

A Collaborative Social Networking Website Inculcating Sentiment Analysis

Nidhi Bipin, Pranika Dharmawat, Vidhi Ambwani, Shivam Sehgal

SVKM's NMIMS Mukesh Patel School of Technology Management and Engineering, Mumbai,
Maharashtra, India

ABSTRACT

In the recent years, social media has become ubiquitous and important for social networking and content sharing. In this paper we have made an attempt to show how Social media no longer has to be an obstacle to studying; it can help students create and manage a study community, make the best use of study time, and find new resources to help them learn and retain knowledge. As social media data has grown more abundant, data can be captured that may potentially represent behavioural patterns in society. In turn, this unstructured social media data can be parsed and integrated as a key factor for predictive intelligence. In this paper along with the framework of the social networking site, we have used certain mechanisms to undergo predictive texting and sentimental analysis using Natural Language Processing.

Keywords : Social Networking, Sentiment Analysis, Patterns, Natural Language Processing, SNS, NMBYTES

I. INTRODUCTION

The utilization of Social Networking Sites has been widespread to the point that they have grabbed the eye of scholastics around the world. SNS are presently being explored by various sociology analysts.. Further, the use of Social Networking Sites (SNS) among the general population of India is obviously expanding, especially among the Indian undergrads. In this manner we built up an informal communication site called NMBYTES so unique offices could be done effectively by means of the systems administration site, in the meantime we have attempted to gather the information of the client and endeavored to make an example and in this manner execute wistful examination.[2] Web applications and thusly, an exceptional development in client created information is balanced for sentiment mining. Information, for example, web-postings, pictures, recordings, and so forth., all express feelings on different subjects and occasions, offer huge chances

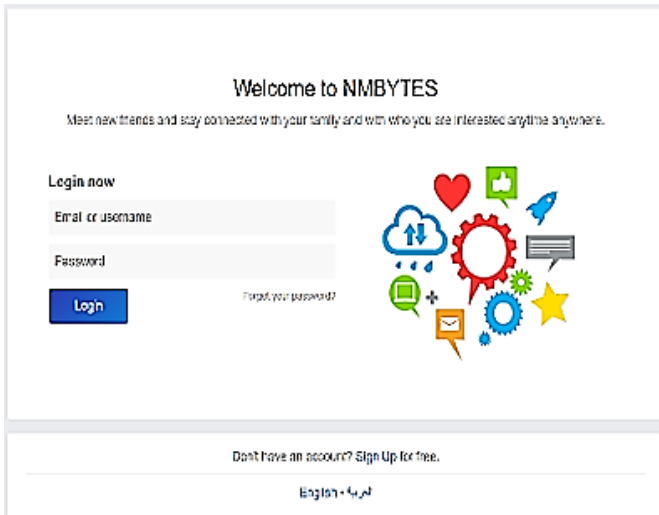
to contemplate and examine human conclusions and opinion. Specifically, we investigate uses of opinion examination and exhibit how conclusion mining in online life can be misused to decide the diverse surveys of the understudies. This paper shows the underlying consequences of an exploration venture whose principle objective is to make a model of information portrayal with regards to interpersonal organizations on the Web. In this undertaking we built up a model programming for assessment investigation of the understudies of our school through the distinctive assignments like posting diverse posts, surveys and numerous different errands. This paper introduces the underlying aftereffects of an exploration venture whose principle objective is to make a model of information portrayal with regards to interpersonal organizations on the Web. In this task we built up a model programming for conclusion examination of our site which is NMBYTES.[1] This is accomplished by the utilization of morphologic examination, and language highlights recognition

helped by ontologies. Explicit targets incorporate the structure of philosophies for sentiment mining, arrangement and grouping, formation of a lingual authority. Using this hybrid unsupervised approach, by combining language processing, lexicon techniques and ontology techniques for the sentiment data structure, we expect to generate classifiers for sentiments and opinions and business intelligence insights that improve the results obtained so far in sentiment analysis and opinion mining without relying on supervised algorithms such as machine learning approaches that requires a costly training phase that may be impeditive for groups with limited resources.[3]

