# **Enhanced Sentiment Analysis by using Social Media**

Shubhangi D. C. PhD Department of Studies in Computer Science and Engineering, Visvesvaraya Technological University Center for PG Studies, Kalaburagi, Karnataka, India

# ABSTRACT

In the current days development of web technology and its growth, there is a huge amount of data present in the web for internet users and a lot of data is generated as well. Internet has become a platform for online learning, exchanging ideas and sharing opinions like Twitter, rapidly gaining popularity as they allow people to share and express their views about the trending topics. In this paper we introduced a new approach to adapt the topic model derived from news to tweets. We proposed the hashtagger+ how quickly can we suggest the hashtag mainly focuses on sentiment analysis of twitter data which is helpful to analyze the information in the tweets where opinions are highly, heterogeneous and are either positive, negative, or neutral in some cases. By using various machine learning algorithms, like Naive Bayes classifier, cold start search algorithm, and hashtagger+ recommendation model.

## **Keywords**

Twitter, machine learning, naive bayes, coldstart search, hash tag.

# 1. INTRODUCTION

Nowadays twitter is one of the most popular social networking and well known platform to express their emotions, opinions and share views about their daily lives. Social media is generating a large volume of sentimental data in the form of tweets. These tweets cover a larger number of various topics, including comments on recent or ongoing events and trending topics. Hashtag (i.e, keyword prefixed with # symbol) has demonstrated its effectiveness in bringing organization to the sparse information in Twitter. Hashtags associated with tweets enhance information diffusion and tweet search as well as facilitate social conversation. The effectiveness of hashtags in tweets. However, is limited by the freedom of users in deciding (i) whether or not to explain tweets with hashtags, and (ii) which hashtags to use (e.g. #modi, #ipl, or #bangalore). A story can have a variety of hashtags and it has a different stages of the story. Suppose, for example,In the Taj hotel attack(a group of terrorists attacks that took place in Mumbai 2008), and a twitter has taken a huge role. The two hashtags were used during this event: #mumbai, #taj hotel attack. For each hashtags dominates the discussion at different time points. #mumbai.the location of the event, is popular at the beginning of the event. Finally #taj hotel attack refers to the series of bombing attacks lasting four days across in Mumbai. This attack which drew widespread global condemnation. This example shows the critical capture the news and the hashtags, as the progress influences the tweeter consideration. The amount of content generated by users is too vast for a normal user to analyze. Likewise in our approach via tweeter text contents we need to capture the sentimental score by using on NLP tool. These contents are mainly opinions, sentiments, appraisals, attitudes, and emotions, which form the core of Sentiment Analysis.

Priyadarshini M. Department of Studies in Computer Science and Engineering, Visvesvaraya Technological University Center for PG Studies, Kalaburagi Karnataka, India

However, Twitter provides hashtags to help label postings and classify them. Hashtags can thus be used to identify particular news stories or trending topics.

# 2. RELATED WORK

Zongyang Ma, Yuan & Gao[1] proposed two PLSA-style point models to display the inactive association connecting tweet substance, client conspiracy, moment in time, and hashtag at subject point. They demonstrated that Hashtag Pivoted Model (HPM) beats Content Pivoted Model (CPM) as far as perplexity measure in the course of broad investigation. Hoang-Vu,et al.[2] With a specific end goal to utilize Twitter as a sensor for news, another way to deal with relate tweets to news stories and themes. Their approach utilizes the material in news to determine a theme demonstrates, and to manage the distinctions in vocabularies. Qi Zhang, et al [3] they suggested the concern of hashtag proposal for micro blogs and presented a novel approach that fuses the various type of hashtags have idiosyncratic propagation into the contemporary interpretation display intended for hashtag suggestion obligation. Yuyun, et al [4] presented a novel deliberation based CNNs designed for playing out the hashtag proposal task. To keep away from hand-create highlights and taking different favorable circumstances of CNNs they presented the proposed system. Bichen, et al [5] A novel consideration based CNNs for playing out the hashtag proposal errand. they proposed a novel consideration based CNN engineering, which comprises of a nearby consideration channel and worldwide channel. Kazi Saidul, et al [6] have presented a study of the cutting edge in programmed key expression extraction. While unsupervised methodologies have begin to match their administered partners in execution. Godin, et al [7] proposed another technique to deal with suggesting hashtags for tweets. The proposed technique accommodates simple ordering and pursuit of tweets. Zhuoye Ding et al [8] introduced a relevant explanation show, that joins the upsides of both topic sculpt and interpretation display. The proposed model use topic particular utterance activated to connect the terminology hole. Then again, it can find the points of tweets by a theme display intended for microblogs. Zhuove, et al [9] introduced a peak meticulous interpretation display, which joins the benefits of both subject model and interpretation demonstrate. Qi Zhang, et al [10] a new technique which joins fleeting plus individual elements into current elucidation demonstrate meant for hashtag suggestion assignment.

