GlobalLogic®

An Introduction to Sentiment Analysis

Ashish Katrekar
AVP, Big Data Analytics

Sentiment analysis and opinion mining have become an integral part of the product marketing and user experience as both businesses and consumers turn to online resources for feedback on products and services. This white paper explores the evolution and challenges of sentiment anaysis, as well as how to best leverage it.

www.globallogic.com

Table of Contents

Introduction

The opinions of others have a significant influence in our daily decision-making process. These decisions range from buying a product such as a smart phone to making investments to choosing a school – all decisions that affect various aspects of our daily life. Before the Internet, people would seek opinions on products and services from sources such as friends, relatives, or consumer reports.

However, in the Internet era, it is much easier to collect diverse opinions from different people around the world. People look to review sites (e.g., CNET, Epinions.com), e-commerce sites (e.g., Amazon, eBay), online opinion sites (e.g., TripAdvisor, Rotten Tomatoes, Yelp) and social media (e.g., Facebook, Twitter) to get feedback on how a particular product or service may be perceived in the market.

Similarly, organizations use surveys, opinion polls, and social media as a mechanism to obtain feedback on their products and services. sentiment analysis or opinion mining is the computational study of opinions, sentiments, and emotions expressed in text. The use of sentiment analysis is becoming more widely leveraged because the information it yields can result in the monetization of products and services.

For example, by obtaining consumer feedback on a marketing campaign, an organization can measure the campaign's success or learn how to adjust it for greater success. Product feedback is also helpful in building better products, which can have a direct impact on revenue, as well as comparing competitor offerings.

This white paper will describe the various types of sentiment classification, explore how to convert unstructured text into structured opinions, and address the current challenges in the field.

Sentiment Classification Levels

Sentiment analysis can occur at different levels: document level, sentence level or aspect/feature level.

Document Level Classification

In this process, sentiment is extracted from the entire review, and a whole opinion is classified based on the overall sentiment of the opinion holder. The goal is to classify a review as positive, negative, or neutral.

Example
"I bought an iPhone a few days ago. It is such a nice phone, although a little large. The touch screen is cool. The voice quality is clear too. I simply love it!"

Is the review classification positive or negative? Document level classification works best when the document is written by a single person and expresses an opinion/sentiment on a single entity.

Sentence Level Classification

This process usually involves two steps:

- Subjectivity classification of a sentence into one of two classes: objective and subjective
- Sentiment classification of subjective sentences into two classes: positive and negative

An objective sentence presents some factual information, while a subjective sentence expresses personal feelings, views, emotions, or beliefs. Subjective sentence identification can be achieved through different methods such as Naïve Bayesian classification. However, just knowing that sentences have a positive or negative opinion is not sufficient. This is an intermediate step that helps filter out sentences with no opinions and helps determine to an extent if sentiments about entities and their aspects are positive or negative. A subjective sentence may contain multiple opinions and subjective and factual clauses.

Example
"iPhone sales are doing well in this bad economy."

Sentiment classification at both the document and sentence levels are useful, but they do not find what people like or dislike, nor do they identify opinion targets.

Aspect/Feature Level Classification

In this process, the goal is to identify and extract object features that have been commented on by the opinion holder and determine whether the opinion is positive, negative, or neutral. Feature synonyms are grouped, and a feature-based summary of multiple reviews is produced.

Converting Unstructured Text into Structured Opinions

Consumer sentiments are mainly expressed in an unstructured format. Text analysis involves taking unstructured data and finding ways to convert it into a more structured format that facilitates analysis of data and getting deeper insights into the sentiments. There are various techniques that can be used to convert unstructured data into a structured format.

One such way is to express an opinion as a quintuple (e_j, a_{jk}, so_{ijkl}, h_i, t_l), where:

e_j is a target entity
a_{jk} is an aspect/feature of the entity e_j
so_{ijkl} is the sentiment value of the opinion from the opinion holder
so_{ijkl} is +ve, -ve, or neutral, or more granular ratings
h_i is an opinion holder
t_l is the time when the opinion is expressed

An entity *e* is a product, person, event, organization, or topic and is represented as an hierarchy of components, sub-components, and so on. Each node is represented as a component and is associated with a set of attributes for the component.

