Prediction of concentration of dispersed phase outlet in rotating disc contactor column

E F Azmi*, N H Miswan1, K A Arshad3 and N A Zainal4

1, 4 Department of Mechanical Engineering Technology, Faculty of Mechanical and Manufacturing Engineering Technology, Universiti Teknikal Malaysia Melaka, Malaysia.
2 Centre of Modelling and Data Science, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Malaysia.
3 Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, Malaysia.

*ez.farhain@utem.edu.my

Abstract. Rotating Disc Contactor (RDC) columns are one of the extractors that used for liquid-liquid extraction. It has an extensive application in various industries. The performances of these columns indicate that they are more efficient and possess better operational flexibility. However, there is still some improving that researchers can do to enhance the performances. This paper presents Support Vector Machine (SVM) and Neural Network (NN) modeling in prediction of concentration of dispersed phase outlet in RDC column. SVM is an exciting Machine Learning technique that learns by example to sign labels to object and can be used for regression as well as classification purpose, while NN is widely used as an effective approach for handling non-linear data especially in situations where the physical processes are not fully understood. The mean square error (MSE) is calculated to compare the result between the two models. The analysis shows that both SVM and NN modeling can predict the concentration of dispersed phase in RDC column but the SVM approach gives better result than the NN approach. Both modelling systems offer the potential for a more flexible and less error in forecasting. Thus, it can help to save time and reducing cost in conducting experiments.

Keywords: Neural Network, Rotating Disc Contactor, Support Vector Machine

1. Introduction

Extraction is a separation of something from another. Extraction process can be classified into four types which are liquid-liquid extraction, solid-liquid extraction, gas-liquid extraction and liquid-vapour extraction. For many years, liquid–liquid extraction or solvent extraction has become one of the key unit operations in process industries [1]. Furthermore, it is often used in the chemical and mining industries and in the downstream recovery of fermentation products (antibiotics, amino acids, steroids) [2]. In addition, liquid-liquid extraction equipment considers the rotating disc contactor (RDC) for refining of vegetable oils, processing of nuclear fuels, and refining of crude petroleum [3].

Application of modeling in RDC has become an interest among researchers in Mathematics and Chemical Engineering. A research uses the method of design of the experiments (DOE) and Multiple Linear Regression (MLR) to analyse the performance of small diameter column RDC using the chemical system involving cumene/ isobutyric asid/ water [4]. In addition, an image analysis
technique is applied to determine the drop size distribution as a function of operating parameters and physical properties [5].

Therefore, in this project, we introduce Support Vector Machine and Neural Network modelling to predict the output (concentration of dispersed phase outlet). We also compare the results from both of the modelling to see the effectiveness. The data is acquired from the researchers at the University of Bradford under contract to Separation Processes Service, AEA Technology, Harwell.

2. Data
The experimental data consists of 256 sets of data with four different inputs and one output. The four inputs data are rotor speed ($N_r$), dispersed phase flow rate ($F_d$), concentration of continuous inlet ($C_{cin}$) and concentration of dispersed inlet ($C_{din}$). These will be treated as independent variables to predict the dependent variable. Dependent variable is the output data named concentration of dispersed phase outlet ($C_{dout}$). A Statistica software is utilized to help with the SVM modelling and a Matlab code is produced to run the NN simulation in this project.

3. Methodology

3.1. Data normalization
The collection of data as explained in the materials section above are chosen for training and testing the SVM and NN model. Before continuing with any modelling steps, it is crucial to normalise the data. The data need to be normalised in order to avoid attributes in greater numerical ranges dominate value in smaller numerical rangers. The normalization process can be done using the following equation,

$$\frac{x}{x_{max}} \leq X_N \leq 1$$

where, $x$ is the value of actual data, $x_{max}$ is the maximum value of the actual data and $X_N$ is the value of data that has been normalised. Then, the data are partitioned into three groups which are train data and test data. The ratio of training and testing data are 0.75:0.25 where 192 data are chosen at random as the training and 64 data as testing. Next, the data are ready for SVM and NN modelling.

3.2. Support vector machine (SVM)
SVM is an exciting Machine Learning technique that was developed by Vladmir Vapnik and his co-workers in the middle of 90’s. It can be used for regression as well as classification purpose. The main idea is to select the line which separates the two or more classes with the maximal distance of different classes. SVM selects the maximum margin separating hyperplane by defining the distance from the separating hyperplane to the nearest expression vector as the margin of the hyperplane.

Any hyperplane can be written as the set of points $x$ satisfying:

$$w \cdot x + b = 0 \tag{2}$$

where the dot product are the points of data, $x$ and $w$ the normal vector to the hyperplane. To find the maximum margin separator, we have to solve the following optimization problem:

$$w \cdot x^c + b > +1 \text{ for positive cases} \tag{3}$$

$$w \cdot x^c + b < -1 \text{ for negative cases} \tag{4}$$

This is tricky but it is a convex problem. There is only one optimum and we can find it without fiddling with learning rates or weight decay or early stopping. Hence, if $w \cdot x^c + b > 0$, it’s a positive case while if $w \cdot x^c + b < 0$, it’s a negative case.

