Weather Forecasting using Deep Learning Techniques

Afan Galih Salman
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
asalman@binus.edu

Bayu Kanigoro
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
Bkanigoro@binus.edu

Yaya Heryadi
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
YayaHeryadi@binus.edu

***Abstract*— Weather forecasting has gained attention many researchers from various research communities due to its effect to the global human life. The emerging deep learning techniques in the last decade coupled with the wide availability of massive weather observation data and the advent of information and computer technology have motivated many researches to explore hidden hierarchical pattern in the large volume of weather dataset for weather forecasting. This study investigates deep learning techniques for weather forecasting. In particular, this study will compare prediction performance of Recurrence Neural Network (RNN), Conditional Restricted Boltzmann Machine (CRBM), and Convolutional Network (CN) models. Those models are tested using weather dataset provided by BMKG (Indonesian Agency for Meteorology, Climatology, and Geophysics) which are collected from a number of weather stations in Aceh area from 1973 to 2009 and El-Nino Southern Oscilation (ENSO) data set provided by International Institution such as National Weather Service Center for Environmental Prediction Climate (NOAA).Forecasting accuracy of each model is evaluated using Frobenius norm. The result of this study expected to contribute to weather forecasting for wide application domains including flight navigation to agriculture and tourism.**

Keywords:deep learning, recurrent neural network, conditional restricted boltzmann machines, convolutional networks, weather

I. INTRODUCTION

Weather forecasting has gained attention many researchers from various research communities due to its effect to the global human life. The current wide availability of massive weather observation data and the advent of information and computer technology in the last decade have motivated many researches to explore hidden pattern in the large dataset for weather prediction.

Weather forecasting is an interesting research problem with wide potential applications ranging from flight navigation to agriculture and tourism. The challenges of weather forecasting, among others, are learning weather representation using a massive volume of weather dataset and building a robust weather prediction model which exploits hidden structural patterns in the massive volume weather dataset.

In the last decade, many significant efforts in weather forecasting using statistical modeling techniques including machine learning have been reported with successful results [1, 2, 3].

The studies on deep belief nets[4] (DBNs),on deep networks[5], on energy-based models[6] have become foundations for the emerging deep learning as deep architecture generative models in the last decade. The word “deep” in deep learning indicates that such neural network (NN) contains more layers than the “shallow” ones as used in typical machine learning models. This multi-layer NN has gained wide research interest after successful implementation of the layer-wise unsupervised pre-training mechanism that is employed to solve the training difficulties efficiently. The “deep” architecture is very significance in compared with shallow models as NN with deep architecture can provide higher learning ability.

The successful applications of deep learning in various domains, which have been reported by various researchers, have motivated its use in weather representation and modeling. The recent study[7], for example, proposes to represent weather using hierarchical feature that are learned from a large volume of weather data using deep neural network (DNN).

The objective of our investigations is to explore the potential of deep learning technique for weather forecasting using rich hierarchical weather representations which are learned from massive weather time series data. The models are tested using a large volume of weather data provided by BMKG (Indonesian Agency for Meteorology, Climatology, and Geophysics) which are collected from a number of weather stations in Aceh area from 1973 to 2009 and El-Nino Southern Oscilation (ENSO) data set provided by International Institution such as National Weather Service Center for Environmental Prediction Climate (NOAA). In this study, several deep learning models for weather forecasting such as: Conditional Restricted Boltzmann Machine (CRBM) and Convolutional Neural Network (CNN) models will be

 978-1-5090-0363-1/15/$31.00

explored and compared with the prominent time series forecasting models such as: Recurrent NN.

The rest of this paper is organized as follows. Chapter 2 describes related of weather forecasting and some method and chapter 3 describe research methodology.

II. Related Works

A. Weather Forecasting Problem

In the last decade, many significant efforts to solve weather forecasting problem using statistical modeling including machine learning techniques have been reported with successful results. Input data for weather forecasting is high-dimensional weather time series data which are collected from a number of weather stations from various regions. The study of Recurrent NN models to forecast regional annual runoff[8]. The study by Chen and Hwang[1] proposes a fuzzy time series model for temperature prediction based on the historical data that are represented by linguistic values. The other study[2] has shown that ensemble of artificial neural networks (ANNs) successfully learns weather patterns based on weather parameters.NN for rainfall forecasting using statistical downscaling[9]. Other research[3] proposes a Chaotic Oscillatory-based Neural Network for short term wind forecasting using LIDAR Data. A long term rainfall forecasting model using Integrated ANN and fuzzy logic wavelet model[10]. The research with the precipitation index drought forecasting using NN, wavelet NN, and Support Vector Regression (SVR) [11]. Rainfall forecasting model based on an ensemble of artificial NN[12].

