Review on Social Media Sentimental/ Opinion Analysis

Mr. Ram B. Ghayalkar1
Assistant Professor,
Shri R. L. T. College of Science,Akola

Dr. D. N. Besekar2
Associate Professor,
Shri Shivaji College, Akola

Dr. G. S. Wajire3
Associate Professor,
Shri Shivaji College, Akola

Abstract:
After introducing of social media sites, information is quite faster transferring from one place to other easily, but with this there are some pros and cons. Sometimes wrong information of rumors spread very fast as well as people sending their views or opinion about person, any product etc, so here opinion or sentimental analysis plays very important role. In this paper research review of sentimental/ opinion data analysis is used to identify and analysis of comments or opinions of different people from different sites for person, issues, events, topics and their attributes or some business review which will then useful in decision making.
Keywords: Information retrieval, sentiment analysis, opinion mining, data mining

Introduction:

Sentiment analysis or opinion mining is the computational study of people's opinions, attitudes,appraisals, and emotions toward entities, individual person, issues, events, topics and their attributes.
"What other people are thinking" is always an essential piece of information during the decision-making process in business intelligence.

For example, In businesses there is importance of public or consumer opinion about product or services.

Now use of social media in everywhere(i.e., reviews, forum discussions, blogs and social networks) on the internet, individuals and organizations are increasingly use of public opinions in these media for their decision making.

It is a very important method in today's world where our maximum work is carried on internet be it communication, reading news, blogs placing reviews about a company, product/software or person. So it becomes very important to detect the exact meaning of the sentence written or else it may lead to disastrous (in many cases completely opposite) understanding of the issue. The essential issues in sentiment analysis are to identify how sentiments are expressed in texts and whether the expressions indicate affirmative (favorable) or negative (unfavorable) perspective toward the subject.[3][4]

Opinion Mining or Sentiment analysis involves creating a system to explore user's opinions made in blog posts, comments, reviews or tweets, about the product, policy or a topic [6].

Components of Opinion Mining
There are mainly three components of Opinion Mining [7]:
• **Opinion Holder/ source**: Opinion holder is the holder of a particular opinion; it may be a person or an organization that holds the opinion. In the case of blogs and reviews, opinion holders are those persons who write these reviews or blogs.
• **Opinion Object**: Opinion object is an object on which the opinion holder is expressing the opinion.
• **Opinion Orientation:** Opinion orientation of an opinion on an object determines whether the opinion of an opinion holder about an object is positive, negative or neutral.
There are two main types of opinions: *regular opinions* and *comparative opinions*.

Different Levels/ Classification of Sentiment Analysis
In general, sentiment analysis has been investigated mainly at three levels [5].
• **Document level:** At this level is to classify whether a complete opinion document expresses a positive or negative sentiment.
• **Sentence level:** At this level the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion.
Objective sentences - that express factual information from sentences
Subjective sentences - that express subjective views and opinions.
• **Entity and Aspect level:** Aspect level performs fine-grained analysis. Aspect level was earlier called feature level or feature-based opinion mining and summarization.

Challenges In Opinion Mining

There are many challenges in Opinion Mining as follows:
- Domain-independence: Good in one domain and poor in other.
- Asymmetry in availability of opinion mining software.
- Detection of spam and fake reviews.
- Incorporation of opinion with implicit word/data and behavior data.
- Mixed Sentences and confusion in affirmative words.
- Way of Expressing the Opinion in different ways for difficult to identify.
- Use of Abbreviations and shortforms, Orthographics Words
- Typographical errors Sometimescause problems while extracting opinions.
- Natural language processing overheads: The natural language overhead like co-reference, ambiguity, Implicitness, inference etc. created difficulties in sentiment analysis tool [5][7][8].

