



PSL  PERSEE

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

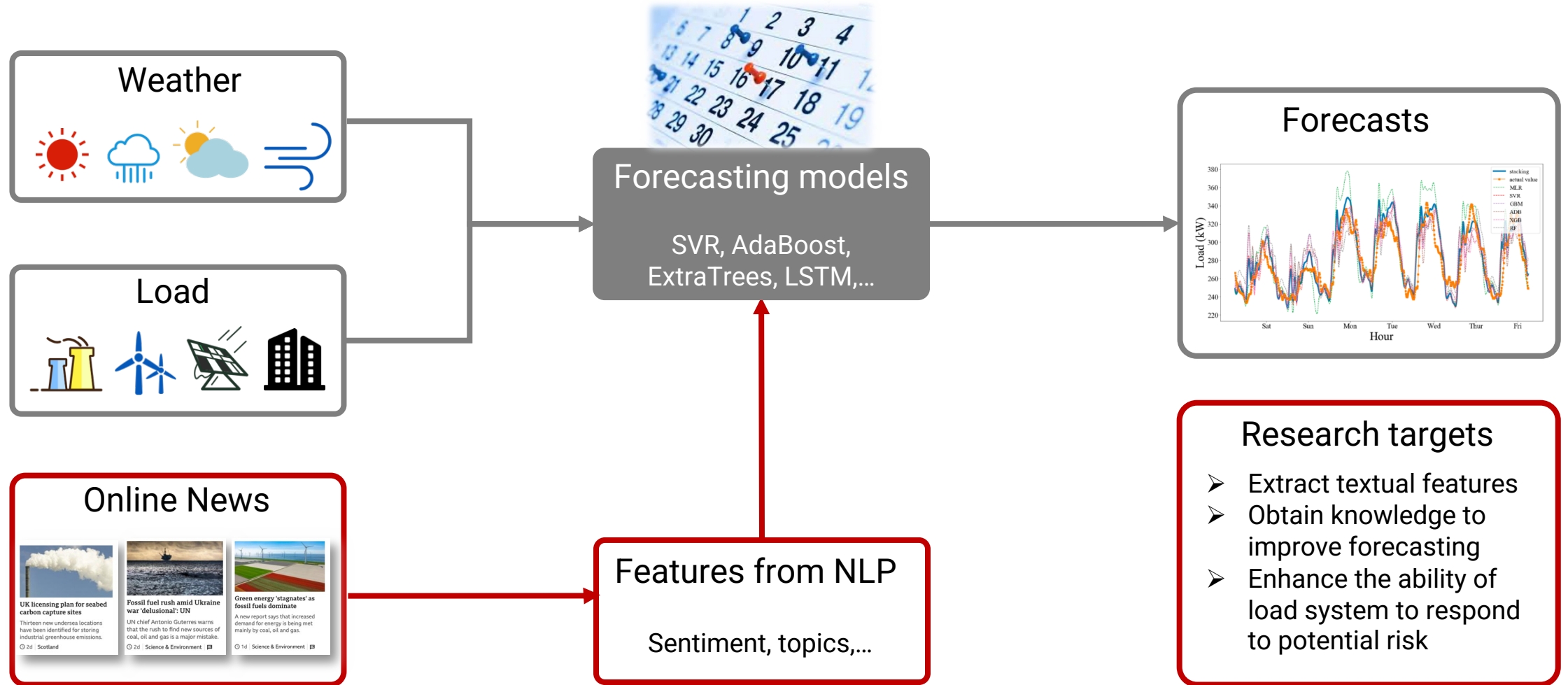
1. GENERAL CONTEXT

2. STATE OF THE ART

3. EARLY RESULTS AND ANALYSIS

4. DISCUSSION

1.1 LOAD FORECASTING WITH EXTERNAL DATA



CONTENTS

1. GENERAL CONTEXT

2. STATE OF THE ART

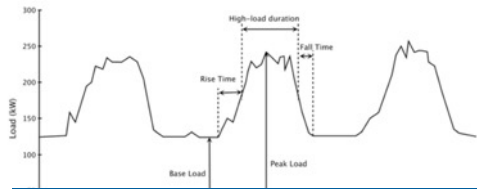
3. EARLY RESULTS AND ANALYSIS

4. DISCUSSION

2.1 ELECTRIC LOAD FORECASTING

Classic models

- Peak load and load shape



Linear relationship

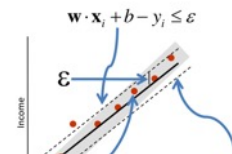
$$\left(1 - \sum_{i=1}^p \varphi_i L^i\right) (1-L)^d X_t = \delta + \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t.$$

Univariate ARIMA,
Multivariate ARIMA,
Seasonal ARIMA,
ARIMA with optimisation...



Machine learning models

- Support Vector Regression

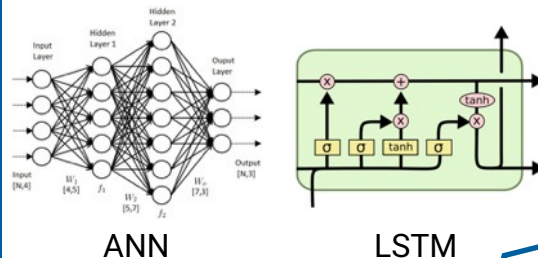


$$\min \frac{1}{2} \|w\|^2$$

$$s.t. y_i - w_1 \cdot x_i - b \leq \varepsilon;$$

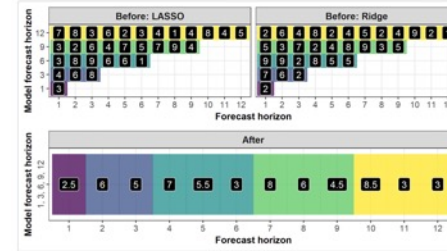
$$w_1 \cdot x_i + b - y_i \leq \varepsilon;$$

Nonlinear relationship



-

Forecast Combination



Robust forecasts

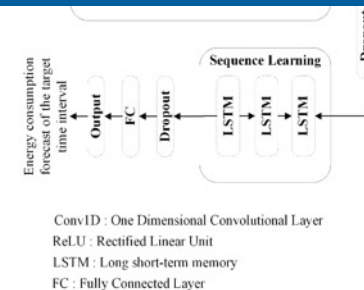
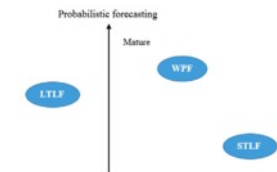


FIGURE 5. Proposed deep CNN-LSTM framework.

