Knowledge-Based Systems 56 (2014) 167–178

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A smart TV system with body-gesture control, tag-based rating and context-aware recommendation

Wei-Po Lee *, Che Kaoli, Jhih-Yuan Huang

Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan

ARTICLE INFO

Article history:
Received 22 February 2013
Received in revised form 9 November 2013
Accepted 9 November 2013
Available online 18 November 2013

Keywords:
Social tagging
Body control
Collaborative filtering
Context-awareness
Mobile multimedia

ABSTRACT

Smart TV, which enables viewers to conveniently access different multimedia contents and interactive services in a single platform, is currently becoming popular. The work presented here develops three new ways to enhance the performance of a smart TV. The first is to design a body control system that recognizes and interprets human gestures as machine commands to control the TV. The second is to create a new social tag-based method to recommend the most suitable multimedia contents for users. Finally, a context-aware platform is implemented to consider different environmental situations to make the best recommendations.

1. Introduction

Smart TV has quickly become popular in the market in recent years. In contrast to traditional TVs that only focus on media broadcasting, smart TV systems can deliver diverse multimedia contents from networked devices directly to the end-users, allowing them to access the contents through a user-friendly interface. Smart TVs also provides interactive Internet-based services, including media-on-demand, social networking, and online gaming. Smart TV continues to develop rapidly to offer increasingly more functions and services. In this work, we address three important issues closely related to smart TV, including the human–machine interface, content recommendations, and context awareness, and we develop the corresponding mechanisms to further enhance the performance of a smart TV system.

Traditionally, people have used device-based control to operate different consumer electronics at home. Researchers then began to implement systems that utilize personal hand-held devices to work as controllers [1,2]. Recently, different methods of human–machine interactions (such as voice-based and gesture-based control) have been proposed to provide natural control on the equipment without any remote control devices [3,4]. Gesture control has many advantages over traditional control methods, such as its high accessibility (with no need of physical contact), high flexibility (adaptable to various applications), and low cost (fixed sensor hardware cost, in contrast to touchscreen cost which increases exponentially while the screen size increases). To move towards an even more natural way for users to interact with machines, in this work we design a body control mechanism through which the smart TV can respond to the human users' actions.

On a smart TV system which can easily present different types of multimedia content to end-users, a large amount of contents can lead to information overload. Therefore, it is necessary to develop personalization techniques to recommend the most suitable content to users [5,6]. Many recommendation methods have been proposed, ranging from content-based user modeling to group-based collaboration. Generally speaking, the collaboration-based approach is considered more efficient and effective than the content-based user modeling approach [7]. The current approach toward organizing and sharing digital content through user-created metadata (i.e., social tags) suggests a potential mean to further improve collaborative recommendation by using metadata to interpret the users' preferences for any specific items. Additionally, we will employ such metadata for multimedia annotation. Also we will use the tag information to analyze the users' preferences and to exploit such information to make collaborative recommendations.

In addition to recommendation techniques that focus on multimedia items, context is an important issue to be considered in personalized recommendations, especially when mobile smart TVs become popular. These handheld devices can collect and analyze context information about their user. In general, context awareness means the ability of the computing systems to acquire and interpret the context information to adapt to the corresponding applications. In other words, context awareness enables a

* Corresponding author. Tel.: +886 73332675.
E-mail address: wplee@mail.nsysu.edu.tw (W.-P. Lee).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.11.007

system to capture a broad range of contextual attributes (such as the user's location, activities, and surrounding environment) to better understand what the user is trying to accomplish and what content suits the user the most within that context [8–10]. By integrating context information into the application service, a recommender system can fulfill the user's needs more efficiently and practically.

Taking the above issues into account, we develop a smart TV system with several unique features. First, we design a Kinect-based system to recognize human body gestures for TV control. Second, we compare different computational methods for making personalized recommendations on multimedia items. Following the current trend of community-based information sharing [11–15], we also integrate social tags to annotate multimedia items to improve recommendation performance. The experimental results show that the tag-based method outperforms other conventional methods. Finally, we implement a context-aware platform to present how we integrate various environmental situations into the proposed method to perform item recommendation accordingly.

2. Background

As mentioned above, due to the tremendous amount of digital multimedia items a smart TV can broadcast, a recommendation mechanism is needed to offer better services. Recommender systems have been advocated in different service domains for many years [16,17]. In general, the recommendation techniques can be categorized into three types: content-based, collaborative filtering, and hybrid methods [16]. The content-based approach predicts the user's preferences for new items based on historical records. Therefore, the most important issue in this approach is constructing a computational model to perform the prediction. Many machine-learning approaches have been applied to construct user models. Nevertheless, the content-based approach largely relies on sufficient samples to construct the model. This approach often recommends items within a specific scope and thus loses item diversity (i.e., ignoring items of unfamiliar classes).

