Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 122 (2017) 300–307

Procedia Computer Science

www.elsevier.com/locate/procedia

Information Technology and Quantitative Management (ITQM 2017)

Forecasting Crude Oil Prices: a Deep Learning based Model

Yanhui Chen[a], Kaijian He[b,d,*], Geoffrey K.F. Tso[c]

[a]School of Economics and Management, Shanghai Maritime University, Shanghai, 201306, China
[b]School of Business, Hunan University of Science and Technology, Xiangtan, 411201, China
[c]Department of Management Sciences, City University of Hong Kong , Kowloon, Hong Kong
[d]School of Economics and Management, Beijing University of Chemical Technology, Beijing, 100029, China

Abstract

With the popularity of the deep learning model in the engineering fields, it has attracted significant research interests in the economic and finance fields. In this paper, we use the deep learning model to capture the unknown complex nonlinear characteristics of the crude oil price movement. We further propose a new hybrid crude oil price forecasting model based on the deep learning model. Using the proposed model, major crude oil price movement is analyzed and modeled. The performance of the proposed model is evaluated using the price data in the WTI crude oil markets. The empirical results show that the proposed model achieves the improved forecasting accuracy.

Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.

Keywords: Crude oil price forecasting; Deep Learning model; ARMA model; Random Walk model

1. Introduction

The crude oil price movements are subject to diverse influencing factors. The dynamic complicated interactions among these factors result in the mysterious behavior of the crude oil price movement, whose characterization and prediction remained one of the most interesting and intriguing research issues in the economic and financial analysis field. Recently numerous empirical studies have revealed the nonlinear nature of economic and financial data, where traditional methods such as linear prediction methods are not able to analyze the complex nonlinear dynamics involved [1]. [2] used the Qual VAR model to model the nonlinear autocorrelation characteristics of WTI crude oil price changes and forecast its future movement. They found that the Qual VAR model outperforms the benchmark Random Walk and VAR model. [3] proposed a Hidden Markov Model (HMM) with threshold effect to model hidden factors influencing the crude oil price movement. They demonstrated that the proposed models outperform the ARMA model, in h-day ahead forecasting exercise. On the other hand, the Artificial Intelligence (AI) and Machine Learning (ML) based approaches received more and more research interests. [4] used the Recurrent Neural Network (RNN) to forecast the crude oil indices. [5] proposed a Genetic Algorithm optimized Neural Network model to forecast the crude oil price fluctuations. They showed this evolutionary neural

*Corresponding author: Tel.:+86-731-58290306
E-mail address: paulhekj@hnust.edu.cn.

1877-0509
Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.
10.1016/j.procs.2017.11.373

network model brings statistically significant performance improvement. [6] provided a comprehensive survey on the AI and ML based crude oil forecasting models.

The deep learning model is a new artificial intelligence paradigm developed beyond the neural network. It has become a popular phenomenon in the computer science and engineering fields such as image recognition, text classification and speech recognition recently. For example, [7] proposed HMM-DNN model for speech synthesis method. In addition to the HMM-DNN model, the CNN model and the RNN model are also applied to the construction of speech recognition models. However, the deep learning model has only been introduced to the financial field quite recently. There are some very early exploratory attempts. For example, [8] proposed a deep learning hierarchical decision models and used it to construct new portfolio with the desired level of performance. Empirical studies show that the proposed model achieves the superior performance. [9] predicted the real estate price by combining the Boltzmann machine with the genetic algorithm. [10] applied the convolution neural network to the biological aspects of DNA-protein transcription prediction. [11] used the deep belief network model to predict the short-term power load in Macedonia from 2008 to 2014. The empirical results showed that the deep belief network model has obvious advantages compared with the traditional forecasting models.

