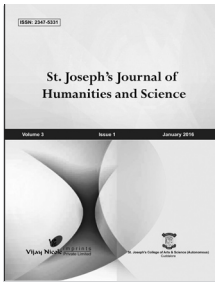
St. Joseph's Journal of Humanities and Science
ISSN: 2347 - 5331
http://sjctnc.edu.in/6107-2/



Popularity Analysis on Centrally Managed Accounts Formed in Social Media Using Big Data

T. Miranda Lakshmi[ab*]
A. Sophia Vilsana[b]
A. Martin[c]
V. Prasanna Venkatesan[d]

ABSTRACT

Social Networking has taken a great leap in analysing the popularity of different things including a product or a person. More importantly many high ranked products and professionals such as politicians, sports man, entertainers, scientists and other personalities are using social media for popularity. The photos, videos, audios, comments and likes about a person or a product is shared among the public for their popularity as well as for marketing. These shared data is analysed using data analytics tool to find the new strategies. Hence the Professionals looks into their opportunities of being noted and watched by public to increase their popularity. The popularity of a post or news is determined by various attributes like shares, likes, comments and recommendations to a friend. There are situations in which fake number of accounts are created in social media. These accounts are centrally managed accounts which creates widespread support and good opinion for a person or a product to increase the popularity. This research designs a model to conduct the popularity analysis and checks the originality of the popularity. Also validates the relationship between popularity and increase in shares for a product using page ranking algorithm.

Keywords: Big data, social media, comments, likes, opinions, popularity analysis, centrally managed accounts.

1. INTRODUCTION

The exact meaning of big data is defined by U.S Congress in August 2012, the big data is a high volume, high velocity, veracity, valuable and variety of data that cannot be handled by traditional techniques and technologies, so it requires advanced techniques and technologies to capture, store, distribute, manage and analyse the information [1]. Big Data analytics can be used to convert unusable data into meaning full one, because big data have large volume, velocity and different type of meaningless data. There are different techniques and methods that are used to convert vacuum data into great insights, they are

[a]Research Scholar, Research and Development Centre, Bharathiyar University, Coimbatore, India.
[b]PG & Research Department of Computer Science, St. Joseph's College of Arts & Science (Autonomous).
[c]Department of Computer Science, Central University of Tamil Nadu, Tamil Nadu.
[d]Professor, Department of Banking Technology, Pondicherry University, Pondicherry, India.
[*]E-mail: cudmiranda@gmail.com, Mobile No.: +91-9600970087.

- Text Analytics
- Audio Analytics
- Video Analytics
- Social media Analytics
- Predictive Analytics

Text Analytics:

Text analytics can be used to collect the exact content of text data for finding the meaning and the patterns of the word. The meaning and pattern of the word is to determine the people feeling about something.

Audio Analytics:

Audio Analytics can be used to collect unstructured audio data to extract these data into useful one. It is used in call centers to identify the product and service issues. Audio analytics is also called as speech analytics.

Video Analytics:

Video analytics can be used to monitor, analyse and extract the data to meaningful information from video streams. It is also called as Video Content Analysis (VCA).

Social Media Analytics:

Social media analytics can be used to collect data from online social networking sites. It provide dynamic network, identifies the dominant objects and find out the interesting patterns of online user's activities.

Predictive Analytics:

Predictive analytics can be used to collect past event information to predict future events using different techniques. Neural networks is a one of the most important technique used in predictive analytics.

1.1 Social Media

Social media is a collection of online interactive techniques and applications; they are social networking sites, blogs, microblogs, social news, wiki, social bookmarking, media sharing and online review sites. Facebook and twitter is a most popular social networking sites among online marketers and online users [1].Social networking users of the world have to develop the life casters, they are happy when sharing their own personal and private trash of their life. Online users share and upload their photographs, beloved dairy flashes, movies, details of personal accomplishments, special rash comments, pleasurable ignorant, likes, dislikes, fears, prejudices, hopes and dreams.

Social media analytics can be used to analyse data from social media and big data analytics can be used to analyse large amount of data in different types (structured, unstructured and semi-structured) from various fields. So social media big data can extract data into meaningful insights for decision making.

This paper describes social media big data analytics to finds how the person is popular on social networking sites based on the post shared on social networking sites and also the way in which the popularity of the person is accessed using comments, shares and likes for the posts made over time.

This paper is organized as follows. First this paper explain the definition of big data and social media, next collection of literature reviews based on popularity, existing and proposed systems, algorithms, tools and techniques for popularity finding. Finally describes the conclusion and future work.

