

Twitter Trend Analysis and Notification

#¹Sanjeetkumar Mehta, #²Vipul Oswal, #³Shivprasad Patil



¹sanjeetkmehta@gmail.com
²vipuloswal34@gmail.com
³shivprasad.patil@sinhgad.edu

#¹²³Department of Information Technology,
 NBN Sinhgad School of Engineering, Ambegaon,
 Pune, Maharashtra

ABSTRACT

Regardless of the extensive acceptance of Twitter internationally, little exploration has investigated the differences among users of different languages. In prior research, the natural tendency has been to assume that the behaviours of English users generalize to other language users. We studied a few thousand-tweet collected over a days' period and found that more than 25 languages were used. Only 34.8% of the tweets were in English. Other popular languages including Japanese, Portuguese, Korean, and Spanish together accounted for 49.3% of the tweets. Examining users of the top 10 languages, we discovered cross-language differences in adoption of features such as URLs, hashtags, mentions, replies, and retweets. This is the first part of our paper where we discuss our work's implications for a more thoughtful and successful analysis of sentiments because the language used does play a vital role in determining the behaviours of the users. Secondly, a majority of the population in India is not aware about all the latest happenings around the world mainly because they are not connected to the Internet. Everyone in and around the world should be cognizant of all the burning topics from all the nicks and corners of the world. In the second part of this paper, we aim to contribute a new and supervised approach of serving the purpose of acquainting the far-reaching explicit users.

Keywords: Language Analysis, Notification.

ARTICLE INFO

Article History

Received: 1st June 2017

Received in revised form :

1st June 2017

Accepted: 4th June 2017

Published online :

6th June 2017

I. INTRODUCTION

Twitter is a free social networking microblogging service that permits registered members to broadcast short posts called tweets. Twitter members can broadcast tweets and follow other users' tweets by using multiple platforms and devices. Tweets and responses to tweets can be directed by mobile client or by posting at the Twitter.com website. Twitter is a broad platform which is trafficked with diverse and revealing briefs. An unsupervised approach is proposed in this paper to model the analysis of tweets and different languages adaptively and is employed to learn semantic representation of tweets to enlighten the far-reaching explicit users. In the first part of this paper, we intend to reveal how different languages used on twitter help in implying the behavioural and cultural differences amongst communities which in turn does affect the sentiment analysis of different communities and users. To the best of our knowledge, our research is the first to

methodically revise how users of different languages behave in Twitter and how it affects the sentiment analysis which can later be deliberated to ensure more practical and accurate analysis of sentiments. We intend to present a small part that will aid in more effectual and dependable analysis of a much larger segment typically the sentiments. We discourse two important queries:

(1) What are the most popular languages used in twitter? Can we validate and outspread this result?

(2) Are there perceptible behaviour variances unveiled by users of different languages? How do users of various languages differ in their insertion of URLs, hashtags, and mentions? What do they intend to use Twitter for? How does this affect the sentiment analysis? [1].

To address these questions, we analysed a few thousand tweets, grouped users into language communities, and then compared key behaviours across communities. In this paper, we report two contributions in our first part.

First, we identify the top 10 languages used in Twitter. Second, we show how significant differences across the top 10 languages in the use of Twitter-specific conventions such as URLs, hashtags, mentions, replies, and retweets reveal their intention of using twitter and how it affects the important analysis of sentiments.

India have a population of 1.22 billion. There are 100 crore mobile users in India (1 billion). This implies that 80% of India is connected to mobile phones. However, only 34% of India's population is connected to the Internet. So, most of the people are oblivious of the latest happenings around them.

According to a report by market research firm eMarketer, in India, just 1.8 percent of the population uses Twitter which will barely budge reaching only 3 per cent by 2018. So, an enormous proportion of population is not cognizant about the most trending subjects in and around the world. This gap can be abridged by making them mindful about the latest happenings through their mobile phones. In the second part of this paper, we intend to specify an approach of creating a portal as a key to the above constraint in order to help make people mindful about latest trends and up-to date topics via the use of Short Message Service(SMS).

II. WORKING PRINCIPLE AND EXPERIMENT

A. Language Analysis

In our study, we have found out that language does play a very vital role in the Sentiment Analysis and is significant in deciding the accuracy of Sentiments analysed. There has been a related work on Sentiment Analysis which is independent of language used [3]. Here, we discuss the working principle supported by an experiment that we had performed.

