

Prologue to Sentiment Analysis and its Challenges

^[1]Pooja Sikchi, ^[2]Himangi Pande

Department of Information Technology, Maharashtra Institute of Technology, Pune.

Abstract – In a most recent couple of years, there has been a colossal consideration towards opinion investigation to remove the plenitude of information that exists on the web-based social networking locales, for example, Facebook, Twitter, and so on. This information is assessment information which can be certain or negative. Different kinds of assumptions are anticipated by various clients. This information comprehends the perspective of clients and furthermore catches the mentality of a client towards the item. We are primarily concentrating on the utilization of content digging for slants grouping. It means to catch or break down the mindset and assumptions of individuals in a specific era or season. This mindset can be glad, tragic, furious, befuddled, and so forth. This investigation would require a specific calculation or a strategy. For this reason, we can utilize existing techniques in view of Support Vector Machine and Naïve Bayes. Utilizing content mining we build a notion vocabulary, in view of interpositions, emojis or acronyms. These are extricated from statuses refreshes by various clients on Facebook. Content mining encourages you to remove subjective data. **Keywords:** We would like to encourage you to list your keywords here. They should be separated by commas.

Keywords: NLP, Sarcasm, Sentiment, Anaphora resolution, Named Entity Recognition

1. INTRODUCTION

Today online networking sites have caught every individual's consideration. Before communicating their feelings with family, companions or partners they refresh their state of mind via web-based networking media locales. This has advanced to a substantial scale. They compose surveys with respect to different items which causes different organizations to break down their client input.

In any decision-making process, what crowd thinks has dependably been a critical snippet of data for the greater part of us. Internet and web have made it feasible for a clique of individuals to cooperate which prompts simple finding and extraction of their sentiments. Correspondence over the web has created different open doors for information mining by means of Natural Language Processing. This aide in estimation order and investigating notions.

Sentiment Analysis is a procedure of distinguishing common dialect to track the mind-sets of open about a specific item or a subject. Sentiment Analysis and opinion mining is the field of concentrate that dissects individuals' conclusions, assumptions, assessments, dispositions, and feelings from composed dialect[12]. Fundamentally, it is the way of deciding the passionate tone behind a progression of words, used to pick up a comprehension of the states of mind, assessments, and feelings communicated inside an online say. Sentiment Analysis requires building a specific framework which performs NLP, text analysis and computational linguistics. This

framework would at first perform content mining i.e., gathering information from audits and posts of facebook, twitter, or other online networking locales. Sentiment analysis is also sometimes referred to as opinion mining. These feelings can be certain or negative. Sentiment Analysis has a capacity to distillate experiences from web-based social networking destinations. This procedure is utilized by numerous associations over the world. Think about an illustration. Obama organization utilized the procedure of feeling mining to check open's consideration towards the battle messages and declarations amid the presidential race in 2012. Notion examination investigations the movements in the web-based social networking destinations and connects it with the share trading system trade.

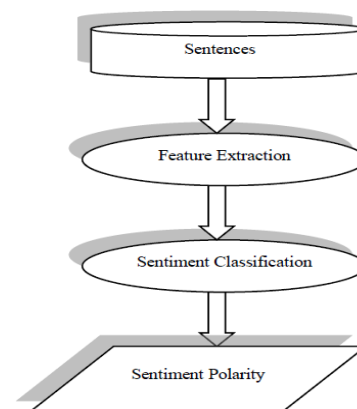


Fig. 1. Process of Sentiment Analysis

Sentiment Analysis is a precise procedure as appeared in figure 1[3]. It starts with analyzing the polarity of

sentences. polarity can be sure, negative or nonpartisan. Next stage is Feature Extraction. This step extracts the features from analyzed sentences. Kinds of highlights are Part of discourse, Presence of terms and their frequencies, Words, and expressions for supposition and refutation. After feature extraction sentiment classification is done. The techniques of Sentiment classification are classified into a lexicon-based approach, machine learning approach and hybrid approach [3]. Lexicon based approach is additionally separated into Dictionary based and corpus-based approach. Machine learning approach is additionally partitioned into supervised learning and unsupervised learning. Algorithms utilized for supervised learning are Naïve Bayes and Maximum Entropy.

Two basic methodologies to detect sentiments from text are Symbolic techniques and Machine Learning techniques[8]. Turney[9] used bag-of-words approach for sentiment analysis. In that approach, connections between the individual words are not considered and a report is spoken to as an insignificant accumulation of words. To decide the general conclusion, slants of each word is resolved and those qualities are joined with some accumulation capacities. He found the extremity of a survey in light of the normal semantic introduction of tuples removed from the survey where tuples are phrases having descriptive words or qualifiers. He found the semantic introduction of tuples utilizing the web crawler Altavista.

