Dawen XIA, Jian GENG, Ruixi HUANG, Bingqi SHEN, Yang HU, Yantao LI, Huaqing LI, 2023. A distributed EEMDN-SABiGRU model on Spark for passenger hotspot prediction. *Frontiers of Information Technology & Electronic Engineering*, 24(9):1316-1331.
https://doi.org/10.1631/FITEE.2200621

A distributed EEMDN-SABiGRU model on Spark for passenger hotspot prediction

Key words: Passenger hotspot prediction; Ensemble empirical mode decomposition (EEMD); Spatial attention mechanism; Bi-directional gated recurrent unit (BiGRU); GPS trajectory; Spark

Corresponding author: Dawen XIA, Huaqing LI
E-mail: dwxia@gzmu.edu.cn, huaqingli@swu.edu.cn
ORCID: Dawen XIA, https://orcid.org/0000-0002-0151-9643; Huaqing LI, https://orcid.org/0000-0001-6310-8965

Motivation

1. Traditional time-series methods do not consider the impact of non-stationary series on prediction accuracy. Furthermore, although the EMD method can reduce the non-stationarity of time series, it still has the problems of end effects and modal confounding.
2. When using neural networks for passenger hotspot prediction, researchers employed the self-attention mechanism to focus on the correlation of the data without considering the spatial correlation between the map road network and passenger hotspots.
3. Although GRU can solve the problems of gradient disappearance and gradient explosion in RNN, the information-dependent GRU method ignores the information context in the road network, with high complexity and a long prediction time. In addition, few researchers used the EEMD method combined with neural network models for passenger hotspot prediction.