Probabilistic Forecasting

GEFCom2014



System risks

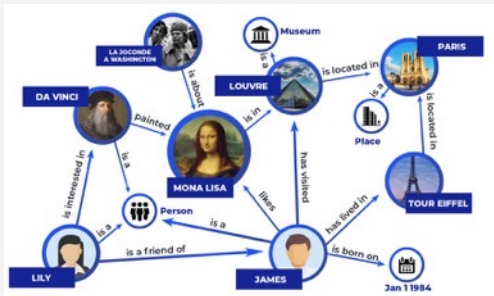
- Quantile regression
- Generalized additive models (GAM)
- Conditional kernel density estimation
- Nadaraya-Watson kernel regression
-

2.2 NATURAL LANGUAGE PROCESSING

Question & answering system



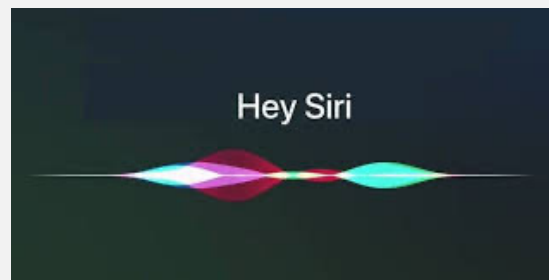
Knowledge graph



Identification of spam emails



Voice Assistant



Definition

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

Key task

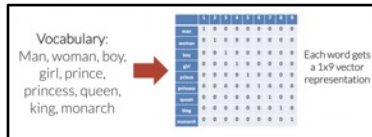
Language model, a process of creating models that predict words or components based on previous words or components.

2.2 NATURAL LANGUAGE PROCESSING

Development of language models

Count based language models

One-hot encoding



High dimension when there are many documents.



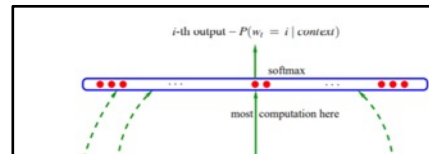
$N = 3$: [this|is|a|sentence] trigrams: [this is a, is a sentence]

TFIDF

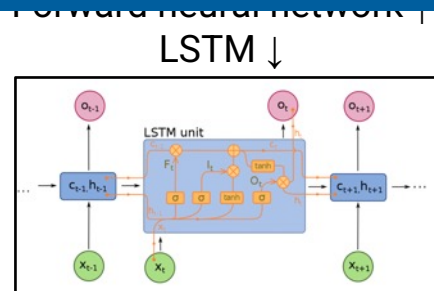
$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j
 df_i = number of documents containing i
 N = total number of documents

Neural language models



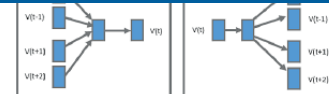
One should train different models for different tasks.



Word embedding

Word embedding obtains a higher-level representation of the probability

Take single words as input, cannot learn the information of words position.



A quick brown fox jumps over the lazy dog

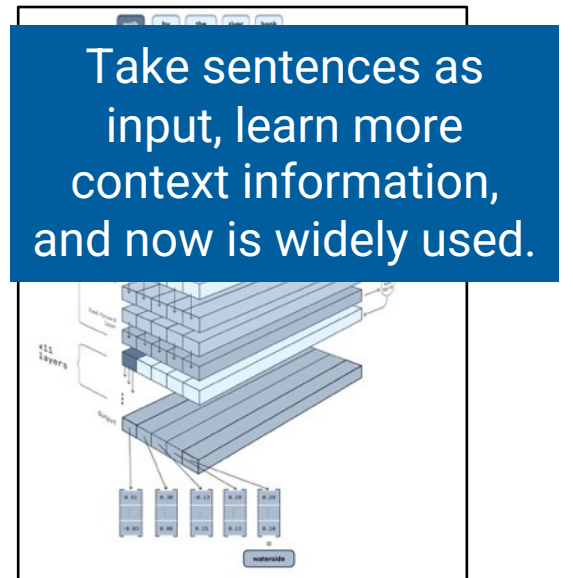
Skip-gram

A quick brown fox jumps over the lazy dog

Word2vec, GloVe,...

Pretrained models

Fine-tune with pretrained models, e.g., BERT, Elmo...



2.2 NATURAL LANGUAGE PROCESSING

Implicit Interpretable Features from Text

Sea level 200,000 p
A report int houses in E flooding in

Eight million Australi urged to turn off light
Australia's energy minister urged households to use li electricity in the face of a p-crisis.

Green energy 'stagnates' as fossil fuels dominate
says that increased energy is being met

UK licensing plan for seabed carbon capture sites
Thirteen new undersea locations have been identified for storing industrial greenhouse emissions.

Fossil fuel rush amid Ukraine war 'delusional': UN
UN chief Antonio Guterres warns that the rush to find new sources of coal, oil and gas is a major mistake.

Sentiments

Polarity

Subjectivity

TextBlob with a massive dictionary

This film is **pretty good**.

Polarity	1.0	0.7
Subjectivity	1.0	0.4

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

Collection of text documents → Dataset → LDA Model → Creation of Topics (1, 2, 3, 4) → Frequency of topics per document (D1, D2, D3, D4)

Parameters: α (Dirichlet parameter), θ (Document-topic distribution), z (Word-topic assignment), w (Observed Word), N (Number of documents), M (Number of words in a given document)

Kinds of News on Facebook

Percent of Facebook News Consumers who Regularly See News on Facebook about...

Entertainment	73%
People & events in my community	65
Sports	57
National govt & politics	55
Crime	51
Health & medicine	46
Local govt & politics	44
Local weather & traffic	42
International news	39
Science & technology	37
Business	31

Based on Facebook News Consumers (N=1,429)
Facebook News Survey Aug. 21-Sept. 2, 2013.
PEW RESEARCH CENTER

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 [...]

Table 1: Excerpt of the Met Office report of Monday the 1st February 2016.

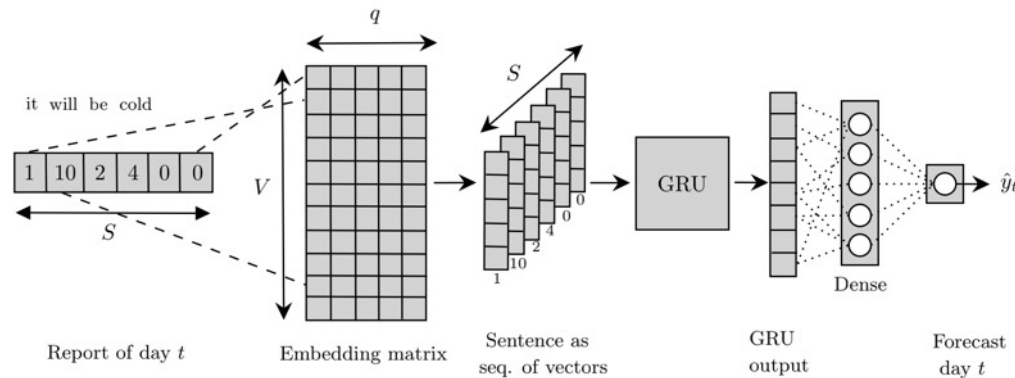


Figure 3: Structure of our RNN. Dropout and batch normalization are not represented.

Method	Encoding	MAPE (%)	RMSE (MW)	MAE (MW)	R^2
Benchmark	-	2.74 ± 0.01	2152.86 ± 5.51	1533.65 ± 3.38	0.96 ± 0.00
RF	TF-IDF	5.40 ± 0.02	4386.67 ± 15.67	3052.75 ± 9.07	0.85 ± 0.00
RF (sel)		5.28 ± 0.01	4168.68 ± 12.13	2977.76 ± 6.28	0.86 ± 0.00
MLP		5.09 ± 0.05	3720.94 ± 27.29	2795.61 ± 25.78	0.89 ± 0.00
LASSO		6.81 ± 0.00	4829.99 ± 0.00	3717.84 ± 0.00	0.82 ± 0.00
CNN	Embedding	6.18 ± 0.34	4493.08 ± 171.40	3359.61 ± 165.39	0.84 ± 0.01
RNN		4.83 ± 0.11	3577.04 ± 64.46	2672.12 ± 52.41	0.90 ± 0.00
Agg (RNN + MLP)	Both	4.66 ± 0.04	3470.53 ± 22.25	2581.94 ± 21.67	0.91 ± 0.00

Table 4: Forecast errors on the net load for the French Dataset.

