International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

Vol. 44 No. 2 May 2024, pp. 183-188

An Application Of Rumor Propagation In Google Trends

Mr. K.L.M. Silva and Prof. R.P.K.C.M. Ranasinghe

Department of Mathematics

University of Sri Jayewardenepura

Nugegoda, Sri Lanka

Abstract—Daley and Kendle's pioneering mathematical basis for modelling rumor propagation around 1965 has paved a strong platform for researchers to come up with new generalizations for practical scenarios of rumor spreading in social networks. In this study, we use a generalization of original DK model, Ignorant-Spreader-Stifler (ISS) model to analyze the cricket related searches in Sri Lanka on Google.

The secondary data used for the study were obtained from Google trends for a five-year period from 2012 to 2017, targeting the three t20 matches held during the period. We introduce two algorithms on Maple mathematical software to estimate the parameters of the ISS model numerically.

We observe that the way people search on web about cricket is roughly analogous to spreading a rumour and forecasting is better when modelling with most recent data than considering more data back in time and we also find that the predicting accuracy is significantly equal of the two models built considering the data in 2012 and 2014.

Keywords—Rumour propagation, ISS model, Google trends, Cricket

I. INTRODUCTION

With the emergence of the social networks, rumors disseminate easily now more than ever. It is well known that rumor propagation has a significant influence on human lives and it can shape the public opinion of a society or a market. Transmission of rumors may have negative sides such as causing panic in some emergency events and could even destroy the credibility of someone.

Long before 'rumor propagation models' was a mainstream discussion, mathematicians knew that there was a similarity between epidemiology and the spreading of an information [1]. Based on that fact, they tried to configure a model for the dissemination of information and in 1965, Daley and kindle were able to put the mathematical basis for it [2]. After the very first classical rumor propagation model proposed by Daley and Kendle around 1965, various researches have been conducted on developing a model for rumor propagation based on situations in the real community.

Google trends is a powerful tool to measure the popularity of trending searches and it is important to examine what people are interested in and what they are mostly curious about for achieving different business and marketing goals. In this work, we attempt to predict the popularity of search term, cricket in Sri Lanka in the vicinity of a T20 world cup match. The ISS model has been used to verify whether cricket related searches on Google behave like a rumor in Sri Lanka and two algorithms were proposed to estimate model parameters in a numerical manner.

II. Theoretical Background

The ISS rumor propagation model is a set of ordinary differential equations described as follows.

$$\frac{dI}{dt} = -k\beta IS \tag{1}$$

$$\frac{dS}{dt} = k\beta IS - kS\alpha(S+R) \tag{2}$$

$$\frac{dR}{dt} = kS\alpha(S+R) \tag{3}$$

In this model, the total population is divided into three subgroups: Ignorants(S), Spreaders(S) and Stiflers(R). Ignorants are the ones who are not aware of the rumor and spreaders are the people who actively spread the rumor while the stiflers become the ones who are not interested in spreading the rumor even though they know it. An ignorant is turned into a new spreader when there is a contact between an ignorant and a spreader and the decaying process of spreading is only caused by the contact of spreader-spreader or spreader-stifler [2].

This model has three parameters.

α :- The probability of decaying of the process of spreading (The stifling rate).

β :- The probability that an ignorant progresses into a spreader (The spreading rate).

k :- The average number of contacts.

Here α and β are the two key parameters of this model and the solution for the spreaders out of this system of ordinary differential equations will depend mainly on these parameters. This model has no analytical solutions; therefore, a numerical approach was used to obtain the set of solutions in modeling the real data and Maple mathematical software was used in order to meet this objective [2].

III. Methodology

3.1. Data Collection

The secondary data used for the analysis, cricket related searches on Google in Sri Lanka were extracted from Google trends for 256 weeks from 11.03.2012 to 26.02.2017.

