REVIEW PAPER ON RANKING OF IMAGES BASED ON CAPTION ON SOCIAL MEDIA

Prof. Suniti Purbey[1], Ankita Lahor[2], Amey Kulkarni[3],

Chetan Walunjkar[4], Ashwini Shahapurkar[5]

[1]*Associate Professor, Computer Engineering, GHRCOEM, Maharashtra, India*
[2345]*UG Student, Computer Engineering, GHRCOEM, Maharashtra, India*

---***---

Abstract – *Now-a-days Instagram, Twitter, Facebook, etc. social media applications are widely used. These applications are very popular because of their simply and easily sharing techniques. Users are updating their data unlimited times. So many researchers started their research to study and analyse it for different purposes, such as data retrieval, advertising strategies, etc. Our target is here to create a customize social site based on caption of images on social media and rank that caption keywords by using TF-IDF. This method is used successfully in this paper. TF-IDF is an useful technique for find and rank multiple used keywords. This technique of data mining based on text mining. Hence this text mining technique used in proposed system. The main objective of this paper is to provide a feature to social media which are mention above, to make searching of images easily based on their captions.*

Key Words: Data mining, Text mining, Social media, TF-IDF, Classification

1. INTRODUCTION

Instagram, Twitter, Facebook, etc. social media platforms are very popular in adolescents as well as old age people. At least one social media account for the user. Day by day sharing of photo is increasing. Their three types of data on social media, like image, text and video. There is need to prove this study, for that text data are used. Text data is the caption or short information of the images uploaded by user. Keywords based on this short information are processed and rank according to their reuse. Text mining method is used for this purpose. And this can be calculated by TF-IDF.

In this proposed system TF-IDF as well as SVM function is used. The SVM function based algorithm is used for training the keywords, based on captions and send that keywords to TF-IDF for calculating their weight and term frequencies and inverse term frequencies. After that result stored in hash map, hash map is used for temporary store the count of keywords and weight of keywords. This technique is easily ranking the keywords.

In this system user and admin both have login authentication. Admin and users both are able to upload images with captions. User can send friend requests and accept friend requests. According to image and captions user and admin both are able see rank of top used keywords

images. In this system text ranking is also done, this result is displaying by generating bar-graph.

This system is giving result in image as well as text form.

2. LITERATURE SURVEY

[1] T.H.Silva and J.Salles in 2013 stated that, the photo sharing system work is done by use of fundamental tool participatory sensing system. This study is for city dynamics potentials. Also explain user's behaviours and a city based available data observes places which presents application to identify regions of interest. And it challenging for unequal photo sharing frequency is temporally.

[2] Manikonda and S. Kambhampati in 2014 stated various insights related to Instagram, these are: 1) 5 different types of users in terms of their uploaded photo, 2) Famous photos categories and last, 3) Followers of user is independent of their shared images on Insta. To identify distinct types of users which are active by using clustering method they used quantitative and qualitative analysis.

[3] Anbarasi M and Buvaneswari in 2014 proposed that, in various databases there is growth in multimedia content, which has issue in access and retrieval in both textual as well as non-textual sources. Information is stored in the cloud environment to resolve this problem. Large and complex data cannot be stored from Internet, also not processed using data processing applications. To extraction of texts and images one idea is involved which is parsing of the web pages. Texts retrieval is complete with keyword extraction whereas feature extraction technique is done for retrieval of images. For clustering purpose K-means algorithm is used. Textual and non-textual data retrieved together.

[4] Rashika Dandel and Monika Gutal in 2015 stated that, the exponential growth in the multimedia content due to Internet. It consist different features like text, images, video and audio etc. To access and retrieval of both textual and non-textual data, there are many problems, so need to solve these problem by using cloud storage. By parsing the web pages they are using extraction of textual and non-textual data. Through feature extraction image retrieval is done and through textual data information retrieval is done. For that K-means algorithm is used.

[5] Z. Mccune and J. Thompson in 2011 stated, Instagram i-phone app is taking as a source of study. The production social photographic study is done by users; they will be examined by analysis. Research is done by creating question, why do they share personal information? This will challenge to culture critics.

[6] C.S. Araujo and D. Silva stated that, one of the social photos sharing service is Instagram. In that users can be used as an examine purpose. Users can like to endorse their photos with many likes and comments, in that rich can get rich phenomenon.

[7] A. Kumar and T. M. Sebastian in 2012 stated that, there has been various user generated content on social media, because of rise of popularity of it. There are millions of users for micro blogging sites. They are sharing their thoughts daily because of its characteristic short and simply format. With the help of twitter people can post real-time reactions under provided services from them. Easily gives opinions about anything, with the help of paradigm. This is done with the help of hybrid approaches i.e. corpus and dictionary based methods. These are used to determine the semantic orientation opinions.

