Personalized Movie Recommendation

Anan Liu[1,2,3], Yongdong Zhang[1], Jintao Li[1]
1. Laboratory of Advanced Computing Research, Institute of Computing Technology, CAS, Beijing, China
2. Robotics Institute, Carnegie Mellon University, Pittsburgh, USA
3. School of Electronic Engineering, Tianjin University, Tianjin, China
{liuanan, zhyd, jtli}@ict.ac.cn

ABSTRACT
Facing the vast amount of novel production in the movie industry, people are in favor of choosing their favorite candidates quickly and previewing movie contents conveniently so as to decide whether they appeal to their personal taste. To meet this growing need, researchers are paying more attention on *Personalization* and *Recommendation*, the new trends of multimedia information retrieval, by integrating content and contextual information. In this paper, we propose a hierarchical framework for personalized movie recommendation. First, movie weekly ranking information is utilized for movie association and recommendation. Then, an integrated graph with both movie content and user preference is constructed to generate dynamic movie synopsis for personalized navigation. The superiorities of the proposed method have two aspects: 1) The prior knowledge independent recommendation scheme is implemented to replace the traditional ranking method for novel information access; 2) Personalized movie synopsis is interactively produced to replace the current movie trailer for preview. The promising results of subjective evaluation indicate that the proposed framework can discover the latent relationship between movies as well as movie highlights and therefore provide personalized movie recommendation to effectively lead movie access in an individualized manner.

Categories and Subject Descriptors
H.3.3 [**Information Search and Retrieval**]: Information Filtering; H.4 [**Information Systems Applications**]: Miscellaneous.

General Terms
Algorithms, Management, Human Factors.

Keywords
Personalized, movie, recommendation, semantic, user preference.

1. INTRODUCTION
The movie industry is an active producer of video. Every year, more than 4,500 movies are released around the world, spanning approximately 9,000 hours of video. With such a massive amount of information, there is a great demand for technologies that enable viewers to access new movies conveniently and therefore facilitate movie propagation.

Recommendation of relevant items to users is a good way to alleviate users' efforts on finding their targets. Therefore, it attracts much interest of researchers. Mukherjee, et al., proposed a user preference modeling method with both voting schemes and text-based learning for movie recommendation [1]. Different from this user model with only users' preference, the methods in [2] integrated multiple contextual information for recommendation. However, there exists the following limitation: 1) The conflicts among user preferences bring much uncertainty for user modeling; 2) The prior knowledge-based method would bring many constrains for the real application; 3) Most of these methods ignore the movie content which is one of the most important factor to attract viewers. Therefore, these methods have not been widely used till now. In the industry area, internet movie social network, such as Internet Movie Database (IMDb) [3] and iTunes Movie Trailers [4], are still the popular method for movie access. These websites usually provide many kinds of ranking information classified by decade, genre, and gender and so on. With these references, new audiences can look for their favorite candidates. Moreover, these websites also provide movie trailers for preview. Therefore, viewers can simply preview the movie trailers of the candidates and make the final decision. Although it seems that the current system works well, there come two important problems with the evaluation on IMDb by 20 volunteers: 1) The simply office box or rental-based ranking can not represent the relationship of movies well and may degrade the potential popular movies; 2) the fixed content and style of the movie trailer may not be suitable for diverse tastes of all the audiences and thereby lead to loss of potential viewers.

To realize the personalized information access in the area of research and practical commercial application in the area of industry [5], we propose a hierarchical framework for personalized movie recommendation. First, the movie weekly ranking information is utilized for movie association and recommendation. With the discovered communities and recommendation scheme, viewers can narrow the candidates for selection. Then, an integrated graph with both movie content and user preference is constructed to generate dynamic movie synopsis for personalized navigation. Thus, viewers can refine the candidates and make the final decision quickly. The main contributions of our work are twofold: 1) The prior knowledge independent hierarchical recommendation scheme is implemented to replace the traditional ranking and searching method facing the situation in which we need to make choices without sufficient personal experience; 2) Personalized movie synopsis is interactively produced to replace the current movie trailer for preview. The promising results of subjective evaluation indicate that the proposed framework can discover the latent relationship between movies as well as movie highlights and therefore provide personalized movie recommendation to effectively lead movie access in an individualized manner.

