

# 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

*And it does not have to be true !!!*

## 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

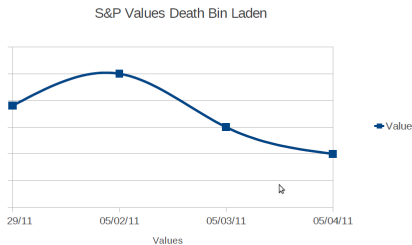
Differences in market reaction?

- ▶ 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  $>1$  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

The screenshot displays the GATE Developer 6.0 interface. On the left, a tree view shows the project structure with 'GATE Document\_0001B' selected. The main window shows a document with several paragraphs of text. Annotations are visible as colored boxes over the text, such as 'CompanyNegative' (red) and 'CompanyPositive' (yellow). A 'Messages' pane on the right lists the annotations with their corresponding features. Below the text, a table shows the details of the selected annotations.

Type	SetStart	End	Id	Features
4 Annotations (0 selected) Select:				
Document Editor Initialisation Parameters				

- ▶ 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

Verb Category	Examples
Obtained	gain(+), add(+), forge(+), win(+), attract(+)
Lost	fire(-), cut(-), cancel(-)
Direction	climb(+), fall(-), boost(+), down(-)



- ▶ Extract adjectives
- ▶ Sort by frequency and score with Sentiwordnet
- ▶ Check adjectives by hand
  
- ▶ Propagate scores by connectives
- ▶ Expand adjectives with Wordnet
- ▶ 2520 Adjectives

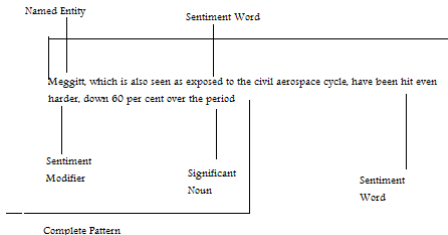
- ▶ Learn Features Associated with Economic Actors and Verbs/Adjectives
- ▶ Typically Nouns: Profits, Costs ....
- ▶ Learnt by Point Wise Mutual Information
- ▶ Capture Words With Statistical Relationship With Economic Actor and Verb / Adjective

Categorization	Examples
Success Measures	footfall, sales, profits, demand
Third Parties	investors, analysts, investors, economists, regulators, consumers

- ▶ Learn Modifiers Associated with Economic Actors and Verbs / Adjectives
- ▶ Typically Adverbs: Sharply, Not, Piffling
- ▶ Learnt by Point Wise Mutual Information
- ▶ Hand Scored

Sentiment modifier categorization	Examples
Maximization	sharply, super, perfectly
Minimization	rickety, piffling, just
Negation	not, none, never

- ▶ 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.

Evaluation Item	Recall	Precision
Sentiment phrase extraction and direction	0.71	0.94
Event phrase extraction and direction	0.84	0.83
Sentiment Target Extraction	0.74	0.74
Event target extraction	0.84	0.77

- ▶ Accurate Headline Classifier
- ▶ Select Stories by Headline
- ▶ Induce Classifier from Data
- ▶ Rules as trainer (S.T Hybrid)
- ▶ Constrained Self-Training

Classifier	Headline	Story Text	Description
Alignment	$0.57 \pm (0.01)$	$0.57 \pm (0.01)$	$0.57 \pm (0.00)$
Hybrid	$0.66 \pm (0.04)$	$0.57 \pm (0.06)$	$0.58 \pm (0.04)$
Rule Trained	$0.77 \pm (0.01)$	$0.60 \pm (0.01)$	$0.65 \pm (0.01)$
S.T. Hybrid	$0.84 \pm (0.01)$	$0.71 \pm (0.01)$	$0.77 \pm (0.01)$



Can we make money? Yes!

Strategy	Voting	Headline	Description	Story Text
Alignment	-12.2%	-24.5%	-20.6%	-0.1%
Hybrid	-10.6%	-10.6%	-10.6%	-10.6%
Rule Trained	16.5%	16.8%	14.8%	-1.2%
S.T. Hybrid	-10.6%	33.8%	-6.5%	-12.3%

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](mailto:Brett.Drury@gmail.com)