Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Granular Computing Concept based long-term prediction of Gas Tank Levels in Steel Industry

Zhongyang Han, Jun Zhao, Wei Wang, Ying Liu and Quanli Liu*

** School of Control Science and Engineering, Dalian University of Technology, China (e-mail: hzy_xyq@163.com)*

Abstract: The converter gas, especially Linz Donawitz converter gas (LDG), is one of the most significant secondary energy resources in a large scale steel plant. For such a thing, the accurate prediction for the gas tank levels largely contributes to the energy optimization operations. Taking the LDG system of a steel plant in China into consideration, a regression model based on the Granular Computing (GrC) is proposed in this study to provide a long-term prediction for the LDG tank levels, in which the data segments are entirely considered for the prediction horizon extension rather than the generic data point-oriented modeling. For being more practical, this study specially granulates the initial data with regard to industrial semantic meaning. And, different from ordinary time series analysis, this method considers the factor related to the gas tank levels. Bearing this in mind, the fuzzy rules by adopting a fuzzy C-means based clustering is established. To verify the effectiveness of the proposed method, a series of practical experiments by using the industrial data coming from the energy data center of this plant are conducted, and the results demonstrate the practicability of the proposed approach.

Keywords: Steel industry; LDG system; prediction; regression model; Granular Computing

1. INTRODUCTION

Generated by converters of steel making process, LDG is always used as a secondary fuel for the manufacturing procedures of steel industry. Since the trend of gas tank levels can be viewed as the reference of the energy balance, the accurate prediction for the gas tank levels is fairly helpful for efficient utilization of LDG. However, due to the complex characteristics in industry, such as nonlinear, time-variation, etc., it is very hard to use traditional modelling methods that were based on the physics or mechanism principles to describe the practical industrial process.

Recently, with the developments of industrial informatics and database technology, most of steel plants had accumulated a large number of energy data, which triggered a series of data-based modelling methods to be presented in literature, especially for the prediction of the by-product gas system in steel industry. Focusing on the coke oven gas and the blast furnace gas, respectively, Zhang et al. (2010) and Zhao et al. (2012a) reported a least square support vector machine (LSSVM) and a multi-kernel based regression approaches to realize the real-time prediction, in which the prediction horizons are limited within 30 to 60 minutes. Yet, the practical structures of both of the gas systems are largely different with that of general LDG system. An echo state network (ESN) combined with a modified grey correlation was proposed in Zhao et al. (2011) for the blast furnace gas prediction, which belonged to a time series based framework. However, the variation of the gas tank levels were usually affected by a number of energy units, and it was rather unreasonable if only adopting such a time series based approach. Besides, a multi-output LSSVM was designed for the prediction of LDG tank levels in Han et al. (2012). Nevertheless, the prediction horizon was also ranged within 1 hour owing to the demand of the calculation accuracy. In fact, a longer prediction horizon is also the concern in industrial practice, which can provide the energy scheduling workers with the scientific guidance for energy optimization.

Granular Computing (GrC), regarded as a new data-based technology, is becoming more and more popular in the domains of data mining and computational intelligence. Smoothly combined with fuzzy theory, rough sets, interval analysis, etc., and instead of data point-oriented research, GrC mainly concentrates on data granules, which can extend the horizon of data analysis Bargiela and Pedrycz (2003). In literature, some valuable deductions and simulations with fuzzy C-means (FCM) clustering under GrC concept was proposed in Pedrycz and Bargiela (2002), where a heuristics based theory explanation was emphasized. Wu et al. (2010) partitioned the fuzzy domain and made prediction with GrC. But the prediction length was still strictly limited due to the iteration mechanism. Besides, Kumar et al. (2007) integrated GrC and SVM to predict the solubility of proteins. However, the results were in terms of some association rules, but not of the consecutive time series values. Then, an approach for long-term prediction was presented in Dong and Pedrycz (2008), which granulated the time series based on its increasing or decreasing trends. While for industrial application, it is more reasonable for data granulation to treat a certain industrial process with a semantic meaning.