MM'09, October 19–24, 2009, Beijing, China.

The remainder of the paper is organized as follows. In Section 2, we present the proposed framework for personalized movie recommendation. Experimental results are illustrated in Section 3, followed by conclusions and future work in Section 4.

2. PERSONALIZED MOVIE RECOMMENDATION

The proposed framework consists of two hierarchies: (1) movie association and recommendation; (2) personalized movie navigation.

2.1 Movie Association and Recommendation

2.1.1 Movie Association

Research in consumer psychology has shown that the product adoption is synonymous to the will to acquire novel information and is influenced by preferences and external environment [6]. For movie recommendation, consumers' behaviors for novelty seeking and conformity seeking can be well accommodated by movie weekly ranking information.

Each movie in the movie social network has two attributes, rank and change of rank. Therefore, we represent each movie (*M*) by:

$$\text{M=M}\left(R_t, \ldots, R_{t-N+1}, R_{AVE}; C_{t,t-1}, \ldots C_{t-N+2,t-N+1}, C_{STD}\right) \quad (1)$$

where, R_t means the rank in the t^{th} week; R_{AVE} means the average rank during *N* adjacent weeks; $C_{t-i,t-i-1}$ means the change of rank between two continuous weeks; C_{STD} means the standard deviation of the changes during N adjacent weeks. We use $\mathrm{M}_{t,r}$ to denote the movie whose rank is *r* in t^{th} week. When the rank information of movies in the current week, T^{th} week, and previous *N-1* weeks is automatically extracted from the website, we keep the following two clusters of movies:

$$C_1 = \{M_{T,r} \mid r \in [1, L]\} \quad (2)$$

$$\begin{aligned} C_2 = \{&M_{t,r'} \mid R(M_{t,r'}) > R(M_{T,r}), \\ &(t \in [T, T-N+1], r \in [1, L], r' \in [1, L'], L < L' \}) \end{aligned} \quad (3)$$

where $R(\bullet)$ means the rank of movie. Cluster C_1 means the top *L* movies in the current week while cluster C_2 means those which used to be in the top L' list and rank higher than at least one of C_1 during the *N* weeks. If there are several movies which have the same rank with the L^{th}, only those with the higher average ranks during the N weeks are kept so that C_1 has *L* members. To discover the trends of rank change, we associate these movies in two methods: 1) Intra-class association: C_1 and C_2 are separately clustered into K_1 *and* K_2 categories depending on their attributes:

$$\begin{cases} C_1 = \{C_{1,k_1} \mid k_1 \in [1, K_1]\} \\ C_2 = \{C_{2,k_2} \mid k_2 \in [1, K_2]\} \end{cases} \quad (4)$$

Therefore, the movies in each category are associated based on their similar trend of rank changes. 2) Inter-class association: C_1 and C_2 are clustered together into K categories as follows:

$$C_{12} = \{C_{12,k} \mid k \in [1, K]\} \quad (5)$$

Then C_{1,k_1} and C_{2,k_2} are associated if $C_{12,k} \cap C_{1,k_1} \neq \varnothing$ and $C_{12,k} \cap C_{2,k_2} \neq \varnothing$ are both satisfied. In this way, the lost movies below top *L* ranks are recovered and associated with the top *L* ones for recommendation.

2.1.2 Movie Recommendation

Based on movie association, we propose a recommendation scheme with three different strategies, including global, related and long tail recommendation, for different applications.

(1)Initialization: Global Recommendation. The one with the highest rank in each cluster C_{1,k_1} of C_1 is firstly recommended. It is intuitive that different center denotes specific trend of rank change and therefore it accommodate viewers' novelty seeking.