B. Recurrent Neural Network

Recurrent neural network (RNN) is an artificial NN[13] used for time series prediction. Elman networks is a class of RNN consists of one or more hidden layer. The first layer has the weight that is obtained from the input layer every layer will receive weight from the previous layer. this Elman network has activation function that can be in the form of any function both continue and discontinue. Delay that is happened the first hidden layer in the previous time (t-1) can be used in the current time (t). The unique of the recurrent neural network is the feedback connection which conveys interference information (noise) at the previous input that will be accommodated to the next input.

Let $x(t)$ and $y(t)$ be input and output time series respectively; the three connection weight matrices are W_{IH}, W_{HH}, and W_{HO}

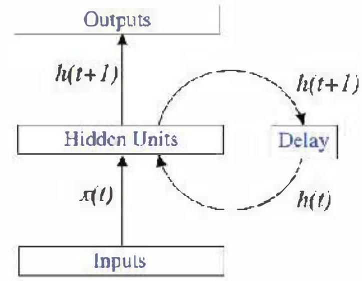

Fig.1. Recurrent Neural Network Model

The RNN parameters are learned using Backpropagation Through Time (BPTT) (Fig.2).

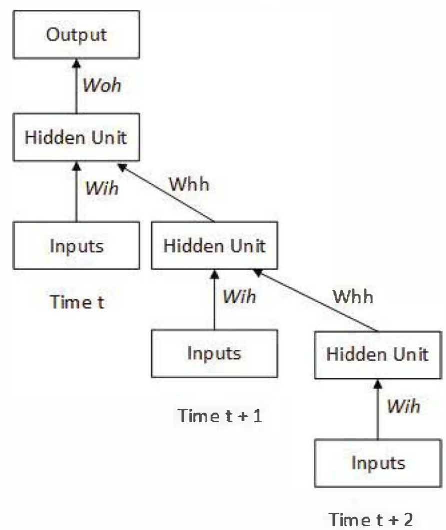

Fig.2. Unfolded Recurrent Neural Network Model

Training sequence starts at time t_0 ends at time t_1 and the sum over time of the standard error function $E_{sse/ct}(t)$at each time-step:

$$E_{total}(t_0, t_1) = \sum_{t=t_0}^{t_1} E_{sse/ce}(t) \quad (1)$$

$$\triangle W_{ij} = -\eta \frac{\partial E_{total}(t_0,t_1)}{\partial w_{ij}} = -\eta \sum_{t=t_0}^{t_1} \frac{\partial E_{total}(t_0,t_1)}{\partial w_{ij}} \quad (2)$$

C. Conditional Restricted Boltzmann Machines (CRBM)

Restricted Boltzmann Machines (CRBMs) are deep learning models to solve various problems such as: collaborative filtering, classification, and modelling motion capture data problems[14].

An RBM defines a joint probability over v and h such that:

$$p(\boldsymbol{v}, \boldsymbol{h}) = \frac{e^{-E(\boldsymbol{v},\boldsymbol{h})}}{Z} \quad (3)$$

 978-1-5090-0363-1/15/$31.00

Where: , v is vector of observed, h is vector of latent (hidden) variables, Z is a normalization constatt and E is an energy function given by:

$$E(\boldsymbol{v},\boldsymbol{h}) = -\boldsymbol{v}^T W h - \boldsymbol{v}^T \boldsymbol{b}^v - \boldsymbol{h}^T \boldsymbol{b}^h \quad (4)$$

In order to obtain $p(v)$, h is marginalized from the joint distribution as follows:

$$p(\boldsymbol{v}) = \frac{1}{Z}\sum_h e^{-E(v,h)} = \frac{1}{Z} e^{-F(v)} \quad (5)$$

Where: W is a matrix of pairwise weights between elements of $v$and h; b^v and b^h are biases for the visible and hidden variables respectively, $F(v)$ is called the free energy and can be computed in time linear proportional to the number of elements in $v$and h.