Given the requirement of the gas system, a data granule (data segment) based modelling approach is proposed in this study for the long-term prediction of the LDG system, in which the factors that influence the tank levels are taken into account. Since the iterative errors of the proposed method can be effectively avoided, the prediction horizon could be extended to a number of hours compared to that of the data points based method. In addition, with regard to the strong dynamic features of the industrial data, the membership grades are effectively determined by using a FCM based clustering approach. To verify the performance of the proposed method, the real gas data coming from a large-scale steel plant are employed, and the experimental results indicate that the prediction accuracy and availability of this method which provides the energy scheduling operations with scientific guidance.

This paper is organized as follows. In Section 2, the structure of LDG system is briefly illustrated, and the long-term prediction modelling for the gas tank levels based on GrC concept is proposed in Section 3. In Section 4, the data experiments and the relevant statistic comparative analysis are involved. Finally, we draw some conclusions in Section 5.

2. PROBLEM DESCRIPTION

A brief structure of the LDG system can be depicted in Fig.1, where it primarily consists of the gas generation units, the transportation system, and the consumption units. The generation units include three converters; the transportation system covers two tanks, the gas mixture stations, the press stations, as well as the whole network pipelines; and the consumption units include the blast furnaces, rolling process, etc. The LDG system structure has relatively specific structure, because the generated LDG needs to be firstly stored into the gas tanks before transported to the consumption units. In such a way, the tank levels are actually affected by not only the generation units, but the other energy users as well.

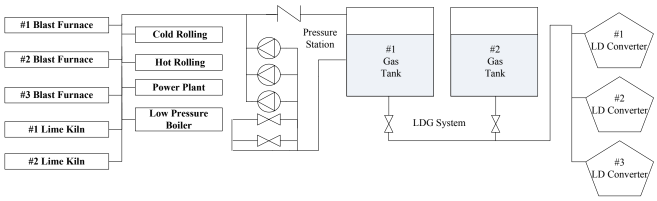

Fig. 1. Network structure of LDG system in steel industry.

In industrial practice, the tanks usually play as the buffers for balancing the gas generation and consumption. However, because of their limitative storage abilities, the scheduling operation has to be real-time implemented in order to maintain the energy balance. In this process, the tank levels will be a very significant evaluation criterion, therefore they are essential and crucial to the energy system in steel industry. Because of the important role of the gas tanks, it could be assumed that if one can estimate their tendencies in advance, then the situations like 'short of supply' or 'waste of energy' would be effectively avoided. Naturally, the prediction work above is benefit for energy saving and the whole production process.

3. REGRESSION MODELING BASED ON GRANULAR COMPUTING CONCEPT

The existing studies on regression modelling for industry mainly focused on the problems that took the data points as the fundamental units, which was applicable for some relatively small-scale problems (Sheng et al. (2013)). However, when facing with a large amount of data, the predicted accuracy can hardly be guaranteed due to the iterative error accumulation, which usually results in the limitation of prediction horizon and the intensive computing time cost.

Given the structure of LDG system, the gas tank levels are affected by a number of energy units, which not only involves the gas generation units, but also some consumption ones. In this study, a proposed data granulation is firstly proposed for providing a scientific data foundation. Then, as for the multiple influence factors, a fuzzy clustering and inference based approach is then designed to model the gas tank levels.

3.1 Data granulation

Data granulation plays an essential role in granular computing based modelling process. In literature, some studies adopted the membership function for granulating (Lin (1999)), and some took the outcomes by clustering as the data granules, see Bargiela and Pedrycz (2008). One can view the above researches as a class of granulation that paid more attention to the data dynamic features. In this study, a novel industrial semantic meaning based data partition method is proposed from the standpoint of engineering application.

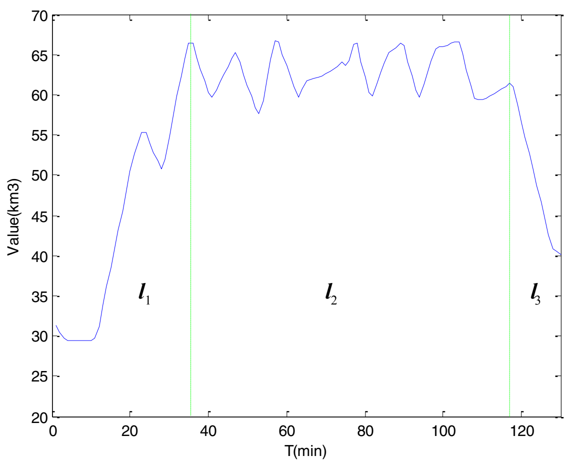

Fig. 2. Data granulation of #1 gas tank.