Method

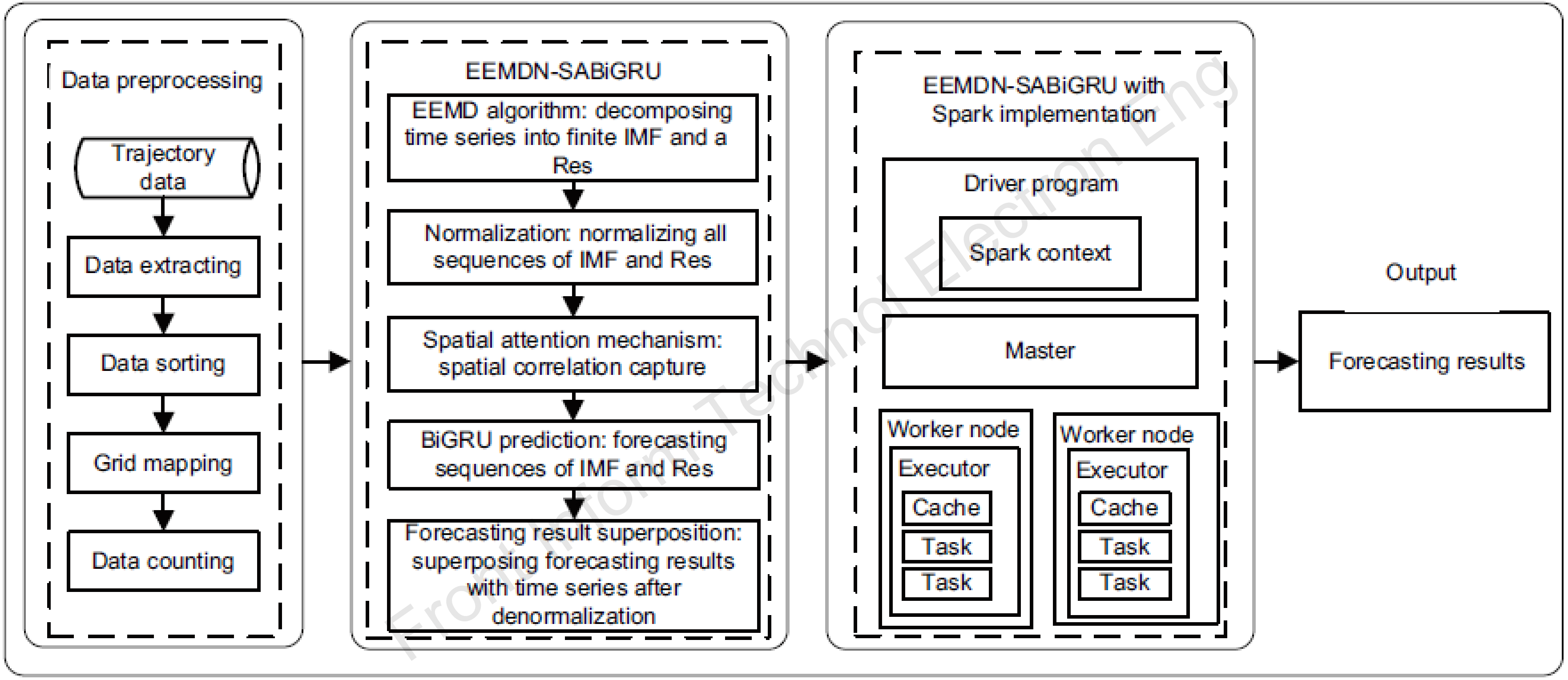

Fig. 1 A distributed EEMDN-SABiGRU model on Spark

The prediction framework based on a distributed EEMDN-SABiGRU model includes data preprocessing, model construction, and model implementation as shown in Fig. 1.

