ARTIFICIAL INTELLIGENCE AND OPINION MINING

SUCHETA PREMCHANDANI[1]*, MANISH PISE[2] AND ASHISH WANKHEDE[3]

Department of Computer Science and Engineering, JDIET, Yavatmal, MS, India.
*Corresponding Author: Email- suchetap02@gmail.com

Abstract- An important part of our information-gathering behavior has always been to find out what other people think. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others. The main goal of this paper is to extracting, classifying, understanding and accessing the opinions expressed in various online news sources. Here opinion mining refers to computational techniques for analyzing the opinions that are extracted from various sources. Current opinion research focuses on business and e-commerce such as product reviews and movie ratings. We developed a framework for analysis with four major stages such as stakeholder analysis, topical analysis, sentiment analysis and stock modeling. During the stakeholder analysis stage, we identified the stakeholder groups participating in web forum discussions. In the topical analysis stage, the major topics of discussion driving communication in the Web forum are determined. The sentiment analysis stage consists of assessing the opinions expressed by the Web forum participants in their discussions. Finally, in the stock modeling stage, we examine the relationships between various attributes of web forum discussions and the firm's stock behavior.
Keywords- Opinion Mining, Sentiment analysis

Introduction

"What other people think" has always been an important section of information for most of us during the decision-making process. Long before knowledge of the World Wide Web became widespread, many of us asked our friends to suggest an auto mechanic or to explain who they were planning to vote for in local elections, requested reference letters regarding job applicants from age group. But the Internet and the Web have now made it possible to find out about the opinions and experiences of those in the huge group of people that are neither our personal associates nor well-known professional critics that is, people we have never heard of. And on the contrary, more and more people are making their opinions available to strangers by means of the Internet.

Many new and exciting social, geo political, and business-related research questions can be answered by analyzing the thousands, even millions, of comments and responses expressed in various blogs such as the blogosphere, forums such as Yahoo Forums, social media and social network sites including YouTube, Facebook, and Flikr, virtual worlds, and Twitter. Opinion mining, a sub-discipline within data mining and computational linguistics, refers to the computational techniques for extracting, classifying, understanding, and assessing the opinions expressed in various online news sources, social media comments, and other user-generated content. Sentiment analysis is often used in opinion mining to identify sentiment, affect, subjectivity, and other emotional states in online text. For example, we might seek to answer these such questions:

-What are the opinions and comments of investors, employees, and activists toward Wal-Mart in light of its cost-reduction efforts and global business practices?

-What was the most successful McDonald's promotional campaign conducted recently in China, and why did it succeed? Which McDonald's product is most preferred by young students in China and why?

Much advanced research in this area has recently focused on several critical areas. In this installment of Trends & Controversies and the next, we review several contributions to this emerging field. The topics covered include how to extract opinion, sentiment, affect, and subjectivity expressed in text. For example, resources online might include opinions about a product or the violent and racist statements expressed in political forums. Researchers have also been able to classify text segments based on sentiment, affect, and subjectivity by analyzing positive or negative sentiment expressed in sentences, the degree of violence expressed in forum messages, and so on.

Opinion target, opinion holder and opinion are the definitions used to extracting opinions from different online sources. An opinion can be expressed in two types. 1. Direct opinion, 2.Comparative opinion. All the opinions are stored in a document. Following are the steps to extracting the opinions.

1)Identify the objects.
2)Feature extraction and synonym grouping.
3)Opinion orientation determination.
4)Integration.

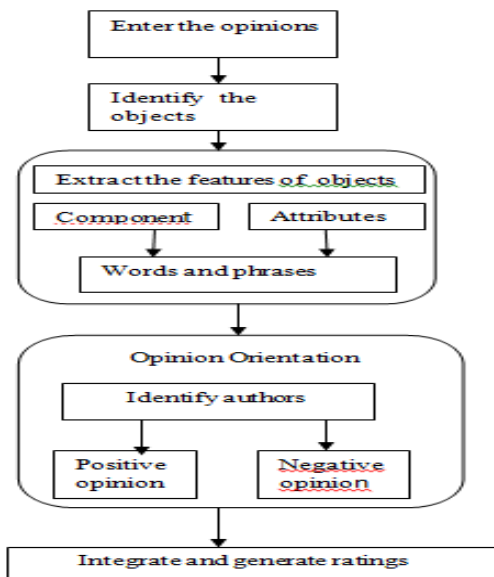Figure below shows the block diagram of the complete process used in extracting the opinion.

