# Sentiment Analysis of Textual Matter

Paramjit Gill

Joint work with Norbert Eke and Abdallah Mohamed

November 2, 2015

- Airline Customer Reviews

- Text Mining and Processing

- Sentiment Analytics

- Results

## Data Science

- Data, Data everywhere

- Open data initiatives: Federal, State, City governments sharing data online

- Crime, traffic, restaurant inspections, property taxes, demographics etc.

- Social and electronic media: Facebook, Twitter, Blogs

- Data Science competitions: Check out `www.kaggle.com`

- A (random) sample from a population of consumers

- They respond to a bunch of questions

- Expensive and time consuming

- Focus on specific issues useful for decision making

## Unsolicited Consumer Reviews

- Utilize online electronic tools and platforms

- Willing (or disgruntled) customers provide feedback

- These reviews convey the real voices of actual customers

- Caution: Possibility of fake reviews

- Deciphering the text review is a monumental task

- Quantitative sentiment analysis can help

# Airline Customer Reviews

- SKYTRAX is a corporate research advisor group for the air transport industry

- www.airlinequality.com provides rating system for the world airlines

- An open forum to write and share review of any airline based on personal experience

- No access to SKYTRAX's in-house database, we used the public information available

# An Air Canada Customer Review

### Air Canada customer review
**Review by Michael Kroeker (Canada) 21st June 2015**

**10/10**

I recently traveled YWG-LHR and returned home from EDI-YWG with AC Rouge. I must say I was pleasantly surprised with how well the trip went in both directions. Outbound flight was delayed about 2 hours due to maintenance, staff kept us updated and were very pleasant. After the aircraft couldn't be fixed they quickly changed to a spare Boeing 777 in YYZ and got us going as quick as they could. Meals on board were not bad, only suggestion would be the continental breakfast could be improved. Seats were comfortable and service was good even though a full flight. Returned on Air Canada Rouge in Premium, seat was adequate, food was good but flight attendants need more training and need to act more professional. Priority bag was first off the carousel and connection was perfect, flights on time no complaints.

| | |
|---|---|
| Aircraft | Boeing 777 |
| Type Of Traveller | Solo Leisure |
| Cabin Flown | Economy |
| Route | YWG-LHR via YYZ |
| Seat Comfort | ✪✪✪✪⊗ |
| Cabin Staff Service | ✪✪✪✪⊗ |
| Food & Beverages | ✪✪✪✪⊗ |
| Inflight Entertainment | ✪✪✪✪⊗ |
| Ground Service | ✪✪✪✪✪ |
| Wifi & Connectivity | ✪✪✪✪⊗ |
| Value For Money | ✪✪✪✪✪ |
| Recommended | ✓ |

# Another Customer Review



**3/10**

## "seat was very uncomfortable"
### W Pollitt (United States) 16th October 2015

Lots of nagging problems. Late Air Canada departure (40 min) from Toronto to Shanghai, only compounded by an even later arrival (85 min) to Shanghai. Paid for an exit seat that was very uncomfortable. Cushion material was bunched and broken down. One of the most uncomfortable seats I have been in. Food is terrible, some not edible. Return flight was better, but still would not recommend. Staff service was excellent.

| | |
|---|---|
| **Aircraft** | Boeing 777 |
| **Type Of Traveller** | Solo Leisure |
| **Cabin Flown** | Economy |
| **Route** | MSP to PVG via YYZ |
| **Date Flown** | September 2015 |
| **Seat Comfort** | ★ ☆ ☆ ☆ ☆ |
| **Cabin Staff Service** | ★ ★ ★ ★ ★ |
| **Food & Beverages** | ★ ☆ ☆ ☆ ☆ |
| **Inflight Entertainment** | ★ ★ ☆ ☆ ☆ |
| **Ground Service** | ★ ★ ★ ★ ★ |
| **Value For Money** | ★ ★ ☆ ☆ ☆ |
| **Recommended** | ✘ |

## Data Mining Challenges

- Automation of web document search

- Mixture of text, numerical and images data

- Example: Convert four star symbols to "4"

- Storage of individual reviews

- One airline at a time

- Data on more than 500 airlines around the world

# Data Mining

Used Java script to access and extract each airline reviews

1. Date of the review
2. Name of the passenger
3. Cabin flown (Economy, Business, First Class)
4. Name of the airline
5. Country of residency of the passenger
6. The rating score (from 1 to 10)
7. Overall recommendation (Yes or No)
8. The review text

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing (NLP), text analysis and computational linguistics to identify and extract subjective information in source materials. (wikipedia.org)

## Methodology

- Polarity of text material as positive, negative or neutral

- Can further be analysed as joy, surprise, happy, sad, fear etc.