In practice, since the gas tanks serve as either a buffer for gas temporary storage or a supplier when the gas amount in use is insufficient, the data of the gas tank level can be divided into three categories. Fig.2 illustrates the data granulation of the #1 gas tank level in this plant: l_1 -- the gas on recycling; l_2 -- relative balance procedure of the gas tank level; l_3 -- the gas in tank on using. Similarly, the data partition of #2 gas tank is shown in Fig.3 which based on the consideration of the industrial process. On the other hand, it is clear that the difference between the gas generation and consumption can greatly affect the tank levels. Here, one can divide the difference situations into two modes, see Fig.4, where the positive $\boldsymbol{d}_1$ refers to the situation that the generation amount is larger than that of the consumption, while the negative $\boldsymbol{d}_2$ is the opposite situation, see also the situations of $\boldsymbol{d}_3$, $\boldsymbol{d}_4$ and $\boldsymbol{d}_5$. These three figures are all for data granulation as the pre-processing part.

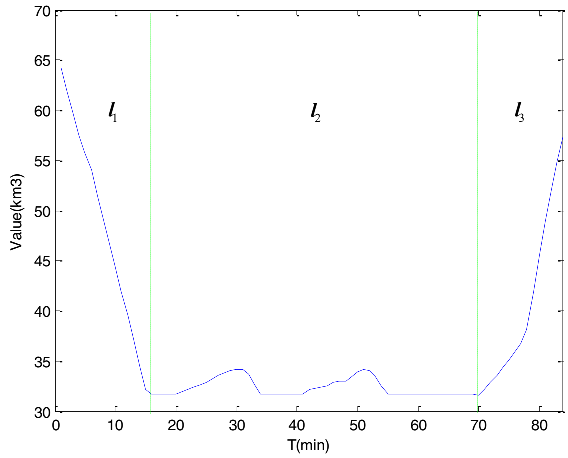

Fig. 3. Data granulation of #2 gas tank.

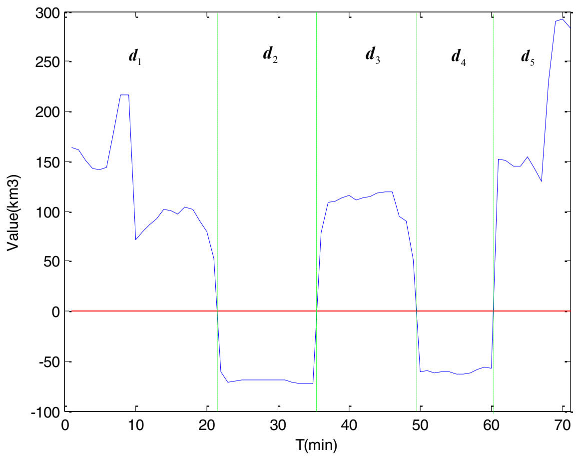

Fig. 4. Data granulation of flow difference.

With the data granulation based on the industrial semantic meaning, the fundamental dataset for the long-term regression modelling can be provided.

3.2 Granular Computing based modelling with FCM

Owing to the complexity of LDG system, as well as the nonlinear characteristic of industrial data, it is hard to establish an accurate membership function for the regression. However, the data clustering based on FCM can also determine the membership grades through unsupervised learning of data itself. Besides, it still exhibits an efficient computational capability even if dealing with a large number of data segments. The modelling process in the GrC framework is as follows.

