Opinion Mining Survey

Trung Huynh - trunghlt@gmail.com

1 Abstract

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 sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object.

This survey covers techniques and approaches that promise to directly enable opinion-oriented information- seeking systems. It includes material on summarization of evaluative text and on broader issues regarding privacy, manipulation, and economic impact that the development of opinion-oriented information-access services gives rise to. To facilitate future work, a discussion of available resources, benchmark datasets, and evaluation campaigns is also provided.

Acknowledgement: Much of this work is cited from Pang and Lee [23]

2 Introduction

Even before the internet, we frequently ask our friends for recommendation or review about stuff we are interested in. The web has enabled collecting experience and opinions from a vast pool of users. The demand of these information interpretation and summarization has been increasing for both individual customers and corporations.

From the surveys ComScore [5], Horrigan [10], 81% of Internet users (60% of Americans) have done online research on a product at least once; 20% (15% of all Americans) do so on a typical day; Among readers of online reviews of restaurants, hotels, and various services (e.g., travel agencies or doctors), between 73% and 87% report that reviews had a significant influence on their purchases; Consumers report being willing to pay from 20% to 99% more for a 5-star-rated item than a 4-star-rated item (the variance stems from what type of item or service is considered); 32% have provided a rating on a product, service, or person via an online rating system, and 30% (including 18% of online senior citizens) have posted an online comment or review regarding a product or service.

For political information, Rainie and Horrigan [26] studied the 31% of Americans - over 60 million people that were 2006 campaign internet users: 28% said that a major reason for these online activities was to get perspectives from within their community; 34% said that a major reason was to get perspectives from outside their community; 27% had looked online for the endorsements or ratings of external organizations.

Kim [14] notes 75,000 new blogs are created daily; 1.2 million new posts each day. Therefore businesses require a new tool to replace traditional methods in order to keep track of consumer opinions.

2.1 Applications to review-related websites

The current search engine can retrieve documents which contain search keywords but can't answer for questions like "How do people think about ipad 2?". In order to answer for these option questions there needs a different treatment to analyse sentiment in clause, sentence and document levels.

Sentiment analysis can also be used to create an automated review and opinion-aggregation machine. Online customers usually do careful research before buying expensive products. Such an opinion-aggregation machine helps customers save time from searching and acknowledges or recommends them about products with good or bad reviews form others.

2.2 Applications as a sub-component technology

Sentiment analysis helps advertisers identify opinion in articles and distribute proper ads to readers. If the author is expressing positive opinion about a product, advertisers can distribute advertisement about that product or related accessories. If the author is expressing negative opinion about that product, advertisers can sell advertisement about its competitors. The effect of improper advertisement can be very low and counter-productive.

2.3 Applications in businesses and government intelligence

Applications in this area include reputation and public relationship management. Companies have been paying a lot of money for manually tracking reputation of their products online. Instead, an automated sentiment analysis machine can save them a significant amount of resource. In addition, these data can also help with trend prediction and proper reaction for corporates to optimise their customer satisfaction.

2.4 Applications across different domains

In politics, understanding what voters are thinking - clarification of politician positions, such as what public figures support or oppose, enhances the quality of information that voters have accesses to. Automatic analysis of the opinions also helps politicians with the most updated opinions people submit about pending policy or government-regulation proposals (Cynthia Farina, Claire Cardie, Thomas Bruce [6], Namhee Kwon, Stuart W. Shulman [19]).

3 Challenges

Consider the following sentence from Mark Twain: "Jane Austen's books madden me so that I can't conceal my frenzy from reader". Regular lexicon methods can't help in this case because "madden" and "frenzy" suggests negative sentiment, while the whole sentiment should be positive. In order to solve the problem of sentiment analysis completely, it is necessary to consider the whole context in an article rather than purely key words.

