The Art of Opinion Mining and Its Application Domains: - A Survey

Dipali V. Talele
ME[CSE],GHRIEM

Chandrashekhar D. Badgujar
Asst. Prof. in Dept. of CSE, GHRIEM

ABSTRACT

The advent of Web 2.0 and Social media content has stirred much excitement and created abundant opportunities for understanding the opinions of the general public and consumers toward social events, political movements, company strategies marketing campaigns, and product preferences. Individuals, businesses and government can now easily know the general opinion prevailing on a product, company or public policy. This paper critically evaluates existing work, presents an opinion mining framework and exposes new areas of research in opinion mining. Overall item sentiment can be expressed based on its sentiment words in general or by specifically identifying its features and the opinions being expressed about them. This leads us to the motivation of the framework for opinion mining and categorizing current literature in such a manner as to make clear, research opportunities. The freedom offered by the web as a platform for presenting opinions on any subject brings with it many new opportunities.

Keywords

Opinion mining, Sentiment Classification, supervised learning, unsupervised learning,.

1. INTRODUCTION

We As e-commerce is becoming more and more popular and interactive, Industry or manufacturing companies that produce new products selling them on the Web 2.0 often ask their customers to review the products which they have purchased to know how their customers feel about these products and this information can be acquired by studying opinions from review portals [2]. At the same time, users or consumers want to know which product to buy so they also read reviews and try to make their decisions. It has been seen that online opinions are getting popular day by day and these opinions represent wealth of information which can be beneficial for the industry as well as consumers [1]. The number of reviews can be in hundreds or even thousands for a popular product. This makes it difficult for a potential customer to read them or to make an informed decision on whether to purchase the product. It also makes it difficult for the manufacturer of the product to keep track of customer opinions. However doing this manually is only possible to a certain extent and time consuming job. As an example, manufacturing organizations prefer information in a format that is easier to use, so automating this process is very useful [3].This is where opinion mining comes to picture. In the web, opinions can be expressed in the form of text, image, audio or video data.

This paper is biased towards text mining as this is a widely researched area. Since this field is very new and much work is currently being done in this area, this paper critically evaluates existing work, presents an opinion mining framework and exposes new areas of research in opinion mining.

Opinion mining aims to extract opinions from information sources (user generated content or user generated media) such as reviews, and present them to the users in a user friendly manner (graphically for example).

This paper is structured into various sections. Section II is the Related Work that presents the existing and related work on opinion mining, depicting the major contribution and limitations of existing techniques. Section III follows with the evaluation of researches. Section IV shows application domains for opinion mining. We conclude our paper with Section V.

2. RELATED WORK

Information available as text format can be broadly classified into two main categories, facts and opinions. Facts are generally objective statements about entities and events. But opinions are subjective statements that reflect people's sentiments or perceptions about the entities and events which is the area of interest for this work. Most of the existing research like information retrieval, Web search, and other text mining and natural language processing tasks on text information processing has been focused on mining and retrieval of factual information. Only a little work has been done on the processing of opinions until recently. Though the number of research interest in this area is growing fast.

Research on opinion mining started with identifying opinion (or sentiment) bearing words, e.g., great, amazing, wonderful, bad, and poor. Many researchers have worked on mining such words and identifying their semantic orientations (i.e., positive or negative). Hatzivassiloglou, V. and Mckeown,K;(1997) in [4], the authors identified several linguistic rules that can be exploited to identify opinion words and their orientations from a large corpus. This method has been applied, extended and improved in [5,6]. In [3,7], a bootstrapping approach is proposed, which uses a small set of given seed opinion words to find their synonyms and antonyms in WordNet (http://wordnet.princeton.edu/). The next major development is sentiment classification of product reviews at the document level [8, 9, 10]. The objective of this task is to classify each review document as expressing a positive or a negative sentiment about an object (e.g., a movie, a camera, or a car). Several researchers also studied sentence-level sentiment classification [6, 11], i.e., classifying each sentence as expressing a positive or a negative opinion. The model of feature-based opinion mining and summarization is proposed in [5, 12]. This model gives a more complete formulation of the opinion mining problem. It identifies the

key pieces of information that should be mined and describes how a structured opinion summary can be produced from unstructured texts. The problem of mining opinions from comparative sentences is introduced in [13, 14].

Recent work of Machova ,K. Penzes,T.(2012) in [19] relates to the problem of opinion analysis. This opinion analysis focuses on opinion mining from web forums. The paper introduces design of the method for extraction of texts from web forums discussions. These texts are carriers of opinions on some subject, which under review. In opinion mining of product reviews, one often wants to produce summery of opinions based on features/attributes. However, for the same feature, people can express it with different words and phrases. To produce meaningful summary, these words and phrases, which are domain synonyms, need to be grouped under the same feature group. Zhai,Z; Liu,B;(2011) in [20] proposes a constrained semi-supervised learning method to solve the problem..

