Unlocking Unstructured Social Media Data in Marketing

William Rand
Assistant Professor of Business Management

In Collaboration with
Kelly Hewett, Roland Rust, and Harald J. van Heerde

NC STATE
Poole College of Management
Bank of America Will Steal Your Parrot And Then Make *You* Look Like The Crazy One

By BESS LEVIN

In its storied history, Bank of America has foreclosed on a few houses. You expect them to foreclose on a few houses. You expect them to foreclose on a few houses if people don’t pay their mortgage. But it’s really something when you hold the contract on someone’s home and then decide that the owner no longer deserves to have the keys to their own home. Even more so when the owner is left with no place to go, and nothing to live on after losing their home.

That is precisely what happened when Bank of America decided to repossess the home of a woman whose mortgage they held. The problem is that the bank decided to repossess the wrong house, and it took a month of the woman living on the streets because of the error.

Bank of America apologised to a woman after one of its contractors removed a dead body from her home and took her parrot while wrongly repossessing it. The bank said it had “zero tolerance for this kind of error” and said it will

The Telegraph

HOMA NEWS NEWS TOPICS HOW ABOUT THAT?

Bank of America apologises for repossessing parrot

Bank of America apologised to a woman after one of its contractors removed a dead body from her home and took her parrot while wrongly repossessing it. The bank said it had “zero tolerance for this kind of error” and said it will

4:37 AM GMT 10 Mar 2010

Forty-six-year-old Angela Iannelli sued the bank on Monday.

She claims her mortgage was up-to-date when one of the banking giant’s contractors damaged furniture, took her pet parrot, Luke, and padlocked her door in October.

In a statement, the bank said it “sincerely apologises” and has tried for months to resolve the issue.

The bank said it has “zero tolerance for this kind of error” and said it will
Managers’ perspectives

“Media is an *echo chamber* . . . you have *advertising* that is perceived one way, *news* perceived another way, and *social media* perceived yet another way and they are different durations.”

Marketing executive in a Fortune 50 FSI
The “echoverse”

Firm Actions

Social Media Word of Mouth

Business Performance

Consumer Sentiment

News Media

$
Industry setting: Top 4 US financial services firms

June 2007 – December 2013
Data

- Firm Communications:
  - All press releases volume and valence
  - All corporate Tweets (volume and valence)
  - Ad spending from Ad$pender (Kantar)
- Traditional Media News Stories (volume and valence)
- Consumer Sentiment – YouGov’s BrandIndex
- Social Media WOM – Tweets about firms (volume and valence)
- Firm Performance – consumer deposits
Textual data overview

• Size of Collections
  ▫ Press releases (n = 5,376)
    ● All of the press releases from the 4 banks over the time period
  ▫ News Articles (n = 65,261)
    ● Provided by Factiva
    ● Drawn from 14 of the largest circulation newspapers in the US
  ▫ Tweets (n = 18,577,733)
    ● Gathered using Gnip queries regarding the 4 banks

• Coding of Data
  ▫ Manual Coding is impossible
    ● Volume and Variety of Data
  ▫ Data coding must be done automatically using state of the art computational linguistics
Automated data extraction

- Conversion of all data from either Microsoft Word documents or PDFs to plain text, and preprocessing of Twitter data
- Data extracted from each document:
  - Date
  - Associated bank(s) - source of the press release, banks mentioned in articles and Tweets
  - Sentiment
- Aggregated data at the week level for each dataset
Twitter data collection

• Both from the bank and about the bank
• Used 72 different Power Track Searches from Gnip
  ▫ 27 BofA, 22 Chase, 15 Wells Fargo, 18 Citibank
• Examples
  ▫ Different Bank Twitter Accounts and Mentions of those Accounts
    ● @BofA_News, @BofA_Help, @BofA_Careers, @BofA_Community, @BofA_Tips,
      @bankofamerica, @MerrillLynch
  ▫ References to the Bank
    ● BAC and (bank or banking), BOA and (bank or banking), “Bank of America”, BofA
  ▫ References to Executives
    ● Ken Lewis and (CEO or BAC or BOA), Moynihan (Brian or CEO or BAC or BOA)
  ▫ References to Products
    ● (BAC or BOA) and loan, (BAC or BOA) and ATM, (BAC or BOA) and mortgage
• All terms were chosen to minimize the probability of false positives but to gather as much as possible (maximizing recall and precision)
Press release and news article sentiment analysis

• Based on LIWC (Linguistic Inquiry and Word Count; Pennebaker et al. 2007)

• Contains a Dictionary of Positive and Negative Emotion Terms, including stems, e.g., ugly and ugliness map to ugl*

• Each target word in a document is examined and see if it matches any of the terms in the dictionaries

• Outputs a percentage of positive and negative words and a word count
Twitter sentiment analysis

• Long Form Sentiment Tools, such as LIWC, do not work for Tweets
  ▫ Misspellings, e.g., “happniess”, “citybank”
  ▫ Acronyms, e.g., LULZ, IMHO
  ▫ Dropped characters, e.g., “hppy”, “