Lets consider another example from Liu [16]:

```
(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...
```

There are several different opinions here. The sentence (2) expresses a positive sentiment about the iphone as general. The sentences (3) and (4) also express positive opinions but about the phone's screen and voice quality. The sentence (5) and (6) express negative opinions of the author's mother on the author and the phone's price. As we can see, sentiment detection is more complicating as in order to detect the opinions of the author on iphone, in each phrase which contains opinions, we need to detect if the opinion holder is the author himself and if he is talking about either the iphone or its features. Therefore the problem of sentiment detection potentially contains difficult subproblems which are entity detection, feature extraction and subjectivity and objectivity classification of opinions.

3.1 Named Entity Recognition (NER)

This is a hard problem that has been investigated widely by researchers around the world. State-of-the-art NER systems for English produce near-human performance Wikipedia [33]. The best systen entering MUC-7 scored 93.39% of f-measure while human annotator scored 97.60% and 96.95% Marsh et al. [17]. However most of NER systems work well in some specific domains and need to be retrained for other domains.

This problem also contains the problems of feature extraction and synonym grouping WebKnox [31]. This part of the problem has remained a major challenge even though there are many attempts to solve it.

3.2 Opinion orientation classification

The most popular approach to this problem is to use opinion words such as *good*, *bad*, *poor*, *brilliant* to predict the sentiment in the context. However this approach is brittle because we need different sets of opinion words for different domain for it to work efficiently. Liu [16] gives an example about a phone battery: "The battery is long", and its camera: "This camera takes a long time to focus". Even though the word "long" is used in both sentences, the former implies a positive sentiment while the latter implies a negative sentiment.

Another the problem with this approach is that it is not easy to build a lexicon that works efficiently. There are unlimited number of expressions that people use to express opinions. Pang et al. [24] show that there are words that we might not think they indicate any sentiment but they appear more frequently in a certain class, i.e "still" surprisingly mainly appears in the positive class.

We also need to build different lexicons for different domains because a bag of words in this domain is positive but can be negative in another domain. For example "go read the book" most likely indicates positive sentiment for book reviews, but negative sentiment for movie reviews.

4 The approaches

4.1 Sentence/phrase-level sentiment analysis

4.1.1 Lexicon based

A typical approach to sentiment analysis is to use a lexicon of positive and negative words and phrases to identify the overall sentiment in a sentence or an article.

Unsupervised approaches

Hatzivassiloglou and Wiebe [9], Turney [30], Pan et al. [21] built a lexicon using some function based on positive, negative polarity and subjectivity with in it.

Hu and Liu [11] use combination of this method and the sentiment polarity of the previous sentence if the scoring function does not indicate a deterministic classification of a given sentence.

Hatzivassiloglou and McKeown [8] use a semi-supervised method to build such a lexicon. Their idea is to classify words by searching for their relationship by looking at conjunctions such as "elegant but over-priced", or "clever and informative". In Hatzivassiloglou and McKeown [8], after clustering, they simply assign the highest average frequency cluster as "positive cluster". The classificaiton precision is more than 90%. In some other work, some *seed words* which are known as possitive or negative are used to predict sentiment orientation of other words in the same clusters or using Wordnet-defined relations.

Pan et al. [21] propose a spectral feature aligning (**SFA**) algorithm to align domain specific words from different domains into unified clusters, with the help of domain-independent words as a bridge. Compared to previous approaches, **SFA** can discover a robust representation for cross-domain data by fully exploiting the relationship between the domain-specific and domain-independent words via simultaneously co-clustering them in a common latent space.

Semi-supervised approaches

The idea is to use the output of an initial simple classifier to feed labeled data into a supervised trainer. The trained models learn certain patterns from the input data and hence can classify a wider set of inputs.

Riloff and Wiebe [27] use an initial classifier to learn patterns of subjective expressions (an interesting example is the noun "fact", as in "The fact is...", exhibits high correlation with subjectivity).

Kaji and Kitsuregawa [12] use a similar method to learn patterns of HTML documents with polarity labels.

Qiu et al. [25] identify relations that link opinion words and targets using a dependency parser and then utilise to expand the initial opinion lexicon and to extract targets based on bootstrapping (**double propagation**). The results show this approach outperforms these existing methods significantly.

Similar work can be found at [32], [28].

