中国科学院深圳先进技术研究院
SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY
CHINESE ACADEMY OF SCIENCES

Multi-Graph Convolutional Network for Relationship-Driven Stock Movement Prediction

ICPR2020

Jiexia Ye, Juanjuan Zhao, Kejiang Ye, Chengzhong Xu

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

State Key Lab of IOTSC, University of Macau

目录

01 Background and Related Work

02 Methodology

03 Experiments

04 Contributions

Part 01

Background and Related Works

Background

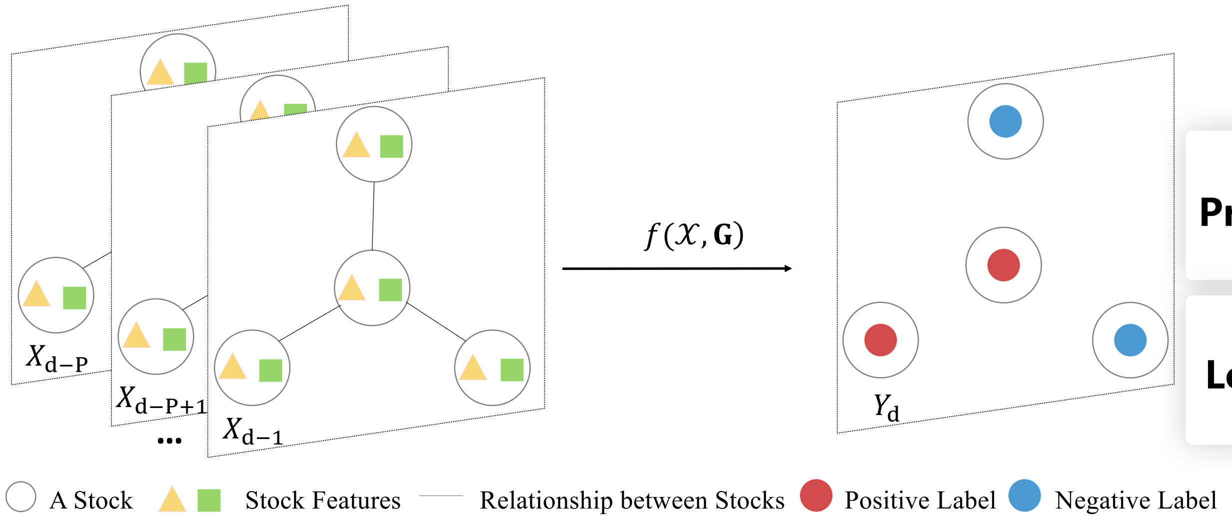
1. A little improvement of stock prediction accuracy might yield a huge gain.
2. Both traditional finance and modern behavior finance believe that fluctuations of stock prices are information-driven.
3. Researches make substantial effort on modeling correlations between various information and stock prices.

Related Works

1. The input features of most works are mainly based on historical market data (e.g. stock prices, trading volume).
2. Other types of information are complementary to enrich the input features, such as public news, texts from social medias.
3. Most of the works above mainly focused on combining a single stock's historical records with other textual information but overlooked the correlations among stocks.

Challenges

1. Design appropriate representation for corporation relationships
2. Design a model without independent instance assumption to extract the cross-correlation among stocks
3. Jointly considering historical observation and the cross-correlation with related stocks for stock prediction



Formulation

Prediction : $\widehat{Y_d} = f([X_{d-P}, \cdots, X_{d-1}], G; \Theta)$

Loss Function : $\mathcal{L} = -\frac{1}{N}\sum_{s=1}^{N}[y_d^s log(\widehat{y_d^s}) + (1 - y_d^s) log(1 - \widehat{y_d^s})]$