There are some of data sets that cannot be separated cleanly. This is because the data may contain an error. SVM is able to deal with this error if we allow a few irregular expression profiles to fall on the “wrong side” of the separating hyperplane. Therefore, SVM algorithm has to be modified by
adding a soft margin. This will let some of the data points to push their way through the margin of the separating hyperplane without affecting the final result. The soft margin works with the help of kernel function.

Kernel function is used to solve problems of non-linear separable data in SVM. Many studies in time series prediction have applied SVM with several selections of kernel functions due to the nonlinearity of the data. The algorithm has common results except that the algorithm fits the maximum-margin hyperplane by replacing every dot product with a nonlinear kernel function. There are various types of kernel functions used in SVM. Table 1 shows some of the common kernel functions used and their equations.

### Table 1. Common kernel functions

| Kernel Functions       | Equations                                                                 |
|------------------------|---------------------------------------------------------------------------|
| Linear                 | $K(x, y) = x \cdot y$                                                     |
| Polynomial             | $K(x, y) = (x \cdot y + 1)^p$                                             |
| Radial Basis Function  | $K(x, y) = \exp\left(\frac{||x - y||^2}{2\sigma^2}\right)$                |
| Sigmoid                | $K(x, y) = \tanh(kx \cdot y - \delta)$                                    |

However, there is no guarantee that one kernel functions will perform well in all kinds of datasets. The choices of kernel functions depend on the problems, parameters and scaling methods.

#### 3.3. Neural network (NN)

NN modeling is just like SVM as it can be introduced when there are problems involving prediction, classification or control. An interconnected group of natural or artificial neurons that uses a mathematical or computational model for information processing based on a connection approach to computation is called Neural Network. They can be used to model complex relationships between inputs and outputs or to find pattern in data. A functional model of the biological neuron has three basic components; weights, actual activity, and activation function.

Weights are modeled by the synapses of the neuron. The value of it represents the strength of the connection between an input and a neuron. Inhibitory connections reflected by negative weight values, while positive weight values assign excitatory connections. The next two components model is the actual activity within the neuron cell. Linear combination is the activity that sums up all the inputs modified by their respective weights. The amplitude of the output of the neuron will control by an activation function. Usually, an acceptable range of output is between 0 and 1, or -1 and 1. This process is described mathematically in figure 1 below.
The interval activity of the neuron from this model is given by the following equation,

\[ v_k = \sum_{j=1}^{p} w_{kj} x_j \]  

where \( w_{kj} \) is the weight and \( x_j \) is the input. Therefore, the outcome of some activation function on the value of \( v_k \) is the output of the neuron, \( y_k \).

One process of training NN from input to the output as explained above is known as forward propagation. To further complete the modelling process, there is a process known as backpropagation. After the forward propagation is finished, the network will compare the error from output produced by the network with the target output. The weights in the input and hidden nodes then are adjusted according to how much each contributes to the overall error.

3.4. Measuring the errors and performances

To verify and determine the best performance of the model, there are several methods that can be applied such as mean absolute error (MAE), mean square error (MSE), mean average prediction error (MAPE), relative mean error (RME) and root mean square error (RMSE). However, MSE is the method that will be used in this project. MSE measures the average of the squares of the errors. The basic difference of MSE compared to other methods is that it penalizes any prediction techniques much more for large forecast errors than for small errors [6]. The MSE formula is given by the following equation,

\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i)^2 \]  

where \( Y_i \) is the value of predicted data, \( X_i \) is the value of actual data and \( n \) is the total data. It is therefore several smaller forecast errors will produce a smaller MSE.

4. Results and discussion

4.1. Support Vector Machine (SVM)

The first step is building the model with the best selection of parameters. The best SVM model is chosen based on their lowest MSE. In searching for the optimal parameters, the performance of

![Diagram of the Neural Network modeling process](image-url)
various SVM models is evaluated. Therefore, this project implements a grid search using cross validation. The number of cross validation fold used is ten in forecasting a daily electricity load [7]. The performance of a model will be the average of the ten validation predictions. In addition, the distribution of data has also been analyzed using correlation coefficient; R. R is a measure of the strength of linear dependence between two variables. After referring past studies and conducting a trial and error process for selection of parameters, the best result for each type of kernel are as in table 2.

