

An Improved Data-driven Soft Sensor Modeling Algorithm Based on Twin Support Vector Regression for Sugar Cane Crystallization

Yanmei Meng*

College of Mechanical Engineering
Guangxi University
Nanning, China
e-mail: gxu_mengyun@163.com
* Corresponding Author

Xiaoyuan Ma

College of Mechanical Engineering
Guangxi University
Nanning, China
e-mail: zomkey03@163.com

Kangyuan Zheng

College of Mechanical Engineering
Guangxi University
Nanning, China
e-mail: zomkey04@163.com

Wenxing Li

College of Mechanical Engineering
Guangxi University
Nanning, China
e-mail: 1551908048@qq.com

Abstract—According to the problem that some key parameters, such as mother liquor supersaturation, mother liquor purity, crystal content and crystal size distribution, cannot be measured on-line during sugar cane crystallization, an improved data-driven soft sensor modeling algorithm based on twin support vector regression is proposed. Following improvements are taken based on traditional data-driven model. The complexity of data-driven model is decreased by adding a regularization term, which can transform empirical risk into structural risk. Computational speed is increased and computational time is decreased efficiently by modifying the size of kernel function matrix. Different punishment weight is given to sample sets according to their own importance, which can increase the algorithm's generalization ability and avoid over-fitting problems to a certain degree. Experimental results show that compared with traditional data-driven soft sensor modeling, this improved algorithm has better prediction result and less prediction error than traditional data-driven modeling method.

Keywords—soft sensor; data-driven modeling; twin support vector regression; punishment weight; structural risk

I. INTRODUCTION

Nowadays, automatic control technology has been widely used in sugar cane crystallization process. However, it's difficult to realize effective control strategies due to the difficulty in measuring some key parameters, such as mother liquor purity, mother liquor supersaturation, crystal size distribution and crystal content. Soft sensor modeling method for cane sugar crystallization based on data-driven is also known as the black box modeling. In 2003, Feyo and Georgieva constructed data-driven model for cane

sugar crystallization based on neural network^[1]. In 2005, Brigitte and Michel took mother liquor purity in sugar crystallization process as research object, using neural network to construct soft sensor model of mother liquor purity^[2]. In addition, Bao Lin constructed soft sensor model based on robust statistics to solve the problem of nonlinear and strong coupling existed in the original data^[3]. In 2006, Kiran Desai and Yogesh Badhe used a novel machine learning algorithm which is called as support vector regression (SVR) to construct data-driven model for online measurement^[4]. From above, data-driven modeling for sugar cane crystallization has caused many research interests^[5-8], but it has not solved the problem that constructed model is lack of robustness and easy to cause over-fitting situation.

According to the problems of over-fitting and decreased training speed in large amount of training data sample, an improved data-driven soft sensor modeling algorithm based on twin support vector regression for sugar cane crystallization is proposed in this paper. The empirical risk minimization of the objective function is turned into structural risk minimization by introducing regularization model, which is aimed at controlling the complexity of data-driven model and has better performance. At the same time, calculating time can be effectively reduced by modifying the training sample size of kernel matrix when solving two QPP problems. And this model can achieve good generalization performance when training large amount of sample data. What's more, the weighted item is introduced to perform different punishment level according to their importance, which can avoid over-fitting problem in a certain extent and have better generalization performance. Finally, the reliability

and superiority of improved algorithm is verified by experiments.

II. ALGORITHM IMPROVEMENT

A. Improvement of Solving Over-fitting Problem

Due to the lack of structural risk, traditional data-driven model is easy to cause over-fitting problem. This paper introduces regularization model $\frac{1}{2}C_1(\mathbf{w}_1^2 + b_1^2)$ and $\frac{1}{2}C_2(\mathbf{w}_2^2 + b_2^2)$ into non-sensitive lower bound and upper bound of the regression function ε_1 , which changes the empirical risk minimization into structural risk minimization. This improvement can reduce the complexity of data-driven model and prevent the over-fitting and sub-optimal problems. The original QPP problems can be modified into following:

