# DESIGN OF INTELLIGENT RECOMMENDATION SYSTEM FOR PROPER REVIEWER SELECTION

Hyung Jun Ahn, Hong Joo Lee, Han Seup Kim, Sung Joo Park
Graduate School of Management, KAIST
hjahn@isir.kaist.ac.kr, iceman@isir.kaist.ac.kr, kimhs@isir.kaist.ac.kr, sjpark@kaist.ac.kr

## ABSTRACT

Selection of proper reviewers for research proposals submitted for funding is crucial in determining the quality of evaluation. However, it has been found that the selection is not satisfactory due to outdated reviewer information, inadequate categorization of reviewers, and changes in research domain of reviewers. In this paper, to help this selection process, we propose an intelligent recommendation system based on information filtering and collaborative filtering techniques.

## I. INTRODUCTION

Most organizations that funds science researches use peer review process to evaluate initial proposals and final reports [2]. In a peer review process, a set of reviewers is selected from a pool of researchers and the selected reviewers go through the evaluation process with given evaluation criteria. Selecting proper reviewers is very important to guarantee successful evaluation and will lead to effective utilization of research funds. However, some survey researches have revealed that much of the dissatisfaction about review process is due to the inadequate selection of reviewers [2] [3] [5].

In most research funding organizations, they select predetermined number of reviewers from their reviewer pool database. In the reviewer pool, registered researchers are assigned with one or more categories of research area and those categories are used to match a proposal or report with a reviewer. The problem with category code method is that the construction of category codes is very knowledge-intensive work, and requires many experts in various research domains. With this reason, the category codes are very costly for management and not updated frequently, resulting in outdated categorization. The management of reviewer pool is also a manual task, and the collection of recent reviewer information is not easy. With this pool problem combined with the category problem, many reviewers may receive an obsolete request for review.

In this paper, we propose an intelligent recommendation system for reviewer selection based on information filtering and collaborative filtering techniques. With information filtering, key words from proposals and reports are extracted and used to link those documents to reviewers. With collaborative filtering, the similarities between researchers are calculated, and a recommendation is made based on the similarity.

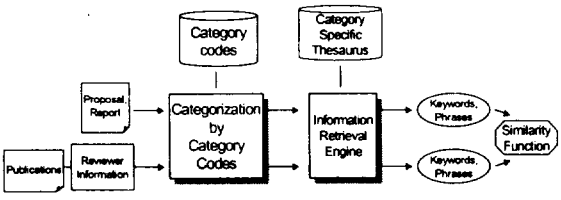
## II. RELATED RESEARCH

Recommendation method such as information filtering or collaborative filtering is used to select similar user for e-commerce shop and recommend

documents that are similar to past work and profile of researcher [7]. Information filtering matches items to the individual user's classification scheme [1]. Information filtering has its root in information retrieval (IR) researches and employs many of the same techniques. Text documents are recommended based on a comparison between their content and a user profile [4]. Weighting scheme, also called relevance feedback [4], and Vector space model that represents a document as the sum of its term vectors are used for information filtering approach. Collaborative filtering makes recommendations based on correlation between personal preferences [6][8]. For each user, a set of "nearest neighbor" users is found based on past ratings [4]. Collaborative filtering uses many statistical analyses, such as Pearson correlation, Spearman rank correlation and variance weighting, to compute the similarity of the users. Many other techniques are used to make recommendations, but these two methods are basic ways and many of the recommendation systems use a combination of two or more techniques.

## III. RECOMMENDATION SYSTEM

### 3-1. Overview




[Figure 1] Recommendation system overview

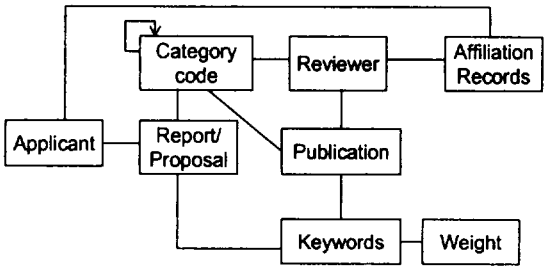
The overall architecture of the recommendation system is given in Figure 1. First, proposals and reports for evaluation are categorized by traditional category codes, and using information retrieval engine with domain specific thesaurus, keywords and phrases are extracted. At the reviewer side, reviewer information including their publications are also categorized and indexed with key words and phrases. The reviewer information along with category code and key words constitute the reviewer pool database that is used persistently afterward. When finding proper reviewers for a specific proposal or report, these keywords and phrases are used to calculate similarity values using a predefined function.

The question can be raised here about why we used traditional category codes together with keyword based recommendation. First, the keyword-based recommendation is based on a similarity function that uses statistics of the keywords, and does not fully represent the semantic contents of documents. For this reason, this approach may always produce erroneous recommendation. Second, by pre-dividing the documents into categories, we can use domain-specific thesaurus for keyword extraction, which enables much higher quality in extraction.

### 3-2. Information filtering

In our recommendation system, we used keyword-based similarity function. In this function, we used frequency statistics to calculate the similarity between a reviewer and a proposal or a report. However, the filtering process has to take into consideration more factors such as category codes, outdated reviewer information, and affiliation record. The data model illustrated in Figure 2 incorporates these requirements for filtering. Reviewers and applicants are associated with their affiliation record such as their institute or university of both current and past situation. These affiliation records are used to prevent recommending a reviewer whose affiliation records are close to the applicant. And the

category codes are assigned to reviewers, their publications, and applicants' reports and proposals.




