Do Surprising Tweets lead to Topics of Importance?

Dhananjay Nakrani, Krupa Pratap Brahmkstri, Varun Varadarajan
{dnakrani, brahmkst, vvaradar}@usc.edu

Abstract

Often it becomes difficult to track the latest trending topics when we follow some domain specific channels on Twitter or even our friends who keep posting tweets related to their own interest. The change is when we come across some surprising and unexpected tweets from these followers. According to Wikipedia, surprise is a brief mental and physiological state, a startle response, experienced by humans as the result of an unexpected event. Therefore, a world that is purely deterministic or predictable in real time for a given observer contains no surprises. Also the data may carry different amount of surprise for different observers or even the same observer taken at different time. Twitter is a social media which produces millions of tweets every day. While some give opinion, some give useful information which can be favorited or retweeted. In this project, we try to predict if tweets that are numerically classified as surprising, become topics of importance. The new and innovative thing about this project is that- in this project, we try to quantify surprise in text/tweet in a particular domain/hashtag. We also try to use surprise in a tweet at the time it was created to determine the overall impact it has on a topic by predicting if it will be important based on the attention it gathers which is directly proportional to the number of likes and number of retweets.

Introduction

Surprise is an integral part of our daily lives. Humans are surprised by a lot of things be it Tsunami waves, or a new Halloween costume. Ideally, surprise comes through our senses like what we see, smell or feel. Uncertainty is necessary for the surprise to exist. Another important point to note is that if one measures surprise, it is with respect to something relative and it depends from observer to observer. We can also be surprised from the things we read. Twitter has a lot of people tweeting and retweeting posts in large numbers. A user typically follows other people, celebrities and domain specific channels to keep a track of the latest news. A user comes across surprising and unexpected tweets from other users/hashtags which user follows. To cite an example, consider a channel tweeting live posts about a soccer match. Among all tweets, the ones which get a lot of attention is when there is a goal, a player being given yellow card or even a player being sent off. Such tweets differ from the other tweets as these get a lot of favorites to it, comments and of course, a lot of retweets. These tweets are nothing but unique tweets from a channel that bring about a feeling of surprise in the users. In this paper, we discuss how we can quantify the surprise expressed by a user based on the content of the tweet. The unit of surprise used is "wow." The goal of this paper is to prove that tweets which have a high "wow" factor for a particular user also captures the user's attention. We measure the attention capturing capabilities of a tweet by checking if a user favorited the tweet or by checking if the user retweeted it.

Related Work

Surprise has primarily been studied in the field of Neuroscience where the effect of surprise on different organs has been extensively studied. Recently, methods to incorporate surprise in recommender system are being researched. The correlation between surprise and attention has been investigated in [Itti, 2005]. The experiments conducted in this paper were on video data. To test the hypothesis, the eye movements of eight naive observers were recorded while the observers watched video clips. A topographic dynamic response map was created to detect the locations in the frames of the videos that the observers gazed on. Evaluation was carried out based on KL Divergence. KL scores for 6 metrics showed significantly different performance levels. KL scores were computed by comparing the number of human saccades landing onto each given range of master map values to the number of random saccades hitting the same range. In our research, we try to prove the hypothesis by checking if a user's attention was captured by a piece of text, by checking if the user favorites or tweets the text.

Dataset

We use the tweets from Fifa World Cup 2014(#FIFA2014) to determine surprise elicited by a tweet for users showing interest in the said hashtag. By using a single domain/hashtag/topic for a dataset, we are trying to eliminate variations that would incorrectly indicate surprise, though they are outliers, because of the various topics present in the tweets the user is interested in. The dataset consists of 19,558,283 tweets and nearly 11,734,800 retweets among them tweeted by over 5 million users.
The Twitter API documentation explains in details about the attributes corresponding to the tweets.


```
{
    "coord_lon": 0,
    "is_HON": false, "is_rt": false,
    "is_NGA": false, "is_ITA": false,
    "text": "If #JamesRodriguez scores for #COL we'll give away a pair of adidas #F50. Follow &amp; RT to enter! #allin or nothing. http://t.co/vilgpqQL7B",
    "is_ECU": false, "is_GER": false,
    "is_CHI": false, "is_POR": false,
    "is_RUS": false,
    "id": 485143250505068545,
    "uid": 16099375,
    "is_IRN": false, "is_JPN": false,
    "is_SUI": false, "is_WorldCup": false,
    "rt_count": 6624,
    "is_GRE": false, "is_geo": false,
    "coord_lat": 0,
    "is_CRC": false, "is_CRO": false,
    "fav_count": 726,
    "is_BRA": false, "is_FRA": false,
    "is_BEL": false, "is_ARG": false,
    "is_ALG": false, "is_MEX": false,
    "lang": "en",
    "is_URU": false,
    "created": 1404502139,
    "is_BIH": false, "is_KOR": false,
    "is_COL": true, "is_AUS": false,
    "is_CIV": false, "is_USA": false,
    "is_CMR": false, "is_ENG": false,
    "is_GHA": false, "is_NED": false,
    "is_ESP": false
}
```


