A Novel Approach To Group Research Proposal and Allocate the Research Reviewer by Using Text-Mining and Clustering

GEETA DANGAR, VIPUL VEKARIYA and DAXA VEKARIYA

Department of Computer Engineering, Noble Group of institutions, Junagadh, India.

ABSTRACT

In any of the organizations many efforts is taken for research project selection.It is necessary for each research proposal, before it given to the appropriate experts, research proposals needed grouping according to their research topics and area of research. In this paper clustering methods and text data mining approaches are used to group research proposals and give it to expert reviewer for reviewing the proposals. This paper uses the real data of BNSF (Beijing Natural Science Foundation).

Key words: Cluster; Data Mining; Expert Review; Text Mining.

INTRODUCTION

Research Project Selection is very important aspect in process of selection the research project selection. For that before we select any of the research proposals it needs to be grouped for the better project selection. So in this paper we introduce one technique for grouping of the research proposals and assign this group to the related reviewers[1].

Fig-1 shows the common method which is generally used by various funding organizations for selection of research project process. According to fig-1, in any organization first step is invites various research proposals. Second Step is: These organizations receive proposals and collect them from various universities or research institutions. Furthermore received proposals or offers are categorized according to their techniques or problems on that they are based. This will help to handover exports for their evaluations. After categorize they are allotted to numerous experts of the related field or technology for their valuable guidance and evaluations. When expert evaluate the proposals they submit their evaluations result to the funding organization. On behalf of these evaluation result funding panel discuss importance of the proposal and afterwards decide the final decision. The final decision may contains where proposals are eligible for funding or not and the limit of fund also be decided. So in this way the research proposal is being choose[2].

Related work

In[1] authors used text mining and clustering technique for group the research proposals selection. Various methods of the research paper selection is using KNN is given in[2]. KNN is the K-Nearest Neighboring technique for clustering; here it is used for cluster the research proposal selection. Ontology based text mining technique also used for group the research proposal selection[3,4].This technique is used with any clustering technique and generates the group of research proposal selection[5]. One another technique calls APN (Analytic Network Process) use for research proposal selection that is given in[6]. For reviewer assignment hybrid approach is there in[7]. Lots of efforts are being taken by organization to give guarantee of the effective research project selection. We have some studies to solve this type of problem from different perspective. For example, Machacha and Bhattacharya presented fuzzy logic method for clustering of the given proposals[8], And Huang et al. employed fuzzy AHP[9] and judgment crisp matrix for clustering. Similarly, Amiri developing the method names adding fuzzy TOPSIS in clustering[10]. Chang and Lee used DEA method with knapsack formulation and fuzzy set theory and clustering[11].

Proposed System

There are some limitations of the algorithm which are used in the existing system so we use
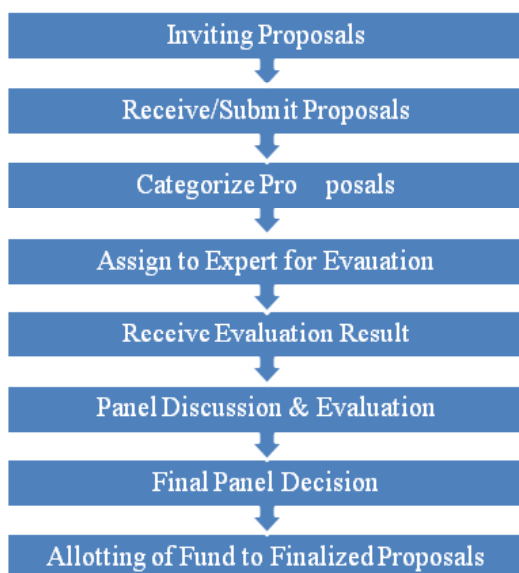

Fig. 1: Process of project selection

some other algorithm in this system. One another limitation of the existing system, it cannot give the review of external reviewer. So we develop a new System which overcomes this limitation.

Limitation of SOM

In SOM we have to define the number of cluster previously and decision of cluster is based on user. This limitation of SOM is overcome by ASOM algorithm which is Adaptive Self Organizing Map Algorithm. This algorithm requires the range of cluster instead of number of cluster. After the range of the cluster is define, it searches the space of the cluster location and the number of cluster using information criterion measure.

