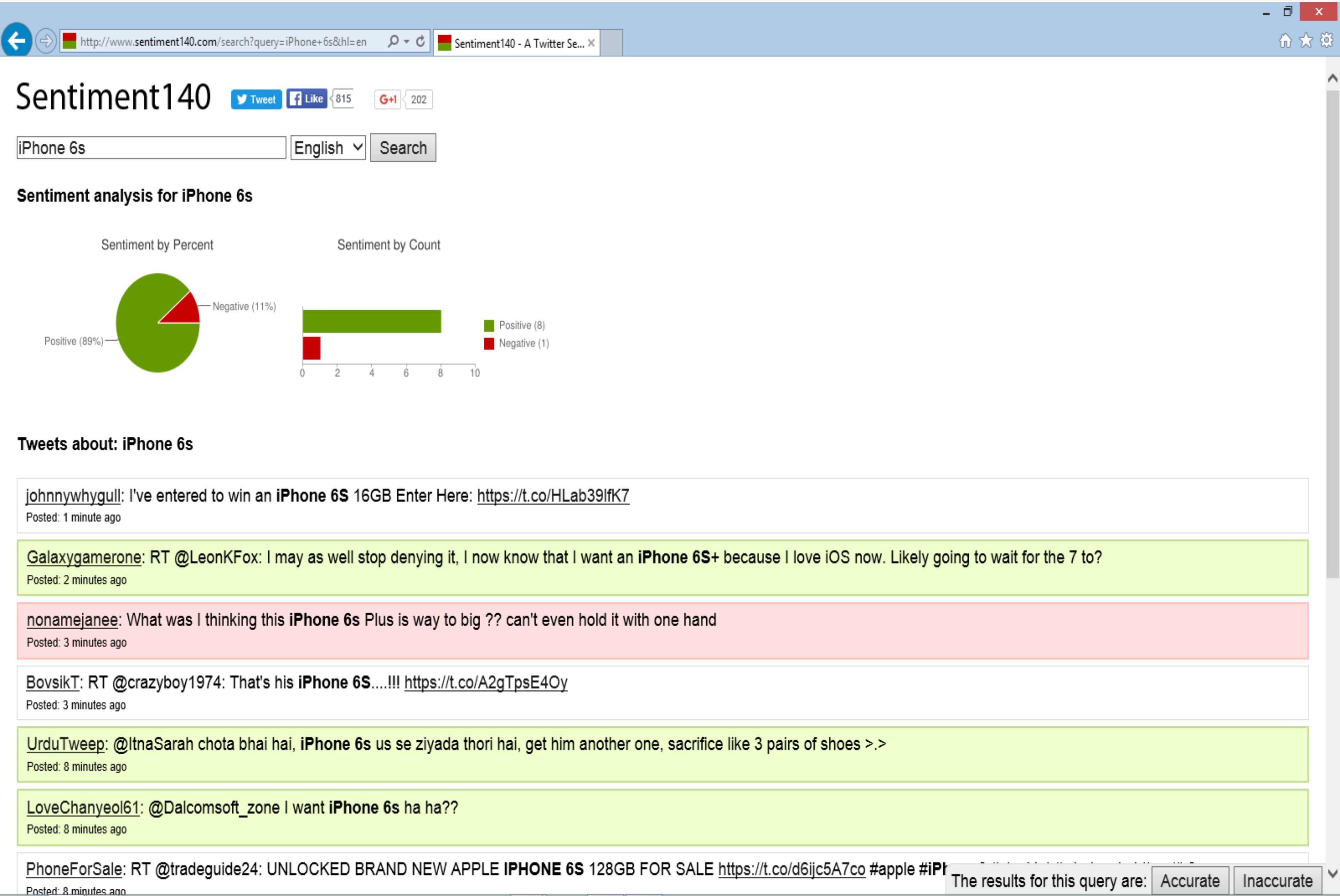
Sentiment Analysis

CIS 600, Spring 2018

February 20, 2018

Sentiment Analysis: An Example



Sentiment Analysis

- **Sentiment analysis**, also called **opinion mining**, is the field of study that analyzes people's:

  opinions, sentiments, evaluations, appraisals, attitudes, and emotions

  towards **entities** such as:

  products, services, organizations, individuals, issues, events, topics, and their attributes.

Sentiment Analysis

- The following different names all refer to sentiment analysis, emphasizing on slightly different but related tasks:

  sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc.

- They now all fall under the umbrella term **sentiment analysis**, although some people in the academia still use opinion mining.

Affect

- **Definition (Merriam-Webster)** a set of observable manifestations of a subjectively experienced emotion

  … *patients* … *showed perfectly normal reactions and affects* …

- *Effect* and *affect* are often confused because of their similar spelling and pronunciation. The uncommon noun *affect*, which has a meaning relating to psychology, is also sometimes mistakenly used for the very common *effect*. In ordinary use, the noun you will want is *effect* (*waiting for the new law to take effect*) (*the weather had an effect on everyone's mood*).

Sentiment Analysis & NLP

- Although linguistics and natural language processing (NLP) have a long history, little research had been done about people's opinions and sentiments before the year **2000**.
- But since then, sentiment analysis has become a very active research area. There are several reasons for that:
  - It has a wide arrange of applications, almost in every domain. This provides a strong motivation for research.
  - It offers many challenging research problems, which had never been studied before.
  - For the first time in human history, we now have a huge volume of opinionated data in social media.
    - In fact, sentiment analysis is now right at the center of the <u>social media research</u>.

Sentiment Analysis & Social Media

- ❖ Social media offers a **World-of-Mouth**
  - ❖ **User-generated content** is teeming with people's feelings & emotions
    - ❖ Personal experiences and opinions about anything in forums, blogs, Twitter, etc.
    - ❖ Postings at social networking sites, e.g., facebook.
    - ❖ Comments about articles, issues, topics, reviews, etc.
- ❖ Social media offers <u>an ocean of opportunities</u>
  - ❖ Social media is on the global scale
  - ❖ The ability to survey a large pool of people easily has enormous potential.

An Ocean of Opportunities

- **Businesses** spend big bucks to find consumer opinions using consultants, surveys and focus groups, etc. to Benchmark products and services

- **Advertisers** want know what type of people are willing to purchase their products, and why.
  - **Ads placements**:
    - Place ads in the social media content
    - Place an ad if one praises a product.
    - Place an ad from a competitor if one criticizes a product
  - **Brand Monitoring**

An Ocean of Opportunities

- For individuals
  - **Consumers** want to make more informed decisions to buy products or to use services
  - **Politicians** want to know public opinions (polls) about their policies and their reputation. Traditional polling methods can take days and weeks to deliver results.
  - **Voters** want to know public opinions about political candidates and issues.
  - **Researchers** want to develop systems that could extract valuable data accurately (and publish more papers)

Representative Applications

Products & Sales

- **Mishne & Glance (2006):** showed that positive sentiment is a better predictor of movie success than buzz counts.
- **Sadikov et al. (2009**): made the same prediction using more sentiment and other features.
- **Liu et al. (2007)**: used sentiment analysis to predict **movie sales**
- **Asur & Huberman (2010), Joshi et al. (2010):** Used Twitter data, movie reviews, and blogs to predict box-office revenues for **movies**.
- **McGlohon et al. (2010):** used sentiment analysis to rank **products and merchants**

Predicting Movie Sales from Blogger Sentiment

Gilad Mishne
Informatics Institute, University of Amsterdam
Kruislaan 403, 1098SJ Amsterdam, The Netherlands
gilad@science.uva.nl

Natalie Glance
Intelliseek Applied Research Center
5001 Baum Blvd, Pittsburgh, PA 15213
nglance@intelliseek.com

They showed that positive sentiment is a better predictor of movie success than buzz counts.

Abstract

The volume of discussion about a product in weblogs has recently been shown to correlate with the product's financial performance. In this paper, we study whether applying sentiment analysis methods to weblog data results in better correlation than volume only, in the domain of movies. Our main finding is that positive sentiment is indeed a better predictor for movie success when applied to a limited context around references to the movie in weblogs, posted prior to its release.

If my film makes one more person miserable, I've done my job.
– Woody Allen

Introduction

Weblogs provide online forums for discussion that record the voice of the public. Woven into this mass of discussion

his unhappiness with Dell as a company reverberated across the blogosphere and into the press, creating a public relations mini-fiasco for Dell.

Hard evidence of the influence of online discussion on consumer decisions is beginning to emerge. An Intelliseek survey of 660 online consumers showed that people are 50 percent more likely to be influenced by word-of-mouth recommendations from their peers than by radio/TV ads [3]. Researchers at IBM reported that online blog postings can successfully predict spikes in the sales rank of books (Gruhl, Guha, Kumar, Novak, & Tomkins, 2005), showing that the raw number of posts about a book was a good predictor.

