

A Novel Identification Method of Two Phase Flow Based on LDA Feature Extraction and GRNN in ERT System

YanJun Zhang^{1,2}

¹ Instrument Science and Technology, post-doctoral research station, Harbin University of Science and Technology, China

² The school of Information Science and Technology, Heilongjiang University, 210#, Harbin, China
zyj716@hotmail

Abstract—Two-phase flow measurement plays an increasingly important role in the real-time, on-line control of industrial processes including fault detection and system malfunction. The flow regime parameter is one of the most important parameters in measurements. This paper proposes a new identification approach for common two phase flow regimes based on Electrical Tomography measurement. LDA feature extraction was employed to extract feature vectors. GRNN was used to train identify the flow regime models. Simulation was carried out for typical flow regimes using the approach. The results show its feasibility, and the results indicate that this method is fast in speed and can identify these flow regimes correctly.

Keywords-electrical resistance tomography; flow regime identification; linear discriminant analysis; general regression neural network.

I. INTRODUCTION

Two-phase flow is a mixed-flow pattern widely found in nature, especially in the chemical, petroleum, electricity, nuclear power and metallurgical industries^[1]. Two-phase flow regime identification plays an increasingly important role in the automation process of energy industry. It can provide valuable information for a rapid and dynamic response which facilitates the real-time, on-line control of processes including fault detection and system malfunction, many experimental and theoretical researches have done in the field of image reconstruction. However, the visualization process cost quite long time so that there are number of challenges in the real applications. An alternative approach to monitor two-phase flow inside a pipe/vessel is to take advantage of identification of flow regimes instead of image reconstruction, especially for those stable and simple Two-phase flows. Two-phase flow regimes not only affect the two-phase flow characteristics and mass transfer, heat transfer performance, but also affect the system operation efficiency and reliability, while the other parameters of the two-phase flow measurements have a great impact. Therefore, on-line identification of two-phase flow regimes are important for oil mixed transportation systems.

In this paper, a Linear Discriminant Analysis was used to extract feature from original data. LDA is closely related to

principal component analysis (PCA).they both look for linear combinations of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. On the other hand, PCA does not take into account any difference in class In the training and identification stages, GRNN was employed.

II. THE FEATURE EXTRACTION BASED ON LDA

This paper uses 12-electrode sensor array, the incentive is in any two adjacent electrodes between the one electrode to the excitation current of the input and one for output, then cycle all the two adjacent electrodes detect electrical potential difference (in addition to two other excitation electrode) can be nine voltage data; is to inspire the next cycle of the electrode, changing the incentives electrode, and can detect the 9 voltage data; to any two adjacent electrodes are incentives once a week to encourage voltage data obtained 12 groups, so that each image acquisition to have $12 * 9 = 108$ elements, the number of independent elements is 54.

Then there is the sample set:

$$V = \{v_1, v_2, \dots, v_n\} \quad (1)$$

Where, n is the index of a flow regime sample, v_i is composed of 12 vectors, each vector have 9 voltage values.

In this paper, core flow, laminar flow, circulation flow and trickle flow are selected to be tested.

The first key step before training and recognition is feature extraction. Linear Discriminant Analysis is employed in this paper. LDA method is widely used in pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events^[2]. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification. LDA is also closely related to principle component analysis(PCA) and factor analysis in that they both look for linear combinations of variables which best explain the data^[3] LDA attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities.

It is supposed that there are M classes, each class has a mean μ_i and the same covariance, then between class variability may be defined by the sample covariance of the class means

$$S_b = \sum_{i=1}^C P_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (2)$$

Where, μ_i is the mean of C_i class means, μ is the mean of all samples, P_i is priori information.

$$S_w = \sum_{i=1}^C P_i S_i \quad (3)$$

Where,

$$S_i = E[(v - \mu_i)(v - \mu_i)^T | x \in C_i]$$

This paper used Fisher's linear discriminant and LDA interchangeably. According to Fisher Criterion Function

$$J(W_{opt}) = \arg \max_w \frac{|W^T S_b W|}{|W^T S_w W|} \quad (4)$$

Where, W_{opt} satisfies

$$S_b W_i = \lambda_i S_w W_i \quad (5)$$

Where, $i = 1, 2, \dots, m$

The class means and covariances are not known. They can, however, be estimated from the training set. Either the maximum likelihood estimate or the maximum a posteriori estimate may be used in place of the exact value in the above equations. Although the estimates of the covariance may be considered optimal in some sense, this does not mean that the resulting discriminant obtained by substituting these values is optimal in any sense, even if the assumption of normally distributed classes is correct. (1)

III. FLOW PATTERN IDENTIFICATION BASED ON GRNN

General Regression Neural Network (GRNN) was first proposed by the specht is a branch of RBF neural network is a nonlinear regression based on the theory of feed-forward neural network model^[4-6]. GRNN structure constitutes a general from the input layer, hidden layer and output layer. Input layer put variables into the hidden layer of the sample does not participate in a real operation. The number of hidden layer neurons is equal to the number of training samples, the weight of the layer functions as a Euclidean distance function. The transfer function of hidden layer is radial basis function, Gaussian function is usually used as the network's transfer function. The third layer of the network output layer is linear, the weight function is normalized dot product weight function.