Part 02

Methodology

Model

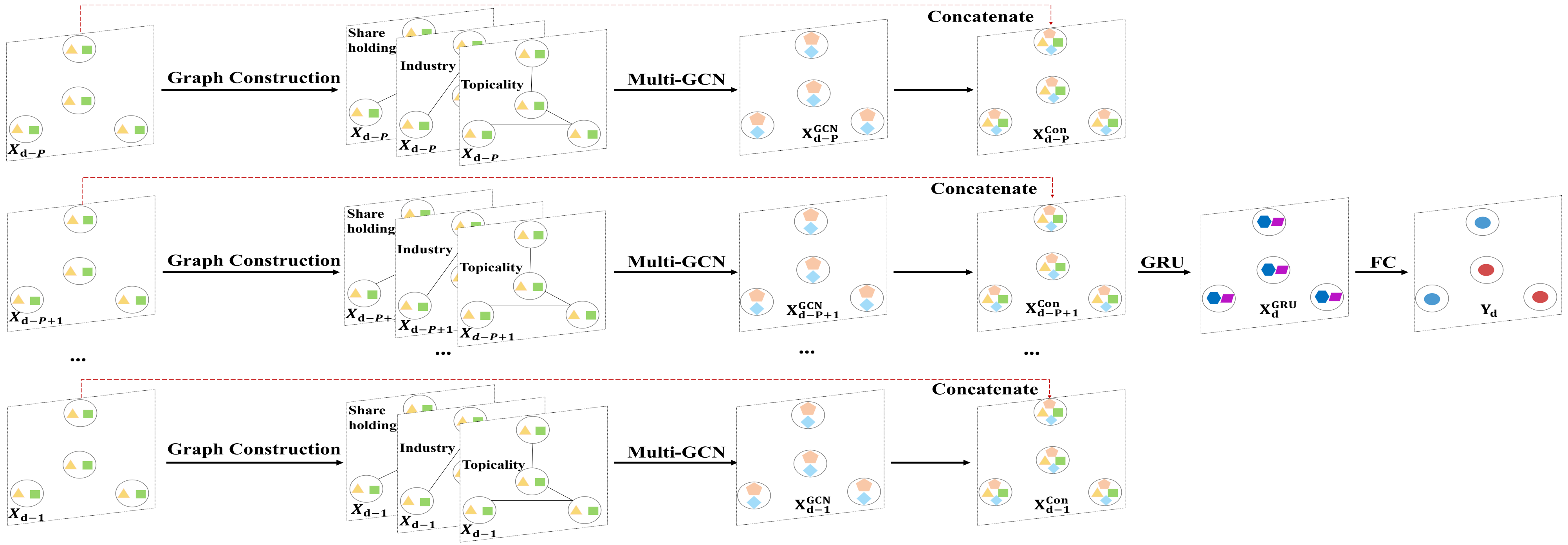
Stock Graphs

Multi-GCN

GRU

The architecture of Multi-GCGRU

中国科学院大学
University of Chinese Academy of Science



1. Construct stock graphs: Shareholding Graph/ Industry Graph/ Topicality Graph
2. Capture cross-effect: At each input step, utilize Multi-GCN to capture cross effect among stocks
3. Extract temporal dependency: The cross-effect features and original features are fed into GRU