However, opinion comes in many flavors: positive, negative, mixed, and neutral mixed in with splashes of sarcasm, wit and irony. Novel techniques in sentiment analysis make it possible to quantify the aggregate level of positive vs. neg-

apparently an early easter is bad for apparel sales. who knew? i'll probably go see "guess who?" this weekend. i liked miss congeniality but the sequel [link to IMDB's page for "Miss Congeniality 2"] looks *awful*. and seattle's too much of a backwater to be showing D.E.B.S. i had to wait forever to see saved! too. mikalah gordon got kicked off american idol last night. while she wasn't the best singer, i wish …
Monday, March 28, 2005 - Miss Congeniality 2: Armed and Fabulous. I know this is overdue, but I wanted to use this opportunity to discuss an important topic. The issue at hand is known as the Sandra Bullock Effect (SBE). This theorem was first proposed by my brother, Arthur, so he is the real expert, but I will attempt to explain it here. The SBE is the degree to which any movie becomes watchable simply by the presence of a particular actor or actress who you happen to be fond of. For example, if I told you that someone made a movie about a clumsy, socially awkward, dowdy female police officer who goes undercover as a beauty pageant contestant to foil some impenetrable criminal conspiracy, you'd probably think to yourself, "Wow that sounds pretty dumb." And you'd be right. However…

Table 4: Typical references to movies in blogs: pre-release (top), and post-release (bottom).

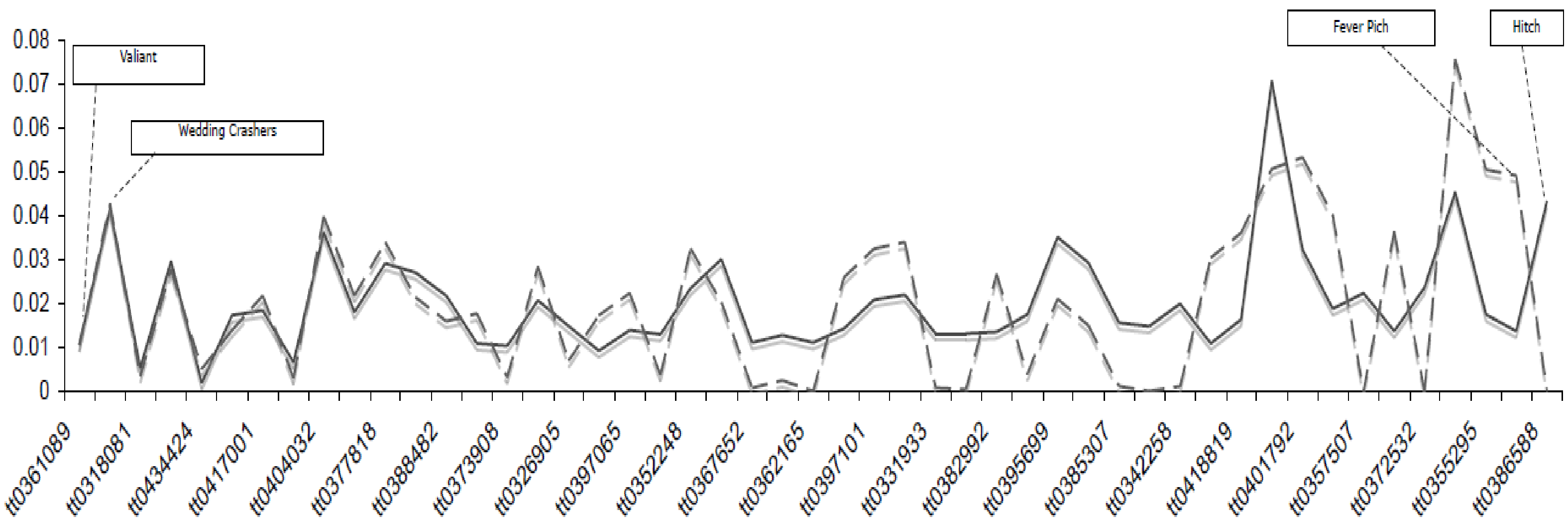


Figure 2: Per-movie comparison of income per screen (blue, continuous line) and positive references (green, dashed line), sorted by degree of correlation. For space reasons, the X-axis shows only the movie IMDB ID.

Eldar Sadikov and **Aditya Parameswaran** and **Petros Venetis**
Computer Science Dept., Stanford University, Stanford, CA 94305
esadikov,adityagp,venetis@stanford.edu

They made the same prediction using more sentiment and other features.

Abstract

Analysis of a comprehensive set of features extracted from blogs for prediction of movie sales is presented. We use correlation, clustering and time-series analysis to study which features are best predictors.

Introduction

In this work, we attempt to assess if blog data is useful for prediction of movie sales and user/critics ratings. Here are our main contributions:

- We evaluate a comprehensive list of features that deal with movie references in blogs (a total of 120 features) using the full `spinn3r.com` blog data set for 12 months.
- We find that aggregate counts of movie references in blogs are highly predictive of movie sales but not predictive of user and critics ratings.
- We identify the most useful features for making movie sales predictions using correlation and KL divergence as

- Weekly box office sales, weeks 1–5

The first two of these variables were collected from `rottentomatoes.com` and the last two were collected from `the-numbers.com`.

We then collected a list of the top 300 Box Office movies of 2008 from `the-numbers.com` and manually filtered the list for the titles that were similar to the common English phrases. For example, searching for 'Wanted' or '21' triggered many false positives that might not refer to the movies. Since such titles could have skewed our results in a non-uniform way, we eliminated such titles. For the remaining 197 movies, we constructed a list of regular expressions and used it to collect posts that referenced movie titles. Since the focus was on US movies, we excluded posts in languages other than English; and to filter out obvious spam posts, we excluded posts that contained more than 30 links. The remaining posts effectively became our working data set from which we extracted our features.

In addition to features that counted movie references

ARSA: A Sentiment-Aware Model for Predicting Sales Performance Using Blogs

Yang Liu[1], Xiangji Huang[2], Aijun An[1] and Xiaohui Yu[2]
[1]Department of Computer Science and Engineering
York University, Toronto, Canada
[2]School of Information Technology
York University, Toronto, Canada
yliu@cse.yorku.ca, jhuang@yorku.ca, aan@cse.yorku.ca, xhyu@yorku.ca

ABSTRACT

Due to its high popularity, Weblogs (or blogs in short) present a wealth of information that can be very helpful in assessing the general public's sentiments and opinions. In this paper, we study the problem of mining sentiment information from blogs and investigate ways to use such information for predicting product sales performance. Based on an analysis of the complex nature of sentiments, we propose Sentiment PLSA (S-PLSA), in which a blog entry is viewed as a

commentaries or discussions on a particular subject, ranging from mainstream topics (e.g., food, music, products, politics, etc.), to highly personal interests [13]. Since many bloggers choose to express their opinions online, blogs serve as an excellent indicator of public sentiments and opinions.

This paper studies the predictive power of opinions and sentiments expressed in blogs. We focus on the blogs that contain reviews on products. Since what the general public thinks of a product can no doubt influence how good it sells, understanding the opinions and sentiments expressed

To better illustrate the effects of the parameter values on the prediction accuracy, we present in Figure 3 the experimental results on a particular movie, *Little Man*. For each parameter, we plot the predicted box office revenues and the true values for each day using different values of the parameter. It is evident from the plots that the responses to each parameter are similar to what is observed from Figure 2. Also note that the predicted values using the optimal parameter settings are close to the true values accross the whole time span. Similar results are also observed on other movies, demonstrating the consistency of the proposed approach for different days.

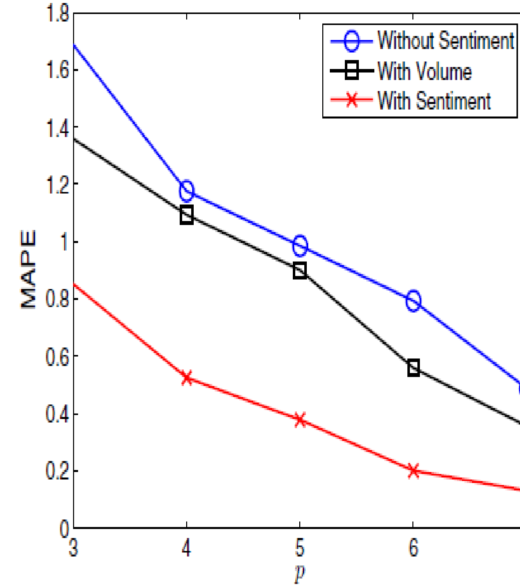


Figure 4: ARSA vs. alternative methods

6.3 Comparison with alternative methods

To verify that the sentiment information captured by the S-PLSA model plays an important role in box office revenue prediction, we compare ARSA with two alternative methods which do not take sentiment information into consideration.

We first conduct experiments to compare ARSA against the pure autoregressive (AR) model without any terms on sentiments, i.e., $y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t$. The results are shown in Figure 4. We observe the behaviors of the two models as p ranges from 3 to 7. Apparently, although the accuracy of both methods improves with increasing p, ARSA constantly outperforms the AR model by a factor of 2 to 3.