Supervised approaches

However lexicon based approaches are not that trivial. Pang et al. [24] built a lexicon applying machine learning techniques based on unigram models that can achieve over 80% in accuracy.

A comparison of supervised and unsupervised methods can be found in [4].

4.2 Sentiment summarization

4.2.1 Supervised approaches

Beineke et al. [3] try to detect the summary quotation of the sentiment in a review from `www.rottentomatoes.com`. They experimented on two learning methods e.g. Naive Bayes and Regularized Logistic Regression to detect the summary quotation basing on some information like the length and location of sentences, dictionary words in the sentences and some other combinations. However the result is not good and the correct percentages of both methods are around or lower than 25%. It does not take into account that the sentiment in a review is expressed in not only one sentence.

Martineau and Finin [18] present a technique named **Delta TF-IDF** which is an intuitive general purpose technique to efficiently weight bag of word scores before classification. The authors then use SVM to train their models for and show that their method improve accuracy for three well known data sets.

Keefe and Koprinska [13] introduce two new feature selection methods and three new feature weighting methods with Naive Bayes and Support Vector Machine. Their methods prove that it is possible to maintain a state-of-the-art classification accuracy of 87.15% while using less than 36% of the features.

Abbasi et al. [1] develop a hybridised genetic algorithm (Entropy Weighted Genetic Algorithm - **EWGA**) that incorporates the information gain heuristic for feature selection. **EWGA** is designed to improve performance and get a better assessment of the key features. The experimental results using **EWGA** with SVM indicate high performance levels, with accuracy over 95% on the benchmark data set and over 93% for both the U.S. and Middle Eastern forums.

4.2.2 Classification based on relationship information

Relationships between discourse participants

Agrawal et al. [2] observe on manual examination of 100 responses in newsgroup that the relationship between two individuals in the "responded-to" network is more likely to be antagonistic. Assuming "respond-to" links imply disagreement, they effectively classify users into opposite camps via graph partitioning, outperforming methods that depend solely on the textual information within a particular document.

Relationships between product features

Snyder and Barzilay [29] utilize agreement information in a task where one must predict ratings for multiple aspects of the same item. They then construct a linear classifier to predict whether all aspects of a product are given the same rating, and combine this prediction with that of individual-aspect classifiers so as to minimize a certain loss function. [29]

Ding et al. [7] study the problems of *entity discovery* and *entity assignment*. *Entity discovery* is the problem of detecting entities explicitly mentioned in sentences while *Entity assignment* is the problem of detecting entity mention by pronouns and language conventions. *Entity discovery* is based on pattern discovery and *entity assignment* is based on mining of comparative sentences. Experimental results demonstrate the effectiveness on both forum posts and a commercial setting.

Sentiment analysis of conditional sentences

Narayanan et al. [20] study sentiment analysis of conditional sentences. Sentiment in conditional sentences are different from normal sentences and hard to determine. For example, in the sentence, *if* *your Nokia phone is not good, buy this great Samsung phone*, the author is positive about *Samsung phone* but does not express an opinion on *Nokia phone*. Conditional sentences can start with also *unless, even if, until, as long as, assuming, supposing, in case, only if*. The authors use SVM to train data from 5 diverse domains and achieve remarkable results.

4.3 Multi-class classification

4.3.1 Relationships between classes

Standard multi-class categorization focuses on capturing the distinct features present in each class, and ignore the fact that "5 stars" is much more like "4 stars" than "2 stars". [22] observe that one-vs-all multi-class categorization scheme can outperform regression for a three-class classification problem (positive, neutral, and negative), perhaps due to each class exhibiting a sufficiently distinct vocabulary, but for more fine-grained classification, regression emerges as the better of the two.

Pang and Lee [22] formulate rating inference as a metric label problem (Kleinberg and Tardos [15]), so that a natural notion of distance between classes ("2 stars" and "3 stars" are more similar to each other than "1 star" and "4 stars" are) is captured explicitly. More specifically, an optimal labelling is computed that balances the output of classifiers that considers items in isolation with the importance of assigned similar labels to similar items.

