Knowledge-Based Systems 98 (2016) 68–75

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

# User identification for enhancing IP-TV recommendation

Zhijin Wang, Liang He*

*Shanghai Key Laboratory of Multidimensional Information Processing, Department of Computer Science and Technology, East China Normal University, Shanghai 200241, China*

ARTICLE INFO

*Article history:*
Received 9 May 2015
Revised 10 December 2015
Accepted 14 January 2016
Available online 10 February 2016

*Keywords:*
User identification
Shared account
IP-TV recommendation

ABSTRACT

Internet Protocol Television (IP-TV) recommendation systems are designed to provide programs for groups of people, such as a family or a dormitory. Previous methods mainly generate recommendations to a group of people via clustering the common interests of this group. However, these methods often ignore the diversity of a group's interests, and recommendations to a group of people may not match the interests of any of the group members. In this paper, we propose an algorithm that first identifies users in accounts, then provides recommendations for each user. In the identification process, time slots in each account are determined by clustering the factorized time subspace, and similar activities among these slots are combined to represent members. Experimental results show that the proposed algorithm gives substantially better results than previous approaches.



## 1. Introduction

In recent years, due to the rapid growth and increasing popularity of Internet Protocol Television (IP-TV) services, IP-TV services have been widely consumed in our daily life, and it has been a common phenomenon that families or roommates share programs after they get back home or dormitory from work. To help users (viewers) benefit from the abundant resources, such as channels, programs and videos, and easily find what they are actually interested in, recommender systems [1] are integrated into the services.

However, the main challenge in developing a IP-TV recommendation system is user identification [2]. Intuitively, the recommendations provided to a shared account, comprising the ratings of two dissimilar users, may not match the interests of either of both users [3]. The use of a single account shared by multiple users poses more personalized requirements in providing programs to this account.

Our goal is to improve the recommendation performance by alleviating the problem of user identification in IP-TV services. However, none of the individual user information [4] can be directly used for the identification, since the interaction between a user and a set-top-box (STB) is very weak. Typically, users do not have easy access to the keyboard, mouse or touch screen. Moreover, the services are indistinctly shared by the users in a shared account. The history logs recorded by STBs contain the following data fields: *AccountId, ProgramId, StartTime, EndTime* and *Genre*. In reality, a log recalls that an account starts and ends a program, and marks a program to a genre.

The assumption is that users within a shared account not only have distinct temporal habits, such as after dinner or at weekends, but also have different preferences for television programs (or genres). There are two questions: (1) how to accurately detect temporal habits over accounts? (2) how to accurately obtain preferences based on the detected temporal habits for a user?

To address these questions, we propose a novel algorithm that consists of a partition process and a consolidation process. In the partition process, the time is divided into several nonoverlapping time slots to present temporal habits of users. More specifically, we use the consumption logs to construct an account-item-time play count tensor. We decompose this tensor into the multiplication of a few (low-rank) latent matrices of accounts, items, time intervals, and a core tensor. And then, time slots are obtained via clustering the latent matrix of time intervals. In the consolidation process, we introduce virtual user to represent preferences of an account in a clustered time slot, and similar virtual users are combined to extract users.

A simple overview of our proposed algorithm is drawn in Fig. 1. The contributions of this paper are summarized as follows:

1. We study the problem of user identification in IP-TV services as mining groups of time slots and preferences within accounts.
2. We propose an algorithm to fulfill the identification task. In this algorithm, we try a tensor factorization based subspace clustering method to discover groups of time slots. And then

* Corresponding author. Tel.: (+86)021-54345153.
*E-mail addresses:* zhijin@ecnu.edu.cn (Z. Wang), lhe@cs.ecnu.edu.cn, zhijin@ecnu.cn (L. He).

http://dx.doi.org/10.1016/j.knosys.2016.01.018
0950-7051/

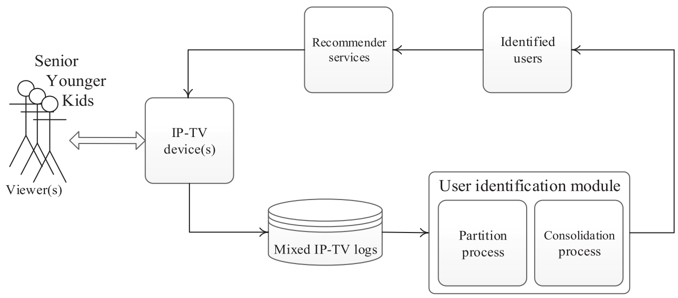


**Fig. 1.** An overview of user identification for IP-TV recommendation.

similar preferences over these time slots are combined to present users.

3. Finally, we demonstrate how this algorithm above can be applied to improve recommendation. Experimental results on a real IP-TV dataset show that our algorithm outperforms comparable methods.

A preliminary result was reported previously [5]. This paper substantially extends this work.