We then proceed to compare ARSA with an autoregressive model that factors in the volume of blog mentions in prediction. In Section 3, we have illustrated the characteristics of this, we experiment with the following autoregressive model that utilizes the volume of blogs mentions. In contrast to ARSA, where we use a multi-dimensional probability vector produced by S-PLSA to represent bloggers' sentiments, this model uses a scalar (number of blog mentions) to indicate the degree of popularity. The model can be formulated as

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \rho_i v_{t-i} + \epsilon_t,$$

where y_t's are obtained in the same way as in ARSA, v_{t-i} denotes the number of blog mentions on day $t-i$, and ϕ_i and ρ_i are parameters to be learned. This model can be trained using a procedure similar to what is used for ARSA. Using

Predicting the Future With Social Media

Sitaram Asur
Social Computing Lab
HP Labs
Palo Alto, California
Email: sitaram.asur@hp.com

Bernardo A. Huberman
Social Computing Lab
HP Labs
Palo Alto, California
Email: bernardo.huberman@hp.com

Used Twitter data, movie reviews, and blogs to predict box-office revenues for movies

***Abstract*—In recent years, social media has become ubiquitous and important for social networking and content sharing. And yet, the content that is generated from these websites remains largely untapped. In this paper, we demonstrate how social media content can be used to predict real-world outcomes. In particular, we use the chatter from Twitter.com to forecast box-office revenues for movies. We show that a simple model built from the rate at which tweets are created about particular topics can outperform market-based predictors. We further demonstrate how sentiments extracted from Twitter can be further utilized to improve the forecasting power of social media.**

I. Introduction

Social media has exploded as a category of online discourse

This paper reports on such a study. Specifically we consider the task of predicting box-office revenues for movies using the chatter from Twitter, one of the fastest growing social networks in the Internet. Twitter [1], a micro-blogging network, has experienced a burst of popularity in recent months leading to a huge user-base, consisting of several tens of millions of users who actively participate in the creation and propagation of content.

We have focused on movies in this study for two main reasons.

- The topic of movies is of considerable interest among the social media user community, characterized both by large number of users discussing movies, as well as a

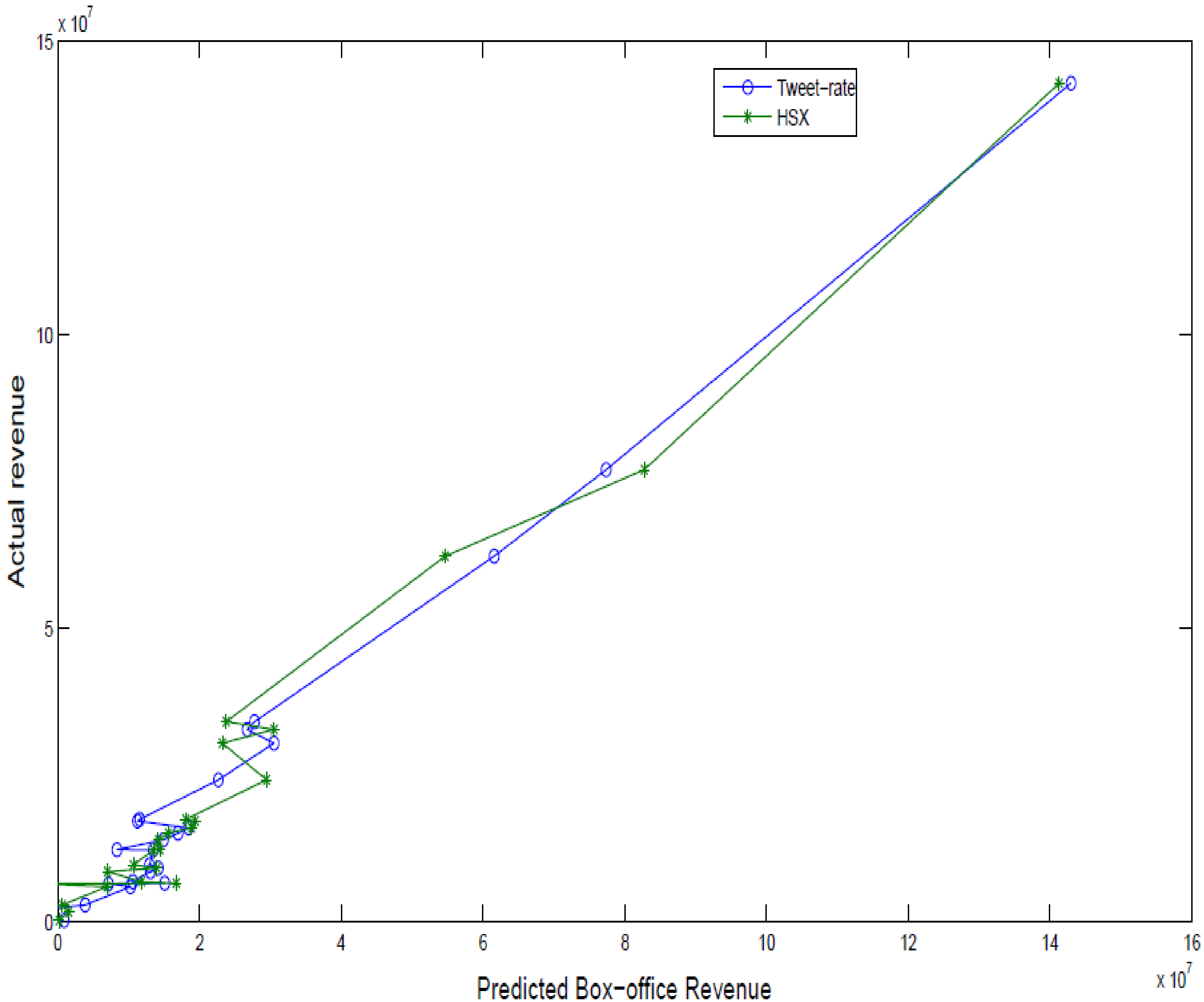


Fig. 6. Predicted vs Actual box office scores using tweet-rate and HSX predictors

Movie Reviews and Revenues: An Experiment in Text Regression*

Mahesh Joshi Dipanjan Das Kevin Gimpel Noah A. Smith
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213, USA
{maheshj,dipanjan,kgimpel,nasmith}@cs.cmu.edu

Abstract

We consider the problem of predicting a movie's opening weekend revenue. Previous work on this problem has used metadata about a movie—e.g., its genre, MPAA rating, and cast—with very limited work making use of text *about* the movie. In this paper, we use the text of film critics' reviews from several sources to predict opening weekend revenue. We describe a new dataset pairing movie reviews with metadata and revenue data, and show that review text can substitute for metadata, and even improve over it, for prediction.

1 Introduction

correlation between actual revenue and sentiment-based metrics, as compared to mention counts of the movie. (They did not frame the task as a revenue prediction problem.) Zhang and Skiena (2009) used a news aggregation system to identify entities and obtain domain-specific sentiment for each entity in several domains. They used the aggregate sentiment scores and mention counts of each movie in news articles as predictors.

While there has been substantial prior work on using critics' reviews, to our knowledge all of this work has used polarity of the review or the number of stars given to it by a critic, rather than the review text directly (Terry et al., 2005).

the dependency relation features (set III) to the n-grams does improve the performance enough to make it significantly better than the metadata-only baseline for per screen revenue prediction.

Salient Text Features: Table 3 lists some of the highly weighted features, which we have categorized manually. The features are from the text-only model annotated in Table 2 (total, not per screen). The feature weights can be directly interpreted as U.S. dollars contributed to the predicted value $\hat{y}$ by each occurrence of the feature. Sentiment-related features are not as prominent as might be expected, and their overall proportion in the set of features with non-zero weights is quite small (estimated in preliminary trials at less than 15%). Phrases that refer to metadata are the more highly weighted and frequent ones. Consistent with previous research, we found some positively-oriented sentiment features to be predictive. Some other prominent features not listed in the table correspond to special effects ("*Boston Globe*: of_the_art", "and_cgi"), particular movie franchises ("shrek_movies", "*Variety*: chronicle_of", "voldemort"), hype/expectations ("blockbuster", "anticipation"), film festival ("*Variety*: canne" with negative weight) and time of re-

	Feature	Weight ($M)
rating	pg	+0.085
	New York Times: adult	-0.236
	New York Times: rate_r	-0.364
sequels	this_series	+13.925
	LA Times: the_franchise	+5.112
	Variety: the_sequel	+4.224
people	*Boston Globe*: will_smith	+2.560
	Variety: brittany	+1.128
	^_producer_brian	+0.486
genre	*Variety*: testosterone	+1.945
	Ent. Weekly: comedy_for	+1.143
	Variety: a_horror	+0.595
	documentary	-0.037
	independent	-0.127
sentiment	*Boston Globe*: best_parts_of	+1.462
	Boston Globe: smart_enough	+1.449
	LA Times: a_good_thing	+1.117
	shame_$	-0.098
	bogeyman	-0.689
plot	*Variety*: torso	+9.054
	vehicle_in	+5.827
	superhero_$	+2.020

Table 3: Highly weighted features categorized manually. ^ and $ denote sentence boundaries. "brittany" frequently refers to Brittany Snow and Brittany Murphy. "^_producer_brian" refers to producer Brian Grazer (*The Da Vinci Code*, among others).

