19th European Symposium on Computer Aided Process Engineering – ESCAPE19
J. Jeżowski and J. Thullie (Editors)

Inferential estimation of kerosene dry point in refineries with varying crudes

Chang Zhou[a], Qiyue Liu[a], Dexian Huang[a,c], Jie Zhang[b]

[a]*Department of Automation, Tsinghua University, Beijing 100084, China.*
[b]*School of Chemical Engineering and Advanced Materials, Newcastle University, Newcastle upon Tyne NE1 7RU, UK, E-mail: jie.zhang@newcastle.ac.uk*
[c] *TNList, Beijing 100084, China*

Abstract

A bootstrap aggregated model approach to the estimation of kerosene dry point in refineries with varying crudes is proposed in this paper. Using on-line measurements of process variables, the feed crude oil is classified into one of the three types: with more light component, with more middle component, and with more heavy component. Bootstrap aggregated neural networks are used in developing the on-line crude oil classifier. Historical process operation data are classified into these three groups. A bootstrap aggregated partial least square (PLS) regression model is developed for each data group corresponding to each type of feed crude oil. During on-line operation, the feed crude oil type is first estimated from on-line process measurements and then the corresponding bootstrap aggregated PLS model is invoked. The overall inferential estimation performance of the bootstrap aggregated PLS estimator integrated with feed crude oil classifier is significantly enhanced.

Keywords: Bootstrap aggregated model, Data-driven models, Inferential estimation, Crude distillation

1. Introduction

Crude oil distillation is a primary process in petro-chemical industry and its operation determines the resource usage efficiency and economic benefits of refineries. In order to properly control refinery operations, it is essential that product quality measurements are available. Since most of the quality indexes can hardly be measured in real-time, various soft-sensor methods have been proposed to estimate these indexes using measurable process variables and have been successfully applied in practice [1, 2]. However, soft-sensing in crude oil distillation with feedstock changes remains a difficult problem because the relationship between the easy to measure process variables and the difficult to measure quality variables varies with the type of crude oil used. Many refineries are operated with mixed sources of crude oil. One possible solution is to develop an inferential estimation model for each type of crude oil. However, this will require many models and, furthermore, crude oil from the same supplier may also vary in the hydrocarbon content. In order to address this problem, this paper proposes a multi-model inferential estimation strategy integrated with crude oil classification for the estimation of kerosene dry point in refineries with varying crudes.

Using on-line measurements of process variables, the feed crude oil is classified into one of the three types: with more light component, with more middle component, and with more heavy component. Historical process operation data are classified into these three groups. A bootstrap aggregated partial least square (PLS) regression model is developed for each data group corresponding to each type of feed crude oil. Each model

has a favourable predictive ability upon the same type of oil but low predictive accuracy upon other types. During on-line operation, the feed crude oil type is first estimated from on-line process measurements and then the corresponding PLS model is invoked.
In the crude oil classifier development, bootstrap aggregated neural networks are used. The inputs to the crude oil classifier are the ratios between product and feed rates. Since the relationship between the classifier inputs and output is nonlinear, a nonlinear model has to be developed using process operational data. Through bootstrap re-sampling of the training datasets, multiple neural network models are developed based on bootstrap re-sampled datasets. A bootstrap aggregated neural network shows better accuracy and generalization capability than a single neural network which can be trapped in a local minimum or over-fit the training data during network training. The overall inferential estimation performance of the bootstrap aggregated PLS estimator integrated with feed crude oil classifier gives much better performance than various single PLS estimators.
The paper is organised as follows. Section 2 presents a simulated atmospheric distillation column in a refinery with varying feed crude oil. On-line crude oil type classification is presented in Section 3. Section 4 presents inferential estimation of kerosene dry point using bootstrap aggregated PLS models integrated with on-line crude oil classification. Some concluding remarks are given in Section 5.

