Sentiment Analysis: Introduction and the State of the Art overview

Adam Westerski

Universidad Politecnica de Madrid, Spain
westerski@dit.upm.es

Abstract. In the following paper we present an overview of the state of the art in the area of sentiment analysis. Although this domain is quite new it already has a considerable amount of contributions. Therefore, we present the general trends and within we only focus on some of the most important topics. Additionally, we present some interesting emerging projects in the contemporary research and point out the key appliances of sentiment analysis. It has to be noted that this paper is meant to be introduction to the topic rather then a comprehensive and full overview.

Keywords: opinion mining, sentiment analysis, NLP

1 Introduction

Recent years have brought the burst of popularity of community portals across the Internet. Alongside of the content created by the portal editors the so called user-generated content is an important part of many websites. Users provide their input not only through discussions and personal notes in various social web spaces (boards, blogs etc) but also on mass scale leave comments and reviews of products and services on numerous commercial websites. The fast growth of such content has not been fully harnessed yet. Information left by the users is often very disorganized and many portals that enable user input leave the user added information unmoderated.

Opinion mining (often referred as sentiment analysis) is an attempt to take advantage of the vast amounts of user generated content. It employs computer processing power to formalize the knowledge taken from user opinions and analyze it for further reuse. Although there are some early works about recognition of subjective texts from early 80s and 90s, the real progress in the area started with the rise of Web 2.0. The new types of Internet content enforced new ways of data management which ,as a consequence, caused new problems and opportunities to arise. Over the last decade a huge increase of interest in the sentiment analysis research is clearly visible [1, 2].

In the following work we try to provide an introduction to the topic of opinion mining through the presentation of the key directions in the research and the associated state of the art in the areas(see Sec. 2,3,4,5). Additionally we present some interesting use cases for the opinion mining and provide a short discussion of their importance mostly with connection to Internet technologies (see Sec. 6).

2 Domain overview

The opinion mining is often associated with another research topic – information retrieval(IR). Nevertheless, opinion mining proves to be a lot difficult task. The primary reason is characteristics of the data sources. In IR, the algorithms operate on factual data, while in opinion mining input data is only subjective information. In practice, this means that opinion mining is needed to go a step further then information retrieval and analyze sentences and phrases deeper with respect to their semantics. During the facts analysis one is interested in simple characteristics and extracting it. In opinion mining the additional task is to determine the nature of opinion: whether it is positive or negative in general; what features does it describe; what features are valued, which are not etc.

As mentioned before, the rise of interest in the area has been caused by the growth of user-generated content on the Web. One of the primary characteristics of such content is its textual disorder and high diversity. The style of writing opinions varies a lot within a single portal but even more if one was to analyze a given topic in the Internet wide scale. Opinions are expressed with informal language. Therefore sentence construction can very a lot depending on the community (which can go even as far as altering grammar within a single language). For instance the product reviews contributed on Amazon about a movie and a computer game based on it can be written totally different and even sometimes not understandable for people outside of the community. Typically, information retrieval techniques achieve best results when applied to highly structured, formalized text, in most cases opinion mining does not have this comfort. In order to give more insight into the problem, in the subsequent subsections we describe various attempts to classify and formalize different opinion types.

2.1 Types of evaluation

In general, there are two main ways to express sentiments: direct opinions and comparisons. Direct opinions usually describe one object and contain some adjectives that refer to it (i.e. “the image quality of this camera is good”). In contrast, the comparative statements mention more then one object and describe some sort of relation (“i.e. the image quality of camera X is much better then camera Y”).

2.2 Types of context

To extract the opinion one has to to know what the opinion is about. Depending on the location/portal the descriptive information can be stated in many different ways. On review portals it is often relatively easy to extract sentiment information but for instance on a forum it is considerably harder to identify the subject of discussion or subject of a single post.

As it is shown further (see Sec. 5), the software that extracts opinions and performs any kind of automatic sentiment recognition is often built for specific contexts. The generic engines and algorithms perform much worse then applications meant only to analyze particular types of text (i.e. movie reviews). Although this is a huge limitation , in practice such techniques still hold a great value even with scope narrowed down to opinion mining posts from a single portal.

