



EMPIRICAL ANALYSIS OF SOCIAL NETWORKS

Predictive Analysis of Tweets – Visualization

Sujay Vittal¹, R Lohit², C O Prakash³

Department of Information Science and Engineering , PES Institute of Technology, Bangalore,

Abstract

Today the World Wide Web is undergoing a subtle but profound shift to Web 2.0, to become more of a social web. The use of collaborative technologies such as blogs and social networking site (SNS) leads to instant online community in which people communicate rapidly and conveniently with each other. We present an empirical study of different tweets obtained from twitter. Our analysis reveals interesting inferences and provides a good visual representation of the same.

Keywords: Empirical Analysis; Twitter; Tweets; Tweepy; Predictive Analysis; Python

I. INTRODUCTION

Empirical Analysis attempts to describe accurately the interaction between the instrument (or the human senses) and the entity being observed. In this scenario, it describes relationship between human behavior and issues that may arise due to any imbalance. [2]

The major roadmap of our study lay on concentrating on achieving the following steps:

Observation: This is the phase that deals with collection and organization of empirical facts; forming hypothesis; Induction: This is the phase which primarily deals with formulating the hypothesis; Deduction: Deducing consequences of hypothesis as testable predictions; Testing: Testing the hypothesis with new empirical material that was obtained in the observation phase; Evaluation: Evaluating the outcome of testing.

We analyzed the tweets of different people/groups of people during the time of crisis. Also, we considered tweets based on a

specific string search. The primary focus was on constructing study patters across large set of data streams.

II. IMPLEMENTATION

A. Empirical Analysis – what?

It attempts to describe accurately the interaction between the instrument (the human senses) and the entity being observed. In this scenario, it describes relationship between human behavior and issues that may arise due to any imbalance. Analysis can be quantitative or qualitative. Learning can happen through direct, indirect observation or through experience. [1]

B. Empirical Analysis – How?

There are three major scenarios that have been taken into consideration. These three scenarios are as mentioned – Can issues be predicted on the basis of Empiricism? How effective are these predictions? And can predictive models be held accountable for predicting imperfections/problems?

As described, the question of how Empirical Analysis is done is answered by working on the lines of the following measures – Observation; Induction; Deduction; Testing; Evaluation.

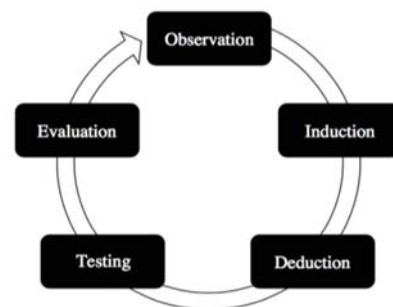


Figure A: Empirical analysis – Phases [3]

III. COMMUNITY DATA AND ASSOCIATED EVENTS

The Goal of this paper is to infer hypothesis from the patters - relate different Tweets with the same inferred pattern (testing). We deduced relationships between patterns and tweets. Say, for problem/situation X, there are Y types of typical tweets. The model would be able to predict the type of problems a particular pattern may indicate and this is the primary goal of this research.

Visualizations shown in the conclusion section give a better picture of the same scenario. The following was the flow diagram that mainly highlights the work that has been done.

Using Tweepy, TextBlob predictive analysis on twitter data was done to build a hypothesis. Tweets were used as data and also were used as benchmarks to compare other tweets.

We are trying to capture the tweets, build the test data and use the same to analyze more tweets. Relevant tests were done on the hypothesis to obtain a strong conclusion.

IV. OBSERVATIONS

The major observations that were made as part of this paper was that there exists a strong correlation between the tweets that originate from certain groups and their cumulative reactions towards a certain happening.

The data was taken over a wide spectrum of topics/events. Every topic had a wide assortment of data associated with it. Analysis was done over a very small duration of time. However small the duration was, since it was live twitter data, we had about five thousand entries within a very short duration of time. Figure A shows the work flow that has been followed.

Sentiments were calculated based on the weights of tweets using TextBlob and sentiment analyzer in python. The same was visually represented using histograms to give a better representation of the whole scenario in figure C and figure D.

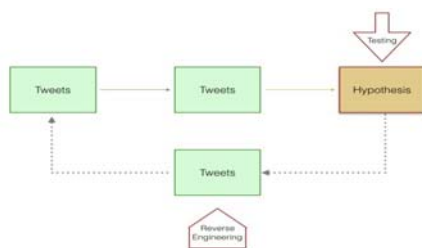


Figure B: Workflow Diagram

V. CHALLENGES

The analysis of social media comes with a set of extremely challenging situations which demand specific actions. The major challenges we faced as part of this paper was the mining of data relevant to our study.

Twitter produces data in extremely large quantities. The major challenges were finding answers to questions such as – How effective are the predictions that are being made? Can issues be predicted on the basis of Empiricism? And can predictive models be held accountable for predicting imperfections/problems?

VI. RESULT AND ANALYSIS

The question about effectiveness was answered quite certainly by our solution which has strong roots for the common thinking abilities of humans. For example considering figure C, when we performed a study with “Oscars2015” as the search string, our predictive analyzer predicted that it made more sense to hold the Oscars in the Americas than any other region. Similar such patterns can be studied. The following shows the graphical representation of the above claimed observation.

The next analysis that was done was based on the Tweets that highlighted Prime Minister Narendra Modi. This is visually represented in figure D. Based on the hundreds of tweets that were generated, the following were the conclusions that were made.

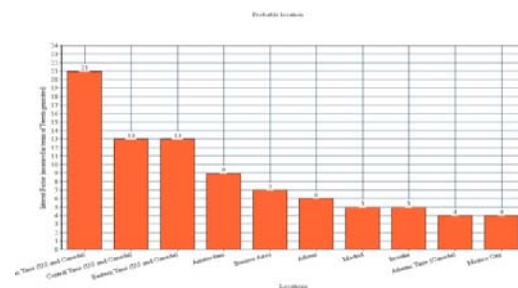


Figure C: Analysis done on “Oscars2015”

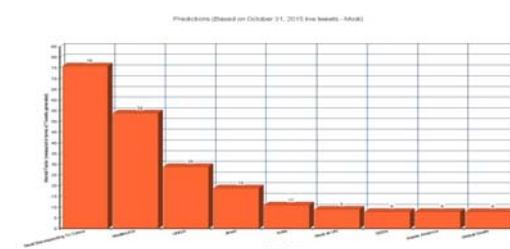


Figure D: Analysis on PMOIndia and PM Narendra Modi

VII. CONCLUSIONS

While doing some relevant literature analysis on Twitter and the people using it, we found an interesting co-relation between the tweets and the relevant activities associated with it. Considering the examples shown in Figure C and Figure D , figure C talks about analysis that was done on #Oscars2015. Looking at the results of the results of the graphical representation, we can conclude that Oscars happening in the Americas is a wise choice than being held in any other part of the world. Similar conclusions can be drawn for various other situations when the data considered is widespread and corelated.

REFERENCES

- [1] M. Schmitt, "Empirical Analysis of Social Networks", January 8, 2009.
- [2] Feng Fu, Lianghuan Liu, Long Wang, "Empirical analysis of online social networks in the age of Web 2.0", Elsevier, July 17, 2007
- [3] Heitink, G. (1999). Practical Theology: History, Theory, Action Domains: Manual for Practical Theology. Grand Rapids, MI: Wm. B. Eerdmans Publishing, p. 233. ISBN 9780802842947