

LINEARNNSEARCH CLASSIFIER USING TOPIC MODELING FOR ACADEMIC SOCIAL NETWORK DATASET

¹G.Ayyappan, ²C.Nalini, ³A.Kumaravel

¹Research Scholar, Department of Computer Science and Engineering, Bharath University, Chennai

²Professor, Department of Computer Science and Engineering, Bharath University, Chennai

³Professor, School of Computing, Bharath University, Chennai

¹ayyappangmca@gmail.com, ²dnalinichidambaram@gmail.com, ³drkumaravel@gmail.com

Abstract: Now a day we are mostly involving the social media. Without social media we can't do any work in the world. It is simply information sharing from one to others. In this paper we focus the K nearest neighbor LinearNNSearch classifications using Topic modeling. We collect all abstracts from the academic social network dataset. The Topic distribution like as Algorithm, data mining, database, Artificial Intelligence, Clinical, Medical Imaging, Image Processing, Biomedical Informatics, Image Processing, Telemedicine. There are ten topics we formed from the actual abstract of academic social network dataset. We measure the academic social network dataset how we apply topic modeling and how we apply k nearest neighbor one of the LinearNNSearch classification and find out which is best method for our research community.

Keywords: Social Network, MeSH, String to vector, Topic modeling, KNN.

1. Introduction

We describe the major tasks in social network analysis data sets used for studying this problem. We find authors belong to which topic in the dataset. We made topic model processes, we converted all abstracts in to topic for using **MeSH(Topic Extraction Tool)**. The Collections of nodes are called each and every person who involved in the social network. The Topic distribution like as Algorithm, data mining, database, Artificial Intelligence, Clinical, Medical Imaging, Image Processing, Biomedical Informatics, Image Processing, Telemedicine.

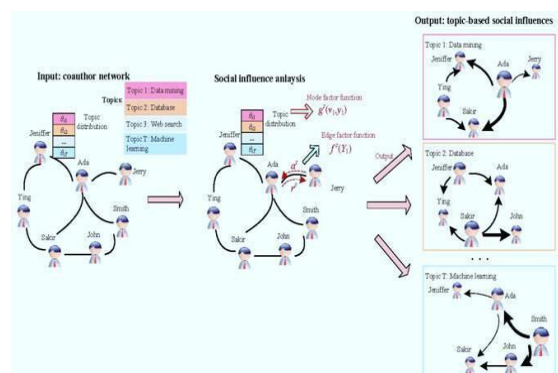


Figure 1. Co-Authors network in Social Network Analysis

Here, the related works in section 2. Materials and Methods in section 3, the result and discussion of various classifiers approaches in section 4 represents. We present the conclusion in section 5. Selecting this dataset which is clean and simple, and preprocessing for appropriate format we follow

2. Literature Survey

Shichao Zhang et al.[1] proposed and developed an enhancement version of kTree method (k*Tree method) to increase speed its test stage by extra storing the information of the training samples in the leaf nodes of kTree, i.e., training samples located in the leaf nodes, their kNNs, and the nearest neighbor of these kNNs. They called the resulting decision tree as k*Tree, It enables to conduct kNN using a subset of the training samples in the leaf nodes rather than all training samples used in the newly kNN methods. It is actually reduces running cost of test stage. the results on 20 real data sets showed that our proposed methods (ie., kTree and k*Tree) are much more efficient than the compared methods in terms of classification tasks.

M.Akhil jabbare et al.[2] Proposed and developed a new algorithm which combines KNN with genetic algorithm for effective classification

methodology. The Genetic algorithms perform global search in composite large and multi modal landscapes and provide optimal result. The output shows that our algorithm enhances the accuracy in diagnosis of heart disease.

Ahmed Alsayat et al.[3] proposed and developed framework is twofold that is to overcome the problem of general K-Means for choosing best initial centroids using Genetic algorithm, and also to maximize the distance between clusters by pairwise clustering using OCD to get an accurate clusters in the calssifications. They used several cluster validation metrics to evaluate the performance of our algorithm method. The analysis shows that the proposed method gives better clustering results and provides a novel use-case of grouping user communities based on their behaviors. Their approach is optimized and scalable for real-time clustering of social media data.

Yu-Ting Wen et.al.[4]proposed and developed an efficient Keyword-aware Representative Travel Route framework that uses knowledge extraction from users' historical mobility records and social networks. They have designed a keyword extraction module to classify the POI-related tags, for effective matching with query keywords. They designed a route reconstruction algorithm to construct route candidates that fulfill the necessities. Santanu Mandal ; et.al.[5] proposed and developed a novel dictionary-based algorithm that uses lexicon-based approach for opinion mining and calculates the sentiment polarity levels. Their algorithm is different from other lexicon-based algorithms in the context that it uses the three degrees of comparisons are positive, comparative , superlative degrees .

3. Materials And Methods

We use weka tool for classifications. Waikato Environment for Knowledge Analysis (Weka) is a suite of machine learning s/w written in Java code, It has developed at the University of Waikato, New Zealand. It is free software licensed under the GNU General Public License.

The k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification. the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression:

- The output is a class membership.
- An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, small).

