

## Soft-Sensing Research on Ethylene Polymerization Based on PCA-SVR Algorithm

X.R. SHI, Q.H. CHI, Z.F. FEI and J. LIANG\*

State Key Lab of Industrial Control Technology, Department of Control Science & Engineering, Zhejiang University, Hangzhou 310027, Zhejiang Province, P.R. China

\*Corresponding author: Tel: +86 571 87951970; E-mail: [jliang@iipc.zju.edu.cn](mailto:jliang@iipc.zju.edu.cn); [shixr2012@163.com](mailto:shixr2012@163.com)

(Received: 6 June 2012;

Accepted: 13 March 2013)

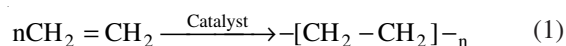
AJC-13112

The important indices indicating product specifications and grades, melt index and density, can not be measured easily in gas phase ethylene polymerization reaction. Thus, a soft-sensing algorithm is proposed which mainly adopts the technique of support vector regression (SVR) to build prediction model. In this model, two important parameters,  $C$  and  $\gamma$  are estimated through particle swarm optimization (PSO). Meanwhile, principal component analysis (PCA) is employed to extract features and by this preprocessing step the prediction accuracy can be improved to varying degree. The experiment results demonstrate that, compared with traditional neural network (NN) and single SVR methods, the proposed PCA-PSO-SVR model can achieve best results and is, hence, suitable for soft-sensing in similar processes.

**Key Words:** Principal component analysis, Support vector regression, Particle swarm optimization, Polymerization reaction.

### INTRODUCTION

Polyethylene (PE), which is widely applied in chemical lines, is a thermoplastic polymer consisting of long chains produced by combining the ingredient monomer ethylene. The chemical equation of this polymerization process is as follows:



Polyethylene is classified into several categories based mostly on its density and branching and the mechanical properties of polyethylene depend significantly on variables such as the type and length of branching, the crystal structure and the molecular weight. As measuring the weight of molten volume is easier than measuring molecular weight, melt index (MI), (the amount of molten polymer that can be squeezed through a standard orifice in 10 min) is usually used to represent the molecular weight. It is also not facile to measure the density of polyethylene resins directly, so soft-sensing methods have been proposed recently to calculate melt index and density in real time to control the grade and quality of products.

Generally, neural networks have been utilized for soft-sensing in many applications<sup>1,2</sup>. Unlike conventional statistical methods, neural networks are data driven, nonparametric, weak models which let "the data speak for themselves"<sup>3</sup>. For example, Li and Liu<sup>4</sup> succeeded in predicting melt index of polypropylene (PP) polymerization process using radial basis function (RBF) network<sup>4</sup>. Radial basis function-neural network (RBFNN) was also employed in Yang's work to calculate melt index and density

of polyethylene<sup>1</sup>. However, methods of this kind suffer from several weaknesses, such as the need for a large number of training samples, difficulty in reaching a global optimum and the danger of over-fitting.

In this study, support vector regression is employed to build prediction model due to its outstanding performance and attractive principle of structure risk minimization<sup>5</sup>. It finds a good balance between empirical error and expected error. This balance eventually leads to better generalization than other neural network models. Additionally, the SVR training process is equivalent to solving linearly constrained quadratic programming problems and the optimal solution is unique thus unlikely to generate local minima just as neural network method. This strategy was adopted in Cao's study for time series forecasting<sup>6</sup>. And a modified SVR method was utilized to predict melt index in Shi's work<sup>7</sup>.

Although SVR model is sufficient to acquire satisfying prediction results, some improvement can still be observed if feature extraction step such as PCA, independent component analysis (ICA) and factor analysis (FA) is carried out previously. For example, Fataei adopted factor analysis to extract variables of more significant information in heavy metal and agricultural toxics monitoring<sup>8</sup>.

