Full Length Research Paper

Near-infrared spectra quantitative analysis for flue gas of thermal power plant based on wavelength selection

Hui Cao[1] and Yan Zhou[2]*

[1]State Key Laboratory of Electrical Insulation and Power Equipment, School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, 710049, China.
[2]School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an 710049, China.

This paper proposed a near-infrared (NIR) spectra quantitative analysis method for flue gas of thermal power plant based on wavelength selection. For the proposed method, the self-adaptive accelerated particle swarm optimization is presented for determining the most representative wavelengths of NIR spectral signals and is combined with partial least square for predicting the various contents of the real flue gas dataset. The proposed method chooses the current own optimal or the current global optimal as the reference state randomly and accelerated updates of the flight velocity by the reference state, then the particle state is updated based on the new velocity self-adaptively. The experimental results of a real flue gas dataset verified that the proposed method has higher predictive ability and could overcome the premature convergence.

Key words: Thermal power plant, fuel gas, near-infrared spectrum, wavelength selection, self-adaptive accelerated discrete particle swarm optimization.

INTRODUCTION

With regards to energy saving in a serious situation of coal and power energy shortages, using nature gas could decrease the power generation cost (Lee and Jou, 2012). The flue gas from the gas generating unit mainly consists of methane (CH_4), carbon monoxide (CO) and carbon dioxide (CO_2). The quantitative analysis of the flue gas could reflect the potential level of environmental pollution. The traditional chemical calibration analysis methods are usually time-expensive. Near-infrared (NIR) spectra is a fast nondestructive technique and has been widely used for analyte quantitative determination (Lillhonga and Geladi, 2011). Partial least square (PLS) is the most popular chemometric method and has been developed

for quantitative analysis of NIR spectral data (Marcio et al., 2011). Since not all wavelengths in the spectrum are equally important model, the wavelength selection is the crucial step of NIR spectroscopy analysis. The most popular wavelength selection method is the uninformative variable elimination by PLS (UVE-PLS) and are better than the statistic-based wavelength selection methods (Centner et al., 1996). Moreover, wavelength selection could be deemed as a combinatorial problem, genetic algorithm (GA) and particle swarm optimization (PSO) combined with PLS are proposed for wavelength selection (Arakawa et al., 2011; Sorol et al., 2010). Nevertheless, GA and PSO may fall into the local minima.

In this paper, we proposed the self-adaptive accelerated particle swarm optimization (SAAPSO) to realize the wavelength selection of NIR spectral for building the analyte quantitative model.

THE PROPOSED METHOD

A wavelength selection method based on SAAPSO is presented for building the concentration prediction model of flue gas. Some notations would be explained beforehand. The number of particles in the population is n and the number of dimensions of each particle is d, namely, d equals the number of the total wavelengths. v_{ij}^k and x_{ij}^k are the velocity and the position of the jth dimension of the ith particle at generation k, where $i \in \{1, 2, \cdots, n\}$, $j \in \{1, 2, \cdots, d\}$ and k is an integer. Moreover, p_i and p_g are the best previous performing state of the ith particle and the global best state, respectively.

For SAAPSO, v_{ij}^k could be updated by $v_{ij}^{k+1} = w \cdot v_{ij}^k + a \cdot r_2 \cdot (p_c - x_{ij}^k)$, where if $0 \le r_1 < f_c$, then $p_c = p_{ij}$, and if $f_c \le r_1 \le 1$, then $p_c = p_{gj}$, p_c is the reference state, r_1 and r_2 are both random number between 0 and 1, p_{ij} and p_{gj} are the positions of the jth dimension of p_i and p_g, respectively, f_c is the selection coefficient, w is the inertia factor and a is the acceleration factor. Then, x_{ij}^k could be updated by the formula (1):

$$x_{ij}^{k+1} = \begin{cases} p_{ij} & \rho < 1/\left(1 + \exp(-v_{ij}^{k+1})\right) \\ p_{gj} & \text{other} \end{cases} \tag{1}$$

where ρ is the random number between 0 and 1.

For wavelength selection, SAAPSO uses the binary coding to encode the location of each particle and each dimension represents a wavelength, where '1' or '0' indicates that the wavelength is selected or dropped, respectively. The velocity and location of each particle are initialized randomly, and then the prediction model is built by PLS with the selected wavelengths. In addition, the root-mean-squares error of cross-validation (RMSECV) is adopted as the fitness function. Aiming at minimizing the RMSECV, the optimal individual of SAAPSO is the solution of the problem after iterations.

