
Article:
Li, F., Zhang, J., Shang, C. et al. (2018) Modelling of a post-combustion CO2 capture process using deep belief network. Applied Thermal Engineering, 130. pp. 997-1003. ISSN: 1359-4311

https://doi.org/10.1016/j.applthermaleng.2017.11.078

Modelling of a Post-combustion CO_2 Capture Process Using Deep Belief Network

Fei Li[a], Jie Zhang[a], Chao Shang[b], Dexian Huang[b], Eni Oko[c], Meihong Wang[c]

[a]School of Chemical Engineering and Advanced Materials, Newcastle University, Newcastle upon Tyne NE1 7RU, UK

[b]Department of Automation, Tsinghua University, Beijing 100084, China

[c]Department of Chemical and Biological Engineering, University of Sheffield, Sheffield S1 3JD, UK

Abstract

This paper presents a study on using deep learning for the modelling of a post-combustion CO_2 capture process. Deep learning has emerged as a very powerful tool in machine learning. Deep learning technique includes two phases: an unsupervised pre-training phase and a supervised back-propagation phase. In the unsupervised pre-training phase, a deep belief network (DBN) is pre-trained to obtain initial weights of the subsequent supervised phase. In the supervised back-propagation phase, the network weights are fine-tuned in a supervised manner. DBN with many layers of Restricted Boltzmann Machine (RBM) can extract a deep hierarchical representation of training data. In terms of the CO_2 capture process, the DBN model predicts CO_2 production rate and CO_2 capture level using the following variables as model inputs: inlet flue gas flow rate, CO_2 concentration in inlet flue gas, pressure of flue gas, temperature of flue gas, lean solvent flow rate, MEA concentration and temperature of lean solvent. A greedy layer-wise unsupervised learning algorithm is introduced to optimize DBN, which can bring better generalization than a single hidden layer neural network. The developed deep architecture network models can then be used in the optimisation of the CO_2 capture process.

Keywords: CO_2 capture; chemical absorption; deep belief network; restricted Boltzmann machine.

Nomenclature

Abbreviations

BANNs	bootstrap aggregated neural networks
BAELM	bootstrap aggregated extreme learning machine
CO_2	carbon dioxide
CCS	carbon capture and sequestration
DBN	deep belief networks

GHG	greenhouse gas
MSE	mean squared error
MEA	monoethamine
RBM	restricted Boltzmann machines
RMSE	root mean squared errors
SLNNs	single-hidden layer feedforward neural networks

Letters

a	mean
$\mathbf{b}$	bias vector of visible layer
$\mathbf{c}$	bias vector of hidden layer
E	energy function
$\mathbf{h}$	vector of hidden unit outputs
m	mass fraction
P	probability distribution function
t	discrete time
u	input variables
$\mathbf{v}$	vector of visible unit outputs
$\mathbf{w}$	weight vector between visible units and hidden units
y	output variables

Greek symbols

α	learning rate
ε	flow rate
Ƞ	capture level
θ	parameter set of a neural network
μ	CO_2 production rate
σ	standard deviation

1. Introduction

Global climate change, as a result of the accelerated build-up of greenhouse gas (GHG) emission in atmosphere, has become a key concern of our society. The main component of GHG gas is carbon dioxide (CO_2). In the past few decades, numerous climate change policies were launched, but nonetheless, annual GHG emission still increased by 1.0 $GtCO_2$-eq (2.2%) per year from 2000 to 2010, compared to 0.4 $GtCO_2$-eq(1.3%) per year, from 1970 to 2000 [1].

A rapidly growing population plus industrialization, with corresponding increase in energy demand, is likely to result in increasing amount of GHG emission. Consequently, the Intergovernmental Panel on Climate Change has proposed that, compared to the emission levels in 1990, a 50% reduction of CO_2 emission is needed by 2050 [2].

The main source of worldwide CO_2 emission is the combustion of fossil fuel, such as petroleum, crude oil, natural gas and coal [3]. Amongst them, coal-fired power plants offer some advantages to operators, not only because of high availability of coal compared to other nature fuels, but also due to its flexible operation to changes in supply and demand [4]. However, the amount of CO2 emission per unit of electricity released by coal-fired power plants is twice as much as their natural gas counterparts [5]. As a result, many researches have been explored to reduce the CO_2 gas emission from coal-fired power plants. Carbon capture and sequestration (CCS) is identified as an appropriate technique for the sustainability of coal-fired power plant, because of its efficiency and effectiveness in reducing CO_2 emission [6]. Amongst the various technologies of CCS, the post-combustion carbon capture technology with chemical absorption has been considered as the most suitable way to reduce CO_2 emission. This is because it can retrofit the existing coal-fired power plant easily and treat flue gas stream with low CO_2 partial pressure [4]. However, it still has some disadvantages, one of which is the large energy requirement for absorbent regeneration. The thermal energy for regeneration usually comes from extracted steam from the low pressure steam turbines, which will reduce the efficiency of the coal-fired power plant. Therefore, it is particularly important to find out the trade-off between CO_2 capture level and energy consumption by using process optimisation. In order to carry out process optimisation, it is necessary to develop an accurate model for the post-combustion carbon capture process.

The models for CO_2 capture processes can be categorized into three groups: mechanistic, statistic and artificial intelligence based models. Lawal et al. [4] have developed a mechanistic model to present how disturbances affect the carbon capture process performance. In their study, the mechanistic model is able to predict the column temperature profile and CO_2 loading with high accuracy. On the other side, it consumes a lot of time consuming and also requires extensive knowledge of the process underlying physics. Further, Zhou et al. [7] have proposed a statistical model to instead of mechanistic model, for predicting heat duty, CO_2 production rate, CO_2 lean loading and capture level. In their statistical analysis, the selection of process variables in the model was affected by experts' opinion. However, the statistical model cannot

