A Dictionary Based Analysis of User’s Sentiment Regarding Indian Premier League

Chayan Paul, Pronami Bora

Abstract: The growth of social media has provided the users with a platform to express their views on numerous themes. Social networking sites like Twitter are considered as large source of users’ sentiment. Twitter has become one of the biggest sources for evaluating sentiment analysis. The shorter and informal nature of the text encourages the users to express their sentiment fast and effectively. The huge amount of data that gets generated mostly in text format can be used for studying user’s sentiment regarding any topic. Indian Premier League (IPL) is a cricket tournament of T20 format that draws a lot of attention from the viewers. Right from the very beginning IPL has remained in the glare for consecutive 12 years. Because of the participation of renowned players from throughout the globe, some famous Bollywood personalities and businessmen, this tournament remains one of the topics for discussion. In this paper, we propose to study the users sentiment related to IPL using twitter data. The tweets related to IPL are proposed to be downloaded and analyzed to find out the sentiment regarding IPL.

Index Terms: Sentiment Analysis, Twitter, IPL, Opinion Mining

I. INTRODUCTION

Sentiment analysis or opinion mining is a research area that got considerable amount of attention, during the last few years. One of the predominant reasons for the development is the rise of microblogging sites and online shopping sites. Twitter has 330 million active users and on a daily basis, there are 500 million tweets are posted in the web [1]. Individuals in different field of work like to express their opinion regarding various issues related to their day to day life. They have a habit of recording these issues more frequently [2]. Social networking sites provide the users with a platform to express the sentiment in short and informal manner. These messages play a significant role when it comes to understanding the user sentiment or opinion about a particular event or entity. People are relying on these sources for information, reviews and opinions about any particular entity. Users of these microblogging sites continuously express their opinion related to various product or services they use in their day to day life. This also gives the manufacturer and the service providers a platform from which they can understand the opinion of the users. The growth of these messages has provided researchers with a good amount of data that can be studied for different purposes and objectives.

Twitter is a microblogging site that allows its registered users to express their views using a limited set of characters. The registered users can also read any message publicly available on twitter. Initially the allowed message length was only 140 characters, but it was raised to 280 characters on November 07, 2017 [3]. Right from its inception in the year 2006, it got popularity because of the reasons like a registered user can follow any public profile in twitter, and directly interact with any celebrity, sportsman, politician, business man etc. Twitter also got its popularity because of the length of the messages. These collection of tweets are often considered as a good source of opinion especially because of huge number of contributors. The compact length of the tweets inspires the users to use more symbols like smileys and truncated forms of the words [4]. Twitter allows its user to put up their views and opinions on various different topics. It can range from their political views to religious view, or it may be a review on a particular product or service. This huge collection of textual data has motivated the researchers working in the area of Natural Language Processing (NLP)[5].

Indian Premier League (IPL) is a twenty20 cricket league played in India among privately owned franchises. Currently it has 8 teams and it is usually played during the month of April and May every year. There are 56 league matches, two semifinals and a final played among the various franchises to decide the final winner. Hence in total there are 59 matches to be played in the tournament [6]. The popularity of the tournament is mainly because of the reasons like India’s huge sporting market of over a billion people, India being the hub of International cricket, the massive fan following for cricket in India and involvement of popular retired cricketers in IPL [7].

Twitter has become one of the most popular platforms to connect with different people. It has become a common trend for politicians, celebrities, sport stars, public figures and many more to connect with their followers through twitter bypassing the normal root of journalists. People involved in professional sports, athletics and the franchises involved in these areas are using the social networking and micro blogging sites progressively [8]. Various cricketers, their respective franchises keep updating their fans and followers through twitter about different events. Twitter users often use this platform to put their opinion regarding any player or a franchise. There can be criticism or appreciations for a player or a franchise. The researchers in this paper endeavor to find out the sentiment score of the users related to Indian Premier League.

II. REVIEW OF LITERATURE:

Sentiment analysis in twitter is a huge area of research. Researchers try to understand the sentiment of people in varied fields using twitter data. Twitter has become successful to provide the users with platform for highly valuable textual data. These
data drew considerable interest of the researchers, in the field of Sentiment analysis [9]. The twitter provides with huge collection of datasets especially in the field of opinion mining and sentiment analysis. Because of the ease of access and the popularity of the format of these messages, users prefer these microblogging sites over the traditional blogs or mailing list [10].

There has been considerable research in the field of opinion mining and sentiment analysis in the recent years. Medagoda et. el [11], presented some widely used methods in sentiment analysis and a detailed review of sentiment analysis in various non English languages like Hindi, Russian and Chinese. Rathan el et [12] proposed a model for sentiment analysis and mining twitter data, which can process features such as emojis, emoticons etc. Their proposed model has the feature of automated training data labeling by using lexicon based approach. Davidov et el [13] proposed sentiment classification framework using supervised learning. In their study, they considered 50 twitter tags and 15 smileys to develop the framework. The proposed framework avoids manual annotation. It can also be used for classification and identification of diverse sentiment texts. Taylor et el [14] extended aspect-based sentiment analysis techniques and applied it to the tourism domain. Using this extension, they tried to discover consumer preferences about different tourism products. Turney et el [15], in their study, proposed a strategy for inferring semantic orientation from semantic association.

