Comparative Study of Forecasting Algorithms for Energy Data

Master Thesis Presentation by
Nishat Fariha Rimi

20th May, 2019

Albert-Ludwigs-Universität Freiburg
Motivation

- Wind and solar energy varies
- What is produced must be used

Benefits of energy demand forecasting—
- Balances supply and demand
- Prevents energy waste
- Reduces operation cost

Comparative analysis of forecasting methods depending on
- Time scale
- Dataset type and sample size

Source: Science direct. smart grid and solar energy

Source: https://www.iass-potsdam.de
Outline

Overview
- Implemented forecasting methods
- Considered forecasting scenarios

Methodology
- Methods
- Forecasting toolbox

Performance analysis
- Performance comparison

Conclusion and Future Work
Goal

➢ **Create**: Structure of a forecasting toolbox

➢ **Compare**: Methods performance according to datasets and forecasting scenarios
Selected Forecasting Approaches

**Statistical Approaches**
- Autoregression
  - Autoregressive moving-average (ARMA)
  - Autoregressive integrated moving average (ARIMA)
- Smoothing
  - Exponential smoothing (ES)
  - Holt-Winters method (HW)

**Machine Learning Approaches**
- Classical Machine Learning
  - K-nearest neighbor (KNN)
  - Random forest (RF)
- Deep Learning
  - Artificial neural network (ANN)
  - Recurrent neural network (RNN)

Considered Scenarios

Performance comparison according to different forecasting aspects.

<table>
<thead>
<tr>
<th>Forecasting horizon</th>
<th>Time scale</th>
<th>Considered time scale</th>
<th>Forecasting sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very short-term</td>
<td>5 min- 1 h</td>
<td>1 h</td>
<td>1</td>
</tr>
<tr>
<td>Short-term</td>
<td>1 h- 24 h</td>
<td>1 d</td>
<td>24</td>
</tr>
<tr>
<td>Medium-term</td>
<td>24 h- weeks</td>
<td>1 w</td>
<td>24*7</td>
</tr>
<tr>
<td>Long-term</td>
<td>month-years</td>
<td>1 m</td>
<td>24<em>7</em>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 m</td>
<td>24<em>7</em>4*3</td>
</tr>
</tbody>
</table>

Training Sample Size

Datasets

PV generation

Electrical load

P. Kuo, Energies, vol. 11, January, pp. 1–13, 2018
Statistical Approaches

Depend on the past values of endogenous variable for forecasting

ARMA \((p,q)\):
- Combination of AR\((p)\) and MA\((q)\) models for stationary time series

ARIMA \((p,d,q)\):
- Transform the non-stationary data into stationary by differencing

ES \((\alpha)\):
- Assign exponentially decreasing weights for past observations

HW \((\alpha, \beta, \gamma)\):
- Design to capture trend and seasonality

Exogenous information and endogenous variables used together

**RF (n\_tree, max\_depth):**
- Constructs multiple decision trees during training

**KNN (k):**
- Searches for a group of k samples nearest based on distance function

**ANN (hidden\_node, hidden layer, epoch):**
- Allows data signals to process the output in one way

**RNN (hidden\_node, hidden layer, epoch):**
- Use internal state (memory) to process sequences of inputs
Optimization of hyperparameter:

Data → Train → Execute → Optimized set of hyperparameter

- **Endogenous**
  - Electrical load
  - PV generation

- **Exogenous**
  - Electrical load time series
  - Hour of the day
  - Diffused Solar radiation
Methodology
Forecasting Toolbox

Files
Optimized parameter of each method

Data pre-processing
Methods

Moving Training

ARMA
ARIMA
RNN
ANN

Testing

Forecast
Save prediction info
Evaluate
Accuracy metrics, Plots
Output
Performance Analysis
Statistical Approaches

- Days Plot – PV Generation
- Forecasting sample size - 1
- Training sample size - 2200

Days Plot – PV Generation
Forecasting sample size - 1
Training sample size - 2200
Performance Analysis
Machine Learning Approaches

➢ Days Plot – PV Generation
➢ Forecasting sample size - 1
➢ Training sample size - 2200
Performance Analysis
All methods

- 1 Day Plot – PV Generation
- **Statistical methods** and **Machine learning methods**
- Forecasting sample size - 1
- Training sample size - 2200
Mean **training time** in **seconds**

**PV generation:** (daily forecasting for 100 training sample size)

<table>
<thead>
<tr>
<th>Model</th>
<th>HW</th>
<th>KNN</th>
<th>ES</th>
<th>ARMA</th>
<th>ARIMA</th>
<th>RF</th>
<th>ANN</th>
<th>RNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.004</td>
<td>0.005</td>
<td>0.006</td>
<td>0.009</td>
<td>0.11</td>
<td>0.29</td>
<td>8.33</td>
<td>32.983</td>
</tr>
</tbody>
</table>

**Electrical load:** (daily forecasting and 100 training sample size)

<table>
<thead>
<tr>
<th>Model</th>
<th>KNN</th>
<th>ES</th>
<th>HW</th>
<th>ARMA</th>
<th>ARIMA</th>
<th>RF</th>
<th>ANN</th>
<th>RNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0084</td>
<td>0.0079</td>
<td>0.25</td>
<td>0.034</td>
<td>0.106</td>
<td>0.35</td>
<td>8.41</td>
<td>48.27</td>
</tr>
</tbody>
</table>

- Training time increases gradually with the increase of training sample sizes

**PC configuration:** Windows 10 computer, with 4 Cores, 8GB of ram, and with 3.4 GHz clock speed.
Predicting Time Comparison

Mean predicting time in seconds

PV generation: (monthly forecasting with 2200 training sample size)

Electrical load: (monthly forecasting with 2200 training sample size)

