To appear in *IEEE Intelligent Systems*, vol(?), no(?), 2010.

Sentiment Analysis: A Multi-Faceted Problem

Bing Liu
Department of Computer Science
University of Illinois at Chicago
liub@cs.uic.edu

Sentiment analysis or opinion mining is the computational study of people's opinions, appraisals, and emotions toward entities, events and their attributes. In the past few years, it attracted a great deal of attentions from both academia and industry due to many challenging research problems and a wide range of applications [1]. Opinions are important because whenever we need to make a decision we want to hear others' opinions. This is not only true for individuals but also true for organizations. However, there was almost no computational study on opinions before the Web because there was little opinionated text available. In the past, when an individual needed to make a decision, he/she typically asked for opinions from friends and families. When an organization wanted to find opinions of the general public about its products and services, it conducted surveys and focus groups. However, with the explosive growth of the social media content on the Web in the past few years, the world has been transformed. People can now post reviews of products at merchant sites and express their views on almost anything in discussion forums and blogs, and at social network sites. Now if one wants to buy a product, one is no longer limited to asking one's friends and families because there are many user reviews on the Web. For a company, it may no longer need to conduct surveys or focus groups in order to gather consumer opinions about its products and those of its competitors because there is a plenty of such information publicly available.

However, finding opinion sites and monitoring them on the Web can still be a formidable task because there are a large number of diverse sites, and each site may also have a huge volume of *opinionated text*. In many cases, opinions are hidden in long forum posts and blogs. It is difficult for a human reader to find relevant sites, extract related sentences with opinions, read them, summarize them, and organize them into usable forms. Automated opinion discovery and summarization systems are thus needed.

In this article, I first give a brief introduction to the field and present some technical challenges. We will see that sentiment analysis is not a single task, but a multi-faceted problem containing many sub-problems. I will then share some of my thoughts on the past and future of sentiment analysis based on my research in the past few years and my experience in the industry for a short while.

1. The Problem of Sentiment Analysis

The research in the field started with sentiment and subjectivity classification, which treated the problem as a text classification problem. Sentiment classification classifies whether an opinionated document (e.g., product reviews) or sentence expresses a positive or negative opinion [2]. Subjectivity classification determines whether a sentence is subjective or objective [3]. Many real-life applications, however, require more detailed analysis because the user often wants to know what the opinions have been expressed on [1, 4]. For example, from the review of a product, one wants to know what features of the product have been praised and criticized by consumers.

Let us use the following review segment on iPhone as an example to introduce the general problem (a number is associated with each sentence for easy reference) [1]:

> "(1) *I bought an iPhone 2 days ago.* (2) *It was such a nice phone.* (3) *The touch screen was really cool.* (4) *The voice quality was clear too.* (5) *However, my mother was mad with me as I did not tell her before I bought it.* (6) *She also thought the phone was too expensive, and wanted me to return it to the shop. …* "

The question is: what we want to extract from this review? The first thing that we may notice is that there are several opinions in this review. Sentences (2), (3) and (4) express three positive opinions, while sentences (5) and (6) express negative opinions. Then we also notice that the opinions all have some targets on which they are expressed. The opinion in sentence (2) is on iPhone as a whole, and the opinions in sentences (3) and (4) are on the "touch screen" and "voice quality" features of iPhone respectively. The opinion in sentence (6) is on the price of iPhone, but the opinion/emotion in sentence (5) is on "me", not iPhone. This is an important point. In an application, the user may be interested in opinions on certain targets, but not on all (e.g., unlikely on "me"). Finally, we may also notice the sources or holders of opinions. The source or holder of the opinions in sentences (2), (3) and (4) is the author of the review ("I"), but in sentences (5) and (6) it is "my mother". With this example in mind, we can define sentiment analysis or opinion mining [1, 4]. We start with the opinion target.

Object and feature: In general, opinions can be expressed on any *target entity*, e.g., a product, a service, an individual, an organization, or an event. We use the term *object* to denote the target entity that has been commented on. An object can have a set of *components* (or *parts*) and a set of *attributes* (or *properties*) [1, 4], which we collectively call the *features* of the object.

A particular brand of cellular phone is an object. It has a set of components (e.g., *battery* and *screen*), and also a set of attributes (e.g., *voice quality* and *size*), which are all called features. An opinion can be expressed on any feature of the object and also on the object itself. For example, in "*I like iPhone. It has a great touch screen*", the first sentence expresses a positive opinion on "iPhone" itself, and the second sentence expresses a positive opinion on its "touch screen" feature.

Opinion holder: The *holder* of an opinion is the person or organization that expresses the opinion.

In the case of product reviews and blogs, opinion holders are usually the authors of the posts. Opinion holders are more important in news articles because they often explicitly state the person or organization that holds a particular opinion.

Opinion and orientation: An *opinion* on a feature f (or object o) is a positive or negative view or appraisal on f (or o) from an opinion holder. Positive and negative are called *opinion orientations*.

With these concepts in mind, we can define a model of an object, a model of an opinionated text, and the mining objective, which are collectively called the *feature-based sentiment analysis model* [1, 4].