$$\begin{cases} \min & \frac{1}{2}C_1(\mathbf{w}_1^2 + b_1^2) + \frac{1}{2}(\mathbf{Y} - e\varepsilon_1 - (K(\mathbf{A}, \mathbf{A}^T)\mathbf{w}_1 + eb_1))^T \\ & (\mathbf{Y} - e\varepsilon_1 - (K(\mathbf{A}, \mathbf{A}^T)\mathbf{w}_1 + eb_1)) + C_3e^T\xi, \\ \text{s.t.} & \mathbf{Y} - (K(\mathbf{A}, \mathbf{A}^T)\mathbf{w}_1 + eb_1) \geq e\varepsilon_1 - \xi, \quad \xi \geq 0, \end{cases} \quad (1)$$

$$\begin{cases} \min & \frac{1}{2}C_2(\mathbf{w}_2^2 + b_2^2) + \frac{1}{2}(\mathbf{Y} + e\varepsilon_2 - (K(\mathbf{A}, \mathbf{A}^T)\mathbf{w}_2 + eb_2))^T \\ & (\mathbf{Y} + e\varepsilon_2 - (K(\mathbf{A}, \mathbf{A}^T)\mathbf{w}_2 + eb_2)) + C_4e^T\eta, \\ \text{s.t.} & (K(\mathbf{A}, \mathbf{A}^T)\mathbf{w}_2 + eb_2) - \mathbf{Y} \geq e\varepsilon_2 - \eta, \quad \eta \geq 0, \end{cases} \quad (2)$$

Where $C_1, C_2, C_3, C_4 > 0, \varepsilon_1, \varepsilon_2 \geq 0$, ξ, η is relaxation vector and e is the unit column vector. The dual form of original optimization problems (1) and (2) can be established by adding Lagrange multipliers α and β :

$$\begin{cases} \max & -\frac{1}{2}\alpha^T \mathbf{Q}(\mathbf{Q}^T \mathbf{Q} + C_1 I)^{-1} \mathbf{Q}^T \alpha + \mathbf{f}^T \mathbf{Q}(\mathbf{Q}^T \mathbf{Q} + C_1 I)^{-1} \mathbf{Q}^T \alpha - \mathbf{f}^T \alpha \\ \text{s.t.} & 0 \leq \alpha \leq C_3 e \end{cases} \quad (3)$$

$$\begin{cases} \max & -\frac{1}{2}\gamma^T \mathbf{Q}(\mathbf{Q}^T \mathbf{Q} + C_2 I)^{-1} \mathbf{Q}^T \gamma + \mathbf{h}^T \mathbf{Q}(\mathbf{Q}^T \mathbf{Q} + C_2 I)^{-1} \mathbf{Q}^T \gamma - \mathbf{h}^T \gamma \\ \text{s.t.} & 0 \leq \gamma \leq C_4 e \end{cases} \quad (4)$$

Among them, $\mathbf{h} = \mathbf{Y} + e\varepsilon_1$.

$$\mathbf{u}_1 = [\mathbf{w}_1^T \quad b_1]^T = (\mathbf{Q}^T \mathbf{Q} + C_1 I)^{-1} \mathbf{Q}^T (\mathbf{f} - \alpha) \quad (5)$$

$$\mathbf{u}_2 = [\mathbf{w}_2^T \quad b_2]^T = (\mathbf{Q}^T \mathbf{Q} + C_2 I)^{-1} \mathbf{Q}^T (\mathbf{h} + \gamma) \quad (6)$$

Final regression function can be obtained as followings:

$$f(x) = \frac{1}{2}(f_1(x) + f_2(x)) = \frac{1}{2}(\mathbf{w}_1 + \mathbf{w}_2)^T K(\mathbf{A}, x) + \frac{1}{2}(b_1 + b_2) \quad (7)$$

The over-fitting problem is solved well by introducing structural risk, and the weight factors C_1 and C_2 is used to balance the regression prediction accuracy and generalization ability. By choosing these two weight factors properly, data-driven model for sugar cane crystallization will have better performance.