# 3. METHODOLOGY

In the existing work they recommended the hashtag deal with the difficulty from either class/topic or learning to rank point of view. The wide adoption of hashtags in Twitter has attracted significant explore consideration. The hashtag recommendation to news it has not effective, efficiency and coverage is not in convenient way. If we look for particular topic to recommend hashtag it will takes the time instead of fraction of sec will takes 5 more minutes to recommend the hashtag.

In this paper we proposed the hashtagger+ to enhances the effective scope of how to rank the system for tweeter hashtags to news articles. By the sentiment analysis we prove that the textual information retrieval techniques mainly focus on processing, searching or analyzing the accurate data present. Sentiment analysis can be defined as a process that automates removal of attitudes, opinions, views and emotions from text, tweets and database sources through Natural Language

Processing (NLP). Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral" from the extraction of tweeter data.

In this section we discuss the proposed system architecture model. We improve our model to address the real time recommendation, a tweeter contains a lot of opinions about the data which are expressed in different ways by the different users. The tweeter dataset used in this work has labeled into positive, negative and neutral thus the sentiment analysis of the data becomes easy to observe the various extracting tweets. In the tweet processing method has a distinctive properties we extract the aspects from the processed data set. Later this aspects are used to compute the positive, negative and natural emotions which are useful for determining the opinion of the individual tweets. Generate the dataset is an important one for solving classification problems. Machine learning techniques require the key features of text for processing. The naive bayes classifier which classify the tweets in the right category or position.

#### Naive Bayes Algorithm:

It is a simple technique for constructing probabilistic classifier and it is examining the set of texts that has been categorized the list of texts to classified to their right position or class. Let t be the tweet and D\* be a class that is assigned to t, where

$$\begin{split} D^* &= arg \; mac_d \; \; P_{NB}(d|t) \\ P_{NB}(d|t) \; &= \; (P(D)) \sum p(f|d)^{ni(d)} \end{split}$$

P(t)

f is a feature, ni(d) and is present in d which represents a tweet. Here, m denotes no. of features. Parameters P(d) and P(f|d) are computed through maximum likelihood information d.To train and classify using Naive Bayes Machine Learning technique.





## A. Cold Start Search Algorithm

This method searches the database for the current tweets from past 12h from the recent point and uses them for the feature extraction tweets. After feature extraction we apply the L2R model this approach will helps in recommending hashtags for a new article. where Sz is the set of articles in the database within the time window  $P_t$ 

#### Algorithm1. ColdStart1: Searching Recent Tweets

**Input:** Given article z, extracted from Query  $Q_a$ , from the database of past article s.

Output: Relevant hash tags Uz-search, Relevant tweets Pz-search

#### Method:

 $P_{a-search=}$ Ø

t=60\*12//the past 12h time slot for the tweeter window

 $P_t$  = Sub set tweets(S, t) //recover all the tweets within time window t.

for tweet  $\in P_t$  do

for phrase  $\in Q_a$  do

if phrase  $\in$  tweet then

 $P_{a-search} = P_{a-search} \cup \{tweet\}$ 

Break

Return Pa-search, Ø

Natural-Language Processing(NLP) is a zone of software engineering and computerized reasoning worried about the connections amongst PCs and human (common) dialects, specifically how to program PCs to productively process a lot of characteristic vernacular information. NLP is described as a difficult issue in software engineering. Human dialect is infrequently exact, or clearly talked. NLP calculations are normally in light of machine learning calculations. Rather than hand-coding substantial arrangements of tenets, NLP can depend on machine figuring out how to rank the more information are examined, the more precise the model will be.

Summarize pieces of content utilizing Summarizer to extricate the most essential and focal thoughts while disregarding unimportant data. Identify the sort of element extricated, for example, it being a man, place, or association utilizing Named Entity Recognition.

Use Sentiment Analysis to recognize the assumption of a string of content, from exceptionally negative to unbiased to extremely positive.

