Document Recommendation using Boosting Based Multi-graph Classification: A Review

Vrushali Deore, Pooja Kamble, Reshma Bendkule, Manisha Dhatrak, Prof. S. W. Jadhav

Department of Computer Engineering Student MET BKC, Nasik, India.
Department of Computer Engineering Professor MET BKC, Nasik, India.

--***---

Abstract - *Every day the mass of information available to use is increase. So we need to increase the ability to efficiently access this information. Text Classification is hard if we do it manually. So we need a tool that classifies text and images more accurately. In Existing system, users need to check various documents to find out similar documents. So the accuracy of related document is very less. It is very time consuming process. The proposed system is a recommendation system with maximum accuracy and minimum time. System recommends documents based on the users query. Also calculate the probability based on classification. Boosting based multi-graph classification technique is used by system to classify the documents. It provides most related documents to the user. System preprocesses the documents and finds the common author relation. System uses multi-graph and article ranking to find the most relevant documents.*

Key Words:-Multi-graph, Boosting, Feature vector citation, Common author relation, Article Ranking, TF-IDF.

1. INTRODUCTION

In today's era a user wants to search any topic then he/she will search on www by giving an input in text format. After searching, many times user won't get the results because same text or words having same meaning. So due to that accuracy of getting correct results of query is less. And users have to search all the links which he gets. So it is very time consuming.

We are going to implement a system which will reduce overhead of user of searching many pages and documents for single query. In the system user upload the document. Then system will perform preprocessing and text mining on data set. So the stop words are removed and important words get mined. According to that, searching performed. System calculates the probability of related documents and multi-graph classification (bMGC) [1] is done. Then system recommends the documents or articles which are closer to input. Users also get the probability of the documents which are more relevant for recommendation.

1.1 Boosting

Boosting is a machine learning meta-algorithm which reduces bias and variance in supervised learning. It is a family of machine learning. Boosting converts weak learners to strong learners. Weak learner classifier is less correlated with true classification. Strong learner classifier is more correlated with true classification. Boosting algorithms consist of learning weak classifiers and adding them strong classifier. The proposed system uses AdaBoost algorithm. AdaBoost is popular machine learning algorithm which adapt to the weak learners.

1.2 Graph Classification

Multi-graph classification problem is viewed as a graph classification problem. In which objects are consider as bag of graphs. Classification of objects is based on the multiple graphs. It can be classified into following two categories:

1. Global Distance Based Approaches: This method is based on the similarities and correlations [2] between two graphs. One drawback of this method is, it is not clear which part of graph is more discriminative for differentiating graphs of different classes.
2. Local Sub graph Feature Based Approaches: This method is based on the frequency of most common sub graph selection which select frequently appearing sub graphs by using frequent sub graph mining methods. One drawback of this method is to handle large graph sets.

To overcome this drawback, some boosting methods [3]–[6] is use sub graph feature as a weak classifier, including some other types of boosting methods [7], [8] for graph classification.

2. LITERATURE SURVEY

2.1 Introducing Docear's Research Paper Recommender System

Docear's recommender system [10] proposed for Docear. Basically, Docear is open source tools which build the literature management tool for searching, organizing and creating literature structure for Researchers and Students.

Docear's has its own digital library. According to user's literature the Docear's literature collection are maps with the Docear's Digital Library. When user request article Docear check that user request in the Digital Library. Based on that user model are created and return the ten recommended articles to user. It only returns ten recommendations to user around 1.8 billion research article.

2.2 The Architecture and Datasets of Docear's Research Paper Recommender system

In this paper, the architecture and four dataset are used to develop the Docear's Recommender System [9]. The system uses the multiple components for implementing architecture of user model and based on that contents are calculated. System uses the Docear's open source tool. There are four dataset contains academic articles, web articles, articles citation and personal libraries. The system uses content-based filtering approach for user's mind-map.