Method	Encoding	MAPE (%)	RMSE (MW)	MAE (MW)	R^2
Benchmark	-	2.28 ± 0.00	1013.83 ± 2.16	754.53 ± 1.45	0.94 ± 0.00
RF	TF-IDF	3.19 ± 0.01	1544.81 ± 3.34	1058.16 ± 2.66	0.85 ± 0.00
RF (sel)		3.22 ± 0.01	1545.04 ± 3.90	1072.00 ± 3.19	0.85 ± 0.00
MLP		3.68 ± 0.13	1651.10 ± 36.46	1209.16 ± 39.34	0.83 ± 0.01
LASSO		4.70 ± 0.00	2257.32 ± 0.00	1519.35 ± 0.00	0.69 ± 0.00
CNN	Embedding	4.23 ± 0.24	1910.73 ± 66.39	1387.22 ± 69.37	0.78 ± 0.02
RNN		3.55 ± 0.08	1695.17 ± 40.82	1182.31 ± 26.41	0.82 ± 0.01
Agg (RF + RF Sel)	TF-IDF	3.18 ± 0.01	1532.75 ± 2.66	1056.55 ± 2.42	0.86 ± 0.00

Table 5: Forecast errors on the net load for the British Dataset.

2.3 FORECASTING WITH TEXTUAL FEATURES

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

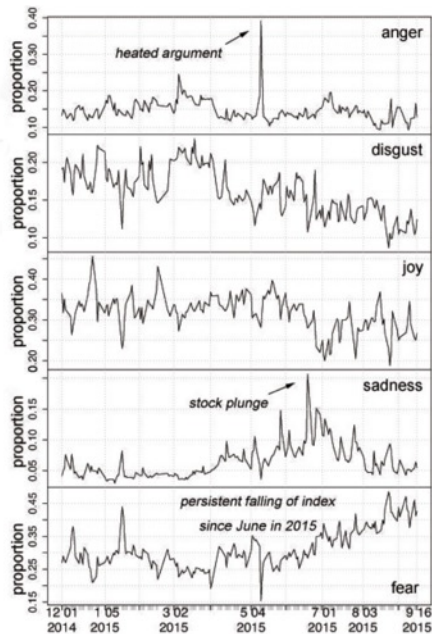


Fig. 2: Time series of each emotion from December 1st 2014 to September 16th 2015.

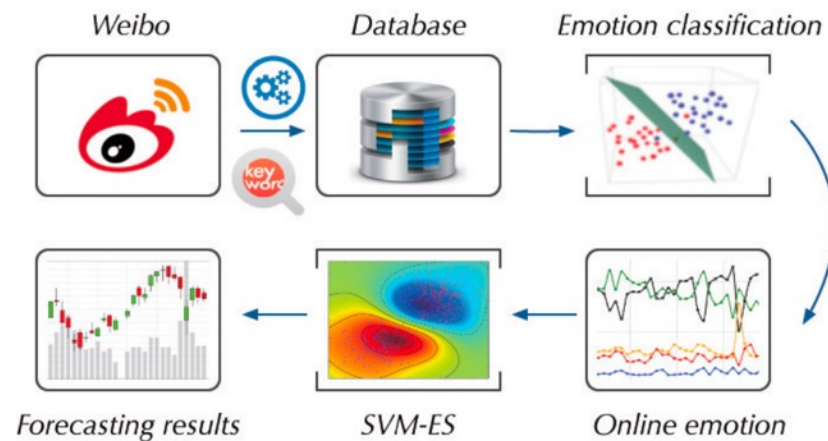


Fig. 5: Framework of realistic application for stock prediction based on SVM-ES.

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

CLOSE (3)	OPEN (3)	HIGH(3)	LOW(3)	VOLUME(3)	CLOSE (2)	OPEN (2)
56.60%	43.40%	64.15%	56.60%	60.38%	60.38%	56.60%

2.3 FORECASTING WITH TEXTUAL FEATURES

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

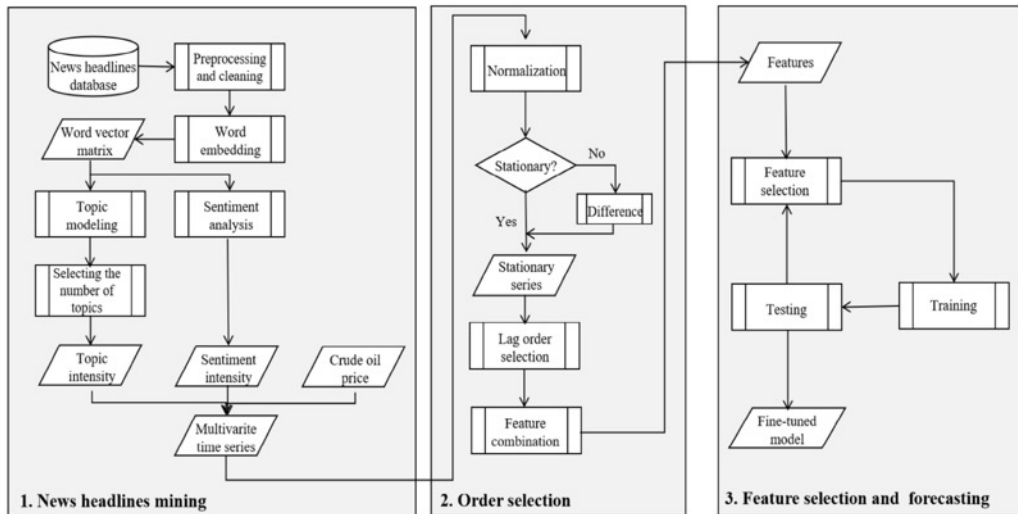


Fig. 3. The framework of crude oil price forecasting.

Table 6

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

	Approach	Number of features	$h = 1$			$h = 2$			$h = 3$		
			rmse	mae	mape	rmse	mae	mape	rmse	mae	mape
rf	no text	3	0.0728	0.0545	0.1018	0.0742	0.0561	0.1042	0.0744	0.0564	0.1052
	textblob	36	0.0681	0.0514	0.0971	0.0691	0.0526	0.0989	0.0676	0.0504	0.0947
	our method	27	0.0683	0.0520	0.0968	0.0714	0.0542	0.1009	0.0720	0.0531	0.0985
svr	svr-Li	28	0.0646	0.0475	0.0885	0.0602	0.0444	0.0827	0.0603	0.0441	0.0821
	no text	2	0.1111	0.1001	0.1765	0.1116	0.0999	0.1761	0.1116	0.1002	0.1766
	textblob	14	0.0586	0.0427	0.0799	0.0597	0.0442	0.0827	0.0593	0.0428	0.0802
	our method	17	0.0584	0.0428	0.0801	0.0596	0.0439	0.0819	0.0593	0.0428	0.0801
arima & arimax	no text	-	0.0565	0.0394	0.0751	0.0566	0.0394	0.0753	0.0566	0.0394	0.0754
	textblob	23	0.0568	0.0401	0.0757	0.0568	0.0404	0.0759	0.0573	0.0407	0.0765
	our method	30	0.0577	0.0412	0.0779	0.0579	0.0417	0.0791	0.0586	0.0422	0.0799
ada	no text	3	0.0594	0.0428	0.0799	0.0599	0.0436	0.0813	0.0602	0.0436	0.0812
	textblob	7	0.0565	0.0397	0.0751	0.0565	0.0397	0.0752	0.0566	0.0397	0.0751
	our method	12	0.0564	0.0396	0.0750	0.0565	0.0396	0.0751	0.0565	0.0396	0.0751

