Identification of Fine Grained Feature Based Event and Sentiment Phrases from Business News Stories

Brett Drury

LIAAD-INESC

May 25, 2011
LIAAD-INESC
Laboratory of Artificial Intelligence and Decision Support
Porto Portugal
Introduction

Lexicons

Learning Features and Modifiers

Grammar Induction

Evaluation
News Can Move Markets !!!

Osama bin Laden's death boosts stock markets
Wall Street climbs to three-year high before falling back, while oil and precious metals drop sharply

Julia Kollewe and Rowenna Davis
guardian.co.uk, Monday 2 May 2011 18.51 BST

Osama bin Laden's death is announced on TV in Tokyo, where the Nikkei index jumped in reaction to the news. Photograph: Issei Kato/Reuters
Six-Year-Old News Story Causes United Airlines Stock to Plummet — UPDATE Google Placed Wrong Date on Story

By Kim Zetter  September 8, 2008 | 3:50 pm | Categories: The Ridiculous

I’m surprised this hasn’t happened before now.

This isn’t a story about security (although it is about securities), but it’s so remarkable I thought I’d include it here anyway.

A worker at a Miami investment advisory firm called Income Securities Advisors, which publishes news alerts that get distributed through the Bloomberg News Service, did a Google search on bankruptcies this morning and got back search results that included a six-year-old story published in the South Florida Sun Sentinel about the 2002 bankruptcy filing by United Airlines.

The employee mistook the news for a current story — despite the date clearly marked on it (see update below) and other information in the article “that would clearly lead a reader to the conclusion that it was related to events in 2002” — and included it in a subscription newsletter that was distributed through Bloomberg.
News analysis refers to the measurement of the various qualitative and quantitative attributes of textual (unstructured data) news stories.

- sentiment
- relevance
- novelty

Expressing information in a numerical manner allows the manipulation of the information contained in news.
(Source: Wikipedia)
What type of information moves markets?

- Events - Entering bankruptcy
- Sentiment - A poor review of a company’s future prospects

Differences in market reaction?

- Events - Short term reaction
- Sentiment - Longer term reaction
Events

- DeBont and Thaler (1985)
- Market Initially Overreacts and Corrects

Example: Reaction of Markets to Bin Laden’s Death
Sentiment

- Lack of dramatic market change
- Longer period of time
- Changes in writing style in company reports
- More accurate predictor than numeric information in company report
Current Approaches

- Supervised Learning
- Large Amounts of Training Data
- Classify News Story
- Assign Relevance to News Story
- Final Score = (Classification Score * Relevance Score)
Lack of training data

News stories may make reference \( \geq 1 \) economic entity

Accurately locate economic entity

Scoring phrases must take into account: negation and sentiment modification

Identify larger phrases which contain smaller phrases
GATE

Rules written in JAPE

”Regular Expressions” for Annotations
- Crawl RSS feeds from free news sites
- Text extracted and sent to Open Calais
- Meta-data appended to each story

News Story Acquisition Pipeline
Crawl RSS ➔ Store Information (headline, date etc) ➔
Extract Text ➔ Send Text to Open Calais ➔ Store RDF
Open Calais Meta-Data

Business Sectors - From Corpus

"Identification, extraction and population of collective named entities"

Entity2010 – Workshop on Resources and Evaluation for Entity Resolution and Entity

Add Entries to Gate Gazetteer

Company List: 2847 — > 42828 entries

USwitch, thinkorswim Inc, easyBus, ZyLAB

telecommunication business, telcoms industry, telco sector
Identify Event Verbs

POS TAG Sentences

Co-occurrence of Verbs with Economic Actors

Sorted by frequency

Verbs verified by hand

Expand with verbs from Levin Categories

Verb Net bounce: drift, drop, float ...

word forms JSpell drop: dropped, dropping, drops ...

330 verbs
► No existing resources for scoring verbs
► hand scored (+1 = positive, -1 = negative)
► positive = 186, negative = 146
► Sorted by frequency
► Verbs verified by hand

<table>
<thead>
<tr>
<th>Verb Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obtained</td>
<td>gain(+), add(+), forge(+), win(+), attract(+)</td>
</tr>
<tr>
<td>Lost</td>
<td>fire(-), cut(-), cancel(-)</td>
</tr>
<tr>
<td>Direction</td>
<td>climb(+), fall(-), boost(+), down(-)</td>
</tr>
</tbody>
</table>
- Extract adjectives
- Sort by frequency and score with Sentiwordnet
- Check adjectives by hand
- Propagate scores by connectives
- Expand adjectives with Wordnet
- 2520 Adjectives
- Learn Features Associated with Economic Actors and Verbs/Adjectives
- Typically Nouns: Profits, Costs ....
- Learnt by Point Wise Mutual Information
- Capture Words With Statistical Relationship With Economic Actor and Verb / Adjective