Why is Language Analysis important and integral to Sentiment Analysis?

Twitter is one of the biggest microblogging services on the internet. Microblogs are short text messages that people use to share all kinds of information with the world. On Twitter, these microblogs are called "tweets", and over 400 million of them are posted every day. They can contain news, announcements, personal affairs, jokes, opinions and more.

Sentiment Analysis is extracting methods of analysing people's opinions from tweets. A single piece of opinion from one person may not seem important, but among the billions of tweets the collection of opinions can form a comprehensive picture. People's opinions and sentiments about products, other people and events in large numbers are invaluable. Companies can use them to gain feedback on their products, to do market research and analyse

customer demands, and for a wide variety of other applications.

Language Analysis is important as different users of diverse communities use different languages and differ in opinions and purposes of using Twitter. These differences in their reasoning, purposes and opinions does relate to their changing behaviours. These behaviours may range from inclusion of URL's or hashtags or mentions to retweeting and replying to certain posts. This does play a significant and crucial role in sentiment analysis. The rate at which a particular user includes URL's or hashtags or mentions and the rate at which he retweets and replies implies and reveals his intent of using Twitter. Specifically, if a user includes URL's and hashtags, that particular user typically is interested in Information Sharing. A mention is generally used to either attract someone's attention or acknowledge someone's association to the content of the tweet. Both are cases of inherently social acts and resemble public conversations in groups of people. Interestingly, some communities such as Korean and Indonesian exhibited more of this social behaviour than others. A reply, a specific form of mention with @username appearing at the beginning of the tweet, is a tweet responding to a previous message. So, different users of varies communities using diverse languages differ in the above mentioned behaviours and so does their intent and purpose of using Twitter. This implies that it is very necessary to consider Language Analysis as a part of Sentiment Analysis in order to have more precise results.

We had performed a small analysis on Twitter which is as follows:

1. Most used languages used for all tweets.

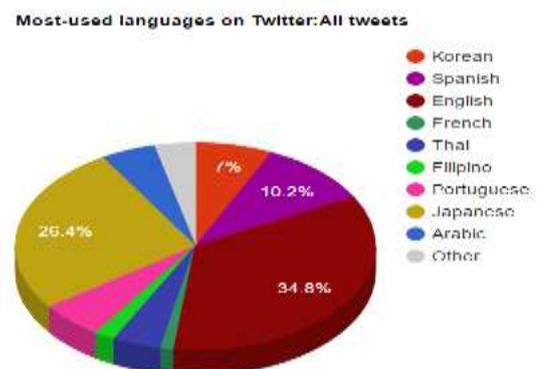


Fig 1: Most used languages on twitter: All Tweets

The above analysis was conducted on a few thousand tweets to understand the most used languages on twitter. Out of these tweets, some tweets were filtered out to be

top tweets on the basis on retweet count and an analysis was done on these top tweets which is as follows:

2. Most used languages for top tweets.

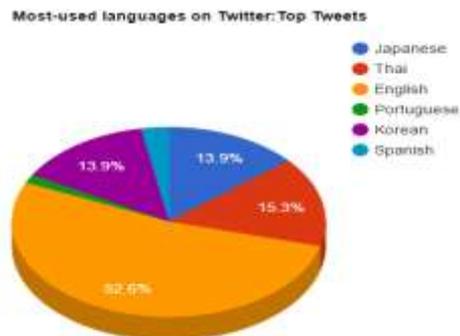


Fig 2: Most used languages on twitter: Top Tweets

The basic purpose of these analyses is to imply that as language differs on Twitter, the behaviour also differs. This has to be taken into consideration while performing sentiment analysis. Some of the variations might be attributed to inherent cultural differences. Others might be due to how long Twitter had been used by a language community, how many people actively used it, whether users were spread out geographically, and how many bilingual brokers there were to spread conventions and practices from one language community to others, etc. Precise results in Sentiment Analysis can thus be obtained and a more systematic approach can be followed with the help of Language Analysis.

