Ensemble Deep Learning for Regression and Time Series Forecasting

Xueheng Qui, Le Zhang, Ye Ren and P. N. Suganthan

Jamsrandorj Dagvadorj
Outline

● Motivation
● Forecasting Models
● Ensemble Deep Learning Model
● Experiment
  ○ Dataset
  ○ Methodology
  ○ Result
● Conclusion
Motivation

- An Ensemble of Deep Belief Learning Network is proposed for regression and time series forecasting
- The proposed method outperforms 4 benchmark methods
- Verify the experiment result
Forecasting Models
Support Vector Regression

- A version of Support Vector Machine for regression [1]
- A data set is given as follows

\[ D = \{(X_i, y_i)\}, 1 \leq i \leq N \]

- The regression function can be defined as

\[ f(X_i) = W^T \phi(X_i) + b \]

Feedforward Back Propagation Neural Network (FNN)

- An Artificial Neural Network is a machine learning model inspired by the central nervous system
Feedforward Back Propagation Neural Network (FNN)

- Feedforward Back Propagation Neural Network (FNN)
  - If the network output does not match the required target response, the weights are adjusted in an adaptive manner so that the error is minimized
  - Designed to minimize the mean square error (MSE)

- Recurrent Neural Network (RNN)
  - Additional neurons in the input layer (context neurons) to accept feedback connections.
Deep Belief Network (DBN)

- A type of deep neural network
- No inter-connection between units in each layer
Ensemble Deep Learning Model
Proposed Method - Ensemble Learning Method

- For regression and time series forecasting, the prediction results can be different when the number of epochs of back propagation training is changed.
- Combine all the outputs generated by FNNs trained with different number of epochs.
Proposed Method - Ensemble Learning Method

1. Train a DBN by using the input data matrix X
2. By setting the back propagation epochs from 100 to 2000 with step size equal to 100. (20 prediction results)
3. Put all the outputs into a matrix $X'$
4. Train a SVR with the expected prediction values $Y$
Proposed Method - Ensemble Learning Method
Experiment
System Details

- CSE Labs UNIX machine [1]
  - Processor: Dell OptiPlex 9020 - Intel Core i7 @ 4 GHz
  - RAM: 32GB

[1] https://cseit.umn.edu/lab/keller-1-250
Models

- SVR
- FNN
- DBN
- DBNs + SVR (the proposed method)
Methodology

- Training and testing values are linearly scaled to $[0, 1]$.
- Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are used to examine the accuracy of the prediction model.

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y'_i - y_i)^2}
\]

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y'_i - y_i}{y_i} \right|
\]
Datasets

- The Mackey-Glass dataset [1]
  - A time series generated by the Mackey-Glass equation
    \[
    \frac{dx}{dt} = \beta \frac{x_t}{1 + x_t^n} - \gamma x, \quad \gamma, \beta, n > 0,
    \]

- The electricity load demand data set from Australian Energy Market Operator (AEMO) [2]
  - The data set of year 2013 in New South Wales (NSW)

- California Housing dataset [3]
  - Generated by collecting information on related variables in California from the 1990 Census

Result
Result

- The Mackey-Glass dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR</td>
<td>0.05354</td>
<td>7.002%</td>
</tr>
<tr>
<td>FNN</td>
<td>0.02786</td>
<td>3.391%</td>
</tr>
<tr>
<td>DBN</td>
<td>0.0164</td>
<td>2.432%</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>0.02297</td>
<td>2.721%</td>
</tr>
</tbody>
</table>

Training time: 24 hours
**Result**

- California Housing dataset

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR</td>
<td>0.72877</td>
<td>27.781%</td>
</tr>
<tr>
<td>FNN</td>
<td>0.66508</td>
<td>24.985%</td>
</tr>
<tr>
<td>DBN</td>
<td>0.66306</td>
<td>23.846%</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Training time: 4+ days (running)
## Result

- The electricity load demand data

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR</td>
<td>345.8633</td>
<td>2.858%</td>
</tr>
<tr>
<td>FNN</td>
<td>99.65179</td>
<td>0.908%</td>
</tr>
<tr>
<td>DBN</td>
<td>1193.39</td>
<td>12.557%</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Training time: ?
Conclusion

● The Mackey-Glass dataset
  ○ The proposed method outperforms SVR, FNN
  ○ DBN outperforms the proposed method
    ■ Parameter tuning is needed

● Training time is longer than other methods
  ○ training is not finished