II. SYSTEM ARCHITECTURE

Social Networking Sites can be comprehensively characterized as web based social spaces intended to encourage correspondence, cooperation, and substance sharing crosswise over systems of contacts. When all is said in done, understudies are progressively keen on utilizing cutting edge innovation in each field contrast with some other age gathering. It is likewise valid on account of mode of correspondence. In the class of clients of SNS the dominant part are youth. In India too this reality is recognized by the different investigations and henceforth today the utilization of SNS has turned out to be substantial piece of the understudy's lives in India. The term Social Networking Sites has been characterized by various creators in a few diverse ways. This leaves the pursuer with a sentiment of being 'uncertain' of its genuine importance. NMBYTES is a stage which may contain suppositions, considerations, realities, references to pictures and other media and, as of late, stream video recorded live and put online by clients. So it is something other than a SNCs in which a client shows and builds their social connections, it is a genuine correspondence direct in which a client can pick its

points and its hub of reference as indicated by his interests and culture.[6] The proposed procedure and our subsequent model gathers, structures and investigates Web data by utilizing a blend of content preparing innovations with a few other etymological strategies, for example, morphological, syntactic and semantic examination guided by the objective space terms provided by our ontologies and the rundown of terms that brings out supposition (with a given extremity and quality esteem)[7] Each article may have at least one terms that can be utilized to distinguish references to that specific item on the sentences extricated. Every notion recognized in the process portrayed is then put away in the philosophy tree structure, as a connection of at least one items in the metaphysics tree to a slant esteem, that can be either impartial, positive or negative. Utilizing this cross breed unsupervised methodology, by joining language preparing, dictionary strategies and philosophy procedures for the notion information structure, we hope to create classifiers for conclusions and assessments and business knowledge experiences that improve the outcomes got so far in feeling investigation. Additionally, the assumption examination strategy utilized needs to consider the prerequisites of the final application in which it will be utilized. There is an imperative contrast between sending a framework working for dialects, for example, English, for which various etymological assets and investigation instruments exist and a framework sent for dialects with couple of such apparatuses or one that is gone for preparing information from an expansive arrangement of dialects [1].



DATA SELECTION

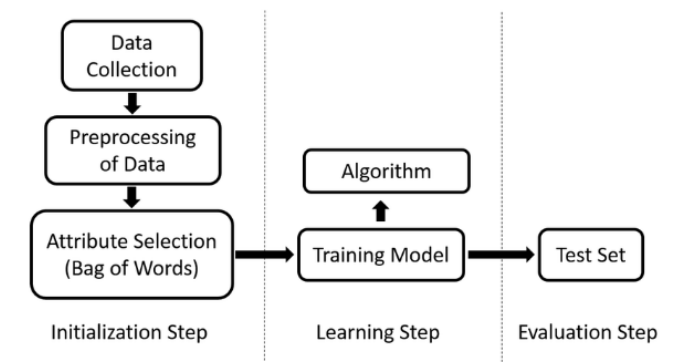
As a Social Media Networking platform NMBYTES is organized as a coordinated chart, in which every client can pursue various different clients (followers), and can be also trailed by different clients (supporters). Therefore, the "follow relationship is asymmetrical , it doesn't require obligatory affirmation, and it is basically used to get every open message distributed by any followee client. Therefore, in our examination we gathered three sorts of information the clients' profiles, posted messages; "pursue" connections among client[4]

SENTIMENT ANALYSIS

Sentiment analysis (a.k.a sentiment classification, opinion mining, subjectivity analysis, polarity classification, affect analysis, etc.) is the multidisciplinary field of study that deals with analyzing people's sentiments, attitudes, emotions and opinions about different entities such as products, services, individuals, companies, organizations, events and topics and includes multiple fields such as natural language processing(NLP), computational linguistics, information retrieval, machine learning and artificial intelligence. It is a set of computational and NLP based techniques which could be leveraged in order to extract subjective information in a given text unlike factual information, opinions and sentiments are subjective. Research on sentiment

analysis has been investigated from different perspectives describes as follows:

- ✓ Document level: The aim here is to determine the overall sentiment of an entire document. For example given a product review, the task is to determine whether it expresses positive or negative opinions about the product. This level looks at the document as a single entity, thus it is not extensible to multiple documents
- ✓ Sentence level: This level of analysis is very close to subjectivity classification and the task at this level is limited to the sentences and their expressed opinions. Specifically, this level determines whether each sentence expresses a positive, negative or neutral opinion.
- ✓ Entity and aspect level: Instead of solely analyzing language constructs (e.g. documents, paragraphs, sentences), this level (a.k.a feature level) provides finer-grained analysis for each aspect(or feature) i.e., it directly looks at the opinions for different aspects itself. The aspect-level is more challenging than both document and sentence levels and consists of several sub-problems[4].



The different bag of words are the review taken into consideration put up by the students.[6]

The pre-processing stage contains the following steps:

- ✓ Repeated punctuation sign normalization In the first step of the pre-processing, we detect repetitions of punctuation signs (“.”, “!” and “?”). Multiple consecutive punctuation signs are

replaced with the labels “multistop”, for the fullstops, “multiexclamation” in the case of exclamation sign and “multiquestion” for the question mark and spaces before and after.