Limitation of EM

When the fraction of missing information is big and dimensionality of the data is large EM does not works properly. Scalability is also problem in that in term of iteration because it requires much iteration to cluster the object. These all the limitation of this algorithm is overcome with the use of K-mean algorithm so we used K-mean instead of EM. The *k*-means algorithm is take the input as, *k*, and partition into a set of objects n, into clusters k. Similarity measurement of is regarding to the *mean* value of the cluster's objects, which is the *centroid* of cluster or *centre of gravity*.

Limitation of DBSCAN

DBSCAN is used to cluster the objects with given input parameters like e(radius of object) and *MinPts(minimum number of object from the radius)*, but it still has limitation , that DBSCAN leaves the decision of select the value parameter to the user. To overcome this difficulty of algorithm we using a clustering method called OPTICS. OPTICS has been computed an augmented *cluster ordering* rather than it produce a data set clustering explicitly, for interactive and automatic clustering analysis in any organization. This cluster ordering method has been use for extraction of basic clustering information (like centres of cluster or arbitrary-shape clusters) also it provide the intrinsic clustering structure for any organization.

Fig-2 shows the block diagram of the proposed system. Grouping of proposals is being done in following steps:
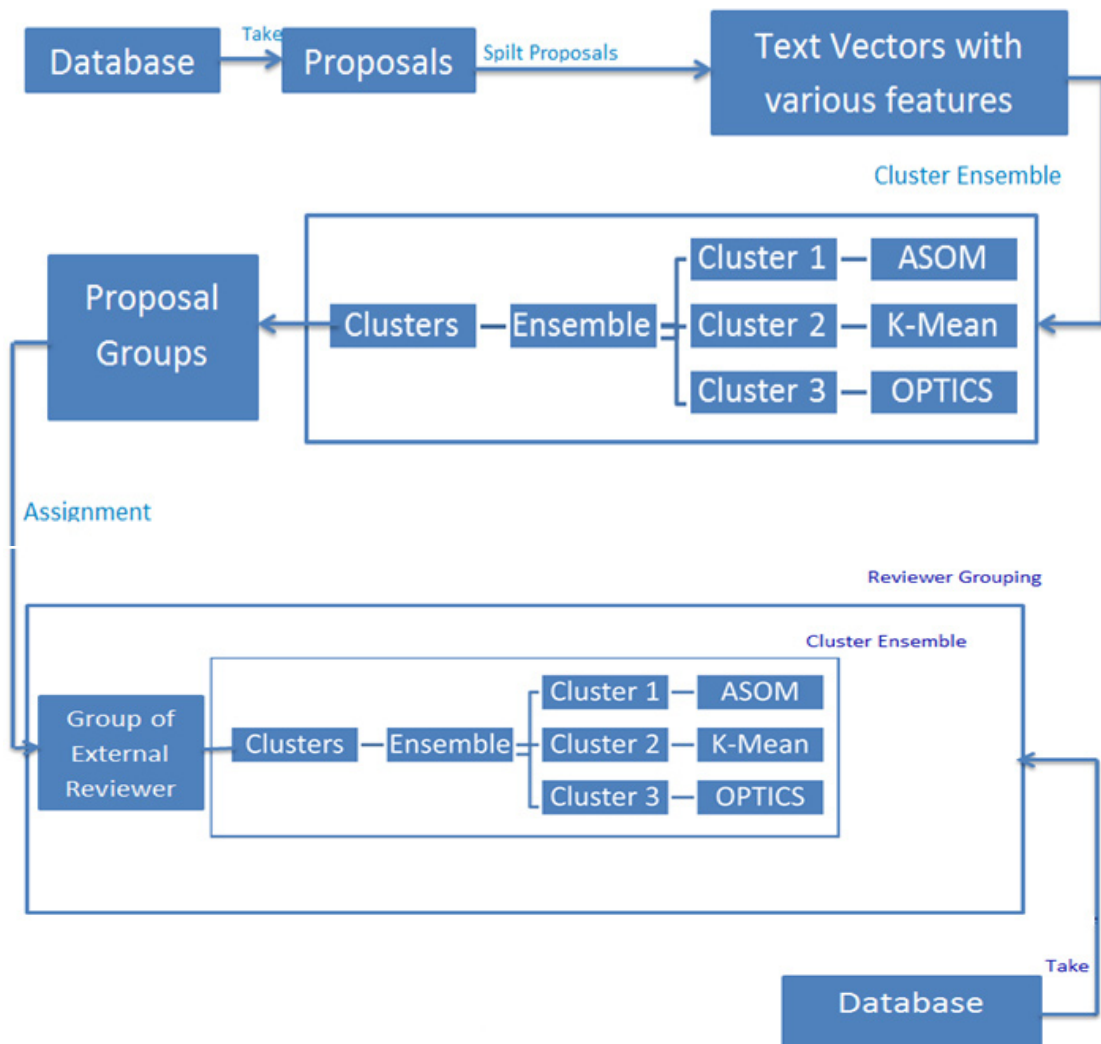
Fig. 2: Block diagram of proposed system

