Integrating text analysis in electricity load forecasting: initial findings from UK

Yun BAI

Supervised by: Andrea MICHIORRI and Simon CAMAL

yun.bai@minesparis.psl.eu

2022.07.11
CONTENTS

1. GENERAL CONTEXT
2. STATE OF THE ART
3. EARLY RESULTS AND ANALYSIS
4. DISCUSSION
1.1 LOAD FORECASTING WITH EXTERNAL DATA

Weather

Load

Online News

Forecasting models
SVR, AdaBoost, ExtraTrees, LSTM,...

Features from NLP
Sentiment, topics,...

Forecasts

Research targets
- Extract textual features
- Obtain knowledge to improve forecasting
- Enhance the ability of load system to respond to potential risk
CONTENTS

1. GENERAL CONTEXT
2. STATE OF THE ART
3. EARLY RESULTS AND ANALYSIS
4. DISCUSSION
2.1 ELECTRIC LOAD FORECASTING

Classic models
- Peak load and load shape

Machine learning models
- Support Vector Regression

Forecast Combination
- Robust forecasts

Probabilistic Forecasting
- GEFSCom2014

Linear relationship
- Univariate ARIMA, Multivariate ARIMA, Seasonal ARIMA, ARIMA with optimisation...

Nonlinear relationship
- ANN, LSTM

Quantile regression
- Generalized additive models (GAM)
- Conditional kernel density estimation
- Nadaraya–Watson kernel regression

System risks
2.2 NATURAL LANGUAGE PROCESSING

Definition
- Analyse natural language
- Empower machines to perform text or speech processing tasks
- Enable and improve human-machine interaction

Key task
Language model, a process of creating models that predict words or components based on previous words or components.
2.2 NATURAL LANGUAGE PROCESSING

Development of language models

Count based language models

- One-hot encoding
- Vocabulary: Man, woman, boy, girl, prince, queen, king, monarch
- High dimension when there are many documents.

Neural language models

- Forward neural network
- LSTM
- Take single words as input, cannot learn the information of words position.

Word embedding

- Word embedding obtains a higher-level representation of the probability distribution of each word from the total amount of text through a deep network model.
- Word2vec, GloVe, ...

Pretrained models

- Fine-tune with pretrained models, e.g., BERT, Elmo...
- Take sentences as input, learn more context information, and now is widely used.

TF-IDF

\[ w_{i,j} = tf_{i,j} \times \log \left( \frac{N}{df_i} \right) \]

- \( tf_{i,j} \) = number of occurrences of \( i \) in \( j \)
- \( df_i \) = number of documents containing \( i \)
- \( N \) = total number of documents
2.2 NATURAL LANGUAGE PROCESSING

Implicit Interpretable Features from Text

**Sentiments**

- **Polarity**
  - This film is pretty good.
  - Polarity: 1.0
  - Subjectivity: 0.7

- **Subjectivity**
  - Polarity of the sentence: \((1.0 + 0.7)/2 = 0.85\)
  - Subjectivity of the sentence: \((1.0 + 0.4)/2 = 0.7\)

**Topics**

What are these news talking about?

**Latent Dirichlet Allocation (LDA) model**
2.3 FORECASTING WITH TEXTUAL FEATURES

Forecasting electricity load in France and UK with textual weather reports [1]

Summary of UK Weather for Monday 01 February 2016.

Overnight, rain pushed northwards across Scotland. This introduced milder conditions across the whole of the U.K., so it was a mild, cloudy night with drizzle and hill fog in the west [...]
2.3 FORECASTING WITH TEXTUAL FEATURES

Forecasting the ups and downs of stock prices with emotions in social media [2].

Fig. 2: Time series of each emotion from December 1st 2014 to September 16th 2015.

Table 3: Accuracies of SVM-ES on realistic application.