Taking the tank levels on current time point $L=\{\boldsymbol{l}_1,\boldsymbol{l}_2,\cdots\cdots,\boldsymbol{l}_N\}$ as an example, where $\boldsymbol{l}_i=\{l_{i1},l_{i2},\cdots\cdots,l_{iw}\}$ is a granule with the length *w*, one can adopt the FCM to find a solution of the following optimization.

$$Q=\sum_{i=1}^{c}\sum_{j=1}^{N}u_{ij}^m d_{ij}^2 \tag{1}$$

$$s.t.\ \sum_{i=1}^{c}u_{ij}=1 \quad j=1,2,\cdots\cdots,N \tag{2}$$

where u_{ij} denotes the membership grades of *j*th granule belonging to *i*th cluster, the matrix constructed by u_{ij} is called as the partition matrix. $1<c\le N$ refers to the constraint of the number of clusters, m is the fuzzy coefficient, d_{ij} is a distance between *j*th granule and the center of *i*th cluster which deployed Euclidean distance in this study Dattorro (2005), i.e.,

$$d_{ij}=\left\|\boldsymbol{l}_j-\boldsymbol{v}_i\right\| \quad i=1,2,\cdots\cdots,c \quad j=1,2,\cdots\cdots,N \tag{3}$$

where v_i refers to the center of *i*th cluster. It is obviously that the optimization can be transformed into a constraint-free problem by introducing the Lagrange multiplier. Computing the derivative of v_{ij} with respect to u_{ij} and making it equal to zero, one can have (4)

$$v_{ij}=\frac{\sum_{k=1}^{N}u_{ik}^m l_{kj}}{\sum_{k=1}^{N}u_{ik}^m} \quad i=1,2,\cdots\cdots,c \quad j=1,2,\cdots\cdots,w \tag{4}$$

It should be stressed here that for the purpose of information integration/unification, the training and testing set for every class of data are computed together, i.e., sharing the same set of the clusters. Then the predicted result and the given sample will be in the same domain.

3.3 Fuzzy Inference

Based on the analysis above, one can designate the relationship between input and output is of second order, elaborated as

$$L_0(t-1),\ D(t)\ \rightarrow\ L(t) \tag{5}$$

Although there were a lot of methods in literature to obtain the fuzzy rules (Kasabov and Song (2002)), such as expert experience based approach (Siler and Buckley (2005)), it was necessary to communicate with the experts or work with a great deal of experiments, which suffered from a complicated process. On the other hand, if describing the data tendency only using ‘rising’ or ‘falling’ linguistics (Dong and Pedrycz (2008)), it is also inapplicable to be used in the gas tank level variations.
In this study, a FCM based clustering gives not only the partition matrix, but also the clustering center for each dataset, which makes the question ‘which cluster the granule to the

most extent belongs to' be solved. As such, the rules can be obtained in the form of

$$R:\ If\ L_0(t-1)\ is\ c_{L_0}, D(t)\ is\ c_D,\ then\ L(t)\ is\ c_L \quad (6)$$

where c_{L_0}, c_D and c_L are the tags of three data sets for identifying which cluster the granule belongs to. For instance, c_{L_0} would be marked as 3 when the maximal membership of $L_0(t-1)$ is toward the 3rd cluster. One can use a $c\times 1$ column vector $\boldsymbol{h}_j = [0\ 0\ 1\ \cdots\ 0]^T$ to quantize the above statement. Finally, a fuzzy rule set with *N* rules in the form of (6) is acquired, together with a set of $\boldsymbol{h}_j$ where $j = 1, 2, \cdots\cdots N$.

3.4 Long term prediction

Given the above fuzzy rules generation, the rules contain of the historical information of LDG system operations. As for the prediction process, one can search the rules that have the same tags as the input $\boldsymbol{l}_{0(N+1)}$ and $\boldsymbol{d}_{(N+1)}$ in the testing set. For summarizing the searching result, a variable considering the output of the rules and the corresponding membership grades u_{rj}^L is as follows:

$$\boldsymbol{p} = \sum_{j=1}^{N} h_j u_{rj}^L \quad (7)$$

where *r* refers to the cluster which $\boldsymbol{h}_j$ tags. It should be mentioned that the u_{rj}^L is the maximal value in the *j*th column of the partition matrix, which also indicates the most grade of belonging towards certain cluster. This paper employs the centroid technique for the final defuzzification, i.e.,

$$\hat{\boldsymbol{p}} = \boldsymbol{p}^T \boldsymbol{V}_L \quad (8)$$

where $\boldsymbol{V}_L$ is the prototype matrix of $\boldsymbol{L}$ through FCM computing, $\hat{\boldsymbol{p}}$ is the predicted value.