2. A simulated atmospheric distillation column in a refinery with varying feed crude oil

The techniques developed in this paper are tested on a simulated refinery with varying feed crude oil. The simulation is carried out in the HYSYS environment. Fig. 1 shows a schematic diagram of an atmospheric distillation column which is one of the major units used in refineries. Three kinds of crude oil are used as varying feed: light, middle and heavy, each of which is produced by setting different assay values to make different hypothetical components that compose the crude oil. A MATLAB programme is used to change the operation condition in a random way to approximate the real industrial process and get a large number of process data automatically. Also, the product flows are carefully set to ensure that product quality constraints are met.

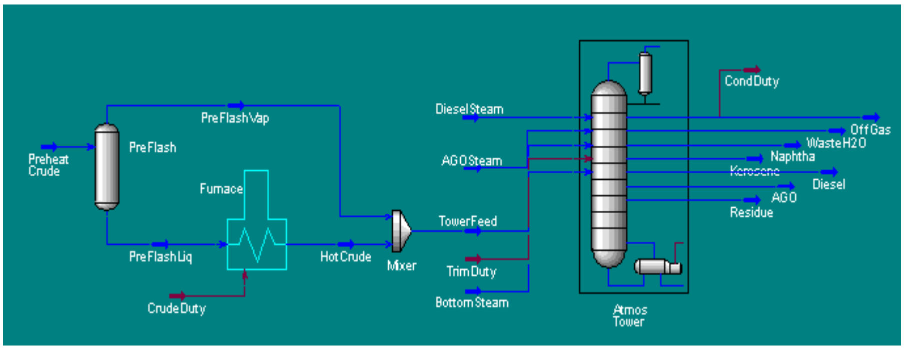

Fig. 1 A HYSYS flow diagram of an atmospheric distillation column

3. Crude oil classification using bootstrap aggregated neural networks

Using on-line measurements of process variables, the feed crude oil is classified into one of the three types: with more light component, with more middle component, and with more heavy component. Fig. 2 shows the relationship between feed crude types and the ratios between products and feed. Fig. 2 indicates that the three ratios could be used in classifying the feed crude oil.

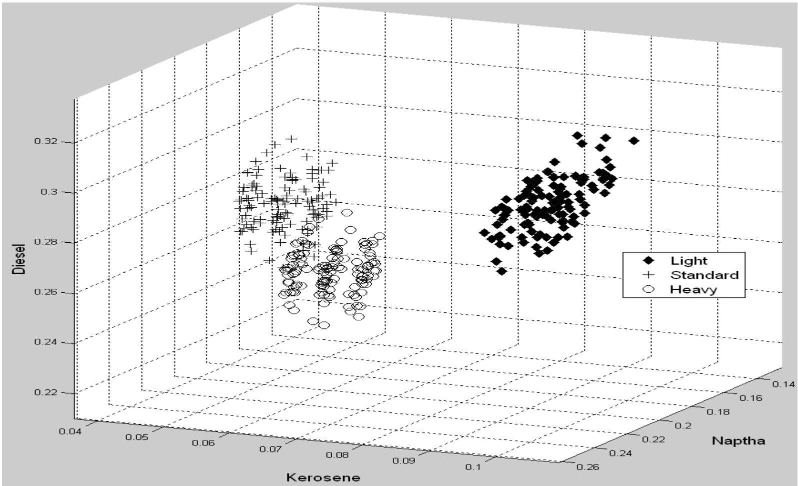

Fig. 2 Relationship between feed crude types and the ratios between products and feed

A linear classifier was first developed and the classification accuracy on the testing data is given in Table 1. It can be seen that the classification accuracy is not very high for the middle and heavy oil. This is due to the fact that the two classes are not linearly separable as indicated in Fig. 2. Thus a nonlinear classifier needs to be developed.

