Background for Case Study Used in Workshop

Fethi Rabhi

School of Computer Science and Engineering
University of New South Wales
Sydney Australia
Preliminaries

• Purpose of lecture
  – Look at domains involved in case study
  – Look at different types of datasets that will be part of workshop

• Before that, we need to:
  – Understand the concepts behind the data
  – Understand the processes and context in which data was generated
Domains involved

1. Finance
   Company
   Financial markets

2. News and Social Media

3. News and Financial markets
1. Finance

Basic concepts:

Company
   Entity that requires funding, usually listed company

Equity
   A share of that company, can be bought or sold

Company code
   A unique identifier that represents the company when processing data

Derivative
   Contracts on future price of company (options, futures)
Financial markets

• Characteristics
  • Allow the buying and selling of financial instruments
  • Financial instrument = equity or derivative
  • Most financial markets trading occurs through electronic exchanges

• Financial market trading cycle
  • Pre-trade analytics
  • Trading
  • Post-trade analytics
  • Settlement and Registry
2. News in Finance
Why are news important in finance?

- News is information that determines prices of financial instruments (Efficient Market Hypothesis)
- News reveal information about intentions and future events (regulatory announcements)
- News contain information that must be acted upon immediately (political news)
- News show people sentiment/mood/behaviour/reactions (social networks)
- Based on news
  - Fund managers adjust their portfolios
  - Traders exploit investment opportunities
  - Analysts publish trends
  - Etc.
Consideration for selecting the best source of news

• Choosing a news provider depends on:
  – Content scope and quality
  – Cost
  – Timeliness
  – Service quality

• Some news providers may offer a high quality service for one country but not for another. Users must be aware of the best source of information for their needs
Different categories of news that relate to finance

- Macroeconomic announcements
- Regulatory announcements
- Press announcements
- Analyst recommendations
- Journalistic commentary
- User generated content
User generated content

• Increasing in importance, users writing and reacting to news events
• Social media provides a universal medium for publishing text, video and images instantly
• Interactions between users themselves provide interesting information (e.g. reactions, voting)
• Very subject to manipulation
Different types of news providers

There are four types of news providers:

1. **Printed media**
   They provide readers with access to some macroeconomic, regulatory, and press announcements, and their journalists issue commentary on many companies.

2. **Newswires**
   There are many companies that utilize technological advances to improve the news dissemination process.
Different types of news providers

3. Market Data Providers
Market Data Providers aim to deliver both prices and news as quickly as possible to their customers. There have always been far more stock transactions than news announcements and thus prices are the core focus for most market data providers. They transmit macroeconomic, regulatory and press announcements, analyst recommendations and journalistic commentary directly to their clients.

4. News Aggregators
By using news aggregators, clients can access content generated from a wide variety of sources from multiple media formats (newspapers, newswires, websites). This type of service allows clients to access a much broader range of content through one easy interface, with an inherent delay from the time the content was made available by the source.
Main sources for user generated contents

- **Twitter**
  - Short text messages. High flow. Not rich meta data

- **Facebook**
  - Rich text content. Highly interconnected.

- **YouTube**
  - Multimedia content.
3. News and financial markets
Investor Sentiment and Asset Price Discovery

On a sunny day...

...people - including investors – tend to be in a better mood and therefore...

...stock prices go up!

Hirshleifer & Shumway (2003)

- found that sunlight is highly significantly correlated with daily stock returns
- confirmed their psychological hypothesis: sunlight affects mood, and mood affects prices
Background…

• Using statistics, or machine learning methods to extract, identify, or otherwise characterize the sentiment content of a text unit

• Sentiment analysis is challenging problem
  – Domain dependent
  – Text type analysed varies in complexity, e.g. financial news are more objective than tweets

• Applications of Sentiment Analysis:
  – Opinion Mining/ Sentiment analysis helps businesses/organisations in monitoring the news/blogs and social media sources
previous studies indicate that the amount of Twitter data is more than sufficient to compute an intraday sentiment index for financial instruments

existing intraday studies comparing trading data and Twitter messages do not employ text mining

difficulty of identifying appropriate Twitter messages that are clearly attributable to financial instruments and express some kind of sentiment
Many approaches

• Buy a news sentiment dataset
  – Thomson Reuters News Analytics (TRNA)
  – Dow Jones Lexicon performs sentiment analysis using a dictionary-based approach
  – RavenPack provides systematic sentiment scoring to text using traditional language processing
  – Microsoft Azure is can also be used to perform sentiment analysis

• Construct your own sentiment dataset
  – Wide range of techniques available
  – E.g. Python NLTK Text Classification (movie reviews)
    – https://newsapi.aylien.com/
### Thomson Reuters News Analytics (TRNA)