As well as above challenges to analysis social media, there are also some other challenges to manage the information:

a.Scraping—social media data are accessible through APIs, but as major dataholders like facebook, Google, twitter, etc are making their 'raw' data accessibility are difficultfor commercial value of the data..

b.Data cleaning—Data cleaning is to cleaning of unstrutured textual data.

c.Data protection—It is related to security of data.

d.Data analytics—foreignlanguages, foreign words, slangs, spelling errors and natural languages. [21]

Methodology

1. **A)Dataset** –Any dataset on where to apply different technique to retrieving information likeSentiment Strength Twitter Dataset or any social / business site datasets.
2. **B)Pre-processing Techniques**
3. **Tokenization**This step breaks the large paragraphs called chunks of text is broken into tokens which are actually sentences. These sentences can further be broken into words.
4. **Normalization**There are many tasks performed simultaneously to achieve normalization. It includes the conversion of all text to either upper or lower case, eliminating punctuations and conversion of numbers to their equivalent words. This increases the uniformity of preprocessing on each text.
5. **Stemming**The stemming process is used to change different tenses of words to its base form this process is thus helpful to remove unwanted computation of words.
6. **Lemmatization**Lemmatization is the process of merging two or more words into single word
7. **Removing Stop Words**Stop words refer to most common words in the English language which doesn't have any contribution towards sentiment analysis.
8. **Noise removal**The datasets taken comes in raw form. We have applied manual cleaning of raw data along with the use of regular expression in NLP used to eliminate noises

C)Feature Extraction

1. **TF-IDF**The term frequency-inverse document frequency (also called TF-IDF), is a well-recognized method to evaluate the importance of a word in a document.
2. **Word N-grams features**
3. The word N-grams feature model is one of most simple and effective representation model for natural language analysis and Twitter sentiment analysis. N-Gram will form the features of text for supervised machine learning algorithms. Some studies have shown state-of-the-art performance for sentiment classification on Twitter data using a unigram model [12], [13].
4. **Twitter specific features**
5. The number of hashtags, emoticons, negation, POS and the presence of capitalized words are used as features.
6. **Word sentiment polarity score features**
7. The word sentiment polarity score is a lexicon-based sentiment feature, and some approaches [14], [15] commonly use it as a sentiment feature for tweet sentiment analysis.
   a. Polarity results for sentiment analysis are usually in the form ofpositive or negative, providing only single-dimensional information in sentiments. Compared to face-to-face communication, this analysis provides the least rich approach to capturing human affective information. To improve the quality of sentiment analysis output, based on the information richness [10]

8. **Word representation features**
9. Learning word vector representations from a large number of unannotated text corpora has recently been used in various natural language processing tasks. [16],[17]

D) Classification Algorithms

For classification of word following algorithm will be useful: Support Vector Machine(SVM), Logistic Regression, Naive Bayes, Decision Tree, K-Nearest Neighbour (KNN), Random Forest

E) Performance Parameters

For performance point of view two important parameter are Accuracy, Precision [18]

Machine Learning In Sentiment Analysis (Ontology Based)

Almost all Machine learning algorithms for sentiment analysis follow standard steps to classify the sentiments as explained below [19]:

1. **Noise Removal**: Cleaning the data to extract relevant data from irrelevant data which increases the ability of an algorithm to predict based on the training data set.
2. **Classification**: Categorizing the data into positive and negative class.
3. **Named Entity Recognition**: In order to predict actual meaning of the comments, it is obvious to extract entities like sender, receiver, and the aspects of their conversation and then classify them as positive or negative.
4. **Subjectivity Classification**: Sentences can be classified into subjective or objective. Subjective are expressions on any attributes, events with the properties. Objectives are opinions that describe feelings toward entities.
5. **Feature Selection**: The features can be unigrams, bigrams, or n-grams with or without punctuations.
6. **Sentiment Extraction**: It can be done using unsupervised and supervised machine learning algorithms such as extra trees classifier, random forest classifier, naïve Bayes, and support vector machine.
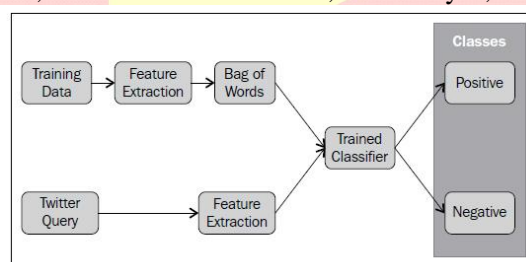
Ontology for Sentiment Analysis(SA)