<table>
<thead>
<tr>
<th>Categorization</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success Measures</td>
<td>footfall, sales, profits, demand</td>
</tr>
<tr>
<td>Third Parties</td>
<td>investors, analysts, investors, economists, regulators, consumers</td>
</tr>
</tbody>
</table>
- Learn Modifiers Associated with Economic Actors and Verbs / Adjectives
- Typically Adverbs: Sharply, Not, Piffling
- Learnt by Point Wise Mutual Information
- Hand Scored

<table>
<thead>
<tr>
<th>Sentiment modifier categorization</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximization</td>
<td>sharply, super, perfectly</td>
</tr>
<tr>
<td>Minimization</td>
<td>rickety, piffling, just</td>
</tr>
<tr>
<td>Negation</td>
<td>not, none, never</td>
</tr>
</tbody>
</table>
Order Independent Triples
- Implemented in JAPE
- Economic Actor, Verb/Adjective, Object
- Microsoft, dropped, profits
- Economic Actor Missing
- Combine Patterns
- Rules: separated by individual token (space, comma etc) or continuation
- Target Location
- Complete Pattern: Economic Actor (EA)
- Partial Pattern: Back to nearest EA
- Exclude third parties
Event Score: determined by verb
Special Features reverse verb scores
Rise in Costs (-), Rise in profits (+)

Sentiment Score: determined by adjective
AVAC Algorithm: adverbs to modified the sentiment score
Gold Standard:

- Identification of phrase
- Differentiation of an event from sentiment,
- Correct identification of target
- Direction of sentiment or event.

<table>
<thead>
<tr>
<th>Evaluation Item</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentiment phrase extraction and direction</td>
<td>0.71</td>
<td>0.94</td>
</tr>
<tr>
<td>Event phrase extraction and direction</td>
<td>0.84</td>
<td>0.83</td>
</tr>
<tr>
<td>Sentiment Target Extraction</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>Event target extraction</td>
<td>0.84</td>
<td>0.77</td>
</tr>
</tbody>
</table>
Accurate Headline Classifier
Select Stories by Headline
Induce Classifier from Data
Rules as trainer (S.T Hybrid)
Constrained Self-Training

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Headline</th>
<th>Story Text</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alignment</td>
<td>0.57 ± (0.01)</td>
<td>0.57 ± (0.01)</td>
<td>0.57 ± (0.00)</td>
</tr>
<tr>
<td>Hybrid</td>
<td>0.66 ± (0.04)</td>
<td>0.57 ± (0.06)</td>
<td>0.58 ± (0.04)</td>
</tr>
<tr>
<td>Rule Trained</td>
<td>0.77 ± (0.01)</td>
<td>0.60 ± (0.01)</td>
<td>0.65 ± (0.01)</td>
</tr>
<tr>
<td>S.T. Hybrid</td>
<td>0.84 ± (0.01)</td>
<td>0.71 ± (0.01)</td>
<td>0.77 ± (0.01)</td>
</tr>
</tbody>
</table>
Can we make money? Yes!

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Voting</th>
<th>Headline</th>
<th>Description</th>
<th>Story Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alignment</td>
<td>-12.2%</td>
<td>-24.5%</td>
<td>-20.6%</td>
<td>-0.1%</td>
</tr>
<tr>
<td>Hybrid</td>
<td>-10.6%</td>
<td>-10.6%</td>
<td>-10.6%</td>
<td>-10.6%</td>
</tr>
<tr>
<td>Rule Trained</td>
<td>16.5%</td>
<td>16.8%</td>
<td>14.8%</td>
<td>-1.2%</td>
</tr>
<tr>
<td>S.T. Hybrid</td>
<td>-10.6%</td>
<td>33.8%</td>
<td>-6.5%</td>
<td>-12.3%</td>
</tr>
</tbody>
</table>
Not published: Rule only system made highest returns

- Very high confidence selections
- Abdication on ambiguous stories
- Very high returns
Sophisticated trading evaluation

- Event and Sentiment Added As Features
- Predict 1, 3, 5 and 10 days ahead
- Currently Running
- Results Evaluation in June / July 2011

We expect:

- Event information improves short term prediction
- Sentiment information improves longer term prediction
- Combination of the two improves general prediction
Conclusion:

- Simple technique
- Functions well on business news
- Business news generally simple
- Fails on complex text (e.g. quotations)
- Domain specific
- Business lexicon changes: recalculate lexicon regularly
Questions

Takk / Obrigado / Thanks
Questions Please !!!
Please Send: Requests for materials, suggestions, extended final work to Brett.Drury@gmail.com