

A Dynamic Approach to Group Research Proposal and Allocate the Research Reviewer by Using TF-IDF and Clustering

Geeta Dangar¹, Dr.Vipul Vekariya²

PG Student, Dept. of CE, Noble Group of Institute, Junagadh, Gujrat, India¹

Associate professor, Dept. of CE, Noble Group of Institute, Junagadh, Gujrat, India²

ABSTRACT: As we all know that great efforts have been taken to guarantee the efficiency of research project selection. Before research proposals are assigned to appropriate experts for evaluation, research proposals needed to be grouped according to their similarities in research topics. For grouping of the research project selection various data mining technique is being used. In this grouping of the proposal clustering is used. In clustering various technique is there from that here mainly three techniques is being used, first ASOM, second K-MEAN, and third is OPTICS. These all method is being used to cluster research proposals and support research project selection. These cluster algorithms are applied to group the proposals and the merits of each algorithm in text clustering are being used.

KEYWORDS: Cluster; Data Mining; Expert Review; Text Mining.

I. 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 reviewers¹.

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.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

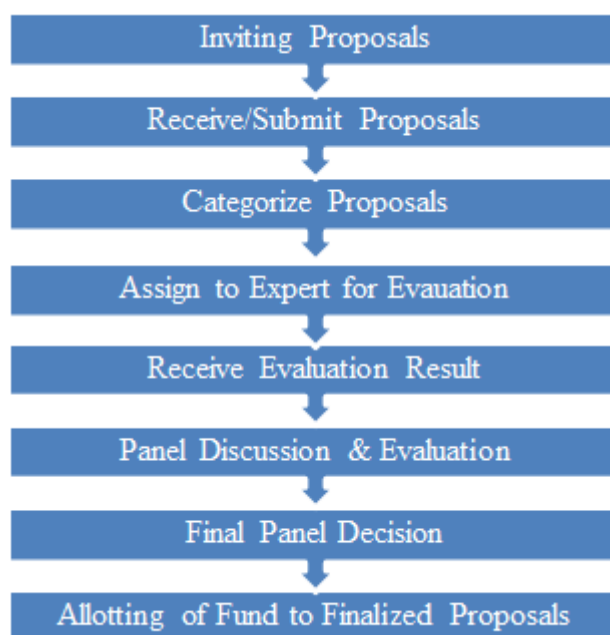


Fig1. Process of Research Project Selection

II. 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 Neighbouring 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].

III. PROPOSED SYSTEM

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

Algorithm which used by previous system are:

- 1) SOM
- 2) EM
- 3) DBSCAN

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

LIMITATION OF THESE ALGORITHMS:

1) Limitation of SOM:

In SOM we need to characterize the quantity of group already and choice of bunch depends on client [13]. This restriction of SOM is overcome by ASOM calculation which is Adaptive Self Organizing Map Algorithm. This calculation requires the scope of bunch rather than number of group. After the scope of the bunch is characterized, it seeks the space of the group area and the quantity of group using information criterion measure.

2) Limitation of EM:

At the point when the part of missing data is enormous and dimensionality of the information is expansive EM does not work legitimately. Adaptability is additionally an issue in that in terms of cycle since it requires much emphasis to group the item. These are all the restrictions of this calculation which are overcome with the utilization of K-mean calculation so we utilized K-mean rather than EM [14]. The k-means calculation takes the information as n , k , and allotment into an arrangement of items n , into bunches k . Comparability estimation of n is with respect to the mean estimation of the group's articles, which is the centroid of bunch or center of gravity.

3) Limitation of DBSCAN:

DBSCAN is utilized to group the items with given data parameters like ϵ (radius of article) and MinPts (minimum number of item from the range), yet despite everything it has a constraint, that DBSCAN leaves the choice of selecting the worth parameter to the client. To conquer this trouble of calculation we utilize a bunching strategy called OPTICS. OPTICS has been registered as an expanded group requesting instead of it delivering an information set bunching expressly, for intelligent and programmed bunching investigation in any association. This bunch requesting technique has been used for extraction of essential grouping data (like focuses of bunch or subjective shape bunches) additionally it gives the inborn grouping structure to any association [15].

Above Fig-2 demonstrates the block diagram of the proposed method. Grouping of propositions is being done as per the following steps:

Step 1: First step is data taken from the database and text mining algorithm is applied to that data.

Step 2: Second step is clustering, for clustering three different clustering algorithms are used, which we describe earlier, afterwards cluster ensemble method is being used.