Step 1: In the first step Segmentation is there for splitting the proposals.

Step 2: After Segmentation that proposal is split, then after text mining technique is used for abstract the keyword and generate the text vector

Step 3: For clustering first 3 clustering algorithm is which we describe previous, then after cluster ensemble technique is used.

Step 4: So finally the grouping of the proposal is completed and it assign to the appropriate reviewer.

CONCLUSION

In existing system no any external reviewer is there for review the project proposals so in proposed system external reviewer is there for review the proposals. Also in proposed system the algorithm used is different because in existing system some algorithms have limitation which can be overcome by these algorithms. In this paper a cluster ensemble-based method is used with text mining for group of research proposals in any institution.

REFERENCES

1. Yibo Wang,Wei Xu,Hongxun Jiang, "Using Text Mining and Clustering to Group Research Proposals for Research Project Selection" , 1530-1605/15 2015 IEEE ,pp.1256 - 1263

2. Devendra Singh Rathore, Dr. R.C.Jain, Babita Ujjainiya, "A Text Mining Method for Research Project Selection using KNN", 978-1-4673-6126-2/13/.00c 2013 IEEE,pp.900-904

3. Prof. Jayvant H. Devare, Dinesh S. Gunjal, Navanath G. Gandole, Mayur S. Nimase, "Ontology Based Text Mining Method to Cluster Proposals for Research Project Selection", Volume 2, Issue 7, March 2015, pg. 2329-2334.

4. Smita Bachal, Prof. S.M. Sangave, " Survey on An Ontology-Based Text-Mining Method to Cluster Proposals for Research Project Selection", **1**(2): (April 2014),pg.160-163.

5. Jian Ma, Wei Xu, Yong-hong Sun, Efraim Turban, Shouyang Wang, and Ou Liu, "An Ontology-Based Text-Mining Method to Cluster Proposals for Research Project Selection", IEEE transactions on systems, man, and cybernetics—Part A: Systems and humans, **42**(3): pp.784-790 (2012).

6. Laura M. Meade and Adrien Presley, "R&D Project Selection Using the Analytic Network Process", *IEEE Transactions on engineering management,* **49**(1): pp.59-66 (2002).

7. Y. Sun, J. Ma, Z. Fan, and J. Wang, "A hybrid knowledge and model approach for reviewer assignment", *Expert Systems With Applications*, **34**(1): pp. 817-824 (2008).

8. L. L. Machacha and P. Bhattacharya, "A fuzzy-logic based approach to project selection", *IEEE Transactions on Engineering Management,* **47**(1): pp. 65–73 (2000).

9. C.-C. Huang, P.-Y. Chu and Y.-H. Chiang, "A fuzzy AHP application in government-sponsored R&D project selection", *Omega-international Journal of Management Science,* **36**(6): pp. 1038–1052 (2008).

10. M. P. Amiri, "Project selection for oil-fields development by using the AHP and fuzzy TOPSIS methods", *Expert Systems with Applications,* **37**(9): 6218–6224 (2010).

11. P.-T. Chang and J.-H. Lee, "A fuzzy DEA and knapsack formulation integrated model for project selection", *Computers & Operations Research,* **39**(1): pp. 112–125 (2012).

12. https://en.Wikipedia.org