Star Quality: Aggregating Reviews to Rank Products and Merchants

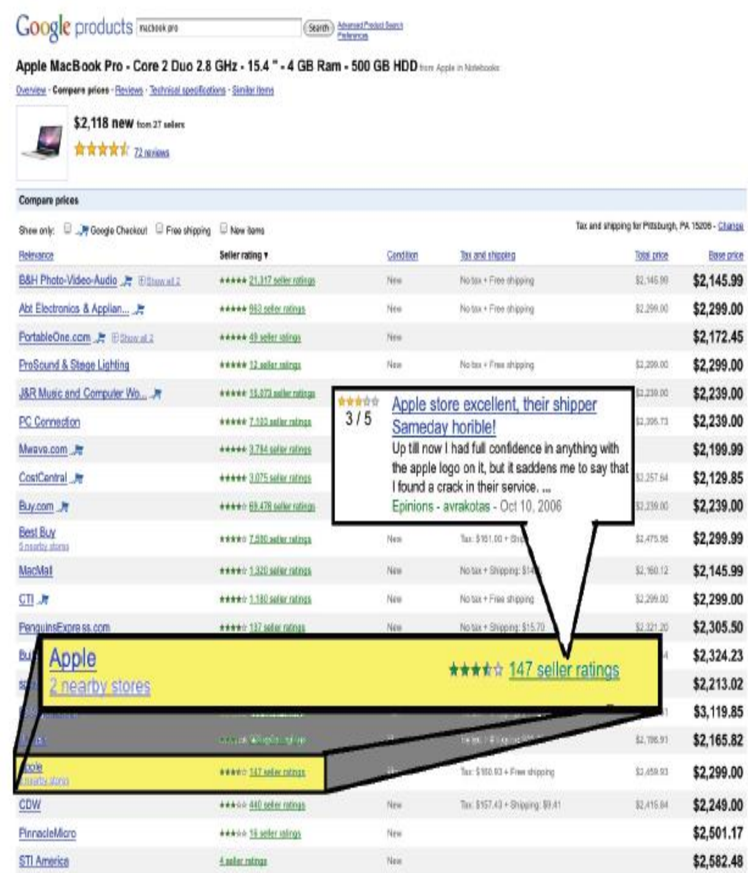
Mary McGlohon*†, Natalie Glance†, Zach Reiter†

mmcgloho@cs.cmu.edu, nglance@google.com, zachreiter@google.com

†Google, Inc, Pittsburgh PA * Carnegie Mellon University, Pittsburgh PA

Abstract

Given a set of reviews of products or merchants from a wide range of authors and several reviews websites, how can we measure the *true quality* of the product or merchant? How do we remove the bias of individual authors or sources? How do we compare reviews obtained from different websites, where ratings may be on different scales (1-5 stars, A/B/C, etc.)? How do we filter out unreliable reviews to use only the ones with "star quality"? Taking into account these considerations, we analyze data sets from a variety of different reviews sites (the first paper, to our knowledge, to do this). These data sets include 8 million product reviews and 1.5 million merchant reviews. We explore statistic- and heuristic-based models for estimating the true quality of a product or merchant, and compare the performance of these estimators on the task of ranking pairs of objects. We also apply the same models to the task of using Netflix



They used sentiment analysis to measure **social influences in online book reviews:** How existing reviews affect new reviews?

Analysis of Social Influence in Online Book Reviews

Patty Sakunkoo and Nathan Sakunkoo

Stanford University, Stanford, CA 94305
{psak, sakunkoo} @ stanford.edu

Abstract

It has been widely recognized that online opinions constitute important informational sources for consumers and producers. The open nature of communication supported by social media, however, raises an important yet unsettled question of whether and how earlier opinions affect those that come after. This poster presents a model to illustrate the relationship between existing and new reviews. Based on 12,500 Amazon reviews, our choice model shows support for the idea that social contagion may be an important mechanism guiding behaviors of online reviewers. The results thus offer novel insights toward a better understanding of contagious behaviors, as well as minority influence, among social media users.

As part of the larger effort to examine the interactions among social media users and how opinions evolve, this poster explores a new interpretation by developing and testing a model for socially-influenced online reviews. The results challenge existing theoretical conjectures above by providing an illustration of how the observed trend can be explained by a *contagious* tendency.

Methods

For preliminary investigation of the aggregate patterns, a simulation was conducted to test whether the existing opinions may prompt or provide bases for a potential reviewer to write a new review. A result from this

Representative Applications

Politics

- O'Connor et al. (2010), **From Tweets to Polls**: tracked how Twitter sentiment was linked with public opinion polls
- Tumasjan et al. (2010): tracked Twitter sentiment to predict election results
- Chen et al. (2010): used sentiment analysis to study political standpoints

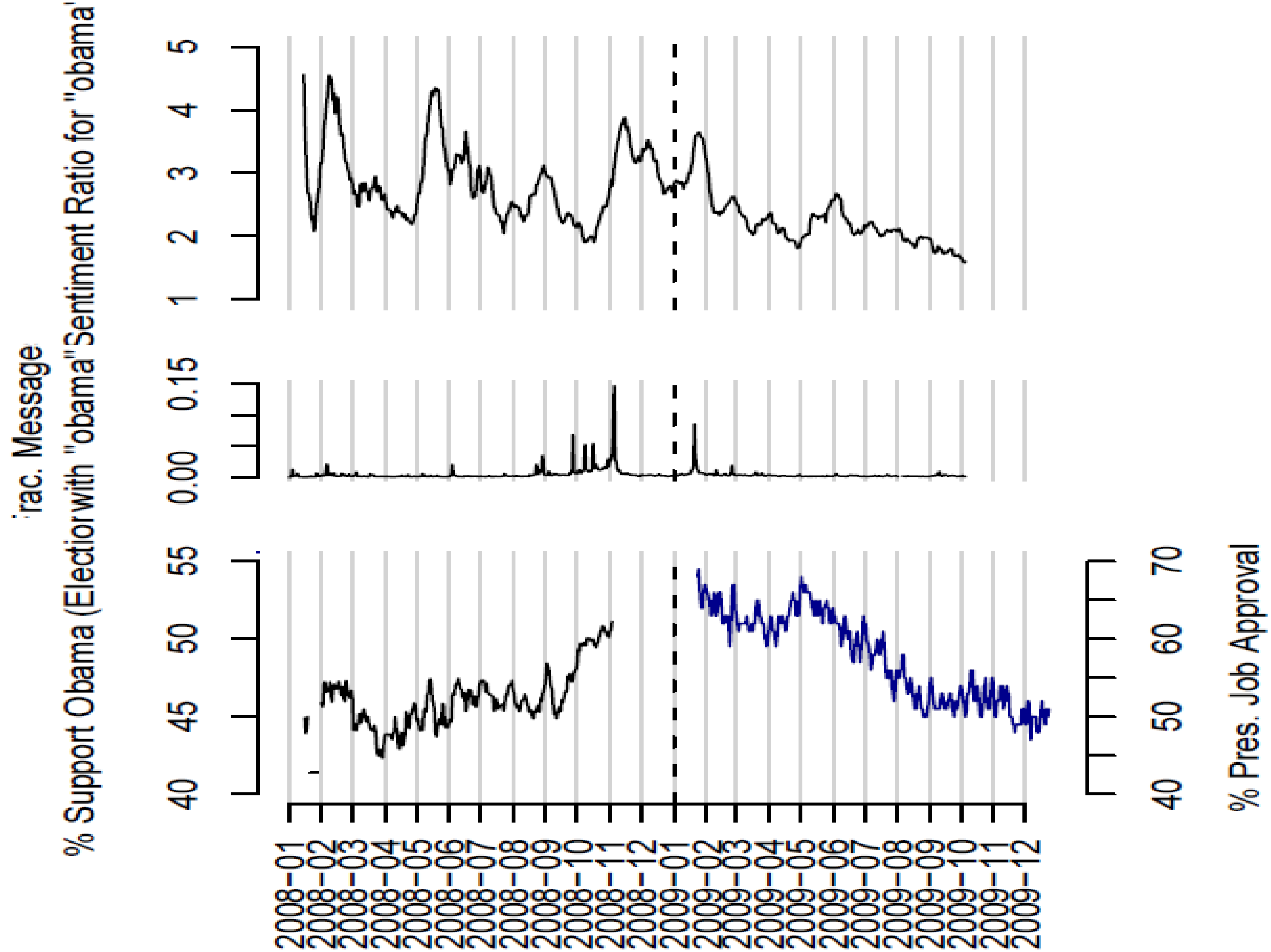


Figure 9: The sentiment ratio for *obama* (15-day window), and fraction of all Twitter messages containing *obama* (day-by-day, no smoothing), compared to election polls (2008) and job approval polls (2009).

Representative Applications

Stock Market

- ❖ Bollen et al. (2011), **Twitter mood predicts the stock market** : used Twitter moods to predict the stock market.

- ❖ Zhang and Skiena (2010), **Trading Strategies To Exploit Blog and News Sentiment**: used blog and news sentiment to study **trading strategies**
  - ❖ Their findings provided them with a sentiment-based trading strategy which gives consistently favorable returns with low volatility over a five year period (2005-2009).

