

User Profiling Based on Sentiment Analysis of Twitter Data

Dr. G Ramesh¹ R. Abirami² S. Padmavathy³ R. Abirami⁴

¹Professor

^{1,2,3,4}Department of Information Technology

^{1,2,3,4}K.L.N College of Engineering, Madurai, India

Abstract— Nowadays candidates are typically victimising the net searches, social networking websites, associate degreed social media to seek out their appropriate jobs where choosing an apt candidate with same set of skills is turning into a difficult task for the recruiters. We measure the practicability of exploitation of twitter knowledge to boost the effectiveness of a recruitment system, particularly for resume classification by using social network analysis methodology. So, the concept of SOCIAL MEDIA ANALYSIS (TWITTER DATA) is recommended to extract the personal data of a candidate to examine the differences depending on the intended use of twitter and their degrees of friendship between twitter and real world friends. Uses text classification to predict personality based on tweets tweeted by users. By analyzing user's interactions, probability of polarity (tweets, reactions or emotions) is determined.

Keywords: Sentiment Analysis, Naive Bayes, Polarity Classification

I. INTRODUCTION

Nowadays people search job opportunity mainly online using LinkedIn, Firstnaukri.com, Guru, Career Builder, Class door, Cool-Works, Monster etc. The key problem is that most of the online job recruiters only provide information about the recruiting process. Candidates have to process among all the data's to select their suitable. The whole procedure is tedious and insufficient. This project develops the online job search portal for candidates to find the jobs by uploading their information in the form of a CV. To make works of human resource(HR) and hiring process easier, we analysis the social network data(Twitter data).

Twitter is a social networking service attracts attention of researches in the past years. It provides many useful, both implicit and explicit social data. An example of the explicit data is user biography consisting of age, gender, and so on. An example of the implicit data is user behaviours.

In Twitter there are two main social actions that user performs: post and tweets. It also allows additional actions like check- in and tag friends. Post and tweets are non-structures data which requires much effort to pre-process. On the other hand, the promising action that relate to user preference identification is 'like' pages and 'share' posts by pages.

Main contribution of this work is to present technique that generates classification of CV based on user performances on social network and also their behaviours. In this work, we using the popular classification algorithms: Naive baiyes(NB) for evaluate performances.

II. RELATED WORKS

There are several works that exploit features in data of social networking services to predict user's personality and/or

interests. Bachrach et al. find association between personality and Twitter usage patterns.

Related work is on analysing and predicting psychodemographic traits based on user interest in twitter, where logistic regression is used for building prediction model[8]. The psycho-demographic is obtained by requesting mechanical truck workers to examine given Twitter account and give opinion on demographic features, e.g., gender, age, religion, income and so on. In comparison, our work is similar to Bachrach et al.on Twitter data and ground truth acquisition. However, our work is more focuses on reconstructing user profiles based on 'tweets' and 'share' behaviours on Twitter public pages, where ground truth is given directly by Twitter volunteered users that give their consent for research purpose.

Recent works on user profiling has been explored. Lu et al.represents user interest profile as concepts from Wikipedia, which is a large and inter-linked online knowledge base, and construct user modelling in Twitter based on their previous tweets and the relevance between a tweet and user interest[9]. Other works construct user profiles based on external source data. The work on use profiling proposed by Abel et al.constructs topic-based and user-based profiling from RDF graph of news articles and microblog posts in Twitter, where news articles are gathered from related websites, for instance CNN.com[10]. And by this way, we guarantee an association between jobs even when they do not share any user interactions. For users who do not have any activities, we model their preferences by classifying their resume content into pre-defined fine-grained job categories, and then recommend them popular jobs under these categories.

A. Job Recommendation

Extracting and analysing the information from the resume is the main objective of the job recruitment industry. These works involve text mining, pre-processing, finding term frequencies and sentiment analysis for matching jobs and candidate profiles. The existing system in this domain focus on candidate selection by human resources which is overcome in this paper by attracting job seekers through job recommendations.