describe the irregular non-linear relationships between process variables. To tackle these problems, they explored a neural network model later and then compared its performance with the previous statistical model [8]. From that research, they found using neural network model was able to predict CO_2 production rate with much higher accuracy than the statistical model. Hence, the artificial intelligence based models become increasingly significant and draws much more attention in modelling the post-combustion CO_2 capture plants. Meanwhile, Sipocz et al. [9] has also used a single-hidden layer feedforward artificial neural network (SLNNs) to model the chemical absorption process and explore the relationship between process variables and quality variables. Similarly, the results in the paper have shown that artificial neural network models are able to simulate the rigorous complex steady state process with minimum time demand and high accuracy. However, the SLNNs model would possibly encounter some problems such as over-fitting of the training data and poor generalization performance. To overcome these shortcomings, Li et al. [10] have presented a new learning algorithm, called bootstrap aggregated neural networks (BANNs), with a combination of several single-hidden layer neural networks together to model post-combustion CO_2 capture process. The modelling technique was verified to predict CO_2 capture level and CO_2 production rate with higher accuracy and reliability than the traditional neural network models, especially performs better in the aspect of long-term predictions. Besides, BANNs model is able to measure prediction reliability by prediction confidence bounds. Later, they have further explored a fast learning algorithm on the basis of BANNs, called bootstrap aggregated extreme learning machine (BAELM) [11]. The structure of ELM is the same as SLFNNs, but the network weights are calculated differently. In ELM, the weights between input and hidden layers are randomly assigned while the weights between hidden layer and output layer are obtained by one-step regression. Their research results show that the training time of BAELM is 5 times lower than that of BANNs. Furthermore, the generalization performance of BAELM is better than BANNs, as indicated by lower mean squared errors (MSE). Nevertheless, all of the above mentioned neural networks have only one hidden layer. They may possibly have difficulty in handling strongly correlated industrial process variables. In other words, the single hidden layer neural networks face difficulties when representing complex and highly-varying relationship and are easy to converge to local optima [12].

A solution to these problems has been proposed, which is a deep multi-layers neural networks model inspired by the structure of human brain [13]. The multiple hidden layers represent different levels of latent features of input data. As a result, it can deal with the data which has

complicated irregular relationship. It is true that, before 2006, the attempts to apply deep multi-layers neural networks were failed because of unsuccessful training strategies. The reason is that the gradient-based method is starting from random initialization, thereby getting stuck near poor solution. Nonetheless, in 2006, Hinton et al. [13] have put forward a greedy layer-wise unsupervised learning algorithm for deep belief networks (DBN), which pre-train one layer at a time in a greedy way. Simply, DBN is a generative model, in which the lower layers represent the low-level features from inputs and the upper layers extract the high-level features that explain the input samples. With the comparison of random initialization, the results show that the initial parameters of DBN are much closer to optimal solutions [14]. Since then, increasing attention has been paid to DBN model and it contributes a lot to image recognition [15] and time series forecasting [16]. However, it has never been applied to modelling of CO_2 capture processes.

The rest of paper is arranged as follows. Section 2 presents the theoretical knowledge of DBNs and their component layers, Restricted Boltzmann Machine (RBM). Section 3 introduces the details of a post-combustion CO_2 capture plant. Then, the comparative result analysis between SLNNs and DBNs is revealed in Section 4. Finally, Section 5 gives conclusions and future works.

2. Deep Belief Networks

Many researches have shown that DBN can produce models with higher accuracy and precision, especially with respect to image recognition [15], time series forecasting [16], and robotics [17]. In this paper, DBNs are used to model a post-combustion CO_2 capture process.

2.1 Restricted Boltzmann Machines

Fig. 1 shows the constituent structure of DBNs, RBMs, which are shallow and two layer-neural nets. The calculations take place in each node of Fig. 1 which represents a neuron-like unit called a node. The nodes of hidden layer are not connected to each other, while they have undirected connections to a layer of vision layer. Theoretically, it is a special type of generative energy based model which can learn probability distribution over its inputs. As there are no connections between hidden units in RBMs, it has an advantage that the hidden unit is conditionally independent to each other. Both visible units (**v**) and hidden units (**h**) are

stochastic binary variable nodes and hypothesis that the joint probability distribution of (**v**, **h**) fits Boltzmann distribution. **v** is connected to **h** through undirected weighted connections. The reason why they are restricted is that, there is no connection between hidden variables or visible variables. A probability distribution p(**v**,**h**) is defined via an energy function, E(**v**,**h**; θ), which can be written as:

$$-\log P(\mathbf{v}, \mathbf{h}) \propto E(\mathbf{v}, \mathbf{h}; \theta) = -\mathbf{b}^T\mathbf{v} - \mathbf{c}^T\mathbf{h} - \mathbf{h}^T\mathbf{W}\mathbf{v} \quad (1)$$

where $\theta = (\mathbf{w}, \mathbf{b}, \mathbf{c})$ is the parameter set, **w** is the weight vector between visible units and hidden units, and **b** and **c** are their bias vectors, respectively. Due to the configuration of RBMs, it is possible to compute the conditional probability distribution, when **v** and **h** are given, as

$$P(h_j|v_i, \theta) = \text{sigm}\left(\sum_{i=1}^{|v|} W_{ij} v_i + b_j\right) \quad (2)$$

$$P(v_i|h_j, \theta) = \text{sigm}\left(\sum_{j=1}^{|h|} W_{ij} h_j + c_i\right) \quad (3)$$

where $\text{sigm}(x) = 1/(1 + e^{-x})$ is the sigmoid function. The parameter θ = (**w**, **b**, **c**) can be calculated using contrastive divergence effectively.

However, the RMBs with binary nodes can only deal with discrete inputs. When inputs are continues values, Gaussian RBMs are suitable to apply [18] as shown in Eq(4).

$$E(\mathbf{v}, \mathbf{h}; \theta) = -\sum_i \frac{(v_i - a_i)^2}{2\sigma_i^2} - \mathbf{c}^T\mathbf{h} - \mathbf{h}^T\mathbf{W}\mathbf{v} \quad (4)$$

where a_i and σ_i are mean and standard deviation respectively of the Gaussian distribution for visible unit i, **v** is the continuous valued input layer, and **h** is the binary layer.