5 Conclusion

Even though an emerge of interest in sentiment analysis have attracted worldwide researchers from both institutions and corporates, a state-of-the-art sentiment analysis machine's performance is still far away from what a human can do, especially for sophisticated documents which include sarcasm, conditional sentences, implications and comparisons. Many sub-problems have been addressed and attempted. The scope of each of these sub-problems is huge and, thus, sentiment analysis still remains a very challenging problem to researchers. However I am positive about the future of sentiment analysis as demanded from industry and its commercial potential. The availability of online user-generated data also enables statistical approaches and, hence, fosters the research of sentiment analysis.

References

[1] Ahmed Abbasi, Hsinchun Chen, and Arab Salem. Sentiment Analysis in Multiple Languages : Feature Selection for Opinion Classification in Web Forums. *Text*, 2007.

[2] Rakesh Agrawal, Sridhar Rajagopalan, Ramakrishnan Srikant, and Yirong Xu. Mining newsgroups using networks arising from social behavior. *Proceedings of the twelfth international conference on World Wide Web - WWW '03*, page 529, 2003. doi: 10.1145/775224.775227. URL `http://portal.acm.org/citation.cfm?doid=775152.775227`.

[3] P Beineke, T Hastie, C Manning, and S Vaithyanathan. Exploring Sentiment Summarization. In Yan Qu, James Shanahan, and Janyce Wiebe, editors, *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text Theories and Applications*, volume 07, pages 12–15. AAAI Press, 2004. URL `http://www.aaai.org/Papers/Symposia/Spring/2004/SS-04-07/SS04-07-003.pdf`.

[4] P. Chaovalit and L. Zhou. Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 00(C):112c–112c, 2005. doi: 10.1109/HICSS.2005.445. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1385466`.

[5] ComScore. Online Consumer-Generated Reviews Have Significant Impact on Offline Purchase Behavior, 2007. URL `http://www.comscore.com/Press_Events/Press_Releases/2007/11/Online_Consumer_Reviews_Impact_Offline_Purchasing_Behavior`.

[6] Erica Wagner Cynthia Farina, Claire Cardie, Thomas Bruce. Using Natural Language Processing to Improve eRulemaking [Project

Highlight]. *Proceedings of the 7th Annual International Conference on Digital Government Research*, pages 4–5, 2006.

[7] Xiaowen Ding, Bing Liu, and Lei Zhang. Entity discovery and assignment for opinion mining applications. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*, page 1125, 2009. doi: 10.1145/1557019.1557141. URL `http://portal.acm.org/citation.cfm?doid=1557019.1557141`.

[8] Vasileios Hatzivassiloglou and Kathleen R. McKeown. Predicting the semantic orientation of adjectives. *Proceedings of the 35th annual meeting on Association for Computational Linguistics -*, pages 174–181, 1997. doi: 10.3115/976909.979640. URL `http://portal.acm.org/citation.cfm?doid=976909.979640`.

[9] Vasileios Hatzivassiloglou and Janyce Wiebe. Effects of Adjective Orientation and Gradability on Sentence Subjectivity, 2000.

[10] John B Horrigan. Online Shopping Internet users like the convenience but worry Findings. *Director*, 2008.

[11] Minqing Hu and Bing Liu. Mining Opinion Features in Customer Reviews. *Science*, 1998.

[12] Nobuhiro Kaji and Masaru Kitsuregawa. Automatic construction of polarity-tagged corpus from HTML documents. *Proceedings of the COLING/ACL on Main conference poster sessions -*, (July):452–459, 2006. doi: 10.3115/1273073.1273132. URL `http://portal.acm.org/citation.cfm?doid=1273073.1273132`.

[13] Tim O Keefe and Irena Koprinska. Feature Selection and Weighting Methods in Sentiment Analysis. *Methods*, 2006.

[14] Peter Kim. Brand Monitoring, Q3 2006. 2006.