**Table 2. SVM model result for prediction of the concentration of dispersed outlet in RDC column**

| Parameters                  | Linear | Polynomial | RBF   | Sigmoid |
|----------------------------|--------|------------|-------|---------|
| Penalty Parameter, C       | 3      | 10         | 6     | 10      |
| Epsilon, ε                 | 0.1    | 0.1        | 0.1   | 0.1     |
| Gamma, γ                   | -      | 0.25       | 0.25  | 0.25    |
| Degree, d                  | -      | 3          | -     | -       |
| Coefficient, k             | -      | 0          | -     | 0       |
| MSE                        | 0.000232 | 0.001080  | **0.000138** | 0.099295 |
| Correlation Coefficient, R | 0.987215 | 0.937471  | **0.992718** | 0.061908 |

From the table, it can be seen that kernel type Radial Basis Function (RBF) gives the best forecast value compared to Linear, Polynomial and Sigmoid kernel. It is because the value of MSE is the least which is 0.000138. It shows that the predicted value of this model is closer to the actual value compared to others. Besides, RBF kernel has a strong positive linear correlation because the value of the correlation coefficient is the largest and nearest to 1.00. Hence, SVM with RBF kernel are selected as the best SVM model and the performance of the best model will be compared with NN modelling later.

4.2. Neural network (NN)

The first step is choosing the suitable number for hidden nodes in the NN model, also other parameters involved which are weight and bias. The process of selecting the parameters value and number of hidden layer involved are by using trial and error method. There is no specific way to determine the best number of input and suitable number of the hidden nodes. The method is applied until the prediction has the less error between the actual data and the prediction data. The process is done in Matlab software. There are two types of transfer functions used in this project which are the Logsig and Purelin functions. The performance of these transfer function is described in table 3.

**Table 3. The MSE and correlation coefficient value for transfer function in NN model.**

| Transfer Function | Logsig | Purelin |
|-------------------|--------|---------|
| Mean Square Error (MSE) | 0.000174 | 0.000566 |
| Correlation Coefficient (R) | 0.99018  | 0.96826  |

Based on table 3, the value of correlation coefficient in Logsig transfer function is larger and it is near to 1.00 compared to Purelin transfer function. The closer to 1.00 shows the better fit of the regression line and Logsig transfer function fit the best. For MSE value, the MSE of Logsig transfer
function is smaller than Purelin transfer function. Therefore, the predicted value from Logsig transfer function is closer to actual value compared to predicted value form the Purelin transfer function.

4.3. Comparative performance of SVM and NN

Both SVM and NN models are compared to assess the overall effectiveness in predicting the concentration of dispersed phase outlet in RDC Column as shown in table 4 below.

|                      | SVM model      | Neural Network model |
|----------------------|----------------|----------------------|
| Mean Square Error (MSE) | 0.000138      | 0.000174             |
| Correlation Coefficient (R) | 0.992718      | 0.99018              |

From the table 4, the predicted values of SVM and Neural Network modelling are closed to the actual values, it is shown that both models can be used to predict the concentration of dispersed phase outlet in RDC column. However, in comparison between these two models, MSE value of the SVM is much smaller than the Neural Network. Besides, the R value for SVM is bigger and nearer to 1 compared to Neural Network modelling. Therefore, we can say that for the data we have, the SVM gives better performance compare to Neural Network.

5. Conclusion

This study attempt to compare the performance of SVM and Neural Network to predict the concentration of dispersed phase outlet in RDC. It is hope that this study can serve as another alternative tool for prediction in any output of RDC column instead the existing prediction techniques and the modelling experiences. Moreover, it is found that the capability of both models in predicting the output data can improve the RDC column performance onwards. We can see that some improvement of the mathematical modelling in the RDC column. Indeed, the researchers can make some improvement in the column component and the input variables.

Acknowledgement

The authors would like to thank Universiti Teknikal Malaysia Melaka (UTeM) for sponsoring this work under short-term grant no. PJP/2018/FTK(16A)/S01642.

References

[1] Hemmati A R Shirvani M Torab-Mostaedi M and Ghaemia A 2015 Hold-up and flooding characteristics in a perforated rotating disc contactor (PRDC) RSC Adv. 5, p. 63025-63033.
[2] Berk Z 2013 Food Process Eng. Technol. p. 287–309.
[3] Adil A A Al-Hemiri and Chalak S Omar 2008 Reduction of extraction efficiency in RDC column using artificial neural network J. Eng. 14, 2.
[4] Ismail W N A Zakaria S A Noor N F M and Ariffin W N M 2015 Experimental design and statistical analysis in rotating disc contactor (RDC) column In AIP Con. Proc. 1691, 1 p. 050010.
[5] Asadollahzadeh M Torab-Mostaedi M Torkaman R and Safdari J 2016 A new model for prediction of drop size distribution in a liquid–liquid extraction column RSC Adv. 6, 86 p. 82496-82504.
[6] Junaida binti Sulaiman 2005 Experimenting The Applicability Of Support Vector Machine Regression In Rainfall Prediction Master Thesis, Universiti Teknologi Malaysia.
[7] Chen BJ Chang M W and Lin CJ 2004 Load forecasting using support vector machines: A study on EUNITE Competition 2001 IEEE Trans. Power Syst. 19, 4 p. 1821-1830.