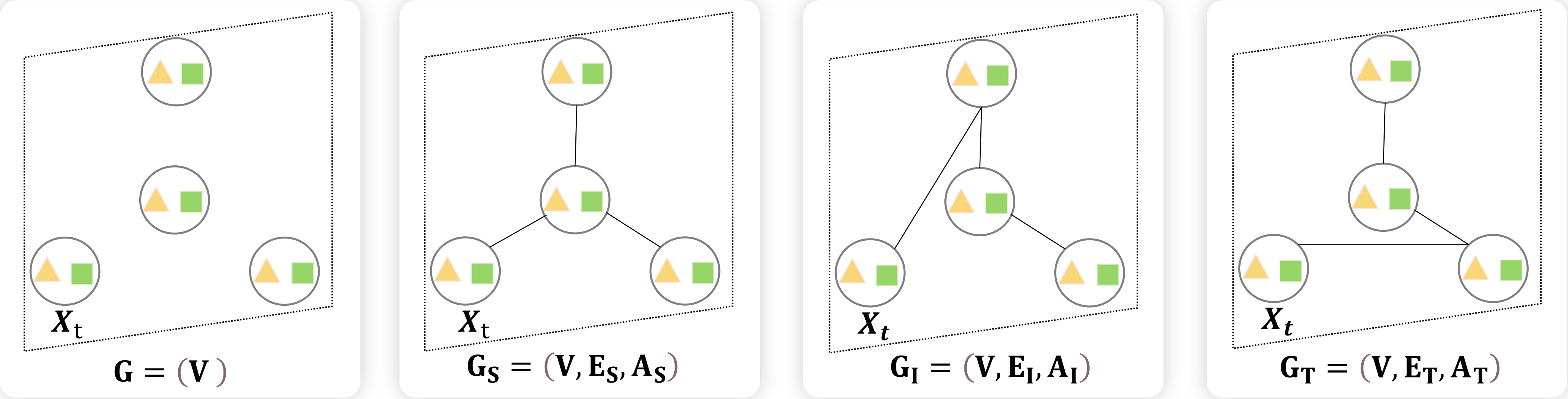
Stock Graphs

Model

Stock Graphs

Multi-GCN

GRU



Stock Collection | Shareholding Graph | Industry Graph | Topicality Graph

1. **Shareholding Graph: The edge weight a_{ij} is shareholding ratio in range of [0,1]**
2. **Industry Graph: The influence from company i to company j in the same industry is denoted as $a_{ij} = \frac{M_i}{M_j}$, where M denotes the firm size**
3. **Topicality Graph: if company i owns M_i topicalities, company j owns M_j topicalities and they share T_{ij} topicalities, the connection strength from i to j is denoted as $a_{ij} = \frac{T_{ij}}{M_i}$**

Multi-Graph Convolutional Network

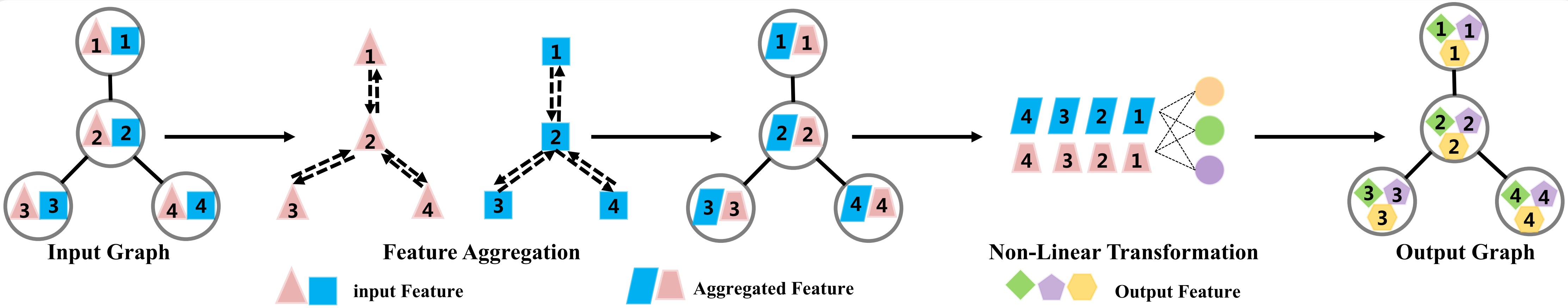
中国科学院大学
University of Chinese Academy of Science

Model

Stock Graphs

Multi-GCN

GRU



Graph Convolutional Layer :

$$H^{(l+1)} = \rho\left(\left(\sum_{k=0}^{K-1} \theta_k L^k\right) H^{(l)} W^{(l)}\right)$$

Multi-Graph Convolutional Layer :

$$H^{(l+1)} = \rho\left(\left(\sum_{k=0}^{K-1} \theta_k \left(\theta_S L_S^k + \theta_I L_I^k + \theta_T L_T^k\right)\right) H^{(l)} W^{(l)}\right)$$

Dynamic Graph Convolutional Layer :

$$H^{(l+1)} = \rho\left(\hat{L} H^{(l)} W^{(l)}\right)$$

Gated Reccurent Unit

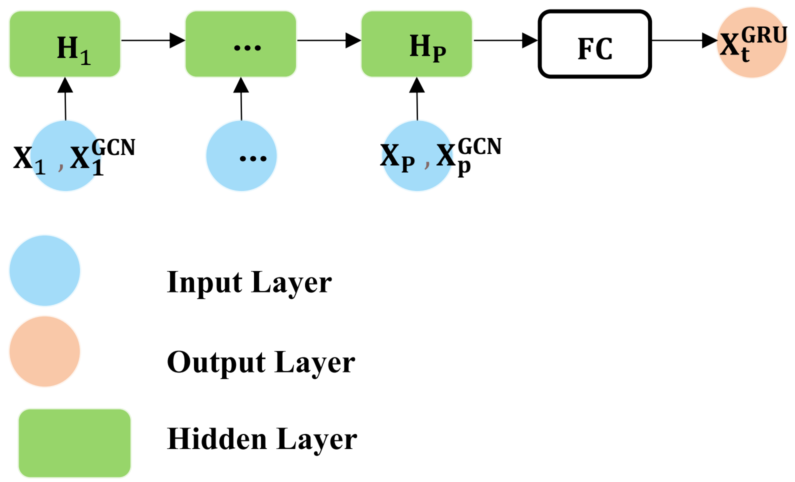


Figure 1 The architecture of our GRU

$$r_t = \boldsymbol{\sigma}([H_{t-1}, X_t, X_t^{\mathbf{GCN}}] \cdot W_r + b_r)$$