2.3 Level of interest

People can express their opinions with different detail. Some will give general information while others will provide more in depth review. Additionally some skip from one product feature to another with only a brief description while others elaborate on certain features a lot more.

This factor has a particular importance during the overall classification of the text orientation (positive/negative). One has to judge if separate sentences refer to the same attribute/object or different. Similarly depending on the user interest one sentence can express many opinions within.

2.4 Querying formula

Depending on the person and the place where people share their opinions, statements and queries can be expressed in a different way. Some users tend to use keywords or short sentences while others provide full text. For example: “iPhone advantages” or “what are the advantages of the iPhone?”

2.5 Type of vocabulary used

Opinions can be expressed in many different ways depending on the manner the vocabulary is used. One can use words that directly refer to the sentence subject (i.e. “I think this product is bad!”) or use affect vocabulary that contains more emotions and can be much harder to recognize (i.e. “I love the way this switch works!” or “I was stunned to see all those special effects”).

Additionally, with regard to vocabulary and grammar, we can variate opinions between stated explicitly (as in using simple language constructs and clear statements) or implicitly (i.e. “This phone fits right into my pocket”).

3 Document level sentiment analysis

Document opinion analysis is about classifying the overall sentiments expressed by the authors in the entire document text. The task is determine whether the document is positive, negative or neutral about a certain object. When applied to a single type of text those techniques typically have a range of accuracy from 70% to 80% depending on amount of human input and type of text [3]. In the rest of the following section we present a number of most representative solutions in the area, some present an novel algorithms [4], while other try to implement approaches proven in other domains.

The work done by Turney [4] on review classification presents an approach based on distance measure of adjectives found in text from preselected words with known polarity (“excellent” and “poor”). The author presents a three step algorithm which processes documents without human supervision. First, the adjectives are extracted along with a word that provides contextual information. Words to extract are identified by applying predefined patterns (for instance: acjective-noun or adverb-noun etc.).Next, the semantic orientation is measured. This is done by measuring the distance from words of known polarity. The mutual dependence between two words is found by analysis of hit count with AltaVista search engine for documents that contain two words in a certain proximity of each other. At the end the algorithm counts the average semantic orientation for all word pairs and classifies a review as recommended or not.

In contrast, Pang et al. [5] present an work based on classic topic classification techniques. The proposed approach aims to test whether a selected group of machine learning algorithms can produce good result when sentiment analysis is perceived as document topic analysis with two topics: positive and negative. Authors present results for experiments with: Naïve Bayes [6], Maximum Entropy [7] and Support Vector Machine algorithms [8]. Interestingly the performed tests have shown results comparable to other solutions ranging from 71 to 85% depending on the method and test data sets.

4 Sentence level sentiment analysis

The sentence level opinion mining is an action that can be associated with two tasks. Initial work is to identify whether the sentence is subjective (opinionated) or objective. The second task is to classify a subjective sentence and determine if it is positive, negative or neutral. Similarly as with document level most techniques use forms of machine learning.

Riloff and Wiebe [9] put most of impact in their work on the task of subjective sentences identification. They propose a method that at bootstrap uses a high precision (and low recall) classifiers to extract a number of subjective sentences. During this phase sentences are labeled by two classifiers: first for high confidence subjective sentences, second for high confidence objective sentences (see Fig. 1). The sentences that are not clearly classified into any category are left unlabeled and omitted at this stage. Both of the classifiers are based on preset list of words that indicate sentence subjectivity. The subjective classifier looks for the presence of words from the list, while the objective classifier tries to locate sentences without those words. According to the results presented by authors their classifiers achieve around 90% accuracy during the tests.

