Information Processing and Management 54 (2018) 529–544

Contents lists available at ScienceDirect

Information Processing and Management

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

Social network data to alleviate cold-start in recommender system: A systematic review

Lesly Alejandra Gonzalez Camacho*, Solange Nice Alves-Souza

Departamento de Engenharia de Computação e Sistemas Digitais, Escola Politécnica da Universidade de São Paulo (EPUSP), Av. Prof. Luciano Gualberto, Travessa 3, 158 - Butantã, São Paulo - SP, 05508-010, Brazil

ARTICLE INFO

Keywords:
Cold start
Social network
Collaborative filtering
Recommender system
Systematic literature review

ABSTRACT

Recommender Systems are currently highly relevant for helping users deal with the information overload they suffer from the large volume of data on the web, and automatically suggest the most appropriate items that meet users needs. However, in cases in which a user is new to Recommender System, the system cannot recommend items that are relevant to her/him because of lack of previous information about the user and/or the user-item rating history that helps to determine the users preferences. This problem is known as cold-start, which remains open because it does not have a final solution. Social networks have been employed as a good source of information to determine users preferences to mitigate the cold-start problem. This paper presents the results of a Systematic Literature Review on Collaborative Filtering-based Recommender System that uses social network data to mitigate the cold-start problem. This Systematic Literature Review compiled the papers published between 2011–2017, to select the most recent studies in the area. Each selected paper was evaluated and classified according to the depth which social networks used to mitigate the cold-start problem. The final results show that there are several publications that use the information of the social networks within the Recommender System; however, few research papers currently use this data to mitigate the cold-start problem.

1. Introduction

In the last few years, the number of services and users have rapidly increased, partly due to the popularity of cloud computing services. By offering a flexible computational architecture, they allow having more and more industrial modern services (Deng, Huang, & Xu, 2014). The prevalence of social network services, mobile devices and large-scale service-oriented systems has produced a vast volume of data. This quantity of data overloads the users with information and hinders finding services that can accomplish their functional requirements (Almohsen & Al-Jobori, 2015; Deng et al., 2014). In this context, Recommender Systems (RSs) have an increasingly important role in helping users to overcome data or service overloading situations, suggesting the most suitable data/service automatically. This happens, for example, in e-commerce services, in which the main idea is to recommend not already acquired items which may be considered relevant to users (Deng et al., 2014; Maniktala, Sachdev, Bansal, & Susan, 2016). Examples in which RSs have been used include movies (Christou, Amolochitis, & Tan, 2016; Moreno, Segrera, López, Muñoz, & Sánchez, 2016), venues of interest (Khalid, Khan, Khan, & Zomaya, 2014; Yin, Cui, Sun, Hu, & Chen, 2014), television programs (Zhang, Chen, & Yin, 2013), online news (Lin, Xie, Guan, Li, & Li, 2014), virtual study groups (Salehi, Nakhai Kamalabadi, &

* Corresponding author at: Departamento de Engenharia de Computação, EPUSP, Av. Prof. Luciano Gualberto, Travessa 3, 158 - Butantã, São Paulo - SP, 05508-010, Brazil.

E-mail addresses: alejandragc@usp.br (L.A. Gonzalez Camacho), ssouza@usp.br (S.N. Alves-Souza).

https://doi.org/10.1016/j.ipm.2018.03.004
Received 5 September 2017; Received in revised form 23 January 2018; Accepted 9 March 2018

Ghaznavi Ghoushchi, 2013) and others. RSs use diverse information collected from the user to identify her/his preferences, as well as demographic information such as gender, age, location, etc. to propose new items that may be useful to her/him. For that, the Recommender System (RS) processes a large volume of data to make more adequate suggestions to users.

Despite the advances in RSs, some problems such as *data sparsity* and *cold-start* still remain open, because there is not a solution that meets the different needs (Al-Hassan, Lu, & Lu, 2015; Bobadilla, Ortega, Hernando, & Bernal, 2012; Huang, Chen, & Chen, 2016). Data sparsity is characterized by a low number of ratings for available items, making it difficult to find a relationship between users and items (Khalid et al., 2014). Cold-start is caused by the lack of both user data and item rating history, which are used as a mechanism to infer users preferences and perform the recommendation (Khalid et al., 2014; Sun, Wang, Cheng, & Fu, 2015). In this paper, we are interested in evaluating only the contributions related to the cold-start problem.