B. Improvement of Rapid Solution Problem

Because all the training data (the number of samples is l) should be used when solving each QPP problem, the computational complexity of model is $O(l^3)$. And this will make it difficult to solve model when traditional data-driven model is extended to large amount of training data

samples. Generally speaking, the more sample data is, the better accuracy data-driven model will have^[9].

This paper selects parts of sample data as the input of kernel function, which makes the original square matrix into rectangular matrix and reduce model solving time^[10-12]. The improved algorithm is able to achieve good generalization performance when training large amount of sample data.

From above 2.1 section, the size of $\mathbf{Q} = [K(\mathbf{A}, \mathbf{A}^T) \quad e]$ is $l \times (l+1)$. And the size of $K(\mathbf{A}, \mathbf{A}^T)$ is $l \times l$. $K(\mathbf{A}_i, \mathbf{A}_j)$ represents value of the i -th row and j -th column. \mathbf{A}_i and \mathbf{A}_j represent data of the i -th row and j -th column. Specific improvements are shown as following steps:

Step 1: l_1 samples are selected randomly from training set, which composes a new training sample set $\bar{\mathbf{A}}_{l_1 \times n}$;

Step 2: Use rectangular matrix $K(\mathbf{A}, \bar{\mathbf{A}}^T)$ with size of $l \times l_1$ to replace square matrix $K(\mathbf{A}, \mathbf{A}^T)$ with size of $l \times l$;

Step 3: Use $K(\mathbf{A}, \bar{\mathbf{A}}^T)$ to transform dual problem $\mathbf{Q} = [K(\mathbf{A}, \mathbf{A}^T) \quad e]$ into $\bar{\mathbf{Q}} = [K(\mathbf{A}, \bar{\mathbf{A}}^T) \quad e]$ with size of $l \times (l_1 + 1)$;

Step 4: Use $\bar{\mathbf{Q}} = [K(\mathbf{A}, \bar{\mathbf{A}}^T) \quad e]$ to replace dual problem $\mathbf{Q}^T \mathbf{Q}$ with $\bar{\mathbf{Q}}^T \bar{\mathbf{Q}}$, which transforms original size of $(l+1) \times (l+1)$ into $(l_1+1) \times (l_1+1)$.

After improvement, the computational complexity of multiple matrix multiplication decreases from $O(2l^3)$ to $O(2ll_1^2 + l^2l_1)$. Suppose that $l = 1000$, $l_1 = 100$, then the computational complexity will be reduced by 10 times. After above improvement, data-driven model uses only part of the sample data and model construction time reduces. This data-driven model is sparse without affecting the model performance, and also has good generalization performance.

C. Improvement of Punishment Weight

Considering that different samples have different efforts on the boundary function area, this paper adds different punishment weight to different samples, which can avoid over-fitting problem in a certain extent. Optimization problems (1) and (2) are modified as followings.

$$\begin{cases} \min & \frac{1}{2}C_1(\mathbf{w}_1^2 + b_1^2) + \frac{1}{2}(\mathbf{Y} - e\varepsilon_1 - (K(\mathbf{A}, \mathbf{A}^T)\mathbf{w}_1 + eb_1))^T \\ & (\mathbf{Y} - e\varepsilon_1 - (K(\mathbf{A}, \mathbf{A}^T)\mathbf{w}_1 + eb_1)) + C_3(e^T\xi + \sigma_1 e^T\xi^*), \\ \text{s.t.} & \mathbf{Y} - (K(\mathbf{A}, \mathbf{A}^T)\mathbf{w}_1 + eb_1) \geq e\varepsilon_1 - \xi - \xi^*, \\ & 0 \leq \xi \leq \varepsilon_1, \quad \xi^* \geq 0 \end{cases} \quad (8)$$

