Opinion Mining Survey

trunghlt@gmail.com

1 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 invididual customers and corporations.

From the surveys ComScore [4], Horrigan [7], 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 retaurants, 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 ragarding a product or service.

For political information, Rainie and Horrigan Rainie and Horrigan [17] studied the 31% of Americans - over 60 million people that were 2006 capaign 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 [10] 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 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 [12]:

```
(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.

2.1 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 [24]. The best systen entering MUC-7 scored 93.39% of f-measure while human annotator scored 97.60% and 96.95% Marsh et al. [13]. 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 as they are just relationships of detected entities [22]. This part of the problem has remained a major challenge even though there are many attempts to solve it.

2.2 Opinion orientation classification

This is the scope where this theis attempts to solve by assuming entity identifications in contexts are given (using available NER systems or manual annotations).

This 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 [12] 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. [16] show that there are words that we might not think they indicate any sentiment but they appears more frequently in a certain class, i.e "still" surprisingly mainly appears in posstive class.

We also need to build different lexicons for different domains because a bag of words in this domain is postive 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.

3 The approaches

3.1 Sentence/phrase-level sentiment analysis

3.1.1 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 [6], Turney [21], Pan et al. [14] built a lexicon using some function based on positive, negative polarity or subjectivity with in it.
- Hu and Liu [8] 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 [5] 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 [5], 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 same clusters or using Wordnet-defined relations.

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.

- [18] 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.
- [9] use a similar method to learn patterns of HTML documents with polarity labels.

- Similar work can be found at [23], [19].

Supervised approaches

- However lexicon based approaches are not that trivial. Pang et al. [16] 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 [3].

3.2 Sentiment summarization

3.2.1 Supervised approaches

- Beineke et al. Beineke et al. [2] 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. '

3.2.2 Classification based on relationship information

Relationsips between discourse participants

- [1] 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

- [20] utilize agreement information in a task where one must predict ratings for multiple asects 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 classifers so as to minimize a certain loss function. [20]

3.3 Multi-classes classification

3.3.1 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". [15] 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 sufficently distinct vocabulary, but for more fine-grained classification, regression emerges as the better of the two.

- [15] formulate rating inference as a metric label problem ([11]), 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 labeling is computed that balacnes the output of classifier that considers items in isolation with the importance of assign similar lables to similar items.

4 Conclusion

References

[1] 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`.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

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

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

[9] 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.

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

[11] 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.

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

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

[14] 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.

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

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

[17] 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.

[18] 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.

[19] 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.

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

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

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

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

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