(2)Update: Related Recommendation & Long Tail Recommendation. If the i[th] center of C_{1,k_i} is selected, the other members in C_{1,k_1} will be recommended depending on their ranks in the related recommendation mood. In addition, the members of the associated cluster in C_2 will also be recommended in the long tail mood. With the two methods, viewers' conformity seeking can be accommodated.

2.2 Personalized Movie Navigation

Because user preference is quite subjective, the uncertainty and conflicts may bring much difficulty for constructing a general user modal. Comparatively, a personalized display of movie content can avoid the problems and would also be helpful to attract potential audiences. Therefore, both movie content and user preference are implemented for personalized movie navigation.

2.2.1 Integrated Graph Construction

First, the multi-model information can be fused for movie highlights extraction. Then, both user preference and semantic information are utilized to generate bipartite networks for the integrated graph construction. User's preference is collected by timing the interval of adjacent selection of highlights by viewers during the interaction between human and computer. If the interval is longer than the preset threshold, the user and highlight are linked as shown in the left part of Figure.1 (a) which means the user likes this highlight. Similarly, semantic detectors are implemented to construct a bipartite network to model the relationship of semantic information and highlights as shown in Figure.1 (b).

We implement the weighting method in [8] to project bipartite network to undirected graph. This method allows the resource flow from one node to another on the bipartite networks and hereby denotes correlation between highlights. As shown in the middle and right parts of Figure.1 (a) and (b), two undirected graphs and corresponding confusion matrixes W1 and W2 are generated. Because the correlation of two highlights depends on both user preference and semantic annotation, the latent relationship between them can be integrated as follows:

$$W = W1 \bullet * W2 \quad (6)$$

where '.*' is inner product. The matrix W represents the relationship of individual highlight in the integrated graph and can be used for information discovery and recommendation.

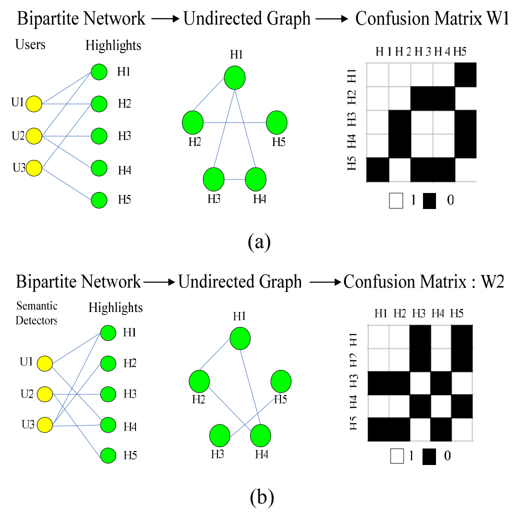


Figure.1. (a) Graph construction with user preference; (b) Graph construction with semantic annotation.

2.2.2 Community Discovery for Navigation

Based on the integrated graph, highlights can be grouped for community discovery. However, complicated linkage in the graph usually lead to highly overlapping cohesive groups of nodes due to human's multiple tastes. To uncover the modular structural of the complex graph, we implement the method in [9] by analyzing the main statistical features of the interwoven sets of overlapping communities to not only discover the k-clique-community but also make the members of the community reachable through well connected subsets of nodes.

By periodically updating integrated graph and communities with latest user selection information, dynamic movie synopsis can be generated with renewed relationship between highlights in an interactive mood. We can implement the same recommendation scheme in Section 2.1.2 for personalized movie navigation.

3. EXPERIMENTS

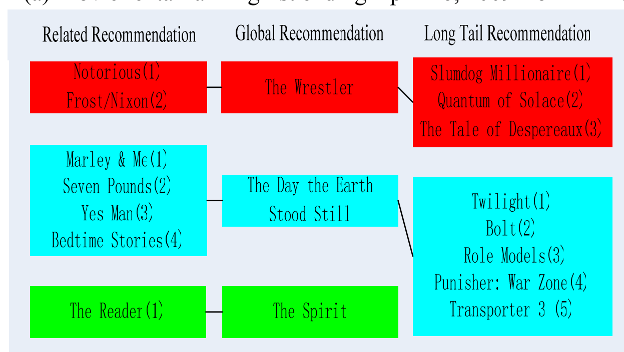
To evaluate the proposed framework for personalized movie recommendation, we choose user study for each hierarchy because both personalization and recommendation are strong subjective tasks and it is difficult for any mechanical comparison or simulation methods to obtain accurate evaluations [10]. In the experiments, we carried out a user study experiment and invited 20 volunteers to give their subjective scores which are quantified into five levels, 5 to 1 respectively denoting from best to worst.