For SAAPSO, the search space could be expanded and the acceleration factor would ensure the convergence rate. Furthermore, the new location determined self-adaptively would overcome the local optimum. In the next section, the experiments results could further verify the effectiveness of the proposed method.

EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed method, a real dataset obtained by measuring the NIR spectra of the field flue gas is used in the experiments. The dataset are obtained during a combustion process and includes 106 samples and each sample consists of a spectrum for a mixture of CH_4, CO and CO_2 with the reference concentration obtained a gas chromatography (GC). The spectra were measured by a GASMET DX4000 Fourier transform infrared (FTIR) gas analyzer. The spectral wave number is 549.44~4238.28 cm^{-1} with a resolution of 7.72 cm^{-1}. Each spectrum contains 473 wavelengths. The concentration ranges of the three analytes are 0~0.4598 ppm, 0~0.4083 ppm and 0~0.3818 ppm, respectively.

In this study, PLS, UVE-PLS, GA combined with PLS (GA-PLS), PSO combined with PLS (PSO-PLS), the neighboring combined with PLS (NPSO-PLS) and SAAPSO combined with PLS (SAAPSO-PLS) are used for the concentrations of CH_4, CO and CO_2 in the gas dataset and compared the effectiveness of these models. The real dataset would be split into the calibration set and the validation set based on the Monte Carlo cross-validation, namely, 80% of the samples in the dataset are randomly selected as the calibration set and the other 20% are regarded as the validation set. To ensure the fairness of the experiment, the size of populations of GA-PLS, PSO-PLS, NPSO-PLS and SAAPSO-PLS were all set to be 30 and the number of iterations is 100. Although GA-PLS, PSO-PLS, NPSO-PLS and SAAPSO-PLS are initialized randomly, the initialization conditions of them are kept consistent for the experiments. For GA, the crossover probability and the mutation probability were set to be 0.7 and 0.05, respectively. For PSO-PLS and NPSO-PLS, two learning factors were both 1.7. For PSO-PLS, NPSO-PLS and SAAPSO-PLS, the inertia factors were all set to be 0.7. For NPSO-PLS, the number of neighbors of particle is 5. For SAAPSO-PLS, the selection coefficient is 0.5 and the acceleration factor is 2. Since these methods are based on PLS, the number of latent variables is determined according to the RMSECV with leave-one-out cross validation. For UVE-PLS, the latent variables are selected for each recalculation.

Similarly, for GA-PLS, PSO-PLS, NPSO-PLS and SAAPSO-PLS, the latent variables are re-determined at each iteration. The RMSECV for the calibration set, the root mean-squared error of prediction (RMSEP) for the validation set, the determination correlation coefficient (R^2), the cross-validation correlation coefficient calculated by with leave-one-out cross validation for the calibration set (R^2_{cv}), the squared correlation coefficient for the validation set (R^2_p), the number of latent variables (N_{lv}) and the compression ratio (CR) are used to assess and compare the predictive ability of the various models, where CR equals $\left(N_t - N_s\right)/N_t \times 100\%$, N_t is the number of total wavelengths and N_s is the number of the selected wavelengths. The experiments are implemented in MATLAB 7.0.4 on a personal computer with an Intel i5-2300 CPU and 3 GB of RAM.

The analytical results are summarized in Table 1. For SAAPSO-PLS, the predictive ability is higher and the compression ratio is largest all alone for different contents. For CH4, R^2 of SAAPSO-PLS are little less than those of UVE-PLS, GA-PLS, PSO-PLS and NPSO-PLS. For CO2, R^2 of SAAPSO-PLS are little less than those of PLS, GA-PLS, PSO-PLS, NPSO-PLS, R^2_{cv} of SAAPSO-PLS are

Table 1. Analytical results for flue gas dataset.