<table>
<thead>
<tr>
<th>Feature</th>
<th>SVM-ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLOSE (3)</td>
<td>56.60%</td>
</tr>
<tr>
<td>OPEN (3)</td>
<td>43.40%</td>
</tr>
<tr>
<td>HIGH (3)</td>
<td>64.15%</td>
</tr>
<tr>
<td>LOW (3)</td>
<td>56.60%</td>
</tr>
<tr>
<td>VOLUME (3)</td>
<td>60.38%</td>
</tr>
<tr>
<td>CLOSE (2)</td>
<td>60.38%</td>
</tr>
<tr>
<td>OPEN (2)</td>
<td>56.60%</td>
</tr>
</tbody>
</table>

Fig. 5: Framework of realistic application for stock prediction based on SVM-ES.
2.3 FORECASTING WITH TEXTUAL FEATURES

Forecasting the WTI crude oil prices with sentiment and topics of futures-related news [3].

Table 6
Forecasting results of multiple methods for crude oil price over h = 1, 2 and 3.

<table>
<thead>
<tr>
<th>Method</th>
<th>Approach</th>
<th>Number of features</th>
<th>$h = 1$</th>
<th>$h = 2$</th>
<th>$h = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>rmse</td>
<td>mae</td>
<td>mapre</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no text</td>
<td>3</td>
<td>0.0728</td>
<td>0.0545</td>
</tr>
<tr>
<td></td>
<td></td>
<td>textblob</td>
<td>36</td>
<td>0.0681</td>
<td>0.0514</td>
</tr>
<tr>
<td></td>
<td></td>
<td>our method</td>
<td>27</td>
<td>0.0683</td>
<td>0.0520</td>
</tr>
<tr>
<td>svr-Li</td>
<td></td>
<td>no text</td>
<td>28</td>
<td>0.0646</td>
<td>0.0475</td>
</tr>
<tr>
<td></td>
<td></td>
<td>textblob</td>
<td>14</td>
<td>0.0586</td>
<td>0.0427</td>
</tr>
<tr>
<td></td>
<td></td>
<td>our method</td>
<td>17</td>
<td>0.0584</td>
<td>0.0428</td>
</tr>
<tr>
<td>arima &amp; arimax</td>
<td>no text</td>
<td>23</td>
<td>0.0565</td>
<td>0.0394</td>
<td>0.0751</td>
</tr>
<tr>
<td></td>
<td></td>
<td>textblob</td>
<td>23</td>
<td>0.0568</td>
<td>0.0401</td>
</tr>
<tr>
<td></td>
<td></td>
<td>our method</td>
<td>30</td>
<td>0.0577</td>
<td>0.0412</td>
</tr>
<tr>
<td>ada</td>
<td>no text</td>
<td>3</td>
<td>0.0594</td>
<td>0.0428</td>
<td>0.0799</td>
</tr>
<tr>
<td></td>
<td></td>
<td>textblob</td>
<td>7</td>
<td>0.0565</td>
<td>0.0397</td>
</tr>
<tr>
<td></td>
<td></td>
<td>our method</td>
<td>12</td>
<td>0.0564</td>
<td>0.0396</td>
</tr>
</tbody>
</table>

Fig. 3. The framework of crude oil price forecasting.
2.3 FORECASTING WITH TEXTUAL FEATURES

Forecasting taxi demand in two regions of New York with the events text online [4].

Table 2
Average results and standard deviations for the Barclays center area over 30 executions of the different approaches.

