Twitter Intention Classification Using Bayes Approach for Cricket Test Match Played Between India and South Africa 2015

Varsha D. Jadhav, P.E.S. College of Engineering, Aurangabad, India
Sachin N. Deshmukh, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India

ABSTRACT

Information retrieval and forecasting in real time is becoming the fastest and most efficient way to obtain useful knowledge of what is happening now, allowing organizations to react quickly when problem appears which help to improve their performance. There is enormous amount of data in the form of tweets. It builds data processing system that creates informative data about the cricket test matches. Using twitter data, the authors find the sentiments or polarity of fans posting tweets related to game. Polarity is given as positive, negative and neutral. The authors also analyze the feelings or emotions of people posting tweets. Emotions are given as anger, disgust, fear, joy, sadness, surprise and unknown. Machine learning algorithm (Bayes) using R technology shows the accuracy when trained with emotion data.

KEYWORDS

API, Bayes, Cricket, Emotions, Intention, MAE, Polarity, Sentiments, Tweets, Twitter

INTRODUCTION

In the early 188s cricket matches were not time limited. There was unlimited time for the teams to chase a target or to bowl their opponents twice. In the timeless matches, only nine innings were declared. In 1039, the fixed-duration format was universally adopted, and in the present day, test matches are played for five consecutive days with three sessions in a day. A team bowls their opponents twice within the time limit of five days. It is like-wise for batting. We summarize the match-state at the end of each day and use the match-state to forecast the match results. These forecasts can then be used by team captains as strategic guidance.

Twitter is popular social media where users create messages called tweets. These tweets express polarity and emotions of the people about a particular topic. This is very helpful because it allows feedback to be aggregated without manual interference.

Intentions can be determined by evaluating a piece of text. Short text messages are challenging to determine the intention. They are usually short in length and are expressed in one sentence or even less. They have misspellings, slang terms, and short form of words. They also include special symbols. Emotions are the facial expressions which are pictorially represented using punctuations and letters. Twitter user use @ symbol to refer to other user on the microblog. This automatically alerts them. Hashtags are usually used to mark the topics to increase the visibility of the topics.

In this paper we present the Bayes model to predict the intentions of the people who tweet on twitter about a specific topic. The topic which we have chosen is the cricket test match played...
between India and South Africa from 3rd December to 7th December 2015. The tweets were extracted from 3rd December to 7th December 2015 at the end of the day. It uses Bayes method twice. Once for determining the polarity and the second time for determining the emotions of the fans. Using polarity intention the match outcomes are predicted. This outcome obtained at the end of the day, serves as strategic guidance to the captains so as to improve the performance of the team.

BACKGROUND OF CRICKET

Cricket is the popular game similar to other bat and ball games. Each team is made up of eleven players with each team taking turns for batting and fielding. Each team takes turn called an inning. The aim of the batting team is to make runs whereas the aim of the fielding team is to get the ten batsmen out, which is called taking wickets. Currently three varieties of cricket matches are played at international level: test cricket, ODI and twenty20 (T20).

LITERATURE SURVEY

To achieve the objective of intention mining the literature survey was done. The literature survey was carried out for cricket match and also for opinion mining. (Yuan, Murukannaiah, Zhang et al., 2014), conducted experiments using the mention and mutual-follow graphs. For the mention graphs, the number of “@” references between users can be viewed as the strength of a tie. (Vikrant Hole, Mukta Takalikar, 2013), proposed a system of scheduled subevents for games using twitter data. They also analyzed the feeling of people posting tweets. Sporting events, specifically FIFA World Cup 2015 and IPL 2015 were summarized. SVM, NaïveBayes and Logistic Regression classifier for classifying sentiment of fans were used. Tweets were fetched using Twitter API v 1.1. The system was tested for different matches IPL2015, qualifier 1, eliminator, qualifier 2 and final match and FIFA2015 final match. (S.M. Crowe, Jennifer Middledorp, 1996), used logistic regression to analyze data from the test cricket series played in Australia from 1977 to 1994. They considered controversy surrounding leg before wicket (LBW) decisions. They collected data from Frindall(1990a,b) and from the test match scoreboards published in the Sydney Morning Herald (1992-94) and the Sydney Sun Herald(1992-94). Logistic regression analysis was used to compare LBW rates for visitors with those for Australia. (Sohail Akhtar and Philip Scarf, 2011), forecasted match outcomes in a test cricket play, session by session. At the start of the session, match outcome probabilities are forecasted using a series of multinomial logistic regression models. Data was obtained from the ESPNcricinfo website (ESPNcricinfo, 2010). The complete dataset (146 matches) relates to all the test matches in the period between November 2005 and March 2010, apart from those matches where the session by session data were not available and in which more than 90 overs were lost to poor weather. International Cricket Council (ICC, 2010) official ratings were used to measure team strength. (Sacheti, A.,Gregory-Smith . I. and Paton, 1998), explored the impact of social pressure exerted by crowds on the decision making of officials in sporting contests by using data from Test cricket matches. They focused on area of decision whether a batsman is LBW out or not. Data from 1000 test matches from 1986 to 2012 was used. Negative binomial regression models were estimated to identify the extent of bias towards the home team. (P. E. Allsopp and Stephen R. Clarke, 2004), applied multinomial regression techniques to rate the cricket teams and analyze outcomes in test cricket. (Bing Liu., May 2012) gave in-depth introduction to Opinions and its related concepts such as sentiments, valuations, attitudes, and emotions that are subjects of study of sentiment analysis and opinion mining. (Blessy Selvam, S.Abirami, 2013), gave a brief survey on the opinion mining framework. (Zheng Chen, Fan Lin, Huan Liu, Yin Liu, Wei-Ying Ma, Liu Wenyin, 2002), surveys probabilistic approaches to modeling information retrieval. In information retrieval (IR), probabilistic modeling is the use of a model that ranks documents in decreasing order of their evaluated probability of relevance to a user’s information needs. (E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse,1998), proposed novel
Visualization and Analysis of Frames in Collections of Messages: Content Analysis and the Measurement of Meaning
www.igi-global.com/chapter/visualization-analysis-frames-collections-messages/63270?camid=4v1a

The Concept of Modularity in the Context of IS Project Outsourcing
www.igi-global.com/chapter/the-concept-of-modularity-in-the-context-of-is-project-outsourcing/184235?camid=4v1a