$$F(v) = -log \sum_h e^{-E(v,h)} = -\boldsymbol{v}^T \boldsymbol{b}^v - \sum_j log\left(1 + e^{b_j^h + v^T W_{.j}}\right) \quad (6)$$

The first practical method for training RBMs[15] is trained using Gradient Descent training algorithm as:

$$\log p(v) = log e^{-F(v)} - log \sum_{v'} e^{-F(v'.)} \quad (7)$$

and differentiating $-\ell(\theta)$with respect to parameter :

$$\frac{\partial -\ell(\theta)}{\partial\theta} = \frac{\partial F(v)}{\partial\theta} - \sum_{v'} \frac{\partial F(v')}{\partial\theta} p(v') \quad (8)$$

A CRBM models the distribution:

$$F(v,u) = -\sum_h log\left(1 + e^{\left(b_j^h + v^T W_j^{vh} + u^T W_j^{uh}\right)}\right) - v^T b^v - u^T W^{uv} v \quad (9)$$

Where:

$$E(v,h,u) = -v^T W^{vh} h - v^T b^v - u^T W^{uv} v - u^T W^{uh} h - h^T b^h \quad (10)$$

The CRBM models the following probability distribution:

$$p(v|u) = \frac{e^{(-F(v,u))}}{\sum_{v'} e^{(-F(v',u))}} \quad (11)$$

Learning in CRBM models involve doing gradient descent in negative log conditional likelihood as follows.

$$\frac{\partial -\ell(\theta)}{\partial\theta} = \frac{\partial F(v|u)}{\partial\theta} - \sum_{v'} \frac{\partial F(v',u)}{\partial\theta} p(v'|u) \quad (12)$$

Once CRBM parameters have been learned, the CRBM can be applied for structured output prediction as reported by [14].

D. *Convolution Network (CN)*

Convolutional network (CN) models[16] , as well as neocognitrons[17,18] known as one of biologically inspired models.

LeCun*et al.*in [17] defines a deep architecture CN may have several stages.In an application of CN using colour images as input, CN usually implemented for domain in feature map would be a 2D array containing a colour channel of the input image (for an audio input each feature map would be a 1D array, and for a video or volumetric image, it would be a 3D array.

The common unsupervised training method for CNN to learn the filter coefficients in the filter bank layers is called Predictive Sparse Decomposition Algorithm[22].

III. Research Methodology

A. *Research Frameworks*

The framework for this study can be described using the following diagram.

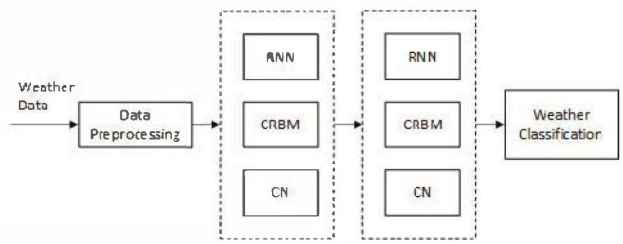

Fig.3. The Research Framework

In this study, the main experiment phases are: training phase of RNN, CRBM, and CN models; and testing each of these models.

B. *Dataset*

The weather data that in this study comprises of:

1) ENSO dataset. ENSO is El Nino Southern Oscillation) dataset comprises of: Wind, Oscillation Index, Sea Surface Temperatur and Outgoing Long Wave Radiation. The data are provided by International Institution such as National Weather Service Center for Environmental Prediction Climate (NOAA).

2) Weather dataset. The weather data set such as Mean Temperature, Max Temperature, Minimum Temperature, Precipitation Temperature, Relative Humidity, Mean Sea Level Pressure, Mean Station Pressure, Visibility , Average Win, Maximum Wind, Wind and Rainfall in Aceh area from 1973 to 2009 provided by BMKG (Indonesian Agency for Meteorology, Climatology, and Geophysics).

Before being used as training dataset, the data are normalized.

C. Experiment Model Training

Three weather forecasting models will be explored in this study which are namely: (i) Recurrence Neural Network (RNN), (ii) Conditional Restricted Boltzmann Machine (CRBM), and (iii) Convolutional Network (CN). Each of these models will be trained and tested using the predetermined weather dataset.

Parameter learning algorithm for each model, for example: gradient descent for CRBM,is implemented to gaintesting error below the predetermined threshold value.