[8] G. Salton and C. Buckley described, in modern society SQL language is used as a database. Their functions are mainly focuses on data processing, used in data mining. Large scale of data processing is become focal point because of fast growth. SQL is used, but to handle large amount of data cloud computing is done. Main purpose is to deal with large amount of data process. In this paper they develop data mining based system. This is directly deal with SQL processing on Hadoop. Hadoop is parallel computing platform. They will have discussion about time's efficiency.

3. Proposed work

This is proposed system architecture with all their modules. In that first user can login with their id and password. Then user is able to upload their images with particular caption, which is suitable for that image. Then Support Vector Machine (SVM) will be training those keywords. After training, keywords are calculated by TF-IDF.
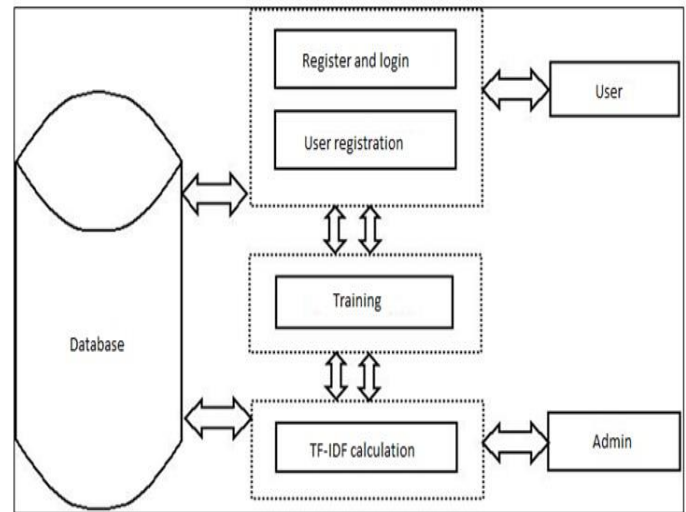

Fig -1: System architecture

In above figure showing that admin is also able to login with their login id and password and able to upload images with caption. Also user can register their information.

In TF-IDF calculated the all terms which are used in documents and in all documents this term is used. And keywords are calculated by multiplication of both TF-IDF. This idea is taking from the literature reviews we have done.

4. CONCLUSIONS

Hence TF-IDF method is very simple and effective. It can easily retrieve the keywords and providing ranking based on user's data. The main topic is showing top ranking of keywords. The higher the keyword is more used keyword of a particular user. This work is still necessary to improve in some manner, like understanding non-English language, slang words etc. Admin also able to generate a graph based text ranking as well as TF-IDF calculated result is displaying in image format.

REFERENCES

[1] T. H. Silva, P. O. S. V. D. Melo, J. M. Almeida, J. Salles, A. A. F. Loureiro "A picture of Instagram is worth more than a thousand words: Workload characterization and application", DCoSS 2013 IEEE International Conference on Distributed Computing in Sensor Systems ,2013.

[2] Y. Hu, L. Manikonda, and S. Kambhampati, "What we Instagram: a first Analysis of Instagram photo content and user types", Eight International AAAI Conference on Webblogs and social Media, 2014.

[3] Anbarasi M, Divya , Buvaneswari , Illakkiya, "Retrieval of Textual and Non-Textual Information in cloud", 2014 IJRET: International Journal of Research in Engineering and Technology , 2014

[4] Rashika Dandel, Monika Gutal, Bhagyashree Dudhal, Poonam Jadhav, Sonali Kadam, ' Survey on Retrieval of Textual and Non-Textual Information in Cloud", 2015 International Journal of Advanced Research in Computer Science and Software Engineering, 2015

[5] Z. Mccune and J. Thompson , "Consumer Production in Social Media Networks: A Case Study of the Instagram iPhone App", 2011 University of Cambridge, 2011

[6] C. S. Araujo, L. P. D. Correa, A. P. C. D. Silva, R. O.Prates, W. Meira, "It is Not Just a Picture: Revealing Some User Practices in Instagram", web mining, 2014 9th Latin American Web Congress , 2014.

[7] A. Kumar and T. M. Sebastian, "Sentiment Analysis on Twitter", Data mining and web mining, 2012 International Journal of Computer Science Issues, 2012.

[8] G. Salton, C. Buckley, "Term-weighted approaches to automatic text retrieval", Data mining, 1988 Information Processing Management: an International journal, 1988.