The architecture of Docear's research paper recommender system client requests recommendation to web services. Mind-map parser process request and store in the mind-map model. In recommendation Engine, using algorithm user model are created and matched with existing model of document dataset. PDF analyzer converts PDF to text, extracting header and create citation. Then indexing to PDF is done to implement paper model. Recommendation and Statistics database display the recommendation articles to use. This is used to recommend the research paper to the clients using the header extraction and citation

3. PROPOSED SYSTEM

To propose a recommendation method, which incorporates common author relations between documents to help generate better recommendations for relevant target users using side-information. Such side information may be of different kinds, such as document provenance information, the links in the document, user-access behavior from web logs, or other non-textual attributes which are embedded into the text document. In this research new guidelines are propose which combines classical partitioning algorithms with probabilistic models in order to create an effective clustering approach.

In proposed system user upload document and based on that system recommends the most relevant documents. System preprocesses the data set. Then meaningful features are extracted from the data set. Then it finds relation between common author and their citation. Historical preferences of researchers are consider. So here using multi-graph classification the documents are classified. It uses graph based ranking to recommend the documents from the data set. Based on that, system recommends documents to user.
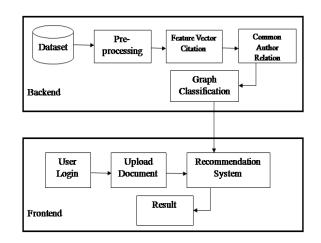

Figure. System Architecture

3. CONCLUSIONS

Thus we proposed the system which uses the multi-graph classification for classifying the documents. User is able to upload the document and based on that system recommends most accurate and relevant documents to user. System uses bMGC algorithm for multi-graph classification of documents. System provides the probability of the recommended documents. So this proposed system is time efficient and provides most accurate result.

REFERENCES

[1] Jia Wu, Student Member, *IEEE*, Shirui Pan, Xingquan Zhu, Senior Member, *IEEE,* and Zhihua Cai "Multi-graph classification using boosting".

[2] J. Wu, X. Zhu, C. Zhang, and Z. Cai, "Multi-instance multi-graph dual embedding learning," in *Proc. 13th ICDM*, Dallas, TX, USA, 2013, pp. 827–836.

[3] T. Kudo, E. Maeda, and Y. Matsumoto, "An application of boosting to graph classification," in *Proc. 18th Annu. Conf. NIPS*, 2004, pp. 729–736.

[4] S. Nowozin, K. Tsuda, T. Uno, T. Kudo, and G. Bakir, "Weighted substructure mining for image analysis," in *Proc. 20th IEEE Conf. CVPR*, Minneapolis, MN, USA, 2007, pp. 1–8.

[5] H. Saigo, S. Nowozin, T. Kadowaki, T. Kudo, and K. Tsuda, "gBoost: A mathematical programming approach to graph classification and regression," *Mach. Learn.*, vol. 75, no. 1, pp. 69–89, 2009.

[6] S. Pan and X. Zhu, "Graph classification with imbalanced class distributions and noise," in *Proc. 23rd IJCAI*, 2013, pp. 1586–1592.

[7] H. Fei and J. Huan, "Boosting with structure information in the functional space: An application to graph classification," in *Proc. 16th ACM SIGKDD*, Washington, DC, USA, 2010, pp. 643–652.

[8] B. Zhang *et al.*, "Multi-class graph boosting with sub graph sharing for object recognition," in *Proc. 20th ICPR*, Istanbul, Turkey, 2010, pp. 1541–1544.

[9] L. Palopoli, D. Rosaci, and G. M. Sarn, A Multi-tiered Recommender System Architecture for Supporting E-Commerce, in Intelligent Distributed Computing VI, Springer, 2013, pp. 7181.

[10] Y.-L. Lee and F.-H. Huang, Recommender system architecture for adaptive green marketing, Expert Systems with Applications, vol. 38, no. 8, pp. 96969703, 2011.