- ✓ Emoticon replacement In the second step of the pre-processing, we employ the annotated list of emoticons from SentiStrength4 and match the content of the reviews against this list. The emoticons found are replaced with their polarity (“positive” or “negative”) and the “neutral” ones are deleted.
- ✓ Lower casing and tokenization. Subsequently, the reviews are lower cased and split into tokens, based on spaces and punctuation signs.
- ✓ Slang replacement The next step involves the normalization of the language employed. In order to be able to include the semantics of the expressions frequently used in Social Media, we employed the list of slang from a specialized site
- ✓ Word normalization At this stage, the tokens are compared to entries in Thesaurus. If no match is found, repeated letters are sequentially reduced to two or one until a match is found in the dictionary (e.g. “perrrrrrrrrrrrrrrrrfeect” becomes “perrfeect”, “perfeect”, “perrfect” and subsequently “perfect”). The words used in this form are marked as “stressed”. [1]
- ✓ Affect word matching Further on, the tokens in the tweet are matched against different sentiment which were previously split into four different categories (“positive”, “high positive”, “negative” and “high negative”). Matched words are replaced with their sentiment label - i.e. “positive”, “negative”, “hpositive” and “hnegative”. A version of the data without these replacements is also maintained, for comparison purposes.
- ✓ Modifier word matching Similar to the previous step, we employ a list of expressions that negate, intensify or diminish the intensity of the sentiment expressed to detect such words in the reviews. If such a word is matched, it is replaced with “negator”, “intensifier” or “diminisher”,

respectively. As in the case of affective words, a version of the data without these replacements is also maintained, for comparison purposes. [3]

III. LANGUAGE BASED PROCESSING METHODS

Alongside sentiment analysis from two unique viewpoints, in particular, vocabulary based, and linguistic analysis. The clearest yet critical markers of assessments are estimation or conclusion words, for example, great, stunning, poor, terrible just as certain expressions and maxims which are utilized to express positive or negative suppositions which has been utilized in our task A notion dictionary (a.k.a feeling vocabulary) is the rundown of such words and states and is important however not sufficient for assumption investigation. Notwithstanding misusing dictionaries, semantic based methodologies we have additionally utilized the linguistic structure of the content for assumption classification. There are two sorts of vocabulary age strategies, in particular, word reference based and corpus based methodologies The first classification begins with a little arrangement of feeling words and grows the vocabulary through bootstrapping a specific word reference while the second class creates the assessment dictionary through learning the dataset. [2]

IV. CONCLUSION

In this project, we introduced a technique to characterize the opinion in surveys, by considering their quirks and adjusting the highlights utilized to their structure and substance. Specifically, we utilized a pre-processing stage to standardize the language and sum up the vocabulary utilized to express slant. This respected spelling, slang, accentuation, and so forth., and the utilization of estimation word references and modifier records to sum up the examples of opinion articulation extricated from the preparation information. We have demonstrated that

the utilization of such summed up highlights significantly improves the consequences of the assumption classification, when contrasted with the best performing methodologies that don't utilize influence word references. Utilizing such a methodology abstains from overfitting the information and, as we have appeared, at tantamount exhibitions on various datasets. In future work, we intend to assess the utilization of higher-request n-grams (3-grams) and skip grams to remove increasingly complex examples of assumption articulations and have the capacity to distinguish all the more definitely the extent of the refutation.

Cite this article as :

Nidhi Bipin, Pranika Dharmawat, Vidhi Ambwani , Shivam Sehgal, "A Collaborative Social Networking Website Inculcating Sentiment Analysis", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 2, pp. 853-857, March-April 2019. Journal URL : <http://ijsrcseit.com/CSEIT1952247>

V. REFERENCES

- [1]. Social Network and Sentiment Analysis on Twitter: Towards a Combined Approach Paolo Fornacciari, Monica Mordonini, Michele Tomaiuolo
- [2]. Sentiment Analysis in Social Media Text Alexandra Balahur European Commission Joint Research Centre Vie E. Fermi 2749 21027 Ispra (VA), Italy
- [3]. Sentiment Analysis on Social Media International Conference on Advances in Social Networks Analysis and Mining
- [4]. Sentiment Analysis of Social Networking Sites (SNS) Data using Machine Learning Approach for the Measurement of Depression
- [5]. International Conference on ICT Convergence, At Jeju Island, Korea
- [6]. Social Media Data Mining For Sentiment Analysis K. C. Khatib¹ T. D. Kamble², B. R. Chendake³ G. N. Sonavane⁴
- [7]. Social Media Sentiment Analysis using Machine Learning Classifiers Bharat Naiknaware¹, Bindesh Kushwaha, Seema Kawathekar
- [8]. Sentiment Analysis for Social Media: A Survey Harshali P. Patil ; Mohammad Atique