For this project, we consider the following tweet attributes: favorite_count, retweet_count, text, user - who posted the tweet and retweeted_status to determine the surprise content of the tweet.

Algorithm

As mentioned in [Itti, 2005], surprise can exist only in the presence of uncertainty which can arise from intrinsic stochasticity or missing information. Also, surprise is related to the expectations of a particular observer i.e a twitter user in our case. What one user finds to be surprising, may not be surprising to a different user. We hence define a surprise threshold which varies for different users. Any tweet whose calculated surprise value if greater than the threshold is most likely to capture a user's attention. This attention a user gives to particular tweet is measured by checking if the user "retweets" or "favorites" that particular tweet.

A user's expectations is determined by calculating his/her prior probability distribution {P(M)} using the existing set of re-tweets already made by the user. Given this prior distribution of beliefs, the effect of the new tweet on the user is to change the prior distribution {P(M)} into posterior distribution {P(M|D)} via Bayes theorem,

$$\forall M \in M, P(M|D) = \frac{P(D|M) * P(M)}{P(D)}$$

Surprise for a given tweet and a given user is then measured as some distance between the prior and the posterior distribution. This distance is best calculated using relative entropy which is also known as information gain or Kullback Leibler(K L) divergence. Thus surprise is mathematically defined as:

$$S(D, M) = KL(P(M|D), P(M))$$
$$= \int_M P(M|D) log(\frac{P(M|D)}{P(M)}) \mathrm{d}m$$

A unit of surprise —a "wow" —may then be defined for a single model tweet as the amount of surprise corresponding to a two-fold variation between P (M |D) and P(M), i.e., as logP(M|D)/P(M) (with log taken in base 2), with the total number of wows experienced for all tweets that a user has retweeted or favorited, obtained through the integration in the above equation.
The process used to calculate surprise elicited by a tweet for a user is calculated as follows:

Step #1: For a user who has shown interest in a particular hashtag/domain, get all the retweets made by the user about that particular topic along with the number of times he has clicked to favourite certain posts in that domain.

Step #2: Construct feature vectors using various features such unigrams, bigrams and hashtags associated with the tweet.

Step #3: For a new tweet with the same hashtag as being, create a feature vector similar to the one mentioned in Step 2. Use equation 1 to calculate information gain for each of the constructed feature-sets. The total surprise as measured in "wows" is calculated as the linear weighted sum of the information gains calculated for the different feature-set.

$$S = \sum_{i=1}^{n} \lambda_i s_i$$

such that, $0 \leq \lambda_i \leq 1$ and $\sum(i * \lambda_i) = 1$

Step #4: If the information gain calculated in Step 3 is greater than a predefined threshold value for the user, the user will most likely be surprised by the new tweet.

Evaluation

For evaluation of our algorithm, we take 20 users who have a large number of retweets in a specific hashtag. For each user we split the retweets made by him/her and use 75% of the tweets to calculate prior probabilities for the user. We will try to obtain the surprise threshold for the user in the particular topic using this 75%. We will then use the above algorithm to calculate the surprise for each tweet in the remaining 25% of the tweets made by the user. For each user, we calculate the accuracy of surprise prediction by taking the ratio of the number of retweets in the 25% whose surprise value is beyond the threshold to the number of retweets in the 25% data. The accuracy calculated is the same as the recall value for surprise calculation as all the tweets that are classified as surprising are true positives and none of the tweets in the 25% are false negatives. As a result, the precision is always 1. Hence, recall is the only measure we consider while evaluating our models

We considered two feature-sets for proving the hypothesis. They are unigrams and bigrams. We discovered that unigrams are better indicators of surprise when compared to bigrams. This difference exists because unigrams are sensitive missing information or new information. Bigrams however indicate large differences in missing or new information in the new and observed data. We used weights 0.65, 0.7 and 0.75 for unigrams and 0.35, 0.3 and 0.25 for bigrams respectively. We discovered that a weight of 0.7 for unigrams and a weight of 0.3 for bigrams best represent surprise and also does not overfit the data.