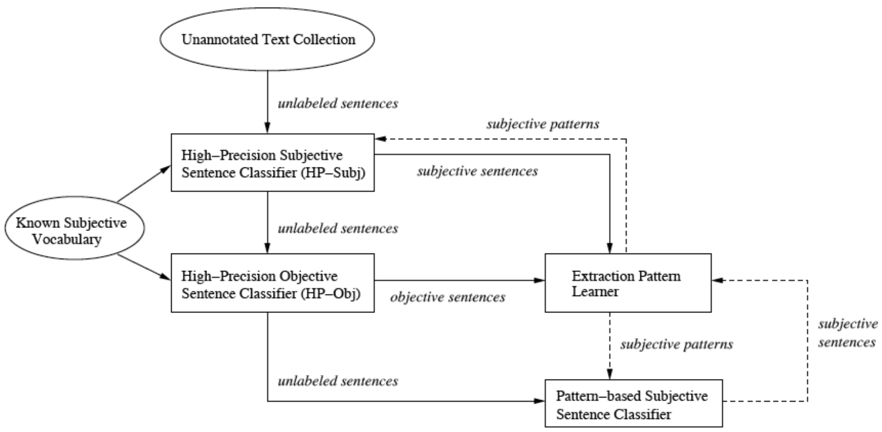

Figure 1. Bootstrap phase [9]

In the second step, the gathered data is used a for training an extraction algorithm that generates patterns for subjective sentences. The patterns are used to extract more sentences in the same text. The presented method has such split in order to increase recall after the initial bootstrap phase (however, as expected, author report the precision to fall between 70-80%).

SYNTACTIC FORM	EXAMPLE PATTERN
<subj> passive-verb	<subj> was satisfied
<subj> active-verb	<subj> complained
<subj> active-verb dobj	<subj> dealt blow
<subj> verb infinitive	<subj> appear to be
<subj> aux noun	<subj> has position
active-verb <dobj>	endorsed <dobj>
infinitive <dobj>	to condemn <dobj>
verb infinitive <dobj>	get to know <dobj>
noun aux <dobj>	fact is <dobj>
noun prep <np>	opinion on <np>
active-verb prep <np>	agrees with <np>
passive-verb prep <np>	was worried about <np>
infinitive prep <np>	to resort to <np>

Figure 2. Extraction patterns learning phase – syntactic templates and corresponding extraction patterns [9]

During the learning phase the algorithm uses a predefined set of syntactic templates that are matched against the subjective sentences (see Fig. 2). After the entire training set is processed the extracted patterns are ranked based on their occurrence frequency and according to some preset conditions only the best patterns are selected for next iteration of base text analysis (see Fig. 3).

PATTERN	FREQ	%SUBJ
<subj> was asked	11	100%
<subj> asked	128	63%
<subj> is talk	5	100%
talk of <np>	10	90%
<subj> will talk	28	71%
<subj> put an end	10	90%
<subj> put	187	67%
<subj> is going to be	11	82%
<subj> is going	182	67%
was expected from <np>	5	100%
<subj> was expected	45	42%
<subj> is fact	38	100%
fact is <dobj>	12	100%

Figure 3. Extraction patterns learning phase – patterns frequency within subjective sentences (in contrast to objective sentences) [9]

Although the presented work does achieve quite good results it only concerns one task put ahead for sentence sentiment analysis. In opposition to it, work done by Yu and Hatzivassiloglou [10] discusses both sentence classification (subjective/objective) and orientation (positive/negative/neutral). For the the first step of sentence classification, authors present test results for three different algorithms:

sentence similarity detection, naïve Bayens classification and Multiple naïve Bayens classification. In the second step of sentence orientation recognition authors use a technique similar to the one used by Turney [4] for document level sentiment analysis (see Sec. 3). The main different is that the algorithm is extended to use more then two ("excellent"/"poor") base words to which all others are compared.