In this paper we propose a new Deep learning based hybrid crude oil price forecasting methodology to model the nonlinear dynamics involved in the crude oil price movement and forecast its future movement at higher level of accuracy. Taking into account both the linear and non-linear characteristics of historical data, we integrate the prediction results of the ARMA model and the prediction results of the deep learning model. Empirical studies have been conducted using the major crude oil prices to evaluate the performance of the proposed model. The superior performance of the crude oil price forecasting model using the deep belief network and recurrent neural network provide the empirical evidence that the market is inefficient in the regional and sub markets. Work in this paper shows methodologically the merit of the deep learning model in tracking and capturing the nonlinearity and dynamic crude oil price movement. When analytic solutions are lacking, it would provide the best approximation to the nonlinear dynamics in the crude oil price movement. It contributes to the literature by providing a new methodology on how the deep learning model can be used to improve the crude oil forecasting accuracy.

The rest of the paper proceeds as follows. In section 2 we provide a brief account of the deep belief network and long short term memory model, two typical deep learning models. We further propose the deep learning based hybrid crude oil price forecasting model. In section 3 we report the results from the empirical studies using the major crude oil prices. We further analyze the results. Section 4 concludes with some summarizing remarks.

2. Methodology

Since the seminal work of Hinton in 2006, Deep Learning received significant research interests from both academia and industry. Actually deep learning algorithms train the neural network model with multiple layers [12]. This method constructs a neural network to describe and explain the intrinsic relationship in historical data[13][14]. Compared with traditional neural network with the same structure, Deep Learning reduces the adverse impact of local minima issue in classic neural network models trained using traditional algorithms. Deep Learning is more suitable to deal with non-linear data than traditional neural networks. It extracts effective features which are helpful to understand the data or signal through multi-layer network. Deep Learning extracts one or more than one data futures from different scope in each layer, and uses these features as the inputs of next layer, so that the abstract features of higher layers are getting through assembling the data features of lower layers. So far the popular deep learning models include Restricted Boltzmann Machine (RBM), Convolutional Neural Networks (CNNs), Deep Belief Networks (DBNs), Deep Recurrent Neural Networks (DRNNs) etc. Among them Deep Belief Networks (DBNs) has demonstrated the excellent performance[15]. It can deal with the issues which are associated with applying back-propagation to deeply layered neural networks traditionally, such as long learning time, necessity of a substantial labeled dataset for training, and inadequate parameter selection techniques that lead to poor local optima[16]. As for the Long Short Term Memory (LSTM) network proposed by Hochreiter and Schmidhuber [17], it can resolve some of the fundamental mathematical difficulties in modeling long sequences. For these reasons, we use DBN and LSTM to forecast crude oil prices to improve prediction accuracy.

2.1. Deep Belief Network

The Deep Belief Network is constituted by a stack of Restricted Boltzmann Machines, which in turn are Boltzmann Machines (BMs) with a single layer of feature-detecting units. Each RBM perceives pattern representations from its lower layer and learns to encode them in unsupervised fashion. To be specific, each RBM gets output through studying the parameters of last RBM when it trains its own parameters dependently. The structure of Deep Belief Network is shown in Figure 1.

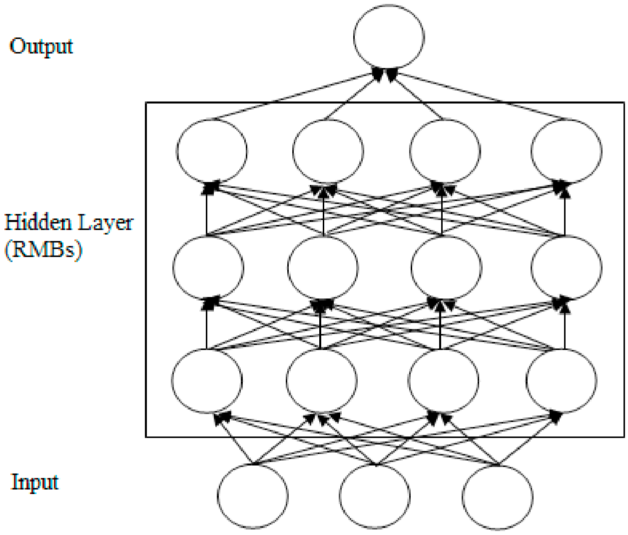

Fig. 1: Deep Belief Network structure

The RBM includes a single visible layer and a single hidden layer. The joint probability for a RBM is defined as (1).