3. EVALUATION OF RESEARCHES

This section of the paper uncovers areas that have not been studied much in the existing literature. The comparison of items and features of products are important and significant in opinion mining of online customer reviews for the benefit of consumers, product distributors and product manufacturers. Having a visual chart or table which depicts the advantages and disadvantages of competing products and services at a glance is very useful. It allows the advantages and disadvantages of a product to its direct competing products to be quickly known, instead of having to spend huge amount of time going through infinite reviews on various websites. It also provides convenience and is less time consuming for product manufacturers to gather market intelligence and product benchmarking information; and provides consumers with a side-by-side and feature-by feature comparisons of consumer opinions on competing products [5]. The visuals allow potential customers to quickly figure out the information that they need before making a purchase decision, and relieves them from the mundane task of going through numerous reviews, especially in today's quick paced and time-pressed society where time is money.

Yee W.LO,DEBI Institute (2009),In [15],illustrate in Fig. 1, most of the research so far is focused on the task of feature extraction and sentiment classification on features. Little work has been done on item extraction, sentiment classification of reviews, sentiment classification on items, and comparison of items and features.

Authors provides us Quicker and more convenient way to pick up areas that are most researched on and areas that are least researched on from the visual chart provided in Fig. 1compared to the reading the illustration in words, which is very useful to our work on survey of related work in opinion mining.z
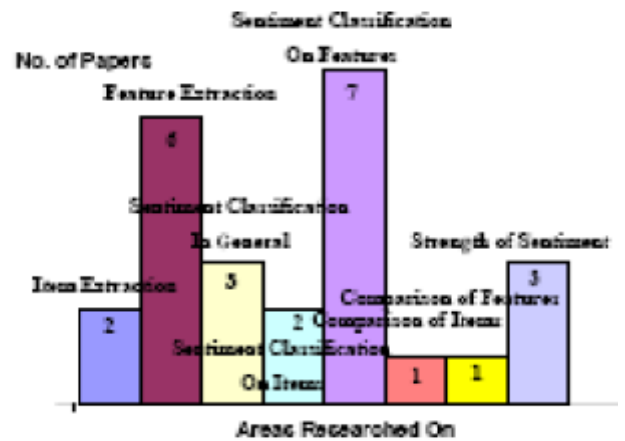

Fig 1: Areas of Opinion mining the papers Focused on

4. OPINION MINING APPLICATION DOMAIN

Opinion mining can be used in various fields to meet varied purposes. In this section, we take the opportunity to present some of the common ones [18].

4.1 Shopping

Perhaps the most popular use of opinion mining is decision support for consumers. Consumers are actively involved in comparison shopping over the internet. Popular web sites like amazon.com allow customers to express their opinions on their websites. Customers can easily view the opinions for products and identify how features between products compare with each other. In some instances, after an opinion has been mined and processed, knowledge is presented to the user in a graphical format for easy comparison of product features. We will consider the comments on an electronic product from an online shop.

"I needed a high-powered laptop for my business needs. Dell offered a variety of products that met my requirements. Their product information was concisely and completely explained, which made my selection process very easy. Their website was easy to use and attractively presented. The merchandise was delivered on time, as per their promise. Shipping charges were very reasonably priced. I was able to take care of all my needs online, thereby not having to use their telephone support service. Their online help and support was excellent. I would recommend Dell Small Business to anyone. [11]".

We note that the opinion is about Dell Small Business and the various features being talked about are delivery time, shipping charges, support and web site navigation.

4.2 Entertainment

Movie goers and home TV viewers can quickly access the opinion on recent releases and popular movies and programs. Currently, there is the internet movie database (IMDB) which provides online reviews for movies as well as TV programs. This acts as a guide for people who are unsure about which movies to watch. We present an extract from the IMDB.