5 Feature based sentiment analysis

The feature level of sentiment analysis is the most detailed study of the text. Being most useful it is also the hardest to perform. The goal is to not only determine text subjectivity and polarity but also what in particular the text author liked or disliked about the object. Typical this objective is split into the following tasks:

- extract object features that are commented
- determine orientation of opinions (positive/negative/neutral)
- group feature synonyms and produce a summary (see Fig. 4)

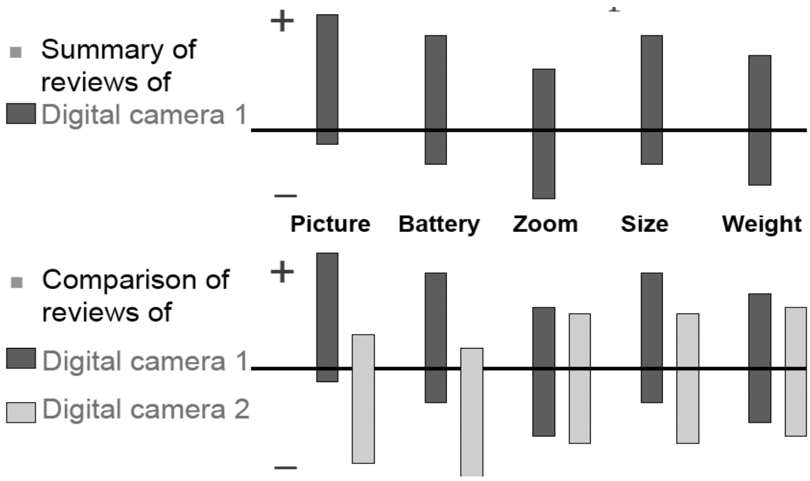

Figure 4. Sample output of the feature based sentiment analysis [3]

Similarly as with both previously described level (see Sec.3 and Sec. 4) often the feature sentiment analysis experiments are conducted only for a single selected text type. Sometimes authors go even further and present methods for specific text format, for instance reviews where positive and negative features are explicitly separated is different areas. Such approach is presented by Hu and Liu in their work about customer reviews analysis [11]. In their research authors present opinion mining based on feature frequency. Only the most frequent features recognized by precessing many review are taken into consideration during summary generation.

6 Opinion mining appliances

This section presents some selected and most prominent use cases for opinion mining techniques. Some have been already related in various research experiments previously mentioned while others still remain a goal to achieve in the future.

6.1 Product benchmarking and market intelligence

The key to selling a product is responding to customers demands in proper time and in the right location. Many companies spend huge money on market analysis and hire external specialized consulting companies. The opinion mining techniques could aid this effort and potentially minimize costs. Market analysis done by specialized companies is needed to take certain amounts of time and effort, while in many cases getting fast access to accurate market data can be a key factor. The right opinion mining tools could create a business advantage for a company to get ahead of its competitors and swiftly react to customer needs.

Additionally opinion mining opens new frontiers. With the immense amounts of community created data on the Internet its analysis becomes impossible or at least very difficult and expensive without some automatic methods. This domain is huge and the amount of appliances possible is vast.

6.2 Advertisement placement

Advertisements over the Internet are best to be placed in places where they can reach the biggest group of potential customers. For instance it is best to advertise selected specialized computer equipment on tech forums, while entire desktop computer sets will find better audience among more common Internet users. Therefore, often topic-based mining techniques are used. Nevertheless sometimes this can be insufficient. For instance one could imagine a situation where a tech review website release a negative review of a product and the topic mining techniques select to display the advertisement of this product next to the review (because topic matches). In such case opinion mining would help to analyze the polarity of the article and not display the ad.

Additionally detection of text polarity and semantics with relation to advertisement topic can help to detect whether content of the website and commercial message contextually fit to each other in other not to bring harm to company reputation or brand popularity. For instance it would be very bad to display a commercial of airlines next to a news post about an airplane crash.

6.3 Individual needs

The opinion mining system could be potentially used by casual Internet users. The aforementioned feature level analysis (see Sec. 5) can be a very good way (if accurate) to provide a summarized view of posts for community review sites (for instance movie reviews or product reviews on Amazon book store). Such small enhancements can greatly improve user experience thus being beneficial for both content consumers and producers.

6.4 Opinion search and retrieval

Opinion search engines could be very beneficial both for individual and corporate users. In theory it could be very to look for opinions just like browsing the current web. For instance one might one to type a query “mobile Nokia” and desire to get only text with user opinions as a result.

