

ANALYSIS ON COGNITIVE SOCIAL DATA FUSION

H.Anandakumar
Assistant Professor

T.Tamilselvan UG
Scholar

S.Nandni
UG Scholar

R.Subashree
UG Scholar

E.Vinodhini
UG Scholar

Department of Computer Science and Engineering,
Akshaya College of Engineering and Technology,
Coimbatore, Tamilnadu, India

Abstract– *Big data stands for effective handling of large amount of data, research, mining, intelligence. In social media large amount of data uploaded every. Social media handle large amount of data like photo, video, songs and so many using big data. When it comes for big data, a large amount of data should be effectively handled. Big data face various challenges like clustering of data, visualizing, data representation, data processing, pattern mining, tracking of data and analysing behaviour of users. In this paper the Emoji in messages are decoded and Unicode will be set. Based on the Emoji the user interest can be understood in a better way. Then another part involves the replacement of repeated data by using the map Reduce algorithm. Mapping of data with key values used to reduce the size of storage.*

Keywords – *Socialmining, Big Data, Mapreduces, Emoji Recognise.*

I. INTRODUCTION

Big data is data set stands for collection of data is emerging with lots of scope. Social mining is the process of extracting useful information from a large collection of data. The social mining will helps to identify a person and also helps in marketing, gain of customers, research, getting details about various products and for wide communication. A large amount of data is shared via social mediathat are stored in big data. In YouTube (type of social media) seventy two hour size video will be uploaded for every minute. Big data is well known for its 4v + value models. It says that with big data it is possible to handle large amount of data (volume) , various types of data that may be structured, unstructured or semi-structured (Varity), timely act of storing and retrieving of data (velocity), high flow of data (variability) and finally the value of data. The big data faces challenges in the field of data storage and analysing a person interest. Cognitive system enables continuous gathering of persons interest and develop the system which acts based on the data collected from the person.

In this paper the decoding of Emoji takes place. The Emoji are given high priority to under the user interested. Emoji are considering as important part of the post or tweets. Identify the Emoji in the tweet and decoding the Emoji. The second most challenges of big data is how the data are stored in the storage. The map Reduce is a type of algorithm used to

reduce the memory by avoiding the unwanted data or repeated data. The key values are arranged for every data stored in the big data. In case the word exists in the big data, the word can be replaced with the key value. This process reduces the overall memory of data base. By this the delay in processing of data can be reduced which increases the efficiency of data storage.

1.1 Methodolgy

Nowadays there is growth of social media and mining increases at larger rates. Those eventually increase the challenge, issues and also the risk factors in handling of data. When it comes for big data, the name itself says the storage should be large in size and also the optimization of the data in the storage. In this paper Unicode of Emoji takes places. It also involves the replacing of repeated data with the key values. It says what essential data are so that unwanted data can be removed. The major crises of analysing algorithms can be reduced. It result in better understanding of the user interested and clustering can be done based on the interests.

II. RELATED WORKS

The cognitive social mining gains the attraction of social media and big data. The cognitive mining helps in better understating of users in the social medial. The users interests are understand more accurate and based on their likes and behaviours their Suggestion varies. Large amounts of data are handled effectively, organized automatically, navigating of data. The data mining involves a large collection of data of different types it can be structured or unstructured. The fusion of data involves among the data which have common behaviour or similarly in patter mining [1]. The inputs cannot be stored directly into the memory it should be analysed before storing or processing. The data processing involves replacing the natural languages with the actual words. For example LOL can be stored as “Laugh Out loud” that can be effectively stored in the data storage [2]. The data that are posted in the twitter are artificially compressed since it allows only maximum of 140 characters at a time, So there must involves the natural data processing, to increase the data storage[1][3].

Everything around us creates an era of big data. To extract useful information from the big data it should be optimised in previous. In order to obtain the result as soon as possible the input data must be optimized. It also says the all new application growth should come under the umbrella of social mining, big data and pattern analysis [4]. The big data also involves demand based driven of information. The big data handles the reciprocal and bipartite social network. An agreed relationship between two or more persons through social network is social media [5].

2.1 Social Cognitive Theory

There is relationship between the person, environment and behaviours. The interaction between people should be a construction of self-efficient for the better understanding of users. Self-efficiency stands for the confidence to do a task [6]. The big data is a large scale system, it follows various algorithms. Retrieve, reuse, revise and relevant are the used in the big data for effective mining [7].

Personal knowledge management involves the connection and establishment of network for communication. It also involves the process of understanding other agents and then works in a coordination among them [8]. The human computer interaction is varying essential for understanding the user interest and dislikes. The embracing the social network involves for effective handling of data [9].