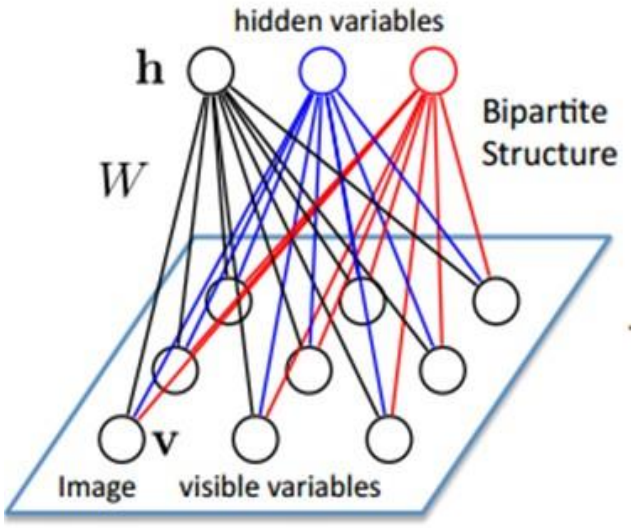

Figure 1: The structure of Restricted Boltzmann Machines

2.2 Learning algorithm for RBMs

As to a RBMs model, **v** is given and **h** is to be estimated. Therefore, learning RBM means to make the probability distribution represented by RBMs (P(**v**)) maximally coincide with training input data, by adjusting parameters θ = (**w**, **b**, **c**).

For S={ v^1, v^2,… v^{n_s} }, n_s is the number of training inputs, v^i = (v_1^i, v_2^i, … , v_{nv}^i)$^{\mathrm{T}}$, i=1,2,…, n_s, and they are independent and identically distributed. The objective of training RBM is to maximise the following likelihood function:

$$\ln \mathrm{L}_{\theta,s}=\sum_{i=1}^{n_s} P(v^i) \tag{5}$$

Gradient ascend is a typical method to maximize Eq(5). It approaches the optimum via iterations, which can be formed as below:

$$\theta \coloneqq \theta + \alpha \frac{\partial \ln(\mathrm{L}_\theta)}{\partial \theta} \tag{6}$$

where α is learning rate. In Eq(6), the calculation of the gradient $\frac{\partial log(\mathrm{L}_\theta)}{\partial \theta}$ is particularly important. To better understand this, the gradient of likelihood function at a single data point v is calculated as:

$$\begin{aligned}\frac{\partial \ln(P(v))}{\partial \theta} &= \frac{\partial}{\partial \theta}(\ln\sum_h e^{-E(v,h)}) - \frac{\partial}{\partial \theta}(\ln\sum_{v,h} e^{-E(\tilde{v},h)}) \\ &= -\frac{1}{\sum_h e^{-E(v,h)}}\sum_h e^{-E(v,h)}\frac{\partial E(v.h)}{\partial \theta} + \frac{1}{\sum_{v,h} e^{-E(\tilde{v},h)}}\sum_{v,h} e^{-E(\tilde{v},h)}\frac{\partial E(\tilde{v},h)}{\partial \theta} \\ &= -\sum_h P(h|v)\frac{\partial E(v.h)}{\partial \theta} + \sum_{v,h} P(\tilde{v},h)\frac{\partial E(\tilde{v},h)}{\partial \theta}\end{aligned} \tag{7}$$

Note that there are two terms called negative term and positive term in Eq(7). The negative term represents the conditional expectation of $\frac{\partial E(v.h)}{\partial \theta}$, given the visible unit v, which is easy to compute. However, the computation of the positive term, which is expectation of $\frac{\partial E(\tilde{v},h)}{\partial \theta}$ for joint distribution $P(\tilde{v},h)$, is an intractable problem. It is causally linked to ($2^{n_v + n_h}$) items in $\Sigma_{v,h}$, giving rise to computation complexity of $O^{(n_v + n_h)}$. Therefore, Gibbs Markov Chain on the pair of variables is usually considered to resolve the problem. However, it is still intricate, because a large quantity of frequency samples is always required to guarantee the precision. Hinton et al. [13] proposed an idea of Contrastive Divergence, which takes initial sample v_0 =

x sampled from the training distribution and arrives the distribution of RBMs with k small steps.

2.3 Unsupervised pre-training and supervised fine tuning of DBN

As the inputs in our case are continuous-valued and not limited to a certain range, the training process of DBN by greedy layer-wise scheme from lower layers to higher layers is demonstrated in Figure 2. The bottom RBM layer is selected as Gaussian units and the remaining hidden layer above this use Binary units. In details, Gaussian RBM h_1 is trained firstly and then taken as the inputs of RBM h_2. Next, the training of h_2 can be accomplished in the same manner. The output of each RMB is the extracting feature of previous output by maximizing its probability. In other words, the high level RMB represents the most representative feature of input data, and the low level RMB is the low-level extraction of input data. It is noted that, during the training process, there is no target variable involved, which is denoted as unsupervised pre-training.

Unsupervised pre-training of DBN is important to improve the model performance. It was interpreted by Bengio [14] as follows: injecting unsupervised training may help to put the parameters of that layer towards the better direction in the parameter space. A greedy layer-wised training algorithm was proposed to train each layer at one time [14]. Specifically, start to learn from the lowest weight matrices and keep all the higher weight matrices tied. In this work, RBM is used to pre-train each layers of DBN networks to lead the initial weights to optimum solution. After the unsupervised steps of DBN are finished, the supervised fine-tuning by back-propagation method is conducted to modify the weights between each different layers. Hinton et al. [13] proposed an idea of wake-up algorithm, which has capability to fine-tune the parameters of all layers together.

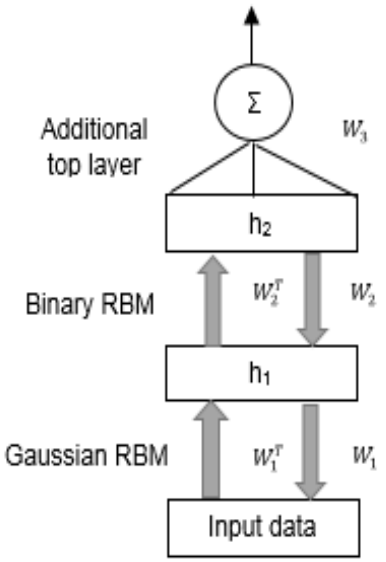

Figure 2: The structure of DBN with continuous-valued inputs.