[Figure 2] Data model used for information filtering

The similarity function is depicted in the following figure. This function is based on vector space representation of reviewer and document profiles. *Pr* represents reviewer profile and *Pd* represents document profile. Both profiles consist of keyword and weight pairs and the weight is normalized such that the total sum of weights makes 1 for each profile. Then the similarity between a reviewer and a document, $S(R_i,D_j)$, is calculated by summing up weight values in *Pr* multiplied by respective importance of each keywords in *Dr*. This implies that the researcher profile is transformed to a similarity value by summing up the contribution of each keyword in the view point of document profile. The weight is calculated by TFIDF scheme [1], which is widely used in information retrieval systems. The similarity between reviewers is calculated in the same manner and used for collaborative filtering.

$Pr_i = \{ (k_1,w_{i1}), (k_2,w_{i2}), \dots \}$

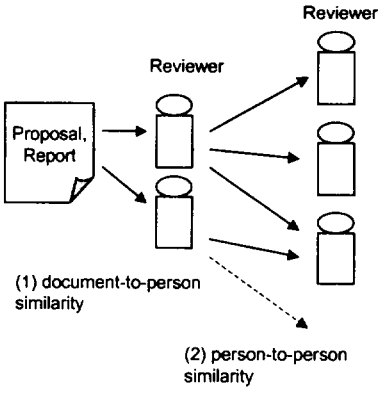
$Pd_j = \{ (k_1,w_{j1}), (k_2,w_{j2}), \dots \}$

$S(R_i, D_j) = \sum w_{ik} * w_{jk}$ for every keyword $k_k$ contained in $Pd_j$, where $(k_k,w_{jk}) \in Pd_j$ and $(k_k, w_{ik}) \in Pr_i$, and if $(k_k, w_{ik}) \notin Pr_i$, we assume $w_{ik} = 0$.

[Figure 3] Similarity function

### 3-3. Collaborative filtering

When there are enough reviewers and their up-to-date information is present in the reviewer pool, information filtering may work well and produce good results in recommendation. However, managing a reviewer pool database to a quality where above conditions are met is not feasible practically in many cases. For this reason, we used collaborative filtering model to make recommendation for such cases with too few candidate reviewers or too poor values of similarity value from selected reviewers. This mechanism is illustrated in Figure 4.




[Figure 4] Collaborative filtering

At the first stage, a document-to-person similarity function is used to find most relevant reviewers, and the search is expanded using person-to-person similarity to find more reviewers.

### 3-4. Enhancing recommendation quality by time-variant weight

Researchers usually expand or change their research area as time goes on. For this reason, an outdated researcher's profile may produce poor recommendation result. To overcome this problem, we used a weighting policy that decreases as time

gap increases between the publication date and current date. For example, assume we apply a 5-year lifecycle for a publication and a linear function for decreasing weights. In this case, keywords extracted from a publication published 5 years ago receive 1/5 weight, and those published 1 year ago receive 4/5 weight. However, depending on the nature of research, the lifecycle must be applied differently. For a research based on a mature technology, the lifecycle must be longer, and in the opposite case, it can be shorter. Also, we may apply different functions to calculate weights, but this needs further research, which is out of the scope of this paper.

## IV. CONCLUSION

We proposed a recommendation system, based on information filtering and collaborative filtering techniques, to help select more relevant reviewers for submitted proposals or reports. The design of data models, similarity function, and filtering mechanisms were introduced.

For further research, we plan to build a prototype system based on the design, gather experiment samples, and test recommendation.

## REFERENCES

[1] Delgado, J., Ishii N. and Ura, T., Content-based Collaborative Information Filtering: Actively Learning to Classify and Recommend Document, in: Matthias Klush, Gerhard Weiss, eds., Cooperative Agents II, Proceedings/CIA'98, pp. 206-215, 1998

[2] Ki-Yong Om, A Study on the Peer Review Processes: Focused on Korea Science and Engineering Foundation(KOSEF), Doctoral Thesis of KAIST in Korean, 1998

[3] Lee Mushin, Son Byoungho and Om Kiyong, Evaluation of National R&D Projects in Korea, Research Policy, Elsevier, Vol. 25, No. 5, pp. 805-818 (1996)

[4] M. Balabanovic, Y. Shoham, Fab: Content-Based, Collaborative Recommendation, Communication of the ACM, 40(3), pp. 66-72, 1997.

[5] McCullough, J., First Comprehensive Survey of NSF Applicants Focuses on their Concerns about Proposal Review, Science, Technology and Human, Vol. 14, No.1, 1989, pp.379-388

[6] N. Glance, D. Arregui and M. Dardenne, Knowledge Pump: Supporting the Flow and Use of Knowledge, Information Technology for Knowledge Management, Eds. U. Borghoff and R. Pareschi, Springer-Verlag, 1998.

[7] Paul Resnic, Hal R. Varian, Introduction (to the special section on recommender systems), Communication of the ACM, 40(3), pp. 56-59, 1997.

[8] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, GroupLens: An Open Architecture for Collaborative Filtering of Netnews, In Proceeding of the ACM Conference on Computer-Supported Cooperative Work (Chapel Hill, NC) 1994.