Polarity simply means emotions expressed in a sentence. Emotions are closely related to sentiments. The strength of a sentiment or opinion is typically linked to the intensity of certain emotions, e.g., joy and anger. Polarity in sentiment analysis refers to identifying sentiment orientation (positive, neutral, and negative) in written or spoken language.

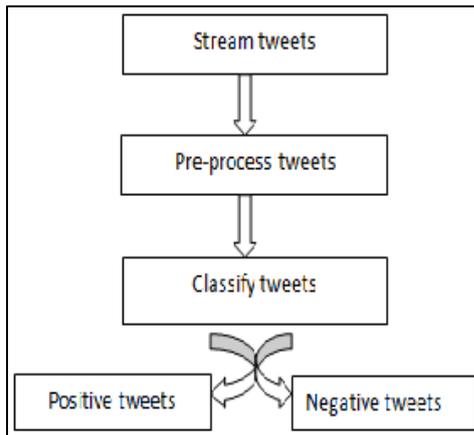


Fig. 1: flow diagram for sentiment analysis

1) *Methodology:*

- 1) Step-1 First we are going to stream tweets in our build classifier with the help of Tweepy library in python.
- 2) Step-2 Then we pre-process these tweets, so that they can be fit for mining and feature extraction.
- 3) Step-3 After pre-processing we pass this data in our trained classifier, which then classify them into positive or negative class based on trained results.

B. *Text Mining*

The first step for this process involved a research how to extract data from Twitter. We find Twitter API.

API that allows us to perform various tasks, being one of them the extraction of data from Twitter platform. This API also provides all core functionalities of Twitter application using Rest web services. For the goal of this paper, we used the Feed connection of the Page object. The Feed represents the list of all Post objects, which contains details like post, likes, comments, polarity (reactions or emotions) and shares. Still, there are some concerns with API like the total number that is called. When the API is called a big number of consecutive time, gets block. For example, if we want get the total number of posts of a page, it will be block at same time. So, the workaround its extract the data with time intervals. After finding or creating Twitter API, they provide consumer key ,consumer secret, access token and access secret. Using these parameters, anyone can make a custom implementation and also can extract data, as we made one to adapt for our research purpose.

C. *Text Pre-Processing*

After mining data from twitter, we start our analysis by breaking the text down into words. Tokenisation is one of the most basic step in text analysis, which is used to split a stream of text into smaller units called tokens, usually words or phrases.

Text pre-processing also involves removing special characters and emoticons from splitted words. Special characters and emoticons are: ☺, ☹, :, :[, :), :(, ☹ and so on.

Tweet Type	Result
Original Tweet	@XYZ I am so happy for you and thanks for joining the dinner with us.
Pre-Processing Tweet	happy,thanks,joining

Table 1: example for text preprocessing of tweets

D. *Term Frequency*

After collecting data and pre-processing some text, we are ready for some basic analysis, the analysis of term frequencies to extract meaningful terms from our tweets. Term frequency analysis gives the most commonly used terms in the document. It is also used as a weighting factor in information retrieval and text mining. The TF (term frequency) of a word is the frequency of a phrase or text (i.e. number of times it appears) in a document.

Term Frequency can be calculated by:

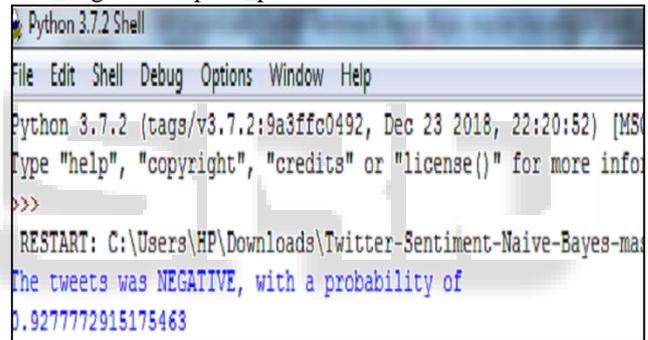
$$TF = \frac{\text{Number of times term T appears in the particular tweet}}{\text{number of terms in that tweet}}$$