3. Case study: Post–combustion CO2 capture plant with chemical absorption

3.1 Case description

A simplified flow diagram of post-combustion CO_2 capture plant with chemical absorption is shown in Figure 3. The flue gas exhausted from the upstream power plant is transported into the bottom of an absorber and contacted counter-currently with lean amine from the top side of the absorber. The solution will react and absorb CO_2 in the flue gas, thereby generating treating gas containing much lower CO_2. Then the rich amine exchanges heat with regenerated amine solution coming from the bottom of regenerator in a heat exchanger, before pressured into a stripper. In the stripper, the pure CO_2 is evaporated and amine solution is regenerated by the heat provided from reboiler. The CO_2 is cooled in the condenser to knock out vaporised solvent and then compressed for downstream storage. The heat supplied by reboiler is coming from the low pressure steam from upstream power plant. The capture of CO_2 in flue gas will lead to a large energy penalty.

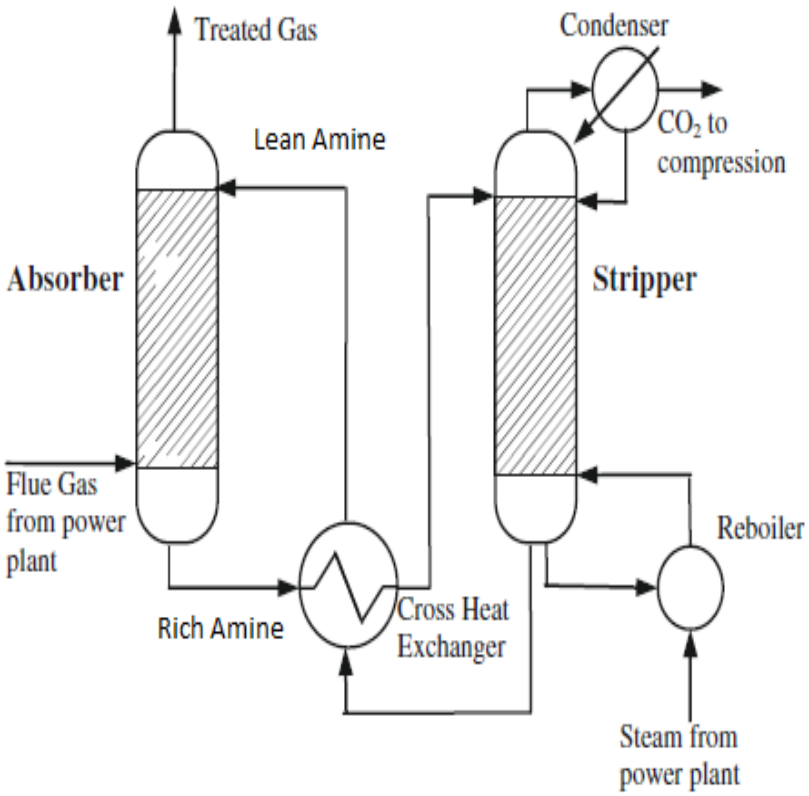

Figure 3. Simplified process flow diagram of chemical absorption process for post-combustion CO_2 capture.

3.2 Variable selection

As can be seen from the process details, there are a huge number of variables in the post-combustion CO_2 capture process. However, in this study, only two quality variables, CO_2 capture level and CO_2 production rate, are considered as the main indicators of the process performance. Capture level is the amount of CO_2 extracted from the inlet flue gas in the absorber. It is calculated as follows:

$$\eta = 1 - \frac{m_{outlet\ co_2} \times \varepsilon_{outlet\ gas}}{m_{inlet\ co_2} \times \varepsilon_{inlet\ gas}} \quad (8)$$

where $m_{outlet\ co_2}$, $\varepsilon_{outlet\ gas}$, $m_{inlet\ co_2}$ and $\varepsilon_{inlet\ gas}$ represent CO_2 mass fraction in gas out of absorber, gas flow rate out of absorber, CO_2 mass fraction in inlet flow gas of absorber, and inlet gas flow rate of absorber, respectively.

CO_2 production rate (μ) is the amount of CO_2 extracted from the flue gas and amine solvent in the regenerator. It is measured at the top of regenerator and calculated as Eq. (9).

$$\mu = \mu_{gas} \times m_{co2} \tag{9}$$

where μ_{co2} and μ_{gas} are production rate of CO_2 and gas from the top of regenerator respectively, and m_{co2} is the CO_2 mass fraction in the regenerating gas in regenerator column.

Eight process variables are selected as model inputs and they are inlet gas flow rate, inlet gas pressure, inlet gas temperature, CO_2 concentration in inlet gas, MEA circulation rate, lean solvent flow rate, lean solvent temperature and reboiler duty, which influence the capture efficiency and process performance [4, 5].

3.3 Neural networks construction

An amount of 1076 data samples were collected from the gPROMS based simulator at University of Hull. The nonlinear dynamic models in this study are developed as following form:

$$y(t)=f[u(t-1), u(t-2), u(t-3), y(t-1), y(t-2)] \tag{10}$$

where y is the process output variables (CO_2 capture level and CO_2 production rate), u represents the process input variables mentioned above, t is discrete time, and $f[\ ]$ is the nonlinear function represented by the neural network.

Prior to building the models, the data should be pre-processed to avoid missing values and outliers. As the variables have different physical units and magnitudes, each variable should be scaled to zero mean and unit variance. In developing the DBN models, all the input data is used for the unsupervised training process to extract their feature, which is stated in Section 2.3. Then, the data samples are randomly split into three sets: training data (64%), test data (16%) and validation data (20%). To evaluate the dynamic model performance, the data of batch 1 are used to further evaluate model prediction performance. Accordingly, two DBN models are constructed for the quality predictions of CO_2 production rate and CO_2 capture level. Cross-validation is used to select the network architecture and both models are found to have the structure of 26-20-17-1. That is the DBN has 26 input nodes, 20 hidden nodes in the first hidden layer, 17 hidden nodes in the second hidden layer, and 1 output layer node. As the neural network learning depends on random initial weights, it is necessary to repeat the training

procedure for several times and the result with least training error is selected. In this study, the training procedure is repeated for 20 times.

4. Results and discussions

In this study, the performance of DNB modelling technique is compared with traditional neural network modelling technique, namely, SLNNs. As mentioned above, the structure of 2 hidden layer is determined for DBN, in which the bottom hidden layer is Gaussian RBM and top hidden layer is binary RBM. The numbers of neurons in these two hidden layers are 20 and 17 respectively, which are determined by considering a range of hidden neuron numbers and inspecting the performance on the testing data. The learning rates for both unsupervised training and supervised training for DBN are selected as 0.1 to avoid low learning speed and local optimisation. SLNNs are trained by the Levenberg-Marquardt optimisation training algorithm with regularisation and cross validation based "early stopping". The regularisation parameter is set as 0.1 as the data are generated from simulation and do not contain much noise. The hidden layer of SLNN contains 20 hidden neurons, which is also determined by considering a range of hidden neuron numbers and inspecting the performance on the testing data.