As the modelling process demonstrated, the unsupervised method FCM takes the responsibility of obtaining the membership grades, which apparently avoids the iteration compared to the data points based method, and gives the predicted result with length *w*.

4. EXPERIMENTS AND ANALYSIS

In this section, this study considers two gas tanks of the LDG system in the plant and employs the practical data as the validation example, where the sample interval is one minute. Considering the industrial demand, one can set the prediction horizon as 360 minutes, which means that the future 360 data points will be predicted. To quantify the prediction quality, the mean absolute percentage error (MAPE) and the root mean square error (RMSE) are adopted as the evaluation criterions,

$$MAPE = \frac{1}{T}\sum_{t=1}^{T} \left|y^T(t) - y_d^T(t)\right| \Big/ y_d^T(t) \times 100\%,$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y^T(t) - y_d^T(t))^2}$$

where T refers to the number of predicted points, $y(t)$ are the predicted values, and $y_d(t)$ are the real values. In order to provide a comprehensive validation for the proposed long-term prediction, two classes of verifications are designed. Here, one can select a class of kernel learning based method, least square support vector machine (LSSVM) , and a typical recurrent neural network, echo state network (ESN) for the comparative validation experiments, which had been applied to the energy system modelling (see, e.g., Zhang et al. 2010; Zhao et al. 2012b).

Table 1 Error and computing time statistics of set 1 for gas tank #1

Methods	MAPE	Root MSE	CT(s)
LSSVM	0.2073	16.9644	1.558
ESN	0.2651	21.6255	1.249
GrC (Time-Series model)	0.2108	17.0688	0.154
Proposed Method	0.0658	6.5440	0.103

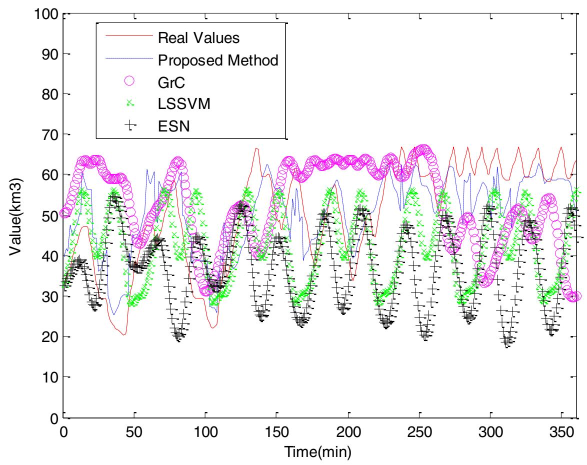

Fig.5. Prediction results of set 1 for gas tank #1

Firstly, to verify the selection of the influence factors on the gas tanks is appropriate, the comparison between the time series based modelling and the proposed relation based approach is carried out. We randomly use one of the experimental results by adopting the #1 tank as an illustrative example, see Fig. 5. It is primarily from this figure that the proposed method presents the better results compared to the other time series concept based methods, where the GrC based one fails to provide the obvious advantages. As for a quantitative comparison, one can see the resulting statistics as listed in Table 1, where the accuracy and the computing time (CT) are listed.

Additionally, the data experiments by using the data of #2 gas tank are shown in Fig.6, and the quantitative results statistics is also listed in Table 2, where the results of the proposed method can achieve the accuracy of about 95%. From the visual results, one can conclude that the prediction for the gas tank levels cannot be effectively solved by time series based analysis.

