Information Processing and Management 56 (2019) 192–211

Contents lists available at ScienceDirect

Information Processing and Management

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

Building user profiles based on sequences for content and collaborative filtering

Pablo Sánchez*, Alejandro Bellogín

Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain

ARTICLE INFO

Keywords:
Hybrid recommender systems
Preference filtering
Content-based filtering
Collaborative filtering
Longest Common Subsequence

ABSTRACT

Modeling user profiles is a necessary step for most information filtering systems – such as recommender systems – to provide personalized recommendations. However, most of them work with users or items as vectors, by applying different types of mathematical operations between them and neglecting sequential or content-based information. Hence, in this paper we study how to propose an adaptive mechanism to obtain user sequences using different sources of information, allowing the generation of hybrid recommendations as a seamless, transparent technique from the system viewpoint. As a proof of concept, we develop the Longest Common Subsequence (LCS) algorithm as a similarity metric to compare the user sequences, where, in the process of adapting this algorithm to recommendation, we include different parameters to control the efficiency by reducing the information used in the algorithm (*preference filter*), to decide when a neighbor is considered useful enough to be included in the process (*confidence filter*), to identify whether two interactions are equivalent (*δ-matching threshold*), and to normalize the length of the LCS in a bounded interval (*normalization functions*). These parameters can be extended to work with any type of sequential algorithm.

We evaluate our approach with several state-of-the-art recommendation algorithms using different evaluation metrics measuring the accuracy, diversity, and novelty of the recommendations, and analyze the impact of the proposed parameters. We have found that our approach offers a competitive performance, outperforming content, collaborative, and hybrid baselines, and producing positive results when either content- or rating-based information is exploited.

1. Introduction

Recommender Systems, even though they have been studied in depth during the last decade, remain a constant source of innovation. With the global growth of the Internet, they have become a necessary tool for a large number of online applications due to their ability to make personalized recommendations by adapting to different types of user profiles aiming to achieve better customer satisfaction (Ricci, Rokach, & Shapira, 2015). In order to produce interesting and personalized suggestions, it is usually necessary to work with large amounts of data (Ricci et al., 2015; Xu, 2018), and depending on how this information is exploited and the type of strategy developed, several types of recommender systems can be distinguished, although the two most classical and widespread techniques are the content-based (CB) and collaborative filtering (CF) ones. The former learns to recommend items similar to the ones the user liked before (de Gemmis, Lops, Musto, Narducci, & Semeraro, 2015), whereas the latter suggests items that users with similar tastes liked in the past (Ricci et al., 2015).

* Corresponding author.
E-mail addresses: pablo.sanchezp@uam.es (P. Sánchez), alejandro.bellogin@uam.es (A. Bellogín).

https://doi.org/10.1016/j.ipm.2018.10.003
Received 6 April 2018; Received in revised form 26 September 2018; Accepted 8 October 2018
Available online 18 October 2018