• We used a 10-fold, cross-validated, Naïve Bayes Classifier (Go, Bhayani, and Huang, 2009)
  ▫ Trained on tweets that contain emoticons, e.g., :) and :( 
  ▫ Tweets are preprocessed
    ● Removed whitespaces
    ● Added a feature for all CAPS
    ● Lowercase everything
    ● Remove punctuation at beginning and end of words, but keep emoticons
    ● Reduce repeated letters, e.g., sooooo becomes so
    ● Remove common stop words
  ▫ In the end wind up with ~8000 features, e.g., common words, timestamps, links
  ▫ Output is -1 to 1, measuring how positive or negative the tweet is
Creation of weekly data

- **Twitter Data**
  - Valence: Sum of the sentiment scores associated with tweets by bank by week
  - Volume: Count of the number of tweets by bank by week
  - Separate out Tweets from the Bank vs. Tweets by others

- **Press Releases and News Articles**
  - Valence: Sum of percentage of positive words minus the sum of percentage of negative words by bank by week
  - Volume: Count of the number of documents by bank by week
Consumer Sentiment

• Data from BrandIndex database (YouGov)

• Weekly measure of YouGov’s overall buzz metric:
  ▫ Which of the following banks are in the news?
  ▫ Are they in the news in a positive or a negative way?
  ▫ Average across thousands of responses
What we observed

The echoverse at work

The case of

Bank of America

Consumer sentiment

Twitter volume

News article volume

June 2007 – December 2013
The brand echoverse

Firm Communications:
- Press Releases
- Advertising
- Twitter posts

Online Word of Mouth:
- Twitter posts

Traditional Media News Stories

Consumer Sentiment
- BrandIndex

Business Outcome:
- Customer deposits

Online Word of Mouth:
- Twitter posts

Consumer Sentiment
- BrandIndex
Statistical Method

- Echoverse conceptualization says all variables are endogenous
- Time lags exist - 1 lag is optimal
- Appropriate methodology is VAR Model: Vector AutoRegressive Model

- Calculate Generalized Impulse Response Functions:
  - Shock one variable by 1 standard deviation
  - Simulate effect on all variables in system
The Estimation Equations

\[
\begin{pmatrix}
\text{News Article Volume} \\
\text{New Article Valence} \\
\text{Consumer Sentiment} \\
\text{Public Twitter Volume} \\
\text{Public Twitter Valence} \\
\text{Company Twitter Volume} \\
\text{Company Twitter Valence} \\
\text{Press Release Volume} \\
\text{Press Release Valence} \\
\text{Advertising Spend} \\
\Delta \text{Customer Deposits}
\end{pmatrix}_{\text{bank } b, \text{week } t} = \sum_{l=1}^{L} B_{bl} 
\begin{pmatrix}
\text{News Article Volume} \\
\text{New Article Valence} \\
\text{Consumer Sentiment} \\
\text{Public Twitter Volume} \\
\text{Public Twitter Valence} \\
\text{Company Twitter Volume} \\
\text{Company Twitter Valence} \\
\text{Press Release Volume} \\
\text{Press Release Valence} \\
\text{Advertising Spend} \\
\Delta \text{Customer Deposits}
\end{pmatrix}_{\text{bank } b, \text{week } t-l} + \text{Controls } + e_{bt}
\]

Controls: Fixed effects for Banks, Competitor Actions, Consumer Confidence, Unemployment Rate; Interest Rate, Time trend
First, the bad news . . . .

1) Bad news spreads fast and wide

2) Online word of mouth hurts firm performance (customer deposits)
1) Bad news spreads fast and wide.

A. More negative news articles lead to more public tweets.

B. More negative public tweets lead to more news articles.
C. More negative consumer sentiment

&

D. More negative company tweets

More news articles
2) Online word of mouth hurts firm performance.

What can companies do?
The good news . . .
A) Press releases are surprisingly effective.

More press releases

More/more positive news articles

More/more positive word of mouth

Higher customer deposits
BUT . . . press release strategy matters.

Let’s lay low and watch.

Let’s keep our name out there

BAD IDEA!

GOOD IDEA!
AND . . . press release language matters.

Categories of words used in press releases:
Differences in press release language emerged.

Tendency to **react** to negativity in the media

**BAD IDEA!**

Tendency to **ignore** negativity in the media

**GOOD IDEA!**
B) Company tweets can calm things down.

More company tweets

Fewer news articles
**BUT . . .** Twitter strategy matters.

- Greater volume
- More personalized
- Customer service-focused; more neutral

**GOOD IDEA!**

- Low volume
- Mass audience
- Promotional; more positive

**BAD IDEA!**
C) Advertising bypasses the echoverse.

1. No impact on news articles
2. No impact on word of mouth
3. No impact on consumer sentiment

We? Advertising
C) Advertising bypasses the echoverse.

BUT Can still influence customer deposits
Thanks!

Questions?

wmrand@ncsu.edu
@billrand