- ❖ Bar-Haim et al. (2011): used sentiment analysis to identify **expert investors** (vs. non-expert)

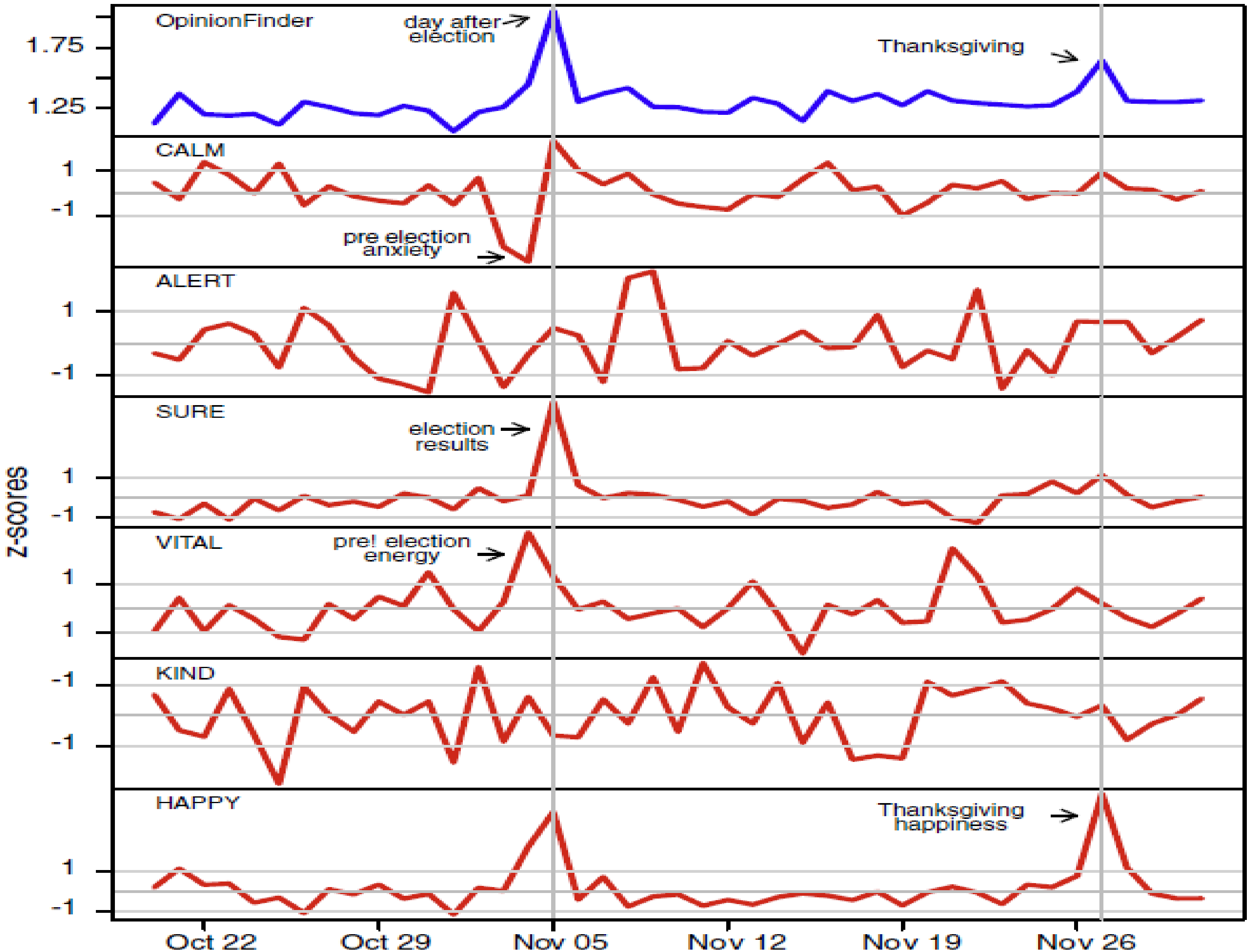
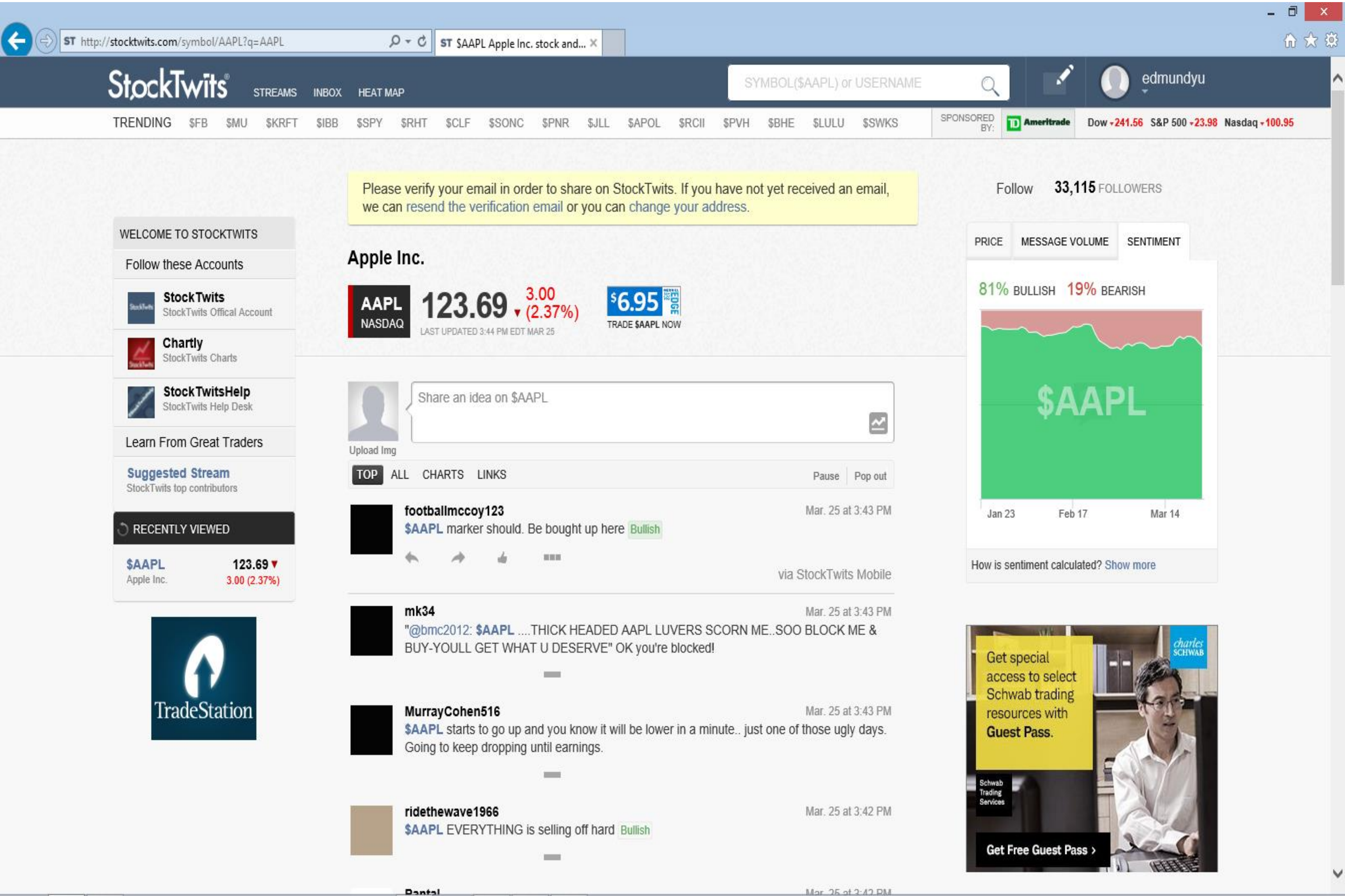