$$\begin{cases} \min & \frac{1}{2}C_2(\mathbf{w}_2^2 + b_2^2) + \frac{1}{2}(\mathbf{Y} + e\varepsilon_2 - (K(\mathbf{A}, \mathbf{A}^T)\mathbf{w}_2 + eb_2))^T \\ & (\mathbf{Y} + e\varepsilon_2 - (K(\mathbf{A}, \mathbf{A}^T)\mathbf{w}_2 + eb_2)) + C_4(e^T\eta + e^T\eta^*), \\ \text{s.t.} & (K(\mathbf{A}, \mathbf{A}^T)\mathbf{w}_2 + eb_2) - \mathbf{Y} \geq e\varepsilon_2 - \eta - \eta^*, \\ & 0 \leq \eta \leq \varepsilon_2, \quad \eta^* \geq 0 \end{cases} \quad (9)$$

Where $C_1, C_2, C_3, C_4 > 0, \varepsilon_1, \varepsilon_2 \geq 0$, and ξ, ξ^*, η, η^* is relaxation vector. e is unit column vector of $l \times 1$ dimension and σ_1, σ_2 is punishment weight coefficient. In order to obtain the solution of original problem, it's necessary to transform a pair of QPP into dual problems, which are shown as followings

$$\begin{cases} \max & -\frac{1}{2}\alpha^T \mathbf{H}(\mathbf{H}^T \mathbf{H} + C_1 I)^{-1} \mathbf{H}^T \alpha + \\ & \mathbf{f}^T \mathbf{H}(\mathbf{H}^T \mathbf{H} + C_1 I)^{-1} \mathbf{H}^T \alpha - \mathbf{f}^T \alpha - \varepsilon_1 e^T \lambda \\ \text{s.t.} & 0 \leq \alpha \leq C_3 \sigma_1 \\ & \lambda \geq 0 \end{cases} \quad (10)$$

$$\begin{cases} \max & -\frac{1}{2}\gamma^T \mathbf{H}(\mathbf{H}^T \mathbf{H} + C_2 I)^{-1} \mathbf{H}^T \gamma + \\ & \mathbf{h}^T \mathbf{H}(\mathbf{H}^T \mathbf{H} + C_2 I)^{-1} \mathbf{H}^T \gamma + \mathbf{h}^T \gamma - \varepsilon_2 e^T \rho \\ \text{s.t.} & 0 \leq \gamma \leq C_4 \sigma_2 \\ & \rho \geq 0 \end{cases} \quad (11)$$

Among them, $\mathbf{h} = \mathbf{Y} + e\varepsilon_2$.

$$u_1 = [\mathbf{w}_1^T \quad b_1]^T = (\mathbf{H}^T \mathbf{H} + C_1 I + \delta I)^{-1} \mathbf{H}^T (\mathbf{f} - \alpha) \quad (12)$$

$$u_2 = [\mathbf{w}_2^T \quad b_2]^T = (\mathbf{H}^T \mathbf{H} + C_2 I + \delta I)^{-1} \mathbf{H}^T (\mathbf{h} + \gamma) \quad (13)$$

Final regression function can be obtained as followings:

$$f(x) = \frac{1}{2}(f_1(x) + f_2(x)) = \frac{1}{2}(\mathbf{w}_1 + \mathbf{w}_2)^T K(\mathbf{A}, x) + \frac{1}{2}(b_1 + b_2) \quad (14)$$

Different punishment weight can avoid over-fitting problem, which can increase robustness of data-driven model for sugar cane crystallization.

III. IMPROVED DATA-DRIVEN SOFT SENSOR MODELING ALGORITHM

The improved data-driven soft sensor modeling for sugar cane crystallization is constructed by selecting seven directly measurable variables as input of data-driven model, including vacuum degree, massecuite temperature, massecuite level, steam pressure, steam temperature, feeding rate and massecuite brix, and choosing two immeasurable variables as output, including mother liquor supersaturation and mother liquor purity. Gauss radial basis kernel function is chosen as the kernel function of data-driven model. Specific modifying steps are shown as following:

- Step 1:* Construct data sample according to the input and output of data-driven model.
- Step 2:* Normalize the sample set and divide them into training set and testing set.
- Step 3:* Optimize the parameters of data-driven model by combining particle swarm algorithm and 90% off cross validation method.
- Step 4:* Obtain the optimal punishment factor, kernel function parameters, non-sensitive loss function coefficient and weight factor.

