Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

Recommender System for Academic Literature with Incremental Dataset

Mahak Dhanda and Vijay Verma*

National Institute of Technology Kurukshetra, Kurukshetra 136 119, India

Abstract

On account of the colossal expansion in the size of research paper repository, the stature of Recommender System has increased, as it can guide the researchers to find papers akin to them from this vast collection. Furthermore, the recommendation methods like collaborative-filtering or content-based do not allow the user's to provide their personalized requirements explicitly; hence the focus is shifted towards the customized Recommender Systems that can scrutinize user's preferences by contemplating their inputs. But the state-of-art recommendation techniques satisfying user's personalized requirements make a strong assumption of static dataset. So, in this work we are going to present a customized Recommender System that can acknowledge the ever growing nature of research paper repository. To accomplish this, the Efficient Incremental High-Utility Itemset Mining algorithm (EIHI), which has been recently introduced in the literature, is used which is specialized to work with dynamic datasets. Experimental results prove that the proposed system satisfies the researcher's personalized requirements and at the same time handles the incremental nature of the research paper repository efficiently.

Keywords: High-Utility Itemset Mining (HUIM); Recommender Systems (RS); Utility-based Recommendation.

1. Introduction

The massive hike in the extent of information over the web has turned it arduous for the users to search for the information admissible to them. Recommender System (RS) has come out as a revolutionary concept to ride out through this situation[14]. It is a tool (software) that provides the users with the suggestions of information that may be useful to them. These suggestions may turn out helpful to the users in many scenarios where decision making is involved ex. selecting books to read, movies to watch etc. A lot of techniques are available for recommendation which are majorly categorized as collaborative filtering[7] and content-based filtering[5]. Collaborative filtering works on the concept of finding out similar users so as to make recommendations, while content-based techniques work on the basis of similarity in features of the item and the user.

RSs have gained a lot of emphasis in the commercial environment[9] and along with it they have proven to have a crucial role in academic literature domain[11] also. As a result of continuous growth in the size of research paper repository (because of lot of papers coming out of journals and conferences each year), it has turned out to be

*Corresponding author. Tel.: +91-8901370565.
E-mail address: mahak0570@gmail.com

troublesome task for the researchers to get papers akin to them. To endure this problem, recommender systems have become necessary in the academic literature domain[11]. Also, as different researchers may have their own personal requirements; customized recommender systems are drawing researcher's attention[3], where users can provide their specific preferences before the recommendation is provided to them. But still the major complication resides due to the dynamic nature of research paper database.

The recommendation technique[3] considering user's preferences, make a stronger assumption about the nature of database to be static. Here, through this work we are going to propose a new customized RS for research paper which can incorporate the dynamic aspect of the dataset. The size of research paper repository may change because of new papers getting added to it time to time and this yearns for a recommender system that can support the dynamic dataset. So, the EIHI algorithm[1] is utilized in our proposed approach, as it is compatible with datasets which are dynamic in nature.

The proposed technique works in two stages 1) first stage is used to filter out the papers pertinent to the topic of interest of the user based on their content; 2) second stage makes sure that the personalized requirements of the user are satisfied (on the basis of usability assigned to the papers with the help of input taken from the user).

The rest of the article is organized as: Section 2 covers Preliminaries for the approach; Section 3 describes the Related work, Section 4 and 5 covers the Proposed approach and Simulation results respectively; and Conclusion is added as Section 6.

2. Preliminaries

EIHI[1] is a High Utility Itemset Mining (HUIM) algorithm, designed to work in case of the incremental datasets. This technique can perform well in case of such dataset because whenever updation is made in the dataset, it can compute the new High Utility Itemsets (HUIs)[2] only by scanning new transactions and previous HUIs (without starting the process again from scratch).

EIHI works by calculating Transactional weighted Utility (TWU)[2] and making utility lists and Estimated Utility Co-occurrence Structure (EUCS)[13] only for the items present in newly added transactions and also by exploring the extensions of only those items[1]. As EIHI is tree-based algorithm it becomes very easy to add new HUIs to the tree as well as to make updations to utilities of existing HUIs[1].

As the academic literature repository is incremental (as new papers may get added to the repository from time to time) and it's not possible to scan the complete repository again and again. Hence we take the advantage of concepts used in EIHI to handle the dynamic nature of this repository. Our proposed approach uses EIHI to mine the High Utility Reference-sets (HURs)[3] and recommend them to the users.

3. Related Work

Recommender Systems have become an important part of academic literature domain because of the ever growing size of research paper repository[16]. Research paper recommender systems are mainly influenced either by the frequency of citations of a paper or by the pertinence of its contents to the user, while making recommendations[11].

The approaches based on frequency of citation may face *cold start* problem and hence recommender system based on Belief-Propagation has been proposed[10]. In order to further improve the quality of recommendation some approaches using semantic data for recommendation have been created[15]. Research paper recommender system supporting diversity has also been developed[6] which uses the concepts of co-authors and dissimilar users to make recommendations.

Also as different researchers may have their own personalized requirements in terms of publishing dates etc., recommender systems based on user's recent interests are gaining importance. To find user's interests, recommender systems may either use the author's published work[4] or can be customized to take input from the user before recommending papers[3]. But the recommendation technique proposed in[3] suffers from dearth of efficiency because it relies on Two-phase algorithm[2] which does not contemplate the incremental nature of research paper repository. To overcome this limitation we are proposing EIHI based recommendation approach having support for dynamic datasets.