2.3 FORECASTING WITH TEXTUAL FEATURES

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

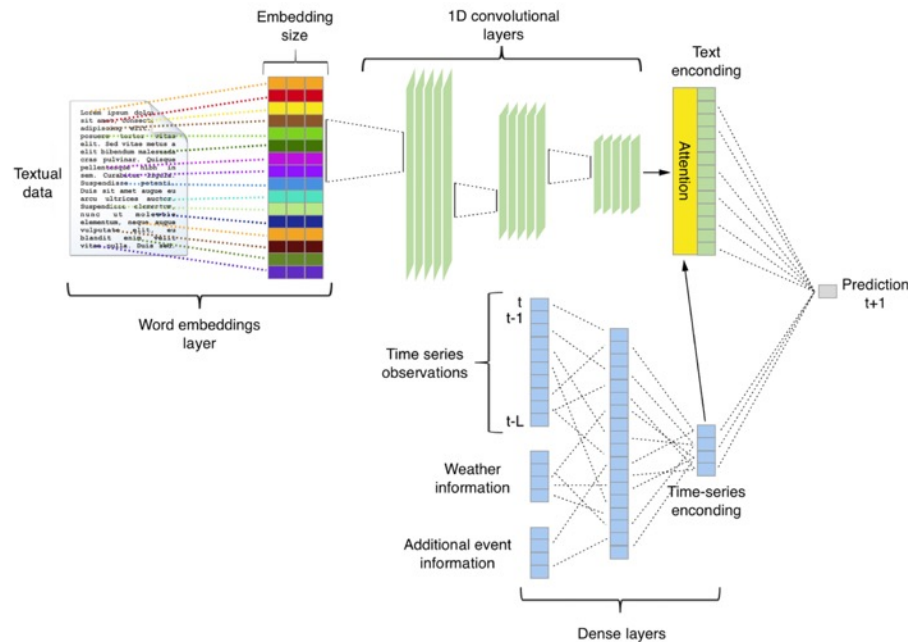


Fig. 4. Proposed neural network architecture with FC layers for modelling the time-series observations (DL-FC).

Table 2

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

Method	MAE	RMSE	MAPE	R ² (× 100)
SVR L	120.7 (± 0.0)	167.5 (± 0.0)	15.4 (± 0.0)	42.8 (± 0.0)
SVR L+W	120.1 (± 0.0)	166.4 (± 0.0)	15.4 (± 0.0)	43.6 (± 0.0)
SVR L+W+E	97.9 (± 0.0)	139.1 (± 0.0)	12.6 (± 0.0)	60.5 (± 0.0)
GP L	145.2 (± 0.0)	199.4 (± 0.0)	16.3 (± 0.0)	18.9 (± 0.0)
GP L+W	127.3 (± 0.0)	176.9 (± 0.0)	15.2 (± 0.0)	36.2 (± 0.0)
GP L+W+E	100.0 (± 0.0)	141.1 (± 0.0)	12.8 (± 0.0)	59.4 (± 0.0)
DL-LSTM L	129.0 (± 2.6)	172.6 (± 4.3)	15.9 (± 0.2)	39.2 (± 3.1)
DL-LSTM L+W	128.9 (± 1.9)	167.8 (± 2.0)	16.5 (± 0.2)	42.6 (± 1.4)
DL-LSTM L+W+E	103.9 (± 1.2)	141.5 (± 1.6)	13.3 (± 0.3)	59.2 (± 0.9)
DL-LSTM L+W+E+T	100.4 (± 1.8)	139.5 (± 2.0)	12.8 (± 0.3)	60.3 (± 1.2)
DL-FC L	120.8 (± 1.1)	161.7 (± 0.6)	15.8 (± 0.3)	46.7 (± 0.4)
DL-FC L+W	119.8 (± 1.0)	160.5 (± 0.6)	15.6 (± 0.2)	47.5 (± 0.4)
DL-FC L+W+E	95.6 (± 0.9)	134.8 (± 0.8)	12.7 (± 0.2)	63.0 (± 0.4)
DL-FC L+W+E+T	93.2 (± 0.8)	132.3 (± 0.9)	12.3 (± 0.1)	64.3 (± 0.5)