Kamps et al. [10] utilized the lexical database WordNet [11] to decide the emotional substance of a word along various measurements. They built up a separation metric on WordNet and decided the semantic introduction of descriptive words. WordNet database comprises of words associated by equivalent word relations.

Machine Learning procedures utilize a training set and a test set for grouping. Training set contains input feature vectors and their corresponding class labels. Utilizing this training set, a grouping model is produced which tries to arrange the input feature vectors into relating class marks. At that point a test set is utilized to approve the model by foreseeing the class names of unseen feature vectors[8].

2. TYPES OF SENTIMENT CLASSIFICATION

a) Document Level: Supervised machine learning approaches are utilized for foreseeing the general opinion of the archive. The survey of the whole document is taken and is prepared with the labeled samples. At last, the

document is marked as either as positive and negative as a whole. This whole procedure involves two stages: 1. Extracting the subjective highlights and converting them into feature vectors. 2. Based on feature vectors, subjectivity is classified [4].

b) Sentence Level: Sentence level records are simply short reports. Documents acquired from surveys are parsed into sentences. The sentences containing the opinion words are extracted and grouped into subjective and objective sentences. Subjective sentences hold opinions while objective sentences containing accurate data. A semantic orientation is done at sentence level by extricating opinion bearing terms, opinion holders and supposition item viewpoint relationship in each sentence. Pronoun determination and substance extraction are a portion of the strategies [1].

c) Word/Phrase level: Generally, adverbs and adjective sort of words have a semantic orientation. This orientation orders the word into positive, negative and neutral classes. The model of feature-based opinion mining and a synopsis is proposed which removes the opinion words from the surveys and arranges them appropriately [1].

3. CHALLENGES IN SENTIMENT ANALYSIS

Online networking destinations have opened numerous entryways for individuals to express their perspectives and conclusions. This data is identified with zones of training, business, tourism, wellbeing and it causes in expanding the span of web exponentially. This is where a large number of individuals speak with each other. The ubiquity of these social media sites has taken the world to an alternate stage. It has brought about a colossal accumulation of individuals suppositions and assessments. Every single individual has an alternate perspective with respect to any point/item. These sentiments are communicated in an unstructured organization. These perspectives can be a reality or a feeling. They are normally anticipated in characteristic dialect on the web. Extraction of helpful data from this convoluted information is a troublesome procedure. In spite of the fact that this aide in a basic decision making process, influencing a computer to act like a human brain and breaking down these conclusions is again a noteworthy test.

Human dialect is convoluted. Any maker can't think about perspectives of every shopper. Extracting this content is past human limit and time. Sentiment Analysis has been

inconspicuous because of intrinsic trouble in separating individual or singleton assessment from long entries. Documents change in their length and make characterization a more troublesome errand. Likewise deciding the subjectivity of a report whether the content in the record is a goal i.e., expressing a factor is subjective i.e., expressing an opinion. The complexity of language is hard to major and investigate. This requires the selection of an apparatus which will extract the assessment and investigate them to get a conclusion. Sentiment Analysis is an augmentation of information mining.

Another challenge is to educate a computer to act like a human brain and look at the incorrect spellings, complex sentences, syntactic subtleties, an expression which are anticipated online is a troublesome procedure. Likewise influencing a machine to figure out how a lexicon can influence a tone is much more troublesome. Think about a sentence, "My cell phone has been lost. Incredible!" Humans can easily distinguish that this sentence is wry and communicates antagonism. Be that as it may, what PC does is, picks the word incredible and order it as positive. Sarcasm is wrongly deciphered as to a great degree positive sentiment. Along these lines, sarcasm is another challenge in sentiment analysis[6]. The present strategy for opinion investigation has a tendency to distinguish the perspective of clients yet bombs because of multifaceted nature in passed on sentences. Understanding the semantics of a sentence is another test for the Semantic web community. Patching up the unstructured content into a shape which can be prepared by a machine is a troublesome undertaking.

Named entity recognition is another challenge faced by sentiment analysis. It is an errand in information mining to recognize and extricate the named substances. Sentiment analysis requires gathering sentiments about entities that matter and empowers utilization in an easy to use way. An NER(Named Entity Recognition) names arrangements of words in which there are names of entities, for example, individuals, associations, areas and so forth approaches prevail with regards to recognizing the extremity of the archive, articulations, words which can influence a sentence. Named entity recognition is what is the person actually talking about? For example 300 Spartans – a group of Greeks/ a group of movies? These methodologies are reliant upon the original source text and fail to get it. An entity is only a name related gestures and perspectives. Utilizing an element we can figure out what a man needs to tell about a specific point. An element is characterized as a subject or a definition a