$$u_t = \boldsymbol{\sigma}([H_{t-1}, X_t, X_t^{\mathbf{GCN}}] \cdot W_u + b_u)$$

$$\hat{H}_t = \boldsymbol{tanh}([r_t \odot H_{t-1}, X_t, X_t^{\mathbf{GCN}}] \cdot W_h + b_h)$$

$$H_t = u_t \odot H_{t-1} + (1 - u_t) \odot \hat{H}_t$$

Part 03

Experiments

Datasets and Experiment Design

Datasets

Model Comparison

Model Ablation

Hyper parameter

- ## Datasets

CSI300, CSI500

Time Range: 2015/07/01-2019/12/30

Input Features: opening price, high price, low price, trading amount

Relationship Features: shareholder and shareholding ratio, industry category, registered capital, topicality.

Labels: closing price rising=1, closing price falling=0

The Split of Dataset

Training Set 70% **Test Set** 20% **Validation Set** 10%

Indexes	Training set	Validation set	Testing set	Total
CSI 500	383,719	54,817	109,633	548,169
CSI 300	225,209	32,173	64,345	321,727

- ## Experiment Design

Does taking the cross effect among stocks into consideration enhance the stock movement prediction?

Does our proposed model provide a better solution to incorporate the cross effect?

Which kind of corporation relationship is more effective for stock prediction and why?

How does the hyperparameter affect the model performance?

Model Comparison

University of Chinese Academy of Science

Datasets

Model Comparison

Model Ablation

Hyper parameter

Input Feature	Models	CSI300 Accuracy	CSI300 Precision	CSI300 Recall	CSI300 F1	CSI300 MCC	CSI500 Accuracy	CSI500 Precision	CSI500 Recall	CSI500 F1	CSI500 MCC
Historical Records	LR	0.5145	0.9746	0.5133	0.6724	0.0228	0.5149	0.9723	0.5148	0.6732	0.0117
	SVM	0.5197	0.9498	0.5165	0.6691	0.0412	0.5253	0.9662	0.5202	0.6763	0.0636
	RF	0.5375	0.9298	0.5271	0.6728	0.0957	0.5433	0.9900	0.5294	0.6899	0.1587
	ANN	0.5191	0.9724	0.5158	0.6740	0.0463	0.5202	0.9900	0.5170	0.6792	0.0576
	LSTM	**0.5435**	0.9756	0.5291	0.6861	0.1443	**0.5461**	0.9662	0.5318	0.6860	0.1384
	GCN-S	0.5472	0.9609	0.5317	0.6845	0.1421	0.5463	0.9675	0.5423	0.6950	0.0717
	Multi-GCGRU	**0.5754**	0.9603	0.5484	0.6981	0.2171	**0.5885**	0.9894	0.5658	0.7199	0.2377

1. Linear Regression performs the worst for it can only capture linear relationship.
2. Among machine learning, random forest perform the best, for it can model randomness in stock price.
3. LSTM performs better than ANN for it can capture the temporal dependency in stock prediction.
4. Our Multi-GCGRU performs the best for it not only captures non-linearity but also temporal dependency and cross-effect among stocks.

Conclusion Cross effect does matter!!!

Model Ablation

中国科学院大学 University of Chinese Academy of Science

Input Feature	Models	CSI300 Accuracy	CSI300 Precision	CSI300 Recall	CSI300 F1	CSI300 MCC	CSI500 Accuracy	CSI500 Precision	CSI500 Recall	CSI500 F1	CSI500 MCC
Historical Records & Corporation Relationships	GCGRU-S	0.5505	0.9321	0.5346	0.6795	0.1338	0.5521	0.9635	0.5458	0.6969	0.0938
	GCGRU-I	0.5598	0.9561	0.5392	0.6895	0.1739	0.5678	0.9814	0.5540	0.7082	0.1655
	GCGRU-T	0.5628	0.9512	0.5412	0.6899	0.1782	0.5751	0.9837	0.5581	0.7122	0.1916
	GCGRU-D	0.5602	0.9442	0.5402	0.6871	0.1667	0.5697	0.9844	0.5549	0.7097	0.1756
	Multi-GCGRU	**0.5754**	0.9603	0.5484	0.6981	0.2171	**0.5885**	0.9894	0.5658	0.7199	0.2377