Bootstrap aggregated neural networks are used in developing a crude oil classifier. Fig. 3 shows a bootstrap aggregated neural network where several neural networks are developed to model the same relationship and are combined together. The individual networks are developed on data sets obtained from bootstrap re-sampling of the original training data [3]. Earlier studies show that an advantage of stacked neural networks is that they can not only give better generalisation performance than single neural networks, but also provide model prediction confidence measures [4]. The aggregated neural network output is given by:

$$f(X) = \sum_{i=1}^{n} w_i f_i(X) \tag{1}$$

where $f(X)$ is the aggregated neural network predictor, $f_i(X)$ is the ith neural network, w_i is the aggregating weight for combining the ith neural network, n is the number of neural networks, and X is a vector of neural network inputs.

For the purpose of comparison, a single neural network based classifier is also developed. Table 1 shows the classification accuracy of different classifier. It can be seen that bootstrap aggregated neural network classifier gives the best classification accuracy while the linear classifier gives the worst performance.

In order to demonstrate the robustness of bootstrap aggregated neural networks, five different netwrok combination schemes listed in Table 2 were studied. The experiments were repeated 20 times with different bootstrap resamples generated. Fig. 4 shows the classification accuracy and their 95% confidence bounds. Fig. 4 clearly indicates that bootstrap aggregated neural networks give much accurate and reliable (with narrower confidence bounds) classifications.

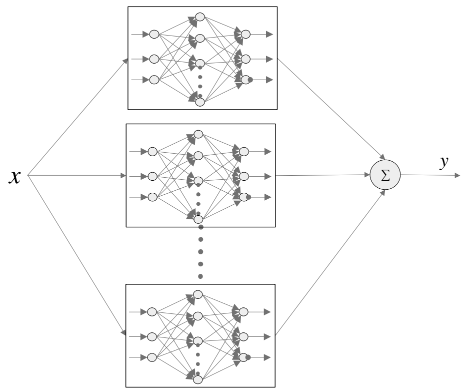

Fig. 3 A bootstrap aggregated neural network

Table 1. Classification accuracy on testing data

	Light	Middle	Heavy	All
Linear	100%	88.22%	97.81%	95.34%
Single network	100%	89.04%	99.18%	96.07%
Aggregated network	100%	90.96%	98.63%	96.53%

Table 2. Network combination schemes

1	Single neural network
2	Median of 20 neural networks
3	Median of 10 neural networks with better performance on training data
4	Average of 20 neural networks
5	Average of 10 neural networks with better performance on training data

4. Inferential estimation with bootstrap aggregated PLS models

Three PLS inferential estimation models, corresponding to the three types of crude oil, were developed. PLS is a powerful modelling technique for situations where the predictors are correlated [5]. The prediction performance of a PLS or PCR (principal component regression) model on unseen data is highly influenced by the number of latent variables or principal components retained in the model. Bootstrap aggregated PCR or PLS models have been shown to be an effective way to obtain robust PCR or PLS models [6].

In the bootstrap aggregated PLS model, 20 re-sampled datasets are produced through bootstrap re-sampling of the training datasets. And 20 different PLS model are developed based on the re-sampled data sets. As the industrial process condition is always changing (i.e. the varying crude), it is hoped that one or more of the models based on the re-sampled datasets is able to reflect the currents condition better than the original training datasets.

