### **ARTICLE IN PRESS**

[m3Gsc;April 17, 2017;20:31]

Computers and Electrical Engineering 000 (2017) 1-10



Contents lists available at ScienceDirect

### Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

# Gene-based Collaborative Filtering using recommender system $\!\!\!\!\!\!^{\bigstar}$

Jinyu Hu<sup>a,b,\*</sup>, Sugam Sharma<sup>c</sup>, Zhiwei Gao<sup>d,\*\*</sup>, Victor Chang<sup>e</sup>

<sup>a</sup> Division of Immunology and Rheumatology, Department of Medicine, Stanford University, USA

<sup>b</sup> Department of Medicine, Veterans Affairs Palo Alto Health Care System, CA, USA

<sup>c</sup> Center for Survey Statistics & Methodology, Iowa State University, IA, USA

<sup>d</sup> Faculty of Engineering and Environment, University of Northumbria, Newcastle upon Tyne, NE1 8ST, UK

<sup>e</sup>Xi'an Jiaotong-Liverpool University, Suzhou, China

### ARTICLE INFO

Article history: Received 29 December 2016 Revised 12 April 2017 Accepted 12 April 2017 Available online xxx

Keywords: Big Data GeneCF GPC HCC Recommender systems Collaborative filtering

### ABSTRACT

The recommender system (RS) has achieved substantial evolution in this information age of the twenty-first century, with no exception to biological domain. While RS has been effectively exploited in analysis of biological data for gene prediction, it has raised interesting research challenges such as how to explore the gene interest (Gi) and recommend the genes for individual patients. To meet these research challenges, we propose a novel TOP-N Gene-based Collaborative Filtering (GeneCF) algorithm based on Gi of patients. The GeneCF algorithm is aimed for matching more accurate recommendations about genes to the patients, with exceptional precision and coverage achieved. The GeneCF algorithm has been tested and evaluated on a hepatocellular carcinoma (HCC) gene expression database. We found that six genes could be the cause of liver cancer: AMP, SAA1, S100P, SPP1 and CY2A7 and AFP. The GeneCF algorithm contributes to help doctors provide smarter, customized care for cancer patients.

© 2017 Elsevier Ltd. All rights reserved.

### 1. Introduction

In the Big Data era today, the Recommendation Systems RS(s) [1] have been envisioned to acquire a huge positive impact in improving the quality of our lives. In the business and consumer world over the last recent years, the RS(s) have attained immense popularity in information science. The increasing maturation of the information space, primarily world-wide-web, and the swift development and evolution of e-commerce space have accelerated the development of RS(s). The RS abstraction can be interpreted as a technology to filtering the individualize details about a particular user in order to distinguish a set of additional items, predicted based on that user's interests. Significant research efforts are still continuing in this direction under the purview of information retrieval and filtering. The evolution of data in the form of Big Data [2,3] that potentially encompasses the biological data, has further driven the RS-oriented research to the next level in this data science age. The data-richness of the associated environment promises better efficiency of RS. Therefore, RS is an effective and extensive approach to study the complexity of science and the complex data. In RS context, the Big Data consists of brows-

\* Correspondence author at: Department of Medicine, Veterans Affairs Palo Alto Health Care System, CA, USA.

\*\* Corresponding author.

http://dx.doi.org/10.1016/j.compeleceng.2017.04.010 0045-7906/© 2017 Elsevier Ltd. All rights reserved.

Please cite this article as: J. Hu et al., Gene-based Collaborative Filtering using recommender system, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.04.010

<sup>\*</sup> Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. M. Karuppiah.

E-mail addresses: jyhu01@gmail.com (J. Hu), sugam.k.sharma@gmail.com (S. Sharma), zhiwei.gao@northumbria.ac.uk (Z. Gao).

2

### ARTICLE IN PRESS

#### J. Hu et al./Computers and Electrical Engineering 000 (2017) 1-10

ing and purchasing history, and other related information are capable of providing the rich feedback to RS(s) that enable them to disseminate the effective, relevant, highly accurate recommendations [3]. In the information science literature, the RS(s) are considered as subclass of the filtering systems that help predict the preference and rating of related items to the users during the purchasing process and aid them in meeting the appropriate commodities [4]. For state-of-the-art RS(s), the collaborative filtering (CF) is evidently the central algorithm [5], which predominantly relies upon the recommended tag users, recommended resources or recommended tags. The user-relied CF has hitherto been witnessed as the highly reliable, most robust and widely used algorithm to develop the efficient commercial RS(s). A concise set of items is intended to be recommended, but the extended set of recommended items is actually expected and required by the users. The research challenge then arises - what should be the appropriate number of items to be recommended to the used? To address this concern, Hu et al. [6] devise a method, called Top-N recommender and illustrate that with examples.

