A novel recommendation approach based on chronological cohesive units in content consuming logs

Jaekwang Kim [a], Jee-Hyong Lee [b],*

[a] *Sungkyunkwan Convergence Institute of Intelligence and Informatics, Sungkyunkwan University, Republic of Korea*
[b] *College of Software, Sungkyunkwan University, Republic of Korea*

A B S T R A C T

We propose a novel recommendation approach based on chronological cohesive units (CCUs) of content consuming logs. Chronological cohesive units are defined as sub-sequences of logs in which items are highly related to each other. We first generate rules for splitting consuming logs into CCUs. We select features which are effective for splitting of consuming logs and combine them into a binary decision tree to generate splitting rules with genetic programming. With the rules, we split content consuming logs into CCUs, and identify strongly associated items in the CCUs. Next items are recommended with an association rule-based approach. The proposed method is evaluated using two-real datasets: web page navigation logs and movie consuming logs. The experiments confirm that the proposed approach is superior to the existing methods in various aspects such as hit ratio, click-soon ratio, sparsity, diversity and serendipity.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Recommender systems are one of the subclasses of information filtering systems that seek to recommend content for users [12,37]. Today, due to the huge amount of available data and the development of effective data processing algorithms, recommender systems have become indispensable tools in various domains, such as movies, books, news, music, and so on [6,7,19,20,35,38,44,45]. Indeed, recommender systems have a particularly significant effect on content businesses. As examples, over a decade ago, two thirds of movie rentals are prompted by recommendations in the case of Netflix. More than 38% of click-throughs on Google news are generated by recommendations, and 35% of sales on Amazon are attributable to recommendations [5].

Most of recommender systems are based on content consuming logs of users which are sequences of the content consumed by users. In a sequential log, neighboring items are associated each other, but all the neighboring items are not equally associated. The example in Fig. 1 shows two web navigational logs of two users. The users visited the same pages, but the orders of the visits differ. User *A* first visited three camera pages and then three smartphone pages, while user *B* navigated camera pages and smartphone pages in a mixed order.

Since the two users have visited the same pages, we may recommend the same pages to the both users if we do not consider the orders. However, the interests of user *A* seem to have changed from cameras to smartphones, while user *B*'s interests remain with photo-taking devices. That is, the log of user *A* can be separated into two sub-sequences in which

---

* Corresponding author.
E-mail addresses: linux@skku.edu (J. Kim), john@skku.edu (J.-H. Lee).

(a) Web navigational log of user *A*
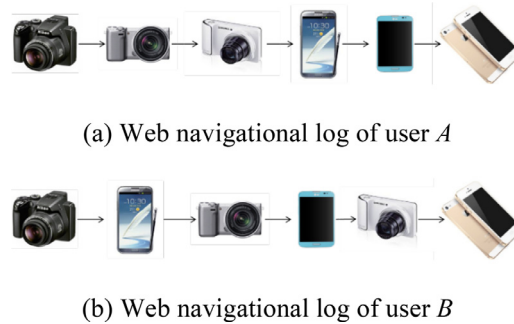

(b) Web navigational log of user *B*

Fig. 1. Two different consuming logs with the same consumed content.

items are more strongly associated. It will be helpful for user *A* to recommend web pages on smartphones considering the recent three pages on smartphones, but for user *B* web pages on photo-taking devices considering all the six pages. If we are able to separate user logs into cohesive sub-sequences which reflect consuming context, such as interest, we can obtain richer information and provide better recommendations for users.

In this paper, we are focusing on the cohesive sub-sequences in sequential logs. Because users consume content within contexts, it is very probable that there are sub-sequences in which contents are strongly associated. The identification of those cohesive sub-sequences facilitates the recommendation of next content to users.

Existing recommendation approaches for sequential logs can be categorized into transaction-based [11,14,15,20,25,29,34,39,41–43], sequence-based [8,21,26], session segmentation-based [9,27,30,33] and time-aware recommendations [3,4,10,31]. Transaction-based approaches consider content consuming logs as sets rather than sequences, so that it cannot aware the cohesive sub-sequences which reflect users' consuming context. On the other hand, sequence-based approaches may generate different recommendation models depending on the orders of consumptions. However, emphasis on modeling consumption orders may generate overfitted and complex models [2], thus the performance of recommendation could be degraded. Session segmentation-based approaches try to separate user logs into sessions to identify items strongly associated each other. However, most approaches relied on expert knowledge to identify sessions. Time-aware approaches generate recommendation lists by predicting temporal behaviors of users. It can recommend novel items, but the model complexity could be high.

The existing methods were unrecognized or limited to considering cohesive subsequences in user logs to recommend items to users. In this paper, we splits content consuming logs into cohesive sub-sequences or chronological cohesive units (CCUs), in which contents have strong associations with each other. It analyzes the CCUs to find strongly associated items and recommends items that have strong associations with the current CCU. Because content consuming logs are split into cohesive units, the associations found in the units are much stronger than the associations found in the original non-split content consuming logs. Thus, we can expect an improvement of the recommendation performance.

The question is how to split the content consuming logs of users into cohesive units. We assume that any cohesiveness in the logs mainly originates from the context in which users consume content. So, we first select various features which may reflect the contexts of content consuming [16]. We combine the features to establish splitting rules using a genetic programming approach. After splitting, we find strongly associated content sets in CCUs and recommend next content based on an association rule-based approach. The association rule-based method is suitable for recommending next content by observing previously consumed content [25].

To implement this approach, it is hard to find out the optimal parameters such as support, confidence values for association rules, and the number of pre-consumed content for recommendations. In this paper, the parameters for association rules are also found by a genetic approach, and the number of pre-consumed content for recommendations is also dynamically determined based on CCUs.

The proposed approach is evaluated using two-real datasets: web page navigation logs and movie consuming logs which have totally different characteristics. The experiments show that our approach is superior to the existing recommendation models in the aspects of hit ratio, click-soon ratio, sparsity, diversity and serendipity.

The rest of the paper is organized as follows. Section 2 presents an overview of related work. Section 3 describes the proposed approach. We present the experiments in Section 4 and conclude in Section 5.

2. Related work

The recommendation approach developed in this paper is concerned with sequential content consuming logs. In this section, transaction-based recommendation, sequence-based recommendation, session segmentation-based recommendation and time-aware recommendation approaches are discussed.