Hybrid Model for Weather Forecasting Using
Ensemble of Neural Networks and Mutual Information

Zahra Zargaran1, Abbas Ahmadi2, Azadeh Mohebi3, Farahnaz Taghavi4

1University of Economic Science, 2Amirkabir University of Technology
3Iranian Research Institute for Information Science and Technology,
4University of Tehran, Tehran, Iran

1. INTRODUCTION AND MOTIVATION

Weather Forecasting is the use of science and technology simultaneously in order to predict weather for a specific location. Because the atmosphere circumstances are perceived incompletely, forecasting results are not usually precise and the existing numerical models that are used for this purpose do not provide accurate predictions. Therefore, there is a need for developing more precise models that can produce better results.

Neural networks (NN) are one of the most powerful artificial intelligence methods that are used for nonlinear modeling and especially for weather forecasting. Many researchers first applied different types of neural networks such as multi-layered perceptron (MLP) or radial basis function (RBF) for predicting the weather [1,2,3]. Then, some researchers proposed to use a collection of neural networks instead, in order to forecast the weather and provide more accurate results [4]. In the filed of using multiple neural networks, modular neural network has captured the attention of many researches rapidly since it is able to improve generalization and learning speed in a network [5]. Many researches have applied the notion of combining multiple networks in order to have more accurate results in the field of prediction and forecasting, however there is a lack in addressing the redundancy issue in combining the results of multiple networks. Although some works have been done to address this issue indirectly, eliminating the redundancy of information in ensemble of networks still remains a challenging task. In this paper, a new approach is proposed based on ensemble of neural networks for predicting the weather temperature. In the proposed approach the redundancy caused as a result of combining multiple networks is reduced notably by using a mutual information approach.

2. PROPOSED MODEL

The proposed model for temperature prediction is composed of four different stages. At the first stage data are normalized and useful features are selected. In the second stage, we propose to have an ensemble of four different neural networks; each is trained based on the data provided by the first stage. Then at the third stage, we propose to use mutual information approach in order to select the best networks. Finally, at the last stage the results of the selected networks are combined.
2.1 **Data Normalization and Feature Selection**

To enhance the accuracy of the results, data are first normalized based on Min-Max method, then stepwise regression method is applied for feature selection. Stepwise regression analysis is one of the most common methods, used for feature selection. Feature selection in this method is done through inclusion or exclusion of feature at each step based on an F statistical test.

2.2 **Ensemble of Neural networks**

Two main categories exist in combining neural networks: modular networks and ensemble of networks. In modular network the problem is divide into parts and each part is solved by one module, while in ensemble of networks, a collection of independent networks are considered and each of them learns and solves the whole problem [9]. The proposed model contains ensemble of four neural networks: MLP, RBF, general regression neural network (GRNN) and time delay neural network (TDNN). The main reason for selecting these networks is their good performance in weather forecasting problems based on previous studies. MLP, which is a feedforward neural network uses the back-propagation algorithm for training. RBF is also a feedforward network and a combination of supervised and unsupervised methods is used in the training task. GRNN is a four-layered neural network and it usually works better that MLP while at the same time, it can be trained faster only within a single iteration [8].

2.3 **Mutual information**

Mutual information (MI) is a linear value that is used for measuring correlation between two linear or nonlinear random variables. MI, \( I(X;Y) \), showing the amount of information being shared by \( X \) and \( Y \), is defined as

\[
I(X;Y) = - \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}
\]

Based on this definition, if \( X \) and \( Y \) are closely related, then \( I(X;Y) \) will be large, and will be zero when there is no relationship [6]. We apply mutual information to calculate the correlation between the output of each network and the desired output. Then, the networks that have produced an output closer to the desired one are selected. The outputs of the chosen networks are then combined together at the last stage of the model based on the Winner Take All strategy. It is notable that, while an ensemble of neural networks is considered in the proposed model, the model itself is also modular, since the learning process is modular via a feature selection task (at the first stage).

3. **RESULTS AND COMPARISONS**

The numerical data used in this research consists of six different features of the atmosphere of the city of Tehran, which are Dew Point, Humidity, Sea Level Pressure (hPa), Visibility (Km), Wind Speed (Km/h) and Precipitation (mm). These features are collected from fall 2011 to fall 2012\. Data are divided into three datasets: one for training, one for validation, and one for test to report the error of the proposed model.

---

1 Data are collected from this website [http://www.wunderground.com](http://www.wunderground.com)
After the feature selection at the first stage of the model, four features out of six features are selected, which are: Dew Point, Humidity, Sea Level Pressure and Visibility. The validation set consists of 48 data samples, which are the features related to the first, tenth, twentieth, and thirty day of each month (in the Persian calendar) within a year.

At the second stage of the model, each of the four networks in the model is trained based on the training data. Then at the third stage, the mutual information between the actual output (real temperature) and the output of each network is calculated, as presented in Table 1. The data in the above table can be used to choose the best networks for the last stage of the model. As can be seen from this table, the MI of TDNN and MLP are very close together and less than the two other networks, so the output of these networks will not be considered in the last stage of the model. Therefore, at the last stage, the output of GRNN and RBF are combined based on the Winner Take All strategy.

In order to validate our proposed model, we have compared our model with other types of models in the literature. We have chosen five different methods to compare with: GRNN, RBF, and three different ensembles of networks presented in [4] and [5].

In Figure 1 the output of GRNN, RBF and the proposed model are compared with the actual real temperature, based on the validation dataset. As can be seen from this figure, the proposed model provides more promising results, much closer to the desired output, as compared to the other methods.

In Figure 2 the output of the ensemble of networks is the weighted average (WA) of all of the networks in the ensemble [4]. In [5] two different methods are presented for combining the networks’ results: the first one is based on a competitive network like SOM, and the second one is based on a voting scheme. Table 2 shows mean-squared error (MSE) between the actual temperature and the estimated one for

Table 1. MI between the output of each network and the actual (real temperature) output

<table>
<thead>
<tr>
<th>MI (GRNN)</th>
<th>MI (TDNN)</th>
<th>MI (RBF)</th>
<th>MI (MLP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7461</td>
<td>1.4476</td>
<td>1.6730</td>
<td>1.4584</td>
</tr>
</tbody>
</table>

Figure 1. Estimated temperature based on the proposed method, GRNN, and RBF as compared with the real data
different methods, based on the testing data. As can be seen from the results in this table, the proposed model can provide better results comparing with other types of methods in terms of MSE.

Table 2. Comparing the performance of the proposed method with other methods in terms of MSE

<table>
<thead>
<tr>
<th>Method</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed method</td>
<td>0.0002</td>
</tr>
<tr>
<td>Voting scheme</td>
<td>0.0004</td>
</tr>
<tr>
<td>SOM</td>
<td>0.0010</td>
</tr>
<tr>
<td>RBF</td>
<td>0.0032</td>
</tr>
<tr>
<td>GRNN</td>
<td>0.0042</td>
</tr>
</tbody>
</table>

![Figure 2. Estimated temperature based on the proposed method as compared to WA method in [4]](image)

![Figure 3. Estimated temperature based on the proposed method as compared to SOM and voting scheme in [5]](image)

4. CONCLUSION

In this research, a new model is proposed for weather temperature prediction based on ensembles of neural networks. The proposed model is also a modular model since a feature selection module is considered at the first stage of the model. In addition, to reduce the redundancy caused by considering ensemble of networks, and also decrease the complexity, we have proposed to calculate the correlations between the output of each network and the actual result, based on a mutual information approach. A comparison between the results of the proposed model and some similar models in the literature, shows an outperformance of the proposed model.

5. REFERENCES