2. LITERATURE REVIEWS

Sharing and receiving the information through social media elements (social media websites like Facebook, Twitter, Research Gate, etc.,) to create interaction among the people is called as social media. Social media based popularity analysis describes the popularity of various subjects such as political, sports, governmental norms and organizations, entertainments and consumer or company products based on social media user's content, shares, likes, tweets and posts on different social media sites.

The number of comments and the number of likes of product brands can be collected from facebook that shows the position of brand post on top of the brand fan page enhances brand post popularity [2]messages, quizzes, information, and other material. The facebook and twitter data can be used for competitive analysis on three different companies to find best company [3].

A model to analyze the different tweets that are shared in the twitter social media website has been created. Even though it is a simple model it out performs market based predictions [4]. There are different ways to predict the popularity of the movie before it is released. One of the way is the activities that are associated with editor and movie watchers with respect to the movie. The different views that are collected from the Wikipedia (online open access repository for knowledge collection for different domains) about the movie can be used for the popularity of the movie [5]. Nowadays social media services like WhatsApp, Facebook, etc., provides an extraordinary services to the society to share the information about different activities that are performed by Government and different organizations. The information which are shared by social media elements reaches better than traditional web sites [6].

Local governments are used to increase transparency and e-participation, opening a real corporate dialog to identify which norms promote highly by online users[7]. The political homophily on Twitter using social network analysis is to classify users as Democrats or as Republicans based on the political content shared to predict maximum shares of political homophily among Democrats or Republicans. In general, Democrats exhibit higher levels of political homophily. But Republicans who follow official Republican accounts exhibit higher levels of homophily than Democrats [8]. The claim is validated based on the social news portal Digg using a developed model of social voting based on the Digg user interface [9]with some items receiving a disproportionate share of attention from users. Predicting which newly-submitted items will become popular is critically important for both companies that host social media sites and their users. Accurate and timely prediction would enable the companies to maximize revenue through differential pricing for access to content or ad placement. Prediction would also give consumers an important tool for filtering the ever-growing amount of content. Predicting popularity of content in social media, however, is challenging due to the complex interactions among content quality, how the social media site chooses to highlight content, and influence among users. While these factors make it difficult to predict popularity \\emph{a priori}, we show that stochastic models of user behavior on these sites allows predicting popularity based on early user reactions to new content. By incorporating aspects of the web site design, such models improve on predictions based on simply extrapolating from the early votes. They validate this claim on the social news portal Digg using a previously-developed model of social voting based on the Digg user interface.", "author" : [{ "dropping-particle" : "", "family" : "Lerman", "given" : "Kristina", "non-dropping-particle" : "", "parse-names" : false, "suffix" : "" }, { "dropping-particle" : "", "family" : "Hogg", "given" : "Tad", "non-dropping-particle" : "", "parse-names" : false, "suffix" : "" }], "id" : "ITEM-1", "issued" : { "date-parts" : [["2010"]] }, "title" : "Using a Model of Social Dynamics to Predict Popularity of News", "type" : "article-journal" }, "uris" : ["http://www.mendeley.com/documents/?uuid=09d70316-8d89-4c12-b7cb-e598377276ed"] }], "mendeley" : { "formatted Citation" : "[9]", "plain Text Formatted Citation" : "[9]", "previously Formatted Citation" : "[9]" }, "properties" : { }, "schema" : "https://github.com/citation-style-language/schema/raw/master/csl-citation.json" }. Popularity analysis can be used to predict the geographical area based on youtube video updated and video's views [10].

Using messages posted by a sample of restaurants on Facebook to measure the message popularity by the numer of "likes" voted by fans. And also to discover "more popular" and "less popular" social marketing messages [11]` Content popularity can be found by tracking the change in the number of links to pages, and the number of times these pages are visited. The popularity of articles can be predicted by using article cited in wikipedia and mendeley library. The popularity of emotional content in news feed can be found by facebook news posts [8].

3. EXISTING SYSTEM

The Research study comprised of Astroturf political campaigns on micro blogging platforms: politically-motivated individuals and organizations that use multiple centrally-controlled accounts to create the appearance of widespread support for a candidate or opinion. The Research also proposed a machine learning framework that combines topological, content-based and crowd sourced features of information diffusion networks on Twitter to detect the early stages of viral spreading of political misinformation.

There are different ways to create fake reviews and distribute through social media websites. For example during election campaign lot of fake accounts are created. These fake accounts are managed by centrally controlled authorities. From the fake accounts widespread support and good opinion are created for the election candidate. These false truth information may play key role to get votes. This will inturn decide the victory of the candidate during the election. Hence it is essential to design a model to detect misinformation which are created by centrally managed account during the early stages of an information sharing. The proposed model considers a machine learning algorithm to detect the creation of fake accounts and centrally managed fake accounts using topological, content-based and crowd sourced features of information diffusion networks.