Fig. 1- Block diagram of the complete process used in extracting the opinion

Object identification
This object identification is important because without knowing the object on which an opinion has been expressed, the opinion is of little use. The problem is similar to the classic named entity recognition problem, but there is a difference. In a typical opinion mining application, the user wants to find opinions on some competing objects, such as competing products or services. Thus, the system need to separate relevant and irrelevant objects. In general, people can express opinions on any target entity like products, services, individuals, organizations, or events. In this paper, the term object is used to denote the target entity that has been commented on. For each comment, we have to identify an object. Based on objects, we have to integrate and generate ratings for opinions. The object is represented as "O". An opinionated document contains opinion on set of objects as {o1, o2, o3… or}.

Feature extraction
An object can have a set of components (or parts) and a set of attributes (or properties) which we collectively call the features of the object. For example, a cellular phone is an object. It has a set of components (such as battery and screen) and a set of attributes (such as voice quality and size), which are all called features (or aspects). An opinion can be expressed on any feature of the object and also on the object itself. With these concepts in mind, we can define an object model, a model of an opinionated text, and the mining objective, which are collectively called the feature-based sentiment analysis model. In the object model, an object "O" is represented with a finite set of features, $Fr = \{fr1, fr2,…, frn\}$ which includes the object itself as a special feature. Each feature fi $\in F$ can be expressed with any one of a finite set of words or phrases Wi = {wi1, wi2, …, wim} which are the feature's synonyms.

Opinion-orientation determination
The next task is to determine whether a verdict contains an opinion on a feature and, if so, whether it is positive or negative. Existing approaches are based on special supervised and unsupervised methods using opinion words and phrases and the grammar information. One key issue is to identify opinion words and phrases (such as good, bad, poor, or great which are helpful to sentiment analysis. However, there are seemly an unlimited number of expressions

that people use to express opinions, and in different domains, they can be considerably different. Even in the same domain, the same word might indicate different opinions in different context.[1]For instance, in the verdict "The battery life is long," long indicates a positive opinion about the battery life feature. However, in the verdict "This camera takes a long time to focus," long indicates a negative opinion.The opinion holder 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. 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. From this opinion orientation we have to determine the type of opinion whether it is direct opinion or comparative opinion.

Direct opinion
A direct opinion is a quintuple (oj, fjk, ooijkl, hi, tl),
where oj is an object, fjk is a feature of the object oj,
ooijkl is the orientation of the opinion on feature fjk of object oj, hi is the opinion holder, and tl is the time when the opinion is expressed by hi. The opinion orientation ooijkl can be positive, negative, or neutral.

Comparative opinion
A comparative opinion expresses a preference relation of two or

more objects based their shared features. A comparative opinion is usually conveyed using the comparative or superlative form of an adjective or adverb, such as "Coke tastes better than Pepsi."

Integration

Integrating these tasks is also complicated because we need to match the five pieces of information in the quintuple. That is, the opinion $ooijkl$ must be given by opinion holder hi on feature fjk of object oj at time tl .To make matters worse, a sentence might not explicitly mention some pieces of information, but they are implied using pronouns, language conventions, and context. Then generate ratings based on above tasks. Thus we can clearly see how holders view the different features of each product.