- If $k = 1, 2, \dots, 10$ then the object is simply assigned to the class of that single nearest neighbor to Tenth nearest neighbor.
- We apply three different methods in K-NN.

a) *LinearNNSearch*

4. Experiments And Results

Table1. LinearNNSearch Classifications Accuracies with Time Take to build the model

KNN	LinearNNSearch	Time Taken to build Model for LinearNNSearch (In Seconds)
1	95.3796	0.03
2	95.5156	0.02
3	95.5483	0.01
4	95.5483	0.02
5	95.5483	0
6	95.5374	0
7	95.5537	0.02
8	95.5537	0.03
9	95.5592	0
10	95.5592	0.01

In this research work we apply various k values in the topic modeling relational data model. We identify the high accuracy based on the below table

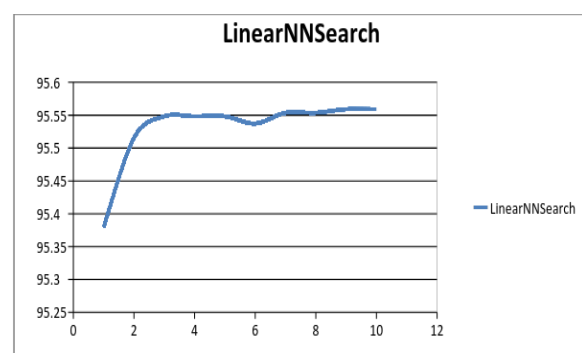


Figure 2. Accuracy Vs k-NN

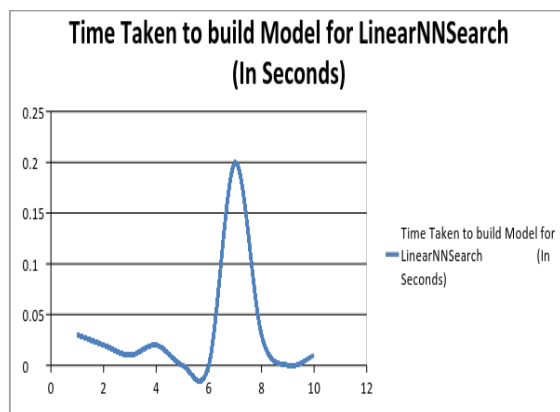


Figure 3. Time consumption to build model Vs k-NN

We apply this LinearNNSearch method we apply k-NN =1 to 10 in the Ibk model in weka tool. When we apply k-NN =1 the output is 95.3796 with time consumption is 0.03. when we apply k-NN =2 the output is 95.5156 with time consumption is 0.02, When we apply k-NN =3,4,5 the results are same result. i.e., the output is 95.5483 but time consumption is like 0.02,0,0 seconds. When we apply k-NN = 6 the output is 95.5374 and time consumption is zero second. When we apply k-NN =7,8 we get same result like as 95.5537 but the time consumption is 0.02 and 0.03 seconds. We find the high accuracy when we apply k-NN =9 and k-NN =10 but, time consumption taken when we apply k-NN =9 is zero seconds. But when we apply k-NN =10 time consumption is 0.01.

5. Conclusion

In this paper we conclude K-Nearest Neighbor Method for classifications for academic social network dataset. We LinearNNSearch is producing the results when we tune k-NN=1,2,...10. But we recommended when we apply k-NN =9. It has zero seconds and high accuracy 95.5592 for this social network dataset.

References

- [1] Shichao Zhang ; Xuelong Li ; Ming Zong ; Xiaofeng Zhu ; Ruili Wang, Efficient kNN Classification With Different Numbers of Nearest Neighbors, IEEE Transactions on Neural Networks and Learning Systems (Volume: PP, Issue: 99) Page(s): 1 - 12,Date of Publication: 23 May 2017.
- [2] M.Akhil jabbar,B.L Deekshatulua, Priti Chandra Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm, Procedia Technology 10 (2013) 85 – 94].
- [3] Ahmed Alsayat ; Hoda El-Sayed [3] Social media analysis using optimized K-Means clustering Software

Engineering Research, Management and Applications (SERA), 2016 IEEE 14th International Conference on 21 July 2016.

[4] Yu-Ting Wen ; Jinyoung Yeo ; Wen-Chih Peng ; Seung-Won Hwang, Efficient Keyword-aware Representative Travel Route Recommendation IEEE Transactions on Knowledge and Data Engineering (Volume: PP, Issue: 99),Page(s): 1 – 1,Date of Publication: 03 April 2017, Print ISSN: 1041-4347

[5] Santanu Mandal ; Sumit Gupta, Research in Computational Intelligence and Communication Networks (ICRCICN), 2016 Second International Conference on 23-25 Sept. 2016 ,IEEE EXPLORER 16 January 2017,Electronic ISBN: 978-1-5090-1047-9.

[6]<https://www.nlm.nih.gov/mesh/MeSHonDemand.html> [Online on 2.5.2017 to 5.5.2017]

[7] <https://aminer.org/lab-datasets/soinf/>

[8] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social Influence Analysis in Large-scale Networks. In Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'2009).

[9] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. ArnetMiner: Extraction and Mining of Academic Social Networks. In Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'2008). pp.990-998.