Table 1: Comparison of modelling results of DBN and SLNNs on CO_2 production rate.

	DBN	SLNNs
Training RMSE (kg/s)	0.000072	0.0016
Validation RMSE (kg/s)	0.000048	0.0042
Testing RMSE (kg/s)	0.000246	0.0038

As to predicting CO_2 production rate, the root mean squared errors (RMSE) values on training, validation and testing data are given in Table 1. It clearly shows that DBN model gives much lower RMSE values than SLNNs model. As a result, DBN model has ability to model the CO_2 capture process more accurately than the SLNN model. This is because DBN can extract the data characteristics by unsupervised learning, thereby accelerate the learning convergence and avoid local minimum. To further prove this point, the data of batch 1 is used to verify the DBN model. Figure 4 compares the one-step-ahead prediction performance on CO_2 production rate by SLNN model (top) and DBN model (bottom). The red dashed line represents the prediction values, while the blue solid line represents actual values. It can be seen clearly that the

predictions from the DBN model are closer to the actual values than the SLNN model. The RMSE values of DBN on batch 1 is 0.000125 kg/s, while that of SLNNs model is 0.000185 kg/s which is slightly higher. When consider long term prediction, Figure 5 shows their performance comparison on the data of batch 1. Obviously, both models can provide quite accurate long range predictions, but the DBN model performs much better. The predictions by DBN model are much closer to the actual values, in which the RMSE value is only 0.000276 kg/s. Compare to the RMSE value of short term prediction, it is not different a lot. However, as to the long term prediction by SLNNs model, the RMSE value is 0.000926 kg/s, which is 4 times higher than that of short term prediction. This demonstrates that DBN model can not only give the accurate short term predictions, but also predict the long term values with high performance.

The RMSE values of CO_2 capture level prediction by DBN and SLNNs models are given in Table 2. As can be seen clearly, the RMSE values of SLNNs model on training, validation and testing data are much higher than those of DBN model. Especially, the RMSE values on validation and testing data by SLNNs model are approximately 10 times higher than those by DBN model. This indicates that DBN model has a better generalisation ability than SLNNs model. The reason is that, the unsupervised training procedure is using principal component analysis (PCA) to analyse the underlying structure of the input data, in which the reduced-dimensionality feature is captured. Therefore, it can extract the most important feature of data and works well in modelling variables.

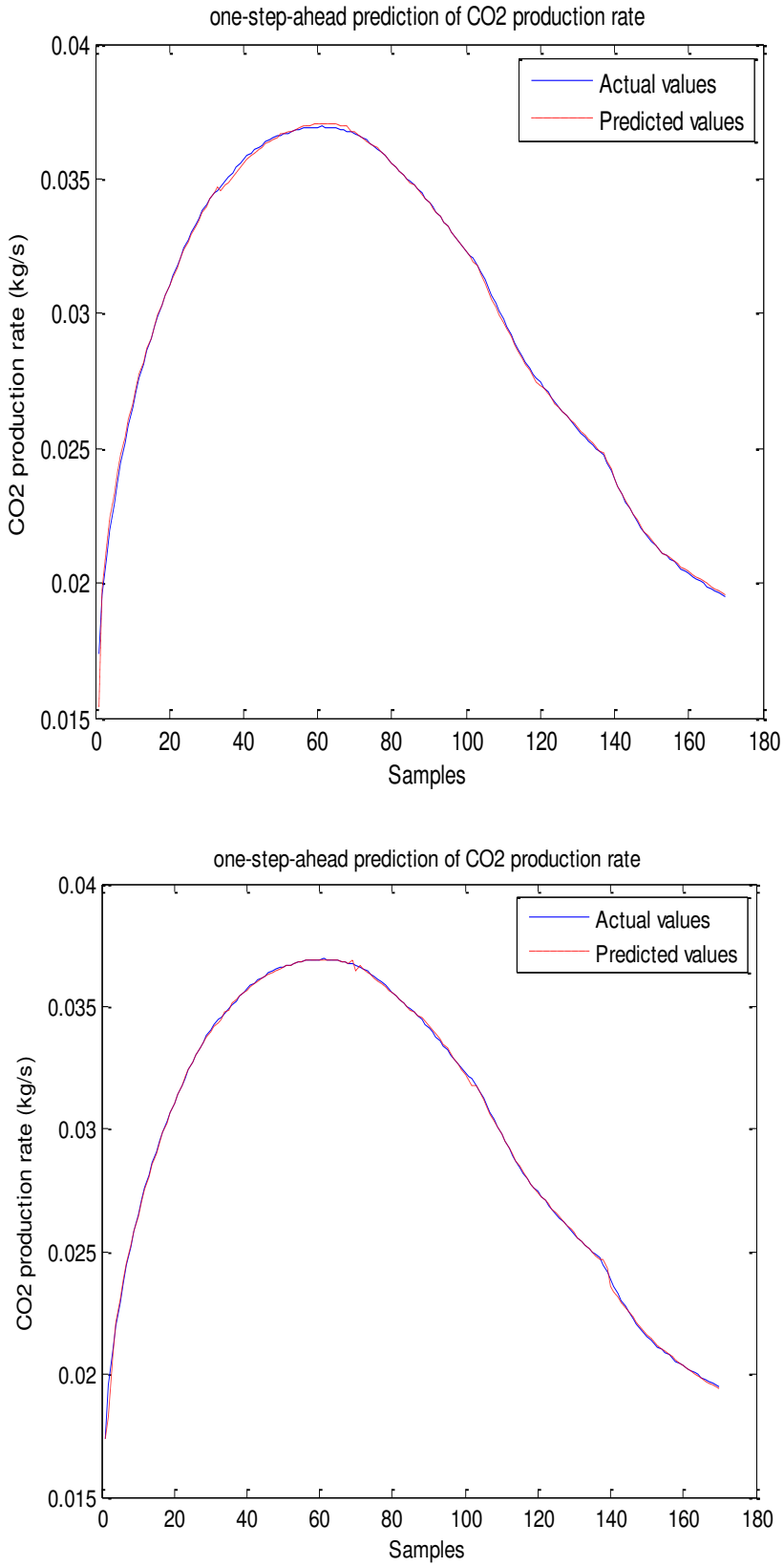

Figure 4: One-step-ahead predictions of CO_2 production rate by SLNNs (top) and DBN (bottom).