In contrast to the content-based method, the collaborative filtering (CF) method does not build a personal model for prediction. There are two major techniques to perform CF methods: the neighborhood methods and latent factor models [7]. The neighborhood methods recommend items to the user according to the evaluations (opinions) of other users with similar tastes (or recommends items similar to the ones with high user ratings in a similar way). In such an approach, therefore, the measurement of similarity between users is most important so that the system can employ a k-nearest neighbor method to find the most similar users to perform the recommendation. The system's prediction of a new item for a user is thus based on a combination of the ratings of the user's nearest neighbors. This approach has been widely used in different applications, for example [2,18].

The above similarity-based methods are very popular because they are intuitive and relatively simple to implement. They also offer useful and important properties: explicit explanation of the recommendations and easy inclusion of new ratings [7,16]. However, standard neighborhood-based methods raise some concerns: for example, the similarity function is often arbitrary; the interactions among neighbors are not taken into account; the weights for each neighbor need to be determined properly; and some users have rated only a very small number of items (i.e., the cold-start problem).

The second type of CF methods, latent factor models, provides an alternative approach by transforming both users and items to the same latent factor space. This space intends to explain the ratings on some implicit factors obtained automatically. Different algorithms have been proposed to derive these factors by minimizing the difference between the predicted ratings and the observed ones (e.g., [19,20]). Matrix factorization (i.e., singular value decomposition or SVD) is one of the most popular latent factor models, in which each user and item is assumed to be represented by a small number of unobserved features. Though latent factor models can provide somewhat more accurate results than the neighborhood methods, still they have some disadvantages. For example, optimizing the often used object function (the sum-of-square of factorization error) in the traditional matrix factorization methods cannot guarantee the consistence of the predicted ratings and the user preferences. Also, the latent factor models often employ a stochastic algorithm (such as gradient decent) to solve the optimization task. As can be observed, when the numbers of users and items increase, the corresponding optimization task will become more difficult to solve.

Some researchers have pointed out that factorization methods can lead to relatively accurate results, while neighborhood methods can have some practical advantages [7]. Therefore, several integrated methods have been suggested to have advantages from both types of methods, for example, to derive a hybrid model that can add a local perspective delivered by the neighborhood method into a global factorization method [7]. Some other techniques are also proposed to extend the capabilities of recommender systems in general, including taking more user and item information into account, incorporating contextual information, and supporting multi-criteria rating [10,16,21]. Meanwhile, to overcome the cold-start problem, the notion of trust has been introduced to improve the recommendation accuracy [22]. Trust plays a major role in exchanging relationships among people; it is often used to decide with whom we share information and from whom we accept information. It is believed that by utilizing a trust network (a social network augmented with trust ratings) in collaborative filtering methods, better results can be obtained [22,23].

In addition to measuring item similarity or user similarity to predict user preferences, we can also analyze how the user prefers specific items from different feature dimensions (i.e., in more detail). Social tags are brief descriptions of items and can be used as features to capture the semantics of the target items, for example [11,24]. These tags are freely supplied by a community of Internet users to aid access to large collections of media [12,13]. Compared with knowledge representation schemes that involve domain experts, the tagging activity shifts the task of classifying domain items from knowledge engineers to Internet users; it is thus a cost-efficient alternative to the popular and precise knowledge ontology in content annotation [14,24]. Additionally, tagging is neither exclusive nor hierarchical. In some circumstances, tagging has advantages over hierarchical taxonomies. Therefore, in this work we choose to use social tags to represent item features, and we propose a new tag-based collaborative filtering (tag-CF) method for multimedia recommendation. In contrast to other works that use tags to annotate target items (e.g., [15,25]), our work uses social tags to predict a user's preference for recommendation.

Context plays an important role in determining the relevance of an application's service (or function) to a user's needs, and any small contextual changes may lead the user to select a different service. Dealing with the context issue involves defining contexts relevant to the application service and identifying the key contexts in which people often use the service. Regarding different mobile applications, context factors can be defined as any information used to characterize the user situation that can influence his decision in requesting a service. There are two types of context factors: personal and environmental [8,26]. Personal context is the personal state or condition of the user himself (such as his emotional