E. *Sentiment Analysis*

Sentiment analysis using Naive Bayes(NB) Algorithm: Naive Bayes Classifier is a classification algorithm that relies on Bayes' theorem. Twitter Sentiment Analysis, defined as using advanced text mining techniques to analyze the sentiment of the text (here, tweet) in the form of positive and negative. Extracted polarity indicates the sentiment where value nearer to 1 means a positive sentiment and values nearer to 0 means a negative sentiment.

$$\text{Prob_positive} = \frac{\text{float(positive_instance)}}{\text{positive_instance + negative_instance}}$$

$$\text{Prob_negative} = 1 - \text{prob_positive}$$



III. CONCLUSION

Sentiment analysis is employed to distinctive people's opinion, perspective and emotional states. The views of the people can be positive or negative. The parts of the phrases are used to extract the sentiment of the text. An adjective plays a crucial role in identifying sentiment from parts of speech. It is difficult to identify sentiment and opinion when words having adjective and adverb are used together. To do the sentiment analysis of tweets, first step is to extract the twitter posts from twitter by user. The system can also compute the frequency of each term in tweet. Using naive bayes algorithm help us to obtain the results.

Twitter is massive supply of information, which make it more attractive for performing sentiment analysis. We perform analysis on around 15,000 tweets total for each party, so that we analyze the results, understand the patterns and give a review on people opinion. We saw totally different party have different sentiment results consistent with their progress and dealing procedure. We also saw how any social event, speech or rally cause a fluctuation in sentiment of people. Provides user's personality which is predicted from text written on Twitter. Provides user's personality which is predicted from text written on Twitter.

Users' degree of reflection between the real world and the TWITTER world was measured by the users' way of developing friendship. Finally, it provides the degree of reflection and the probability of the polarity. It also can be used for finance, marketing, reviewing and many more.

IV. FUTURE ENHANCEMENT

In this project, We tried to show the basic way of classifying tweets into positive or negative category using Naive Bayes as baseline. We could further improve our classifier by trying to extract more features from the tweets, trying different kinds of features, tuning the parameters of the naïve Bayes classifier, or trying another classifier all together.

REFERENCES

- [1] K. Shin and J. Lee, "A job applicants' résumé verification method using a social network analysis Using Facebook like as LinkedIn for a recruiting," 2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), Malabe, 2017, pp. 1-5
- [2] B. Y. Pratama and R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM," 2015 International Conference on Data and Software Engineering (ICoDSE), Yogyakarta, 2015, pp. 170-174.
- [3] Zaroor, M. Maree and M. Sabha, "JRC: A Job Post and Resume Classification System for Online Recruitment," 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), Boston, MA, 2017, pp. 780-787.
- [4] Teixeira and R. M. S. Laureano, "Data extraction and preparation to perform a sentiment analysis using open source tools: The example of a Facebook fashion brand page," 2017 12th Iberian Conference on Information Systems and Technologies (CISTI), Lisbon, 2017, pp. 1-6.
- [5]] R. Jiamthaphaksin and T. H. Aung, "User preferences profiling based on user behaviors on Facebook page categories," 2017 9th International Conference on Knowledge and Smart Technology (KST), Chonburi, 2017, pp. 248-253.
- [6] <https://medium.com/@koshut.takatsuji/twitter-sentiment-analysis-with-full-code-and-explanation-naive-bayes-a380b38f036b>.
- [7] Cantador, Iván & Fernandez-Tobias, I & Bellogín, Alejandro & Kosinski, Michal & Stillwell, David. (2013). Relating Personality Types with User Preferences Multiple Entertainment Domains. CEUR Workshop Proceedings. 997.
- [8] H. Parveen and S. Pandey, "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm," 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Bangalore, 2016, pp. 416-419.
- [9] M. Kumar and A. Bala, "Analyzing Twitter sentiments through big data," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 2628-2631.