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE</th>
<th>RMSE</th>
<th>MAPE</th>
<th>$R^2$ ($\times 10^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR L</td>
<td>120.7 (± 0.0)</td>
<td>167.5 (± 0.0)</td>
<td>15.4 (± 0.0)</td>
<td>42.8 (± 0.0)</td>
</tr>
<tr>
<td>SVR L + W</td>
<td>120.1 (± 0.0)</td>
<td>166.4 (± 0.0)</td>
<td>15.4 (± 0.0)</td>
<td>45.6 (± 0.0)</td>
</tr>
<tr>
<td>SVR L + W + E</td>
<td>97.9 (± 0.0)</td>
<td>139.3 (± 0.0)</td>
<td>12.6 (± 0.0)</td>
<td>60.5 (± 0.0)</td>
</tr>
<tr>
<td>GP L</td>
<td>145.2 (± 0.0)</td>
<td>199.4 (± 0.0)</td>
<td>16.3 (± 0.0)</td>
<td>18.9 (± 0.0)</td>
</tr>
<tr>
<td>GP L + W</td>
<td>127.3 (± 0.0)</td>
<td>176.9 (± 0.0)</td>
<td>15.2 (± 0.0)</td>
<td>36.2 (± 0.0)</td>
</tr>
<tr>
<td>GP L + W + E</td>
<td>100.0 (± 0.0)</td>
<td>141.3 (± 0.0)</td>
<td>12.8 (± 0.0)</td>
<td>50.4 (± 0.0)</td>
</tr>
<tr>
<td>DL-LSTM L</td>
<td>129.0 (± 2.6)</td>
<td>172.6 (± 4.3)</td>
<td>15.9 (± 2.0)</td>
<td>39.2 (± 3.1)</td>
</tr>
<tr>
<td>DL-LSTM L + W</td>
<td>128.9 (± 1.9)</td>
<td>167.8 (± 2.0)</td>
<td>16.5 (± 0.2)</td>
<td>42.6 (± 1.4)</td>
</tr>
<tr>
<td>DL-LSTM L + W + E</td>
<td>103.0 (± 1.2)</td>
<td>141.5 (± 1.6)</td>
<td>12.3 (± 0.3)</td>
<td>59.2 (± 0.9)</td>
</tr>
<tr>
<td>DL-LSTM L + W + E + T</td>
<td>100.4 (± 1.8)</td>
<td>139.5 (± 2.0)</td>
<td>12.8 (± 0.3)</td>
<td>60.3 (± 1.2)</td>
</tr>
<tr>
<td>DL-FC L</td>
<td>120.8 (± 1.1)</td>
<td>161.7 (± 0.6)</td>
<td>15.8 (± 0.3)</td>
<td>46.7 (± 0.4)</td>
</tr>
<tr>
<td>DL-FC L + W</td>
<td>119.8 (± 1.0)</td>
<td>160.5 (± 0.6)</td>
<td>15.6 (± 0.2)</td>
<td>47.5 (± 0.4)</td>
</tr>
<tr>
<td>DL-FC L + W + E</td>
<td>95.6 (± 0.9)</td>
<td>134.8 (± 0.8)</td>
<td>12.7 (± 0.2)</td>
<td>63.0 (± 0.4)</td>
</tr>
<tr>
<td>DL-FC L + W + E + T</td>
<td>93.2 (± 0.8)</td>
<td>132.3 (± 0.9)</td>
<td>12.3 (± 0.1)</td>
<td>64.3 (± 0.5)</td>
</tr>
</tbody>
</table>

$L$ for lags, $W$ for weather, $E$ for if the event present, $T$ for textual information.

Fig. 4. Proposed neural network architecture with FC layers for modelling the time-series observations (DL-FC).
CONTENTS

1. GENERAL CONTEXT
2. STATE OF THE ART
3. EARLY RESULTS AND ANALYSIS
4. DISCUSSION
3.1 MODELING

**Level 1**
- Electric load from ENTSOe transparency platform
- Textual data from BBC news

**Level 2**
- Normalize Data;
- Fill missing values;
- Resample data as given frequency;
- Design models and parameters.

**Level 3**
- Statistical methods
  - Count features
  - Words frequency
  - Sentiment scores
  - Topic distribution
  - GloVe embeddings
  - Granger test
- Forecast and evaluation
- Select features useful for load forecasting
3.2 ACQUISITION AND DATA ANALYSIS

Electric load data [5]

News text data [6]
3.3 TEXTUAL FEATURES

Features obtained by statistical methods

**Count features**

- Count numbers of words, sentences, pieces of news, unique words, non-stopping words, ‘electric’ words
- Count average length of sentences and news (with unit words)
- Section type (Asia, Europe, Scotland, …)
- 27 features selected for titles, descriptions and text bodies

**Word frequency**

- Count the number of documents that a certain word appear.
- Select words appear in more than 500 titles, 1000 descriptions and 10000 text bodies, ensuring final number of words is balanced.
- 227 words for titles, 178 words for descriptions, 330 words for text bodies
3.3 TEXTUAL FEATURES

Features obtained by NLP techniques

**Sentiment scores**
- Compute the polarity and subjectivity scores of a document with TextBlob dictionary.
- Consider the quantile (1~5) of scores.
- Description statistics of both scores. (min, max, mean, std)
- 18 features selected for titles, descriptions and text bodies.

