Aspect-Based Sentiment Analysis of online reviews

Suhailla Zainudin, Zohreh Madhoushi and Abdul Razak Hamdan
Data Mining and Optimization Research Grup
Center for Artificial Intelligence Technology
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
What is sentiment analysis?

Sentiment analysis and opinion mining is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. (Liu 2012)

The movie was fabulous!
Factual

The movie stars Mr. X

The movie was horrible!
Sentimental
Why sentiment analysis?

- With huge volume of opinionated text, Normal users and organizations have difficulty summarizing opinions.
- This information is unstructured, with lower quality, full of noise and spams and it is not something that is easily machine processable.
- Sentiment analysis is hard and a thriving research area in NLP, ML, data and text mining.
Why online reviews?

Pre Web
- Friends and relatives
- Acquaintances
- Consumer Reports

Post Web
“...I don’t know who...but apparently it’s a good phone. It has good battery life and...”
- Blogs (google blogs, livejournal)
- E-commerce sites (amazon, ebay)
- Review sites (CNET, PC Magazine)
- Discussion forums (forums.craigslist.org, forums.macrumors.com)
- Social Networks (Twitter, Facebook)
Why online reviews? Cont....

- There are too many reviews to read

What is the best digital camera?
Do people like camera $X$? or dislike it?

Reviewers

- Poor picture quality
  - Disappointing battery life

- The batteries are great
  - It is a little expensive

- Lovely picture quality
  - The battery life is OK
Aspect-Based Sentiment Analysis (ABSA)

Google product
ABSA

Two main tasks for ABSA:

- aspect extraction
- aspect polarity estimation/Sentiment rating
Background of study
Word2Vec (uses Deep Learning)

- A framework for learning word vectors (Mikolov 2013)
Doc2Vec (uses Deep Learning)

- A framework for learning paragraph vector (Mikolov 2014)
Deep learning

- Learn feature hierarchies, in which features from higher levels of the hierarchy are formed by the composition of lower level features.
- Automatically learning features at multiple levels of abstraction can be regarded as learning complex functions.
- Example-multilayer neural networks, which consists of several layers of non-linear operators for the composition of inputs from lower layers.
ABSA methods

- Aspect term extraction methods
- Sentiment rating methods
- Language rule mining methods
- Sequential model methods
- Topic model methods
- Lexicon-based methods
- Supervised learning methods

References:

(Hu & Liu 2004)
(Zhuang et al. 2006)
(Du & Tan 2009)
(Moghaddam & Ester 2010)
(Hai et al. 2011)
(Qiu et al. 2011)
(Lai & Asnani 2014)
(Zhang et al. 2014)
(Poria et al. 2014)
(Liu et al. 2015)

(Hu and Liu 2004)
(Ding et al. 2008)
(Moghaddam & Ester 2010)
(Jakob & Gurevych 2010)
(De Albornoz et al. 2011)
(Jiang et al. 2011)

(Choi & Cardie 2010)
(Shariati & Moghaddam 2011)
(Zhao et al. 2015)
(Brody & Elhadad 2010)
(Zhao et al. 2010)
(Moghaddam & Ester 2011)
(Chen et al. 2014)
ABSA methods cont...

- **Sequential models** are not suitable in this study due to their supervised nature which makes them domain dependent.
- **Topic models** are too statistic centric.
- Based on our literature review most of the works in ABSA use **language rule models** for this task.
- In this study we focus on language rule methods.
Language rule models

- Finds frequent nouns/noun phrases
- Finds adjectives in a window of 5-6 words to create aspect – sentiment pair.

• Liu 2004
• Ding 2008
• Liu 2014
• Lek 2013
• Marrie 2014

Example:

Image quality of this camera which I bought it in a reasonable price is excellent

Uses dependency parser to find aspect – pair.

• Zhuang 2006
• Wu 2009
• Qiu 2011
• Lizhen 2014
ABSA challenges

Some of various challenges from the book of liu 2012 that make the problem of ABSA hard:

1. Using different words or phrases to express the same aspect, e.g.,
   • Photo quality is a little better than most of the cameras in this class.
   • That gives the SX40 better image quality, especially in low light, experts say.

2. using different sentiments for expressing the same polarity, e.g.,
   • For a camera of this price, the picture quality is amazing.
   • I am going on a trip to France and wanted something that could take stunning pictures with, but didn’t cost a small fortune.
ABS A challenges… Cont.

3. Reviews include a large amount of irrelevant information.
   - I have owned Canon power shot pocket cameras exclusively over the years.
   - I have fat hands but short fingers.

4. While explicit aspect/sentiment extraction is easy, extracting implicit ones is difficult.
   - This mp3 player is very affordable.
   - I bought this mp3 for almost nothing!
   - After a twenty-one mile bike ride a four mile backpacking river hike, the size, weight, and performance of this camera has been the answer to my needs.
   - The grip and weight make it easy to handle and the mid zoom pictures have exceeded expectation.
Aspect extraction problem
Problem example

- Results are **limited to the window size**.

  *Example:* *Image quality* of this camera which I bought it is in a reasonable price is **excellent**.

- Reviews are **full of irrelevant information**.

  *Example:* “I have fat **hand** and short **fingers**.”
Problem example

• **Explicit aspect**
  
  *Example*: “The *picture quality* of this phone is great”.

• **Implicit aspect**
  
  *Example*: “This car is so expensive.”  →  ‘*price*’
  
  *Example*: “This phone will not easily fit in a pocket”.  →  ‘*size*’
Polarity prediction problem

Problem with current models

- Supervised models
  - Assign one polarity for one sentence
  - Cannot work in different domain/need lots of training data in each domain

- Lexicon-based Models
  - Cannot detect domain dependent orientation of aspects
  - Detect limited number of implicit sentiment
  - Few models detect implicit sentiment

Objective

- Finding more accurate polarity
- Developing a model for different domains
Problem Example

- More than one aspect in a sentence
  Example: The image quality is good but the price is high.

- Domain dependent orientation of opinions
  Example: quite is positive in restaurant domain but negative in MP3 player domain.

- Explicit sentiment
  Example: After a twenty-one mile bike ride a four mile backpacking river hike, the size, weight, and performance of this camera has been the answer to my needs.
Research framework

**Literature review**
- Finding product review datasets (Labeled and unlabeled) & ground truth
- Literature review on SA to find a gap and propose a solution

**Algorithm development**
- Re-implmenting word2Vec using dataset R
- Proposed aspect extraction phase on dataset M and S
- Re-implmenting Doc2Vec using dataset R
- Proposed Polarity prediction phase on dataset M, D and S

**Evaluation**
- Comparing the results with baselines
- Using Precious, Recall, F-measure and Significant as evaluation criteria.
(Our Proposed method) aspect extraction

Aspect extraction \(\rightarrow\) extract all nouns, adjectives, adverbs and verbs as candidate aspect+ uses Word2Vec (mikolov 2013)
(Our Proposed method) sentiment rating

Sentiment rating uses Doc2Vec + (Hu and Liu 2004)

Doc2Vec as unsupervised phase and Logistic regression as supervised phase (mikolov 2014)
Conclusions

The proposed improvements are valuable:

For Consumers:
- Easing the process of decision making when purchase products or services by providing a decomposed view of rated aspects

For producers:
- Source of consumer feedback.
- Benchmark products and services
- Save lots of money they spend to obtain consumer opinions, using survey, focus group and consultants.

For other systems:
- Opinion summarization systems
- Opinion question answering systems
- Recommendation systems (to provide explanations for recommendation)
- Advertising system (to place an ad of a product with similar rated aspects)
- Many business tasks related to sale management, reputation management, and public relations