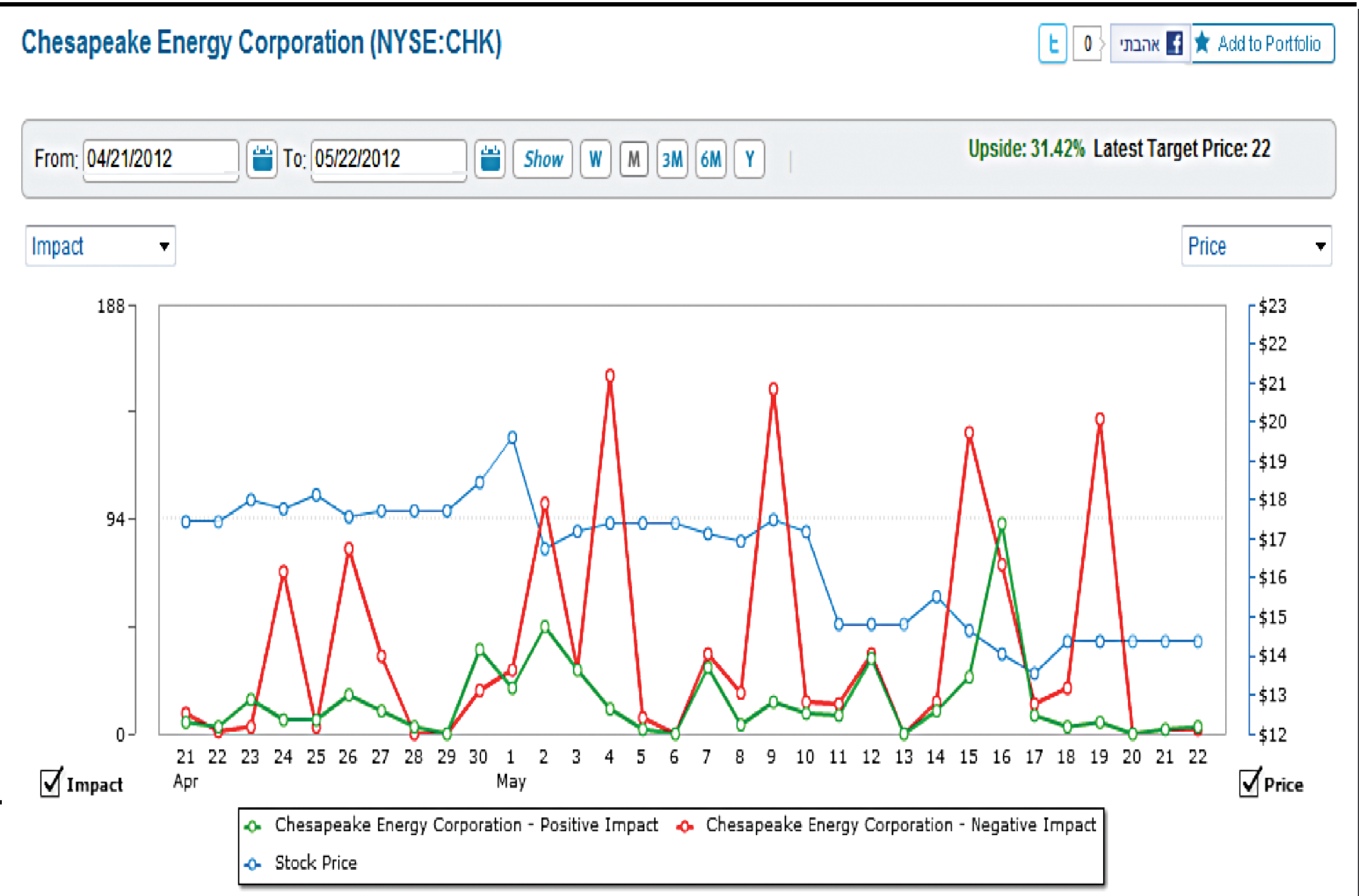
Fig. 2. Tracking public mood states from tweets posted between October 2008 to December 2008 shows public responses to presidential election and thanksgiving.

StockTwits



http://stocktwits.com/symbol/AAPL?q=AAPL
$AAPL Apple Inc. stock and...
StockTwits
STREAMS
INBOX
HEAT MAP
SYMBOL($AAPL) or USERNAME
edmundyu
TRENDING $FB $MU $KRFT $IBB $SPY $RHT $CLF $SONC $PNR $JLL $APOL $RCII $PVH $BHE $LULU $SWKS
SPONSORED BY: Ameritrade
Dow 241.56 S&P 500 23.98 Nasdaq 100.95
Please verify your email in order to share on StockTwits. If you have not yet received an email, we can resend the verification email or you can change your address.
Follow 33,115 FOLLOWERS
WELCOME TO STOCKTWITS
Follow these Accounts
StockTwits
StockTwits Offical Account
Chartly
StockTwits Charts
StockTwitsHelp
StockTwits Help Desk
Learn From Great Traders
Suggested Stream
StockTwits top contributors
RECENTLY VIEWED
$AAPL 123.69
Apple Inc. 3.00 (2.37%)
TradeStation
Apple Inc.
AAPL
NASDAQ
123.69
3.00 (2.37%)
LAST UPDATED 3:44 PM EDT MAR 25
$6.95
TRADE $AAPL NOW
PRICE
MESSAGE VOLUME
SENTIMENT
81% BULLISH 19% BEARISH
$AAPL
Jan 23 Feb 17 Mar 14
How is sentiment calculated? Show more
Share an idea on $AAPL
Upload Img
TOP ALL CHARTS LINKS
Pause Pop out
footballmccoy123
Mar. 25 at 3:43 PM
$AAPL marker should. Be bought up here Bullish
via StockTwits Mobile
mk34
Mar. 25 at 3:43 PM
"@bmc2012: $AAPL ....THICK HEADED AAPL LUVERS SCORN ME..SOO BLOCK ME & BUY-YOULL GET WHAT U DESERVE" OK you're blocked!
MurrayCohen516
Mar. 25 at 3:43 PM
$AAPL starts to go up and you know it will be lower in a minute.. just one of those ugly days. Going to keep dropping until earnings.
ridethewave1966
Mar. 25 at 3:42 PM
$AAPL EVERYTHING is selling off hard Bullish
Get special access to select Schwab trading resources with Guest Pass.
charles SCHWAB
Schwab Trading Services
Get Free Guest Pass ›

Sentiment Graph (http://www.thestocksonar.com).



The Stock Sonar
See through the current

The Stock Sonar - Sentiment Service of US Stocks

Data

Published by:	Digital Trowel Inc.
Categories:	Capital Markets, Business and Finance
Date added:	10/10/2011

Get support for this offering

The Stock Sonar Sentiment Service provides sentiment scores for public companies trades on the US stock market. The Stock Sonar retrieves, reads and analyzes information from a wide variety of online sources including articles, blogs, press releases and other publicly available information based on an in-depth understanding of a text's meaning. The platform's advanced capabilities enable intelligence discerning of positive and negative nuances and events quantifying them to provide investors with immediate insights. The Stock Sonar Sentiment Service presents the following data for each public company: Daily Sentiment Score, Daily Positive Score, Daily Negative Score. The Daily Sentiment Score is a figure between -1 and 1 (where -1 is most negative and 1 is most positive). The Daily Sentiment Score is a weighted average of that day's news articles sentiments. The Daily Positive Score is the weighted sum of all

Transactions/month	Price
500	$0.00 per month
1,000	$30.00 per month
5,000	$129.00 per month
25,000	$599.00 per month
50,000	$999.00 per month

Service documentation
Service client developer guidelines

RESOURCES

Microsoft PowerPivot For Excel 2010
Learn how to use Microsoft PowerPivot for Excel 2010, with this and other DataMarket data, to create compelling self-service BI solutions.

mailto:support@digitaltrowel.com

The Stock Sonar — Sentiment Analysis of Stocks Based on a Hybrid Approach

[1,2]**Ronen Feldman**, [2]**Benjamin Rosenfeld**, [2]**Roy Bar-Haim** and [2]**Moshe Fresko**
[1]School of Business Administration, The Hebrew University of Jerusalem, Jerusalem, ISRAEL
[2]Digital Trowel, Airport City, ISRAEL
ronen.feldman@huji.ac.il,{grur,roy,moshe}@digitaltrowel.com

Abstract

The Stock Sonar (TSS) is a stock sentiment analysis application based on a novel hybrid approach. While previous work focused on document level sentiment classification, or extracted only generic sentiment at the phrase level, TSS integrates sentiment dictionaries, phrase-level compositional patterns, and predicate-level semantic events. TSS generates precise in-text sentiment tagging as well as sentiment-oriented event summaries for a given stock, which are also aggregated into sentiment scores. Hence, TSS allows investors to get the essence of thousands of articles every day and may help them to make timely, informed trading decisions. The extracted sentiment is also shown to improve the accuracy of an existing document-level sentiment classifier.

1 Introduction

trast, TSS provides precise *sentiment extraction*: highlighting of positive and negative expressions within the article text, as well as extraction of positive and negative business events. Extracted sentiment provides the user an *explanation* for the article score as well as an effective *summary* of multiple news articles. As we show in this paper, the sentiment extracted by TSS can also be used to improve document-level sentiment classifiers.

Another limitation of previous methods is concerned with the level of linguistic analysis required to correctly predict sentiment. Current systems usually employ sentiment lexicons and machine-learning algorithms that operate at the word or phrase level. Such methods typically fail to model compositional expressions, e.g. correctly classifying *"reducing losses"* as positive, but *"reducing forecasts"* as negative. Furthermore, it is often necessary to go beyond the

2 Architecture

TSS collects thousands of articles from thousands of sources every day. The articles are collected via stock-specific RSS feeds, so that each article is associated with one or more stocks. The same news is often repeated by multiple sources. Currently we do not attempt to identify these duplicates, except for the obvious case where articles have the same title and date. However, we assume that the number of times a story is repeated is indicative for its significance, and therefore keeping these duplicates is beneficial.

The collected articles are first cleaned so that the main body of the article is maintained and the extraneous content (such as ads, links to other stories, etc.) is deleted. The module in charge of the extraction of the main textual content from the HTML pages is based on a supervised machine learning approach and a visual training module as described in (Rosenfeld, Feldman, and Ungar 2008). The output of this module is plain text. Each article is analyzed separately for each of its associated stocks. We shall refer to the stock for which the article is currently analyzed as the *main company*. Based on the extracted sentiment, an article score is computed, and article scores are then aggregated into daily scores for each ticker. The rest of this section details sentiment extraction and scoring.