Features Used:

- Hashtags: The Twitter community uses tokens introduced by a hashmark (#) to label the topical content of tweets.
- Mentions: A Twitter user can include another user's screen name in a post, prepended by the @ symbol.
- URLs: To extract URLs from tweets by matching strings of valid URL characters that begin with 'http://.'
- Phrases: To make meme from the various text shared in the tweet, metadata, punctuation and URLs are removed.

4. RESEARCH APPROACH

The research approach finds the popularity person on social networking sites based on the number of shares, likes etc. The stepwise procedure is shown here.
Step 1: Extraction of Popular Posts

Having fixed the time frame, the first step is to start extracting relevant popular social networking information from 3 data sets based on social networking news posts, facebook metrics and blogs.

Step 2: Data processing

Data can be analysed using R programming language.

Step 3: Application of Classifiers

Page Rank Link Analysis Algorithm is applied separately, and classification of the test data is carried out based on the existing training data that has been constructed beforehand.

Step 4: Finding percentage of Popularity

The overall percentage of shares, likes and comments about each team is calculated based on the results of Step 3, separately for the resultant data.

5. METHODOLOGY

The methodology describes the data collection, data analytical tool and algorithm for finding the popularity among online users.

5.1 Data Collection

There are three different data set such as facebook metrics data, popularity of news data and blogs data is collected from UCI repository. In total, over 39645 data records and 61 attributes are used, each and every posts in dataset belong to different categories, such as sports, spatial, entertainment, political.

5.2 Analytical Tool

R is an open source programming language and software environment for statistical calculating and graphics. It was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand.R and its libraries implement a wide variety of statistical and graphical techniques, include linear and nonlinear modeling, classical statistical tests, time-series analysis, classification and clustering. R can be used for data analyse.

Data analysis can be done using 8 parameters values, they are shares, average positive polarity, minimum positive polarity, maximum positive polarity, average negative polarity, minimum negative polarity, maximum negative polarity and the title of the posts.

It finds the relationship between these parameters for finding the popularity among others. The comparisons are made by using plots.

Coding for Data Analytics:

1. *library(readr)*
2. *colors<- c("Red","blue")*
3. *OnlineNewsPopularity<-read_csv("D:/iot/Thesis implementation/SOPHIA/OnlineNewsPopularity/OnlineNewsPopularity/OnlineNewsPopularity.csv")*
4. *View(OnlineNewsPopularity)*
5. *plot(OnlineNewsPopularity$self_reference_avg_sharess,OnlineNewsPopularity$shares,col = OnlineNewsPopularity$shares,pch = 16,cex = 2)*
6. *plot(OnlineNewsPopularity$avg_positive_polarity,OnlineNewsPopularity$avg_negative_polarity,col=OnlineNewsPopularity$shares,pch = 16,cex = 2)*
7. *barplot(OnlineNewsPopularity$avg_positive_polarity,OnlineNewsPopularity$avg_negative_polarity,col = "blue")*
8. *plot(OnlineNewsPopularity$avg_positive_polarity,OnlineNewsPopularity$min_negative_polarity,col="blue")*
9. *plot(OnlineNewsPopularity$avg_positive_polarity,OnlineNewsPopularity$max_negative_polarity,col=OnlineNewsPopularity$shares,pch = 16,cex = 2)*
10. *plot(OnlineNewsPopularity$min_positive_polarity,OnlineNewsPopularity$avg_negative_polarity,col=OnlineNewsPopularity$shares,pch = 16,cex = 2)*
11. *plot(OnlineNewsPopularity$min_positive_polarity,OnlineNewsPopularity$min_negative_polarity,col=OnlineNewsPopularity$shares,pch = 16,cex = 2)*
12. *plot(OnlineNewsPopularity$min_positive_polarity,OnlineNewsPopularity$max_negative_polarity,col=OnlineNewsPopularity$shares,pch = 16,cex = 2)*

The graph which is depicted in Figure 1 shows the relationship between average positive polarity and maximum negative polarity.