- A document assigned a sentiment score based on polarity of sentences

- Different documents can be compared quantitatively

# Text Processing

- OpenNLP: an open source C toolbox to parse the textual matter (Northedge, 2006)

- The Apache OpenNLP library is a machine learning based toolkit for the processing of natural language

- Natural language processing tasks: tokenization, sentence segmentation, part-of-speech tagging etc.

- Break text material into sentences

- Tokenization: break a sentence into words, phrases, or other meaningful tokens

- In total 261,106 sentences of reviews from 32,412 customers - a review is 8 sentences long on average

## Opinion Mining

- The "keywords" or "bag-of-words" approach

- Detecting and counting words (uni-grams) labelled as "positive" or "negative"

- However, single words are not sufficient to extract the sentiments expressed in a sentence

- Example: "Meals on board were not bad"

- Modifiers: words, phrases, or clauses function as adjectives or adverbs to describe a word or make its meaning more specific

- Use of "bi-grams" substantially enhanced the sentiment extraction efforts

# Sentiment Vocabularies

- Affective Norms for English Words (ANEW)
- SenticNet (http://sentic.net/)
- SentiWordNet (http://sentiwordnet.isti.cnr.it/)
- WordNet-Affect (http://wndomains.fbk.eu/)
- Opinion Lexicon List (Hu and Liu, 2004)

## Bag of Words

- General purpose 5,000 Negative words and 2,100 Positive words (Hu and Liu, 2004)

- Double negative bigrams created by adding "not", "no", "didn't", "don't" etc. in front of negative words

- About 300 special negative bi-grams: "below expectations", "sinking feeling", "cold food", "lost luggage"

- About 200 special positive bi-grams: "look after", "decent deal", "fresh food", "wide selection"

# Accuracy Checks

- Each sentence assigned a sentiment score
  Sentscore = Postive unigrams + Positive bigrams + 2*Double
  negative bigrams - Negative unigrams - Special negative bigrams

- We manually read and sorted 1000 sentences as positive,
  negative and neutral

- Sentence-wise accuracy around 70%

- Sarcastic sentences most difficult
    *By time you buy ticket, then meal, pay extra bag
    etc, you could have purchased economy ticket
    another airline and had all extras free and excellent
    service.*

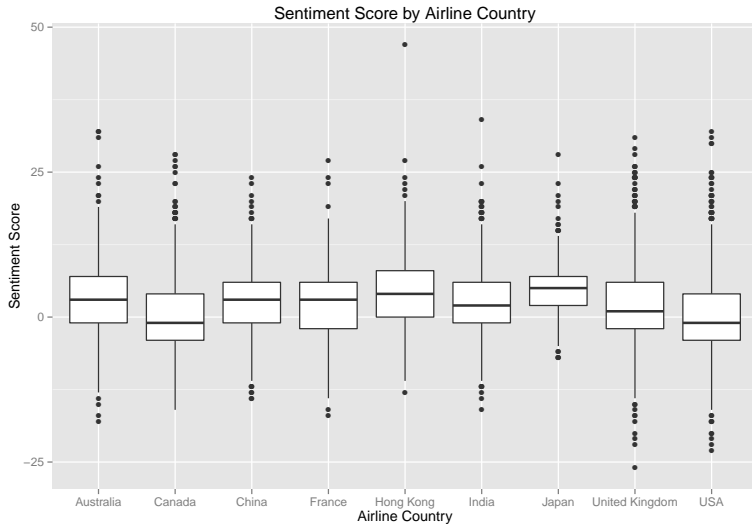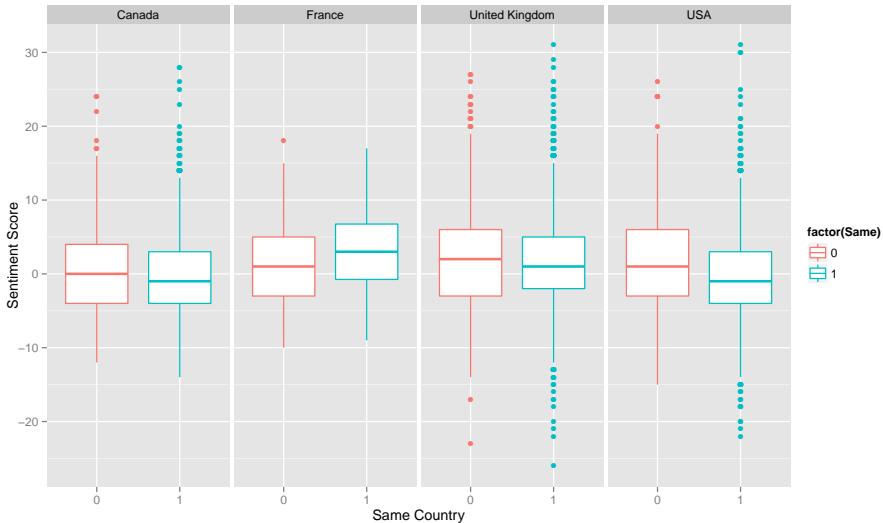- Commercial packages claim 80 to 90% accuracy

