Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 8, August 2015, pg.124 – 127

RESEARCH ARTICLE

Research Proposal Selection Using Clustering and Ontology Based Text Mining Method

Pravin Shinde[1], Sharvari Govilkar[2]

[1]Information Technology, Mumbai University, India

[2]Information Technology, Mumbai University, India

[1] spravinshinde@gmail.com; [2] sgovilkar@mes.ac.in

Abstract— Selection of research project is a very important process for government and for private research funding agency. When the agency or research institution received lots of research proposals then according to the similarity these can be grouped in research area and then assigned to the respective experts or guide for review. This approach of grouping the research proposal based on keywords and matching the similarities between research areas is done manually. However the exact research area of particular domain with respect to proposal cannot be accurate because subjective views and possible misinterpretation of applicants. Hence text mining methods are used to classify text documents which are having rich information.

This approach presents text mining methods to cluster the research proposals using ontology, which is based on research areas and similarities between them. For clustering the research proposals this approach is effective and efficient. The optimisation model is also used to balance and regroup the clusters of research proposals. The external reviewer is clustered according to their expertise and then these clustered proposals are assigned to respective reviewer.

Keywords— Clustering, Classification, Self Organizing Map, Text Mining, Optimization, Latent Semantic Indexing

I. INTRODUCTION

Text mining is also known as text analytics or text data mining. Text analytics is the process of extracting highly important and qualitative information from text and this happens by using trends and patterns of that text, for example statistical pattern learning. Generally text mining have the process of input text structuring, patterns are derived within this structured data, and output is generated by interpretation of these patterns. In this approach text mining methods can be implemented on ontology for research proposal selection which gives effective and optimum result.Selection of research projects plays very important role at research funding agencies in many private and government sector. It is a challenging task that involves multiple processes, like it starts with a funding agency by calling for proposals (CFP) and it is distributed to related institutions such as research institutions or universities. To funding agency the research proposals are submitted and they are assigned to respective experts for review. After collection of the review results these proposals are ranked.

In any research funding agency five to six reviewers are appointed to review each proposal so as to ensure reliable and accurate opinions on proposals. If large numbers of proposals are there to select, it is necessary to form a group of these proposals according to similarities or correlation in research disciplines and then these groups are assigned to relevant reviewers.

This is very time consuming process, hence to reduce the time text mining methods use classification of text documents approach. This approach cluster the research proposal using ontology based text mining methods, which is based on research area and their similarities. This approach for clustering the research proposals is

efficient and effective. The optimization model in this method considers applicant's characteristics which are used to balance the research proposals by geographical area, cluster the reviewers based on their expertise of research areas and assigns grouped research proposals to reviewers parallel. The results can also be used to enhance the effectiveness and efficiency of research proposal selection process at government or private research funding agencies.

II. **METHDOLOGY**

The proposed Ontology based Text Mining method (OTMM) is based on the statistical methods and optimizations model. This methodology is based on the following phases. In first phase, the research ontology is constructed using the keywords of research proposals. In second phase the proposals are classified using the sorting algorithm with the help of ontology. In next phase the classified proposals are clustered into groups with the help of self organizing map algorithm. After clustering the large size groups they are balanced using optimization model. Then this optimized clusters proposals are assigned to the respective experts. All this phases are explained below

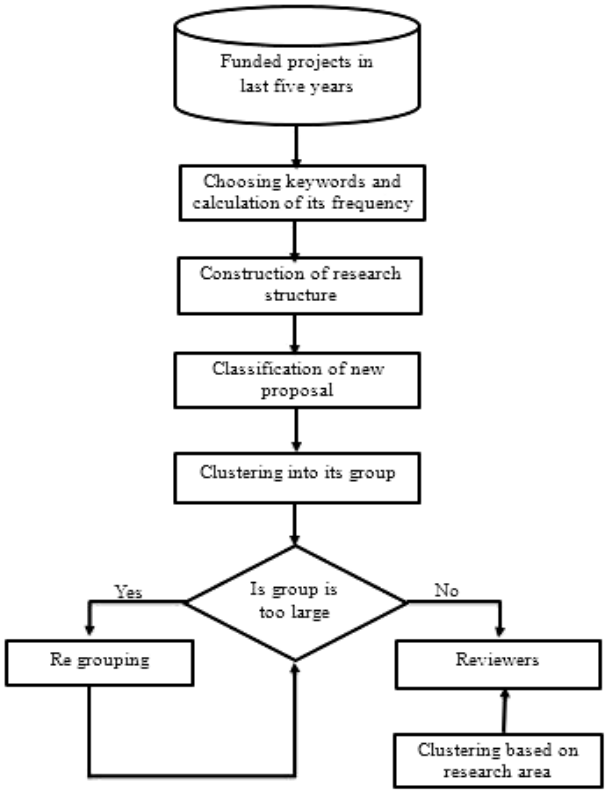

Fig. 2.1 Proposed Approach

2.1 Research ontology construction

Research ontology is a public concept set of research project management domain it considered as a domain specific ontology. The research paper of different discipline can be shown under this research ontology. This research ontology can be created by following steps.

2.1.1 *Creating the research topics of the discipline Ak.*

Keywords of the particular research proposals are collected and their frequency is counted. And this frequency and keywords create the feature set like (Nok, IDk, year, {(key1, freq1), (key2, freq2)… (keyk, freqk)}). Where Idk is discipline code and Nok is number of kth record.