➢ predicting time has increased gradually with the increase of prediction horizons

PC configuration: Windows 10 computer, with 4 Cores, 8GB of ram, and with 3.4 GHz clock speed.
Performance Analysis

Indicators

- Accuracy Metrics
  - R value
  - RMSE
  - MAE
  - MAPE

- Forecasting Timescale
  - Hour
  - Day
  - Week
  - Month
  - 3 Months

- Training Sample Size
  - 50
  - 100
  - 200
  - 400
  - 800
  - 1600
  - 2200

- Electrical demand

- PV generation

- ARMA
- ARIMA
- ES
- HW
- KNN
- RF
- ANN
- RNN

Computer Science Department, Albert-Ludwigs-Universität Freiburg
RMSE Comparison for PV Generation

Hourly

☐ ARIMA, ARMA
☐ ES

Prediction horizon: 1H, Error measurement: RMSE

Methods
- arima
- arma
- artificial neural network
- exponential smoothing
- holt winters
- k nearest neighbour
- random forest
- recurrent neural network
MAE Comparison for PV Generation

- RF, KNN
- HW, RNN (1600, 2200)
- ES, ARMA

Daily

Prediction horizon: 1D, Error measurement: MAE

Methods
- arima
- arma
- artificial neural network
- exponential smoothing
- holt winters
- k nearest neighbour
- random forest
- recurrent neural network

Sample size: 50, 100, 200, 400, 800, 1600, 2200

MAE: 0.0, 0.5, 1.0, 1.5, 2.0
## RMSE Comparison Summary

### Photovoltaic Dataset

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Models</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hourly</strong></td>
<td>✓ ARIMA, ARMA</td>
<td>- ES</td>
</tr>
<tr>
<td></td>
<td>✓ RF, KNN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>✓ HW, RNN (1600, 2200)</td>
<td>- ARMA, RNN (small sample)</td>
</tr>
<tr>
<td><strong>Daily</strong></td>
<td>✓ RF, KNN, HW, RNN (1600, 2200)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>✓ HW (1600,2200), RNN (2200)</td>
<td>- RNN (small sample)</td>
</tr>
<tr>
<td><strong>Weekly</strong></td>
<td>✓ RF, KNN, HW (1600,2200), RNN (2200)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>✓ RF, KNN, RNN (1600,2200)</td>
<td>- HW</td>
</tr>
<tr>
<td><strong>Monthly</strong></td>
<td>✓ RF, KNN, RNN (1600,2200)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>✓ RF, KNN, RNN</td>
<td>- HW</td>
</tr>
</tbody>
</table>

### Electrical Load Dataset

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Models</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hourly</strong></td>
<td>✓ ARIMA, ARMA</td>
<td>- HW</td>
</tr>
<tr>
<td></td>
<td>✓ RF, KNN, HW</td>
<td></td>
</tr>
<tr>
<td></td>
<td>✓ RNN, HW (&gt;400)</td>
<td>- ARMA, ES</td>
</tr>
<tr>
<td><strong>Daily</strong></td>
<td>✓ RF, KNN, HW (&gt;400)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>✓ Similar to Daily</td>
<td>- Similar to Daily</td>
</tr>
<tr>
<td><strong>Weekly</strong></td>
<td>✓ RF, KNN, RNN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>✓ HW (&gt;200)</td>
<td>- ARMA, ES</td>
</tr>
</tbody>
</table>

---

20.05.2019

Computer Science Department, Albert-Ludwigs-Universität Freiburg
## MAE Comparison Summary

### Photovoltaic Dataset

<table>
<thead>
<tr>
<th>Period</th>
<th>Methods</th>
<th>Similar to Previous Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hourly</td>
<td>✓ ARIMA, ARMA, − ES</td>
<td>−</td>
</tr>
<tr>
<td>Daily</td>
<td>✓ RF, KNN, ✓ HW, RNN (1600, 2200), − ES, ARMA</td>
<td>−</td>
</tr>
<tr>
<td>Weekly</td>
<td>✓ Similar to Daily, − Similar to Daily</td>
<td>−</td>
</tr>
<tr>
<td>Monthly</td>
<td>✓ RF, KNN, ✓ RNN (sample size &gt; 800), − HW</td>
<td>−</td>
</tr>
</tbody>
</table>

### Electrical Load Dataset

<table>
<thead>
<tr>
<th>Period</th>
<th>Methods</th>
<th>Similar to Previous Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hourly</td>
<td>✓ ARIMA, ARMA, − HW</td>
<td>−</td>
</tr>
<tr>
<td>Daily</td>
<td>✓ RF, KNN, ✓ RNN, HW (&gt;400), − ARMA, ES</td>
<td>−</td>
</tr>
<tr>
<td>Weekly</td>
<td>✓ Similar to Daily, − Similar to Daily</td>
<td>−</td>
</tr>
<tr>
<td>Monthly</td>
<td>✓ RF, KNN, ✓ RNN, ✓ HW (&gt;200), − ARMA, ES</td>
<td>−</td>
</tr>
</tbody>
</table>
Conclusion

Comparative analysis of eight forecasting methods –

❑ Prediction horizon
❑ Training sample size
❑ PV generation and electrical load

Hourly

❑ ARMA and ARIMA – optimum choice
❑ Computation time of ARMA < Computation time of ARIMA

Daily, weekly and monthly

❑ RF and KNN
❑ Computation time of KNN < Computation time of RF

HW and RNN - large dependency on sample size
Future Work

Adding other forecasting approaches e.g.
- Support vector regression
- Gaussian process regression etc.

Optimizing parameter with dynamic optimization function
- Genetic algorithm

Training with –
- Datasets of shorter time interval like 15 or 30 minutes
- More datasets / applications
Thank You