2.2 Relationship Model

It is a process of extraction of various types of information from large data storage. It is used for effective handling of wanted data. The big data follow various types of algorithm for handling data. K-means clustering [10], It involves grouping of data of k-clusters data sets. Every k cluster belongs to k centres. Thus the clustering is done to the nearest centric. Neural Network consists of nonlinear dependency functions. Sequential clustering involves grouping of data of similar data types[12]. No tool automatic identification of personal data, it will be done by repeated analysing of data. It can be used for the preparation, selection of data, scaling the size of data, collecting of large amount of data and then processing [13].

2.3 Clustering

Clustering stands for the grouping of data. The clique stands for the networks that have very close similarity with neighbour data and can be accessed quickly. N-cliques is a long stringy grouping of data. It helps for the better understanding. K-palorit's intype of algorithm for the better understanding of similar data[14].

2.4 Like Mining

Social media become a medium to share the likes and dislikes of users. It helps in better understanding of the users. It is possible to create a network of likeminded peoples [15]. It is possible to generate a solution for the non-linear data can be reduced[15]. Social decision support was developed as computational support to analyse the users [15][16].

III. PROPOSED SYSTEM

The proposed system involves the analysis of Emoji the post and to reduce the memory size by using map reduce algorithm. It is divided into four modulo which are explained below

3.1 Social Intelligence

Social media is the place where the users can share their views and interest. Based on the shares it is possible to decide about the users. Nowadays Emoji plays a vital role in deciding about the person. Peoples used to post data with the Emoji. In the phase of analysis the Emoji are consider as important part (deciding Emoji helps to understand about the user). For every Emoji Unicode will be assigned. The Emoji will of negative form or positive form. If a sentence followed by negative Emoji makes the sentence negative. Let us consider a instance post by a person is "I like mango(inverted hands)". Now the Unicode for the inverted hand is negative value. So the sentence will be reformed as "I hate apple". This data will be stored in the users' details.

3.2 Social Handover

In the social media the data about the users are shared with the company or service providers for marketing based on user interest. The interest of users are collected using the like and dislike buttons in social media. If the user is not analysed correctly then the user will be getting unwanted data and message. For instance if a person post I like dog (angry smile), it indicates that the person hates dogs. So the person data will not be shared to any of the dog consultancy. It saves the time and efficiency of social media. The interest of persons is also understood in the effective way. Now the clustering of data take place only with the peoples having similar interest.

3.3 Social Decision Making

This phase involves in the reducing the storage by using the map reduce algorithm. By using the map reduce algorithm the repeated values are replaced by their key values. For every data in the big data can be store effectively by using the key values for each data. A temporary memory can be created to store the repeating data. Whenever the data is needed it can be replaced by the key values. It reduces the size of memory. In

this phase the data processing will takes place. It involves the text analysis which removes the unwanted data. For instant “I like apple” in this post like apple will be stored and repeated words will be replaced with the unique. Similarly there is post say I like apple, it denotes two things it may be apple phone or fruit apple. By using the Emoji of phone it can be identified as apple phone. Then the suggestions for the person will be given related to the apple phone. Thus the user can be understand in more effective way and able to get clear visualization of the users

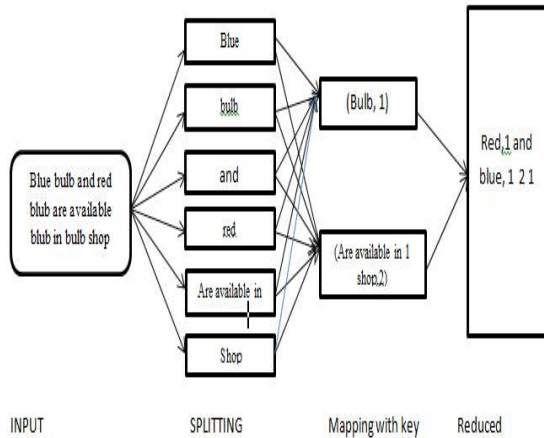


Fig 1 : Proposed Approach

3.4 Cognitive Mining

In social media there are large number of data of large size are shared every day. In this data it is possible to have the repeating data. The repeating data will leads to the waste usage of memory. If the repeating data are found then repeating data can be replaced with their reference. From the large storage sub domain storage will be created for users based on their interest and like. Thus handling of large number of users of similar likes can be handled effectively with low memory capacity. For instance thousands of people have similar likes, allowing them for direct accesses for the domain storage will slow down the processing speed of social media. If a sub domain is created for users of similar choice, it can reduce the over count of users accessing main domain. The functioning of sub domain will be same as the main domain storage.

