MA-LSTM: A Multi-attention Based LSTM for Complex Pattern Extraction

Jingjie Guo,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences.

Abstract:

Recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behavior, which makes it applicable to tasks such as handwriting recognition or speech recognition. However, the RNN relies heavily on the automatic learning ability to update parameters which concentrate on the data flow but seldom considers the feature extraction capability of the gate mechanism. In this paper, we propose a novel architecture to build the forget gate which is generated by multiple bases. Instead of using the traditional single-layer fully-connected network, we use a Multiple Attention (MA) based network to generate the forget gate which refines the optimization space of gate function and improve the granularity of the recurrent neural network to approximate the map in the ground truth. Credit to the MA structure on the gate mechanism. Our model has a better feature extraction capability than other known models. MA-LSTM is an alternative module which can directly replace the recurrent neural network and has achieved good performance in manyareas that people are concerned about.

Experiments:

The focus of our study is the wide applicability and better predictions of MA-LSTM. Our aim is to compare the model with the standard RNN network, but not to achieve state of-the-art results. Therefore, our experiments are designed to make sure that the comparison is fair and our model is applied in different scenarios and domains which can provide sufficient support for our multiple bases theory. We conduct experiments with our model in three representative areas: Spatiotemporal Data, Computer Vision and Natural Language Processing, which have been developing in data science for a long time and is still growing rapidly. On the spatiotemporal data, we experiment with our model on Traffic Prediction . At the same time, we experiment on Handwritten Recognition in CV and Word-Level Language Model in NLP.

TABLE I
STAOC AND BASELINE MODELS PERFORMANCE

Data	Method	Start		End	
		rmse	mape	rmse	mape
Taxi	ARIMA	36.53	22.21%	27.25	20.91%
	LR	28.51	19.94%	24.38	20.07%
	MLP	26.67	18.43%	22.08	18.31%
	XGBoost	26.07	19.35%	21.72	18.70%
	LinUOTD	28.48	19.91%	24.39	20.03%
	ConvLSTM	28.13	20.50%	23.67	20.70%
	DeepSD	26.35	18.12%	21.95	18.15%
	ST-ResNet	26.23	21.13%	21.63	21.09%
	DMVST-Net	25.74	17.38%	20.51	17.14%
	STDN	24.10	16.30%	19.05	16.25%
	STAOC(ours)	**23.56**	**15.63%**	**18.43**	**15.56%**
Bike	ARIMA	11.53	26.35%	11.25	25.79%
	LR	10.92	25.29%	10.33	24.58%
	MLP	9.83	23.12%	9.12	22.40%
	XGBoost	9.57	23.52%	8.94	22.54%
	LinUOTD	11.04	25.22%	10.44	24.44%
	ConvLSTM	10.40	25.10%	9.22	23.20%
	DeepSD	9.69	23.62%	9.08	22.36%
	ST-ResNet	9.80	25.06%	8.85	22.98%
	DMVST-Net	9.14	22.20%	8.50	21.56%
	STDN	8.85	21.84%	8.15	20.87%
	STAOC(ours)	**8.54**	**21.25%**	**8.12**	**20.70%**

Model:

Figure 1 is the architecture of MA-LSTM. Figure 2 is the architecture of Forget Gate net. We propose a new RNN structure: MALSTM. We combine the attention mechanism and RNN. To reduce calculation time, we use the couple link to reduce the number of the parameters which shows a better performance compared with the common LSTM. In order to generate a forget gate, we use a fully connected layer and conversion function to generate multiple bases. Then we use the attention mechanism to fuse them and generate forget gate, as shown in Figure 1 and Figure 2.

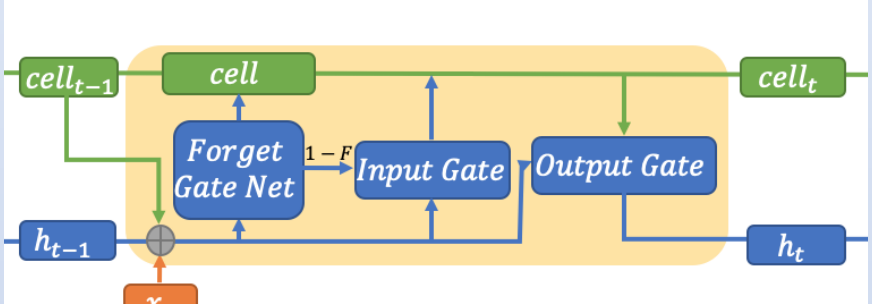

Figure 1

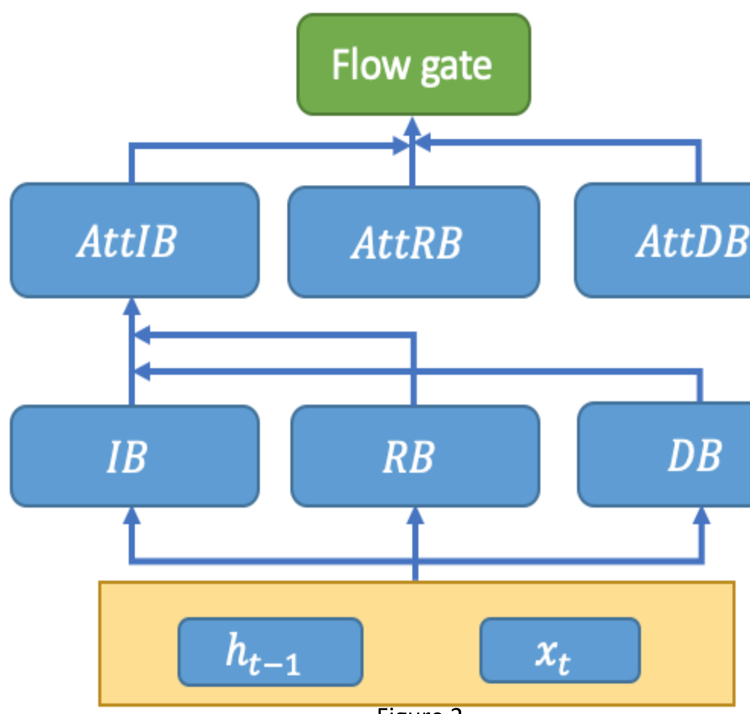

Figure 2

TABLE II
MA-LSTM AND BASELINE MODELS PERFORMANCE

Dataset	Method	Tess Loss	Test Accuracy
Mnist	LSTM	0.07298	98.56%
	GIFG	0.07043	98.28%
	GRU	0.07739	98.45%
	MA-LSTM(ours)	**0.06722**	**98.93%**

TABLE III
MA-LSTM AND BASELINE MODELS PERFORMANCE

Dataset	Method	Tess Loss	Test Perplexity
PTB	LSTM	4.50	90.47
	GRU	4.52	91.02
	On-LSTM	4.48	88.52
	MA-LSTM(ours)	**4.47**	**87.07**

Conclusion:

An novel model is proposed to extract the diverse pattern information which hid deeply in the data. Compared with the traditional RNN, our experiments on the three main areas with sequence data processing show a better performance and enhanced the feature extraction ability of the traditional LSTM.