**Sentiment STL scores**
- Decompose time series into trend, seasonal and residual terms with STL.
- 4 features all types of news.
3.3 TEXTUAL FEATURES

Features obtained by NLP techniques

**Topic distributions**
- Learn the topic distribution with LDA.
- 50 for titles, 43 for descriptions, 47 for text bodies.

**GloVe embeddings**
- Represent each word with a 100d GloVe embedding pre-trained model.
- Average the words vectors in a documents to get the text vectors.
- 100 features all types of news.
3.4 SENTIMENT FEATURES AND SELECTION

Granger causality test

\[
x_t = c_1 + \sum_{i=1}^{3} \alpha_{1,i} y_{t-i} + \sum_{i=1}^{3} \beta_{1,i} x_{t-i} + \epsilon_{x,t}
\]

\[
y_t = c_2 + \sum_{i=1}^{3} \alpha_{2,i} y_{t-i} + \sum_{i=1}^{3} \beta_{2,i} x_{t-i} + \epsilon_{y,t}
\]

---

<table>
<thead>
<tr>
<th>Textual features type</th>
<th>Num of features</th>
<th>Num of selected features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count features</td>
<td>27</td>
<td>6</td>
</tr>
<tr>
<td>Word frequency</td>
<td>227</td>
<td>30</td>
</tr>
<tr>
<td>Sentiment scores</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>Sentiment STL scores</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Topic distributions</td>
<td>50</td>
<td>7</td>
</tr>
<tr>
<td>GloVe embeddings</td>
<td>100</td>
<td>17</td>
</tr>
<tr>
<td>Count features</td>
<td>27</td>
<td>6</td>
</tr>
<tr>
<td>Word frequency</td>
<td>178</td>
<td>30</td>
</tr>
<tr>
<td>Sentiment scores</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>Sentiment STL scores</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Topic distributions</td>
<td>43</td>
<td>4</td>
</tr>
<tr>
<td>GloVe embeddings</td>
<td>100</td>
<td>23</td>
</tr>
<tr>
<td>Count features</td>
<td>27</td>
<td>5</td>
</tr>
<tr>
<td>Word frequency</td>
<td>330</td>
<td>65</td>
</tr>
<tr>
<td>Sentiment scores</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Sentiment STL scores</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Topic distributions</td>
<td>47</td>
<td>15</td>
</tr>
<tr>
<td>GloVe embeddings</td>
<td>100</td>
<td>9</td>
</tr>
</tbody>
</table>
3.5 DETERMINISTIC FORECASTING

SVR, AdaBoost, LSTM models on RMSE index

---

**SVR model for forecasting on RMSE**

- Titles-count
- Titles-sentiment
- Titles-word freq
- Titles-topic
- Titles-glove
- Des-count
- Des-sentiment
- Des-word freq
- Des-topic
- Des-glove
- Des-all
- Body-count
- Body-word freq
- Body-topic
- Body-glove
- Body-all
- All text
- No text
- Body-sentiment

**AdaBoost model for forecasting on RMSE**

- Titles-count
- Titles-sentiment
- Titles-word freq
- Titles-topic
- Titles-glove
- Des-count
- Des-sentiment
- Des-word freq
- Des-topic
- Des-glove
- Des-all
- Body-count
- Body-word freq
- Body-topic
- Body-glove
- Body-all
- All text
- No text
- Body-sentiment

**LSTM model for forecasting on RMSE**

- Titles-count
- Titles-sentiment
- Titles-word freq
- Titles-topic
- Titles-glove
- Des-count
- Des-sentiment
- Des-word freq
- Des-topic
- Des-glove
- Des-all
- Body-count
- Body-word freq
- Body-topic
- Body-glove
- Body-all
- All text
- No text
- Body-sentiment
3.5 DETERMINISTIC FORECASTING

SVR, AdaBoost, LSTM models on MAE index

SVR model for forecasting on MAE

AdaBoost model for forecasting on MAE

LSTM model for forecasting on MAE

DETERMINISTIC FORECASTING

SVR, AdaBoost, LSTM models on MAE index
3.5 DETERMINISTIC FORECASTING

Relationship analysis

<table>
<thead>
<tr>
<th>Models</th>
<th>RMSE Textual features type</th>
<th>Improvement percentage</th>
<th>MAE Textual features type</th>
<th>Improvement percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR</td>
<td>Des-GloVe embeddings</td>
<td>5.44%</td>
<td>Des-GloVe embeddings</td>
<td>4.99%</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>Des-GloVe embeddings</td>
<td>3.74%</td>
<td>Des-GloVe embeddings</td>
<td>4.09%</td>
</tr>
<tr>
<td>LSTM</td>
<td>Titles-Word frequency</td>
<td>8.12%</td>
<td>Body-count features</td>
<td>11.47%</td>
</tr>
</tbody>
</table>

- 23 features from 100 in Des-GloVe embeddings
- dim3/dim4/dim5/dim9...