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

CONTENTS

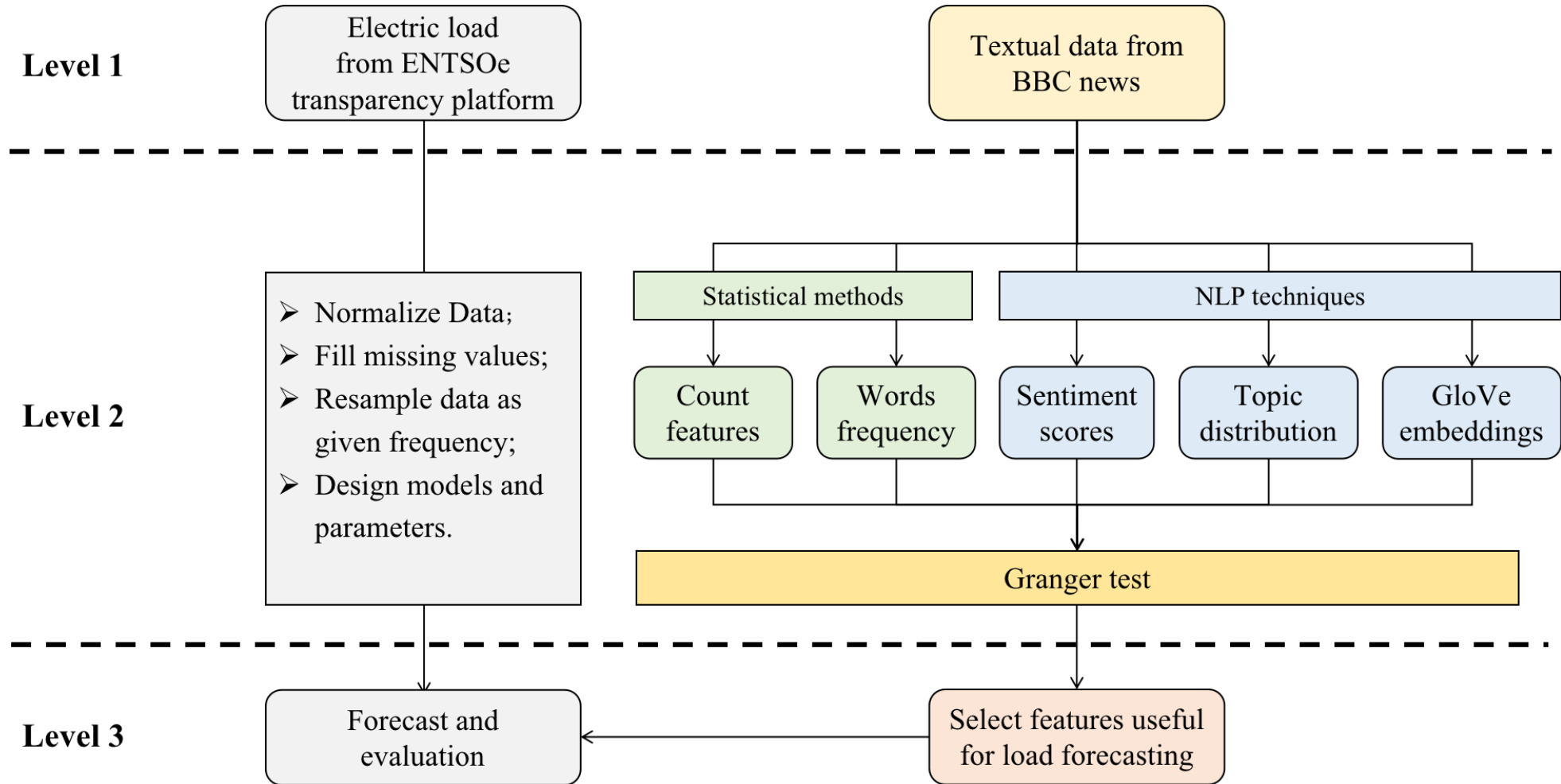
1. GENERAL CONTEXT

2. STATE OF THE ART

3. EARLY RESULTS AND ANALYSIS

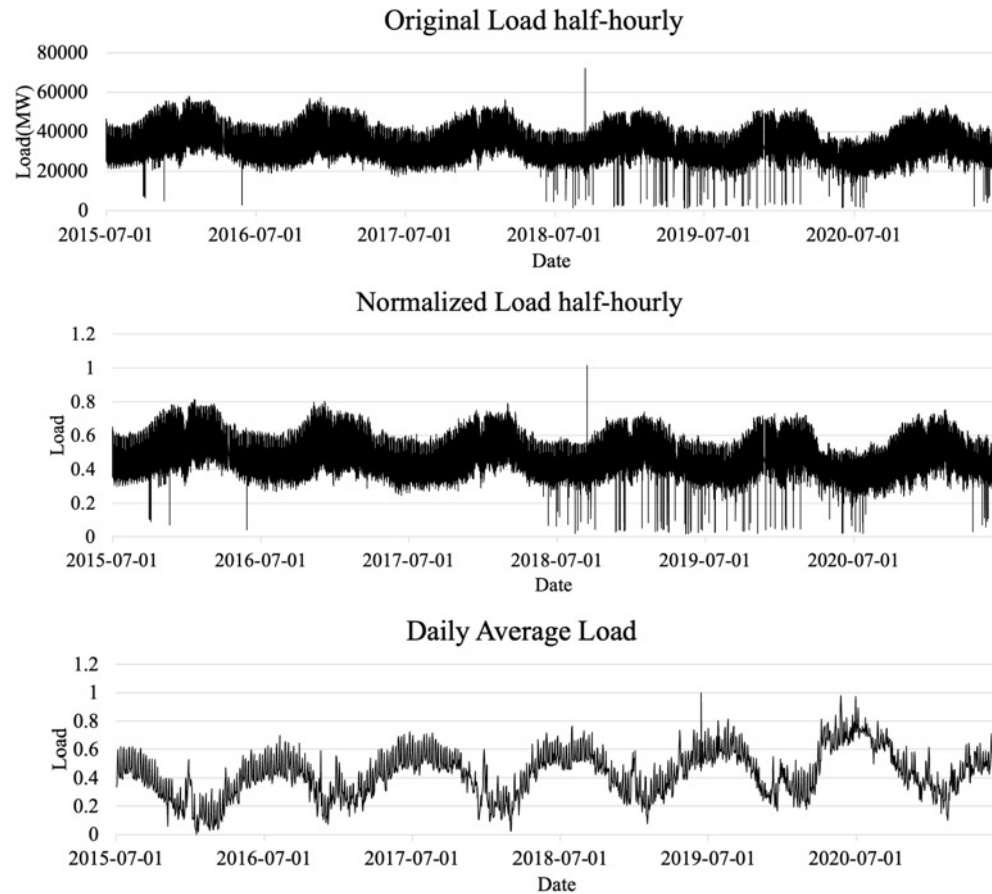
4. DISCUSSION

3.1 MODELING

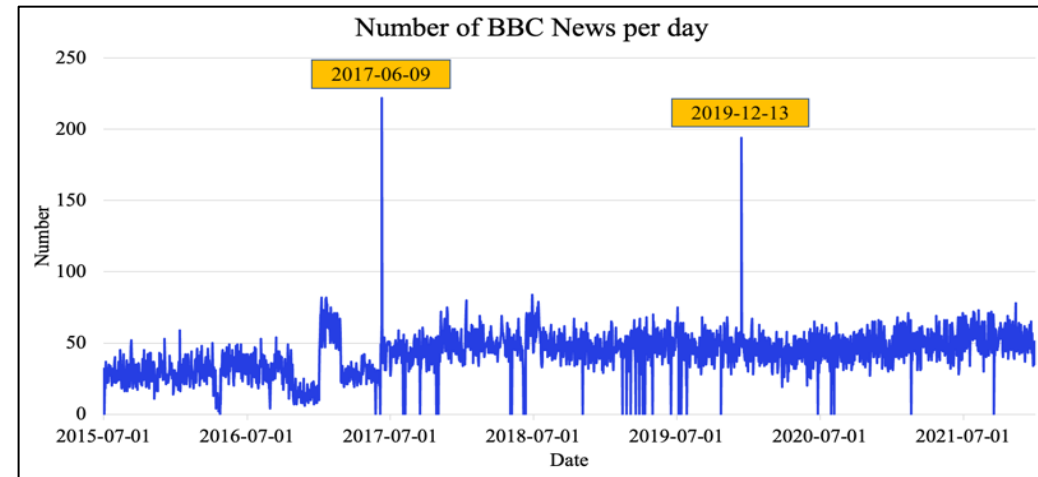
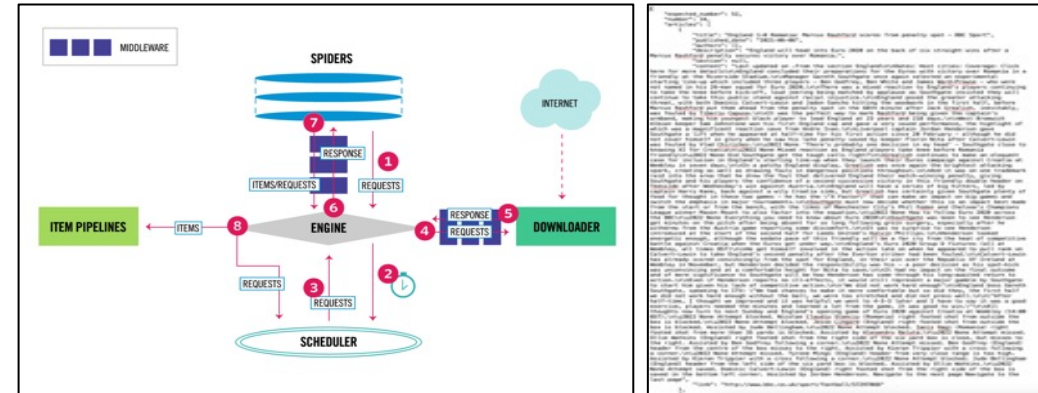


3.2 ACQUISITION AND DATA ANALYSIS

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.