Step 5: Construct the original optimization problem of improved data-driven model.

Step 6: Construct and solve the Lagrange equation of the original problem, which transforms the original problem into dual problem.

Step 7: A new subset of the training samples is selected to replace all the training samples, which transforms the kernel function square matrix into rectangular matrix.

Step 8: Solve the dual problem of nonlinear data-driven model and obtain the optimal solution.

Step 9: The optimal solution of original problem is obtained by Eqs. (12) and (13).

Step 10: Construct final objective regression function of data-driven model for sugar cane crystallization by using Eq. (14).

Step 11: Generate data-driven model for sugar cane crystallization.

Step 12: Determine prediction output of test sample input by combining with test sample.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

212 data samples are selected as sample set of data-driven model for sugar cane crystallization, and this sample set is divided into training set and testing set after normalization. Among them, the number of training set and testing set is 140 and 72 relatively^[13]. Training set is used to optimize model parameters and construct data-driven model, and testing set is used to validate and evaluate the constructed data-driven model. Two data-driven models are constructed by parallel mode because this model belongs to a multi-input and multi-output problem. y_1 and y_2 represent mother liquor supersaturation and purity relatively.

Fig. 1 shows the compared result of data-driven model before and after improvement, the output of which is mother liquor supersaturation. Fig. 2 shows the compared prediction result of data-driven model before and after improvement, the output of which is mother liquor purity. Table 1 shows the compared performance result of data-driven model before and after improvement, the output of which is mother liquor supersaturation. Table 2 shows the compared performance result of data-driven model before and after improvement, the output of which is mother liquor purity.

From Fig. 1 and Fig. 2, it can be seen that the improved data-driven soft sensor model for sugar cane crystallization has better prediction result than original model, which can better fit the actual value and have smaller prediction error range than original model. From Table 1 and Table 2, the improved data-driven model has smaller RMSE, MAE, MAPE, SSE, SSE/SST and higher SSR/SST (R2) compared with original model, which indicates that the improved data-driven modeling algorithm has better prediction performance and generalization performance than original model. This improved algorithm is validated to have great superiority.