B. Notifying top news and trends via SMS.

Twitter is a microblogging service where a large population in and around the world is constantly updating latest trends with the help of short posts called as tweets. There is a large extent of varied and real-time information on Twitter. This is due to nature of microblogs on which people post real time messages about their opinions on a variety of topics, discuss current issues, complain, and express positive sentiment for products they use in daily life. In fact, companies manufacturing such products have started to poll these microblogs to get a sense of general sentiment for their product. Many times, these companies study user reactions and reply to users on microblogs.

Also, we have already mentioned about how an abundant part of the population is equipped with mobile phones. We also know that there is large evolution the field of wireless technologies and use of internet is splurging widely but a surplus number of people do not have constant and instant access to the internet due to which they are not connected with all the hot and trending news

and events in the nicks and corners of the world. It is very important, in today's century, that all people be connected to and be aware of all the happenings. In this paper, we propose to analyse the top tweets and trends as in part 1 and then notify the populous about the trends via Short Message Service(SMS). [2]

What is the importance of SMS?

Modern mobile telephony has produced a number of path breaking applications. But among the various applications and services, text messaging is still the universal platform for the masses. Among the various facilities available within the mobile communication system, SMS (Short Message Service) is the oldest and most used till date. This is basically because it does not require special downloads as it is already available on all the mobile phones. In principle, text message can be used either as a one-way communication to provide the user information such as reminder, alert, etc., or as a two-way communication that enables the user to send and receive information (such as question and answer). Event notification (through SMS) is a well-known way of notifying users about an event scheduled to take effect within a particular period in an institution. Moreover, Mobile text messages are an excellent aid for communication when there is a need to submit information also at long distances or without well working communication system or infrastructure, or when the people cannot physically meet the staff that is concerned, provided that the cost of the text message is very low and it is available to practically everyone.

III.FUTURE SCOPE

We have analysed the tweets that are trending on twitter and then we will provide a feasible approach to let the users' select their domains of interests. Accordingly, the registered users will be updated about the trends and news of their interests via the use of SMS. [4]

Also, general notifications will be forwarded to all of the users irrespective of their interests which can be disasters or epidemics, emergency crisis and basic important information that everyone must be well aware of.

IV. CONCLUDING REMARKS

While there has been much work on characterizing general Twitter usage, we know of no large-scale in-depth studies that compare the behaviours of different language users in Twitter. Since Internet usage is a global phenomenon, studies of how users perceive and behave on social websites will become increasingly important. In this paper, we have used the phrase "language community" somewhat loosely, but it is clear that

languages serve as barriers in information diffusion. In prior research, the natural tendency has been to assume that the behaviours of English users generalize to other language users. However, since many communication tools are global, we need to examine critically whether indeed users of non-English languages behave in similar ways. In this paper, we make two key contributions. We identified the most popular languages used in Twitter. The communities differed considerably on using specific Twitter conventions. Our findings can help designers of cross-cultural communication tools to take into account the differences between languages. These differences in behaviours must be critically considered in order to precisely help in developing accurate analysis reports of sentiments.

In the second part, we have seen how people not connected to the Internet can be made aware of all the trends and news in and around the world. Using this, we intend to help a massive population by making them aware about the latest happenings and trends in the society. Awareness about natural calamities, epidemics and the necessary precautions can be spread amongst a number of people via SMS.

ACKNOWLEDGMENT

The authors would like to thank Mr. Lalit Bhattacharaya of the Persistent Systems and Professor Shiv Prasad Patil of the Information Technology Department, NBN Sinhgad School of Engineering, Pune for their continued and unconditional guidance. Both exemplary teachers and magnificent personalities, we consider ourselves lucky to have been advised by them and to have worked under their supervision. Without their co-operation, the preparation of this paper would not have been possible.

REFERENCES

- [1]. Lichan Hong, Gregorio Convertino, and Ed H. Chi, "Language Matters in Twitter: A Large-Scale Study", Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.
- [2]. Oludare Olaleye, Ayodele Olaniyan, Olalekan Eboda, Adeleke Awolere, "SMS-Based Event Notification System", Journal of Information Engineering and Applications ISSN 2224-5782 (print) ISSN 2225-0506 (online) Vol.3, No.10, 2013.
- [3]. Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau, "Sentiment Analysis of Twitter Data".

- [4]. Ivan Marcin, Sam Shiu, "Extracting Topic Trends and Connections"