Personal Movie Recommendation Visualization from Rating Streams

Kodzo Webga *   Aidong Lu †

University of North Carolina at Charlotte

ABSTRACT

Online ratings from consumers are important in improving user shopping experiences at almost all popular e-commerce stores. This work discusses the challenges of providing effective personal visualization of movie recommendation based on personal watch history and on-line rating streams. We present our current visualization design which layouts movies based on their statistical features, recommendation results from collaborative filtering, and the rating records. We have developed a prototype system and provided several example results to demonstrate our approach for users with different profiles. We expect that the approach of personal visualization of movie recommendation can be extended to many similar applications in online e-commerce stores and online social networks.

Index Terms: K.6.1 [Computer Graphics]—Application K.7.m [Information Interfaces and Presentation]—User Interfaces

1 INTRODUCTION

This work studies the problem of movie recommendation for individual users based on their personal watch history in on-line rating datasets. Personal movie recommendation is one popular application of rating streams and it can be extended to a number of similar scenarios, such as music and book recommendations. In fact, online e-commerce stores are widely available and have formed an important part of our everyday lives. The online stores often collect ratings from users and recommend similar items according to user profiles; however they seldom provide visualizations for users to browse and search recommendation results. The personal visualization of recommendations can help users digest such a large amount of data produced by our everyday life and enable individual users to organize and explore movies based on their personal watch experiences.

While recommendation problems are common in online e-commerce stores and research fields of data mining and machine learning, personal visualizations of movie recommendations are rare. Movie datasets such as Internet Movie Database (IMDb), have been visualized using graph visualization methods [7, 10, 12], which are capable of providing advanced visualization and visual analytics functions. However, they are not designed for personal visualization of general users. In the following, we describe three key challenges for personal movie recommendation from online rating datasets.

- Personal visualization - Different from visualization for professional analysts, personal visualization of movie recommendation should be intuitive for general users to understand.

- Visual recommendation - Providing advanced yet simple visual analytics functions for users to gain insights on the background recommendation mechanisms and explore movie recommendations.

- Online streaming - Handling real-time rating streams at unexpected speeds, data amounts and filtering relevant information for local machines to process and store. Both the results of a recommendation and the recommendation visualization should be adjusted across time as well.

To develop a practical personal visualization system of movie recommendations, we also expect several additional challenges as follows. As the rating datasets are collected from individual users, this application has a potential privacy issue for users who prefer not to publish their user profiles but are willing to share their ratings. There can be some extreme cases requiring time-critical recommendations, especially when the new products are available to sell and the amounts are limited. As rating records often come with review comments, text analysis should also be considered in the recommendation visualization. At the end, the ambiguity issues raised from the recommendation algorithms may affect the exploration of recommendation results significantly.

In this work, we present our current visualization design for personalized movie recommendation toward addressing the three key challenges. We start by describing the application requirements and our design considerations. Then, we present our approach of embedding statistical features of movie ratings and recommendation results on a 2D layout, which is straightforward for general users to connect the relationships of movies and ratings. Specifically, each movie is located by the average rating separated by the rating columns and a distance measurement from a singular value decomposition spaces (SVD) for statistical data features. We have developed a prototype system for handling real-time rating streams and integrated interactive exploration functions for users to visualize the viewing history of the user, recommended movies, and a succinct overview of the entire movie database. We use example results for users of different watch profiles to demonstrate key interaction functions our prototype system can provide for a personal movie visualization.

The remainder of the paper is organized as follows. We first present the related work on personal visualization and recommendation visualization in section 2. Section 3 describes the application requirements and our visualization design. Section 4 provides example results and discusses our approach from different aspects of design considerations. Finally, we conclude this work and describe the future work in Section 5.

2 RELATED WORK

We briefly describe the related work of the personalized movie recommendation from the aspects of personal visualization and recommendation visualization.

2.1 Personal Visualization

Huang et al. [8] provided a survey of personal visualization and personal visual analytics (PV&PVA) from related VIS and HCI fields and they developed a taxonomy of design dimensions for a coherent vocabulary for PV&PVA. In addition, several recent work has been devoted to personal visualization. For example, Choe et al. [3] mapped the visual annotations for several insight types and discussed four areas for the design of personal visualization systems.