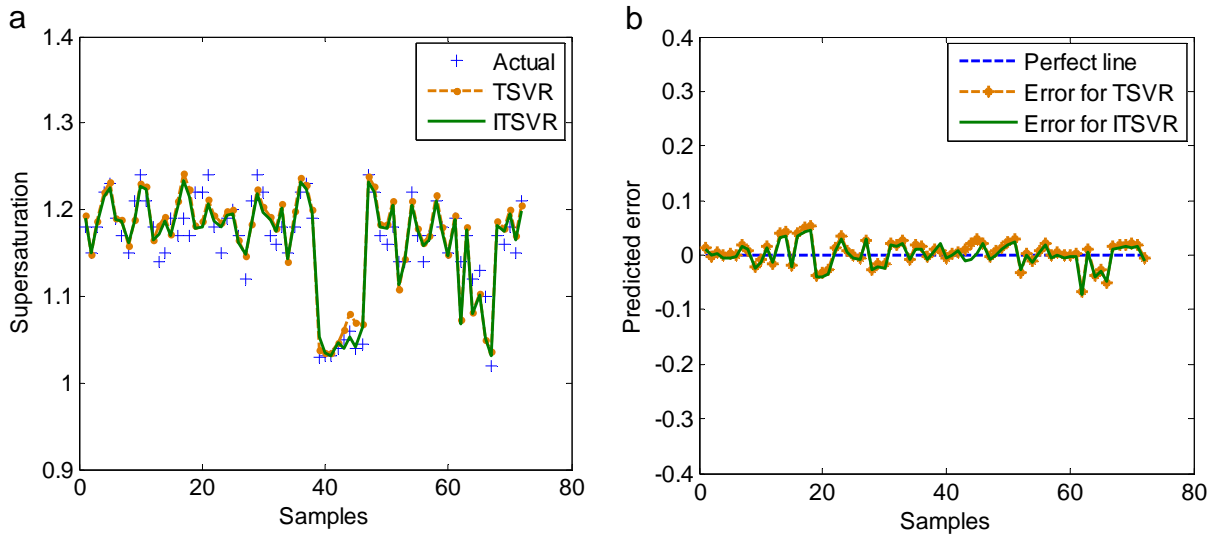


Figure 1. Comparison of prediction result of the original data-driven model and the improved model with mother liquor supersaturation as the output: (a)prediction result, (b)prediction error

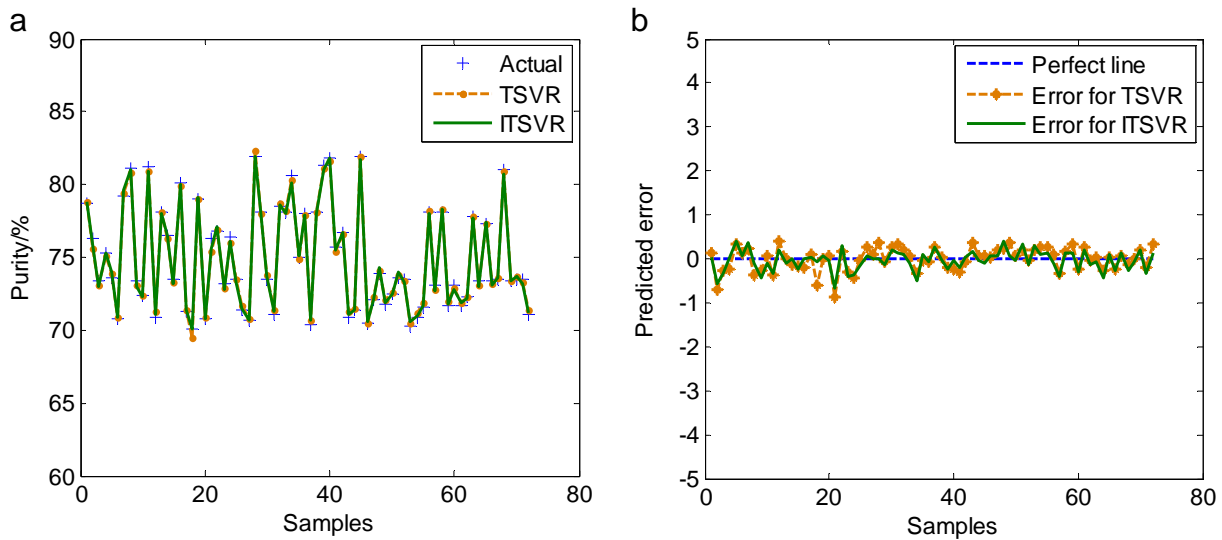


Figure 2. Comparison of prediction result of the original data-driven model and improved model with mother liquor purity as the output: (a)prediction result, (b)prediction error

TABLE 1. COMPARISON OF THE PERFORMANCE RESULT OF THE ORIGINAL DATA-DRIVEN MODEL AND THE IMPROVED MODEL WITH MOTHER LIQUOR SUPERSATURATION AS THE OUTPUT

Algorithms	Training time (s)	RMSE	MAE	MAPE	SSE	SSE/SST	SSR/SST
TSVR	0.1521	0.0231	0.0179	1.5462	0.0383	0.1679	0.8321
ITSVR	0.1076	0.0217	0.0165	1.4299	0.0339	0.1484	0.8516

TABLE 2. COMPARISON OF THE PERFORMANCE RESULT OF THE ORIGINAL DATA-DRIVEN MODEL AND THE IMPROVED MODEL WITH MOTHER LIQUOR PURITY AS THE OUTPUT

Algorithms	Training time (s)	RMSE	MAE	MAPE	SSE	SSE/SST	SSR/SST
TSVR	0.1323	0.2640	0.1915	0.2511	5.0184	0.0057	0.9943
ITSVR	0.1014	0.2329	0.1801	0.2410	3.9049	0.0044	0.9956