D. Performance Evaluation

Model performance will be evaluated using k-fold Cross-validation technique which runs as follows. First, the weather dataset will be divided randomly into k folds. Each of k-chuck contains approximately m/k of the total data. Compute accuracy (in percent) to estimate the classifier error using the following formula:

$$E = \frac{\sum_{i=1}^{k} n_i}{m} \times 100 \quad (13)$$

Let $E_1,\ldots,E_t$ be the accuracy estimates obtained in t runs. The estimate for the model performance is an error e with standard deviation σ where:

$$e = \frac{1}{t}\sum_{j=1}^{t} E_j \quad (14)$$

$$V = \frac{1}{t-1}\sum_{j=1}^{t}\left(E_j - e\right)^2, \sigma = \sqrt{V} \quad (15)$$

E. Research Roadmap

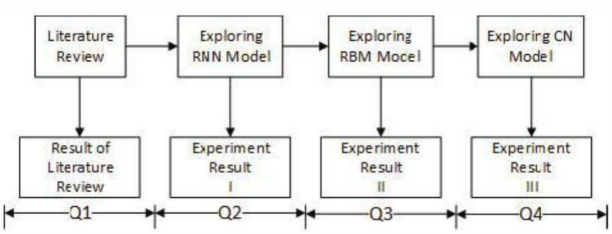

Fig.4. The Research Roadmap

As can be seen in the above figure, this research is planned to be completed in 2 years (4 Quarters).

IV. Result

As a preliminary study has been implemented to explore Recurrent NN using heuristically optimization method for rainfall prediction based on weather dataset comprises of ENSO variables [8]. In this preliminary study, the training dataset are divided into two subsets. The first experiment with training dataset is: 75% data training and 25% data testing. The second experiment with training dataset is divided into: 50% data training and 50% data testing.

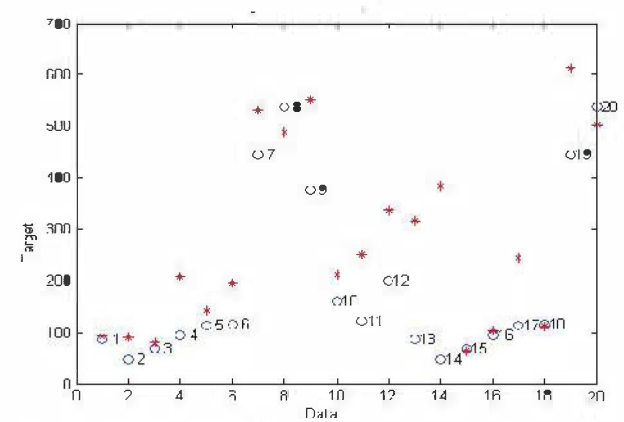

Fig.5. Result of First Experiment

In both experiments, the ENSO variables: wind, SOI,SST and OLR used as independent variables and rainfall as a forecasted variable. Among tested leap 1, 2 and 3, the leap 1 give the best prediction result with resilient backpropagation. The resulted maximum R2 value from the first experiment is 84,8% with RMSE value is 125. While, the maximum maximum R^2value from the second experiment is 59,9% within the R^2 with RMSE value is 155,29.

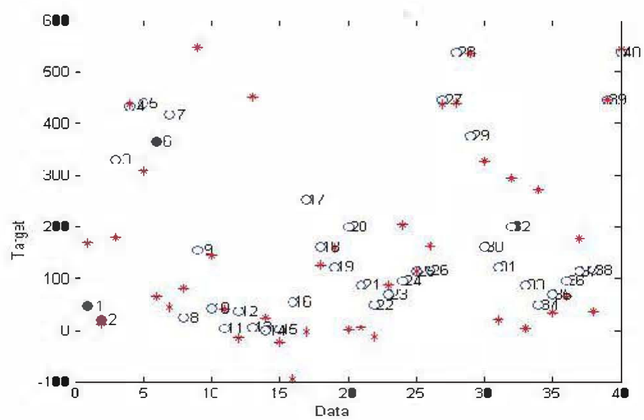

Fig.6. Result of Second Experiment

V. Conclusion

In preliminary experiment recurrent neural network has implemented using heuristically optimization method for rainfall prediction based on weather dataset comprises of ENSO variable. The result is recurrent neural network can be applied in prediction rainfall with adequate accuracy level. We hope that in the next experiment we can be implement the deep Learning methods such as : Conditional Restricted Boltzmann Machine (CRBM) and Convolutional Neural Network (CNN) that offer the accurate representation, classification and prediction on a multitude of time-series problems and compared with the shallow approaches when configured and trained properly.