$$P_\theta(v,h) = \frac{1}{Z(\theta)} e^{-E(v,h,\theta)} = \frac{1}{Z(\theta)} \prod_{i,j} e^{W_{ij} v_i h_j} \prod_i e^{b_i v_i} \prod_j e^{a_j h_j} \tag{1}$$

Where the set of visible units is v, the set of hidden units is h. $\theta = \{W, a, b\}$ indicates the parameters, Z is the partition function, which is given by:

$$Z(\theta) = \sum_{v,h} e^{-E(v,h,\theta)}$$

and E is the energy of the joint configuration (v,h), which is defined as (2).

$$E(v,h,\theta) = -a^T h - b^T v - v^T W h = -\sum_{i,j} v_i W_{ij} h_j - \sum_j a_j h_j - \sum_i b_i v_i \tag{2}$$

Where W is a matrix which represents the symmetric interaction between visible unit i and hidden unit j while b_j and a_j are their bias terms. The DBNs greedily train each layer from the lowest to the highest, while each RBM layer independently learns its parameters and the previous layer's activations are used as the inputs of the next layer.

2.2. Long Short Term Memory Network (LSTM)

Long Short Term Memory Network is a new type of Recurrent Neural Network (RNN) [17]. As shown in Figure 2, LSTM contains many subnetworks in a recursive manner. The subnetworks are known as memory cells and each memory cell contains one or more self-loop memory units and three multiplication units. Three multiplication units are represented by input gate, output gate and forget gate. They control the flow of information. Cells are connected recurrently to each other, replacing the usual hidden units of ordinary recurrent networks.

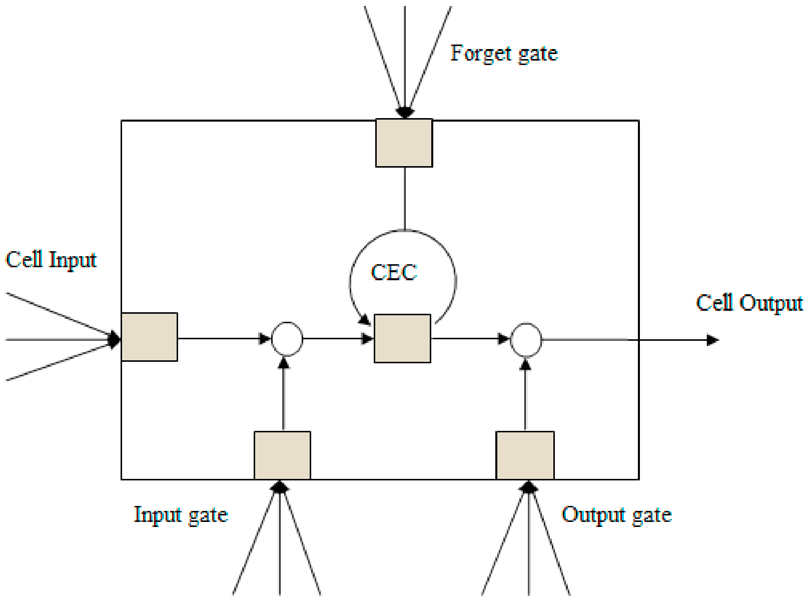

Fig. 2: One memory cell in Long Short Term Memory Networks

For one hidden layer in LSTM, activation function is used in forward propagation, and gradient is used in backward propagation. We use w_{ij} to represent the weight of information transfer from neuron i to j. Moreover, we use I, H and C to represent the number of cells in input layer, the number of cells in hidden layer, and the number of memory cells separately. At time t the input value of neuron j is a^t_j and through the processing of activation function, the output is represented by b^t_j.