We also tried calculating surprise using another distance metric called the Jensen-Shannon divergence. We found that KL divergence performs the

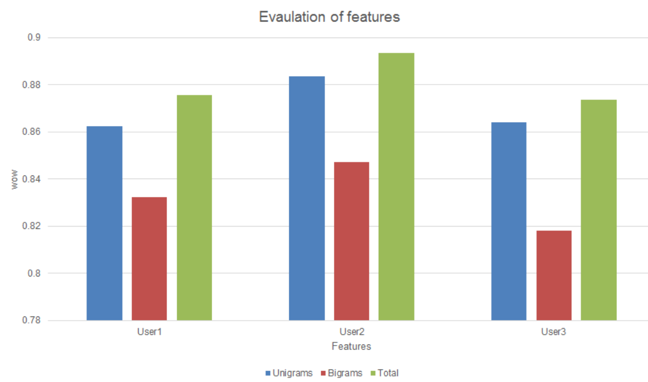

Figure 1: "wow" factor for 3 users.

same as Jensen-Shannon Divergence in most cases. Also, the value of Jensen-Shannon is bounded by 1 whereas KL divergence can sometimes be positive infinity. For our experiments, we have used KL divergence.

Following is the result of the accuracy of the top 3 users:

Features-User	**User-1**	**User-2**	**User-3**
Unigram	0.8624	0.8836	0.8641
Bigram	0.8325	0.8471	0.8182
Total	0.8756	0.8934	0.8738

Example: Let us try to understand this with the help of an example: Our initial dataset has the following tweets:

T1= "I am just a little bit concerned that Rooney appears to have put on a bit of weight #ENG"
T2= "Thank God! An unfounded weight has been lifted from Rooney's shoulders #ENG"
T3= "Still time to snatch defeat from jaws of victory. Sorry. draw #overweight #overpaid #ENG" For this

dataset, we construct a Unigram frequency distribution table like follows:

Unigram	**Frequency Distribution**
little	$\frac{1}{23}$
bit	$\frac{1}{23}$
.	.
.	.
#ENG	$\frac{3}{23}$

Now, when a new tweet comes up, let us see how it is related to the initial dataset and if so, what is the information gain?

Tweet= "They have already defeated two of the heavy weights. It is possible #CRC"

Computing its initial frequency distribution:

Unigram	**Frequency Distribution**
already	$\frac{1}{7}$
defeat	$\frac{1}{7}$
.	.
.	.
#CRC	$\frac{1}{7}$

Now let us see whether this tweet has any relation or anything common with respect to the initial dataset:

Unigram	**Frequency Distribution**
little	$\frac{1}{29}$
bit	$\frac{1}{29}$
.	.
.	.
#CRC	$\frac{1}{29}$

This is basically done to normalise the probability distribution. This gives a KL divergence of 0.83.

Conclusion

In this paper, we described our work on calculating surprise in text documents. We used tweets and determined that surprising tweets are also the tweets that capture user's attention. The direct consequence of proving the above hypothesis is that if a large audience finds a given tweet surprising then the topic of the tweet will be trending. We used various feature sets to calculate surprise. We found unigrams are good indicators of surprise. We also discovered that bigrams also indicate surprise but they do not indicate subtle differences in information gain.

In our future research, we want to consider other features such as time a tweet was created. We also want to create an graph of events/topics and use the structure of the graph as a feature to measure surprise. We also want to experiment with various distance metrics other than KL divergence and Jensen-Shannon distance.

The applications of measuring surprise in text is many. For example, we can manipulate the content shown to the user on a website based on how surprising it is and thus attract more users to the website. Our research can also find applications in various other fields such as Marketing and Education as attracting human attention is very important.

Reference

[Itti, 2005] L. Itti, P. Baldi. Bayesian Surprise Attracts Human Attention. In Press, Proc. Neural Information Processing Systems (NIPS 2005) [Kaminskas, 2014]M Kaminskas, D. Bridge. Measuring Surprise in Recommender Systems. Workshop on 'Recommender Systems Evaluation: Dimensions and Design' (REDD 2014) [Kullback, 1959] Kullback S. (1959), Information Theory and Statistics. 1959