Sentiment Analysis: A complex Problem

Sentiment analysis is the computational study of people's opinions, appraisals, and emotions toward entities, events and their attributes. In the past few years, this field has attracted a great deal of attention from both the academic world and diligence due to many challenging research problems and a range of applications. Opinions are important because whenever people need to make a assessment, they want to hear others' opinions. The same is factual for organizations. However, few computational studies on opinions existed prior to the Web because there was little opinionated text available. In the past, when making a decision, individuals usually asked for opinions from friends and families. When an organization wanted to hit upon opinions of the general public about its products and services, it conducted surveys and focus groups. However, with the rapid 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 mercantile sites and express their views on almost anything in discussion forums and blogs, and at social network sites. Hence, individuals are no longer limited to asking friends and families because of the surplus of user-generated product reviews and opinions available on the Web. In turn, companies might no longer need to conduct surveys or focus groups to gather round consumer opinions about its products and those of its competitors because there's plenty of such information publicly available. However, finding opinion sites and monitoring them on the Web is a difficult task because there are numerous, diverse sources, each of which might also have a huge volume of opinionated text. In many cases, opinions are veiled in long forum posts and blogs, so 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 formats. Automated opinion detection and summarization systems can address this need. In this article, a brief introduction to this problem has been given and some technical challenges have been presented. As we will see, sentiment analysis is not a solitary problem, but a combination of many facets or subproblems. This article introduces and explains some of these dilemmas.

Sentiment-Analysis Dilemma

The research in the field begin with sentiment and subjectivity classification, which treated the problem as a text classification problem. Sentiment classification classifies whether an opinionated document or verdict expresses a positive or negative opinion [2].Subjectivity classification determines whether verdict is subjec-

tive or objective [3].Many real-life applications, however, require more thorough analysis because users often want to know the subject of opinions.[1,4] For example, from a product review, users want to know which product features patrons have praised and criticized. To explore this common problem, let's use the following review segment on Motorola Phone as an example:

• I bought an Motorola Phone two days ago.

• It was such a nice phone.

• The touch screen was really good.

• The voice quality was clear too.

• However, my mother was angry with me as I did not tell her before I bought it.

• She also thought the phone was too expensive, and wanted me to return it to the shop.

The question is, what do we want to extract from this evaluation? 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 or emotions. We can also see that all the opinions are expressed about some targets or objects. For example, the opinion in sentence 2 is on the Motorola Phone as a whole, and the opinions in sentences 3 and 4 are on the Motorola Phone's touch screen and voice quality, respectively. Importantly, the opinion in sentence 6 is on the phones' price, but the opinion/emotion in sentence 5 is about "me," not the phone. In an application, the user might be interested in opinions on certain targets but not necessarily on user-specific information. Finally, we can also see the sources or holders of opinions. The source or holder of the opinions in sentences 2, 3, and 4 is the review of the author, but in sentence 5 and 6, it is "my mother."

Definitions

With this example in mind, we can now delineate the sentiment analysis or opinion mining problem. We start with the opinion target. In general, people can express opinions on any target entity—products, services, individuals, organizations, or events. In this framework the term object is used 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 together call the features of the object. For example, a particular brand of cell phone is an object. It has a set of components (such as battery and screen) and a set of attributes (such as voice quality and size), which are all called features (or aspects).
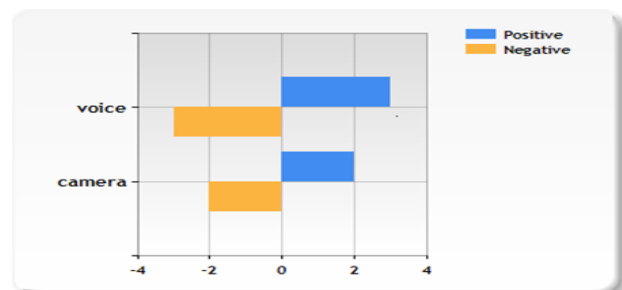

Fig. 2- Visual comparison of feature-based opinion summaries of cellular phone. Each bar above the x-axis shows the number of positive opinions on a feature (given at the top), and the bar below shows the number of negative opinions on the same feature.

An opinion can be expressed on any feature of the object and also on the object itself. For instance, "I like the iPhone. It has a great touch screen," the first sentence expresses a positive opinion on the iPhone itself, and the second sentence expresses a positive opinion on its touch screen feature. We have collected the opinion regarding the camera and the voice quality of a particular cellular phone. We collected the number of likes and dislikes from the opinion holder. The figure below shows the graph which contains the opinion regarding the feature of the cellular phone.

