Next Cashtag Prediction on Social Trading Platforms with Auxiliary Tasks

Chung-Chi Chen, Hen-Hsen Huang, Hsin-Hsi Chen

http://cjchen.nlppfin.com
Task Design
The interests of investors

$AAPL$ support identified $198.8 \ldots$ next move to $215$

Bullish

$TVIX$ making a new 52 week low.

Bearish

$SPX$ 3145 is still possible, the level is the 1.618% extension within The Grand Super Cycle, Ideal Target.
Next Cashtag Prediction

We use the post(s) of $user_i$ at time $t$ to represent the preference of user $user_i$. With the help of this information, we predict the cashtags that $user_i$ will mention within next 5 days.
Auxiliary Task

• **Hottest cashtag prediction** \((A_{hot})\)
  This task is aimed at predicting which cashtag will have the highest growth rate of being mentioned within the following T days.

• **Most profitable cashtag prediction** \((A_{profit})\)
  This task is aimed at predicting which cashtag will have the highest return within the following T days.
Representation & Model
Representation of Interest, Analysis, Chart

Tweet(s) of user\_i at t  \rightarrow  Embedding of user\_i

All tweets about cashtag\_j at t  \rightarrow  Embedding of cashtag\_j  \rightarrow  Representation of the analysis of all users for cashtags  \rightarrow  Representation of the price information of all cashtags
Attentive Capsule Network (ACN) & The Joint Model
Experiment
## Dataset

<table>
<thead>
<tr>
<th>Description</th>
<th>Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of users in the experimental set</td>
<td>126,369</td>
</tr>
<tr>
<td>Avg. number of hashtags mentioned per tweet</td>
<td>4.85</td>
</tr>
<tr>
<td>The first date in the training set</td>
<td>May 27, 2018</td>
</tr>
<tr>
<td>The last date in the training set</td>
<td>Jan. 2, 2019</td>
</tr>
<tr>
<td>Number of instances in the training set</td>
<td>97,740</td>
</tr>
<tr>
<td>The first date in the test set</td>
<td>Jan. 3, 2019</td>
</tr>
<tr>
<td>The last date in the test set</td>
<td>Feb. 23, 2019</td>
</tr>
<tr>
<td>Number of instances in the test set</td>
<td>21,538</td>
</tr>
</tbody>
</table>
## Experimental Results

<table>
<thead>
<tr>
<th>Model</th>
<th>hit@2</th>
<th>hit@3</th>
<th>hit@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint ACN</td>
<td>69.03%</td>
<td>74.01%</td>
<td>80.33%</td>
</tr>
<tr>
<td>-Attention</td>
<td>67.30%</td>
<td>71.95%</td>
<td>78.04%</td>
</tr>
<tr>
<td>-CapsNet</td>
<td>66.90%</td>
<td>72.05%</td>
<td>78.87%</td>
</tr>
<tr>
<td>-$A_{hot}$</td>
<td>68.96%</td>
<td>73.78%</td>
<td>80.24%</td>
</tr>
<tr>
<td>-$A_{profit}$</td>
<td>68.74%</td>
<td>73.60%</td>
<td>79.89%</td>
</tr>
</tbody>
</table>
Conclusion & Future works

Experimental results show
• the effectiveness of the proposed ACN model
• the usefulness of the corresponding auxiliary tasks

The proposed task can be extended to
• user grouping
• dynamic interest prediction
• market information prediction such as price movement prediction and market volatility prediction
FinNum-2: Numeral Attachment in Financial Tweets

Numeral Attachment:

$\text{NE}$ OK NE, last time oil was over $\text{65}$ you were close to $\text{8}$. Giddy-up…

- Shared task in NTCIR 2020
- Identify the relation between the cashtag and the numeral
- The pilot dataset is available now.
- More than 15,000 instances in 10,000 unique tweets will be available.
- Macro-F1 score is adopted for evaluating the experimental results.

http://finnum.nlpfin.com
Thank you!