Method

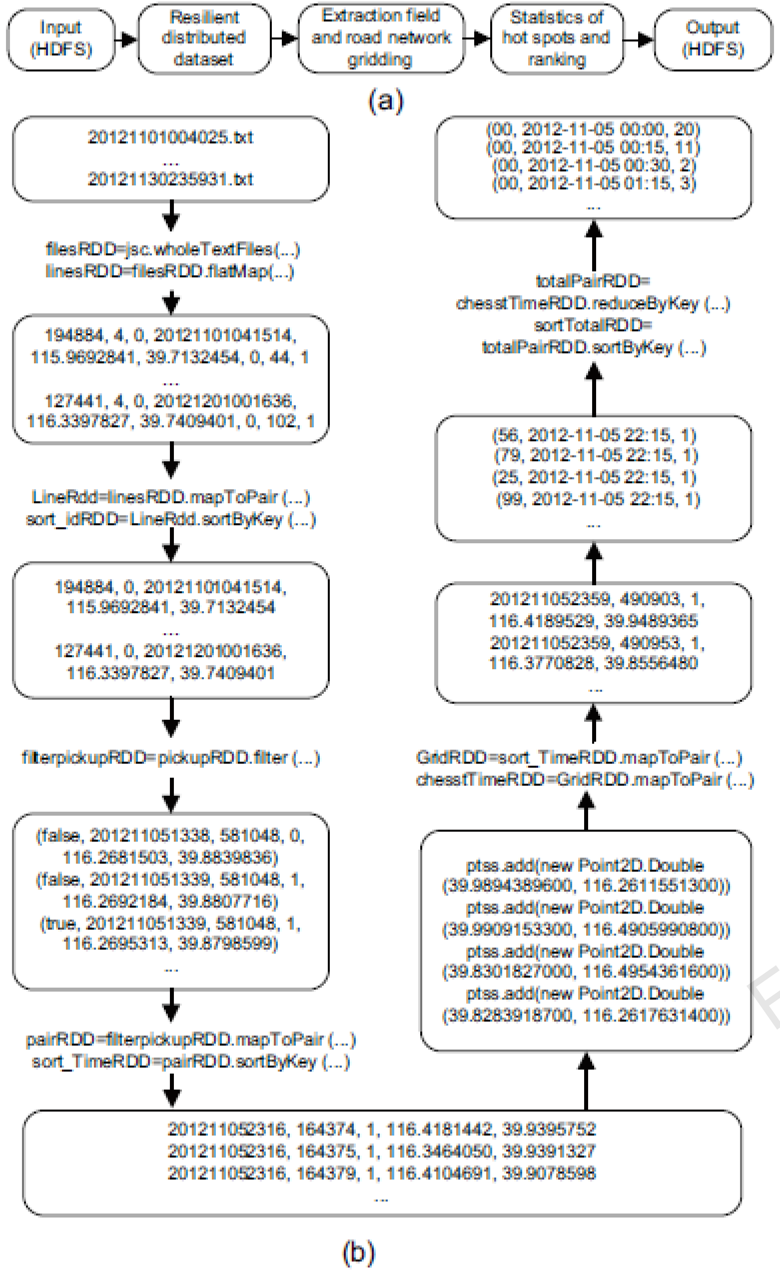

Fig. 2 Process of data preprocessing: (a) data preprocessing; (b) data flow

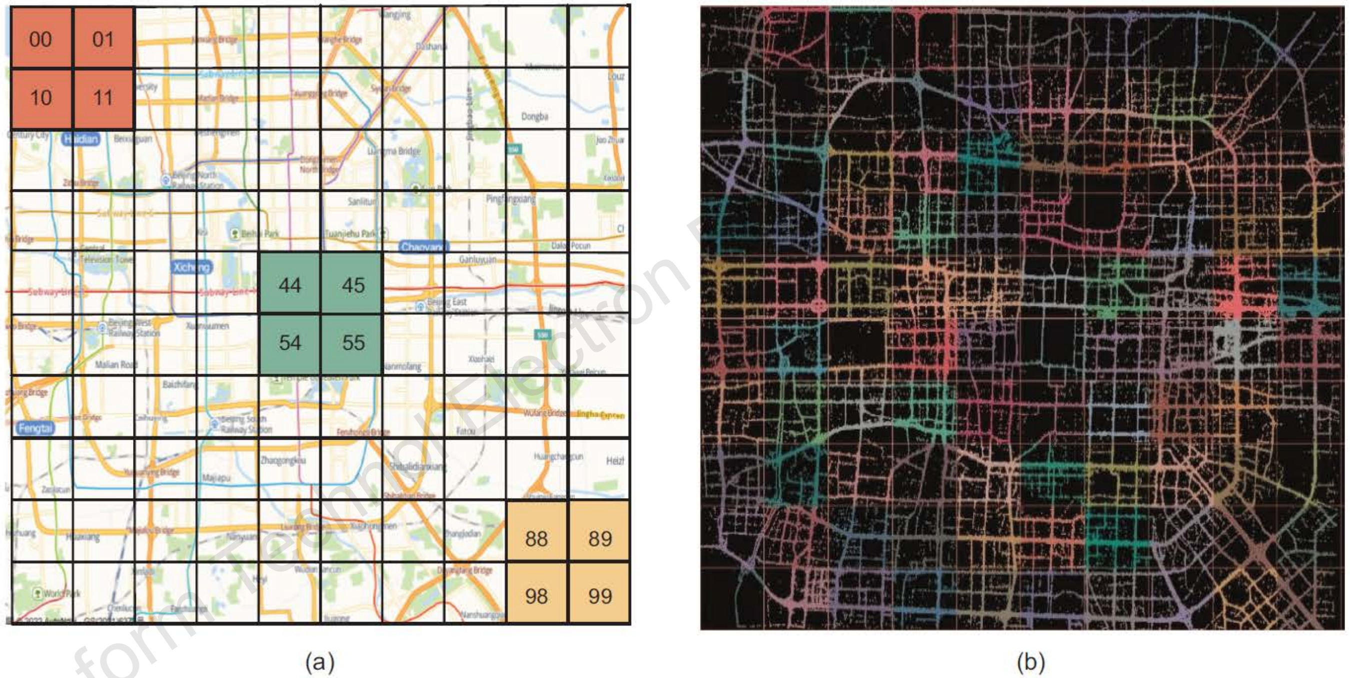

Fig. 3 Road network grid: (a) 10×10 grid; (b) road network with 10×10 grid

The process of data preprocessing is illustrated in Fig. 2.The sorted data are mapped or matched into these latitude and longitude ranges using the sorted data, and the latitude and longitude of the sorted data gridded as 10×10 are illustrated in Fig. 3.