Example
Review from XYZ on 7/8/2013 - "I purchased a Galaxy 5S phone. It is a great phone overall. The screen resolution is cool and has a good battery life."

This opinion can be expressed by the following quintuples:
(Galaxy 5S, screen, +, XYZ, 7/8/2013)
(Galaxy 5S, battery, +, XYZ, 7/8/2013)

Quintuples form the basis for opinion summarization. The objective of the quintuple is to convert the unstructured data to a more structured form that can be used for further analysis. The initial step is to discover all quintuples and find the five attributes required by the quintuplet. Once the data is in a more structured form, it is much easier to analyze and perform sentiment analysis. Once the quintuples have been extracted, they can be fed to visualization and analysis tools.

Aspect Based Sentimental Analysis (ABSA)

ABSA is based on identifying aspects of given target entities and estimating the sentiment polarity for each mentioned aspect. This can be decomposed into two tasks: aspect extraction and aspect sentiment classification.

Aspect extraction pertains to recognizing aspects of the entity, and more generally can be seen as an information extraction task. Aspect sentiment classification determines whether the opinions on different aspects are positive, negative or neutral.

While lexicon-based approaches use a list of aspect-related sentiment phrases as the main resource, the key issue for learning methods is to determine the scope of each sentiment expression, if it covers the aspect in the sentence.

Aspect Extraction

To identify all the aspect terms present in a sentence, all highly frequent phrases across reviews (e.g. food) should be found and filtered by rules like "occurs right after sentiment word" (e.g. great food). Then a set of phrases that occur frequently can be built. Another approach is to determine all the aspects in advance and find them in the reviews. For a restaurant, the aspects could be: food, service, value, décor.

Example
"The food was great, but the service was slow."
Aspects: food, service

Sentiment Classification

Words express various kinds of sentiments that may be positive, negative, strong, or weak. To perform sentiment analysis, it is important to understand the polarity of words and classify sentiments into categories such as positive, negative, or neutral. This task can be accomplished through the use of sentiment lexicons. There are different types of sentiment lexicons available that have words classified as having positive or negative sentiments.

Examples
The General Inquirer (http://www.wjh.harvard.edu/~inquirer)

LIWC (Linguistic Inquiry and word count) (http://www.liwc.net)

MPQA Subjective Cues Lexicon (http://www.cs.pitt.edu/mpqa/subj_lexicon.html)

Bing Liu's Opinion Lexicon (http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar)

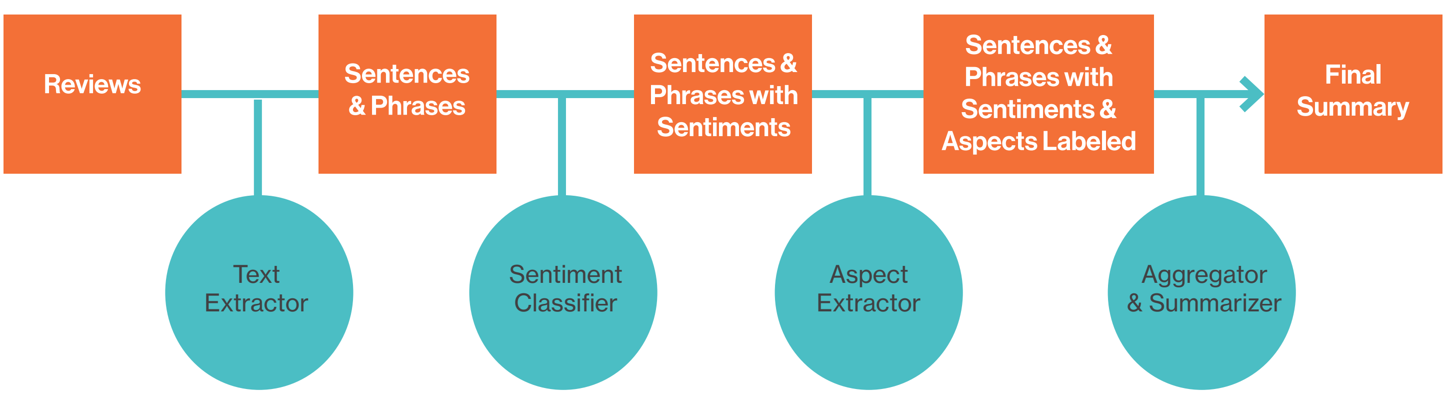
SentiWordNet (http://sentiwordnet.isti.cnr.it/)