#### Reuters NewsScope Sentiment Engine: data from an automated news engine

<table>
<thead>
<tr>
<th>IDN_TIME</th>
<th>BCAST_REF</th>
<th>ITEM_ID</th>
<th>RELEVANCE</th>
<th>SENTIMENT</th>
<th>SENT_POS</th>
<th>SENT_NEUT</th>
<th>SENT_NEG</th>
<th>ITEM_TYPE</th>
<th>BCAST_TEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>27 MAY 2011 15:48:33.120</td>
<td>LTO.MI</td>
<td>rna.service.1.2011-05-27_15.48.32.nWEA3699.T1.1cdb1446</td>
<td>1</td>
<td>1</td>
<td>0.457817</td>
<td>0.392742</td>
<td>0.149441</td>
<td>ARTICLE</td>
<td>BRIEF-Lottomatica CFO to leave company at end of May</td>
</tr>
<tr>
<td>27 MAY 2011 15:48:52.229</td>
<td>OVTI.O</td>
<td>rna.service.1.2011-05-27_15.48.51.nl3E7GR14.T1.2fdd2bea</td>
<td>1</td>
<td>-1</td>
<td>0.196533</td>
<td>0.113616</td>
<td>0.689852</td>
<td>ARTICLE</td>
<td>Shares of OmniVision fall on dependence on Apple</td>
</tr>
<tr>
<td>27 MAY 2011 15:48:52.229</td>
<td>2498.TW</td>
<td>rna.service.1.2011-05-27_15.48.51.nl3E7GR14.T1.2fdd2bea</td>
<td>0.218218</td>
<td>-1</td>
<td>0.164928</td>
<td>0.29081</td>
<td>0.544263</td>
<td>ARTICLE</td>
<td>Shares of OmniVision fall on dependence on Apple</td>
</tr>
<tr>
<td>27 MAY 2011 15:48:52.229</td>
<td>MMI.N</td>
<td>rna.service.1.2011-05-27_15.48.51.nl3E7GR14.T1.2fdd2bea</td>
<td>0.218218</td>
<td>-1</td>
<td>0.164928</td>
<td>0.29081</td>
<td>0.544263</td>
<td>ARTICLE</td>
<td>Shares of OmniVision fall on dependence on Apple</td>
</tr>
<tr>
<td>27 MAY 2011 15:48:52.229</td>
<td>MSI.N</td>
<td>rna.service.1.2011-05-27_15.48.51.nl3E7GR14.T1.2fdd2bea</td>
<td>0.218218</td>
<td>-1</td>
<td>0.164928</td>
<td>0.29081</td>
<td>0.544263</td>
<td>ARTICLE</td>
<td>Shares of OmniVision fall on dependence on Apple</td>
</tr>
<tr>
<td>27 MAY 2011 15:48:52.229</td>
<td>6758.T</td>
<td>rna.service.1.2011-05-27_15.48.51.nl3E7GR14.T1.2fdd2bea</td>
<td>0.218218</td>
<td>1</td>
<td>0.464075</td>
<td>0.39339</td>
<td>0.144615</td>
<td>ARTICLE</td>
<td>Shares of OmniVision fall on dependence on Apple</td>
</tr>
<tr>
<td>27 MAY 2011 15:48:52.229</td>
<td>AAP.L.O</td>
<td>rna.service.1.2011-05-27_15.48.51.nl3E7GR14.T1.2fdd2bea</td>
<td>0.218218</td>
<td>-1</td>
<td>0.214353</td>
<td>0.117423</td>
<td>0.856222</td>
<td>ARTICLE</td>
<td>Shares of OmniVision fall on dependence on Apple</td>
</tr>
<tr>
<td>27 MAY 2011 15:48:52.229</td>
<td>005930.KS</td>
<td>rna.service.1.2011-05-27_15.48.51.nl3E7GR14.T1.2fdd2bea</td>
<td>0.218218</td>
<td>1</td>
<td>0.464075</td>
<td>0.39339</td>
<td>0.144615</td>
<td>ARTICLE</td>
<td>Shares of OmniVision fall on dependence on Apple</td>
</tr>
<tr>
<td>27 MAY 2011 15:48:54.269</td>
<td>OILS</td>
<td>rna.service.1.2011-05-27_15.48.54.nl3E7GR2FO.T1.acb33b6e</td>
<td>1</td>
<td>1</td>
<td>0.725917</td>
<td>0.115735</td>
<td>0.019269</td>
<td>ALERT</td>
<td>Indore Soybean, Soy oil, Soy meal Close- May 27</td>
</tr>
<tr>
<td>27 MAY 2011 15:49:56.385</td>
<td>MNI.N</td>
<td>rna.service.1.2011-05-27_15.49.56.nWNA9715.A2.cf73e8b6</td>
<td>0.57735</td>
<td>1</td>
<td>0.852683</td>
<td>0.46301</td>
<td>0.144615</td>
<td>ALERT</td>
<td>RPT-MOODY'S CHANGES MCCLATCHY'S &lt;MNI.N&gt; RATING OUTLOOK TO POSITIVE ON MIAMI LAND SALE</td>
</tr>
</tbody>
</table>

The sample set above contains ten selected variables out of 89.
Figure 1 Sentiment analysis techniques
Other types of sentiment analysis

- Image and video analysis
  - Face features recognition
  - Crowd behaviour analysis
  - Extracting and analysing text from multimedia

- Internet of things
  - Sensor networks (presence, health, consumption etc.)
  - Intelligent devices (buildings, cars, etc.)
Summary

- Financial markets give information about:
  - Company
  - Country
  - Industry sector

- Text sources give information about:
  - An event
  - A company
  - A campaign

- Project looking at developing software that provide information on:
  - How news affect markets?
  - How markets affect news?

- Many applications:
  - Market trading and investment
  - Planning and forecasting
Summary of Learning Outcomes

• Learning a new domain
  – Financial markets
  – News processing
  – Social media data

• Application development
  – Building a data analysis dashboard for traders
  – Incremental building (from API to application)
  – Exploiting publicly available data (teams free to use additional data sources)

• Software Development Skills
  – Business Analysis
  – API design and Implementation
  – Coding and Testing
  – Software Reuse