VideoTopic:
Modeling User Interests for Content-Based Video Recommendation

Qiusha Zhu, Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL, USA
Mei-Ling Shyu, Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL, USA
Haohong Wang, TCL Research America, Santa Clara, CA, USA

ABSTRACT

With the vast amount of video data uploaded to the Internet every day, how to analyze user interests and recommend videos that they are potentially interested in is a big challenge. Most video recommender systems limit the content to metadata associated with videos, which could lead to poor recommendation results since the metadata is not always available or correct. On the other side, visual content of videos contain information of different granularities, from a whole video, to portions of a video, and to an object in a video, which are not fully explored. This extra information is especially important for recommending new items when no user profile is available. In this paper, a novel recommendation framework, called VideoTopic, that targets at cold-start items is proposed. VideoTopic focuses on user interest modeling and decomposes the recommendation process into interest representation, interest discovery, and recommendation generation. It aims to model user interests by using a topic model to represent the interests in the videos and then discover user interests from user watch histories. A personalized list is generated to maximize the recommendation accuracy by finding the videos that most fit the user’s interests under the constraints of some criteria. The optimal solution and a practical system of VideoTopic are presented. Experiments on a public benchmark data set demonstrate the promising results of VideoTopic.

Keywords: Internet, Metadata, Recommendation, Recommender Systems, VideoTopic

1. INTRODUCTION

As the exponential wide use of digital devices, Internet video traffic will be 55 percent of all consumers’ Internet traffic in 2016, up from 51 percent in 2011, as reported by Cisco. It would take over 6 million years to watch the amount of videos that will cross global IP networks each month in 2016 (Cisco, 2012). Users are overloaded by the choices of so many videos that a smart recommender system is on demand, which could provide the recommendation lists personalized to each user’s interests.

Research on recommendations are generally proceeded along three dimensions: content-based recommendation that focuses
on analyzing the content of items; collaborative filtering that utilizes user profiles, such as ratings or clicks, to recommend items for like-minded users; and hybrid recommendation that incorporates both approaches. Due to the superior performance in Netflix competition (Lowe, 1999), latent factor model (LFM) was adopted in many state-of-the-art recommendation models (Koren, Bell, & Volinsky, 2009; Rendle & Schmidt-Thieme, 2008; Salakhutdinov & Mnih, 2008). These approaches belong to the collaborative filtering category, which involve analyzing user profiles, typically in the form of the user-item matrix. However, in many situations, user profiles are not available or very sparse, especially for online videos, since a large proportion of the users browse videos anonymously. As a result, dealing with the cold-start problem is inevitable. The cold-start problem describes the scenarios in recommender systems when user profiles are not available, which commonly arises at the beginning of a recommender system. Thus, for cold-start items, i.e., items without any user behavior data, collaborative filtering would fail. Some recently proposed frameworks bring the contents of items into consideration, such as (Agarwal & Chen, 2009; Gantner, Drumond, Freudenthaler, Rendle, & Schmidt-Thieme, 2010), which extended LFM to incorporate item features, user features, and global features in its model. These approaches can only handle cold-start problem to some extent for they rely on factorizing the user-item matrix or use it to optimize the models. If all the items are cold-start items, which are very common in real applications, especially when launching a new recommender system, these improved approaches would still fail and only content-based recommendation can be adopted at this stage before enough user profiles can be gathered.

Currently, the content in most recommender systems is limited to the metadata associated with items (Cheng, Dale, & Liu, 2008). It represents items as feature vectors, and user interests or preferences are discovered by analyzing these textual features (Davidson et al., 2010; Zhao et al., 2012). For video recommendations, Netflix and Hulu use movie genres, sub-genres, or the combination of them to describe and organize user interests. Jinni constructs movie genome by expert knowledge and online reviews to describe each movie. Similar work is conducted in music recommendation. Pandora Radio’s core technology is the music genome project, which uses about 400 attributes to describe the properties of songs and artists. It allows Pandora Radio to recommend songs that possess similar musical traits as the songs liked by the users, and thus it needs very little user behavior information to get started. One shortcoming is that constructing the genome requires expert knowledge. Another shortcoming is that using textual features to describe items requires the textual information about items to be accurate. However, most of the online videos are unorganized nowadays, which means their textual information could be incomplete, non-existent, or even incorrect. For these videos, either a lot of efforts need to be spent to manually annotate them, or automatic tagging methods have to be applied (Larson et al., 2011; Siersdorfer, San Pedro, & Sanderson, 2009). Otherwise, these systems would produce poor results. Meanwhile, user interests in a video are multifaceted. A user might be interested in the plot, the characters in the video, or the visual appearance of some scenes. In these cases, a deeper analysis to the video content is needed in order to recommend videos that a user is really interested in, which can hardly be captured by the video metadata. Moreover, users are usually in full control of when to start and end watching. It is very often that a user watches a part of the video because he or she only wants to watch that part. Unfortunately, metadata that describes the whole video cannot capture and interpret these watch patterns in a finer granularity. Given these inevitable problems of using metadata alone, visual semantic content mining could provide extra information from a visual point of view and help better capture user interests.

To recommend personalized videos, not only the information other than the metadata needs to be considered, the granularity of user interests is also worth examining. Compared to regular items, such as those products sold on
Location-Based Network Resource Management
Handbook of Research on Mobile Multimedia (pp. 139-164).
www.igi-global.com/chapter/location-based-network-resource-management/20963?camid=4v1a

A High Quality Audio Coder Using Proposed Psychoacoustic Model
(2012). Signal Processing, Perceptual Coding and Watermarking of Digital Audio:
Advanced Technologies and Models (pp. 102-114).
www.igi-global.com/chapter/high-quality-audio-coder-using/56064?camid=4v1a