Sentiments		
	Positive	**Negative**
Adjectives	attractive, superior	inefficient, risky
Verbs	invents, advancing	failed, lost
Nouns	opportunity, success	weakness, crisis
Multi-word expressions	exceeding expectations, falling into place	chapter 11,pull back
Neutral expressions	in the worst case, best practice	
Sentiments Modifiers		
Emphasis	Huge, incredible, highly	
De-emphasis	mostly, quite	
Reversal	far from, cut, no	

Table 1: Lexicon components

2.1 The CARE Extraction Platform

evaluating the sentiment of each token. Furthermore, because the three components operate within the CARE framework, they can utilize its CRF classifier, which flexibly connects them to the state of the art NER and POS taggers.

The TSS rulebook was developed by a team of three linguistic engineers, assisted by two financially-trained domain experts, over a period of five months. We now turn to a short description of its individual components.

Category	Example (polarity)
Legal	ArvinMeritor (ARM), a maker of integrated systems, rose after it won an antitrust suit against electrical power gear maker **Eaton (ETN)**. ⊖
Analyst Recommendation	On June 23, Caris & Co. reiterated its "buy" rating on **CRM** and increased its price target to $115 from $100. ⊕
Financial	**Western Digital** earnings climb 35% ⊕
Stock Price Change	**SandRidge Energy** fell 3.0 percent to $6.24. ⊖
Deals	The U.S. Army's Mission Installation Contracting Command has awarded **Northrop Grumman Corporation (NYSE:NOC)** a contract to provide logistics support at Fort Eustis, Va. ⊕
Mergers and Acquisitions	On Monday, Ramius LLC has offered to acquire all the outstanding shares of **CYPB** for $4.00 per share in cash. ⊕
Partnerships	Stennis has partnered with **Orbital Sciences Corporation** to test the AJ 26 engines that will power the first stage of the company's Taurus II space launch vehicle. ⊕
Product	**The fast-food giant** recalled 12 mil drinking glasses that contain cadmium ⊖
Employment	**Heidrick & Struggles** Appoints New Hedge Fund Leadership Team. ⊕

Table 2: Types of extracted events. Main company is marked in bold. ⊕/⊖ represent positive/negative polarity.

nounced today its new treatment for multiple sclerosis was found effective in clinical trials". It defines in (1) a positive product event. The primary "anchors" these rules utilize are *"treatment"* nouns (5), followed by a positive adjective (6).

2.5 Sentiment Relevance

When analyzing sentiment for a given company, it is crucial to assess that the sentiment indeed refers to that company. O'Hare et al. (2009) suggested to consider only a window of *N* words around each mention of the main company in the article, and showed that it improves polarity prediction as compared to considering the whole article for the company. Our experiments on a training corpus confirmed that identification of relevant sections is crucial to obtain reasonable precision, and that the distance from a mention of the main company is a good predictor of relevance. However, we found two additional cues for relevant that were not considered by O'Hare at el.:

1. *Directionality* - sentiments that appear after the main company are more likely to be relevant than sentiments preceding it.

2. *Other entities* - entities that appear between the main company mention and the sentiment are good indicators for irrelevance.

Eventually, we implemented a relevance strategy that considered only sentiments that appear after the main company

Representative Applications

- Hong and Skiena (2010), **The Wisdom of Bookies? Sentiment Analysis vs. the NFL Point Spread**: tracked the relationships between the NFL betting line and public opinions in blogs and Twitter
  - The American Football betting market provides a particularly attractive domain to study the nexus between public sentiment and the wisdom of crowds.
  - Based on the text data from LiveJournal blogs, RSS blog feeds captured by Spinn3r, Twitter, and traditional news media.
  - A strategy based on their finding, betting roughly **30 games per year,** identified winner roughly **60%** of the time from 2006 to 2009, well beyond what is needed to overcome the bookie's typical commission (53%)

Mohammad (2011), **From Once Upon a Time to Happily Ever After**: tracked emotions in <u>Brothers Grimm fairy tales</u>

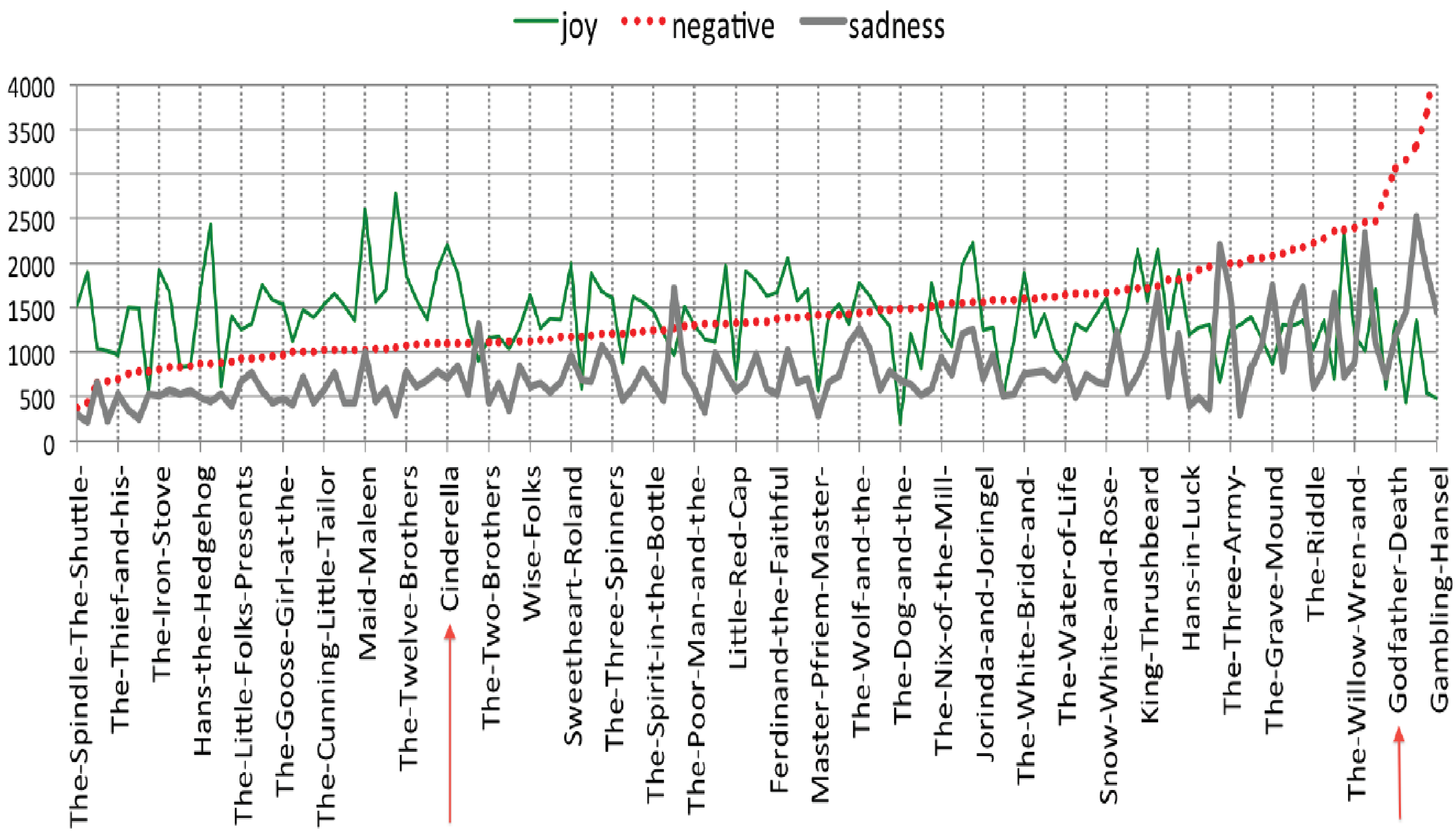


Figure 9: The Brothers Grimm fairy tales arranged in increasing order of negative word density (number of negative words in every 10,000 words). The plot is of 192 stories but the x-axis has labels for only a few due to lack of space. A user may select any two tales, say *Cinderella* and *Godfather Death* (follow arrows), to reveal Figure 10.