2.1.2 Research ontology construction.

The research ontology is categorized and developed according to research area and then it is further divided into some narrower discipline area. Lastly it leads to research topics of the various disciplines by using the feature set, given at step 1.

2.1.3 Research ontology update.

The research ontology is updated on the yearly basis according change of feature set. The classification of research proposal and clustering can be done effectively and efficiently by the following way.

2.2 *Classification of new research proposals*

. In this phase a simple sorting algorithm is used for classification of research proposal in their appropriate discipline areas and it can be done by using research ontology. Here in this simple sorting algorithm, K is discipline areas where Ak and Pi denotes the area k and proposal respectively, Sk denote the set of proposal which having area k. by using this parameters the sorting algorithm can be implemented

2.3 Text mining clustering of research proposals

After classification the proposals are clustered in each discipline with text mining methods. The clustering approach contains few steps like text document collections, preprocessing, encoding, vector dimension reduction and text vector clustering. The details information of each field is given below

2.3.1 *Collection of text document.*

The research proposal is classified according to the discipline areas whereas the proposal documents in each discipline Ap(p =1,2,…, P). This proposal document is collected for preprocessing of text document.

2.3.2 *Preprocessing of text document.*

The text document is usually non structured inorder to analyse, extract and identify the keywords we are using domain specific research ontology. In this phase tokesisation can be carried out.

2.3.3 *Encoding of text document.*

In the preprocessing the documents are segmented, after that they are get transferred to featured vector representation V = (v1, v2, . . . , vM), where vi (i = 1, 2, . ., M) is the term frequency and inverse document frequency encoding of the keyword wi and M is the number of features selected.

2.3.4 *Vector dimension reduction*

This function is used for reducing the large size vector into smaller size by selecting most important keywords in terms of frequency. The latent semantic indexing is used to reduce the size of vector with preserving meaning.

2.3.5 *Text vector clustering*.

This method uses the Self Organizing Map (SOM) algorithm to cluster feature vector which is based on research area. The SOM is an unsupervised learning neural network model that is used for clustered the input data having similarities.

2.4 Balancing and regrouping of research proposals

After clustering the proposals, some of the clustered data needs to be balance and regrouped properly. May be few clustered group are very large because the number of proposals, so in order to balance them the applicants characteristics are considered like the affiliated universities or departments. The compositions of proposal at cluster should be diverse. Sometimes reviewer handled the proposals improperly (having same affiliation or department of the universities) which causes poor group composition. At this point the evaluation of proposals which have poor group composition may feel uncomfortable or confused for reviewers, so to avoid this, it is advisable to consider applicants characteristics to produce diversity at each proposal groups.

2.5 Assign to external reviewer.

The information retrieved by knowledge based agent is assigned to external reviewers where reviewer's research area, experience will be collected before. According to their research area and experience the reviewer will be clustered. But there may be some ambiguous occur because the reviewer may be specialized in more than one domain. For example, the reviewer will be specialized in data mining and network security areas. So while clustering the reviewer, their priority of research area is taken into consideration.

III. CONCLUSIONS

This proposed approach is using the ontology based text mining methods, where it is used for grouping the research proposal and assigning them to the reviewers systematically. The domain specific research ontology is created for concept term by categorizing them with respective discipline area and to form relation between them. This approach is used for clustering and balancing the research proposals by facilitating text mining and optimization technique, using similarities and applicant's characteristics. The propose approach can be used at government or private research funding agencies.

REFERENCES

[1] L. Razmerita, "An ontology-based framework for modeling user behavior—A case study in knowledge management," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 4, pp. 772–783, Jul. 2011.
[2] L. L. Machacha and P. Bhattacharya, "A fuzzy-logic-based approach to project selection," IEEE Trans. Eng. Manag., vol. 47, no. 1, pp. 65–73, Feb. 2000.
[3] A. J. C. Trappey, C. V. Trappey, F. C. Hsu, and D. W. Hsiao, "A fuzzy ontological knowledge document clustering methodology," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 3, pp. 806–814, Jun. 2009.
[4] Y. H. Sun, J. Ma, Z. P. Fan, and J. Wang, "A group decision support approach to evaluate experts for R&D project selection," *IEEE Trans. Eng. Manag.*, vol. 55, no. 1, pp. 158–170, Feb. 2008.
[5] A. D. Henriksen and A. J. Traynor, "A practical R&D project-selection scoring tool," IEEE Trans. Eng. Manag., vol. 46, no. 2, pp. 158–170, May 1999.
[6] L. M. Meade and A. Presley, "R&D project selection using the analytic network process," IEEE Trans. Eng. Manag., vol. 49, no. 1, pp. 59–66, Feb. 2002.
[7] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 586–600, May 2000.
[8] M. Nagy and M. Vargas-Vera, "Multiagent ontology mapping framework for the semantic web," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 4, pp. 693–704, Jul. 2011.
[9] C. Lu, X. Hu, and J. R. Park, "Exploiting the social tagging network for web clustering," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 5, pp. 840–852, Sep. 2011.
[10] http://www.nsfc.gov.cn/publish/portal1/ accessed on 02/01/15 at 2pm.
[11] http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/text.htm#DMCON011 accesed on 12/02/15 at 3pm.