The rest of this paper is organized as follows. Section 2 provides a brief review of related work on IP-TV recommendation. Section 3 gives the problem definition and notations. Section 4 describes our proposed approach to carry out identification task. Section 5 shows the settings in our experiments. Section 6 presents the experimental results and analysis. Finally, Section 7 concludes the paper.

## 2. Related work

The work in this paper closely relates to the research areas: collaborative filtering and context-aware recommendation. We present the most relevant previous work in each of them.

### *2.1. Collaborative filtering*

Collaborative filtering (CF) has been the most popular technique for recommender systems [6,7]. Typically, collaborative filtering approaches can be divided into two types: memory-based methods as well as model-based approaches. Memory-based methods focus on using predefined similarity calculation functions to find similar users or items for generating predictions [8]. Memory-based methods can be further classified as user-based [9–11] and item-based [12–14] approaches based on whether similar users or similar items are used. In contrast, the model-based approaches use the observed ratings to train a predefined learning model, and the unobserved ratings are then predicted via the trained model. Algorithms in this category include but not limited to clustering model [15], the aspect models [16], the Bayesian hierarchical model [17], the ranking model [18], etc.

Recently, a particular group of methods, referred to as matrix factorization methods, have become dominant in the field [19]. The performance of the group of methods for the rating prediction problem has been tested in Netflix Prize Contest [20,21] and the KDD CUP 2011 [22]. Matrix factorization methods normally seek to factorize the user-item rating matrix into two low rank latent matrices of users and items, and then utilize the factorized matrices to make further predictions. The factorized latent matrix of user is also employed to cluster groups of users with similar tastes. Zhang et al. [3] believed that users within a household have similar interests, and applied several clustering algorithms to identify groups of users as households. Matrix factorization methods have been further extended to incorporate content metadata information [23], so that the rich side information of users and items beyond the

**Table 1**
Notations and semantics.

| Notation | Semantic |
|---|---|
| $A$, $I$ | set of accounts, set of items |
| $U$ | set of identified users |
| $U(a)$ | set of identified users within account $a$ |
| $u_{ah}$ | the $h$th identified user within account $a$ |
| $P$ | period (a.k.a, set of sub-period) |
| $p_k$ | the $k$th sub-period within $P$ |
| $v_{ak}$ | virtual user within account $a$ in $p_k$ |
| $s_{ah}$ | several sub-period consumed by $u_{ah}$ |
| $I(v_{ak})$ | set of items consumed by account $a$ in $p_k$ |
| $\mathcal{C}$ | account-item-time play count tensor |
| $T$ | time dimension of $\mathcal{C}$ (a.k.a, time interval set) |
| $\mathcal{M}$ | factorized core tensor |
| $X$, $Y$, $Z$ | latent matrix of accounts, latent matrix of items, latent matrix of time intervals |
| $c_{ait}$ | play count of account $a$ to item $i$ in time interval $t$ |
| $\hat{c}_{ait}$ | predicted play count of account $a$ to item $i$ in time interval $t$ |
| $d_{aik}$ | duration of account $a$ consumes item $i$ in $p_k$ |
| $r_{aik}$ | implicit rating of account $a$ to item $i$ in $p_k$ |
| $G$ | preference similarity graph of virtual users |
| $S_{vv'}$ | preference similarity between virtual user $v$ and $v'$ |
| $\rho$ | threshold of preference similarity |
| $\lvert\cdot\rvert$ | length of a set |

user-item relations can be exploited for improving recommendation. In addition, the matrix factorization framework has also been developed for the top-$K$ recommendation problem in domains with implicit feedback data [24,25], and the binary rating technique is employed to represent users' implicit preferences.

However, these methods cannot be directly used to alleviate our problem, since a television is indistinctly shared by multiple users. In this paper, we consider users within a shared account do not have common interests and temporal habits in IP-TV services.

### *2.2. Context-aware recommendation systems*

The context-aware recommender systems (CARS) have received lots of attentions. Early work in CARS utilized contextual information (e.g., demography, location and time) for pre-processing, where the context drives data selection, or post-processing, where the context is used to filter recommendations [26]. Said et al. [27] used time to split an user into two contextual user profiles, and recommendations are provided to each contextual user profiles. A significant portion of recent work has focused on incorporating context variables into the matrix factorization methods [28–30]. Due to the success of matrix factorization for modeling the user-item relations, one major group of approaches exploited the tensor factorization techniques [31] for modeling the 3-way user-item-context relations [32–34]. A tensor in this case is a generalization of matrix from 2-dimension to $n$-dimension. Another contribution for modeling the contextual information is factorization machines [35,36], which models the interactions between each pair of entities in terms of their latent factors, such as user–user, user-item, user-context interactions.

The work in this paper also builds upon tensor factorization models. The latent matrix of the context (time) is factorized and clustered to detect temporal habits.

## 3. Problem definition and notations

Table 1 gives the main notations used in this paper.

To start with, let us consider a common scene that multiple users share a common account in IP-TV services. As Fig. 2 shows,