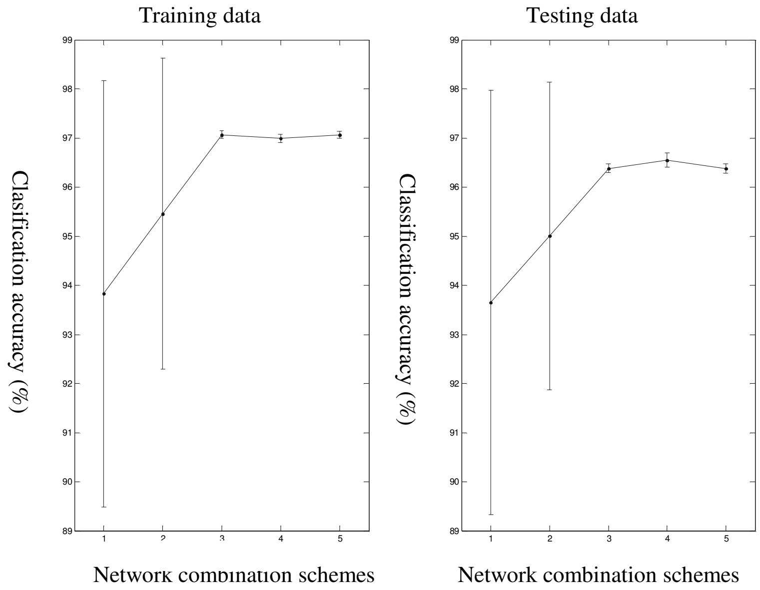

Fig. 4 Classification accuracies and their 95% confidence bounds

The inferential estimation model uses the following 16 measured process variables as its inputs: top stage temperature, diesel temperature, AGO temperature, kerosene temperature, preheat crude flow rate, reflux temperature, feed temperature, five product and feed flow rate ratios, reflux ratio, and three middle draw heat ratios. Simulated data were obtained using the HYSIS simulation described in Section 2. For building each of the PLS models corresponding to the three types of feed crude oil, the training data set contains 50 sample and the testing data set contains 400 samples. The reason for using a small training data set is from a practical consideration that the laboratory analysis data for kerosene dry point is usually limited.

Table 3 shows the root mean squared errors (RMSE) of the developed models on the testing data, using both single PLS and bootstrap aggregated PLS. In Table 2, models I to III are developed using light oil data, middle oil data, and heavy oil data respectively. The results reveal that bootstrap aggregated PLS performs better on the whole, especially when a model is used on other types of crude oil (e.g. model II used on Light dataset). Thus the bootstrap aggregated PLS model will not give significantly large errors when the crude oil is occasionally misclassified.

5. Conclusions

A bootstrap aggregated PLS model inferential estimation approach integrated with on-line crude classification is developed for kerosene dry point estimation. Bootstrap aggregated neural networks are used to on-line classify the feed crude oil into three types using the ratios between on-line measured product and feed rates. It is shown that bootstrap aggregated neural network gives better classification accuracy than a single neural network. A bootstrap aggregated PLS inferential estimation model is developed for each type of crude oil. The results demonstrate that the bootstrap aggregated PLS model inferential estimation approach gives better performance than a single estimation model.

Table 3. RMSE on testing data

Models	I		II		III	
	single	multiple	single	multiple	single	multiple
Light	2.38	2.28	93.40	12.91	89.89	15.82
Middle	45.49	31.08	2.44	2.47	268.04	89.15
Heavy	187.63	41.32	230.16	59.95	2.38	2.24
All	111.48	53.47	143.41	42.24	163.23	63.90

Acknowledgements

The research is supported by National High-tech 863 Program of China (2006AA04Z168 and 2007AA04Z193) and UK Department for Innovation, Universities and Skills under the UK/China Fellowship for Excellence programme.

References

[1] J. Zhang, Chemical Engineering & Technology, 29 (2006) 560-566.
[2] Y. Li, Q. Li, H. Wang, and N. Ma (1995). Particle Soft Sensing Based on LS-SVM and Its Application to a Distillation Column. in Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06), Jinan, China. Vol 1, 177-182.
[3] B. Efron, The Jackknife, the Bootstrap and Other Resampling Plans; Society for Industrial and Applied Mathematics: Philadelphia, 1982.
[4] J. Zhang, E.B. Martin, A.J. Morris, & C. Kiparissides, Computers & Chemical Engineering, 21 (1997) s1025-s1030.
[5] P. Geladi and B. R. Kowalski, Analytica Chimica Acta, 185 (1986), 1-17.
[6] M. Ahmed and J. Zhang (2003). Improved Inferential Feedback Control through Combining Multiple PCR Models, in Proceedings of The 2003 IEEE International Symposium on Intelligent Control, Houston, Texas, U.S.A., 878-883.