Method

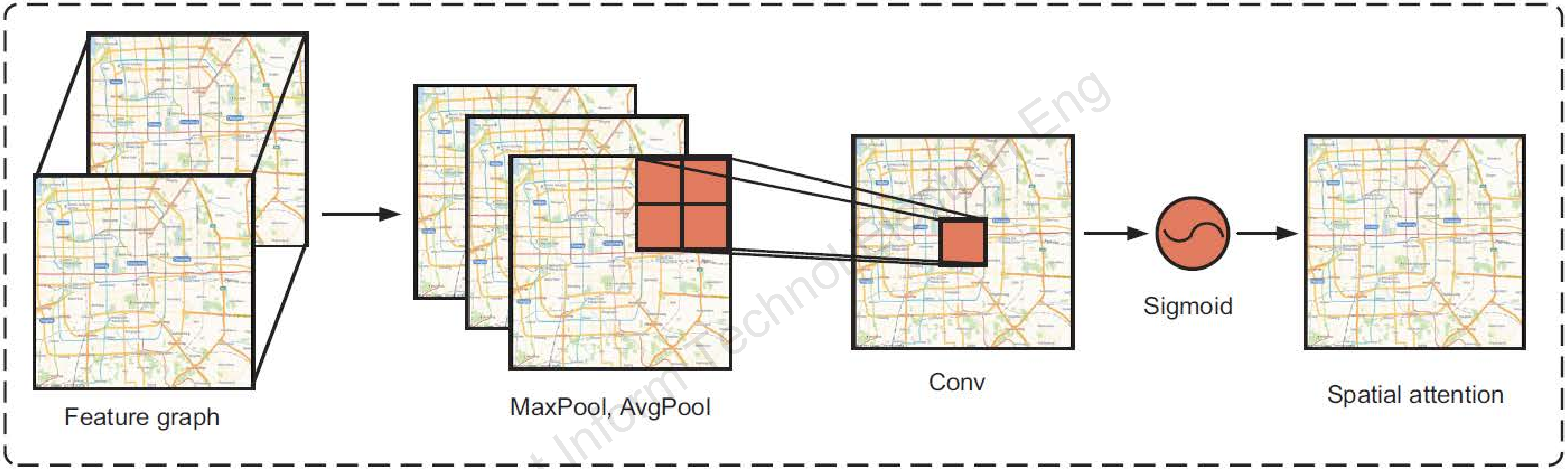

Fig. 5 Spatial attention mechanism

$$M(\boldsymbol{F}) = [\mathrm{AvgPool}(\boldsymbol{F}), \mathrm{MaxPool}(\boldsymbol{F})] \qquad M_s\boldsymbol{F} = \sigma(f(M(\boldsymbol{F})))$$

The process of building the spatial mechanism module is plotted in Fig. 5.