If we assume the input is time series x with length T, the functions at different layers in a typical forward propagation LSTM is defined as follows [18].

The input and output values of input gate l at time t are defined as (3) and (4).

$$a^t_l = \sum_{i=1}^{I} w_{il} x^t_i + \sum_{h=1}^{H} w_{hl} b^{t-1}_h + \sum_{c=1}^{C} w_{cl} s^{t-1}_c \tag{3}$$

Where w_{il}, w_{hl} and w_{cl} represent the weight of information transfer from input layer, hidden layer and memory cell to input gate separately.

$$b^t_l = f\left(a^t_l\right) \tag{4}$$

Where f refers to the activation function of input gate.

The input and output values of forget gate ϕ are defined as (5) and (6).

$$a^t_\phi = \sum_{i=1}^{I} w_{i\phi} x^t_i + \sum_{h=1}^{H} w_{h\phi} b^{t-1}_h + \sum_{c=1}^{C} w_{c\phi} s^{t-1}_c \tag{5}$$

Where $w_{i\phi}$, $w_{h\phi}$ and $w_{c\phi}$ represent the weight of information transfer from input layer, hidden layer and memory cell to forget gate separately.

$$b^t_\phi = f\left(a^t_\phi\right) \tag{6}$$

Where f refers to the activation function of forget gate.

The memory cell c is defined as (7) and (8).

$$a_c^t = \sum_{i=1}^{I} w_{ic} x_i^t + \sum_{h=1}^{H} w_{hc} b_h^{t-1} \tag{7}$$

Where w_{ic} and w_{hc} represent the weight of information transfer from input layer and hidden layer to memory cell separately.

$$s_c^t = b_\phi^t s_c^{t-1} + b_l^t g\left(a_c^t\right) \tag{8}$$

Where s_c^t refers to the state information of memory cell c at time t, g refers to the input activation function of memory cell.

The input and output values of output gate ω are defined as (9) and (10).

$$a_\omega^t = \sum_{i=1}^{I} w_{i\omega} x_i^t + \sum_{h=1}^{H} w_{h\omega} b_h^{t-1} + \sum_{c=1}^{C} w_{c\omega} s_c^{t-1} \tag{9}$$

Where $w_{i\omega}$, $w_{h\omega}$ and $w_{c\omega}$ represent the weight of information transfer from input layer, hidden layer and memory cell to output gate separately.

$$b_\omega^t = f\left(a_\omega^t\right) \tag{10}$$

The output of memory cell is defined as (11).

$$b_c^t = b_\omega^t h\left(s_c^t\right) \tag{11}$$

Where h refers to the output activation function of memory cell.

2.3. A Deep Learning based Hybrid Crude oil price Forecasting Model

The deep learning based hybrid crude oil price forecasting model is divided into two stages: the individual forecasting stage and the hybrid forecasting stage. In the individual forecasting stage, we make forecasts based on either time series model and Deep Learning model individually. On one hand, the time series model is used as the baseline model to capture the linear data feature. It may include Random Walk and ARMA models. On the other hand, the deep learning model is used to capture the nonlinear data feature. There are many different deep learning models. Two typical deep learning models are the deep belief network model and the long short term memory model. The forecasting of the nonlinear components is conducted in the following steps:(1) In the typical time series framework, the AIC minimization criterion is used to choose the optimal lag order; (2) Determine the deep learning network structure, including the number of hidden layers, the activation function of the neurons, the learning rate, and the number of trainings. Use the training data to train the network parameters; (3) Forecast the crude oil price for the test set with the trained deep learning network structure, using the rolling windows method.