3.1 Evaluation on the First Hierarchy

We collected weekly movie rental information during 8 weeks (N=8) from March 2, 2009 to April 26, 2009. We took top 10 movies (L=10) in the latest week and top 50 movies (L' =50) in each week for consideration. Thus C_1 can C_2 respectively contain 10 and 16 members. Each movie can be represented with 17-D feature vector by equation (1). We respectively clustered C_1 , C_2 and C_{12} into 3 categories (K_1=K_2=K=3) with Kmeans algorithm. The comparison between movie rental ranking list ending April 26, 2009 and recommendation list based on the first hierarchy are shown in Figure.2. We set 4 criteria for evaluation:

Rank	Title	Rank	Title
1	The Wrestler	15	Twilight
2	The Day the Earth Stood Still	17	Bolt
3	Notorious	18	Role Models
4	The Spirit	19	Punisher: War Zone
5	Marley & Me	20	Transporter 3
6	Frost/Nixon	21	Australia
7	Seven Pounds	22	Milk
8	The Reader	23	Rachel Getting Married
9	Yes Man	24	Beverly Hills Chihuahua
10	Bedtime Stories	25	Body of Lies
11	Slumdog Millionaire	26	In the Electric Mist
12	Quantum of Solace	28	Changeling
13	The Tale of Despereaux	31	Nights in Rodanthe

(a) Movie rental ranking list ending April 26, 2009 from IMDb



(b) Recommendation list with the proposed scheme (Different colors denotes associated movies; Value beside titles means the rank in each category)

Figure.2 Comparison between the current ranking list and the one based on the 1st hierarchy.

Table 1. Subjective evaluation for 1st hierarchy

Method	I	II	III	IV
IMDb	3.3	4.0	2.1	3.5
Proposed	4.3	4.2	4.6	4.5

(Note: The scores above are the average by 20 participants)

I Time-consuming: whether users can quickly find favorite movie candidates;

II Acceptable: whether the method is friendly for human-computer interaction;

III Helpful: whether the method can recover the potential popular movies;

IV Overall: the overall evaluation for the method considering 3 criteria above.

The evaluation on both recommendation lists by the volunteers is used for comparison. From Table.1 we can see that movie association and recommendation can friendly facilitate viewers to find the candidates quicker without increasing much operation and moreover it can recover the potential popular movies by long tail recommendation. Therefore, the proposed method got much higher overall score than IMDb.

3.2 Evaluation on the Second Hierarchy

For the experiment in the second hierarchy, we selected 6 unfamiliar movies for the 20 volunteers. The details of each movie are presented in Table.2. For pre-processing a movie, video structurization, including: shot boundary detection, key frame extraction, and scene boundary detection in [11], was implemented on it. Then movie highlights were extracted and annotated with semantics based on the methods in [7].

Table 2. Detailed information about each movie

Movie Title	a	b	c	d	e	f
Runtime (min)	110	120	113	145	131	134
Movie Genre	Action/Drama		Action/Sci-fi		Action/War	

(a: Fearless; b: Crouching Tiger, Hidden Dragon; c: The Matrix 1; d: Minority Report; e: Enemy At The Gates; f: Wind Talkers;)

Table 3 Subjective evaluation for 2nd hierarchy

Method	I	II	III	IV
Trailer	1	5	3.4	3.6
Proposed	4.8	4.2	4.5	4.6

(Note: The scores above are the average by 20 participants for 6 movies)