Let the training sample input matrix T and output matrix P as follow:

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1Q} \\ p_{21} & p_{22} & \dots & p_{2Q} \\ \vdots & \vdots & \vdots & \vdots \\ p_{R1} & p_{R2} & \dots & p_{RQ} \end{bmatrix}, \quad T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{1Q} \\ t_{21} & t_{22} & \dots & t_{2Q} \\ \vdots & \vdots & \vdots & \vdots \\ t_{s1} & t_{s2} & \dots & t_{sQ} \end{bmatrix} \quad (12)$$

P is the training sample of input variables, T the training samples of the output variable, R is the dimension of input variables, S is the dimension of the output variable, Q is the training set samples. Q hidden layer neurons corresponding to the threshold as follows:

$$b_1 = [b_{11}, b_{12}, \dots, b_{1Q}] \quad (13)$$

Where $b_{11} = b_{12} = \dots = b_{1Q} = \frac{0.8326}{spread}$, spread is the pace

of expansion of the radial basis function. The output of hidden layer neurons have the following formula:

$$a^i = \exp(-\|c - p_i\|^2 b_1), i = 1, 2, \dots, Q \quad (14)$$

Where $p_i = [p_{i1}, p_{i2}, \dots, p_{i3}]$, $a^i = [a_1^i, a_2^i, \dots, a_Q^i]$.

The hidden layer and output layer connection weights W is taken as the output matrix for training set in GRNN, When hidden layer and output layer neuron connection weights determined, we can calculate the output of output layer neurons, with the following formula:

$$n^i = \frac{LW_{2,1} a^i}{\sum_{j=1}^Q a_j^i}, i = 1, 2, \dots, Q \quad (15)$$

$$y^i = \text{purelin}(n^i) = n^i, i = 1, 2, \dots, Q \quad (16)$$

IV. EXPERIMENTS AND CONCLUSIONS

The approach based on PCA and ELM identifier is tested on full pipe, empty pipe, laminar flow, trickle flow, the core flow and circulation flow, Samples of flow regimes used in the experiments are as follows:

1. Empty pipe and Full pipe

50 samples for empty pipe and full pipe were simulated respectively.

2. The core flow, laminar flow and circulation

Each flow regime was divided into three groups according to the fluid phase holdup. The core flow were classified into three groups according to fluid contents holdup 20%, 50%, 80% as the figure. 50 samples is trained for each group respectively.

3. Trickle each flow regime

Trickle Each flow regime was divided into according to the different phase holdup has also been divided into three groups, which contains 5% water, 10% and 20%, 50 samples were trained.

The simulations are carried out in MATLAB 6.5 environment running in a Intel Core 2, 2.93GHZ CPU. Experiments results are as Table 1.

The average accuracy rate is 97%. It is noted that the accuracy rate of identification of trickle flow is significantly

lower than other flow regimes, it can be explained by the complexity of trickle flow. It is observed that the fluid phase holdup is bigger, identification accuracy is lower for core flow, circulation flow, laminar flow and trickle flow. The average time-consuming of recognition is less than 65 second. The speed can satisfied with the real time control.

V. ACKNOWLEDGEMENTS

This work is partially funded by Postdoctoral Fund of Heilongjiang Province(LBH-Z11107).

TABLE I. THE IDENTIFICATION ACCURACY RATE

Flow Regime	Samples	TC for T \R /ms	Accuracy Rate
Empty pipe	50	4848\61	100%
Full pipe	50	4834\63	100%
Core 20%	50	4847\62	98%
Core 50%	50	4845\65	98%
Core 80%	50	4856\66	98%
circulation 20%	50	4853\64	100%
circulation 50%	50	4861\64	96%
circulation 80%	50	4852\65	96%

laminar 20%	50	4861\67	98%
laminar 50%	50	4853\65	96%
laminar 80%	50	4859\64	96%
Trickle5%	50	4852\64	94%
Trickle 15%	50	4858\65	94%
Trickle 20%	50	4849\65	94%

REFERENCES

- [1] Yanjun Zhang, Yu Chen, A Novel PCA-SVM Flow pattern Identification Algorithm For Electrical Resistance Tomography System, In Proceeding(s) of the Advances in Intelligent and Soft Computing , Volume 160, 2012.
- [2] Yu, H., Yang, J., A direct LDA algorithm for high-dimensional data with application to face recognition, Pattern Recognition, Volume 34 (10), pp.2067–2069.
- [3] Friedman, J. H. (1989). Regularized Discriminant Analysis. Journal of the American Statistical Association (American Statistical Association) 84 (405): 165–175.
- [4] Xing Xiaozhe, Luo Wei. State estimation in power system based on GRNN algorithm. J. Heilongjiang Electric Power. 33(1), 50-54 (2011).
- [5] Liu Zunxiong, Zhou Tianqing. : Application of GRNN Model Based on SSA in Financial Time Series. 28(2), 29-33(2011).
- [6] WANG Hongwei ,LIN Jianliang. : Sales Prediction Based on Improved GRNN. 32 (1), 153–157(2010).