Relationship of load and 3rd dimension from GloVe model
3.5 DETERMINISTIC FORECASTING

Relationship analysis

<table>
<thead>
<tr>
<th>Models</th>
<th>RMSE</th>
<th>Improvements percentage</th>
<th>MAE</th>
<th>Improvement percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Textual features type</td>
<td></td>
<td>Textual features type</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Improvement percentage</td>
<td></td>
<td>Improvement percentage</td>
<td></td>
</tr>
<tr>
<td>SVR</td>
<td>Des-GloVe embeddings</td>
<td>5.44%</td>
<td>Des-GloVe embeddings</td>
<td>4.99%</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>Des-GloVe embeddings</td>
<td>3.74%</td>
<td>Des-GloVe embeddings</td>
<td>4.09%</td>
</tr>
<tr>
<td>LSTM</td>
<td>Titles-Word frequency</td>
<td>8.12%</td>
<td>Body-count features</td>
<td>11.47%</td>
</tr>
</tbody>
</table>

- 30 features from 227 in Titles-Word frequency
3.5 DETERMINISTIC FORECASTING

Relationship analysis

<table>
<thead>
<tr>
<th>Models</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Textual features type</td>
<td>Improvement percentage</td>
</tr>
<tr>
<td>SVR</td>
<td>Des-GloVe embeddings</td>
<td>5.44%</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>Des-GloVe embeddings</td>
<td>3.74%</td>
</tr>
<tr>
<td>LSTM</td>
<td>Titles-Word frequency</td>
<td>8.12%</td>
</tr>
</tbody>
</table>

- 5 features from 27 in Body-count features
- ‘section_Europe’, ‘section_Family_Education’, ‘section_Scotland’, 'section_UKPolitics', ‘section_US_Canada’
3.6 PROBABILISTIC FORECASTING

Results from ExtraTrees Regression

<table>
<thead>
<tr>
<th>Textual features type</th>
<th>Improvement percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRPS Des-GloVe embeddings</td>
<td>4.80%</td>
</tr>
<tr>
<td>Sharpness Titles-sentiment</td>
<td>9.36%</td>
</tr>
</tbody>
</table>

For Titles-sentiment:

- 6 features from 18
- polarity_range(0.6,0.8], polarity_range(0.8,1], polarity_mean, subject_range(0.4,0.6], subject_title_max, subject_title_min'

Relatively positive sentiment, and the extreme values of subjective will help improve the sharpness in probabilistic forecasting.
CONTENTS

1. GENERAL CONTEXT
2. STATE OF THE ART
3. EARLY RESULTS AND ANALYSIS
4. DISCUSSION
4.1 DISCUSSION

• Design of a relatively novel time series forecasting model for electrical loads, considering news text features.

• A number of interpretable and non-interpretable features were obtained from hundreds of thousands of news items as exogenous variables in the time series using simple statistics and natural language processing techniques.

• Granger causality tests were used to filter out text features that would be useful for load forecasting.

• Results for deterministic forecasting showed that text features improved long-term forecasting (h=365) by about 5% and that global word vector representations (GloVe) from news descriptions (abstracts) are more effective.

• Results for probabilistic forecasting showed that feeding all Granger-tested text features into the ExtraTrees model performed clearly better on the Pinball loss metric. The CRPS and Sharpness metrics, on the other hand, preferred Glove word vectors from news descriptions and sentiment features from news headlines, respectively.
4.2 REFERENCES


Thanks for listening

Integrating text analysis in electricity load forecasting: initial findings from UK

Yun BAI

Supervised by: Andrea MICHIORRI and Simon CAMAL

yun.bai@minesparis.psl.eu