### 4. RESULT AND DISCUSSION

The results show that the proposed method is significantly better than the existing methods. Analysis was done on this labeled datasets using various feature extraction technique. We collected dataset containing positive, negative or neutral tweets. Those dataset were trained data and was classified using Naive Bayes Classifier. To determine sentiment of tweets we collected data using tweeter API. Those data were stored in the dataset of a tweeter. To check the polarity of set tweet we train the classifier with the help of trained data. Later we will apply the hash tag recommendation model can be predicted by taking into account considerations such as positive, negative or neutral. Then our complete proposed performance will be represents them in a graph.



Figure 1:Home page

Logging in, (or logging on or signing in or signing on), is the process by which an individual gains access to a computer system by identifying and authenticating themselves.



Figure 2:Connection status page

If any User want to get the tweets from twitter, they must create an application on Twitter API. After that it will generate the Secret Key and Token. Through API User will get Secret Key and Token to Collect Tweets from Twitter.



Figure 3: Collect tweets

Based on a Query, the tweets from the User Timeline are filtered. As there are a large number of tweets available, the topic based tweets have to be filtered by using a query.



**Figure 4:Display Tweets** 

Based on NLP tool, sentiment scores are calculated for each and every tweets have a sentiment score. So, we have to calculate the sentiments score based on the NLP tools. NLP will calculate the score based on positive negative and neutral tweets.



Figure 5:Sentiment graph page

The use of a recommender system is to analyze the negative node. After that, the recommender system will use the analyzed information to condition of the negative node.

## 5. CONCLUSION

In this paper we proposed a new approach hashtagger+ to recommend the hashtag for the tweeter data. The proposed method provides the highly effective scope how to rank the system. Our frame work obtains high scope for the trending topics in tweeter and also focus on the machine learning method in order to improve the accuracy of sentiment classification. In future work can be extended by enhancing various techniques and applying on other social media.

#### 6. **REFERENCES**

- [1] Z. Ma, A. Sun, Q. Yuan, and G. Cong, "Tagging your tweets: A probabilistic modeling of hash tag annotation in twitter," in Proc. 23<sup>rd</sup>ACMInt. Conf. Conf. Inform. Knowledge. Manage, 2014, pp. 999–1008.
- [2] T.-A. Hoang-Vu, A. Bessa, L. Barbosa, and J. Freire, "Bridging vocabularies to link tweets and news," in Proc. Int. Workshop Web Databases, 2014.

- [3] Y. Gong, Q. Zhang, and X. Huang, "Hash tag recommendation using dirichlet process mixture models incorporating types of hashtags," in Proc. Empirical Methods Natural Language Process., 2015, pp. 401–410.
- [4] Y. Gong and Q. Zhang, "Hashtag recommendation using attention-based convolutional neural network," in Proc. 25th Int. JointConf. Artif. Intell., 2016, pp. 2782–2788.
- [5] B. Shi, G. Ifrim, and N. Hurley, "Learning-to-rank for real-time high-precision hash tag recommendation for streaming news," in Proc. 25th Int. Conf. World Wide Web, 2016, pp. 1191–1202.
- [6] K. S. Hasan and V. Ng, "Automatic key phrase extraction: A survey of the state of the art," in Proc. 52nd Annu. Meeting Assoc. Compute. Linguistics, 2014, pp. 1262– 1273.
- [7] F. Godin, V. Slavkovikj, W. De Neve, B. Schrauwen, and R. Van deWalle, "Using topic models for twitter hashtag recommendation, "in Proc. 22nd Int. Conf. World Wide Web, 2013, pp. 593–596.
- [8] Z. Ding, X. Qiu, Q. Zhang, and X. Huang, "Learning topical translation model for microblog hash tag suggestion," in Proc. 23rd Int. Joint Conf. Artif. Intell., 2013, pp. 2078–2084.
- [9] Q. Zhang, Y. Gong, X. Sun, and X. Huang, "Time-aware personalized hash tag recommendation on social media," in Proc. 25th Int. Conf. Compute. Linguistics, 2014, pp. 203–212.
- [10] T.-Y. Liu, "Learning to rank for information retrieval," Foundations Trends Inform. Retrieval, vol. 3, pp. 225– 331, 2009.
- [11] S. Sedhai and A. Sun, "Hashtag recommendation for hyperlinkedtweets," in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2014, pp. 831–834.
- [12] F. Xiao, T. Noro, and T. Tokuda, "News-topic oriented hash tag recommendation in twitter based on characteristic co-occurrence word detection," in 12th Int. Conf. Web Eng., 2012.
- [13] O. Phelan, K. McCarthy, and B. Smyth, "Using twitter to recommend real-time topical news," in Proc. 3rd ACM Conf. Recommender Syst., 2009, pp. 385–388.