

Chapter 6

A Survey on Data Mining Techniques in Research Paper Recommender Systems

Benard Magara Maake

Tshwane University of Technology, South Africa

Sunday O. Ojo

Tshwane University of Technology, South Africa

Tranos Zuva

Vaal University of Technology, South Africa

ABSTRACT

In this chapter, the authors give an overview of the main data mining techniques that are utilized in the context of research paper recommender systems. These techniques refer to mathematical models and tools that are utilized in discovering patterns in data. Data mining is a term used to describe a collection of techniques that infer recommendation rules and build models from research paper datasets. The authors briefly describe how research paper recommender systems' data is processed, analyzed, and then, finally, interpreted using these techniques. They review different distance measures, sampling techniques, and dimensionality reduction methods employed in computing research paper recommendations. They also review the various clustering, classification, and association rule-mining methods employed to mine for hidden information. Finally, they highlight the major data mining issues that are affecting research paper recommender systems.

DOI: 10.4018/978-1-5225-8437-7.ch006

1. INTRODUCTION

Recommender systems are lately gaining significant roles in information filtering search. In the field of research paper recommender systems, various data mining techniques have been utilized to perform various tasks. This chapter intends to highlight the use of data mining and associated methods that have been used in research paper recommendation. We partly adopt the data mining steps and methods for recommender systems as highlighted by (Amatriain, Jaimes, Oliver, & Pujol, 2011) in the recommender systems handbook by (Ricci, Rokach, & Shapira, 2011) to represent the various data mining methods and technologies that were employed at various levels of computing research paper recommendations. Data mining in this context consists of three main steps namely: Data preprocessing stage, Data analysis stage and the Result interpretation stage. We may not have a crisp separation and categorization of some of the methods and algorithms since most of them overlap.

This review chapter is organized according to the following sections: The chapter introduction and overview is presented in Section 1. A summary of data preprocessing methods and measures as utilized in research paper recommender systems is presented in Section 2. Classification algorithms utilized by research paper recommender systems are highlighted in Section 3. Section 4 presents clustering algorithms, while Section 5 presents other approaches to classification. Section 6 presents the main data mining issues facing research paper recommendation, whereas Section 7 concludes the chapter.

Figure 1 highlights data mining features, approaches, and processes utilized in research paper recommender systems (RPRS). It represents the three main data mining steps which are consecutively applied during the processing of data, and they include data preprocessing step, data analysis step and finally, the results interpretation step. This chapter, however, dwell much on the first two steps, data preprocessing and data analysis steps since they actively utilize various data mining techniques.

2. DATA PREPROCESSING IN RPRS

Data preprocessing is an important step in machine learning and information retrieval because it screens data for any problems to prevent the possibility of producing misleading results after the processing process. Real-world datasets in the field of RPRS were generally incomplete (Gupta & Varma, 2017), noisy (Bogers & Van den Bosch, 2008; Bollen & Van de Sompel, 2006; Dong, Tokarchuk, & Ma, 2009; J. He, Nie, Lu, & Zhao, 2012; Y. Liang, Li, & Qian, 2011; McNee et al., 2002; Torres, McNee, Abel, Konstan, & Riedl, 2004; Tran, Huynh, & Hoang, 2015; Wu, Hua, Li,

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/a-survey-on-data-mining-techniques-in-research-paper-recommender-systems/232427

Related Content

Skills of Library and Information Science Professionals in MOOCs Environment

Anna Kaushik (2018). *Library and Information Science in the Age of MOOCs* (pp. 205-219).

www.irma-international.org/chapter/skills-of-library-and-information-science-professionals-in-moocs-environment/204195

Extending TAM to Understand Library User Acceptance of E-Books in Tanzania

Daniel Ntabagi Koloseni, Herman Mandariand Vincent T. Msonge (2021). *International Journal of Library and Information Services* (pp. 46-63).

www.irma-international.org/article/extending-tam-to-understand-library-user-acceptance-of-e-books-in-tanzania/277425

Development and Practice of Research Support Services in Peking University Library

Yong Tangand Chunhong Zhang (2019). *International Journal of Library and Information Services* (pp. 22-39).

www.irma-international.org/article/development-and-practice-of-research-support-services-in-peking-university-library/228176

Afar Print: Future Reading in Digital Medias

R. Natarajan (2015). *Handbook of Research on Inventive Digital Tools for Collection Management and Development in Modern Libraries* (pp. 327-339).

www.irma-international.org/chapter/afar-print/133972

Absorptive Capacity and Smart Library

Adebowale Jeremy Adetayo (2022). *Handbook of Research on Emerging Trends and Technologies in Librarianship* (pp. 310-326).

www.irma-international.org/chapter/absorptive-capacity-and-smart-library/295204