CH_4	PLS	UVE-PLS	GA-PLS	PSO-PLS	NPSO-PLS	SAAPSO-PLS
RMSECV	103.34	74.346	57.275	59.799	63.422	55.194
RMSEP	132.02	145.54	126.64	132.27	138.6	114.82
R^2	0.9975	0.9999	0.9998	0.9998	0.9998	0.9996
R^2_{cv}	0.9955	0.9957	0.9986	0.9985	0.9982	0.9987
R^2_p	0.9908	0.9907	0.9941	0.9912	0.9905	0.9942
CR(%)	0	57.505	46.723	50.74	47.569	82.241
N_{lv}	8	26	28	25	27	19

CO	PLS	UVE-PLS	GA-PLS	PSO-PLS	NPSO-PLS	SAAPSO-PLS
RMSECV	61.055	30.041	40.409	39.208	42.101	18.85
RMSEP	95.483	42.005	74.162	78.901	69.259	36.562
R^2	0.9993	0.9999	0.9994	0.9995	0.9991	0.9999
R^2_{cv}	0.9906	0.9977	0.9959	0.9961	0.9955	0.9991
R^2_p	0.9751	0.9954	0.9844	0.9826	0.9865	0.9963
CR(%)	0	82.875	53.066	46.934	48.203	84.355
N_{lv}	21	28	27	28	24	25

CO_2	PLS	UVE-PLS	GA-PLS	PSO-PLS	NPSO-PLS	SAAPSO-PLS
RMSECV	44.648	36.263	31.362	33.454	34.66	30.607
RMSEP	49.167	41.326	41.982	44.579	43.242	41.241
R^2	0.9996	0.9994	0.9996	0.9995	0.9996	0.9995
R^2_{cv}	0.9984	0.9989	0.9992	0.9991	0.9990	0.9991
R^2_p	0.9955	0.9970	0.9971	0.9965	0.9969	0.9970
CR(%)	0	76.321	52.854	48.203	51.163	78.647
N_{lv}	12	12	12	11	12	11

little less than those of GA-PLS and PSO-PLS and R^2_p of SAAPSO-PLS is only less than that of GA-PLS. The prediction accuracy of SAAPSO-PLS is higher.

Figure 1(a) to (c) show the scatter plots of measured value vs. predicted value of SAAPSO-PLS for CH_4, CO and CO_2 respectively. Because the data almost are positioned around the diagonal line, the prediction accuracy of SAAPSO-PLS model is high. Figure 1(d) and (f) show the iterations curves of GA-PLS, PSO-PLS, NPSO-PLS and SAAPSO-PLS for CH_4, CO and CO_2, respectively. At the beginning of iteration, SAAPSO-PLS would explore the new search area by the reference state determined randomly and the search efficiency would be limited in a certain extent. Nevertheless, because of including the acceleration mechanism, SAAPSO-PLS begins to converge after several iterations and has the fastest convergence rate overall. In summary, the experiments results verify that SAAPSO-PLS could be adopted for quantitative analysis of real NIR spectral data successfully and has a powerful predictive ability.

Conclusion

representative wavelengths and is combined with PLS for

predicting the various contents of the real flue gas of thermal power plant. The proposed method has some advantages as follows. First, the prediction model of the various contents of flue gas dataset built by the proposed method is more accurate. Second, the proposed method explores new areas to avoid the premature convergence in a certain extent. Third, the proposed method accelerated finds the optimal solution with faster convergence speed. The experiments results verify that the predicative ability and the search performance of the proposed method.

Conflict of Interests

The author(s) have not declared any conflict of interests.