### EXPERIMENTAL

**Support vector regression:** For linear support vector regression, a decision function has the following form:

$$f(x) = wx + b \quad (2)$$

$x$  is input data.  $w$ ,  $b$  are parameters to be determined. SVR estimates the function by minimizing the regularized risk function:

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{l} \sum_{i=1}^l L_{\varepsilon}(y_i, f(x_i)) \quad (3)$$

$y$  is output variable. The first term  $\|w\|^2$  is called the regularized term and minimizing it will make a function as flat as

possible. The second term  $\frac{1}{l} \sum_{i=1}^l L_{\varepsilon}(y_i, f(x_i))$  is the empirical

error which is measured by the  $\varepsilon$ -insensitive loss function defined as follows:

$$L_{\varepsilon}(y, f(x)) = \begin{cases} |y - f(x)| - \varepsilon & (y - f(x)) \geq \varepsilon \\ 0 & \text{Otherwise} \end{cases} \quad (4)$$

$l$  is the number of data points and  $C$  is referred to as the regularization constant which plays a trade-off between the regularized term and the empirical error.

While introducing margin slack variable  $\xi$  and  $\xi^*$ , eqn. 3 can be rewritten as:

$$\begin{aligned} \min & \frac{1}{2} \|w\|^2 + C \frac{1}{l} \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s.t. } & y_i - wx_i - b \leq \varepsilon + \xi_i \\ & -y_i + wx_i + b \leq \varepsilon + \xi_i \\ & \xi_i, \xi_i^* \geq 0, i = 1, \dots, l \end{aligned} \quad (5)$$

This is the primal problem and its dual problem is:

$$\begin{aligned} \max & -\frac{1}{2} \sum_{i,j=1}^l (a_i - a_i^*)(a_j - a_j^*) \langle x_i, x_j \rangle \\ & - \varepsilon \sum_{i=1}^l (a_i + a_i^*) + \sum_{i=1}^l y_i (a_i - a_i^*) \\ \text{s.t. } & \sum_{i=1}^l (a_i - a_i^*) = 0; a_i, a_i^* \in [0, C] \end{aligned} \quad (6)$$

where  $a_i, a_i^*$  are Lagrange multipliers.

For nonlinear regression, the main idea is to map the input data  $x$  into a higher dimensional feature space by some function  $\Phi(x)$  and calculate the linear regression there<sup>5</sup>, thus the decision function is of the following form:

$$f(x) = w\Phi(x) + b \quad (7)$$

and inner product  $\langle x_i, x_j \rangle$  in eqn. 6 is replaced by  $K(x_i, x_j)$ .  $K(x_i, x_j)$  is defined as the kernel function. In this paper, RBF kernel is employed and in general it is the reasonable first choice<sup>9</sup>. However, there are two parameters to be determined while choosing RBF kernel:  $C$  and  $\gamma$ .  $C$  is the regularization variable mentioned above which tries to get a balance between empirical risk and good generalization.  $\gamma$  is the kernel parameter for RBF kernel function which is defined as:

$$K(x, y) = e^{-\gamma \|x - y\|^2} \quad (8)$$

The issue of parameters selection is discussed in the following section. Through SVR, we can construct the soft computing model by which MI and density can be calculated according to the input process data. Further, the feature extraction step should be introduced for the dimension of original data is as large as up to 29 and there must exist correlation and noise in these data. By feature extraction, the prediction precision is expected to be improved to some extent.

**Principal component analysis:** Principal component analysis is designed to extract features from inputs. It transforms data from high dimension space into a lower one wherein the components are uncorrelated. So, principal component can describe original data with reduced form and minimum loss of important information<sup>10</sup>. With a given data set  $X \in \mathbb{R}^{m \times n}$ ,  $m$  refers to the number of observations and  $n$  denotes the number of process variables or model indicators, the PCA decomposition of  $X$  is as follows:

$$X = \sum_{i=1}^k t_i p_i^T + E \quad k \leq n \quad (9)$$

The vectors  $p_i \in \mathbb{R}^{n \times 1}$  are the loading vectors and they are orthogonal to each other. Vector  $t_i \in \mathbb{R}^{m \times 1}$  is principal component, also called score vector. It is the projection of data  $X$  onto vector  $p_i$ . That is:

$$t_i = X p_i \quad (10)$$

$E \in \mathbb{R}^{m \times n}$  is residual matrix representing the PCA model prediction error. Usually it is caused by measurement noise, so if we discharge it by only picking up the first several principal components, we can extract essential features from original data. *i.e.*, we can get less dimension data and better prediction performance.

