Predicting Financial Market Direction Using Social Media Data

Benedict Augustine¹, Baodan Zhang¹, Dr. Yuri G Balasanov²,*, Dr. Sema Barlas³,*

1. Candidate for MS in Analytics, Graham School, University of Chicago 2. Supervisor and Faculty of MSc in Analytics program 3. Director of MSc in Analytics program

---

**Research Objective**

- Examine whether relationship exists between Sentiment Indicators and Market Direction
- Validate whether these Sentiment Indicators are leading and adds dimension to current Technical Indicators
- Better predict market direction using this additional information

**Interesting Findings**

![Graphs showing positive and negative articles ratio vs Russell Index](image)

**Preprocessing**

Before Preprocessing

<table>
<thead>
<tr>
<th>Original Data</th>
<th>Preprocessed Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Text</td>
</tr>
<tr>
<td>Tokenization</td>
<td>Tokenization</td>
</tr>
<tr>
<td>Part of Speech</td>
<td>Part of Speech</td>
</tr>
<tr>
<td>Tagging</td>
<td>Tagging</td>
</tr>
<tr>
<td>Classification</td>
<td>Classification</td>
</tr>
</tbody>
</table>

**Classification Example**

**Example: Complicated Scenario**

- Standard publicly available dictionaries e.g. Bill McDonald's, Harvard Inquirer etc.
- Machine Learning to build dictionary
- Manually further customize

**Methodology**

- Information Source: Social Media Data
- Data Cleaning/Pre-processing
- Sentiment Classification Engine
- Daily Sentiment Measures

**Sentiment Dictionary**

Association rule discovery:

<table>
<thead>
<tr>
<th>Rules</th>
<th>Support</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>[AWESOME] =&gt; {1}</td>
<td>0.011289867</td>
<td>1.773717146</td>
</tr>
<tr>
<td>{lamest} =&gt; {0}</td>
<td>0.000141123</td>
<td>2.292461986</td>
</tr>
<tr>
<td>{hate,lamest} =&gt; {0}</td>
<td>0.000141123</td>
<td>2.292461986</td>
</tr>
</tbody>
</table>

**Research Result**

Leading relationship exists between Sentiment and Russell 2000 PC1 vs Russell 2000 Close Price

- Positive Articles Ratio vs Russell Index
- Negative Articles Ratio vs Russell Index

---

**Findings**

**Relationship Hypothesis**

- Sentiment Indicators did not exhibit strong (short term) relationship with S&P 500
- Strong, leading, short term relationship with the Russell 2000 Index

**Predict Market Direction Hypothesis**

- Able to predict the direction of Russell 2000 for next trading day with 80% accuracy, 30% better than just using Technical Indicators

**Information Source**

- Latest Articles related to specific tickers
- Market related to broader market
- Long ideas derived from Latest Articles
- Short Ideas derived from Latest Articles

**Sentiment Classification Engine**

- Different from Standard dictionary based classifiers
- Identifies Entity
- Matches Sentiment with the Entity
- Accuracy rate was 84% vs Standard Probabilistic approach was 51%

**Conclusion & Further Improvements**

- Additional data needed
- Weight the sentiment measures based on stock weights in the index.
- Adjust for Seasonality
- Remove lagging effect from leading information
- Use time warping techniques to give different weights to different periods
- Try additional machine learning techniques