To conduct effective evaluation for personalized navigation, we set 4 criteria for reference:

I. Personalized: whether the method can satisfy individual interests;

II. Acceptable: whether the method is friendly for human-computer interaction;

III. Attractive: whether the display of the movie content attracts audiences;

IV. Overall: the overall evaluation for the method considering the 3 criteria above.

The results that volunteers evaluated on both movie trailers from IMDb and personalized navigation are used for comparison. From Table 3, we can get the following results:

- Personalized: The proposed navigation method can interactively recommend movie highlights depending on viewers' selection. Comparatively, traditional trailer can only show the fixed movie content. Therefore the former strongly outperforms the later on personalization.

- Acceptable: Although the proposed navigation method needs more operation, it is still acceptable for three reasons: (1) viewers can choose highlights based on their interests with the recommendation methods although facing unfamiliar movies; (2) movie highlights benefit the understanding of movie content; (3) the duration of one highlight is under the patience of viewers.

- Attractive: The proposed navigation is more attractive because the interactively and dynamically generated synopsis can avoid the loss of the potential audiences by the movie trailers with fixed content and style.

- Overall: Although movie trailer is a widely used method for movie access, it can not satisfy viewers' great need for conveniently previewing novel movies. By recommending highlights to viewers, users can easily navigate the movie content depending on personal interest and quickly decide whether the entire movie is worth watching. With the proposed recommendation scheme, the system can perform much better to satisfy human's diverse taste.

4. CONCLUSION AND FUTURE WORK

In this paper, we propose a hierarchical framework for personalized movie recommendation. The information of movie weekly ranking is utilized for movie association and recommendation. Moreover, both movie content and user preference are integrated to generate dynamic movie synopsis for personalized navigation. The promising results of subjective evaluation indicate that the proposed framework can not only avoid the problems of user preference collection and conflicts but also facilitate movie access with personalized recommendation.

In the future, we will focus on the following challenging aspects: (a) Exploring movie ranking information for movie association and recommendation; (b) Developing algorithms of latent relationship discovery with content and contextual information for movie highlight categorization and recommendation.

5. ACKNOWLEDGMENTS

This work was supported by National Basic Research Program of China (973 Program, 2007CB311100), Beijing New Star Project on Science & Technology (2007B071), Co-building Program of Beijing Municipal Education Commission.

6. REFERENCES

[1] R. Mukherjee, et al., A Movie Recommendation System - An Application of Voting Theory in User Modeling. *User Modeling and User-Adapted Interaction*, pp. 5-33, 2004.

[2] C. Ono, et al., A Context-Aware Movie Preference Model Using a Bayesian Network for Recommendation and Promotion. *User Modeling*, pp. 247-257, 2007.

[3] http://www.imdb.com.

[4] http://www.apple.com/trailers.

[5] Nicu Sebe, Qi Tian, Personalized Multimedia Retrieval: The New Trend?, In *proceeding of MIR'07*, Germany.

[6] Sabri Boutemedjet, Djemel Ziou, A graphical Model for Context-Aware Visual Content Recommendation. *IEEE Transaction on Multimedia*, Vol. 10, No.1, Jan., 2008.

[7] Anan Liu, Sheng Tang, Yongdong Zhang, et al., A Hierarchical Framework for Movie Content Analysis: Let Computers Watch Films like Humans. In *proceeding of CVPR2008 Workshop on new SLAM2008*, America.

[8] Tao Zhou, Jie Ren, Matús Medo, et al., Bipartite network projection and personal recommendation. *Physical Review E* 76(4): 046115, 2007.

[9] Palla et al., Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 2005.

[10] Yu-Fei Ma, Lie Lu, Hong-Jiang Zhang, et al, A User Attention Model for Video Summarization. In *proceeding of the tenth ACM international conference on Multimedia*, 2002.

[11] Sheng Thang, Yong-Dong Zhang, Jin-Tao Li, et al, Rushes Exploitation 2006 By CAS MCG. In *proceeding of TRECVID Workshop*, USA, 2006.