"Christopher Nolan's second bundle of joy "The Dark Knight" EXCEEDED all of my expectations!!! I can HONESTLY tell you that: as good as Jack Nicholson was in Batman'89 he is CHILD'S PLAY compared to this Joker. He is sadistic, psychotic, and downright SCARIER and PSYCHOLOGICALLY disturbing than the previous incarnation of The Clown Prince of Crime and Ledger gives it his all to do him justice. The action is great, and the plot is deeper and engrossing. [11]"

From the previous opinion, we observe that capital letters and exclamation signs are being used to emphasize emotions. Furthermore, we observe that the first comment is positive and refers directly to the movie, "The Dark Knight". However, subsequent statements refer to the actors and it is their attributes that are being mentioned. The last statement mentions the film once again. This kind of opinion, which revolves between the actors and the movie, is relatively simple for a human reader to understand but not so for a machine. Therefore, this presents some complexity to machine language learning. It is evident that two objects are being described, the movie and actors. Although words with a negative connotation (sadistic, psychotic, and disturbing) are being used to express positive aspects of the movie, it does not mean that the film is not highly recommended but rather, just an illustration of the complexity that exists for machine language learning. Moreover, most opinions about movies are expressed in this way.

4.3 Government

Governments can mine the prevailing opinions on public policy. Election candidates can become more knowledgeable about specifics of the opinion poll. This knowledge can assist politicians to identify where their strengths and weaknesses lie according to their electorate. Consider the following political opinions that have been expressed.

"Expect more inflation. More unemployment. Really, we need some better selection process. Who chooses these people? They make history by raising rates for the first time in the lead up to an election. I have no confidence in the published figures." [12]

A quick glance at these terms indicates a sense of dissatisfaction among the electorate. Furthermore, key areas of concern are addressed in terms of what is lacking and what the expectations are.
Issues that deal with public policy normally categorize voters into one of three groups, for, against or neutral. A good example is the statement, *"I think this all seems extremely harsh. Boredom, if anything, is a sign of intelligence." [12]* A statement of this kind makes it clear that the opinion is for the motion. The advantage of opinion mining over traditional opinion polls like telephone polls is that it can be determined why electorates are for or against a proposal. Most web sites, particularly those whose fundamental objective is to provide news, have a facility for web users to express their opinions on their websites.

4.4 Research and Development

Product reviews can be used by manufacturing companies to improve features and provide a platform for innovation. Web based applications could offer platforms for customers to design products and submit the designs to the manufacturing companies. An approach of this nature could significantly assist in establishing features that are liked by customers. Consider the following review for an electronic product,

"The click wheel is HORRIBLE and completely lacks response and sensitivity." [13]

This is a negative opinion being expressed about the click wheel. The use of upper caps signifies to the reader the extent of disappoint. If opinion mining is able to detect emotions of this kind being expressed in evaluative text, it will prove to be very beneficial. This will act as an indicator on how the product has been received by a consumer. However, after expressing negative opinions on the product features, a statement such as *"although I am really disappointed this is probably still the best high capacity music player on the market [13]".* This positive statement indicates to the R & D department and marketing departments that the music player is still the best in the market and it is the high capacity which is favored by customers. Further examples can be obtained from web sites like bizrate.com and epinion.com.

4.5 Marketing

Companies can now make savings on marketing expenses by requesting for reviews on their websites and peer review websites. This eliminates the need for business consultants to conduct surveys as companies can now have all the data they need online. The advent of the internet has brought along with it new ways of marketing. Family and friends can now recommend products/services to each other or seek more knowledge about a product or service before committing themselves. It is analogous to the traditional word of mouth marketing of products and services. To encourage postings and recommendations among peers, marketers normally offer incentives like discounts for recommendations that turn into purchases [14]. Let us consider the following recommendation for a tourism resort,

"It is a land of contrasts and majesty, Africa at its most wild and unexplored [12]."

Positive opinions of this kind can improve a products rating. WAYN (where are you now) is a social network that lets friends track each other during their travels, posting photographs and describing their adventures and experiences.

4.6 Education

In e-learning systems, user's opinions can be used to evaluate academic institutions and academics. Academics can know the sentiment on courses based on sentiment analysis of opinions expressed by students. This can help to improve service delivery and bolster marketing campaigns. Unit coordinators can know what students think about their team members and tutors by requesting them to provide online reviews as a part of course requirement. Curtin University of Technology offers units in which students must submit weekly peer reviews and also offers discussion forums. Such frameworks act as rich sources of user generated content that can be mined. Research findings by [15] reveal that e-learning systems adoption by tertiary institutions is still in its early

phases. Members of the legal fraternity use legal blogs (blawgs) to express opinions. According to [16], students post their experiences online and law professors provide comprehensive analysis of court cases and post their findings to the community. Legal researchers can benefit from different opinions that are posted for a legal issue. Let us consider a sample opinion from an academics point of view.

"My research is improving my analytical, problem solving skills and ability to plan my own work. The feedback from the supervisor is valuable. The computing facilities are excellent. However, the monthly down load quota is too low for conducting research without being exceeded. [17]"

The overall opinion is for research (object) and it is positive. The features to extract an opinion on would be the supervisor, computing facilities and download quota. When data from different institutions of this kind is recorded and made readily available, it can be used for comparison purposes.