Ontology-based Sentiment Analysismodel consists of classes, subclasses, objects, and their properties. The objects are stored in the database, and ontology model can be queried using query language like SPARQL to retrieve the data. Using ontology, a dynamic model can be developed that can predict objects regardless of the domain being used. Features of the domain are extracted by building ontology which helps in getting the refinedSentiment Analysis. [19]

Future Scope

A drawback of such information overload is sometimes evident inusers' inability to find important information whenever to need.

User-generated content allows collective understanding, which is a massive machine-human knowledge processing function capable of managing chaotic volumes of information. [11]

There are several approaches by which the data can be obtained for research anddevelopment. One of these approaches is from Open Data Portals. The open dataportals provide authentic data sets. The datasets can be downloaded from different portals in multiple formats including XML, CSV, JSON andmany others. [9]

Future Work:

The following section discusses the work that will be implemented with future releases of the Web Application.

1. First, users' awareness of identity management online may increase overtime. The "manipulation" of personal image online may need to be taken into considerations in future work.

2. Second, There are a different other issues hidden in the "long tail". Future work can be done to design more sophisticated algorithms in order to reveal the hidden information in the "long tail". Several of these issues may be of great interest to education researchers and practitioners.

3. Often times, manual analysis is time consuming not only because of the time spent on analyzing the actual data, but also the time spent on cleaning, organizing the data, and adapting the format to fit the algorithms. We plan to build a tool based on the workflow proposed here combining social media data and possibly student academic performance data.

4. Possible future work could analyze student's generated content other than texts (e.g. images and videos), on social media sites other than Twitter (e.g. Facebook, Tumbler, and YouTube). Future work can also extend to students in other majors and other institutions. [20]

Automated opinion mining and summarization systems are thus needed, as subjective biases and mental limitations can be overcome with an objective sentiment analysis system. The issue of detecting opinion spam or fake reviews. Finally, we also introduce the research topic of assessing the utility or quality of online reviews.

Conclusion

Sentimental or opinion analysis is a major research area which focus on the understanding and behavior of the society, it is also very important for business to understand the feedback and expectation from the users to improve the quality of product. In this paper explained the procedure to analyzing the opinion using different techniques, also gives insight on recent updates and innovation in sentimental analysis using recent ontology based machine learning approach explored better decision making.

References

1. Bing Liu, Lei Zhang "*A Survey of Opinion Mining and Sentiment Analysis*" ,*Mining Text Data*, DOI 10.1007/978-1-4614-3223-4_13, © Springer Science+Business Media, LLC 2012

2. Vijay Rana, Vijay RanaAnalysis of Web Mining Technology and Their Impact on Semantic Web International Conference on Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity (CIPECH14) 28 & 29 November 2014

3. Prof Shivani Desai, PriyankBhatt(11BIT044), VrajSolanki(11BIT059), Semantic and Sentiment Analysis, **IJCSC** Volume 5 • Number 1 March-Sep 2014 pp.57-60 ISSN-0973-7391

4. Tetsuya Nasukawa, Jeonghee Yi: Sentiment Analysis: Capturing Favorability Using Natural LanguageProcessing,(2003).

5. AsmitaDhokrat, Sunil Khillare, C. NamrataMahender, Review on Techniques and Tools used for Opinion Mining, International Journal of Computer Applications Technology and Research Volume 4– Issue 6, 419 - 424, 2015, ISSN:- 2319–8656

6. S. ChandraKala and C. Sindhu, October 2012 "Opinion Mining and Sentiment Classification: A Survey," ICTACT Journal On Soft Computing, Volume: 03, Issue: 01.