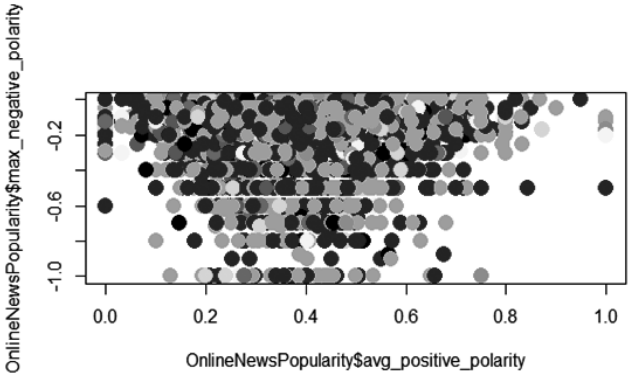


Fig.1 Relation between Average Positive Polarity and Maximum Negative Polarity

The graph which is depicted in Fig. 2 represents self reference minimum shares with respect to self reference maximum shares.

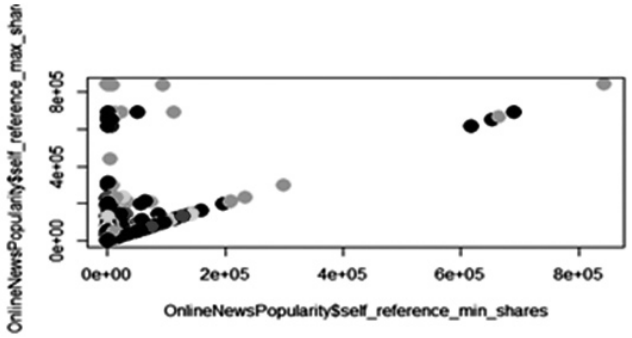


Fig. 2 Represents Relationship between Self Reference Minimum Shares with Respect to Self Reference Maximum Shares

The graph which is represented in Fig. 3 shows the relationship between self reference average shares and shares.

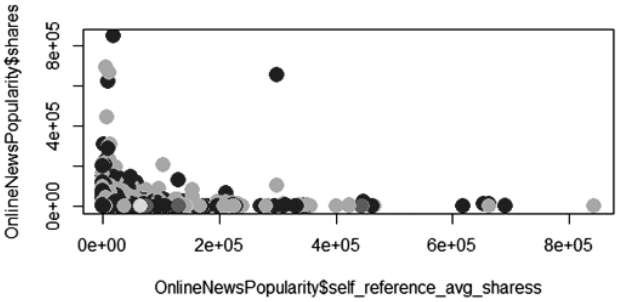


Fig. 3 The Graph Shows Self Reference Average Shares and Shares

The PageRank is one of the most well-known big data based unsupervised learning approach because of its effectiveness and simplicity. It is sufficient for scalability of increased data and to manage big datasets. PageRank is implemented using MatLab that will be involved in the prediction of popularity of Posts with public response and views.

6. CONCLUSION

Big data analytics on social media plays a significant role in the life and career of every individual. This paper designs a model to conduct the popularity analysis and checks the originality of the popularity. Also it validates the relationship between popularity and increase shares for a product using page ranking algorithm. It is useful for the corporate companies who spend time in investing money only on those people who are popular among the people in the society. This work can be further extended using expert system.

REFERENCES

[1] "holzener social media."

[2] L. De Vries, S. Gensler, and P. S. H. Leeflang, "Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing," *J. Interact. Mark.*, vol. 26, no. 2, pp. 83–91, 2012.

[3] W. He, S. Zha, and L. Li, "Social media competitive analysis and text mining: A case study in the pizza industry," *Int. J. Inf. Manage.*, vol. 33, pp. 464–472, 2013.

[4] P. Aula *et al.*, "Social-Customer-Relationship-Management (Social-CRM): Anwendung & Technologie," *Wirtschaftsinformatik*, vol. 263, no. 5, pp. 56–71, 2012.

[5] M. Mestyán, T. Yasseri, and J. Kertész, "Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data," *PLoS One*, vol. 8, no. 8, 2013.

[6] A. Edmondson, D. Garvin, and F. Gino, "Is Yours a Learning Organization?," *Harv. Bus. Rev.*, p. 10, 2008.

[7] F. Flores, "No Title," pp. 1–23.

[8] E. Colleoni, A. Rozza, and A. Arvidsson, "Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data," *J. Commun.*, vol. 64, no. 2, pp. 317–332, 2014.

[9] K. Lerman and T. Hogg, "Using a Model of Social Dynamics to Predict Popularity of News," 2010.

[10] A. Brodersen, S. Scellato, and M. Wattenhofer, "YouTube Around the World : Geographic Popularity of Videos," *Proc. 21st Int. Conf. World Wide Web*, pp. 241–250, 2012.

[11] B. Yu, M. Chen, and L. Kwok, "Toward predicting popularity of social marketing messages," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6589 LNCS, pp. 317–324, 2011.