**PCA-PSO-SVR:** Integrating the advantages of SVR to build prediction model and PCA to extract crucial features, we can have a combinational algorithm PCA-SVR. Still, it is not known in advance which  $C$  and  $\gamma$  are the best for special problem, consequently some model determining (parameter search) must be done. Its goal is to select good  $C, \gamma$  so that the model can predict unknown data more accurately.

In this study, we use particle swarm optimization to find good  $C, \gamma$  pair for its outstanding performance in parameter optimization area. This algorithm was first proposed by Eberhart and Kennedy and was designed to simulate social behaviour such as the foraging movement of a bird flock or fish school<sup>11</sup>. It optimizes a problem by having a population of candidate solutions called particles, then moving these particles around in the search area according to the preset simple formulae over the particle's position and velocity. For PSO details, refer to related documents<sup>11-13</sup>.

Fig. 1 illustrates the process of building PSO based PCA-SVR prediction model. The original process data  $X$  is first transformed to  $X'$  through PCA. The SVR prediction model is then constructed using  $X'$  and the output variable, MI or density in this paper. PSO is employed to estimate the important model parameters  $C$  and  $\gamma$ .

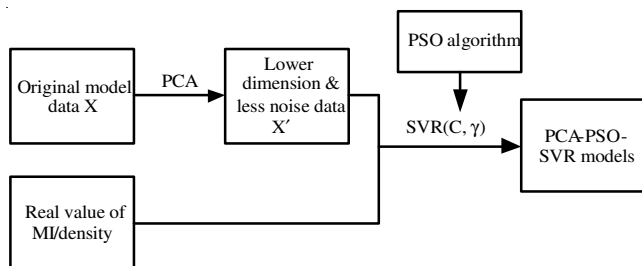


Fig. 1. Flowchart of building PCA-PSO-SVR models

**Model variables and calculation:** Fig. 2 presents the polymerization of ethylene in a fluidized-bed reactor with a heterogeneous Ziegler-Natta catalyst<sup>14</sup>. Ethylene, butene, hydrogen and catalyst are fed continuously into the reactor. The gas phase transports heat out of the reactor through a recycling system through which the recycle stream is pumped back into the bottom of the reactor together with the fresh feeds. Fluidized polymer product is removed from the base of the reactor through a discharge valve and excessive pressure and impurities are removed at the top of the reactor through a bleed.

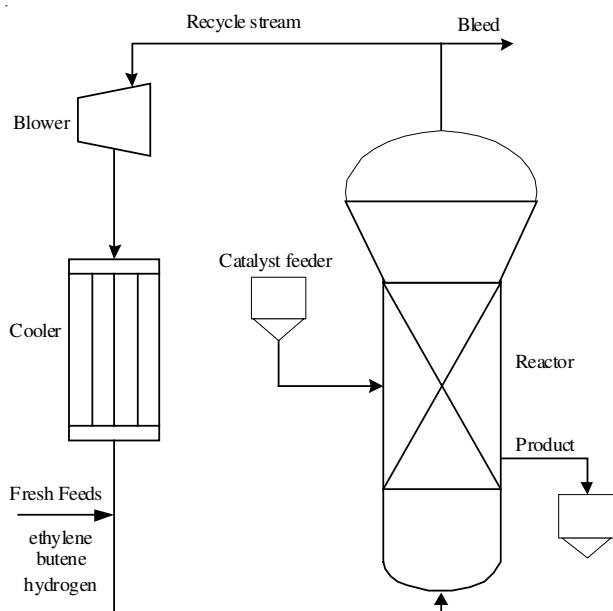


Fig. 2. Gas phase polyethylene reactor system

In our experiments, 29 process variables as shown in Table-1 were selected to built the prediction model for MI and density. Obviously there exist noise in these variables and correlation among them for these variables should conform to various inside physical and chemical mechanisms. Further, these inside mechanisms are complicated so in this paper, mathematic but analytic approach is considered.