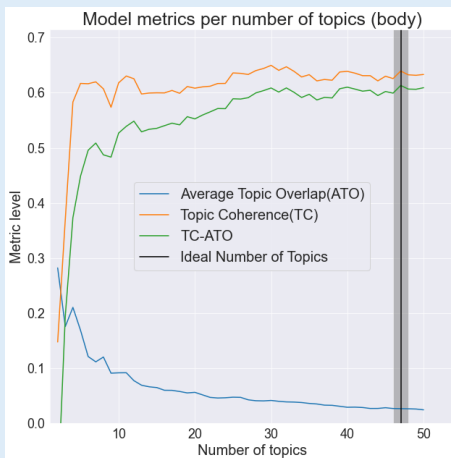


Table 1: Topics mined by LDA model on des data.

Topic ID	Keywords	Summarization
1	say, England, new, Johnson, university, boring, help, student, work, test	Univeristy and students
2	year, first, Olympics, win, Britain, day,time, gold, great, new	Olympics
3	league, Manchest, city, unit, premier, football, manage, club, Liverpool, season	Football
4	US, president, state, say, Trump, force, group, UK, Syria, Islam	America
5	Brexit, ministry, say, EU, leader, UK, deal, parties, labour, may	Election
6	policy, say, found, court, attack, two, kill, die, charge, man	Social events
7	say, government, UK, firm, secretary, new, social, could, use, plan	Business
8	people, say, home, million, work, staff, test, help, uk, pay	Covid influence
9	year, old, man, two, arrest, police, say, murder, boy, open	Social events
10	world, final, England, Cup, BBC, win, first, news, beat, one	World Cup
11	people, say, die, fire, kill, one, police, injure, near, London	Social events
12	star, say, die, age, front, show, page, film, TV, US	Film and actor
13	say, Covid, England, health, coronavirus, school, first, case, year, 19	Covid-19

- 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

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

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

Granger causality comparisons			
Model	Regression	X Coefficients	Wald test
Restricted	$y_t = c_2 + \sum_{i=1}^3 \alpha_{2,i} y_{t-i} + \epsilon_{y,t}$	$\beta_{2,1} = \beta_{2,2} = \beta_{2,3} = 0$	Null hypothesis
Unrestricted	$y_t = c_2 + \sum_{i=1}^3 \alpha_{2,i} y_{t-i} + \sum_{i=1}^3 \beta_{2,i} x_{t-i} + \epsilon_{y,t}$	At least one of $\beta_{2,1}, \beta_{2,2}, \beta_{2,3} \neq 0$	Alternative hypothesis

titles

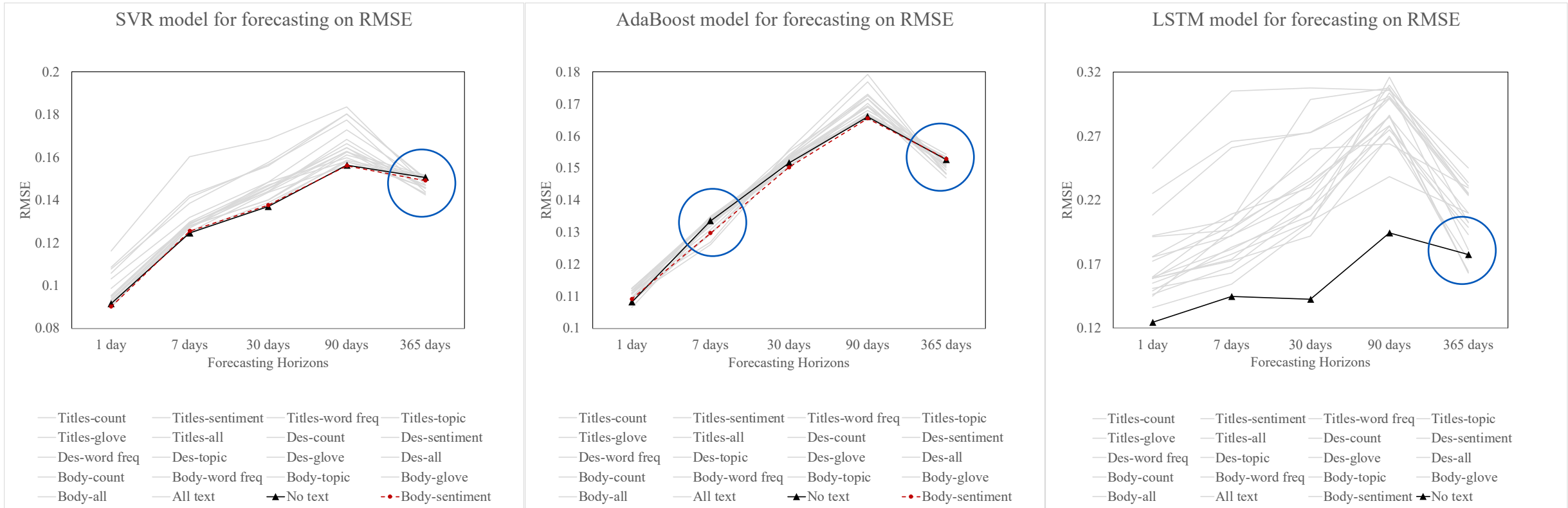
des

body

Textual features type	Num of features	Num of selected features
Count features	27	6
Word frequency	227	30
Sentiment scores	18	6
Sentiment STL scores	4	0
Topic distributions	50	7
GloVe embeddings	100	17
Count features	27	6
Word frequency	178	30
Sentiment scores	18	2
Sentiment STL scores	4	0
Topic distributions	43	4
GloVe embeddings	100	23
Count features	27	5
Word frequency	330	65
Sentiment scores	18	4
Sentiment STL scores	4	0
Topic distributions	47	15
GloVe embeddings	100	9

