

Predicting Demographic User Using Social Network Site

Dr Kamaljit Kaur ^{#1}, Ripanjit kaur ^{*2}

^{#1} Assistant Professor Department of Information Technology, ^{#2} PG Student
Guru Nanak Dev College of Engineering, Ludhiana, Punjab, India

Abstract - Publicity is any promotional communication regarding an organization and it is not a paid form of communication. The main purpose of publicity is to promote themselves. On a social network site such as Facebook[1], Twitter[2], Skype[3] some contents are attract to more visitor. It is very difficult for user to predict the publicity of person from large number of post and messages. In this paper we apply a FCM clustering technique on the database to predict the cluster of demographic user (male, female), after that applying the popularity index formula on high, medium low clustering percentage and show how many male like to person 1 and person 2.

Keywords — Popularity, FCM, P-index.

I. INTRODUCTION

Publicity is not a way of mass media communication to attracts the buyer to sell product. It is a way of gaining public attention by a social media like Twitter, Facebook, Instagram etc. That include the huge amount of post etc. Publicity makes a public relation. According to William J. Stanton, "Publicity is any promotional communication regarding an organization and/or its product where message is not paid for the organization benefit from it."

The purpose of publicity is to promote themselves. For example sale promotion is carried out for different purposes. It include promotion of new product, employee activities (brilliant or achievement of staff), new policies. It shows the highlighting of company profile and product its main goal to make a company image. Publicity is non paid form. It is just a arranging some function calling conference or other media to promote things. For example, to predict which product become more popular between a Garnier and Lakme company? We predict easily who become more popular from product sale. Another example, in politics, a one person carried out different seminar, conference or lecture at many area before election to promote ourselves than anti person. For promotion they make a video of ourselves and promote on social media like Facebook and Instagram. Unlike in tradition media such as newspaper, magazine, television also provide an information or knowledge to user. We used data mining techniques to predict the publicity between two users based on facebook data. Data mining is a procedure to evaluate the useful

relationship from large database. To resolve this problem, we apply an algorithm FCM (Fuzzy clustering means), FCM is the subpart of clustering technique.

In this research work, we predict the same kind of work but in a greater amount with the help like SNS such as Facebook. A question arises why I am prefer Facebook? Because there are no. of people uses Facebook than other site.

II. METHODOLOGY

In this paper, develop the system using MATLAB Software. It is the programming language implement by MathWorks and is simple and easy way to make learning fast and efficient. First, load the data in the Matlab. We used algorithm FCM (Fuzzy clustering means) FCM is the subpart of clustering technique. First we collected the data from facebook media. In pre-processing, extract the main feature from database. Apply clustering technique on different count and predict the cluster. Predict high, medium low clustering percentage using weighted formula.

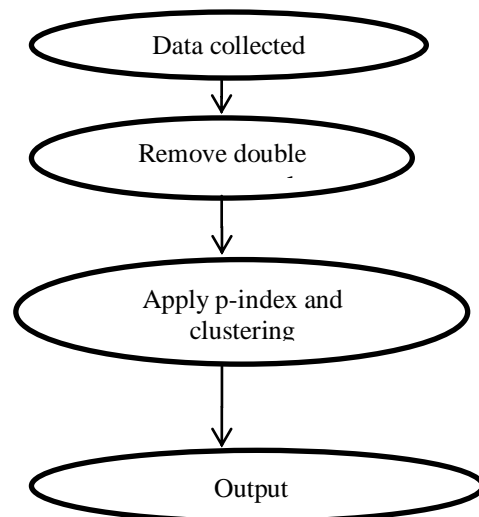


Figure: Flow diagram for various steps

A. Fuzzy Clustering means algorithm

FCM algorithm is developed by Dunn and improved by Bezdek. Fuzzy clustering Means is also called as soft clustering and also version of k-means. Aim of

k-means clustering is partition the cluster into minimize cluster variance or square error function.

B. P-index

Applying this formula on the database(Person 1 and Person 2) we predict the popularity index for high, medium and low.

1) **Pre - processing phase:** This phase include tokenization and parts of speech tagging (whether male or female like)

2) **Tokenization:** Tokenization is a process used to make token from the input count. These token are usually called as terms or words. Tokenization is used to predict the high, medium, low clustering percentage of male and female data.

3) **Remove punctuation:** In this step, punctuation with double quotes are removed from the tokenized sentence and converted into a digit. These words are removed from the input text to predict the male/female data correctly.

4) **Parts of speech tagging:** Parts of speech tagging assign the particular tags to each count in the input text. It is a process used to classify how many male or female comments for person 1 or person 2.

III. RESULT

To design a system, there are different steps which are implemented for predicting the political popularity among persons. In the Pre processing phase, data is collected from online source (social network site) such as Facebook. For predicting the demographic user (male and female) between two person using algorithm FCM (Fuzzy clustering means) FCM is the subpart of clustering technique. The system performed a task to predict a demographic user (male female). Number of count (such as haha_count , angry_count, like-count, wow-count, thankful-count, love-count) is taken as input text next apply the clustering technique. Then this text converted into high, medium and low clustering percentage. Using p-index to predict how many male and female like person 1. In last step we compare (high, medium, low) popularity index of person 1

IV. CONCLUSION

Publicity is any promotional communication regarding an organization and/or its product where message is not paid for the organization benefit from it. To resolve the problem of predicting publicity, we apply an algorithm FCM (Fuzzy clustering means), FCM is the subpart of clustering technique. In this paper work, we used to predict the cluster from matrix and show them in graphical form. Then it compares the high, medium and low popularity index to predict political popularity between two persons named as Person 1 and Person 2. Similarly, the popularity of any two personalities can be

recognized by providing the required data to the system.

REFERENCES

- [1] Youmna Borghal et. al, 2009 “Characterizing and Modeling Popularity of User-generated Videos”.
- [2] Meeyoung Cha, Haewoon Kwak et al, 2009, “Analyzing the Video Popularity Characteristics of Large-Scale User Generated Content Systems”, IEEE/ACM transactions on networking, Vol. 17, no. 5.
- [3] Jonah Berger et.al, 2010, “Positive Effects of Negative Publicity: When Negative Reviews Increase Sales”, Market science, Vol. 29, No. 5.
- [4] Vishal Gupta, Gurpreet Singh Lehal, 2010, “A Survey of Text Summarization Extractive Techniques”, Journal of emerging technologies in web intelligence, VOL. 2, NO. 3.
- [5] Aniket Rangrej et al, 2011, “Comparative Study of Clustering Techniques for Short Text Documents”.
- [6] Alexandru Tatar et.al, 2011, “Predicting the popularity of online articles based on user comments”.
- [7] Charu C. Aggarwal et. Al, 2012, “On the use of Side Information for Mining Text Data”, IEEE,
- [8] Kevin W. Boyack et al, 2013, “Improving the accuracy of co-citation clustering using full text” , journal of the american society for information science and technology.
- [9] Henrique Pinto et. al, 2013, “Using Early View Patterns to Predict the Popularity of YouTube Videos”.
- [10] Alexandru Tatar et.al, 2014, “A survey on predicting the popularity of web content”, Journal of Internet Services and Applications.
- [11] Yaser Keneshloo et.al, “Predicting the Popularity of News Articles”.
- [12] Mariam Adedoyin-Olowe, Mohamed Medhat Gaber et. Al, “A Survey of Data Mining Techniques for Social Media Analysis”.
- [13] Saïd Assar et.al, 2015, “Using text clustering to predict defect resolution time: a conceptual replication and an evaluation of prediction accuracy”, Springer.