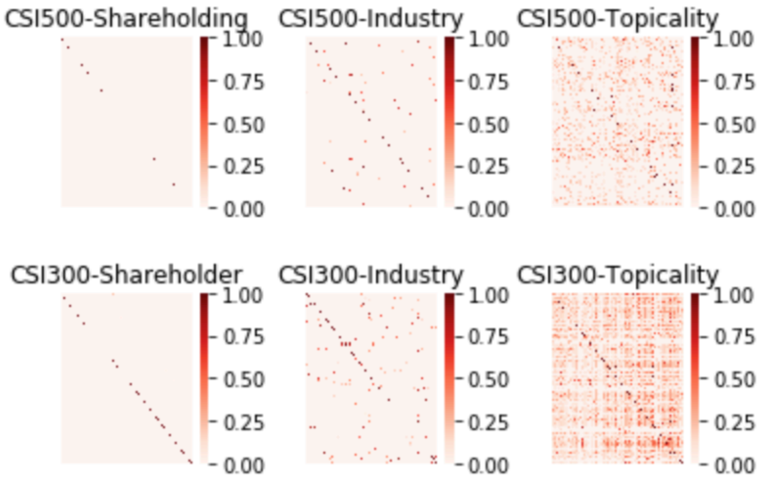


Figure1 Visualization of Shareholding/Industry/Topicality Matrices

1. Shareholding effect performs worst. Topicality effect performs best
2. The denser the matrix is, the more information it carries.
3. Three matrices perform better than any single matrix.
4. The performance of data-driven matrix is similar to a single Pre-defined matrix but with less data.

Conclusion

(1) More pre-defined matrices can improve prediction accuracy.
(2) When data is insufficient, data-driven matrix can be considered.

Hyper-parameter Analysis

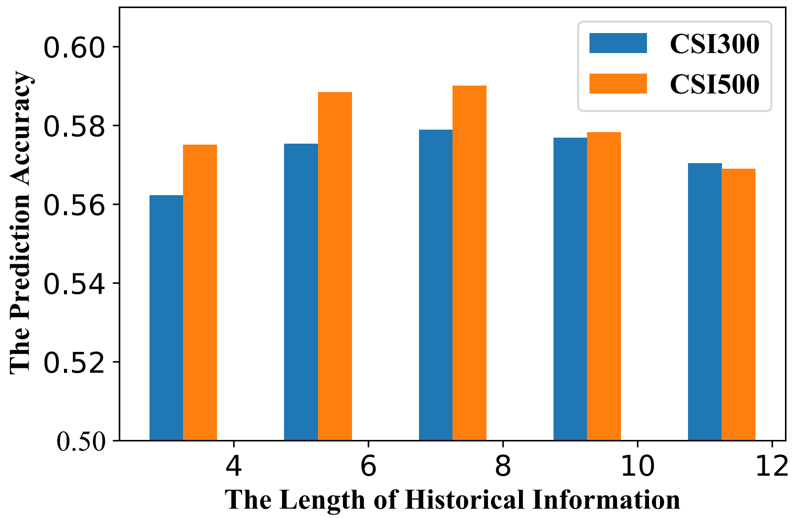
University of Chinese Academy of Science

Datasets

Model Comparison

Model Ablation

Hyper parameter



MULTI-GCGRU WITH DIFFERENT LAG SIZES

Length	CSI300 ACC	CSI300 MCC	CSI500 ACC	CSI500 MCC
3-days	0.5623	0.1513	0.5752	0.1742
5-days	0.5754	0.2171	0.5885	0.2377
7-days	**0.5790**	0.2196	**0.5901**	0.2821
9-days	0.5769	0.1869	0.5783	0.1965
11-days	0.5705	0.1378	0.5691	0.1221

1. The input length of 3 days performs worst. Perhaps information is insufficient.
2. The input length of 7 days performs best.
3. More than 7 days, the accuracy becomes worse.

Conclusion The best input length should be not too short or too long and depends on the datasets.

博学笃志 格物明德

Part 04

Contributions

Contributions

中国科学院大学
University of Chinese Academy of Science

Compared with previous works, our contributions are as follows:

1. We take cross-effect among stocks into consideration, instead of historical observations of only single stock.
2. We novelly design industry/topicality graph to represent cross-effect among stocks.
3. We also explore a data-driven matrix to get rid of expert knowledge.
4. We utilize graph convolution network to capture cross-effect and GRU to capture temporal dependency in stock price.
5. Our Multi-GCGRU is flexible to consider more valuable pre-defined relationships.

Thanks for Listening