

31005_ML_Assignment_3

October 8, 2019

Haozhe.Tong SID: 13019841

1 Introduction

I will discuss the possibility of predicting whether a Twitter message could be spread as a meme or not in this paper. The term “meme” firstly emerged from “The Selfish Gene” written by Richard Dawkins in 1976, it was defined as “the unit of culture transmission” (Dawkins, 1976 cited in Díaz C and Mauricio C, 2013). With the advance of the internet, the concept of “meme” is extended to the internet layer. The internet meme refers to a concept or idea that spread from one person to another through the internet (Díaz C and Mauricio C, 2013), it could be an image, an email, a video or any other things popular among the net-users.

In this paper, I discuss the reason why it is hard to predict whether a Twitter message will be spread as meme or not or what kind of message will be spread. And I present two approaches that used to predict whether a tweet will go viral or not. In the last section, I try to discuss the ethical and social consequences if a pattern exists and it is implemented in society.

2 Reason

In this paper, I discuss the reason why it is hard to predict whether a Twitter message will be spread as meme or not or what kind of message will be spread. And I present two approaches that used to predict whether a tweet will go viral or not. In the last section, I try to discuss the ethical and social consequences if a pattern exists and it is implemented in society.

On the other hand, the popular internet meme has changed all the time. Such popularity of a current meme could be suddenly transformed and non-directive that people cannot discover a pattern. In 2019, for example, the best memes are “Ugandan Knuckles”, “Super Bowl Selfie Kid”, “Gym Kardashian” However, the best memes for 2018 are “Don’t say it”, “A Star is Born memes”, “Squinting woman” and so on. There are no similar memes according to the comparison of these two years. I believe that the biggest obstacle to solving this problem comes from humans themselves.

3 Prediction Approaches

There are two approaches can be used to predict whether a tweet will go viral or not with data analytics. However, these approaches are still conceptual which means they cannot yet predict the result accurately.

3.1 Naive Bayes model

The purpose of the Naive Bayes model is to calculate the possibility to classify samples based on the Bayesian theorem (Maximilian, Gjergji & Felix 2013). For this case, this model can calculate a joint probability of $NR(x) > T$ and the feature vector (x) for a tweet x following:

$$P(NR(x) > T, (x)) = P((x)|NR(x) > T)P(NR(x) > T)$$

Because there is no workable standard to judge the viral tweet, a predefined threshold T that refers to the criteria of determining whether a tweet is viral or not will be required. $NR(x)$ represents the number of retweets.

Naïve Bayes model is about the tweet feature, in other words, this model can be used to predict what kind of tweet could go viral. Therefore, it is required a random sample from a related dataset to train this prediction model. The samples can be generated from TREC 2011 microblog track and it can be separated into two groups, one for training and one for testing. A threshold T is a pre-determined condition to distinguish between non-viral and viral tweets, the different values of T could cause the different predicted results for the same sample. Hence, this model can be validated by comparison with the accuracy of the predicted result under different thresholds T .

3.2 Time-sensitive modeling

A Time-sensitive model is built basing on the Passive-Aggressive (PA) algorithm. PA algorithm maintains a linear decision boundary, which is properly classified within a certain range for each new instance while keeping the decision boundary as close as possible to the old one (Petrovic, Osborne & Lavrenko 2011). Time-sensitive models will build different models based on when the tweet was created and predict the result through the passive-aggressive algorithm.

The way to validate this model is similar to the way that validates the Naive Bayes model. The time-sensitive model needs to assume that there are specific rules about what is being retweeted at every time of a day (Petrovic, Osborne & Lavrenko 2011). Because this model will create different models for a different time, the samples at different times should be divided into some subsets of data and then undergo PA algorithm to predict whether the tweet in a specific time will go viral or not. The result will be compared with the entire dataset and calculate the accuracy.

4 Ethical and Social Consequence

A viral tweet could propagate false information which could negative effects. A viral tweet means that a tweet is shared by lots of people. If there is a template that enables a tweet to become viral discovered, all the kinds of the tweet could be viral and viewed by many net-users for better or worse. A repeated slander makes others believe. In this case, an advertising company can spread the fake information of their product by using the template, and then the product with fake information could be true after virally spread. It could harm society if there are many products with false information spread.

People could restrict themselves if a standard pattern has been constructed. The concept of virality includes a sense of competition. Each popular internet meme is created by people which could be unique and nonredundant. This situation could be changed when the pattern of viral tweet come up. All the tweets or memes generate basing on this patter might be similar and redundant not only on Twitter but other social media and the innovation ability of people could also be decreased.

5 Reference

- Díaz C, Mauricio C. 2013, 'Defining and characterizing the concept of Internet Meme'. CES Psicología, vol.6, no.2, pp.82-104.
- Maximilian,J., Gjergji, K. & Felix, N. 2013, 'Analyzing and predicting viral tweets', 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web, pp.657-664.
- Petrovic, S., Osborne, M. & Lavrenko, V. 2011, 'Rt to win! predicting message propagation in twitter', Fifth International AAAI Conference on Weblogs and Social Media.
- Santiago, A. 2019, 15 of 2018's most exceptional memes, Mashable. viewed 6 October 2019, <https://mashable.com/article/best-memes-of-2018/>.
- 20 of the Best Memes of 2019 (So Far!) 2019, Lifewire. viewed 6 October 2019, <https://www.lifewire.com/best-memes-4160652>.

5.1 Link to GitHub

https://github.com/HaozheTong/UTS_ML_2019_SID13019841_ASSIGNMENT.git