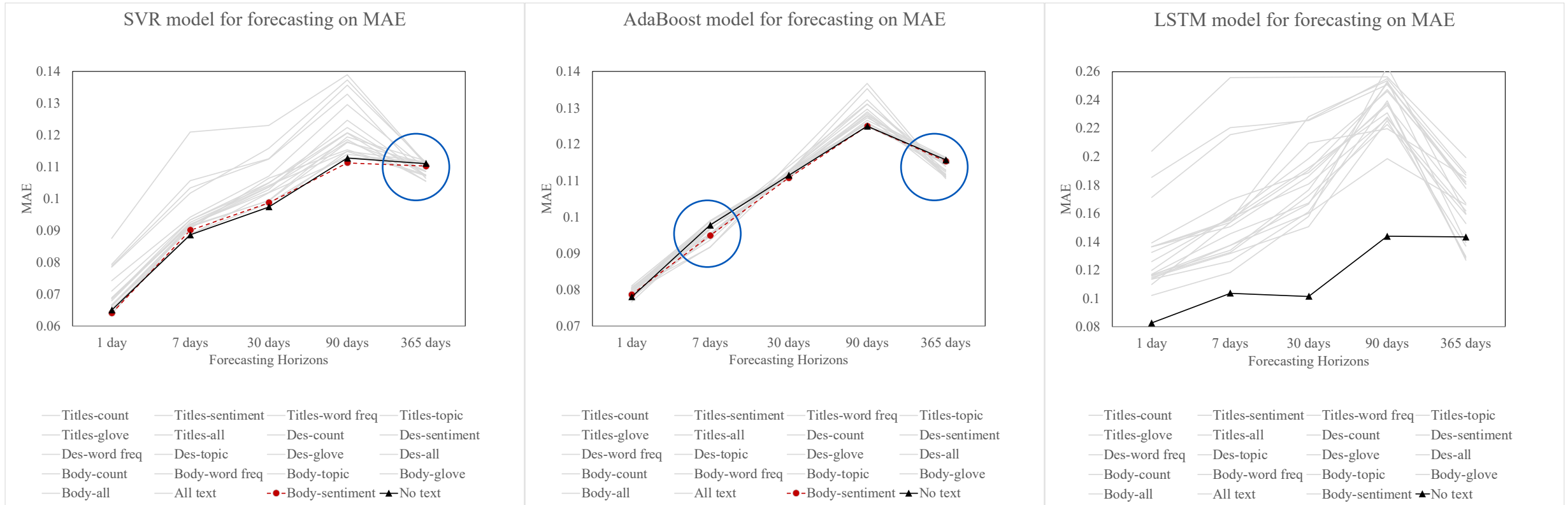
3.5 DETERMINISTIC FORECASTING

SVR, AdaBoost, LSTM models on RMSE index



3.5 DETERMINISTIC FORECASTING

SVR, AdaBoost, LSTM models on MAE index

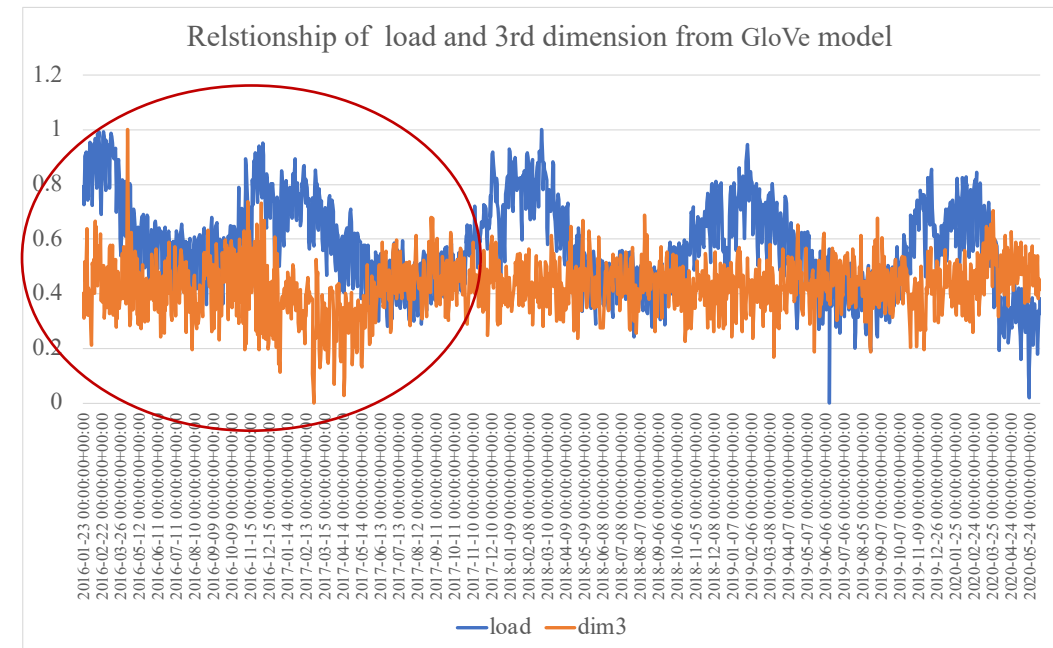


3.5 DETERMINISTIC FORECASTING

Relationship analysis

Models	RMSE		MAE	
	Textual features type	Improvement percentage	Textual features type	Improvement percentage
SVR	Des-GloVe embeddings	5.44%	Des-GloVe embeddings	4.99%
AdaBoost	Des-GloVe embeddings	3.74%	Des-GloVe embeddings	4.09%
LSTM	Titles-Word frequency	8.12%	Body-count features	11.47%

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

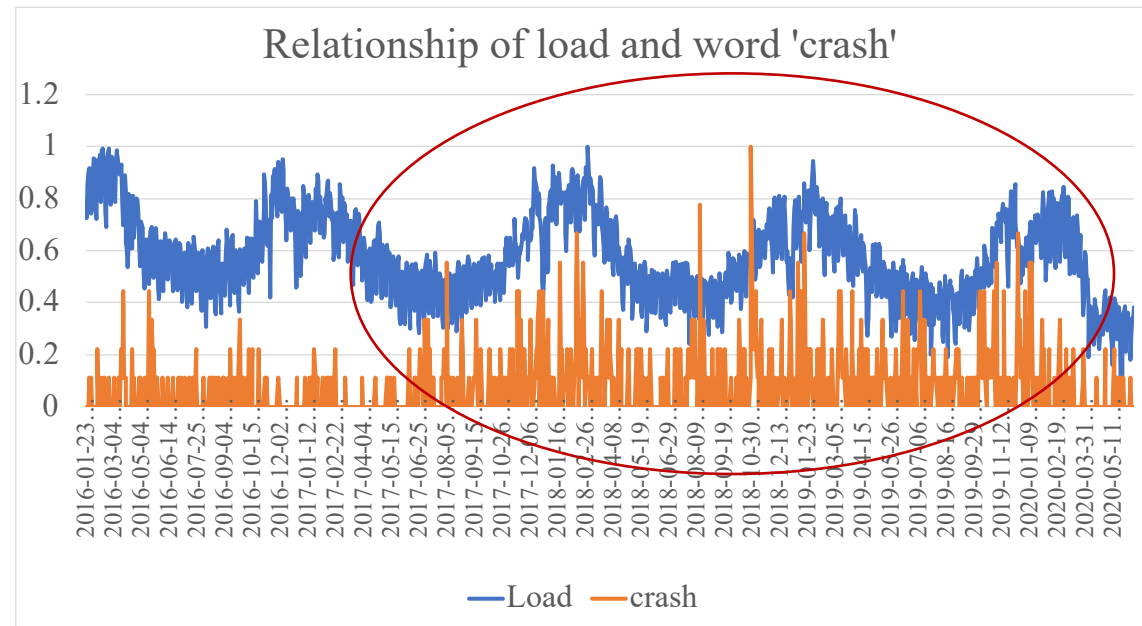


3.5 DETERMINISTIC FORECASTING

Relationship analysis

Models	RMSE		MAE	
	Textual features type	Improvement percentage	Textual features type	Improvement percentage
SVR	Des-GloVe embeddings	5.44%	Des-GloVe embeddings	4.99%
AdaBoost	Des-GloVe embeddings	3.74%	Des-GloVe embeddings	4.09%
LSTM	Titles-Word frequency	8.12%	Body-count features	11.47%

- 30 features from 227 in Titles-Word frequency
- 'warn', 'life', 'while', 'new', 'england', 'cup', 'attack', 'people', 'face', 'headline', 'job', 'mps', 'nhs', '2', 'show', 'british', 'liverpool', 'child', 'abuse', 'pay', 'jail', 'happen', 'crash', 'trial', 'cities', 'use', 'risk', 'rule', 'universe', 'vaccine'