We have 456 samples collected from industrial field with the frequency one sample per minute. The whole samples were divided into two parts, 256 for training prediction models and the other 200 for model evaluating. As shown in Fig. 3, we tested the same data with three different models: RBFNN, unique SVR and PSO based PCA-SVR to inspect their prediction precision separately and some conclusions are listed in following section. In Fig. 3, variable MI is taken as an instance and method for density is exactly the same.

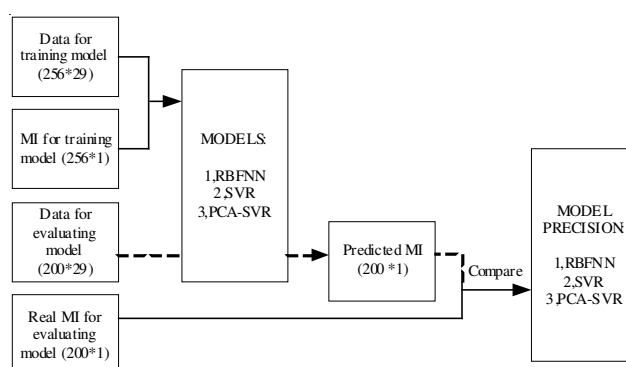


Fig. 3. Scheme of inspecting three models' prediction precisions for MI

## RESULTS AND DISCUSSION

For variable MI, we built three prediction models: RBF neural network (Yang's method<sup>1</sup>), SVR with RBF kernel<sup>9</sup> and PSO based PCA-SVR. Results of three models are presented in subfigure a, b and c of Fig. 4.

As can be seen from Fig. 4, the prediction precision becomes better and better from RBF model to PSO based PCA-SVR. Subfigure a, b and c of Fig. 5 presents the prediction accuracy of different models for density. As for the prediction performance, we can draw the same conclusion with that of MI.

TABLE-1  
29 VARIABLES FOR BUILDING PREDICTION MODEL

No.	No description (Unit)	No.	No description (Unit)
1	Inlet temperature (°C)	16	Input rate of aluminum triethyl (kg/h)
2	Reactor temperature (°C)	17	Flow rate of product (t/h)
3	Outlet temperature (°C)	18	Molar ratio of hydrogen to ethylene
4	Ethylene temperature (°C)	19	Molar ratio of butene to ethylene
5	Butene temperature (°C)	20	Rotation rate of catalyst feeder A (r/min)
6	Reactor pressure (kPa)	21	Rotation rate of catalyst feeder B (r/min)
7	Ethylene pressure (kPa)	22	Weight of bed (t)
8	Pressure drop of tank A (kPa)	23	Bed level (%)
9	Pressure drop of tank B (kPa)	24	Density of upper product (kg/m <sup>3</sup> )
10	Pressure drop of distribution plate (kPa)	25	Density of bottom product (kg/m <sup>3</sup> )
11	Flow rate of recycle gas (t/h)	26	Concentration of ethylene in recycle gas (mol %)
12	Flow rate of ethylene (t/h)	27	Concentration of butene in recycle gas (mol %)
13	Flow rate of butene (kg/h)	28	Concentration of hydrogen in recycle gas (mol %)
14	Flow rate of hydrogen (kg/h)	29	Concentration of nitrogen in recycle gas (mol %)
15	Flow rate of helium (kg/h)	—	—