Secondly, as we assume that the crude oil price movement have both linear and nonlinear dynamics, we combine the forecasts from the ARMA model together with the forecasts from the deep learning models to produce the crude oil price forecasts.

$$y_t = \omega_{lm} \widehat{r_{lm}} + \omega_{nlm} \widehat{r_{nlm}} \tag{12}$$

Where ω refers to the weights for different forecasts. lm refers to the linear models such as ARMA and Random Walk model and $\widehat{r_{lm}}$ refers to the estimate value based on linear models. nlm refers to the employed nonlinear models such as the Deep Belief Network or Long Short Term Model (LSTM) and $\widehat{r_{nlm}}$ refers to the estimate value based on nonlinear models. In this paper we assume the equal contribution that both linear and nonlinear models made, i.e. 0.5 for both coefficients ω.

3. Empirical Studies

This paper uses the historical data of the WTI crude oil market to evaluate the performance of the proposed models. The data are downloaded from the website of Quandl, which is a public domain data warehouse (https://www.quandl.com/). The time span for each crude oil price is from July 23, 2007 to February 24, 2017, with a total of 2409 daily observations. We divided the original data set into two parts, where 70% of the data in the whole dataset was used as a training set for the parameters estimation of deep belief network and LSTM, and the remaining 30% of the data as a test set to evaluate the performance of the models. The data is log differenced to remove the trend factor. It is further standardized during the deep learning training process. The standardization process is reversed at the end of modeling process to produce the actual forecasts. Deep learning models used in this paper are implemented in the TensorFlow framework proposed by Google as well as various python packages such as keras, pandas, etc. The descriptive statistics of mainstream crude oil prices were analyzed by descriptive statistics such as mean, standard deviation, skewness, kurtosis and Jarque-Bera normality test and BDS test as in Table 1.

Table 1: Descriptive statistics and statistical tests

Statistics	$Mean_{\times 10^{-4}}$	Standard Deviation	Skewness	Kurtosis	p_{JB}	p_{BDS}
WTI	4.0207	0.0251	-0.0831	7.6730	0.001	0

Results in Table 1 show the complex nonlinear dynamic data features in the crude oil price data, whose distribution deviates from the usual normal distribution. As far as the downside risk measure is concerned, the skewness is non-zero. The significant kurtosis value indicates the probability of extreme event occurrence. The p value is less than 0.05 cutoff value for both JB test and BDS test. This indicates that the distribution of the crude oil price movement does not conform to the normal distribution and the crude oil price movement may contain nonlinear dependence in data. The parameters of deep learning models are determined as follows. The lag is set to 2, identified using AIC minimization criteria during the time series modeling phase. As for DBN model, the number of hidden layers is set to 4, the number of hidden neuron is set to 100, the input node is 2, the output node is 1, the activation function is sigmod, the learning rate in the model is set to 0.01, the maximum number of iterations is set to 2000 times. As for LSTM model, the number of hidden layers of the LSTM model is set to 2, the hidden neuron is set 4, the input node is 2, the output node is 1, and each node is set to 4, the ADAM algorithm is used as the optimization algorithm of the model, the Mean Squared Error (MSE) method as the target loss function of the model.

Table 2: Out-of-sample Performance Comparison of Different Models

Statistics	RW	ARMA	DBN	RW-DBN	ARMA-DBN	LSTM	RW-LSTM	ARMA-LSTM
$MSE_{\times 10^{-4}}$	5.3235	5.4753	5.3236	5.3224	5.3811	5.5219	5.3868	5.4504
CW_{RW}	N/A	0.9726	0.3655	0.3655	0.9612	0.7794	0.7794	0.9360
CW_{ARMA}	0	N/A	0	0	0	0.0345	0.0004	0.0345

Results in Table 2 show that the optimal performance is achieved using RW-DBN model when the MSEs are the smallest among all models. In general, the MSEs are lower from DBN based model than those from LSTM based model. This indicates the superior forecasting power of DBN model when compared to the LSTM model, in the WTI crude oil market. The superior performance of DBN based model against the benchmark ARMA model is statistically significant. However, it is not statistically significant when the predictive power is tested against the RW model. We found in general the combination of Random Walk or ARMA model with the deep learning model would lead to lower MSE values, which indicates the improvement in the forecasting accuracy. Interestingly the forecasting accuracy of ARMA based combination is lower than the Random Walk based combination. For example, the MSE for ARMA-DBN model is lower than MSE of RW-DBN model.