Model of an object: An object o is represented with a finite set of features, $F = \{f_1, f_2, \ldots, f_n\}$, which includes the object itself as a special feature. Each feature $f_i \in F$ can be expressed with any one of a finite set of words or phrases $W_i = \{w_{i1}, w_{i2}, \ldots, w_{im}\}$, which are *synonyms* of the feature.

Model of an opinionated document: A opinionated document d contains opinions on a set of objects $\{o_1, o_2, \ldots, o_r\}$ from a set of opinion holders $\{h_1, h_2, \ldots, h_p\}$. The opinions on each object o_j are expressed on a subset F_j of features of o_j. An opinion can be either one of the following two types:

1. **Direct opinion**: A *direct opinion* is a quintuple $(o_j, f_{jk}, oo_{ijkl}, h_i, t_l)$, where o_j is an object, f_{jk} is a feature of the object o_j, oo_{ijkl} is the orientation of the opinion on feature f_{jk} of object o_j, h_i is the opinion holder and t_l is the time when the opinion is expressed by h_i. The opinion orientation oo_{ijkl} can be positive, negative or neutral.

2. **Comparative opinion**: A *comparative opinion* expresses a preference relation of two or more objects based on some of their shared features. It is usually conveyed using the *comparative* or *superlative* form of an adjective or adverb, e.g., "*Coke tastes better than Pepsi*". Due to space limitations, we will not discuss such opinions in this article (see [1] for more details).

Objective of sentiment analysis on direct opinions: Given an opinionated document d,

1. Discover all opinion quintuples $(o_j, f_{jk}, oo_{ijkl}, h_i, t_l)$ in d, and
2. Identify all synonyms (W_{jk}) of each feature f_{jk} in d.

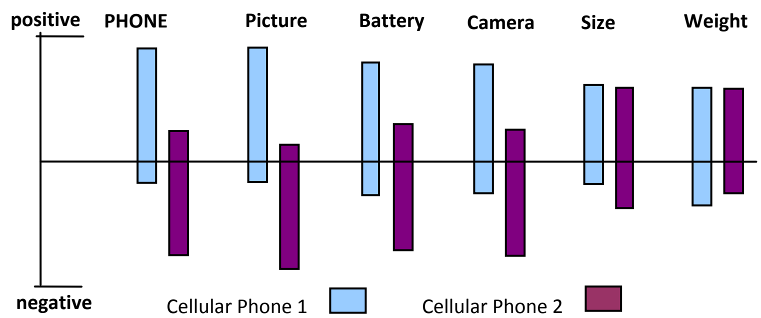

Figure 1. Visual comparison of feature-based opinion summaries of two cellular phones

In practice not all five pieces of information in the quintuple need to be discovered for every application because some of them may be known or not needed. For example, in the context of online forums, the time when a post is submitted and the opinion holder are all known as the site typically displays such information. We will not discuss them further in this article.

Applications: A simple way to use the results is to produce a *feature-based summary* of opinions on an object or multiple competing objects [1, 4]. Figure 1 shows the summary of opinions on two competing cellular phones along different feature dimensions. In the figure, each bar above the *X*-axis in the middle shows the number of positive opinions on a feature (given at the top), and the bar below the *X*-axis shows the number of negative opinions on the same feature. We can clearly see how consumers view different features of each product. "PHONE" represents the phone itself. Phone 1 is clearly a better product.

2. Technical Challenges

The objective of opinion mining gives us a good clue of the main tasks involved and technical challenges. None of the problems is solved. Let us use a more complex example blog to discuss them:

> "(1) *Yesterday, I bought a Nokia phone and my girlfriend bought a moto phone.* (2) *We called each other when we got home.* (3) *The voice on my phone was not clear.* (4) *The camera was good.* (5) *My girlfriend said the sound of her phone was clear.* (6) *I wanted a phone with good voice quality.* (7) *So I was satisfied and returned the phone to BestBuy yesterday*."

Object identification: The objects to be discovered in this blog are "moto" (Motorola) and "Nokia". This problem is important because without knowing the object on which an opinion has been expressed, the opinion is of little use. The issue is similar to the classic *named entity recognition* problem. However, there is a difference. In a typical opinion mining application, the user wants to find opinions on some competing objects (e.g., products). The system thus needs to separate relevant objects and irrelevant objects. For example, "BestBuy" is not a competing product name, but the name of a shop.

Feature extraction and synonym grouping: In the example above, the phone features are "voice", "sound", and "camera". Although there were attempts to solve this problem, it remains to be a major challenge. Current research mainly finds nouns and noun phrases. Although the recall may be good, the precision can be low. Furthermore, verb features are common as well but harder to identify. To produce a summary similar to the one in Figure 1, we also need to group synonym features as people often use different words or phrases to describe the same feature (e.g., "voice" and "sound" refer to the same feature in the above example). This problem is also very hard. A great deal of research is still needed [1].