5. CONCLUSION

This paper presents the current state of the art in opinion mining. On the one hand, we discover that research has been concentrated towards finding out the sentiment on an item and classifying it as thumbs up or down. On the other hand, little research has been undertaken towards web queries that provide knowledge about an explicit item and its comparison with other items. Although reviewers would benefit more from a comparison between items and/or item features, our research findings reveal a different reality. To the best of our knowledge, minimal research has been done in this area. Opinion mining has many application domains including science & technology, entertainment, education, politics, marketing, accounting, law, research and development. We hope that researchers will explore these areas in the future. Since there is possibility of employing various methods and combining them to be used in single situation to achieve better and more accurate outcome and results are also demonstrate, in our future work we plan to further improve and refine existing techniques.

6. REFERENCES

[1] M. Hu and B. Liu, "Mining Opinion Features in Customer Reviews," presented at 19th National Conf. on Artificial Intelligence, 2004.

[2] Wilson T, "Just how mad are you? Finding strong and weak opinion clauses," presented at In Proc. 19th National Conference on Artificial Intelligence. 2004.

[3] L. B. Hu M, "Mining and summarizing customer reviews.," presented at Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining 22:2525, 2004.

[4] Hatzivassiloglou, V. and McKeown, K., "Predicting the Semantic Orientation of Adjectives.", ACLEACL' 97, 1997.

[5] Ding, X., Liu, B. and Yu, P. "A Holistic Lexicon-Based Approach to Opinion Mining." Proceedings of the first ACM International Conference on Web search and Data Mining (WSDM'08), 2008. Conference on Empirical Methods in Natural Language Processing (EMNLP'06), 2006.

[6] Popescu, A.-M. and Etzioni, O. " Extracting Product Features and Opinions from Reviews." Proceedings of the 2005, Conference on Empirical Methods in Natural Language Processing (EMNLP'05), 2005.

[7] Kim, S. and Hovy, E. "Determining the Sentiment of Opinions." Proceedings of the 20th International Conference on Computational Linguistics (COLING'04), 2004.

[8] Wiebe, J. and Riloff, E. "Creating Subjective and Objective Sentence Classifiers from Un-annotated Texts." Proceedings of International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'05), 2005

[9] Wilson, T., Wiebe, J. and Hwa, R. " Just How Mad Are You? Finding Strong and Weak Opinion Clauses." Proceedings of National Conference on Artificial Intelligence (AAAI'04), 2004.

[10] Liu, B., Hu, M. and Cheng, J. "Opinion Observer: Analyzing and Comparing Opinions on the Web." Proceedings of International World Wide Web Conference (WWW'05), 2005.

[11] Pang, B., Lee, L. and Vaithyanathan, S. " Thumbs up? Sentiment Classification Using Machine Learning Techniques." Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP'02), 2002.

[12] Turney, P. "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews." ACL'02, 2002.

[13] Ganapathibhotla, G. and Liu, B. "Identifying Preferred Entities in Comparative Sentences." Proceedings of the 22nd International Conference on Computational Linguistics (COLING'08), 2008.

[14] Jindal, N. and Liu, B. "Mining Comparative Sentences and Relations." Proceedings of National Conference on Artificial Intelligence (AAAI'06), 2006.

[15] Yee W.LO,DEBII and Vidyasagar POTDAR,DEBII,"A Review of Opinion Mining and sentiment Classification Framework in Social Networks",2009 3rd IEEE International conference on Digital Ecosystems and Technologies.

[16] Hsinchun Chen, University of Arizona, "AI and opinion mining − part II", 1541-1672/10/$26.00 © 2010 IEEE IEEE INTELLIGENT SYSTEMS Published by the IEEE Computer Society

[17] Wei Huang, Xin Chen, Haibo Wang, "Product Information Retrieval Based on Opinion Mining", 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010),978-1-4244-5934-6/10/$26.00 ©2010 IEEE

[18] Haji Binali, Vidyasagar Potdar, Chen Wu," A State Of Art Of Opinion Mining And Its Application Domains", Digital Ecosystem and Business Intelligence Institute, Australia

[19] Machova , K.; Penzes, T.," Extraction of Web Discussion texts for opinion analysis", Applied Machine Intelligence and Informatics(SAMI) 2012-IEEE

[20] Zhai, Z; Liu,B.; Wang, J," Product Feature Grouping for Opinion Mining Using Soft-Constraints and EM", Intelligent Systems, IEEE-2011