*e-mail: kwegba1@uncc.edu
†e-mail: aidong.lu@uncc.edu

Wang et al. [15] presented three distinct personal visualization designs for visualizing Facebook user data and suggested the importance of an illustrative design. Our design of the personalized movie recommendation follows the guidelines from the previous work by providing a simple yet flexible visualization design.

2.2 Recommendation Visualization

Recommendation algorithms have been applied to quite a number of applications related to personable visualization, such as for images [5], music [2], and movies. Most approaches are from the fields of data mining and machine learning, where network visualizations are often adopted. For example, Luo et al. [11] used hyperbolic and multi-modal view to visualize the recommendation list. Kermarrec et al. [9] used SVD-like matrix factorization and PCA for global mapping of movie ratings from high dimensions to a two-dimensional space. Crnovrsanin et al. [4] proposed a task-based and information-based network representation for users to interact and visualize a recommendation list. Vlachos et al. [14] used bipartite graphs and minimum spanning trees to explore and visualize recommendation results of a movie-actor dataset. Gretarsson et al. [6] visualized recommended users in social networks with node-link diagram and grouped relevant nodes on the recommendation list in parallel layers. Our approach also adopts the node-link diagram on a newly designed two-dimensional space (SVD-based distance / rating) for general users of a movie recommendation visualization.

3 PERSONAL VISUALIZATION OF MOVIE RECOMMENDATION

We start with describing several aspects of the personal visualization for designing our approach. We then describe our current approach in details.

Integration of automatic recommendation algorithms and intuitive visual interface. To provide a personal visualization which is intuitive for users to understand and powerful to achieve advanced analysis functions, automatic recommendation algorithms should be integrated in the visual interface. It is preferred that some user interaction functions for exploring recommendation results or adjusting recommendation preferences are integrated with the background recommendation algorithms to reveal additional insights to the movie dataset.

Personalized visualization. As users may have very different profiles and tastes on movie selections, the visual interface should be highly customizable for adjusting the visual parameters as well as levels-of-details. It is ideal to vary the visualization by adjusting different aspects ranging from simple parameters such as the genres of movies, released years, average ratings, the number of recommendation items, to complex features of movies and recommendations according to user watch history.

Simple and meaning visualization. The visualization should be designed for general users who may not have previous visualization experiences. It is ideal that the visualization and interactive functions are self-explanatory. Such visualization could replace the recommendation list by providing insights between the recommended items and movies a user has watched through several simple interactions for specifying personal tastes.

We design our personal visualization of movie recommendation based on the considerations described above. For recommendation algorithms, we adopt a popular collaborative filtering approach using a singular value decomposition (SVD) [1]. The recommendation list is computed through finding movies with high ratings from users with similar profiles.

Our visualization approach modifies the node-link diagram on a 2D panel designed for personal recommendation with rating streams. Each movie is mapped to the panel according to the overall average rating for its horizontal position, and a distance between the movie and the user in selected SVD space for its vertical position. As shown in the examples in the next section, the movies closer to the top of the space may possess stronger relationships to the user and the movies closer to the right of the space have higher average ratings.

The details of the visual parameters are set as follows. All the movie nodes are presented using circular shapes and user nodes are hidden. The size of a node is based on the number of users who have rated it. Thus larger movie nodes indicate more popular movies. The nodes are colored based on the movie category. The movies that have been watched by the user are visualized with a doughnut around the nodes and the color of doughnut is greener for a higher rating value.

To compute the recommendation list, we first map each user into the SVD feature space by multiplying the adjacency matrix of rating records M by the V component of SVD results. We then compute the cosine similarity between the active user and all other users using the results of $M \times V$. Next, we automatically select the list of users that have a positive similarity with the current user. We choose the movies those users rated positively ($ratings \geq 3$) but have not been watched by the current user. To provide a strong connection of those movie nodes in the visualization space, we compute the minimum spanning three of a complete graph of the recommended movies using Prim's algorithm, as shown in Figures 1 and 2.