Method

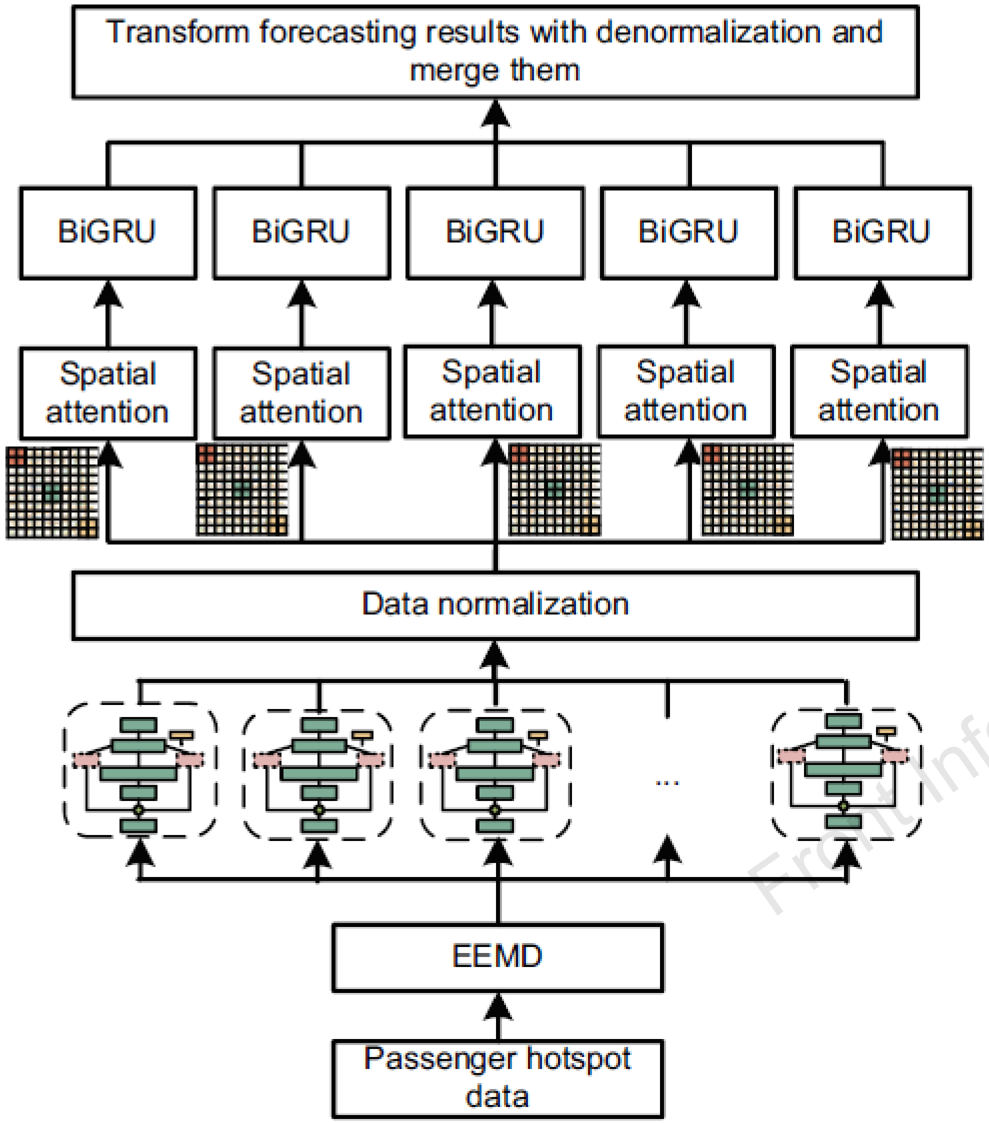

Fig. 6 Implementation process of the EEMDN-SABiGRU model

- Step 1: data decomposition. The EEMD algorithm decomposes the 15-min interval passenger hotspot data in the grid to obtain a finite number of IMFs and a Res sequence.
- Step 2: data normalization. Normalize the IMF and Res, and map them to a range of [0, 1].
- Step 3: model prediction. The normalized IMF and Res are inputted into the BiGRU model with spatial attention for prediction.
- Step 4: result superposition. The prediction values are denormalized, and the values are summed to obtain the final prediction results.