7. HaseenaRahmath, 2014 "Opinion Mining and Sentiment Analysis - Challenges and Applications," International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 5.

8. Liu, B., 2010 "Sentiment Analysis and Subjectivity," Appeared in Handbook of Natural Language Processing, Indurkhya, N. &Damerau, F.J. [Eds.].

9. Manu Bansal , "SENTIMENT ANALYSIS FROM SOCIAL MEDIA LIVE FEEDS USING UNSTRUCTURED DATA MINING", International Journal of Computing and Corporate Research, ISSN (Online) : 2249-054X Volume 5 Issue 5 September 2015

10. ZirunQi , Veda C. Storey , WaelJabr, "Sentiment Analysis Meets Semantic Analysis: Constructing Insight Knowledge Bases", https://www.researchgate.net/publication/282008175,

11. Conference Paper · December 2015

12. KawaljeetKaurKapoor, KuttimaniTamilmani, Nripendra P. Rana, PushpPatil, Yogesh K. Dwivedi, Sridhar Nerur, "Advances in Social Media Research: Past, Present and Future", Springer open access publication InfSyst Front (2018) 20:531–558

13. Y. Bengio, ``Learning deep architectures for AI,'' *Found. Trends Mach.*

14. *Learn.*, vol. 2, no. 1, pp. 1_127, 2009.

15. A. Pak and P. Paroubek, ``Twitter as a corpus for sentiment analysis and

16. opinion mining,'' in *Proc. LREC*, vol. 10. 2010, pp. 1320_1326.

17. M. Thelwall, K. Buckley, and G. Paltoglou, ``Sentiment strength detection

18. for the social Web,'' *J. Amer. Soc. Inf. Sci. Technol.*, vol. 63, no. 1,

19. pp. 163_173, 2012.

20. G. Paltoglou and M. Thelwall, ``Twitter, MySpace, Digg: Unsupervised

21. sentiment analysis in social media,'' *ACM Trans. Intell. Syst. Technol.*,

22. vol. 3, no. 4, pp. 1_19, 2012.

National Conference on Recent Advances in Physical and Mathematical Sciences (NCRAPMS-2020)
Organizer :- Shri Shivaji College Of Arts , Commerce & Science, Akola

18th Jan. 2020

23. ZHAO JIANQIANG , GUI XIAOLIN, AND ZHANG XUEJUN, "Deep Convolution Neural Networks for Twitter Sentiment Analysis", IEEE Access, VOLUME 6, 2018 .

24. ByeongkiJeong, Janghyeok Yoon, Jae-Min Lee, "Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis", International Journal of Information Management, 2017

25. RavinderAhuja, Aakarsha Chug, ShrutiKohli, Shaurya Gupta, PratyushAhuja, "The Impact of Features Extraction on the Sentiment Analysis", International Conference on Pervasive Computing Advances and Applications – PerCAA 2019, ScienceDirect, Procedia Computer Science 152 (2019) 341–348

26. K. N. Manasa and M. C. Padma, "A Study on Sentiment Analysis on Social Media Data", Springer, https://doi.org/10.1007/978-981-13-5802-9_58

27. K. C. Khatib, at alT. D. Kamble, B. R. Chendake, G. N. Sonavane , "Social Media Data Mining For Sentiment Analysis", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056, p-ISSN: 2395-0072 Volume: 03 Issue: 04 | Apr-2016, Page 373-376

28. Sheela Singh1, Priyanka Arya2, Alpna Patel3 and Arvind Kumar Tiwari, "Social Media Analysis through Big Data Analytics: A Survey", 2nd International Conference On Advanced Computing And Software Engineering (ICACSE-2019)

Aayushi International Interdisciplinary Research Journal (ISSN 2349-638x) (Special Issue No.66)
Impact Factor 6.293   Peer Reviewed Journal   www.aiirjournal.com   Mob. 8999250451

43