Sentiment Score and Overall Rating

Sentiment Score and Recommendation
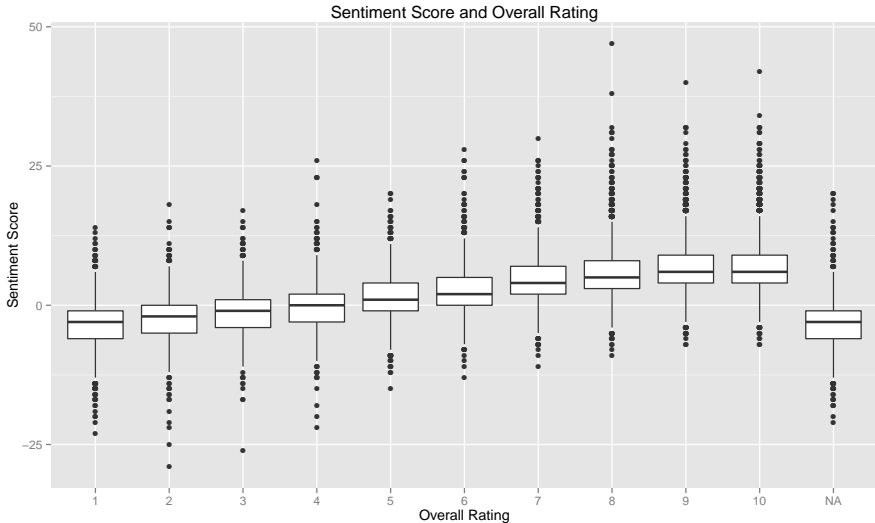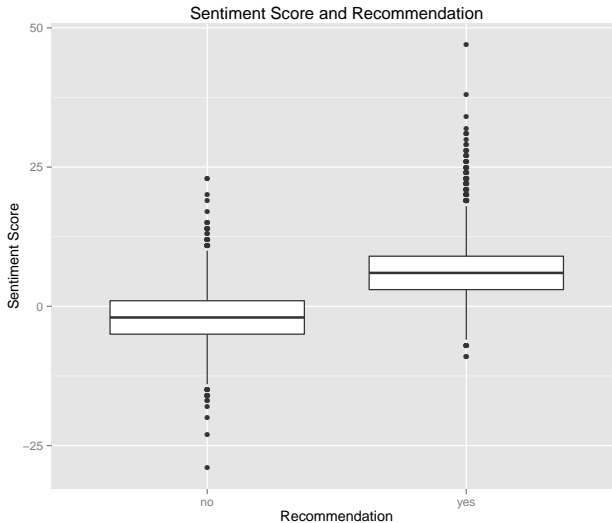
# Analysis Results

# Future Plans

- Need better machine learning algorithms for sentiment extraction

- Temporal trends: How does sentiment change around Christmas rush and long weekends?

- Before and after events such as job action, major announcements such as adding a surcharge

- Compare no-frills (charter) and full service airlines

- Can we detect fake reviews?