The online lexicons may not be sufficient because they may not have enough words or do not pertain to the domain or topic of discussion. In such a case, a new sentiment lexicon can be built. This can be done using techniques such as a semi-supervised lexicon, which leverages a small amount of information (e.g., a few labeled examples or a few hand-built patterns) to bootstrap a complete lexicon through the learning of lexicons. In a bootstrapping approach, a high-precision classifier is first used to identify some subjective and objective sentences. A set of patterns is learned from these identified subjective and objective sentences. The learned patterns are then used to extract more subjective and objective sentences. This process can then be repeated until the desired lexicon is built.

Finding Sentiments for Aspects

For this scenario, let's analyze reviews for a particular restaurant. One customer posted the review, "The food was great, but the service was slow." The figure below shows system components for finding sentiments for aspects.

Fig 1. System Overview: Finding Sentiments for Aspects



Finding sentiments for aspects can be divided into four subtasks:

1. **Text extraction:** extract sentences and phrases from the reviews

2. **Sentiment classification:** run a sentiment classifier on each extracted sentence and phrase to determine if it is positive, negative, or neutral

3. **Aspect extractor:** perform aspect term extraction for the ones that express a sentiment and determine the polarity for each of the aspects identified (e.g., food, service)

4. **Aspect polarity aggregation:** group the sentiments for aspects together and produce a final summary (e.g., food: positive, service: negative)

Challenges in Sentiment Analysis

However, there are still some challenges to overcome before sentiment analysis can become a more perfect tool. For example, human judgment is still far more accurate as a gauge in sentiment analysis. Automated systems cannot differentiate sarcasm from sincere text, nor can they always correctly analyze the specific contextual meaning of a word. Use of acronyms like "lol" or word abbreviations also pose interpretation challenges.

Furthermore, mixed opinions such as "I like the printer quality, but the size is too big," can be difficult to classify, as is identifying genuine feedback in a general comment like, "I surprised my wife on her birthday with this new Samsung phone, and she just freaked out!" It's also unlikely that an automated system could identify biased or fake reviews on a product or service.

Conclusion

Sentiment analysis is an evolving field with a variety of use applications. Although sentiment analysis tasks are challenging due to their natural language processing origins, much progress has been made over the last few years due to the high demand for it. Not only do companies want to know how their products and services are perceived by consumers (and compare to competitors), but consumers want to know the opinions of others before making buying decisions.

The growing need for product insights – and the technical challenges currently facing the field –will keep sentiment analysis and opinion mining relevant for the foreseeable future. Next-generation opinion mining systems need a deeper bind between complete knowledge bases with reasoning methods inspired by human thought and psychology. This will lead to a better understanding of natural language opinions and will more efficiently bridge the gap between unstructured information in the form of human thoughts and structured data that can be analyzed and processed by a machine.

The result: intelligent opinion mining systems capable of handling semantic knowledge, making analogies, continuous learning, and detecting emotions – leading to highly efficient sentiment analysis.

References

1. Sentiment Analysis and Opinion Mining by Bing Liu (http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.html)

2. Sentiment Analysis by Professor Dan Jurafsky (https://web.standford.edu/class/cs124/lec/sentiment.pdf)

3. S. Blair-Goldensohn, Hannan, McDonald, Neylon, Reis and Reynar 2008 – Building a Sentiment Summarizer for Local Service Reviews (http://www.ryanmcd.com/papers/local_service_summ.pdf)

About GlobalLogic Inc.

GlobalLogic is a full-lifecycle product development services leader that combines deep domain expertise and cross-industry experience to connect makers with markets worldwide.Using insight gained from working on innovative products and disruptive technologies, we collaborate with customers to show them how strategic research and development can become a tool for managing their future. We build partnerships with market-defining business and technology leaders who want to make amazing products, discover new revenue opportunities, and accelerate time to market.

For more information, visit www.globallogic.com

GlobalLogic®

Contact

Emily Gunn
Marketing Communications
emily.gunn@globallogic.com