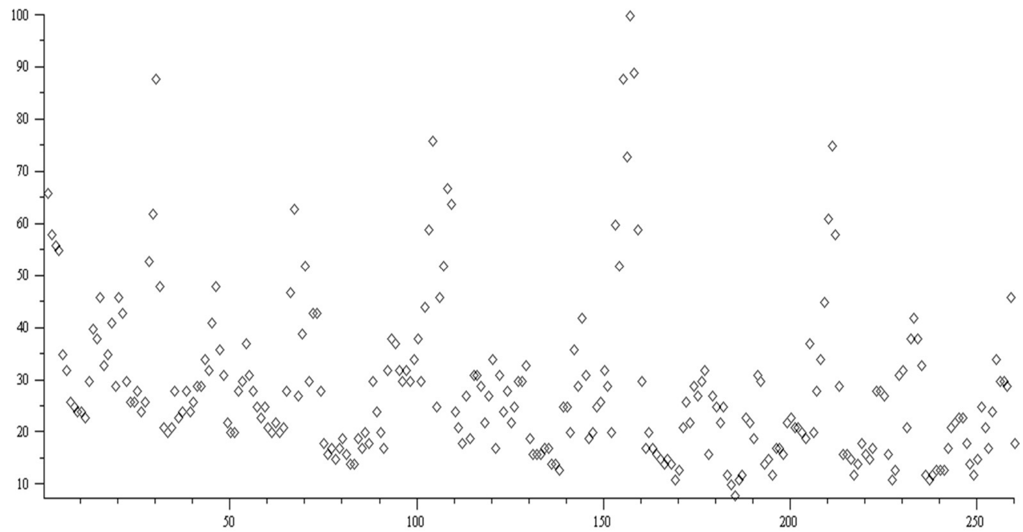

Fig .1. The scatter plot diagram for the real data obtained from Google trends under the search term "cricket in Sri Lanka"

Google trends helps to find the search interest of a particular topic typed on Google browser over a chosen period in a selected region. First, it takes the search volume of a topic and then that search term is divided by the total number of searches. Then, this ratio is scaled on a range of 0-100 and that is represented by the y axis of the chart in Fig 1. Along the x axis, the time is represented broken by dates. The peak points of the chart always occur near the world cup dates.

3.2. Numerical Approach Through Algorithms

The parameters of the model were estimated through minimizing the sum of absolute errors (SAE) at each time point for the available data. The following two algorithms were developed in order to obtain the optimum estimates for the model parameters.

1. Random finder Algorithm (RF) -

   Inputs: the real data set, the number of combinations of parameter values expected to consider, the initial spreader population

   Output: the estimates of the model parameters with the sum of absolute errors.

   This algorithm returns the set of values for the parameters which give the least sum of absolute errors by setting the values for the parameters ad hocly at each iteration.

2. Fine tuner algorithm (FT) -

   Inputs: lower and upper bounds to the parameter values obtained from the Random Finder Algorithm, step sizes, real data set, initial spreader population

   Output: the optimized estimates for the model parameters with the sum of absolute errors.

   This algorithm is an improved version of the random finder, constructed to tackle the optimized model parameters in a more systematic way. Here, at each iteration, an alpha value is fixed and then all the beta values within the specified step size are concerned within the given boundaries while the absolute error summation is recorded. The same process is repeated by moving to the next consecutive alpha value. Then the set of values for the parameters are returned with the lowest sum of absolute errors. Therefore, the accuracy is much higher compared to the previous algorithm.

Firstly, different weights of data from 2012 and 2014 T20 matches were systematically concerned to train the model and the parameters were estimated numerically through the proposed algorithms. Then, the model was tested for all combinations of weighted data by forecasting the t20 world cup in 2016. The results can be seen in the next section.

IV. Results And Analysis

4.1. Summary results of the two algorithms

Table 1. The estimated model parameters

Weights		Random finder			Fine tuner			MAE of
2012	2014	α	β	k	α	β	k	2016 T20
1	0	0.0677	0.3689	3.0458	0.06	0.3923	3.426	0.0933
0.9	0.1	0.0841	0.4631	3.02	0.07	0.442	2.982	0.0924
0.8	0.2	0.0677	0.3689	3.0458	0.0705	0.4305	3.009	0.0915
0.7	0.3	0.0696	0.4477	2.9412	0.08	0.4702	2.705	0.0904
0.6	0.4	0.14	0.791	1.577	0.14	0.791	1.577	0.0893
0.5	0.5	0.1564	0.8381	1.4871	0.1399	0.7614	1.6088	0.0882
0.4	0.6	0.14	0.74	1.62	0.14	0.74	1.62	0.0877
0.3	0.7	0.14	0.74	1.579	0.14	0.74	1.579	0.0889
0.2	0.8	0.1564	0.8381	1.4871	0.1	0.51	2.24	0.0882
0.1	0.9	0.104	0.4885	2.2322	0.14	0.69	1.62	0.0875
0	1	0.1189	0.4759	2.1013	0.17	0.81	1.351	0.0868

Predictive accuracy of the model was measured through the mean absolute error (MAE) of the forecasted data for t20 world cup in 2016. Table 1 represents the estimated values of the model parameters and the corresponding weights of the trained data and the MAEs of the tested data. The highest MAE value is recorded as 0.093 when only 2012 train data is considered and the lowest MAE is 0.0868 when the complete weight is given to 2014 train data. It's clear that MAEs are decreasing downwards in the table 1, when more weights are given to 2014 train data.

4.2. Testing Equality of MAEs

In this section, we test for the equality of predictive accuracy of the two forecasting models based on the 2012 (MAE=0.0933) and 2014 (MAE=0.0868) framework. We conduct a hypothesis testing on MAEs in a bootstrap approach as in the following way [3].

$$H_0\colon\ MAE_{2012} - MAE_{2014} = 0 \quad vs\ H_1\colon\ MAE_{2012} - MAE_{2014} \neq 0$$

For the test, 5000 bootstrapping replicates are used and the mean (t) is taken to be the bootstrapping statistic.