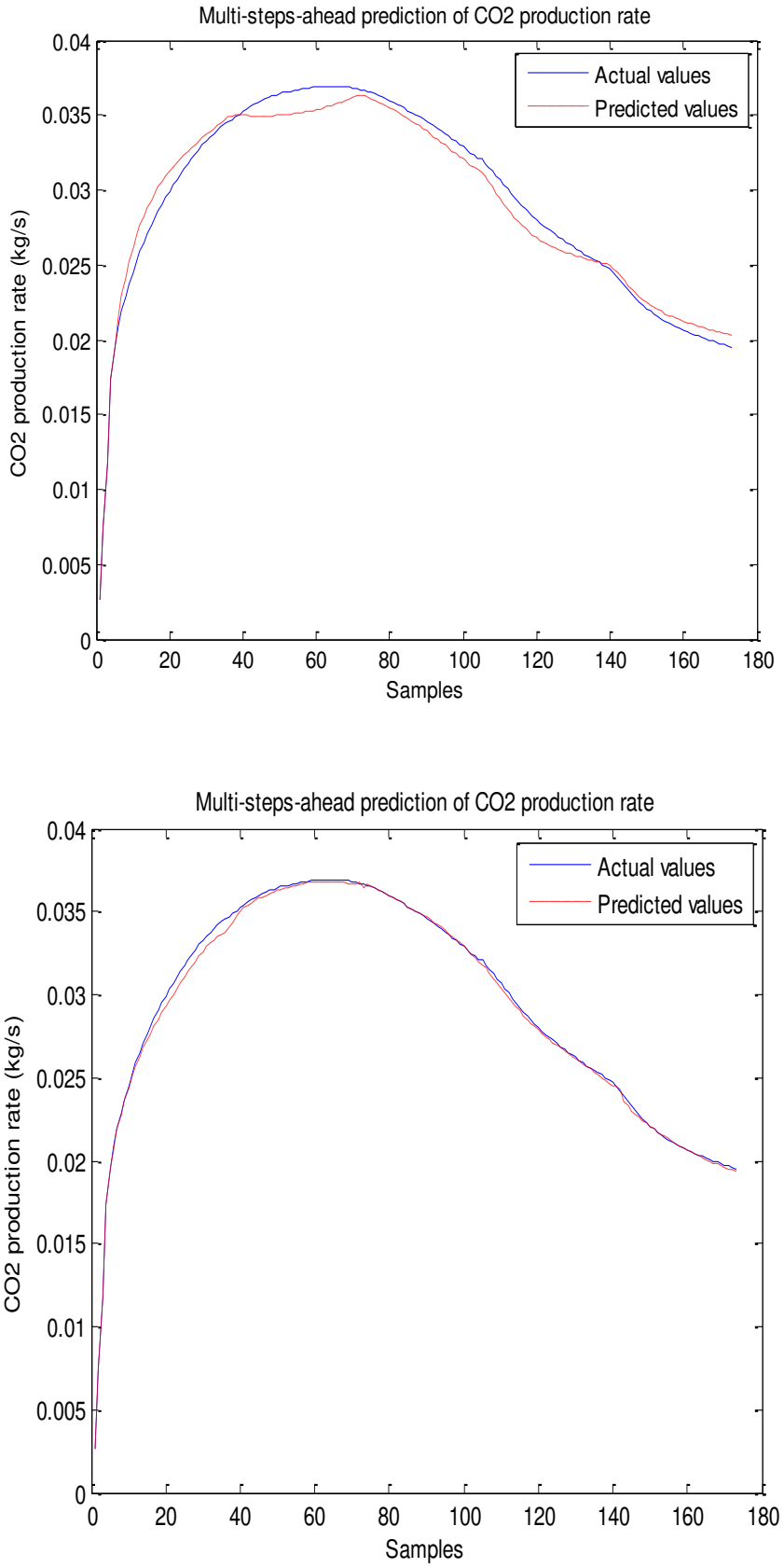

Figure 5: Multi-step-ahead predictions of CO_2 production rate by SLNNs (top) and DBN (bottom).

The one-step-ahead prediction performance comparison of SLNN and DBN models on batch 1 is shown in Figure 6. The predictions shown as red dashed curves are almost identical to the true values shown as blue solid curves in both plots. However, slightly large prediction errors are seen clearly in the top plot when there are step changes in inputs. This demonstrates that the DBN is able to catch the underlying feature of the data and represent the dynamics of process accurately. In details, the RMSE values of the DBN and SLNN models are 0.0259% and 0.0312% respectively, and the former one is little lower than the latter one. Figure 7 shows the long range predictions of CO_2 capture level by the SLNN and DBN models. It can be seen clearly from the top graph, the multi-step-ahead predictions of SLNN model are not accurate as a result of some extremely large errors. However, for the DBN model, the long range predictions are much closer to the actual values of CO_2 capture level. In details, the RMSE values of long range predictions by DBN and SLNN models are 0.0891% and 0.3544%, respectively. The former one is about 4 times lower than the latter one. It further proves that the DBN model is able to predict with higher generalization ability than the SLNN model. For model predictive control and real-time optimisation applications, the long range prediction ability is generally of much greater importance.

Table 2: Comparison of modelling results of DBN and SLNNs on CO_2 capture level.