6.5 Opinion spam detection

The opinion spam is a direct result of the user generated content popularity. The opinions given by the users about various products and services have gained huge commercial value over the recent years. The modern Internet is being ruthlessly used just like other media as a battle front for clients in between companies and corporations. Therefore, not surprisingly, the systems that enable to post opinions are often abused. Fake or misinforming comments are posted to mislead the potential client into buying or not buying a product. This can be done both automatically but also by humans. In small scale opinion content is easy to moderate, however on big and popular forums, message boards or even internet shops this can be a very hard task. Systems that detect bogus product comments could improve the credibility of any community portal thus increasing the potential revenue. Nevertheless this domain is still not discussed that much in the open.

One reason is that misleading clients and subtly deceiving them to buy a particular product stead of others is a target of every profit oriented company on the market. Therefore, it is in direct interest of such companies to use any legal means to criticize all competitive products, and make this criticism look as credible as possible (even if its not). On the other hand it would be extremely useful to be able to eliminate all false comment about own products (and have this ability exclusive).

Secondly, it has to be noted that the problem of detecting opinion spam is even harder task then opinion mining itself. One has not only to detect sentences with opinions or types of opinions but also judge which opinion is correct and which is deliberately formed false. In some situations this can be impossible to determine for human not to even mention any machine AI like techniques.

7. Conclusions

A lot research is being conducted in the domain and many attempts to approach the topic have been presented. Nevertheless most of the approaches are still very naïve and it seems that the contemporary results can be improved in many ways in almost every domain presented in the following paper. The trend that seems to be followed by many is to improve or port their research results into various domains to make the solutions more independent or at least more easily adjustable to new context.

Additionally a very important and especially motivating factor is the potential business benefit that can come from fully functional sentiment analysis systems (see Sec. 6). Therefore, not surprisingly, the interest in the topic of both small specialized companies (Nielsen [12], Biz360 [13], Cymfony [14] and in recent years many others) and large corporations (i.e. IBM TAKMI system [15]) is growing very rapidly. It is certain that this trend will continue and it is very possible that it will drive the future of opinion mining.

References

1. Sentiment Classification bibliography, http://liinwww.ira.uka.de/bibliography/Misc/Sentiment.htm
2. Opinion Mining bibliography by Jan Wiebe, http://www.cs.pitt.edu/~wiebe/subjectivityBib.html
3. Liu, B., Opinion Mining and Summarization, World Wide Web Conference, Beijing, China, 2008
4. Turney, P., “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews.”, ACL 2002, 417-424
5. Bo Pang and Lillian Lee and Shivakumar Vaithyanathan, “Thumbs up? Sentiment Classification using Machine Learning Techniques”, Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002, 79-86
6. Lewis, D., “Naive (Bayes) at forty: The independence assumption in information retrieval.” In Proc. of the European Conference on Machine Learning (ECML), 1998
7. Berger, A., Della Pietra, S., J. Della Pietra, V., "A maximum entropy approach to natural language processing", Computational Linguistics, 1996
8. Joachims, T., "Text categorization with support vector machines: Learning with many relevant features.", In Proc. of the European Conference on Machine Learning (ECML), 1998
9. Riloff, E., Wiebe, J., “Learning Extraction Patterns for Subjective Expressions”, Proceedings of the 2003 Conference onEmpirical Methods in Natural Language Processing(EMNLP), Japan, Sapporo, 2003
10. Yu, H., Hatzivassiloglou, V., “Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences”, Proceedings of the 2003 Conference onEmpirical Methods in Natural Language Processing(EMNLP), Japan, Sapporo, 2003
11. Hu, M., Liu, B., “Mining and Summarizing Customer Reviews”, Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA, 2004
12. Nielsen Online homepage, http://www.nielsen-online.com/
13. Biz360 homepage, http://www.biz360.com/
14. Cymfony homepage, http://www.cymfony.com/
15. IBM TAKMI – Text mining system homepage, http://www.trl.ibm.com/projects/textmining/takmi/takmi_e.htm