Opinion orientation classification: This task determines whether there is opinion on a feature in a sentence, and if so, whether it is positive or negative. Existing approaches are based on supervised and unsupervised methods. One of the key issues is to identify opinion words and phrases (e.g., *good*, *bad*, *poor*, *great*), which are instrumental to sentiment analysis. The problem is that there are seemly an unlimited number of expressions that people use to express opinions, and in different domains they can be significantly different. Even in the same domain, the same word may indicate different opinions in different contexts [1]. For example, in the sentence, "*The battery life is long*" "long" indicates a positive opinion on the "battery life" feature. However, in the sentence, "*This camera takes a long time to focus*", "long" indicates a negative opinion. Also, sentence (6) in our example blog above seemingly expresses a positive opinion, but it does not. There are still many problems that need to be solved [1].

Integration: Integrating the about tasks is also complex because we need to match the five pieces of information in the quintuple. That is, the opinion oo_{ijkl} must be given by opinion holder h_i on feature f_{jk} of object o_j at time t_l. To make matters worse, a sentence may not explicitly mention some pieces of information, but they are implied due to pronouns, language conventions, and the context.

To deal with these problems, we need to apply NLP techniques in the opinion mining context, e.g., parsing, word-sense disambiguation, and coreference resolution. We use coreference resolution as an example to give a glimpse of the issues. For our example blog, to figure out what is "my phone" and what is "her phone" in sentences (3) and (5) is not a simple task. Sentence (4) does not mention any phone and does not have a pronoun. The question is which phone "the camera" belongs to. Coreference resolution is a classic problem in NLP. There is still no accurate solution from the research community.

3. My Perspective about the Past and Future

I would now like to share some of my thoughts on the past and future of the field based on my research and practical application experiences.

The Past

The research community has studied almost all main aspects of the problem. The most well studied sub-problem is opinion orientation classification (i.e., at the document level, sentence level and feature level). The existing reported solutions are still far from perfect. The main issue is that the current studies are still coarse. Not much has been done on finer details. For example, on opinion classification, there are many conceptual rules that govern opinions [1], and there are even more expressions (possibly unlimited) that can convey these concepts. However, little in-depth study has been done on many of them. On feature extraction and synonym grouping, they remain to be very challenging. Object extraction is probably the easiest because many existing information extraction algorithms can be applied. Integration and matching of all 5 pieces of information in the quintuple is still lacking, which is probably not surprising as the research community likes to focus on individual sub-problems. This leads us to the question of sentiment analysis accuracy, i.e., what is the accuracy of the current state-of-the-art algorithms? This question is not easy to answer because there are so many sub-problems. Although for some individual sub-problems researchers have annotated data for benchmark testing, there is still not a comprehensive public domain corpus that can be used to evaluate all tasks in a unified way.

The Future

Building on what have been done so far, I believe that we just need to conduct more refined and in-depth investigations, and to build integrated systems that try to deal with all the problems together because they are all needed in applications, and their interactions can help solving each individual problem. I am optimistic that the problems will be solved to a satisfactory level in the next few years for widespread applications. In fact, we may already begin to see the light at the end of the tunnel. For instance, based on our tests using 10 diverse data sets, the system that we are building (called Opinion Parser) can achieve 80-90% of accuracy on feature-based opinion orientation classification. It is also able to perform integration to a good extent based on several automated discovery functions.

On real-life applications, to provide a completely automated solution is nowhere in sight. However, it is possible to devise effective semi-automated solutions. The key is to fully understand the whole range of issues and pitfalls, cleverly manage them, and determine what portions can be done automatically and what portions need human assistance. In the continuum between the fully manual solution and fully automated solution, we can push more and more toward automation.

Beyond what have been discussed so far, it is also important to deal with the issue of opinion spam (e.g., fake reviews). Opinion spam refers to writing fake or bogus reviews that try to deliberately mislead readers or automated systems by giving untruthful positive and/or negative opinions in order to promote some target objects and/or to damage the reputations of some other objects [5]. Detecting such spam is vital as we go forward because heavy spam can make sentiment analysis useless for applications.

Finally, despite the challenges, the field has made significant progresses over the past few years. This is evident from the large number of start-up companies that provide sentiment analysis or opinion mining services. There is a real and huge need in the industry for such services. This practical need and the technical challenges will keep the field vibrant and lively for years to come.

References

[1]. B. Liu. *Sentiment Analysis and Subjectivity*. *Handbook of Natural Language Processing*, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010.

[2]. B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis." *Foundations and Trends in Information Retrieval* 2(1-2), pp. 1–135, 2008.

[3]. J. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin, "Learning Subjective Language," *Computational Linguistics*, vol. 30, pp. 277–308, September 2004.

[4]. M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (KDD), pp. 168–177, 2004.

[5]. N. Jindal, and B. Liu. "Opinion Spam and Analysis." *Proceedings of the ACM Conference on Web Search and Data Mining* (WSDM), 2008.

Bing Liu is a professor of computer science at University of Illinois at Chicago. He received his PhD in Artificial Intelligence from University of Edinburgh. Contact him at liub@cs.uic.edu.