Step 3: Third step is assignment, in that the grouping of the proposition is being completed and it allocates to the appropriate reviewer

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

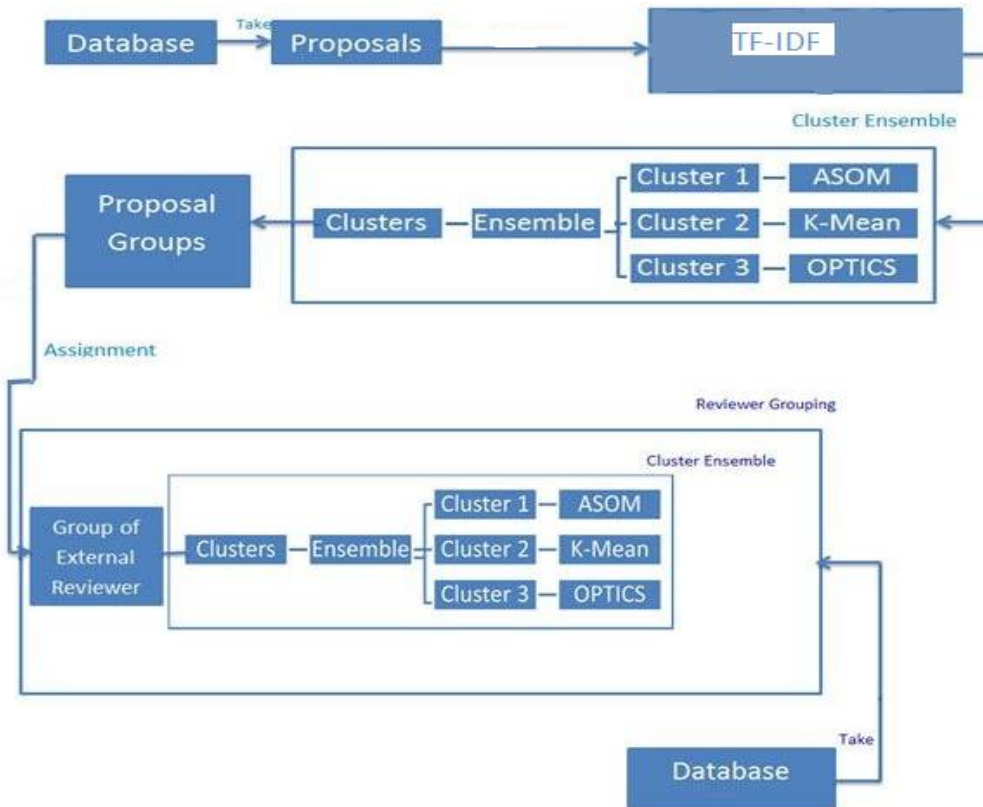


Fig2. Process of Research Project Selection

IV. EXPERIMENTAL ANALYSIS

To confirm the efficiency of the proposed system, several experimentations are accompanied using the previous research propositions. The quantity, which consists of 1001 research propositions with 9 different research area or branch, has been carried out in the experiential study. Moreover, to calculate the execution of the clustering algorithms, F measure is being used, because F measure is a standard for measuring the quality of text clustering and other data mining or text mining algorithm. For previously defined research topic t and cluster c , the Precision and Recall is being defined as follows:

$$\text{Precision}(c,t) = n(c,t)/n_c$$

$$\text{Recall}(c,t) = n(c,t)/n_t$$

where $n(c,t)$ be the amount of research propositions in both topic t and cluster c , n_c be the number of research propositions in cluster c , and n_t be the number of research proposals in cluster c .

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

The F measurement can be denoted as follows:

$$F = \sum_i \frac{n_i}{n} \max_j \{F(i, j)\}$$

$$\text{where } F(c, t) = \frac{2 \text{Recall}(c, t) * \text{Precision}(c, t)}{\text{Recall}(c, t) + \text{Precision}(c, t)}$$

Below figure shows the result comparison of the both existing system and system which is being introduced:

Table 1 Experimental Result of the System

Method \ NRP	100	300	500	1000
ASOM	0.41	0.54	0.64	0.76
K-MEANS	0.44	0.52	0.57	0.61
OPTICS	0.42	0.48	0.42	0.73
ENSEMBLE	0.61	0.49	0.65	0.66

Table 2 Experimental Result of previous System

Method \ NRP	100	300	500	1000
SOM	0.43	0.53	0.61	0.76
EM	0.59	0.47	0.61	0.68
DB-SCAN	0.39	0.47	0.39	0.61
ENSEMBLE	0.51	0.64	0.67	0.80