thing or an item. We can discuss a man, we can discuss an item, we can discuss a theme, and so forth. For depicting these things we utilize certain sort of words. These words may be certain or negative. The stature of positivity and negativity must be controlled by considering every one of the general population sees. Individuals may be in thousands, in lakhs. Gathering reviews about these individuals is a simple undertaking because of web-based social networking locales accessible. In any case, disentangling them is a troublesome errand. This is because of the sarcasm feature which is utilized and it really makes the reader confounded. So you can think how a machine will comprehend this component. Influencing a machine to go about as a human brain isn't a simple subject. A machine needs to think in every one of the bearings whether the individual is certain or negative. Be that as it may, to accomplish this we need to utilize certain tools which are available for sentiment analysis. These devices ought to have the capacity to accumulate the watchwords, compose them and afterward discover a conclusion. So it isn't a simple errand.

Anaphora resolution is another challenge. It is the issue of settling a thing or a pronoun. For instance: " We watched the movie and went to dinner, it was dreadful" This sentence is discussing what. What it really alludes to? The sentiments of a man enlighten u everything regarding that individual.

Parsing is another challenge. It implies what is the subject and object of the sentence? Which one is the verb and which one is the adjective is hard to decide?

Recognizing counterfeit surveys and spam information is another strenuous task[6]. The web and online networking contain both authentic and spam information. For a powerful feeling examination, it is important to dispose of the spam substance before preparing. This should be possible by distinguishing copies, by recognizing exceptions and by thinking about the notoriety of reviewer[4].

Grammatically incorrect words are the another challenge. There are numerous methodologies that analyze sentiments yet barely any work achieved on linguistic mistakes. The results of slant examination can be enhanced if these sorts of mistakes can be mapped to remedy words[13].

Credibility/Beaviour/Homophily is another challenge. Behaviours in web-based social networking are just seen by the traces they leave in online networking. We once in a while watch the driving variables that reason these behaviors; nor would we be able to talk with people as to behaviors. Regardless of whether a conduct is investigated on online networking and related examples

are gathered, it's hard to check the legitimacy of these behavioral examples. Assessment turns out to be more challenging for businesses in which imperative choices are to be mentioned in light of objective facts of individual behavior[13].

Handling Noise and Dynamism is another problem faced in sentiment analysis. Online networking information is gigantic, uproarious, unstructured, and dynamic in nature, and in this way novel difficulties emerge, presents agent examine issues of mining social media. Recognizing and expulsion of noisy data is a challenging task.

4. CONCLUSION

Sentiment Analysis helps in mining a lot of information from online networking and different sources. This paper outlines different difficulties of sentiment analysis like named entity recognition, anaphora resolution, sarcasm, fake reviews, spam data, and so on. It likewise talks about different levels of feeling examination like document level, sentence level, word/phrase level.

REFERENCES

1. P. Balaji, O. Nagaraju, and D. Haritha, "Levels of sentiment analysis and its challenges: A literature review," 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, 2017, pp. 436-439.
2. Doaa Mohey El-Din Mohammed Hussein, "A survey on sentiment analysis challenges," 2016
3. Chandni, Nav Chandra, Srishti Gupta, Renuka Pahade, "Sentiment Analysis and its challenges," 2015 International Journal of Engineering Research and Technology (IJERT).
4. Haseena Rahmath P, "Opinion Mining and Sentiment Analysis- Challenges and Applications," 2014 International Journal of Application or Innovation in Engineering & Management (IJAIEM).
5. A. M. Abirami and V. Gayathri, "A survey on sentiment analysis methods and approach," 2016 Eighth International Conference on Advanced Computing (ICoAC), Chennai, 2017, pp. 72-76. doi: 10.1109/ICoAC.2017.7951748
6. Mr. Saifee Vohra, Prof. Jay Teraiya, "Applications and Challenges for Sentiment Analysis: A survey," 2013 International Journal of Engineering Research and Technology (IJERT).
7. Z. Nanli, Z. Ping, L. Weiguo, and C. Meng, "Sentiment analysis: A literature review," 2012 International Symposium on Management of Technology (ISMOT), Hangzhou, 2012, pp.572-576.
8. M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, 2013, pp. 1-5.
9. P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417-424, Association for Computational Linguistics, 2002.
10. J. Kamps, M. Marx, R. J. Mokken, and M. De Rijke, "Using wordnet to measure semantic orientations of adjectives," 2004.
11. C. Fellbaum, "Wordnet: An electronic lexical database (language, speech, and communication)," 1998.
12. Bing Liu, "Sentiment Analysis and Opinion Mining," in Sentiment Analysis and Opinion Mining , 1, Morgan & Claypool, 2012, pp.167-
13. H. P. Patil and M. Atique, "Sentiment Analysis for Social Media: A Survey," 2015 2nd International Conference on Information Science and Security (ICISS), Seoul, 2015, pp. 1-4.