Major results

Table 3 Comparisons of models in different datasets using the 00-grid

Dataset	MOE	LSTM	EMD-LSTM	EEMD-LSTM	GRU	EMD-GRU	EEMD-GRU	EMDN-GRU	CNN	BP	EEMDN-SABiGRU
1-day	MAPE (%)	28.900	12.000	5.700	25.400	10.900	8.100	3.600	31.300	9.500	1.500
	MAE	18.099	2.197	2.453	15.607	3.970	2.657	1.336	14.972	4.571	0.736
	RMSE	23.397	2.508	4.017	20.512	4.756	3.257	1.637	16.028	4.901	0.736
	ME	48.642	6.477	7.943	44.560	9.842	6.029	3.488	25.524	7.834	2.116
5-day	MAPE (%)	12.300	11.500	12.000	10.800	29.300	26.900	6.800	20.800	9.500	3.100
	MAE	3.529	2.197	0.966	3.489	3.626	3.599	1.808	8.843	3.584	0.449
	RMSE	6.503	2.508	1.291	7.002	4.057	3.976	2.518	10.598	4.420	0.563
	ME	38.176	6.477	6.020	42.404	10.926	22.635	9.080	26.963	11.868	1.983
10-day	MAPE (%)	5.000	18.500	8.100	9.500	22.800	21.900	4.400	21.400	10.700	2.500
	MAE	2.378	2.695	1.395	4.395	3.957	2.973	1.990	10.154	3.495	0.705
	RMSE	5.397	4.398	2.477	7.340	6.211	4.213	2.984	12.440	4.437	0.978
	ME	50.297	28.234	18.028	61.684	37.755	21.053	14.305	38.353	15.135	3.392
15-day	MAPE (%)	6.300	18.800	17.500	25.600	31.500	22.300	20.000	16.600	15.700	2.800
	MAE	2.407	2.199	2.009	3.256	2.972	2.156	2.015	8.735	3.652	0.265
	RMSE	4.777	3.756	3.100	6.996	5.109	3.776	3.614	10.633	4.595	0.361
	ME	79.900	19.632	27.407	94.321	31.845	30.907	12.538	39.290	19.754	2.193
20-day	MAPE (%)	5.800	18.300	8.800	16.600	42.500	11.700	16.300	16.100	16.700	1.500
	MAE	1.275	2.729	0.986	3.421	4.528	1.419	1.828	7.950	3.025	0.208
	RMSE	2.137	3.821	1.440	4.891	5.840	2.115	2.559	9.821	3.864	0.294
	ME	21.752	19.704	15.155	33.022	26.693	18.093	12.392	28.836	13.480	1.779
25-day	MAPE (%)	16.100	16.000	12.100	17.100	29.000	56.700	40.800	15.500	17.300	2.000
	MAE	2.166	2.307	1.735	3.560	5.115	4.482	3.520	8.223	3.119	0.362
	RMSE	4.893	4.987	3.698	7.443	6.942	6.636	5.009	10.278	4.080	0.673
	ME	112.576	77.116	63.480	129.499	96.555	84.950	56.937	45.614	23.005	9.526
30-day	MAPE (%)	7.600	24.700	40.400	17.300	68.200	36.000	8.400	16.500	20.600	2.800
	MAE	2.693	4.071	3.850	2.552	9.013	4.464	2.006	9.385	3.290	0.396
	RMSE	8.194	6.094	6.002	10.144	9.989	8.294	4.209	12.105	4.407	0.670
	ME	118.708	64.821	62.146	133.717	82.860	80.284	29.887	51.783	23.545	9.091