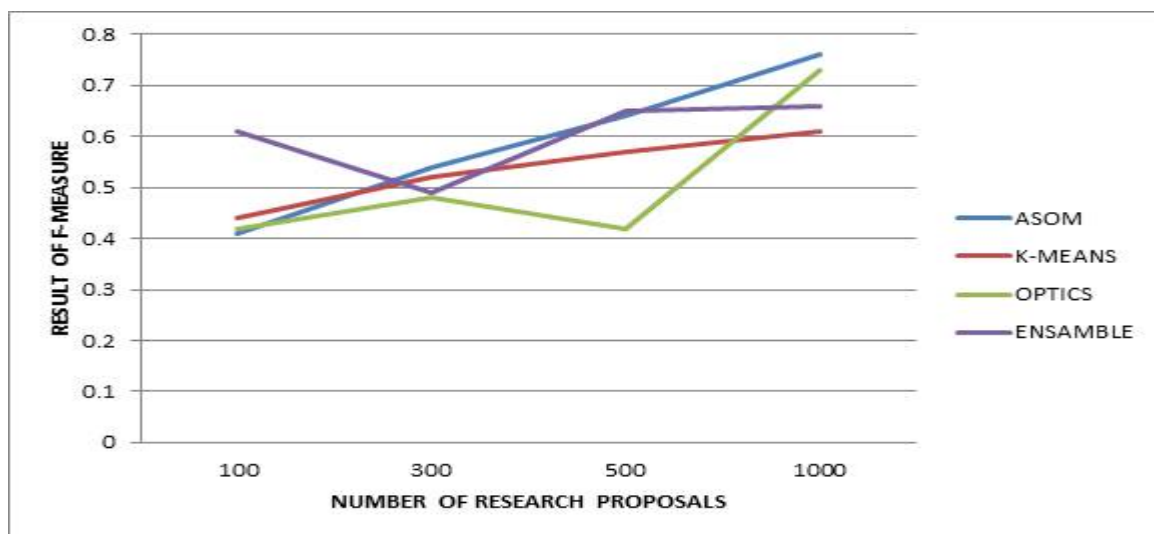


Figure 3 Performance of System

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

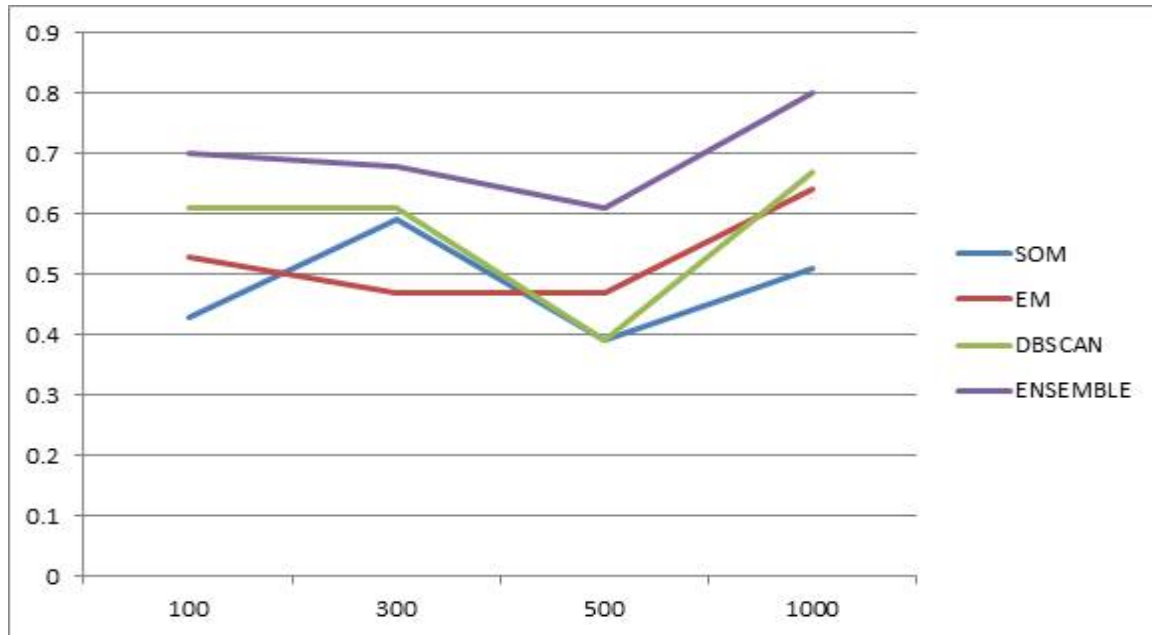


Figure 4 Performance of previous System

V. CONCLUSION AND FUTURE SCOPE

In present framework no any exterior analyst has been there for audit the venture recommendations so in recommended framework exterior commentator has been there for review the proposition. Additionally in recommended framework the algorithm utilized is being distinctive on the grounds that as a part of existing framework a few algorithms have confinement that could be overcome by these algorithms. In this system an ensemble-based strategy is utilized along with data mining for gathering of examination proposition in many foundation

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).

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

- [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>
- [13] Jaiwei Han and Micheline Kamber , "Data Mining concepts and Techniques" 2nd edition.
- [14] Soumi Ghosh, Sanjay Kumar Dubey, " Comparative Analysis of K-Means and Fuzzy C-Means Algorithms", (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, Vol. 4, No.4, 2013.
- [15] Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, Jörg Sander, " OPTICS: Ordering Points To Identify the Clustering Structure", Presented by Chris Mueller November 4, 2004.