	DBN	SLNNs
Training RMSE (%)	0.0222	0.1475
Validation RMSE (%)	0.0314	0.3681
Testing RMSE (%)	0.0321	0.2572

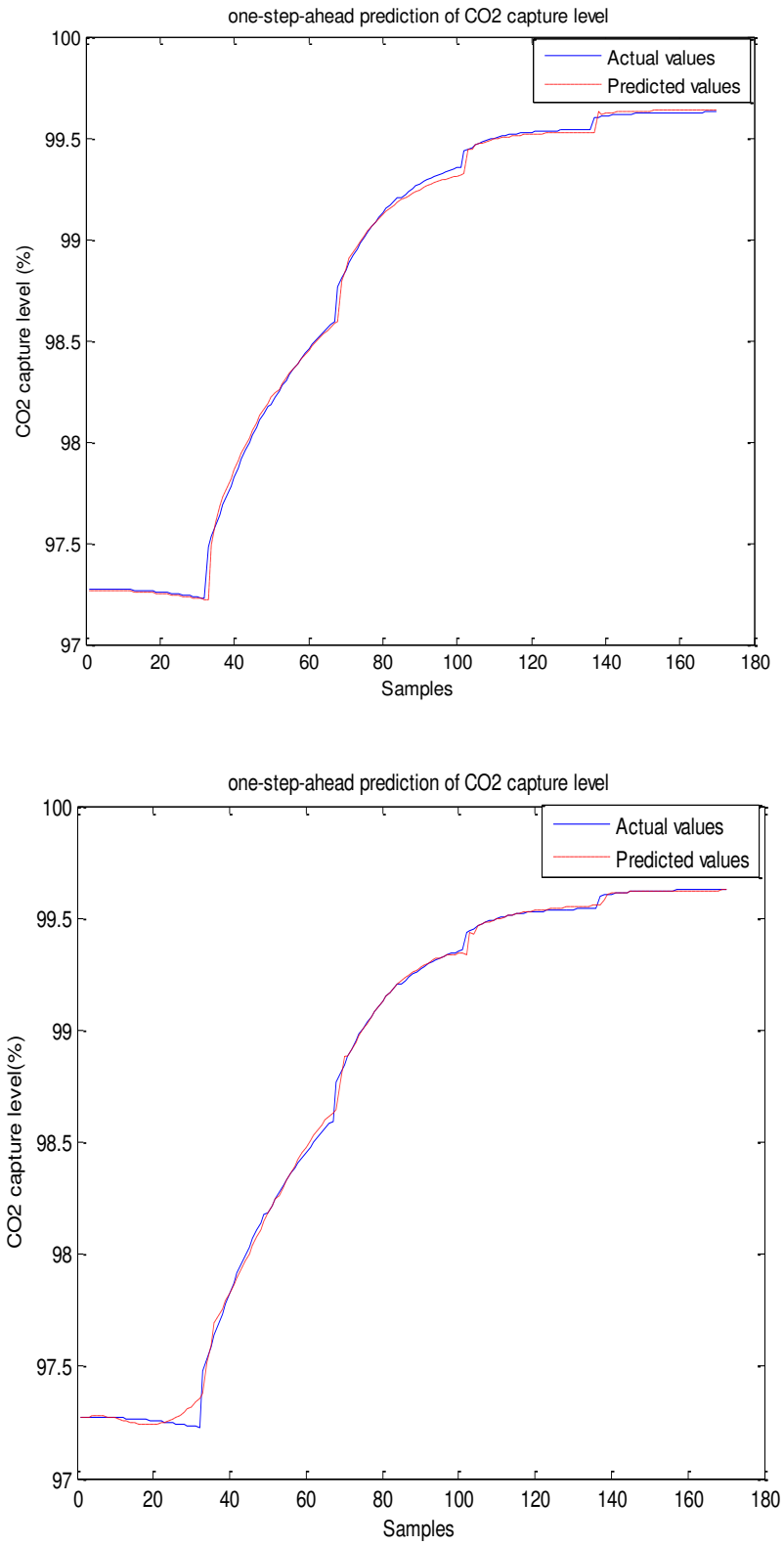

Figure 6: One-step-ahead predictions of CO_2 capture level by SLNNs (top) and DBN (bottom).