[15] Jon Kleinberg and Éva Tardos. Approximation algorithms for classification problems with pairwise relationships: metric labeling and Markov random fields. *Journal of the ACM*, 49(5):616–639, September 2002. ISSN 00045411. doi: 10.1145/585265.585268. URL `http://portal.acm.org/citation.cfm?doid=585265.585268`.

[16] Bing Liu. Sentiment Analysis : A Multi-Faceted Problem. *Science*, (1), 2010.

[17] Elaine Marsh, Dennis Perzanowski, and Ralph Grishman. MUC-7 EVALUATION OF IE TECHNOLOGY : Overview of Results MUC-7 Program Committee. *Program*, (April), 1998.

[18] Justin Martineau and Tim Finin. Delta TFIDF : An Improved Feature Space for Sentiment Analysis. *Artificial Intelligence*, pages 258–261, 2009.

[19] Eduard Hovy Namhee Kwon, Stuart W. Shulman. Multidimensional Text Analysis for eRulemaking. In *the 2006 international conference on Digital government research ACM*, pages 157–166. San Diego, California, 2006.

[20] Ramanathan Narayanan, Bing Liu, and Alok Choudhary. Sentiment analysis of conditional sentences. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing Volume 1 - EMNLP '09*, page 180, 2009. doi: 10.3115/1699510.1699534. URL `http://portal.acm.org/citation.cfm?doid=1699510.1699534`.

[21] Sinno Jialin Pan, Xiaochuan Ni, Jian-Tao Sun, Qiang Yang, and Zheng Chen. Cross-domain sentiment classification via spectral feature alignment. *Proceedings of the 19th international conference on World wide web - WWW '10*, page 751, 2010. doi: 10.1145/1772690.1772767. URL `http://portal.acm.org/citation.cfm?doid=1772690.1772767`.

[22] Bo Pang and Lillian Lee. Seeing stars : Exploiting class relationships for sentiment categorization with respect to rating scales. In *ACL*, number 1, 2005.

[23] Bo Pang and Lillian Lee. Opinion Mining and Sentiment Analysis. *Foundations and TrendsÂ® in Information Retrieval*, 2(1â2):1–135, 2008. ISSN 1554-0669. doi: 10.1561/1500000011. URL `http://www.nowpublishers.com/product.aspx?product=INR&doi=1500000011`.

[24] Bo Pang, Lillian Lee, Harry Rd, and San Jose. Thumbs up ? Sentiment Classification using Machine Learning Techniques. *Language*, 2002.

[25] Guang Qiu, Bing Liu, and Chun Chen. Opinion Word Expansion and Target Extraction through Double Propagation. *Computational Linguistics*, (July), 2010.

[26] Lee Rainie and John Horrigan. Election 2006 Online The number of Americans citing the internet as the source of most of their political news and information doubled since the last mid- term election Findings. *Director*, 2007.

[27] Ellen Riloff and Janyce Wiebe. Learning extraction patterns for subjective expressions. *Proceedings of the 2003 conference on Empirical methods in natural language processing* -, pages 105–112, 2003. doi: 10.3115/1119355.1119369. URL `http://portal.acm.org/citation.cfm?doid=1119355.1119369`.

[28] Ellen Riloff, Janyce Wiebe, and Theresa Wilson. Learning subjective nouns using extraction pattern bootstrapping. *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003* -, pages 25–32, 2003. doi: 10.3115/1119176.1119180. URL `http://portal.acm.org/citation.cfm?doid=1119176.1119180`.

[29] Benjamin Snyder and Regina Barzilay. Multiple Aspect Ranking Using the Good Grief Algorithm. 2007.

[30] Peter D Turney. Thumbs Up or Thumbs Down ? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Computational Linguistics*, (July):417–424, 2002.

[31] WebKnox. Named Entity Definition, 2011. URL `http://www.webknox.com/blog/2010/09/named-entity-definition/`.

[32] Janyce Wiebe and Ellen Riloff. Creating Subjective and Objective Sentence Classifiers from Unannotated Texts. 2005.

[33] Wikipedia. Named Entity Recognition. URL `http://en.wikipedia.org/wiki/Named_entity_recognition`.