ACKNOWLEDGEMENT
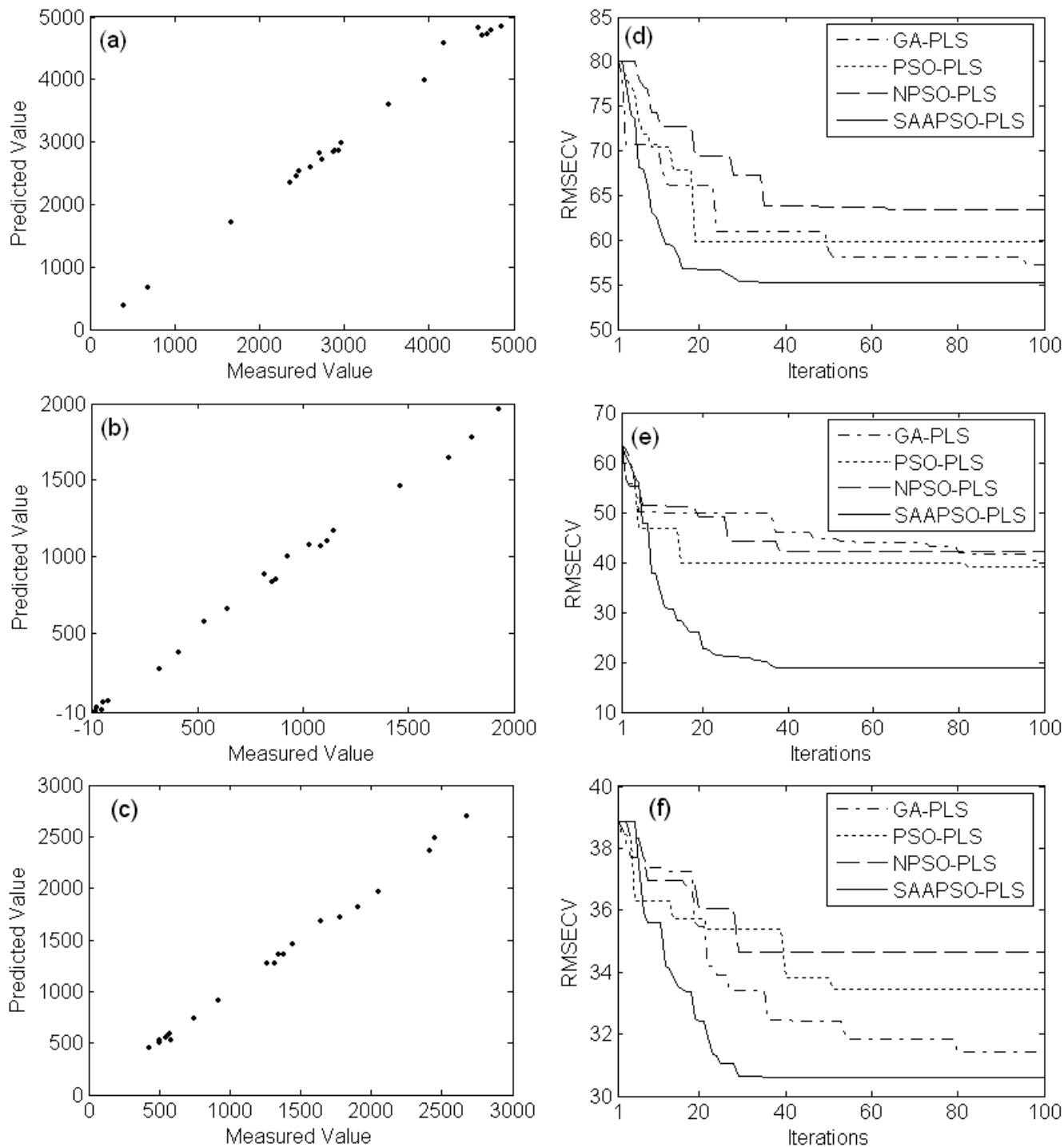
Figure 1. Scatter plot of measured value vs. predicted value and iterations curves for flue gas dataset.

In the paper, SAAPSO is presented for determining the University.

REFRENCES

Lee CL, Jou CJG (2012). Integrated methods to improve efficiency of furnace burning recovered tail gas. Int. J. Hydrogen Energy. 37(8):6620-6625. http://dx.doi.org/10.1016/j.ijhydene.2012.01.027

Lillhonga T, Geladi P (2011). Three-way analysis of a designed compost experiment using near-infrared spectroscopy and laboratory measurements. J. Chemometrics 25(4):193-20. http://dx.doi.org/10.1002/cem.1371

Marcio VR, Jamile BS, Maria FP, Leonardo SGT (2011). A novel approach for development of a multivariate calibration model using a

Doehlert experimental design: Application for prediction of key gasoline properties by near-infrared spectroscopy, Chemometrics and Intelligent Laboratory Systems 107(1):85-193.

Centneret V, Désiré-Luc M, Onno EN, Sijmen de J, Bernard MV, Cécile S (1996). Elimination of uninformative variables for multivariate calibration. Anal. Chem. 68(21):3851-3858. http://dx.doi.org/10.1021/ac960321m; PMid:21619260

Arakawa M, Yosuke Y, Kimito F (2011). Genetic algorithm-based wavelength selection method for spectral calibration. J. Chemometrics 25(1):10-19. http://dx.doi.org/10.1002/cem.1339

Sorol N, Eleuterio A, Santiago AB, Alejandro CO (2010). Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice: A test field for variable selection methods, Chemometrics and Intelligent Laboratory Systems 102(2):100-109. http://dx.doi.org/10.1016/j.chemolab.2010.04.009