We have developed a prototype system with two components. One is a server component that handles online rating streams and processes time consuming operations such as SVD decomposition. The other is an interface for users to visualize and analyze movie recommendations. The server component is built with Windows Communication Foundation (WCF). We used the incremental SVD package [1] to decompose real-time rating streams. Note that the details of the server description is skipped due to the paper limit and our concentration on personal visualization interface.

We have also developed several interaction functions to filter the movie nodes by the movie category, the release date, the similarity with a selected movie, and the number of recommended movies. The user can also explore these movies by clicking on IMDb page link that would take him or her directly to the IMDb page of the specific movie for further information.

4 RESULTS

We have tested our prototype system using the MovieLens100K dataset [13]. The dataset has 100K ratings from 1-5 and 1,682 movies from different categories (or genres) rated by 943 users. The ratings are from September 19th, 1997 through April 22nd, 1998. Each user has rated at least 20 movies during the entire period. The following shows several example results of our approach for users with different profiles.

4.1 Example Results

The first set of results shown in figure 1 demonstrated the variety of relevant movies for the same user in different features space of the SVD. It also highlights the personalized taste of each user.

As shown in the top left image in figure 1, we can observe in the first dimension of SVD that the first user does not like comedy movies but enjoys majority drama and romance movies. For example, the user likes the comedy-romance movies "Truth About Cats and Dogs, The 1996", and "Sleepless in Seattle (1993)". We can find a similar movie - the drama-romance movie "Phenomenon (1996)" - with similar average rating and a similar distance from the user. However, in the second and fifth dimensions of the SVD features space shown in the bottom left image in figure 1, the result reveals different set of movies that are closer to the same user. The romance and drama movies dominate this set as well. Other than the comedy-horror movie "Yang Frankenstein (1974)" that is

the closest to the user, we can find movies such as comedy-romance movie "Annie Hall (1977)", drama movie "Gandhi 1982", and other drama and romance movies the user have rated with high scores.

The second user in the first SVD dimension space (top right image in figure 1) does not have a preference over any movie genre. From the movies the user has rated, there is no consistency in ratings among any genre of movies. For example, we can see that the user like horror movie "Shining The (1980)", the comedy-romance movie "Sense and Sensibility (1995)", and the drama movie "To Kill a Mockingbird (1962)". However, the user does not like the drama-romance movie "Phenomenon (1996)" nor drama movie "Trainspotting (1996)" etc. Overall, the user does not like comedy movies according to the rating records. This user profile is also represented by the diversity of the genre of movies in this SVD features space. A similar result can be seen in the second and fifth SVD features space (bottom right image in figure 1). In this space, the movies the user has rated spread out vertically in the 2D panel and are mixed with movies the user has not watched yet. The user may use our visualization to explore recommended movies. For example, the user may like the drama movie "Margaret's Museum (1995)" since it has a similar average rating value and the same distance from the user with the drama movie "One Flew Over the Cuckoo's Nest (1975)" the user likes. Our results demonstrate that personalized recommendation visualization is unique to each user.

The second set of results is for the interaction and streaming of rating records, as the recommendations need to be updated through time as well. Figure 2 shows our visualization of a user in different time frames. The first top left image shows the result at the initial time frame and the bottom left image shows a result of a set of interactions the user has performed. The user first set the movie similarity slide bar (not shown in the image) to 60 percent. Then, the user selected a reference movie to the drama movie "Field of Dreams (1989)". This indicates that the user is interested in movies with positive ratings by at least 60 percent of similar users who also like the movie "Field of Dreams (1989)". We can see that the system presents a number of movies fitting the taste of the user. Among which, there are three movies the user has not watched yet: a drama movie "Mr. Holland's Opus (1995)", a drama movie "It's a Wonderful Life (1946)", and a comedy movie "Birdcage (1996)". The two images on the right of figure 2 show the results of the same user at the 10th time frame and the result of similar user interactions. These results contain more movies compared to the previous time frames. Other than the three movies the user has not rated yet at the initial frame, two additional movies are in the list for the user to explore: drama movies "Phenomenon (1996)" and "Time to Kill, A (1996)".