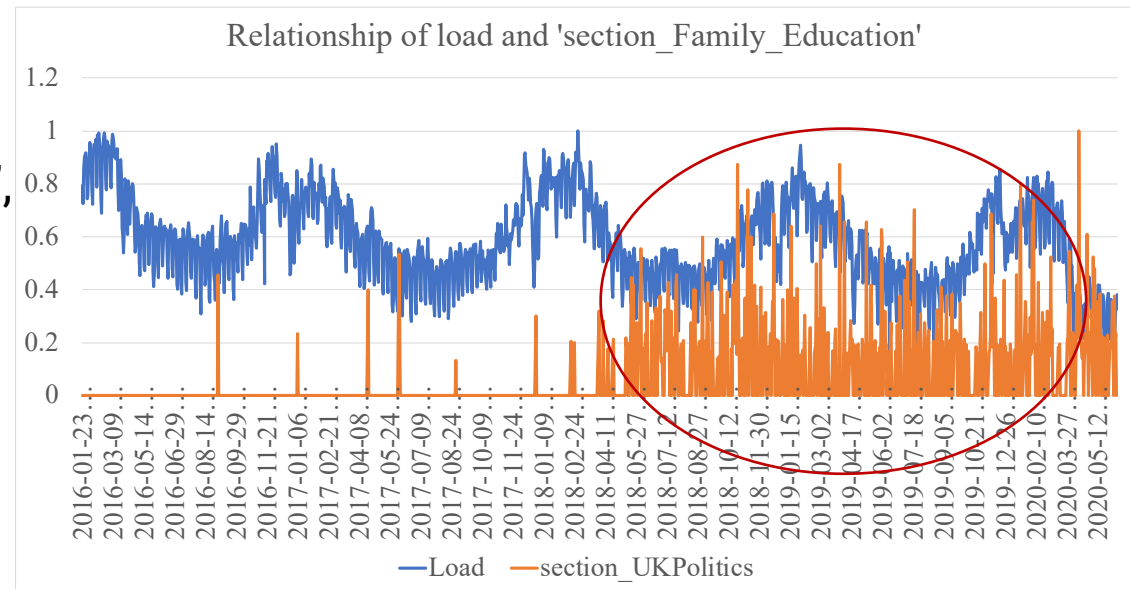


3.5 DETERMINISTIC FORECASTING

Relationship analysis

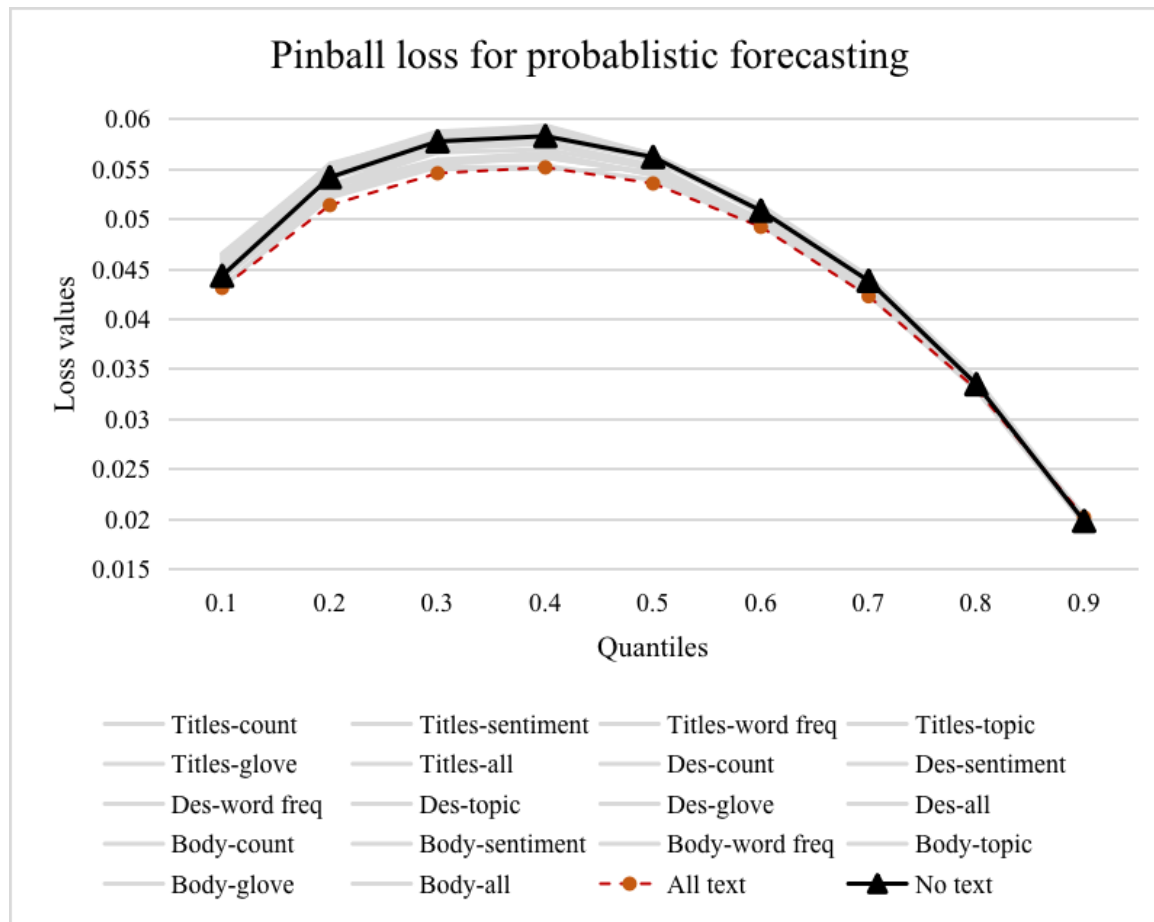
Models	RMSE		MAE	
	Textual features type	Improvement percentage	Textual features type	Improvement percentage
SVR	Des-GloVe embeddings	5.44%	Des-GloVe embeddings	4.99%
AdaBoost	Des-GloVe embeddings	3.74%	Des-GloVe embeddings	4.09%
LSTM	Titles-Word frequency	8.12%	Body-count features	11.47%

- 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



	Textual features type	Improvement percentage
CRPS	Des-GloVe embeddings	4.80%
Sharpness	Titles-sentiment	9.36%

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

- [1] Obst David, et al. "Textual data for time series forecasting." *arXiv preprint arXiv:1910.12618* (2019).
- [2] Zhenkun Zhou, Jichang Zhao, and Ke Xu. "Can online emotions predict the stock market in China?." *International conference on web information systems engineering*. Springer, Cham, 2016.
- [3] Yun Bai, Xixi Li, Hao Yu, Suling Jia. "Crude oil price forecasting incorporating news text." *International Journal of Forecasting* 38.1 (2022): 367-383.
- [4] Rodrigues Filipe, Ioulia Markou, and Francisco C. Pereira. "Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach." *Information Fusion* 49 (2019): 120-129.
- [5] Entsoe transparency platform: <https://www.entsoe.eu>
- [6] BBC news dataset: <https://github.com/LuChang-CS/news-crawler>



PSL  PERSEE

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