Human Action Prediction Using Sentiment Analysis on Social Networks

Tatenda Duncan Kavu, Chinhoyi University of Technology, Chinhoyi, Zimbabwe
Tinotenda Godknows Nyamandi, Chinhoyi University of Technology, Chinhoyi, Zimbabwe
Allela Chirinda, Chinhoyi University of Technology, Chinhoyi, Zimbabwe
Talent T. Rugube, Chinhoyi University of Technology, Chinhoyi, Zimbabwe
Kudzai Zishumba, Chinhoyi University of Technology, Chinhoyi, Zimbabwe

ABSTRACT

There is a rapid increase of mass demonstrations in different locations worldwide triggered by social networks discussions, as witnessed in the USA, Egypt, and South Africa. This paper challenges the underutilization of social media to detect people’s mood and to predict their actions based on their sentiments. Recent published work has demonstrated utility of sentiments on Twitter to predict outcomes of different events, so to come up with the geographical action prediction tool the authors utilized geocodes, sentiment analysis, probability theory, and logistic regression. The tool informs relevant authorities like governments to know the state of people’s moods. Entities like business enterprises also benefit from this tool in their plans, especially in avoiding unnecessary costs due to infrastructure destruction.

KEYWORDS

Forecast, Geocode, Mood Detection, Prediction, Sentiment Analysis

1. INTRODUCTION

Social media platforms are not utilized to detect people’s moods and predict their actions as they are used to send messages world over to start demonstrations and nation uprisings. Social media is being used by terrorist groups while there are no methods used to tap into this information (Schulson, 2015). Sentiment, opinion, and action are three essential aspects of user attitude which have drawn much attention in various existing research work. Investigating them together not only brings unique challenges, but can also help to better understand a user’s online behavior and benefit a set of applications related to online campaign and recommender systems. None of the existing research on opinion and sentiment analysis predicts the likelihood of taking an action based on the current sentiment/opinion toward a topic (Gao et al., 2014).

The aim of this research is to detect people’s mood using sentiments on social media based on their geographical location and capture their likelihood of taking an action. To fulfill the goal, we designed a machine learning tool that crawls social media to gather sentiments for analysis. We utilized an established theoretical framework in psychology and marketing on attitudes and attitude models, where attitude is defined as a unified concept containing three aspects: “feelings”, “beliefs”, and “actions” (Eagly & Chaiken, 1993). Moreover, the Logistic regression and Probability theory models were used to predict action based on the analyzed sentiments. The computational tool further
implements topic extraction and mathematical functions, to detect feelings, beliefs, and actions from users’ sentiments. The Web holds an enormous amount of opinionated data which, if correctly scrutinized, can significantly help policymakers, social scientists, businesses, and society to a great extent (Verma et al., 2014).

A human action prediction tool can be an excellent source of information and can provide insights that determine marketing strategy, improve campaign success, improve product messaging, improve customer service and generate leads. The study of sentiment analysis, if conducted properly, is exceptionally complex and is actually a field of study, not just a feature in a social media tool. The main contribution in this article is the prediction of a major or minor action considering a geographical location without being affected by location of an opinion writer. This tool can be used by relevant authorities such as government for infrastructure protection, analyzing people’s opinions about pending policies or government-regulation proposals.

The remainder of the article is organized as follows: Section 2 shows the related work which looks at the available technologies as well as the progress that has been made in the area of concern (i.e., human action prediction). This section also outlines the significance of investigating challenges previously experienced by researchers and possible solutions. Section 3 demonstrates the software and hardware development tools used in the implementation and development of the Human action prediction tool. Algorithms which this tool followed are also elaborated in this section. Section 4 describes the architecture and system design of our tool. The strategy used to derive the final results is described in Section 5. The discussion and analysis of results in Section 6 indicates the importance of the tool not only recognized in Zimbabwe but globally within various business environments. Section 7 explains the implications of the findings. The section titled ‘Recommendations’ is the 8th section and it demonstrate other areas where this tool can be relevant. Section 9 is the future work. Finally, section 10 consists of concluding remarks.

2. RELATED WORK

2.1. Studies Using Lexicon-Driven Methods

O’Connor et al. (2010) used the Multi-Perspective Question-Answering (MPQA) sentiment lexicon to identify sentiments in tweets mentioning former US President Barack Obama. The MPQA lexicon is a collection of 535 English-language news articles from a variety of news sources manually annotated for opinions and other private states (i.e., beliefs, emotions, sentiments, speculations, etc.). To classify a tweet, they simply counted if it contains more positive or negative words according to the sentiment lexicon. Even though this is a very simple approach, they report significant correlation between the aggregated sentiment in tweets and public opinion polls. Thelwall et al. (2010) proposed a lexicon-based algorithm called SentiStrength, which assigns a polarity (positive/negative) and corresponding strength value between 1 and 5 to a given text. In addition to their list of 298 positive and 465 negative terms annotated with polarity and strength values, an algorithm (SentiStrength) uses lists of emoticons, negations, and boosting words in the decision process. To deal with emphatic lengthening the authors proposed a 3-step method to reduce words to their standard form. When comparing SentiStrength to various machine learning classifiers on Myspace comments, it performs better in classifying negative comments than positive comments.

2.2 Studies on Term Frequency—Inverse Document Frequency

To determine whether or not each comment is on-topic (relevant to the focal story), a standard information retrieval technique called “TFIDF” (Term Frequency–Inverse Document Frequency)
Replicating Silicon Valley in China: An Industrial Upgrade and Expansion Approach
www.igi-global.com/article/replicating-silicon-valley-in-china/79232?camid=4v1a

Perceived Usefulness and Ease of Use of Mobile Government Services: The Moderating Impact of Electronic Word of Month (eWOM)
www.igi-global.com/article/perceived-usefulness-and-ease-of-use-of-mobile-government-services/242988?camid=4v1a