the signal.

The application result can be seen in Fig. 2. With the multiple model soft sensor based on AP, GPR and BCM, we obtain not only the prediction value but also the prediction variance with confidence. In other words, we get a Gaussian distribution prediction with confidence. The result shows that the multi-model soft sensor method has a good tracking precision performance and can guide production better.

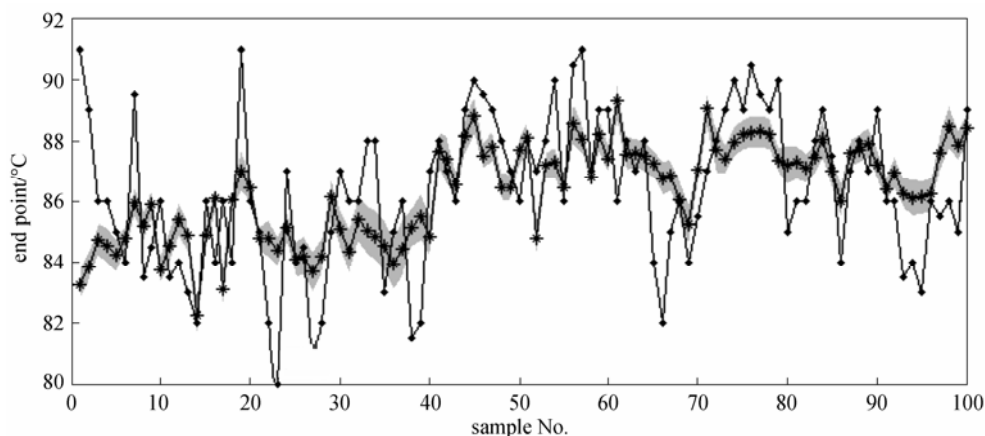


Figure 2 Applications result of multiple model soft sensor based on AP+GP+BCM
 ■ predictive variance; * predictive output; ◆ measured output

7 CONCLUSIONS

In this article, a multiple model soft sensor algorithm is proposed. The algorithm is comprised of three steps. First, AP clustering arithmetic is used to cluster the training samples according to different work modes. Second, GPR is used to build a nonlinear model for each sub cluster and gives Gaussian distribution predictions. Third, the Gaussian predictions are combined by Bayesian committee machine. The simulation results show that it is superior to the soft sensor based on sliding window or traditional multiple model soft sensor.

REFERENCES

- 1 Liu, J.L., "On-line soft sensor for polyethylene process with multiple production grades", *Control Engineering Practice*, **15**, 769–778 (2007).
- 2 Feng, R., Zhang, Y.Z., Song, C.L., Shao, H.H., "A multiple model approach to modeling based on fuzzy support vector machines", *Journal of Shanghai Jiaotong University*, **8** (2), 137–141 (2003).
- 3 Zhang, Y., Li, N., Huang, D., "Soft sensor of polyester reaction based on multiple neural network model", *Journal of East China University of Science and Technology (Natural Science Edition)*, **31** (2), 208–211 (2005).
- 4 Gui, W.H., Li, Y.G., Yang, C.H., Chen, Z.S., "Distributed SVM based on improved clustering algorithm and its application", *Control and Decision*, **19** (8), 852–856 (2004).
- 5 Du, W.L., Qian, F., Liu, M.D., Zhang, K., "4-CBA soft sensor based on fuzzy CMAC neural networks", *Chin. J. Chem. Eng.*, **13** (3), 437–440 (2005).
- 6 Zhang, Y., Su, H.Y., Liu, R.L., Chu, J., "Fuzzy support vector regression model of 4-CBA concentration for industrial PTA oxidation process", *Chin. J. Chem. Eng.*, **13** (5), 642–648 (2005).
- 7 Wang, J.L., Yu, T., Jin, C.Y., "On-line estimation of biomass in fermentation process using support vector machine", *Chin. J. Chem. Eng.*, **14** (3), 383–388 (2006).
- 8 Frey, J., Dueck, D., "Clustering by passing messages between data points", *Science*, **315** (5814), 972–976 (2007).
- 9 Rasmussen, C.E., Williams, C.K.I., *Gaussian Process for Machine Learning*, MIT Press, USA, 8–31 (2006).
- 10 Likar, B., Kocijan, J., "Predictive control of a gas-liquid separation plant based on Gaussian process model", *Comp. Chem. Eng.*, **31**, 142–152 (2007).
- 11 Tresp, V., "A Bayesian committee machine", *Neural Computation*, **12**, 2719–2741 (2000).
- 12 Tresp, V., "The generalization Bayesian committee machine", In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, NY, USA, 130–139 (2000).
- 13 King, M.J., "How to lose money with inferential properties", *Hydrocarbon Processing*, **83** (10), 47–52 (2004).
- 14 Fortuna, L., Graziani, S., Rizzo, A., Xibilia, M.G., *Soft Sensors for Monitoring and Control of Industrial Processes*, Springer Verlag, London (2006).
- 15 <http://www.springer.com/engineering/robotics/book/978-1-84628-479-3>
- 16 Schladt, M., Hu, B., "Soft sensors based on nonlinear steady-state data reconciliation in the process industry", *Chemical Engineering and Processing*, **46**, 1107–1115 (2007).