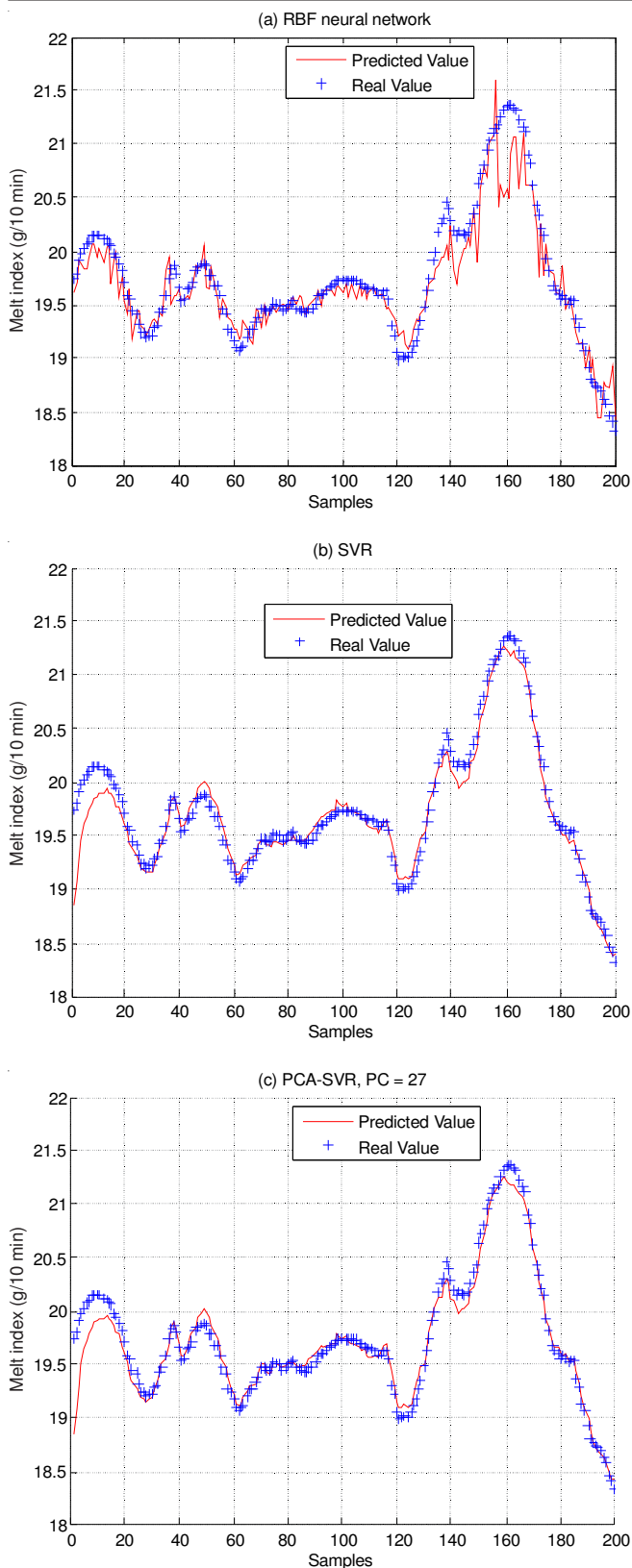


Fig. 4. MI prediction results of three models

The differences between predicted values and observed values are measured with root mean square error (RMSE) which is defined as<sup>15</sup>:

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

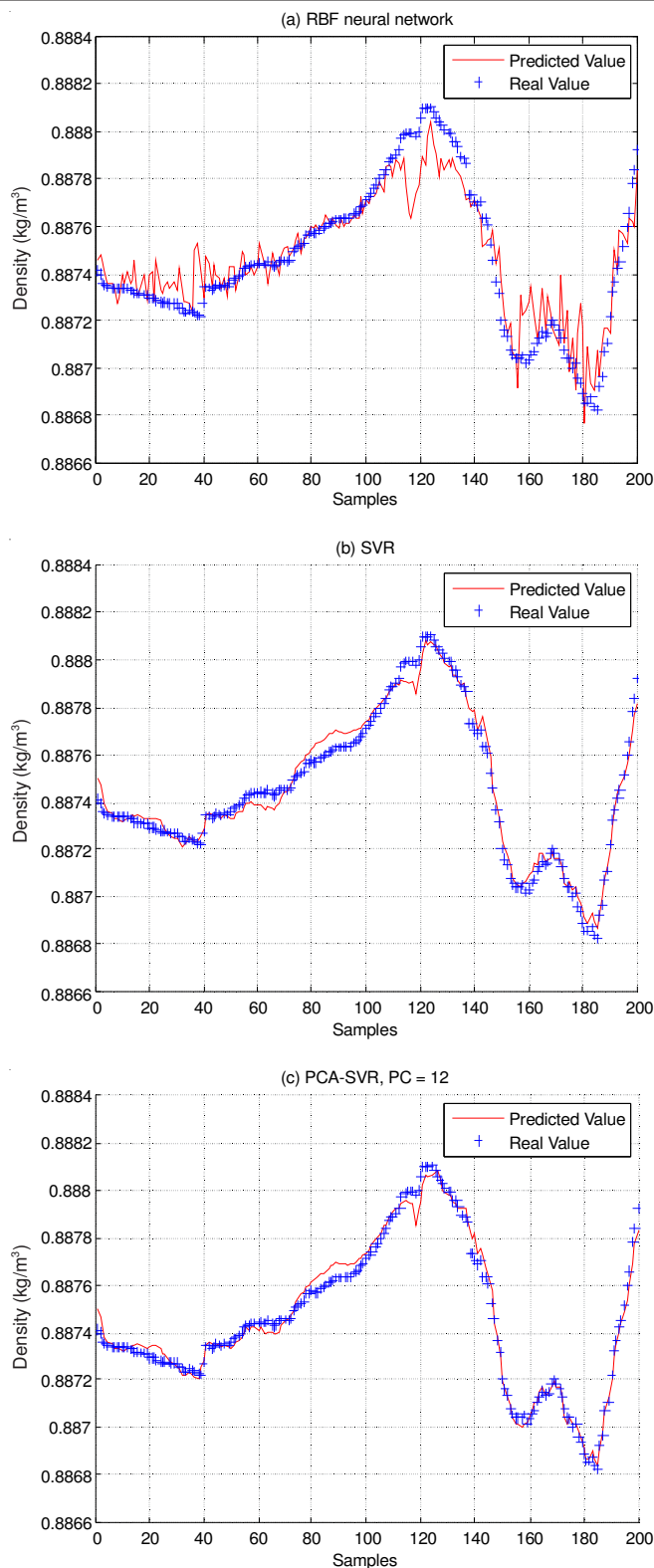


Fig. 5. Density prediction results of three models

where  $N$  is the number of samples.  $y$  and  $\hat{y}$  are predicted values and real values respectively. The RMSE results are listed in Table-2 together with some model parameters such as:  $C$ ,  $\gamma$ .

PC number in Table-2 is the number of principal components which should be determined deliberately to acquire optimal preprocess effectiveness. Ideally, the number should be

TABLE-2  
RMSE VALUES OF DIFFERENT MODELS

Model	RMSE of MI	RMSE of density ( $\times 10^{-4}$ )
RBFNN	0.2195	1.1527
SVR (Best C/best $\gamma$ )	0.1446 (2.8249/0.1)	0.4369 (1.3659/0.1)
PCA-SVR (PC Number/best C/best $\gamma$ )	0.1364 (27/1.15/0.1)	0.3710 (12/2.5471/0.1)

selected to provide maximum feature representation of original data. In experiments, we chose the optimal one through cross validation. Fig. 6 demonstrates that if 12 principal components are selected for measuring density, the best representation of original data and thus the optimal prediction precision can be achieved.

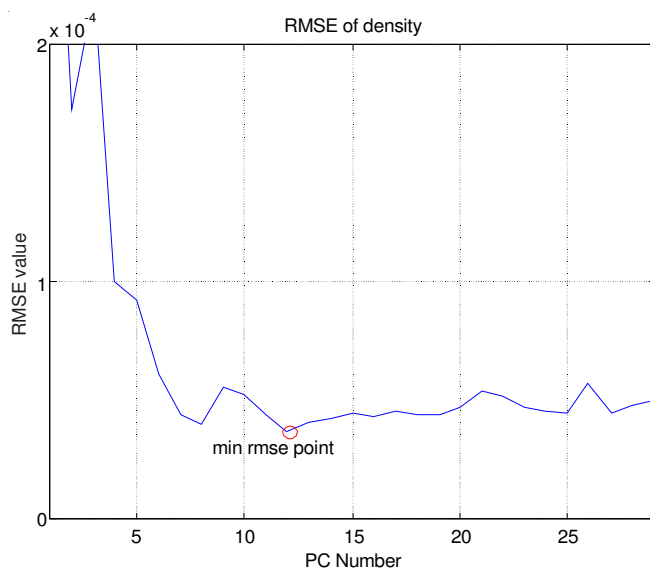


Fig. 6. Number of PCs has effect on density's prediction accuracy

It should be clarified that PSO, as an intelligent algorithm, is applied here to seek for ideal parameters. There are also other approaches to achieve this target among which the grid search method is a typical but expensive one. It just searches in the parameters space exhaustively with proper step size. Compared with this method, PSO is more explainable and in our experiments, better results can be obtained as listed in Table-3. The advantage of PSO in such applications is testified.

