ASSIGNMENT 3: TAKE HOME EXAM

Question 3

Twitter and Predicting Virality

This report outlines a possible prediction model that can be utilised to ensure a Twitter message will spread as a meme and become viral. This assignment discusses the potential ethical and social concerns and consequences of this study.

Spring 2019

09/10/2019

31005 Machine Learning Spring 2019

Assignment 3: Take Home Exam

Table of Contents

https://github.com/mkalw/UTS_31005_A3_12595512

1. Introduction

Twitter is a social media and microblogging platform that launched in 2006 and has become a popular service used by millions around the globe including well-known figures such as celebrities, company executives and politicians. The service provides insights into many topics of discussion such as the public views on current affairs, events and people. Users are limited to 280 characters in a ‘tweet’ which is a message that is posted and broadcasted to the users’ followers. Tweets can be ‘retweeted’ which refers to the posting of another users’ post to one’s own account and thus circulates the message to many other users of Twitter. The circulation of tweets through re-posting increases the exposure of the post and therefore if a high volume of users have retweeted the post, the message will spread as a meme and become viral. Virality refers to the rapid and widely circulated content featured on the internet, in particular, social media thus becoming well-known. Twitter marketing campaigns provide many incentives for advertising companies as it is quick, easy, has worldwide reach, and provides targeted advertising and flexibility. Twitter campaigns are also typically low budget compared to traditional promotional and advertising methods such as billboards and television advertisements thus enticing marketing and advertising companies to discover the criteria and techniques to achieve a viral campaign.

2. Complexities of the Problem

It is difficult to predict the virality of a Twitter message due to the unpredictable nature of a message going viral. There have been many cases were users with few followers have suddenly become viral. The time it takes for something to go viral can also vary greatly as a message many already exist but become viral a few months or years later due to current events.

3. Approach

3.1 Generalised Linear Model

The features chosen to be analysed through this model consist of the number of followers, tweet length, hashtags and mentions, sentiment, emotional divergence and individual sentiments. These featured were selected and then analysed through graphing them against the average number of retweets to see if a linear relationship existed. The dataset, “Tweets2011corpus”, used in this approach was from a reliable and reputable source, the National Institute of Standards and Technology (NIST) which gathered a sample of around 16 million tweets from the social media platform. The Weka toolkit is utilised in this study to both train and test the dataset.

4. Validation

The generalised linear model is a highly suitable methodology for this particular problem that is being predicted. The main features chosen for this model include number of followers, tweet length, hashtags and mentions, sentiment, emotional divergence and individual sentiments. These features all have a linear relationship with the average number of retweets, given outliers are removed from the graph, and therefore is it highly advantageous

to implement the generalised linear model. This prediction model trains by learning the weights for each of the previously mentioned features which correlate to the significance of each feature. The model has revealed that the most significant and impactful feature of a viral tweet is the number of followers of a user. The number of mentions and URLs are the next most influential features in predicting whether a Twitter message will go viral or not.

The study has ensured an accurate outcome to their study through looking at all aspects of a tweet, structure, content and sentiment, to train the model to provide a high-quality prediction. These model also do not simplify independence assumptions to provide reliable results.

5. Consequences

5.1 Ethical Consequences

Studies that focus on creating viral tweets can exploit user data and breach privacy as user's data and online behaviour is being analysed. Companies have little regulation in terms of false advertising and misleading marketing online compared to promoting through traditional media. Moreover, it is much harder to control and police promotional content online than it is through traditional media as there is an overwhelming volume of content that needs to be regulated to ensure users are receiving accurate and reliable messages from marketers.

5.2 Social Consequences

The information gained through successful predictions can shape and influence the marketing and advertising campaigns produced by companies to specifically target people in the most successful and efficient way. Furthermore, targeted advertising may not provide the audience with a holistic view and perspective and may narrow their opinions and reduce social connectedness and inclusivity.

5. References

Jenders, M., Kasneci, G. & Naumann, F. 2013, 'Analyzing and predicting viral tweets', WWW '13 Companion Proceedings of the 22nd International Conference of the World Wide Web, pp. 657-664.

Hong, L., Dan, O. & Davison, B. 2011, 'Predicting Popular Messages in Twitter', WWW '11 Proceedings of the 20th International Conference Companion on World Wide Web, pp.57-58

Yang, J. & Counts, S. 2010, 'Predicting the Speed, Scale and Range of Information Diffusion in Twitter', Proceedings of the 4th International Conference Companion on Weblogs and Social Media, pp. 23-26