Representative Applications

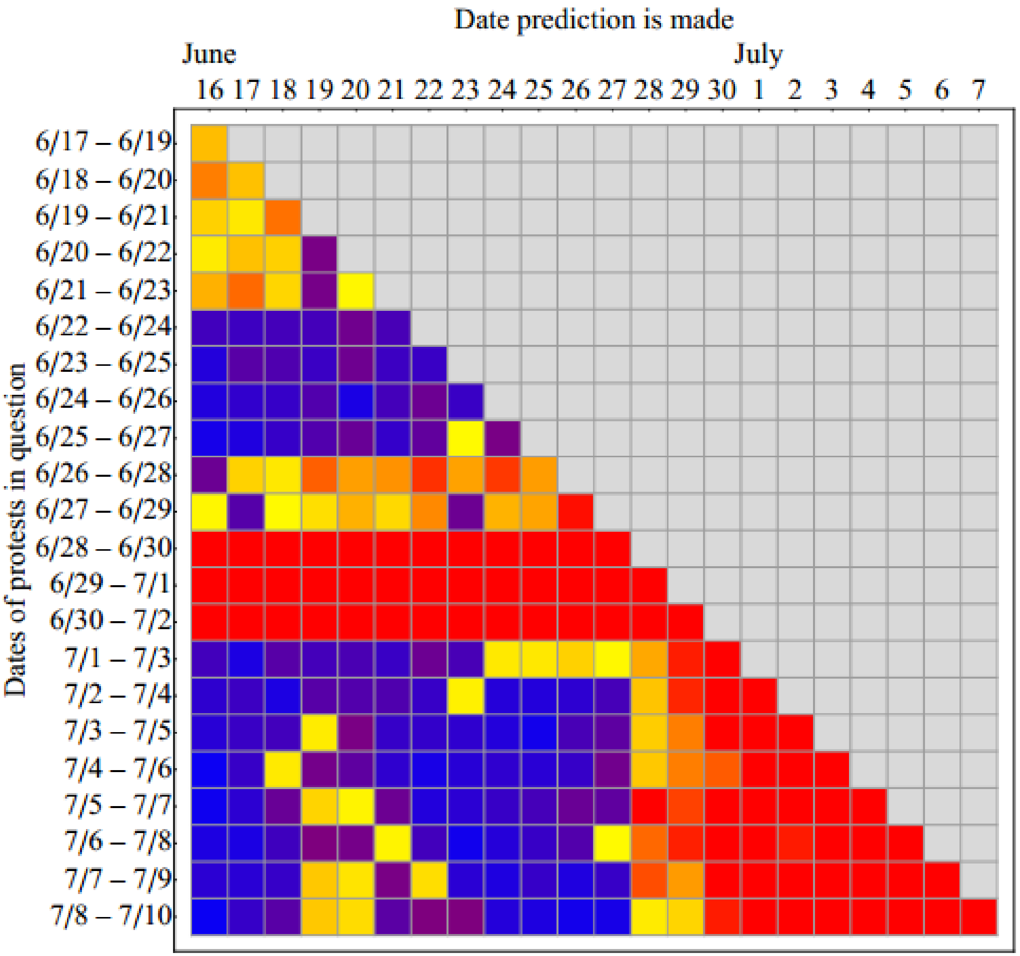
❖ Mohammad and Yang (2011): tracked sentiments in emails see how genders differed on emotional axes. (e.g. women prefer words from the joy–sadness axis, whereas men prefer terms from the fear–trust axis.)

[redacted]

[redacted]

Representative Applications

- Based on public data collected from over 300,000 open content web sources in 7 languages, ranging from mainstream news to government publications to blogs and social media.
- Predictions of protests in Egypt around the time of the coup d'etat.
- Yellow to red mark positive predictions and blue to purple negative, with redder colors indicating more positive <u>votes</u>.



Predicting Crowd Behavior with Big Public Data

Nathan Kallus
Massachusetts Institute of Technology
77 Massachusetts Ave E40-149
Cambridge, MA 02139
kallus@mit.edu

ABSTRACT

With public information becoming widely accessible and shared on today's web, greater insights are possible into crowd actions by citizens and non-state actors such as large protests and cyber activism. We present efforts to predict the occurrence, specific timeframe, and location of such actions before they occur based on public data collected from over 300,000 open content web sources in 7 languages, from all over the world, ranging from mainstream news to government publications to blogs and social media. Using natural language processing, event information is extracted from content such as type of event, what entities are involved and in what role, sentiment and tone, and the occurrence time range of the event discussed. Statements made on Twitter about a future date from the time of posting prove particularly indicative. We consider in particular the case of the 2013 Egyptian coup d'état. The study validates and quantifies the common intuition that data on social media (beyond mainstream news sources) are able to predict major events.

sibility to public information on the web that future crowds may now be reading and reacting to or members of which are now posting on social media can offer glimpses into the formation of this crowd and the action it may take.

News from mainstream sources from all over the world can now be accessed online and about 500 million tweets are posted on Twitter each day with this rate growing steadily [14]. Blogs and online forums have become a common medium for public discourse and many government publications are offered for free online. We here investigate the potential of this publicly available information online for predicting mass actions that are so significant that they garner wide mainstream attention from around the world. Because these are events perpetrated by human actions, they are in a way endogenous to the system, enabling prediction.

But while all this information is in theory public and accessible and could lead to important insights, gathering it all and making sense of it is a formidable task. We here use data collected by Recorded Future (www.recordedfuture.com). Scanning over 300,000 different open content web sources in

3 Levels of Sentiment Analysis

- In general, sentiment analysis has been investigated mainly at three levels:
  - Document Level
  - Sentence Level
  - Entity & Aspect/Feature level

3 Levels of Sentiment Analysis

- **Document level**
  - The task at this level is to classify whether a whole document expresses a positive or negative sentiment.
    - For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product.
  - This level of analysis assumes that each document expresses opinions <u>on a single entity</u> (e.g., a single product)
  - Thus, it is not applicable to documents which evaluate or compare multiple entities.

3 Levels of Sentiment Analysis

- **Sentence level**
  - This level determines whether each sentence expressed a positive, negative, or neutral opinion.
    - Neutral usually means no opinion.
  - Related to subjectivity classification, which distinguishes objective/factual sentences from subjective sentences that express subjective views and opinions.
    - However, many objective sentences can also imply opinions:

      "We bought the car last month and the windshield wiper has fallen off. "
    - Conversely, many subjective sentences may not express any opinion or sentiment:

      "**I think** he went home after lunch."

3 Levels of Sentiment Analysis

- **Entity & Aspect/Feature level**
  - Both the document-level and sentence-level analyses do not discover what exactly people liked and did not like
  - Aspect level directly looks at the opinion itself, instead of documents, paragraphs, sentences, etc.
  - It is based on the idea that an opinion consists of a **sentiment** (positive or negative) and a **target** (of opinion).
  - Realizing the importance of opinion targets also helps us understand the sentiment analysis problem better:

    "Although the service is not that great, I still love this restaurant."

    - This sentence is positive about the restaurant (emphasized), but negative about its service (not emphasized).

3 Levels of Sentiment Analysis

- **Entity & Aspect/Feature level**
  - In many applications, opinion targets are described by entities and/or their different **aspects**:

    "The iPhone's **call quality** is good, but its **battery life** is short."

    - This sentences evaluates two aspects: **call quality** and **battery life**, of iPhone (entity).
    - The sentiment on iPhone's call quality is positive, but the sentiment on its battery life is negative.
    - The call quality and battery life of iPhone are the opinion targets.

2 Types of Opinions

- 2 types of opinions need to be differentiated as well:
  - **Regular opinions**: express sentiments only on an particular entity or an aspect of the entity:

    "Coke tastes very good."

    - This sentence expresses a positive sentiment on the taste aspect of Coke.
  - **Comparative opinions**: compares multiple entities based on some of their shared aspects:

    "Coke tastes better than Pepsi."

    - This sentence compares Coke and Pepsi based on their taste aspect, and expresses a preference for Coke

Sentiment Lexicon

- The most important indicators of sentiments are **sentiment words**, also called opinion words.
  - These are words that are commonly used to express positive or negative sentiments.
    - Good, wonderful, and amazing are positive sentiment words
    - Bad, poor, and terrible are negative sentiment words.
  - Apart from individual words, there are also phrases and idioms:

    "It cost me an arm and a leg."
  - A list of such words and phrases is called a **sentiment lexicon** (or opinion lexicon).
  - Over the years, researchers have designed numerous algorithms to compile such lexicons. (**Sentiment Analysis and Opinion Mining**, Chapter 6; **Sentiment Analysis**, Chapter 7).

Sentiment Lexicon

- Although sentiment lexicons are important or even necessary for sentiment analysis, they are not sufficient:
  - A positive or negative sentiment word may have opposite orientations in different application domains:

    "This camera **sucks**."

    "This vacuum cleaner really **sucks**."
  - A sentence containing sentiment words may not express any sentiment:

    "Can you tell me which Sony camera is **good** ?"

    "If I can find a **good** camera in this shop, I will buy it."

    - Both these sentences contain the sentiment word "good," but neither expresses a positive or negative opinion on any specific camera.

Sarcastic Sentences

- **Sarcastic sentences** with or without sentiment words are hard to deal with:

  "What a great car! It stopped working in two days."

  - Sarcasms are not so common in consumer reviews about products and services, but are very common in political discussions, which make political opinions hard to deal with.

- Many sentences without sentiment words can also imply opinions:

  "This washer uses a lot of water."

  - This sentence implies a negative sentiment about the washer since it uses a lot of resource (water).

  "After sleeping on the mattress for two days, a valley has formed in the middle."

  - This sentence expresses a negative opinion about the mattress.

Sentiment Analysis & NLP

- Sentiment analysis touches every aspect of **NLP**, such as coreference resolution, negation handling, and word sense disambiguation, which are **not** solved problems in NLP.
  - However, sentiment analysis is a highly restricted NLP problem because the system does not need to fully understand the semantics of each sentence or document but only needs to understand some aspects of it, i.e., positive or negative sentiments and their target entities.
  - We will use NLTK to serve our practical NLP needs.