The forecasted values from the proposed models as well as the original returns are illustrated in Figure 3.

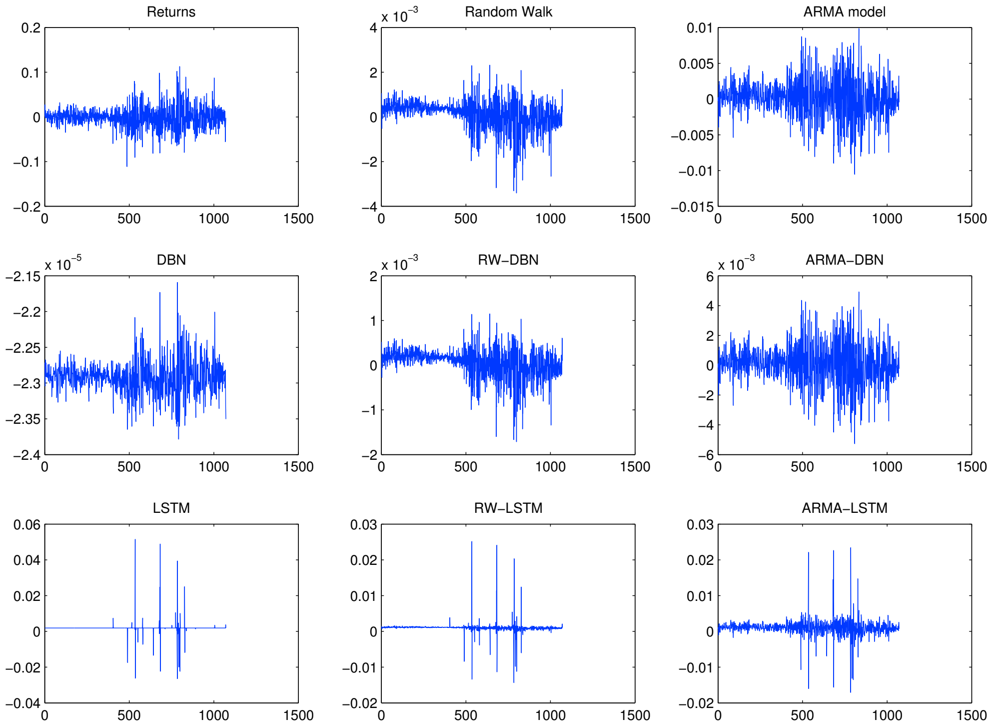

Fig. 3: Forecasted returns of deep learning based forecasting model

It can be seen from Figure 3 that DBN based model tracks the actual returns more closely than the LSTM model. LSTM produces forecasts with significantly larger fluctuation range, with very large value of transient events. This observation is consistent with our analysis of the results in Table 2. In the fast changing crude oil data, the LSTM model may not adapt to changes fast enough to incorporate the new changes available. The market is mostly dominated with shorter term memory behavior.

4. Conclusion

In this paper, we apply the emerging deep learning model to the crude oil price forecast. More specifically we identified two particular deep learning models, i.e. the deep belief network and the recurrent neural network, to be useful in modeling the nonlinear dynamics in the crude oil price movement. We construct a hybrid model that combines the forecasts from ARMA model as well as the forecasts from the deep learning models. We have conducted the empirical studies using the representative WTI crude oil market. We found the introduction of Deep Learning model in the crude oil price models lead to the improved forecasting accuracy. Work in this paper implies that there is exploitable forecasting opportunity in the crude oil price movement. More accurate modeling of the nonlinear dynamics in the crude oil price movement is critical to the further understanding of the determinant underlying the crude oil price movement. In the meantime, we found that the performance of the deep learning model is very sensitive to the parameters. Increasing model complexity with more hidden layers and hidden neurons may not necessarily lead to higher level of nonlinear modeling accuracy. This performance constraint may be attributed to the limited types of deep learning model attempted. It merits further research in constructing some innovative forecasting models based on different types of deep learning models.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (NSFC No. 71671013, No.71201054, No. 71701127), and Humanities and Social Sciences Youth Foundation of Ministry of Education of China (No. 16YJC790026).