Hu et el [16] conducted their study in three different steps, (1) mining features of the products given by the customers; (2) detecting opinion sentences in each review and their polarity; (3) summarization. They also proposed some novel techniques to perform these tasks. They demonstrated the effectiveness of the process through experimental results. Singh et el [17] presented a detailed review of various articles dealing with the research. Gopalkrishnan et el [18] applied neural network based methods for determining the opinion of users from social web in health care domain. Tumajjan et el [19] conducted a study with three aims, i.e., First they examined whether Twitter has any significant role in carrying peoples political opinion. Second, they evaluated if Twitter messages reflect the recent political ecosystem to a greater details. Third, they explored if the analysis of twitter can be helpful in predicting a winner in the elections. Turney [20] proposed an algorithm which can be used to classify reviews as recommended or not recommended. The classification of the reviews is defined by the average semantic orientation of the various words available in the review. Mostafa [21] conducted a systematic exploration of a large corpus of 1,00,000 tweets related to halal foods for finding the opinion related to halal foods. Along with the sentiment involved in each of the tweets he also analyzed the geographic location of the users who posted the tweets.

III. CORPUS COLLECTION AND PREPROCESSING

The corpus for analysis was collected from Twitter by using the R packages like twitteR [22] and ROAuth [23]. These two packages provide us with the functionalities to extract and store the tweets related to topics of one’s interest. All the tweets related to Indian Premier League were downloaded. Different keywords like IPLt20, iplt20, IndianPremierLeague were used to extract the relevant tweets. Using this process around 3,500 tweets were downloaded and stored for further processing and analysis. After initial preprocessing steps, the collection of tweets was processed using Syuzhet [24] for calculating and visualizing the sentiments in the tweets. The get_sentiment function allows getting the sentiment involved in the sentences. The get-sentiment function comes with five different methods, i.e., syuzhet, bing, afinn, nrc, and stanford. Among the five available methods, syuzhet is the default method. Each of the different methods has slightly different process for calculating the sentiment involved in sentences. A detailed analysis of the available corpus was done using the package and results are presented in the following section.

IV. RESULTS

Figure 01, represents a word cloud drawn from the corpus. The word cloud was drawn using 200 words that have frequency of at least 50. Frequency was kept at 50 so that the important words appear in the word cloud. As seen from the figure, ipl, cricket, team, test are some of the most frequently appearing words in the corpus.

Figure 02: Graphical representation of the sentiment calculated using four methods
Each of the methods calculate the sentiments with slightly different scale, hence there is a slight difference between the graphs of the different methods. In overall there are much more positive sentiments involved in the corpus compared to the negative sentiment. This observation can also be validated by using the sum of the vectors used for these different methods. The respective sums of the different methods are given in the table 01 below:

<table>
<thead>
<tr>
<th>Sl No</th>
<th>Method</th>
<th>Sum</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>snyuzhet</td>
<td>1092.95</td>
<td>0.3179029</td>
</tr>
<tr>
<td>2</td>
<td>bing</td>
<td>896</td>
<td>0.2606166</td>
</tr>
<tr>
<td>3</td>
<td>afin</td>
<td>2206</td>
<td>0.6416521</td>
</tr>
<tr>
<td>4</td>
<td>nrc</td>
<td>1318</td>
<td>0.3833624</td>
</tr>
</tbody>
</table>

Table 01: overall emotional valence of the tweets for different methods

V. CONCLUSION AND FUTURE WORK:

In this paper, we tried to calculate and present the sentiment expressed by the users regarding Indian Premier League, in twitter. The overall sentiment score in the tweets were found to be highly positive. Although there are some amount of negative sentiment seen as evident from the graphical representations. But these negative sentiments are almost negligible in comparison to the positive sentiments expressed by the users. This echoes that Indian Premier League remains one of the most liked events throughout the years.

In future work, a detailed analysis of the sentiment expressed by users every year can be collected and a time series analysis of the sentiments regarding Indian Premier League can be done to understand the change in popularity of IPL throughout the years.

REFERENCE


AUTHORS PROFILE

Dr. Chayan Paul is working as Associate Professor in the Department of Computer Science, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India. He completed his PhD from Assam University. He has published a good number of papers in national and international journals. He has also attended a good number of national and international conference. His research interest include Machine Learning, NLP, Sentiment Analysis.

Pronami Bora is an Assistant Professor in the Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India. She is currently pursuing her Ph. D from the Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India. She has a good record of publication in both national and international journals. She has also attended a good number of national and international conferences. Her research interest includes 3D printed microstrip patch antennas for wireless applications.