For CF, the calculation algorithm serves as the potential constituent and the key step and these algorithms help perform the similarity calculation; the Pearson correlation coefficient, cosine similarity and correction cosine similarity are some of the general methods that may be considered. Hu et al. [7,8] explore some correlations that can be effectively used to develop the single mode social network. And, it is discovered that a similarity algorithm, based on the single and dual mode social network is simple, effective and visible [9] to structure the gene's relationship and is used in our work. In this research, we exploit the TOP-N recommender algorithm on biological data and develop a method that helps investigate, how many genes should be recommended to the doctor to achieve maximum precision and coverage in the process of diagnosing the cancer. The proposed method recommends N genes as target and we dub this novel method as Gene-based Collaborative Filtering (GeneCF).

There are several general indicators that are used to evaluate the correctness, effectiveness, accuracy and other similar attributes of the algorithm. Furthermore, the precision, recall, average absolute error, average accuracy rate, AUC (precision-recall curve) and P@n (precision at n) are inclusive in scientific measurement [10]. Additional parameters such as coverage, diversity, and the average popularity can also be taken into consideration [11]. In the context of this work, for the biological gene (Big) data, we propose some interesting indicators- Gene Precision (GP), Gene Penalty Precision (GPP), Gene Coverage (GC) and Gene Penalty Coverage (GPC). We compare their RNA expression of the healthy samples and samples with liver cancer to investigate the recommended genes. To validate our approach, the genes and samples from the cancer staining of patients are used to verify our scientific results.

The rest of the paper is structured as follows. Section 2 consists of the freshly developed related work. Section 3 contains the research methods used in this work. Section 4 explores the experimentation and results, which is followed by the discussion in Section 5. Finally, the paper is concluded in Section 6.

### 2. Related work

#### 2.1. Recommender system and target genes

Recommended system (RS) seeks to provide the users with a related information based on the currently browsing records. It is very ubiquitous in the e-commerce sites such as Amazon, and each page contains a list of items viewed or purchased by other users. According to the purchase history, Amazon also can recommend some other products to the customers resulting in additional purchase due to the strong influence of the recommender system. The RS is positive to lead to a revolution in the browsing and mining biological Big Data. The same principle would provide users and patients with a list of recommended biological or medical objects, thus, the RS could be implemented and used in the biology field. For example, it can create the networks for the immunity research [12], and also can work for the metabolic pathways [13].

Collaborative filtering (CF) is the main technique used in the recommender systems, which is based on a scientific computing repository, such as browsing record, and bioinformatics data [14]. Additionally, the CF is used to build a QSAQ model for biological targets and multiple cell lines [15]. Other biology data study, such as protein-protein interactome relationship [16,17], also can be analyzed by using the CF approach.

Motivated by the success of collaborative filtering (CF)-based approaches to the problem of personalized-recommendation on large, sparse rating matrices, this research work aims at implementing a highly efficient CF-based approach for gene interactome mapping (Fig. 1). To achieve this, we first propose a CF framework for it and under this framework, we model the given data into an interactome weight matrix by using social networks, where the feature-vectors of involved genes are extracted. With them, we design the coefficient to model the inter-neighborhood similarity among the genes involved, for taking the mapping process. However, there are some differences between the RS(s) of traditional business versus biological gene domains (Fig. 2). Generally, the traditional collaborative filtering algorithm is based on the resource recommendation. The Top-N algorithm [5] provides a movie example for its background; if a user watches one movie, it is impossible to recommend this movie to the same user again. In other words, only the missed movie can be recommended to the users. For the case of the biological gene data, whether the gene has been missed out for investigation, we only focus on the genes that highly match to CF. In this paper, the Top-N Gene-Based Collaborative Filtering algorithm (GeneCF) was proposed.

Please cite this article as: J. Hu et al., Gene-based Collaborative Filtering using recommender system, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.04.010

Download English Version:

## https://daneshyari.com/en/article/6883576

Download Persian Version:

https://daneshyari.com/article/6883576

Daneshyari.com