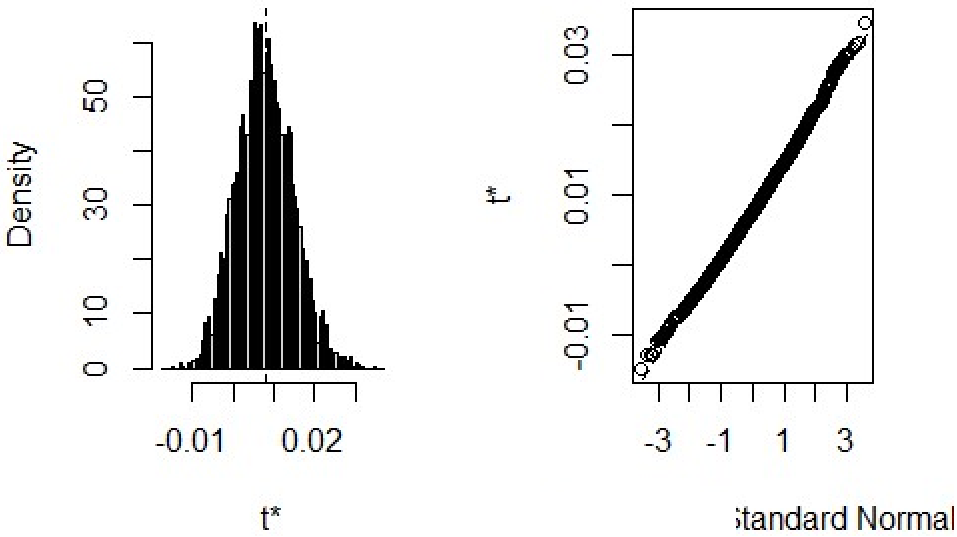

Fig 2. The bootstrap sampling distribution of mean

the sampling distribution of the mean difference of two forecasts seem to be roughly Normal (Fig 2) and the observed mean value, applied to the original data is 0.007824 (Fig 3).

```
ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = df, statistic = dfmean, R = 5000)


Bootstrap Statistics :
      original       bias    std. error
t1* 0.007823841 2.798549e-05 0.006893393
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 5000 bootstrap replicates

CALL :
boot.ci(boot.out = bootstrap, type = c("norm", "basic", "perc",
    "bca"))

Intervals :
Level      Normal              Basic
95%   (-0.0057,  0.0213 )   (-0.0065,  0.0205 )

Level     Percentile            BCa
95%   (-0.0049,  0.0221 )   (-0.0043,  0.0228 )
Calculations and Intervals on Original Scale
```

Fig 3. The R output of the bootstrap object

Since 0 is inside the all types of CI (Normal: (-0.0057, 0.0213) / Basic: (-0.0065, 0.0205) / Percentile: (-0.0049, 0.0221) / BCa: (-0.0043, 0.0228)), we do not reject the null hypothesis and conclude that there is no significant difference between the predicting ability of the two forecasting models.

V. Discussion

Random Finder algorithm (RF) was mainly designed for identifying a possible range for the estimates of the model parameters through an ad hoc approach while the Fine Tune algorithm (FT) can generate more accurate results within the range established by the RF through a systematic approach. The accuracy of the FT can be increased by entering a lower step size but it also increases the evaluation time causing Maple to crash. And since these model parameters lie in the real number set (R) which is a complete space, finding the actual estimates using this procedure is impossible due to the comparatively low processing speed and cache memory of a general use computer.

The study was limited to 3 world cup data and to use a more general form of rumor propagation modeling (ISS model). Customizing the characteristics of the ISS model to accommodate cricket related searches on Google can be done as future work to increase the accuracy in prediction.

VI. Conclusion

There is a similar pattern between the trend that people search for the game of cricket on Google in the vicinity of a cricket world cup and the spreading of a rumor with respect to ISS rumor propagation model. Searching results on Google had been sufficiently spread as a rumor near a cricket world cup since β values had surpassed α values. Hence, the number of searches on cricket in Sri Lanka near a world cup can be modeled with a low forecasting error by estimating the model parameters of the ISS model through a numerical approach. Further, the forecasting was better when modeling with most recent data than considering more data back in time but we prove that there is no significant difference statistically between the predicting accuracy of the two models 2012 and 2014. The ISS model predicts well the cricket related trends for a short time lapse but not good for predicting long term trends.

References

[1]Goffman, W & Newill, VA 1964, ‘Generalization of epidemic theory: an application to the transmission of ideas’, *Nature*, vol. 204, no. 4955, pp. 225–228.

[2]Daley, DJ & Kendall, DG ‘Epidemics and rumours’, *Nature*, vol. 204, no. 4963, pp. 1118-1964.

[3]Boiroju, Naveen & Yerukala, Ramu & Rao, Manneni & Malkareddy, Krishna. (2011). A bootstrap test for equality of mean absolute errors. 6.