V. CONCLUSIONS

An improved data-driven soft sensor algorithm is proposed based on twin support vector regression. The improved model has better generalization performance after adding regularization term. And the computational speed is increased by modifying the size of kernel function sample set, which decreases computational time and complexity. Over-fitting situation is avoided in an extent by adding different punishment weight according to their different importance. Experimental results shows that this improved algorithm have good prediction result and superiority.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (No. 51465003), Innovation Project of Guangxi Graduate Education (No. YCSZ2014029), Scientific and Technological Project of Nanning City (No. 20131079), and Guangxi Manufacturing System and Advanced Manufacturing Technology Laboratory Project (No. 14-045-15S08).

REFERENCES

- [1] P. Georgieva, M.J. Meireles, and SF. de. Azevedo, "Knowledge-based hybrid modelling of a batch crystallisation when accounting for nucleation, growth and agglomeration phenomena," *Chemical Engineering Science*, vol. 58, pp. 3699-3713, Aug 2003.
- [2] B. G. Perez, M. Benne, and C. Bonnacaze, "Industrial multi-step forward predictor of mother liquor purity of the final stage of a cane sugar crystallization plant," *Journal of Food Engineering*, vol. 66, pp. 361-367, Feb 2005.
- [3] B. Lin, B. Reche, JKH. Knudsen, and SB. Jorgensen, "A systematic approach for soft sensor development", *Computers and Chemical Engineering*, vol. 31, pp. 419-425, May 2007.
- [4] K. Desai, Y. Badhe, SS. Taambe, and BD. Kulkarni, "Soft-sensor development for fed-batch bioreactors using support vector regression," *Biochemical Engineering Journal*, vol. 27, pp. 225-239, Jan 2006.
- [5] DC. Psychogios, and LH. Ungar, "A hybrid neural network-first Principles approach to Process Modeling", *AIChE Journal*, vol. 38, pp. 1499-1511, Oct 1992.
- [6] C. Kambhampati, JD. Mason, and K. Warwick, "A stable one-step-ahead predictive control of non-linear systems," *Automatica*, vol. 36, pp. 485-495, Apr 2000.
- [7] EP. Nahas, MA. Henson, and DE. Seborg, "Nonlinear internal model control strategy for neural network models", *Computers & Chemical Engineering*, vol. 16, pp. 1039-1057, Dec 1992.
- [8] LAP. Suárez, P. Georgieva, and SF. De. Azevedo, "Nonlinear MPC for fed-batch multiple stages sugar crystallization," *Chemical Engineering Research and Design*, vol. 89, pp. 753-767, Jun 2011.
- [9] P. Georgieva, LAP. Suarez, and SF. De. Azevedo, "Time Accounting Artificial Neural Networks for Biochemical Processes," in *Intelligent Systems: From Theory to Practice Studies in Computational Intelligence*, vol. 299, pp. 181-199, 2010.
- [10] J.J. Macias, P. Angelov, and X.W. Zhou, "A method for predicting quality of the crude oil distillation", *Proc. of the 2006 Int. Symposium on Evolving Fuzzy Systems*, Sep 2006, pp. 214-220.
- [11] M. Shimojo, A. Namiki, M. Ishikawa, R. Makino, and K. Mabuchi, "A tactile sensor sheet using pressure conductive rubber with electrical-wire stitched method," *IEEE Trans. on Sensors*, vol. 5, No. 4, Oct 2004, pp. 589-596.
- [12] Y. Shi, R. C. Eberhart, "Empirical study of particle swarm optimization C," *Proceedings of the 1999 Congress on Evolutionary Computation*. Washington: IEEE, July 1999, pp. 1945-1950.
- [13] M. Von. Stosch, J. Peres, and SF. De. Azevedo, "Modeling biochemical networks with intrinsic time delays: a hybrid semi-parametric approach," *BMC Systems Biology*, vol. 4, pp. 131, Sep 2010.