The opinion holder is the person or organization that expresses the opinion. In the case of product review and blogs, opinion holder is usually the author of the post. Opinion holders are more important in news articles because they often unambiguously state the person or organization that holds a particular opinion. An opinion on a feature fr (or object o) is a positive or negative view or appraisal on fr (or o) from an opinion holder. Positive and negative are called opinion orientations. With these concepts in mind, we can define an object model, a model of an opinionated text, and the mining objective, which are together called the feature-based sentiment analysis model [1,4]. In the object model, an object o is represented with a finite set of features, $Fr = \{fr1, fr2, …, frn\}$, which includes the object itself as a particular feature. Each feature $fri \in Fr$ can be expressed with any one of a finite set of words or phrases $Wi = \{wi1, wi2, …, wim\}$, which are the feature's synonyms. In the opinionated document model, an opinionated document d contains opinions on a set of objects $\{o1, o2, … , or\}$ from a set of opinion holders $\{h1, h2, …, hp\}$. The opinions on each object oi are expressed on a subset Fi of features of oi. An opinion can be one of the following two types: A direct opinion is a quintuple (oi, fik, ooijkl, hi, tl), where oi is an object, fik is a feature of the object oi, ooijkl is the orientation of the opinion on feature fik of object oi hi is the opinion holder, and tl is the time when the opinion is expressed by hi. The opinion orientation ooijkl can be positive, negative, or neutral. A comparative opinion expresses a preference relation of two or more objects based their mutual features. A comparative opinion is usually conveyed using the comparative or superlative form of an adjective or adverb, such as "Coke tastes better than Pepsi". Therefore, given an opinionated document d, the point of sentiment analysis (or opinion mining) is twofold:

-discover all opinion quintuples (oi, fik, ooijkl, hi, tl) in d and

-identify all synonyms (Wik) of each feature fik in d.

In practice, not all five pieces of information in the quintuple need to be discovered for every application because some of them might be recognized or not needed. For instance, in the context of online forums, the site typically displays the time when a post is submitted and identifies the opinion holder. Figure 2,shows the detailed process of extracting sentiment features evolved during the identification of opinion.
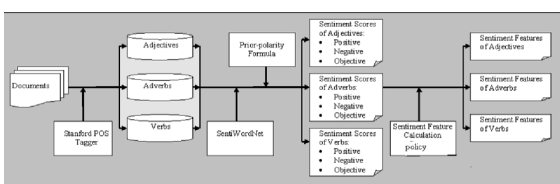

Fig. 3- The detailed process of extracting sentiment features

Conclusion

Our goal in this survey has been to cover techniques and approaches that promise to directly enable opinion-oriented information-seeking systems, and to convey to the reader a sense of our excitement about the intellectual richness and breadth of the area. Previous sentiment classification studies have used either the machine learning approach or the semantic orientation approach. Few have combined both approaches to improve sentiment classification performance. A crucial issue, however, is ensuring the correct proportions of positive and negative opinions on each feature. Hence, the system errors should be balanced so that they do not destroy the natural distribution of positive and negative opinions. Finally, despite these difficulties and challenges, the field has made significant progress over the past few years. This is evident from the large number of start-up companies that provide sentiment-analysis and opinion mining services. A real, substantial need exists in industry for such services.This practical need and the technical challenges will keep the field vibrant and lively for years to come.

References

[1] Abbasi A. et al. (2008) *IEEE Trans. Knowledge and Data Eng.*, 20(9), 1168-1180.

[2] Esuli A. and Sebastiani F. (2006) *5th Conf. Language Resources and Evaluation*, 417-422.

[3] Chen H. (2006) *Intelligence and Security Informatics for International Security*.

[4] Nye J. (2004) *Soft Power: The Means to Success in World Politics*.

[5] Chen H. (2010) *IEEE Intelligent Systems*, 25(1), 68-71.

[6] Antweiler W. and Frank M. (2004) *J. Finance*, 59(3), 1259-1295.

[7] Liu B. *Handbook of Natural Language Processing, 2nd edition*.

[8] Indurkhya N. and Damerau F.J. (2010) *Chapman & Hall*, 627-666.

[9] Pang B. and Lee L. (2008) *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.

[10] Wiebe J. et al. (2004) *Computational Linguistics*, 30, 277-308.

[11] Hu M. and Liu B. (2004) *ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, 168-177.

[12] Jindal N. and Liu B. (2008) *Conf. Web Search and Web Data Mining*, 219-230.