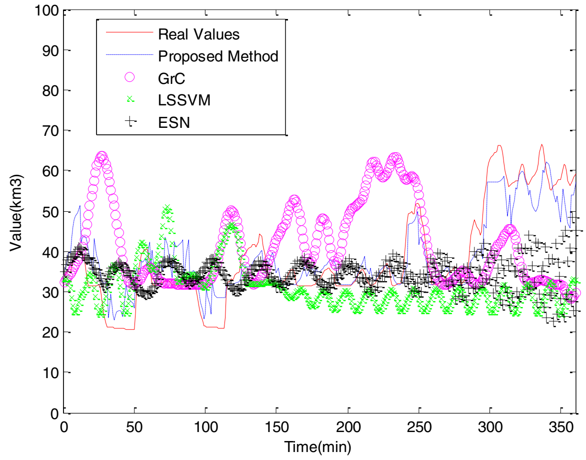

Fig.6. Prediction results of set 1 for gas tank #2

Table 2 Error and computing time statistics of set 1 for gas tank #1

Methods	MAPE	Root MSE	CT(s)
LSSVM	0.1782	15.8729	1.842
ESN	0.1395	13.1051	1.376
GrC (Time-Series model)	0.2078	17.2753	0.312
Proposed Method	0.0492	4.8868	0.204

Secondly, this study conducts the comparative experiments by employing the factors relationship based modelling. The alternatives for this comparison adopt also the above classes of data-based approaches. Similarly, the random selected results by using the data of the two tank levels are illustrated as Fig.7 and Fig.8, respectively. It is noticeable from Fig.7 that the LSSVM performs well at the very beginning period, but with the time goes on, the results tend to be far from the real values owing to the iterative errors. While, the results of ESN are not well in nearly 360 minutes. In comparison, the proposed method contains no iteration while modelling, which can successfully predict the trend of the tank levels in the future period of 360 points within less computing cost. As for Fig.8, it can be seen that LSSVM presents frequently unstable situations, which also fails to track the real values sometimes. While, although ESN acts a bit better than LSSVM, it still has to be low accuracy compared to the proposed method. And, Table 4 also presents the quantitative statistics of the results, which indicate that the effectiveness and the efficiency of the proposed method can completely satisfy the requirements for energy optimization and scheduling.

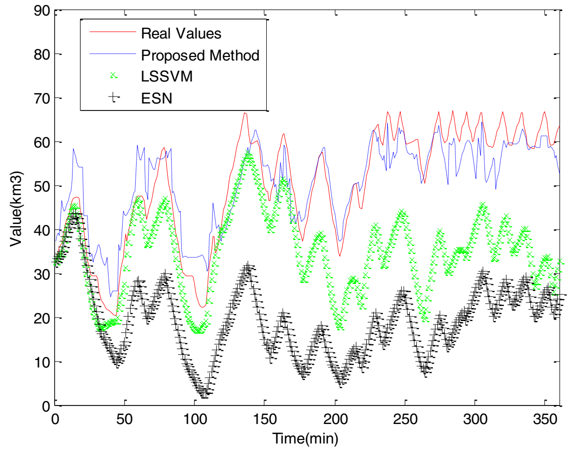

Fig.7. Prediction results of set 2 for gas tank #1

Table 3 Error and computing time statistics of set 2 for gas tank #1

Methods	MAPE	Root MSE	CT(s)
LSSVM	0.2278	18.3104	1.325
ESN	0.4608	33.1793	0.981
Proposed Method	0.0659	6.5440	0.063

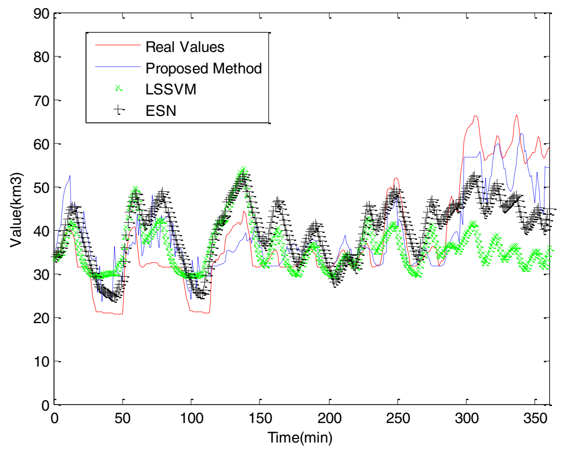

Fig.8 Prediction results of set 2 for gas tank #2

Table 4 Error and computing time statistics of set 2 for gas tank #2

Methods	MAPE	Root MSE	CT(s)
LSSVM	0.1275	11.7726	1.421
ESN	0.1132	9.7757	1.014
Proposed Method	0.0492	4.8868	0.062

