Echolytics
Social Media Analytics

Saghar Iranzadeh Tamaddon & Chloe Reynolds
Overview

- Project
- Results
- Future Work
Project

Saghar Iranzadeh Tamaddon & Chloe Reynolds
Goal

- Using social computing and sentiment analysis techniques to retrieve, organize, and analyze people's judgments about social and environmental activities of companies according to their tweets.

- Using the outcome of such effort to rate businesses according to people's perception rather than objective criteria.
What is CSR and why is it important?
Data Source

- Twitter (Phase 1)
  - Jan 2012 – Feb 2012 *
  - 6 companies in 2 industries

- Facebook

- Blogs (news, experts, discussions, etc.)

Saghar Iranzadeh Tamaddon & Chloe Reynolds
Computing “Goodness” from Social Media Data

Methodology

Needs Requirements
- Interviews
- Comparative Analysis
- Existing Ratings

Context Definition
- Training sets must be industry and company-specific

Data Preparation
- Data Collection
- Data Cleansing
- Data Organization

Natural Language Processing
- Training-Set Creation through Crowdsourcing and manual verification
- Python NTK

Data Analysis
- Callouts of interesting findings

Data Visualization

Scoring
- Ranking

Reporting
- Scoring
- Ranking
- Platform Creation
- Data Visualization

Saghar Iranzadeh Tamaddon & Chloe Reynolds
Steps Taken

- Data collection and meta-data appending
- Data cleansing and organization
- Define CSR (by crowdsourcing)
- Define contexts (industry and company-specific, sentiment – by crowdsourcing)
- Training Set Creation for NLP Classifier
Steps Taken

- NLP classification on tweets
  - Refers to company or not (yes/no/undetermined)
  - Sentiment (+/-/0/undetermined/mixed)
  - Relates to CSR (yes/no/undetermined)

- NLP to identify Top Terms (words and bigrams, POS/chunking) and Top Hashtags

- Manual review of data

- Visualizations and website creation
Classifying Tweets

Tweets:

Bout to be so fat #panera

@laurroggg You're on drugs if you think #panera makes you feel fat!

I love Chipotle more than McDonald’s

Rate that tweet!

- Refers to company?
- CSR related?
- Sentiment?

Saghar Iranzadeh Tamaddon & Chloe Reynolds
Tweet:
I have no more chipotle chips :( 

Rate that tweet!
- Refers to company?
- CSR related?
- Sentiment?
Tweet:

Clorox makes me ITCH

Rate that tweet!

- Refers to company?
- CSR related?
- Sentiment?
Tweet:

I would eat elsewhere but #Panera lists calorie counts #healthiereating

Rate that tweet!

- Refers to company?
- Is CSR related?
- Sentiment?
Tweet:

RT @HyperBrendan: RT @abcnews: Jobs go as iconic brands Mortein & Dettol stop being made in Australia from July http://t.co/y1Xbj8v5 | | Libs #blamecarbontax

Rate that tweet!

- Refers to company?
- CSR related?
- Sentiment?
Tweet:

Maybe if Panera wasn't so damn delicious, I would 1) have more money and 2) consume half as many carbs. 
#paneraiisafoodgroup

Rate that tweet!

- Refers to company?
- Is CSR related?
- Sentiment?
Tweet:

So tired just got done colorix ing the bathroom and I put Lysol together well that made a bomb and I I can't breath my lungs r clogged lol

Rate that tweet!

- Refers to company?
- Is CSR related?
- Sentiment?
Rate That Tweet!

Tweet:

McDonald's to Stop Using 'Pink Slime' in Hamburgers in US.

Rate that tweet!

- Refers to company?
- Is CSR related?
- Sentiment?
Results
Echolytics analyzes tweets, eliminates noise, and ranks companies according to the public’s perception of them.

### Top Do-Good Companies

<table>
<thead>
<tr>
<th>Rank</th>
<th>Company Name</th>
<th>Goodness</th>
<th>CSR Buzz</th>
<th>Details</th>
<th>Echo Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Panera</td>
<td>4</td>
<td>4</td>
<td></td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>McDonalds</td>
<td>4</td>
<td>4</td>
<td></td>
<td>87</td>
</tr>
<tr>
<td>3</td>
<td>Chipotle</td>
<td>4</td>
<td>4</td>
<td></td>
<td>76</td>
</tr>
</tbody>
</table>

Saghar Iranzadeh Tamaddon & Chloe Reynolds
Imperfect Science

Results: Feasible but not an exact science
Operationalization Choices

- Using Twitter and only Twitter
- Which companies are chosen and which industries affect everyone’s scores (future all Top 100)
- Date range
- Company aliases
Procedures Used

- Data collection – English
- Data cleansing - unicode characters and emoticons
- CSR definition – depends on audience
- Training set must be larger because CSR tweets are only a small percent of all tweets
- Manual oversight of NLP program to identify Top Terms and Top Hashtags
We gathered all tweets contained these companies names (and a few aliases) between 1/1/12 and 2/29/12

Define CSR (by crowdsourcing)

NLP classification on tweets
- Refers to company or not (yes/no/undetermined)
- Sentiment (+/-/0/undetermined/mixed)
- Relates to CSR (yes/no/undetermined)
Future Work
Post-Experiment Steps

- Needs Assessment
- Usability Testing
- Further automation of automatable pieces
- Guidelines/tips for human intel procedures
- Training Set supervision
Many other NLP features would be possible

- more sophisticated topics identification, topic comparisons and tracking, trends, “viral”ity, amount and speed of spread
- user analyses
  - Geography
  - indiv vs company
  - Influencers
  - speech patterns-user characteristics mapping
- separate retweets from non-retweets for all reports
Identifying Tweets about environmental or social responsibility:
Among Tweets that contain certain company names, does also containing word "cancer" predict (at least 90% of the time) that the Tweet is about the company's environmental or social responsibility?

![Graph showing the percentage of tweets containing the word "cancer" for different companies.](image-url)
Audience

- Investors (selected for Phase 1)
- Consumers
- Consumer Influencers
- Watchdog Entities
- Corporations
Questions?