Major results

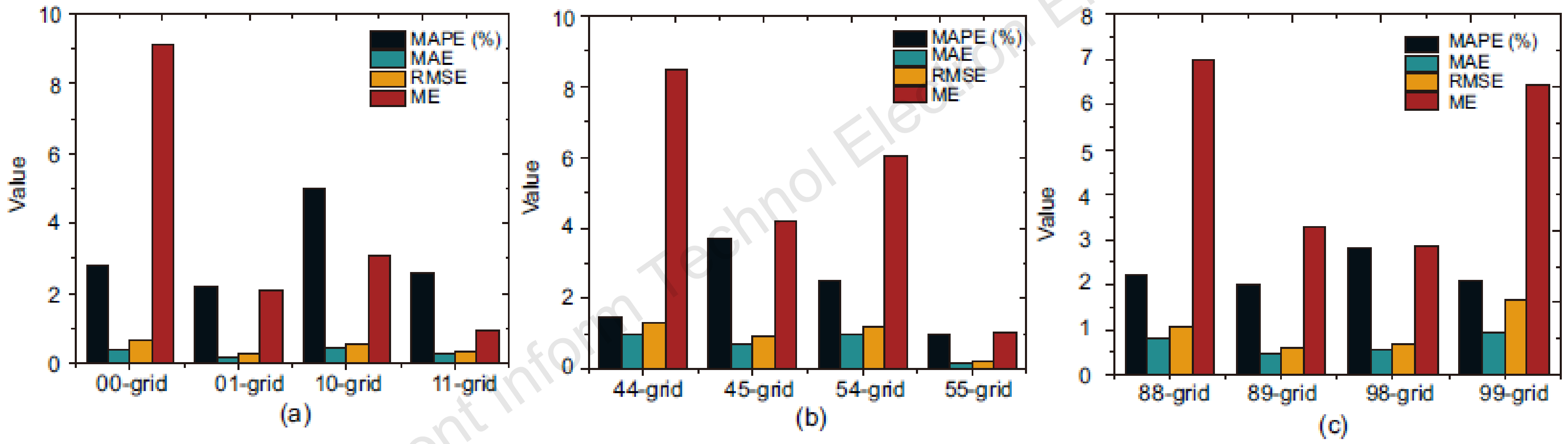

Fig. 9 Comparisons of MOE values for EEMDN-SABiGRU under different grids with the 30-day dataset: (a) 00-grid; (b) 55-grid; (c) 99-grid