5. CONCLUSIONS

Since it is very significant for LDG system in steel industry to acquire the variable values or the trends of the gas tank levels in advance, this study proposes a class of long-term prediction method based on granular computing concept, which can effectively avoid the iterative error and accelerate the predicting process. By using the granular computing, the prediction horizon is largely extended to a few hours, which greatly satisfy the industrial prediction demands for the energy optimization and decision making. To verify the performance of the proposed approach, the practical data coming from the energy data center of a steel plant in China are employed, and the experimental results indicate that the proposed factors relationship based modelling provides the remarkable performance and computing cost for the industrial application.

ACKNOWLEDGEMENTS

This work is supported by National Natural Sciences Fundation of China (61104157,61034003,61273037), National High-Tech R&D Program (2013AA040703), and the Fundamental Research Funds of China for the Central Universities (No. DUT12ZD214). The cooperation of the energy center of Baosteel Co. Ltd, China is greatly appreciated.

REFERENCES

Bagiela, A., Pedrycz, W. (2003). *Granular computing: an introduction*. Springer, USA.

Bargiela A., Pedrycz W (2008). Toward a theory of granular computing for human-centered information processing. *IEEE Transactions on Fuzzy Systems*, 16(2), 320-330.

Dattorro J. (2005). *Convex optimization and Euclidean distance geometry*. Meboo Publishing, USA.

Dong R., Pedrycz W. (2008). A granular time series approach to long-term forecasting and trend forecasting. *Physica A: Statistical Mechanics and its Applications*, 387(13), 3253-3270.

Han Z.Y., Liu Y., Zhao J., Wang Wei. (2012). Real time prediction for converter gas tank levels based on multi-output least square support vector regressor. *Control Engineering Practice*, 20(12), 1400-1409

Lin T. Y. (1999). Granular computing: Fuzzy logic and rough sets. *Computing with Words in Information/Intelligent Systems 1*, 183-200. Physica-Verlag HD, Heidelberg.

Kasabov N. K., Song Q. (2002). DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *IEEE Transactions on Fuzzy Systems*, 10(2), 144-154.

Kumar P., Jayaraman V.K., Kulkarni B.D. (2007). Granular support vector machine based method for prediction of solubility of proteins on overexpression in escherichia coli. *Pattern Recognition and Machine Intelligence*, 406-415. Springer, Berlin Heidelberg.

Pedrycz W., Bargiela A. (2002). Granular Clustering: A Granular Signature of Data. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 32(2), 212-224.

Sheng C., Zhao J., Wang W., Leung H. (2013). Prediction Intervals for a Noisy Nonlinear Time Series Based on a Bootstrapping Reservoir Computing Network Ensemble. *IEEE Transactions on Neural Networks and Learning Systems*, 24(7), 1036-1048

Wu F.M., Li Y., and Yu F.S. (2010). Fuzzy Granulation Based Forecasting of Time Series. *Advances in Intelligent and Soft Computing*, 511-520. Springer, Berlin Heidelberg.

Siler W., Buckley J. J. (2005). *Fuzzy expert systems and fuzzy reasoning*. Wiley. com, USA.

Zhang X.P., Zhao J., Wang W., Cong L. Q., Feng W. M., Chen W. C. (2010). COG holder level prediction model based on least square support vector machine and its application. *Control and Decision*, 25(8), 1178-1183.

Zhao J., Wang W., Liu Y., Pedrycz W. (2011). A two-stage online prediction method for a blast furnace gas system and its application. *IEEE Transactions on Control Systems Technology*, 19(3), 507-520.

Zhao J., Liu Y., Zhang X., Wang W. (2012). A MKL based on-line prediction for gasholder level in steel industry. *Control Engineering Practice*, 20(6), 629–641.

Zhao J., Liu Q., Wang W., Pedrycz W., Cong L.Q. (2012). Hybrid neural prediction and optimized adjustment for coke oven gas system in steel industry. *IEEE Transactions on Neural Networks and Learning Systems*, 23(3), 439-450.