REFERENCES

[1] Chen, S.-M., and J.-R. Hwang. "Temperature prediction using fuzzy time series." Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 30.2 (2000): 263-275.

[2] Maqsood, I., M. R. Khan, and A. Abraham. "An ensemble of neural networks for weather forecasting." Neural Computing & Applications 13.2 (2004): 112-122.

[3] Kwong, K. M., Liu, J. N. K., Chan, P. W., and Lee, R. "Using LIDAR Doppler velocity data and chaotic oscillatory-based neural network for the forecast of meso-scale wind field," 2008. CEC 2008.(IEEE World Congress on Computational Intelligence). IEEE Congress on (pp. 2012-2019), 2008.

[4] Hinton, G., S. Osindero, and Y.-W.Teh. "A fast learning algorithm for deep belief nets." Neural computation 18.7 (2006): 1527-1554.

[5] Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H. "Greedy layer-wise training of deep networks.,"Advances in neural information processing systems, vol. 19 (153) 2007.

[6] Ranzato , M., Y., Boureau, L., Chopra, S., &LeCun, Y. "A unified energy-based framework for unsupervised learning," In Proc. Conference on AI and Statistics (AI-Stats), vol. 20, 2007.

[7] Liu, J. N. K., Y. Hu, Y. He, P. W. Chan, and L. Lai. "Deep Neural Network Modeling for Big Data Weather Forecasting." Information Granularity, Big Data, and Computational Intelligence. Springer International Publishing, 2015. pp. 389-408.

[8] Salman. "Recurrent Neural Network Modelling using Heuristic Optimization for Rainfall Forecasting using ENSO Variables," InstitutPertanian Bogor, Thesis, 2006.

[9] Normakristagaluh P., "Artificial Forecasting in Statistical Downscaling," InstitutPertanian Bogor, Thesis, 2004.

[10] Afshin. "Long Term Rainfall Forecasting by Integrated Artificial ,Neural Network-Fuzzy Logic-Wavelet Model in Karoon Basin," Scientific Research and Essay vol 6(6),pp.1200-1208; 2011.

[11] Belayneh, A. "Standard Precipitation Index Drought Forecasting Using Neural Networks,Wavelet Neural Networks, and Support Vector Regression,"Hindawi Publishing Corporation Applied Computational Intelligence and Soft Computing Volume, Article ID 794061, 13 pages doi:10.1155/2012/794061; 2012.

[12] Gong, Z. "An Ensemble of Artificial Forecasting," ICTer, 2012

[13] Rumelhart, D., G. Hinton, and R. Williams. "Learning internal representations by error propagation," In Parallel Dist. Proc., pp. 318-362. MIT Press, 1986.

[14] Mnih, V., H. Larochelle, and G. Hinton. "Conditional restricted Boltzmann machines for structured output prediction," In UAI 27, 2011.

[15] Hinton, G. E., and R. R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." Science vol. 313.5786, pp. 504-507, 2006.

[16] LeCun, Y., and Y.Bengio. "Convolutional networks for images, speech, and time series." The handbook of brain theory and neural networks 3361: 310, 1995.

[17] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. "Gradient-based learning applied to document recognition," Proceedings of the IEEE, 86(11), pp. 2278-2324, 1998.

[18] Fukushima, K. "Neocognitron: A hierarchical neural network capable of visual pattern recognition." Neural networks 1.2 pp. 119-130, 1988.

[19] Simard, P. Y., D. Steinkraus, and J. C. Platt. "Best practices for convolutional neural networks applied to visual document analysis." 2013 12th International Conference on Document Analysis and Recognition. Vol. 2. IEEE Computer Society, 2003.

[20] LeCun, Y., F.J. Huang, and L.Bottou. "Learning methods for generic object recognition with invariance to pose and lighting." Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. Vol. 2. IEEE, 2004.

[21] Collobert, R., and J. Weston. "A unified language processing: Deep neural networks with multitask learning." Proceedings of the 25th international conference on Machine learning. ACM, 2008.

[22] Kavukcuoglu, K., M. Ranzato, and Y. LeCun, "Fast inference in sparsecoding algorithms with applications to object recognition," Tech. Rep.,2008, tech Report CBLL-TR-2008-12-01.