Major results

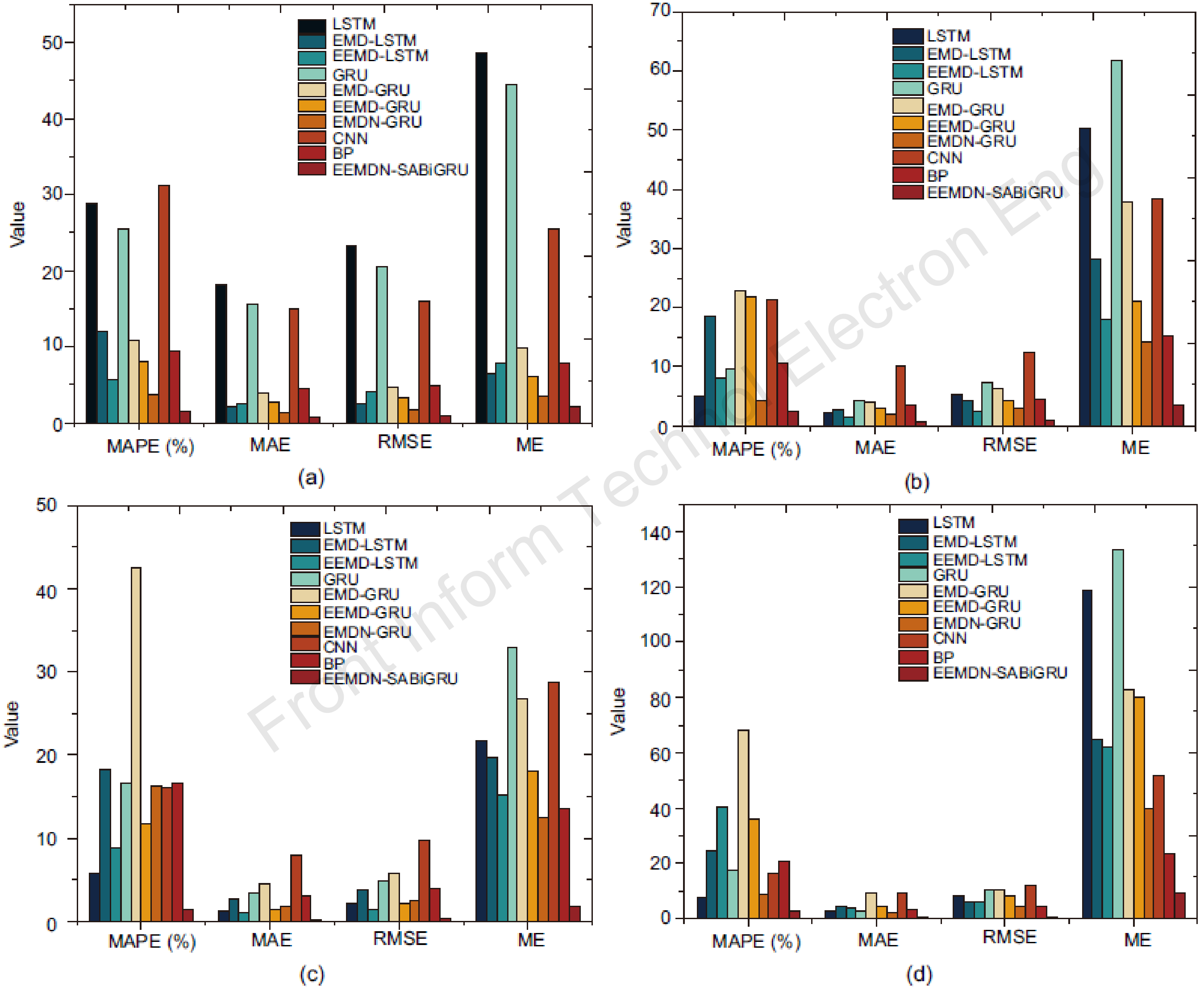

Fig. 10 Comparisons of MOE values for different models with different datasets in the 00-grid: (a) 1-day; (b) 10-day; (c) 20-day; (d) 30-day

Conclusions

1. An EEMDN method is proposed to reduce the influence of non-stationary time series on the prediction performance and to solve the IMF confusion problem of the EMD.
2. A spatial attention mechanism is constructed to capture spatial correlation, extract the number of passengers getting on and off in the grid, form the grid’s spatial weights, and improve the performance of passenger hotspot prediction.
3. A BiGRU model is incorporated to deal with the problem that GRU can obtain only forward contextual information but ignores backward contextual information, which improves the accuracy of feature extraction.

Dawen XIA is currently a professor at the College of Data Science and Information Engineering & Key Laboratory of Pattern Recognition and Intelligent Systems of Guizhou Province, Guizhou Minzu University, Guiyang, China. He received his PhD degree from the College of Computer and Information Science & College of Software, Southwest University, Chongqing, China, in June 2016. From November 2019 to November 2020, he was a visiting scholar supported by China Scholarship Council with the Rutgers, the State University of New Jersey, USA. His research interests include big data analytics, artificial intelligence, and data mining.

Jian GENG is currently working toward his MS degree at the College of Data Science and Information Engineering, Guizhou Minzu University, Guiyang, China. His research interests include big data analytics, data mining, and machine learning.

Huaqing LI is currently a professor at the College of Electronic and Information Engineering & Key Laboratory of Nonlinear Circuits and Intelligent Information Processing of Chongqing, Southwest University, Chongqing, China. He received his BS degree in information and computing science from Chongqing University of Posts and Telecommunications, Chongqing, China, in 2009, and his PhD degree in computer science and technology from Chongqing University, Chongqing, China, in 2013. He was a postdoctoral researcher with the School of Electrical and Information Engineering, University of Sydney, Sydney, Australia, from September 2014 to September 2015, and the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from November 2015 to November 2016. His main research interests include artificial intelligence, nonlinear dynamics and control, multi-agent systems, and distributed optimization. He serves as a regional editor for *Neur Comput Applic* and an editorial board member for *IEEE Access*.