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: \( \hat{Y}_d = f([X_{d-p}, \ldots, X_{d-1}], G; \Theta) \)

Loss Function: \( L = -\frac{1}{N} \sum_{s=1}^{N} [y_d^s \log(\hat{y}_d^s) + (1 - y_d^s) \log(1 - \hat{y}_d^s)] \)
Part 02
Methodology
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**

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 $a_{ij} = \frac{T_{ij}}{M_i}$. 

**Graph Definitions**:
- $G = (V)$
- $G_S = (V, E_S, A_S)$
- $G_I = (V, E_I, A_I)$
- $G_T = (V, E_T, A_T)$

$X_t$, $G$, $S$, $A$ represent different components of the graphs.
Multi-Graph Convolutional Network

Graph Convolutional Layer:
\[ H^{(l+1)} = \rho \left( \sum_{k=0}^{K-1} \theta_k L^k \right) H^{(l)} W^{(l)} \]

Multi-Graph Convolutional Layer:
\[ H^{(l+1)} = \rho \left( \sum_{k=0}^{K-1} \theta_k \left( \theta_S L_S^k + \theta_I L_I^k + \theta_T L_T^k \right) \right) H^{(l)} W^{(l)} \]

Dynamic Graph Convolutional Layer:
\[ H^{(l+1)} = \rho \left( LH^{(l)} W^{(l)} \right) \]
Gated Recurrent Unit

Figure 1 The architecture of our GRU

\[ r_t = \sigma([H_{t-1}, X_t, X_t^{GCN}] \cdot W_r + b_r) \]
\[ u_t = \sigma([H_{t-1}, X_t, X_t^{GCN}] \cdot W_u + b_u) \]
\[ \hat{H}_t = \tanh([r_t \odot H_{t-1}, X_t, X_t^{GCN}] \cdot W_h + b_h) \]
\[ H_t = u_t \odot H_{t-1} + (1 - u_t) \odot \hat{H}_t \]
Part 03
Experiments
Datasets and Experiment Design

**Datasets**

- **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**

<table>
<thead>
<tr>
<th>Indexes</th>
<th>Training Set</th>
<th>Validation Set</th>
<th>Testing Set</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSI 500</td>
<td>383,719</td>
<td>54,817</td>
<td>109,633</td>
<td>548,169</td>
</tr>
<tr>
<td>CSI 300</td>
<td>225,209</td>
<td>32,173</td>
<td>64,345</td>
<td>321,727</td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>Input Feature</th>
<th>Models</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LR</td>
<td>0.5145</td>
<td>0.9746</td>
<td>0.5133</td>
<td>0.6724</td>
<td>0.0228</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>0.5197</td>
<td>0.9498</td>
<td>0.5165</td>
<td>0.6691</td>
<td>0.0412</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>0.5375</td>
<td>0.9298</td>
<td>0.5271</td>
<td>0.6728</td>
<td>0.0957</td>
</tr>
<tr>
<td></td>
<td>ANN</td>
<td>0.5191</td>
<td>0.9724</td>
<td>0.5158</td>
<td>0.6740</td>
<td>0.0463</td>
</tr>
<tr>
<td></td>
<td>LSTM</td>
<td>0.5435</td>
<td>0.9756</td>
<td>0.5291</td>
<td>0.6861</td>
<td>0.1443</td>
</tr>
<tr>
<td>Historical Records</td>
<td>GCN-S</td>
<td>0.5472</td>
<td>0.9609</td>
<td>0.5317</td>
<td>0.6845</td>
<td>0.1421</td>
</tr>
<tr>
<td></td>
<td>Multi-GCGRU</td>
<td><strong>0.5754</strong></td>
<td><strong>0.9603</strong></td>
<td><strong>0.5484</strong></td>
<td><strong>0.6981</strong></td>
<td><strong>0.2171</strong></td>
</tr>
</tbody>
</table>

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

#### Table: Model Ablation Comparison

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

#### Figure 1: Visualization of Shareholding/Industry/Topicality Matrices

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.
The input length of 3 days performs worst. Perhaps information is insufficient.

The input length of 7 days performs best.

More than 7 days, the accuracy becomes worse.

The best input length should be not too short or too long and depends on the datasets.
Part 04
Contributions
Contributions

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