There is an active line of research to solve the difficulties associated with the above problems, and a variety of techniques have been proposed, such as the use of machine learning methods, approximation theory and various heuristics applied to different areas of recommendation (Al-Hassan et al., 2015; Alhamid, Rawashdeh, Dong, Hossain, & Saddik, 2016; Bobadilla et al., 2012; De Campos, Fernández-Luna, Huete, & Rueda-Morales, 2010; Katakis, Tsapatsoulis, Mendez, Triga, & Djouvas, 2014; Khalid et al., 2014; Xu, Fu, & Gu, 2016). In each of these approaches, different sources are explored to increase information about users - items to reduce data sparsity and to mitigate the cold-start problem (Barjasteh, Forsati, Masrour, Esfahanian, & Radha, 2015; Sun et al., 2015).

Social networks are sources of information that provide valuable data to establish the users preferences regarding items, besides identifying relationships of trust between users and the influence of one user on others. All such information can be extremely useful to make recommendations to users more accurately and objectively (Deng et al., 2014; Derczynski et al., 2015; Jiang et al., 2015), helping to alleviate the cold-start problem.

The Systematic Literature Review(SLR) presented herein was motivated by the results from previous works (Prando, 2016; Prando, Contratres, Alves-Souza, & deSouza, 2017), which showed that social networks data is a good information source for RSs to mitigate the cold-start issue.

Some filtering techniques used in the recommendation process in RSs are Collaborative filtering-based(CF), Content-based(CB) (De Campos et al., 2010; Huang et al., 2016), demographic and hybrid. CF-RSs identify groups of people with similar preferences to target user and recommend items these people like. Conversely, CB-RSs determine users' preferences and recommend similar items users previously preferred or browsed. Demographic Filter is based on the principle that users with certain common attributes, such as gender, age, education level, among others, also have common preferences. Hybrid Filter combines two or more filtering approaches to process recommendations (Bobadilla, Ortega, Hernando, & Gutiérrez, 2013; Jain, Grover, Thakur, & Choudhary, 2015; Kaššák, Kompan, & Bieliková, 2016).

While CB-RSs consider the behavior of previous users for recommendation (Huang et al., 2016), CF-RSs explore the assessments of similar users to the target user, increasing the amount of information for the recommendation process. Thus, the CF technique has been the most commonly approach used in RSs.

This paper introduces a SLR of researches, published in the last seven years, about CF-RSs that use data from social networks as main sources of information to alleviate cold-start problems.

This paper is organized as follows. Section 2 presents the main aspects of previous works in the literature review regarding correlate themes. Section 3 details the process followed to carry out the SLR here presented. Section 4 shows the SLR results and analyzes them. Section 5 introduces the main aspects of the papers classified as relevant to the SLR. Finally, the conclusion is shown in Section 6.

2. Related works

Themes on RSs that use social networks as an information source for the recommendation process are quite current, given the large number of papers in recent years (Alhamid et al., 2016; Christou et al., 2016; Katakis et al., 2014; Lim & Finkelstein, 2012; Maniktala et al., 2016; Xu et al., 2016). Moreover, literature review papers have been presented to evidence the interest in these subjects.

Rastogi and Singh (2016) presented an overview of RSs that use social network data to improve the quality of recommendations. This paper proposed categorizing socio-contextual information of social networks into explicit and implicit user-item information and analyzes the main aspects of generic RSs. However, the research was not a SLR because it did not show the steps followed for compiling articles, such as keyword definition, inclusion and exclusion criteria, and databases searched. In addition to that, it was not the focus of the research to identify articles that address the cold-start problem.

Aznoli and Navimipour (2017) made an SLR about the mechanisms of RSs used in cloud computing. In this research, RS filtering techniques were classified into four main categories: collaborative, demographic-based, knowledge-based and hybrid filtering. It also presented a comparison of these techniques in terms of scalability, availability, accuracy, and trust attributes. However, this research neither considered the cold-start problem nor the use of social network information in RSs.

Rahayu et al. (2017) presented a systematic review on RSs for the e-Portfolio domain, classifying articles according to the type of recommendation (Personalization or Business) and techniques employed. However, this review did not specifically research articles that mitigated the cold-start problem or made use of social networks. The authors proposed associate weights with questions to evaluate the selected articles. This strategy was also adopted by our SLR.

The main contribution of the SLR here presented is the collection of papers that address the use of social network data to mitigate the cold-start problem in RSs. Thus, this SLR is specific and, to the best of our knowledge, it differs from other reviews and briefly allows to show the trends of RSs in the adoption of social information as an important strategy to improve the quality of the recommendations.