References

[1] F. Shen, J. Chao, J. Zhao, Forecasting exchange rate using deep belief networks and conjugate gradient method, Neurocomputing 167 (2015) 243 – 253.
[2] R. Gupta, M. Wohar, Forecasting oil and stock returns with a qual var using over 150 years off data, Energy Economics 62 (2017) 181–186.
[3] D. M. Zhu, W. K. Ching, R. J. Elliott, T. K. Siu, L. M. Zhang, Hidden markov models with threshold effects and their applications to oil price forecasting, Journal of Industrial and Management Optimization 13 (2) (2017) 757–773.
[4] J. Wang, J. Wang, Forecasting energy market indices with recurrent neural networks: Case study of crude oil price fluctuations, Energy 102 (2016) 365–374.
[5] H. Chiroma, S. Abdulkareem, T. Herawan, Evolutionary neural network model for west texas intermediate crude oil price prediction, Applied Energy 142 (2015) 266 – 273.
[6] H. Chiroma, S. Abdul-kareem, A. S. M. Noor, A. I. Abubakar, N. S. Safa, L. Shuib, M. F. Hamza, A. Y. Gital, T. Herawan, A review on artificial intelligence methodologies for the forecasting of crude oil price, Intelligent Automation and Soft Computing 22 (3) (2016) 449–462.
[7] Z. H. Ling, L. Deng, D. Yu, Modeling spectral envelopes using restricted boltzmann machines and deep belief networks for statistical parametric speech synthesis, IEEE Transactions on Audio, Speech, and Language Processing 21 (10) (2013) 2129–2139.
[8] J. B. Heaton, N. G. Polson, J. H. Witte, Deep learning for finance: deep portfolios, Applied Stochastic Models in Business and Industry 33 (1) (2017) 3–12.
[9] M. H. Rafiei, H. Adeli, A novel machine learning model for estimation of sale prices of real estate units, Journal of Construction Engineering and Management 142 (2) (2016) 04015066.
[10] H. Y. Zeng, M. D. Edwards, G. Liu, D. K. Gifford, Convolutional neural network architectures for predicting dna-protein binding, Bioinformatics 32 (12) (2016) 121–127.
[11] A. Dedinec, S. Filiposka, A. Dedinec, L. Kocarev, Deep belief network based electricity load forecasting: An analysis of macedonian case, Energy 115, Part 3 (2016) 1688 – 1700, sustainable Development of Energy, Water and Environment Systems.
[12] G. E. Hinton, R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504.
[13] Y. Bengio, Learning Deep Architectures for AI, Now Publishers, 2009.
[14] H. Lee, R. Grosse, R. Ranganath, A. Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in: International Conference on Machine Learning, 2009, pp. 609 – 616.
[15] G. E. Hinton, S. Osindero, Y. W. Teh, A fast learning algorithm for deep belief nets, Neural Computation 18 (7) (2014) 1527–1554.
[16] I. Arel, D. C. Rose, T. P. Karnowski, Deep machine learning - a new frontier in artificial intelligence research [research frontier], IEEE Computational Intelligence Magazine 5 (4) (2010) 13–18.
[17] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9 (8) (1997) 1735.
[18] F. A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: Continual prediction with lstm. neural computation 12(10): 2451-2471, Neural Computation 12 (10) (2000) 2451–2471.