IV. CONCLUSION

The social media is developing continuously with increasing customers and large volumes of data are shared. The usages of Emoji in social media have increased in larger scale. In this paper the problem of storage size can be reduced. By using the map reduce algorithm is used to reduce the memory. Only reference will be given to the repeating data. Thus the

analysis of Emoji plays a vital role in social media. By considering the Emoji in the message it is possible to understand the users likes and dislikes. In social media cognitive involves getting data about the user and predicts the user action. It so removes the unwanted data and repeated data. Efficiency, performance and mining speed everything will be increased.

Reference

- [1] Information Fusion, Social big data: Recent achievements and new challenges Gema Bello-Orgaza, Jason J. Jungb, David Camacho <http://dx.doi.org/10.1016/j.inffus.2015.08.005> 1566-2535
- [2] Big Data Analytics: Recent Achievements and New Challenges Dr Anand Mohan International Journal of Computer Applications Technology and Research International Journal of Computer Applications Technology and Research Volume 5–Issue 7, 460-464, 2016, ISSN:2319–865
- [3] Empirical Analysis of Data Mining Techniques for Social Network Websites S.G.S Fernando, COMPUSOFT, An international journal of advanced computer technology, 3 (2), February-2014 (Volume-III, Issue-II)
- [4] Guidelines for Social Media Mining for Innovation Purposes Robert Eckhoff, Mark Markus, Markus Lassnig, and Sandra Schön ,The Seventh International Conference on Information, Process, and Knowledge Management ISBN: 978-1-61208-386-5
- [5] R.L.Grossman, Y.Gu, J.Mambretti, M.Sabala, A.Szalay, K.White, An overview of the open science data cloud, in: Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing, HPCDC'10, ACM, New York, NY, USA, 2010, pp.377–384, doi:10.1145/1851476.1851533.
- [6] A. Roshini and H. Anandakumar, "Hierarchical cost effective leach for heterogeneous wireless sensor networks," Advanced Computing and Communication Systems, 2015 International Conference on, Coimbatore, 2015, pp. 1-7. doi: 10.1109/ICACCS.2015.7324082
- [7] S. Divya, H. A. Kumar and A. Vishalakshi, "An improved spectral efficiency of WiMAX using 802.16G based technology," Advanced Computing and Communication Systems, 2015 International Conference on, Coimbatore, 2015, pp. 1-4. doi: 10.1109/ICACCS.2015.7324098
- [8] M. Suganya and H. Anandakumar, "Handover based spectrum allocation in cognitive radio networks," Green Computing, Communication and Conservation of Energy (ICGCE), 2013 International Conference on, Chennai, 2013, pp. 215-219. doi: 10.1109/ICGCE.2013.6823431
- [9] H. Anandakumar and K. Umamaheswari, Supervised machine learning techniques in cognitive radio networks during cooperative spectrum handovers, Cluster Computing (2017), 1–11. doi: 10.1007/s10586-017-0798-3
- [10] E.R.Sparks, A.Talwalkar, V.Smith, J.Kottalam, X.Pan, J.E.Gonzalez, M.J.Franklin, M.I.Jordan, T.Kraska, MLI: an API for distributed machine learning, in: Proceedings of IEEE 13th International Conference on Data Mining, Dallas, TX, USA, December 7-10, 2013, pp.1187–1192, doi:10.1109/ICDM.2013.158.

- [11] A.K.Jain,M.N.Murty,P.J.Flynn,Dataclustering:areview,ACMComp ut.Surv.31(3)(1999)264–323, doi:10.1145/331499.331504.
- [12] B.Larsen,C.Aone,Fast and effective text mining using linear time document clustering,in: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and datamining, KDD'99, ACM, NewYork,NY, USA,1999,pp.16–22, doi:10.1145/312129.312186.
- [13] Suriya M,Sugandhanaa M,Vaishnavi J,Dhivya Bharathy P,"A Survey on Cognitive Handover between the terrestrial and Satellite Segment", 2016 IJAICT Volume 2, Issue 11, March 2016 Doi:01.0401/ijaict.2016.11.03 Published on 05 (04) 2016
- [14] Vamsi krishna tumuluru,Ping Wang,Dusit Niyato,wei song"performance analysis of cognitive radio spectrum access with prioritized traffic" IEEE Communications Letters (Volume: 18, Issue: 7, July 2014)
- [15] G.Ezrasastry, S.Tamilarasan, P.Kumar", Dynamic Resource Allocation in Cognitive Radio Networks". International Journal of Computer Science Trends and Technology (IJCS T) – Volume 4 Issue 3, May - Jun 2016
- [16] Apurvavarathe,sriramgulkarnee,"literature survey on routing protocols for cognitive radio network" International Journal of Innovative Research in Advanced Engineering (IJRAE) ISSN: 2349-2163 Volume 1 Issue 10 (November 2014)