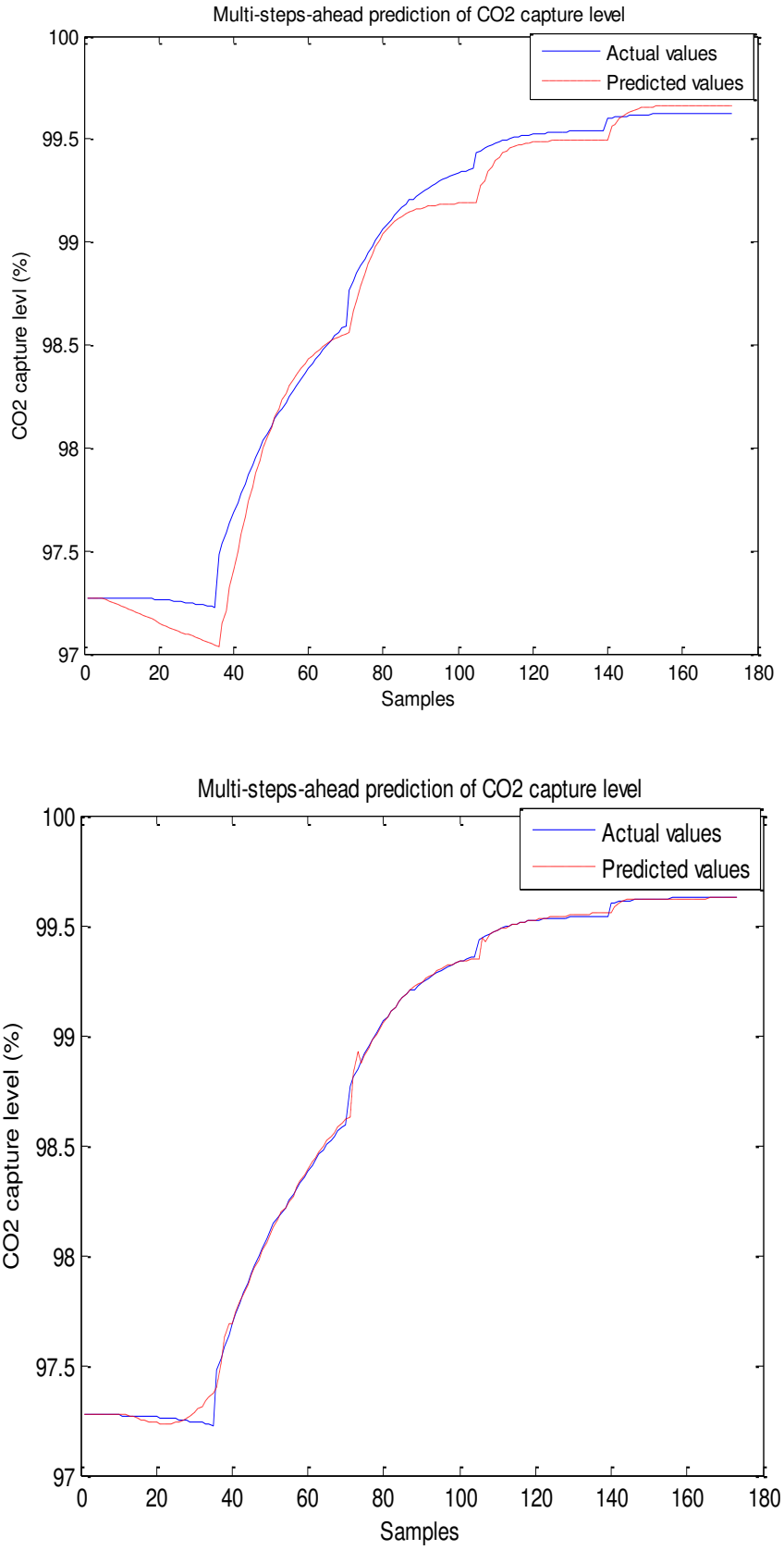

Figure 7: Multi-step-ahead predictions of CO_2 capture level by DBN (top) and SLNN (bottom).

5. Conclusions

The deep learning technique is employed as a new modelling method for post-combustion CO_2 capture plant, as it is identified as a more accurate model than the conventional single-hidden layer neural network model. The developed dynamic model predict CO_2 production rate and CO_2 capture level, which are two key parameters to represent the performance of post-combustion CO_2 capture plant. The advantages and characteristics of DBN are analysed in details. The results indicate that DBN can extract nonlinear latent variables, making the neural networks as a latent variable model. Compared with the SLNN model, DBN model is shown to have better model accuracy and generalisation ability, especially for multi-step ahead predictions. This is because that the unsupervised stage can avoid the model being trapped into local minima and overcome the over-training problem. Therefore, it can be treated as a powerful tool to instead of SLNNs model for post-combustion CO_2 capture process. Nevertheless, it still exits some problems. For instance, the training of DBN procedure requires much more time than SLNN. As well, the modelling parameters is expected to be adjusted for the further results improvement.

Acknowledgement

The work was supported by the EU through the project "Research and Development in Coal-fired Supercritical Power Plant with Post-combustion Carbon Capture using Process Systems Engineering techniques" (Project No. PIRSES-GA-2013-612230) and National Natural Science Foundation of China (61673236).

Reference

[1] M. Anderson, Climate Change 2014: Synthesis Report, Library Journal 141(9) (2016) 28-28.

[2] B. Metz, and Intergovernmental Panel on Climate Change. Working Group III. Climate change 2007 : mitigation of climate change : contribution of Working Group III to the Fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge ; New York, Cambridge University Press, (2007).

[3] P. Freund,. Making deep reductions in CO2 emissions from coal-fired power plant using capture and storage CO2, Proceedings of the Institution of Mechanical Engineers Part A-Journal of Power and Energy 217(A1) (2003) 1-7.

[4] A. Lawal, M. Wang, P. Stephenson, G. Koumpouras and H. Yeung, Dynamic modelling and analysis of post-combustion CO2 chemical absorption process for coal-fired power plants, Fuel 89(10) (2010) 2791-2801.

[5] A. Lawal, M. Wang, P. Stephenson and H. Yeung, Dynamic modelling and simulation of CO_2 chemical absorption process for coal-fired power plants, Computer Aided Chemical Engineering, 27 (2009) 1725-1730.

[6] Metz, B. and Intergovernmental Panel on Climate Change. Working Group III. (2005). IPCC special report on carbon dioxide capture and storage. Cambridge, Cambridge University Press, for the Intergovernmental Panel on Climate Change.

[7] Q. Zhou, C. W. Chan, P. Tontiwachwuthikul, R. Idem, D. Gelowitz, A statistical analysis of the carbon dioxide capture process, International Journal of Greenhouse Gas Control 3(5) (2010) 535-544.

[8] Y. X. Wu, Q. Zhou, C.W. Chan, A comparison of two data analysis techniques and their applications for modeling the carbon dioxide capture process, Engineering Applications of Artificial Intelligence 23(8) (2010) 1265-1276.

[9] N. Sipocz, F. A. Tobiesen, M. Assadi, The use of Artificial Neural Network models for CO2 capture plants, Applied Energy 88(7) (2011) 2368-2376.

[10] F. Li, J. Zhang, E. Oko, M. H. Wang. Modelling of a post-combustion CO2 capture process using neural networks, Fuel 151 (2015)156-163.

[11] F. Li, J. Zhang, E. Oko, M. H. Wang, Modelling of a post-combustion CO2 capture process using extreme learning machine, 21st International Conference on Methods and Models in Automation and Robotics, (2016) 1252-1257.

[12] Y. Bengio, Y. LeCun, Scaling learning algorithms towards AI, Large-Scale Kernel Machines, (Eds.) L. Bottou, O. Chapelle, D. DeCoste, J.Weston, MIT Press, 2007, 34.

[13] G. E. Hinton, S. Osindero, Y. W. The, A fast learning algorithm for deep belief nets, Neural Computation 18(7) (2006) 1527-1554.

[14] Y. Bengio, Learning Deep Architectures for AI, Foundations and Trends in Machine Learning 2 (2009) 1-127.

[15] B. Liao, J. Xu, J. Lv, S. Zhou, An image retrieval method for binary images based on DBN and Softmax classifier, IETE Tech. Re. 32(4) (2015) 294-303.

[16] Y. Ren, J. Mao, Y. Liu, Y. Li, A novel DBN model for time seires forecasting, IAENG International Journal of Computer Science 44(1) (2017) 79-86.

[17] M. Z. Uddin, J. Kim, A robust approach for human activity recognition using 3-D body joint motion features with deep belief network, KSII Transactions on internet and Information Systems 11(2) (2017) 1118-1133.

[18] C. Shang, F. Yang, D. Huang, W. Lyu, Data-driven soft sensor development based on deep learningtechnique, Journal of Process Control 24 (2014) 223-233.