TABLE-3  
SEARCH METHODS FOR MEASURING DENSITY: PSO AND GRID SEARCH

Model	Best C	Best $\gamma$	RMSE of density ( $\times 10^{-4}$ )
PCA-SVR (Grid search)	0.7579	0.3789	0.4395
PCA-SVR (PSO based)	2.5471	0.1000	0.3710

## Conclusion

A PCA-PSO-SVR algorithm for soft-sensing is proposed, in which SVR is proved to be more competent in predicting PE's product indices than traditional neural network method.

In previous works<sup>1</sup>, we have found that increasing noise would make PCA more contributive. Also, in this study, the prediction precision can be improved by 15.1 % for density and 5.57 % for MI (Table-2) with the assistance of PCA. But it should be noticed that the improvements of MI and density are not equivalent or similar, especially, the optimal PC number for measuring density is 12 while 27 is proper for MI. This demonstrates two facts: there exist correlations among process variables; the correlation between input variables and output variable could be unintentionally ignored while using PCA method *i.e.*, even if there is strong linear relationship between MI and 29 process variables, PCA method is not able to decorrelate this correlation. This phenomenon may open the door to introducing other strategies such as partial least squares (PLS) to extract features of both input and output variables. This may be the next research subject of interest.

In this kind of applications, PSO is utilized rightly for conventional grid search method is time expensive and PSO has acquired better results in experiments. In conclusion, the proposed PCA-PSO-SVR algorithm is suitable for soft-sensing in ethylene polymerization and similar processes.

## ACKNOWLEDGEMENTS

This research work was supported by the National Natural Science Foundation of China under Grant No. 61174114, the Research Fund for the Doctoral Program of Higher Education in China under Grant No. 20050335018 and the National High Technology Research and Development Program of China under Grant No. 2007AA04Z168.

## REFERENCES

1. M. Yang, B. Hu, Z.S. Fei, P.Y. Zheng and J. Liang, *Chin. J. Sci. Instrum.*, **31**, 481 (2010).
2. L. Fortuna, S. Graziani and M.G. Xibilia, *Control Eng. Practice*, **13**, 499 (2005).
3. P. Ramasubramanian and A. Kannan, *Soft Comput.*, **10**, 699 (2006).
4. J.B. Li and X.G. Liu, *J. Appl. Polym. Sci.*, **119**, 3093 (2011).
5. Y.L. Wang, Y.D. Hu, G.D. Li and Q.S. Wei, *Asian J. Chem.*, **19**, 4512 (2007).
6. L.J. Cao and F.E.H. Tay, *IEEE Trans. Neural Networ.*, **14**, 1506 (2003).
7. J. Shi and X.G. Liu, *J. Appl. Polym. Sci.*, **101**, 285 (2006).
8. E. Fataei, S.M. Monavari, A.H. Hasani, A.R. Karbasi and S.A. Mirbagheri, *Asian J. Chem.*, **22**, 2991 (2010).
9. C.W. Hsu, C.C. Chang and C.J. Lin, A Practical Guide to Support Vector Classification, National Taiwan University, Taipei (2008).
10. S. Yarel, *Asian J. Chem.*, **21**, 4054 (2009).
11. J. Kennedy and R. Eberhart, In Proceedings of IEEE International Conference on Neural Networks: Particle Swarm Optimization, Perth, Australia, p. 1942 (1995).
12. M. Clerc and J. Kennedy, *IEEE Trans. Evolut. Comput.*, **6**, 58 (2002).
13. Y. Shi and R.C. Eberhart, In Proceedings of IEEE International Conference on Evolutionary Computation: A Modified Particle Swarm Optimizer, San Diego, p. 69 (1998).
14. K.B. McAuley and J.F. MacGregor, *AIChE J.*, **38**, 1564 (1992).
15. X. Bao and L.K. Dai, *Asian J. Chem.*, **22**, 4511 (2010).