5 CONCLUSION AND FUTURE WORK

In this work, we discuss the challenges for personal visualization of movie recommendation based on online rating streams. We provide our current visualization approach which is based on a recommendation analysis with collaborative filtering and interactive visualization on a newly designed 2D space. We demonstrate our approach with a prototype streaming system and example results for users with different profiles. Our approach can be applied to the recommendation visualization of other products with individual rating records, such as music and images.

It is challenging to design intuitive visualization for general users. While our current design requires additional improvements and user studies for its effectiveness, we preserve the similarity between movie nodes and the user from the SVD spaces. Also, we transform the locations of movies and users from high-dimensional SVD spaces to a 2D panel, which is easy to understand and interact. In the future, we will continue to explore other options to improve our design.

We will also continue to investigate SVD approaches to re-

veal additional insights of the statistical data features, as SVD approaches are widely used in automatic recommendation algorithms. We are interested in developing semantic visual analysis approaches for personal visualization, as it may significantly improve the intuitiveness of visualization design for general users. We also plan to expand our streaming platform to take online streams and develop heterogeneous multi-processing solutions to handle rating streams of various types.

REFERENCES

[1] C. Baker, K. Gallivan, and P. Van Dooren. Incpack software: Low-rank incremental svd methods.

[2] D. Bogdanov, M. Haro, F. Fuhrmann, A. Xambo, E. Gomez, and P. Herrera. A content-based system for music recommendation and visualization of user preferences working on semantic notions. In *Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop on*, pages 249–252, June 2011.

[3] E. K. Choe, B. Lee, and M. Schraefel. Characterizing visualization insights from quantified selfers' personal data presentations. *Computer Graphics and Applications, IEEE*, 35(4):28–37, July 2015.

[4] T. Crnovrsanin, I. Liao, Y. Wuy, and K.-L. Ma. Visual recommendations for network navigation. In *Proceedings of the 13th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'11, pages 1081–1090, 2011.

[5] J. Fan, D. Keim, Y. Gao, H. Luo, and Z. Li. Justclick: Personalized image recommendation via exploratory search from large-scale flickr images. *Circuits and Systems for Video Technology, IEEE Transactions on*, 19(2):273–288, Feb 2009.

[6] B. Gretarsson, J. O'Donovan, S. Bostandjiev, C. Hall, and T. Höllerer. Smallworlds: Visualizing social recommendations. *Computer Graphics Forum*, 29(3):833–842, 2010.

[7] B. Herr, W. Ke, E. Hardy, and K. Borner. Movies and actors: Mapping the internet movie database. In *Information Visualization, 2007. IV '07. 11th International Conference*, pages 465–469, July 2007.

[8] D. Huang, M. Tory, B. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *Visualization and Computer Graphics, IEEE Transactions on*, 21(3):420–433, March 2015.

[9] A.-M. Kermarrec and A. Moin. Data Visualization Via Collaborative Filtering. Research report, Inria, Feb. 2012.

[10] J. Lee, G. Noh, and C. kwon Kim. Analysis amp; visualization on movie's popularity and reviews. In *Big Data and Smart Computing (BIGCOMP), 2014 International Conference on*, pages 189–190, Jan 2014.

[11] H. Luo, J. Fan, D. A. Keim, and S. Satoh. Personalized news video recommendation. In *Advances in Multimedia Modeling*, pages 459–471. Springer, 2009.

[12] B. Nemeth, G. Takacs, I. Pilaszy, and D. Tikk. Visualization of movie features in collaborative filtering. In *Intelligent Software Methodologies, Tools and Techniques (SoMeT), 2013 IEEE 12th International Conference on*, pages 229–233, Sept 2013.

[13] G. Research. Movielens100k: Movie rating dataset.

[14] M. Vlachos and D. Svonava. Recommendation and visualization of similar movies using minimum spanning dendrograms. *Information Visualization*, 2012.

[15] S. Wang, Y. Tanahashi, N. Leaf, and K.-L. Ma. Design and effects of personal visualizations. *Computer Graphics and Applications, IEEE*, 35(4):82–93, July 2015.
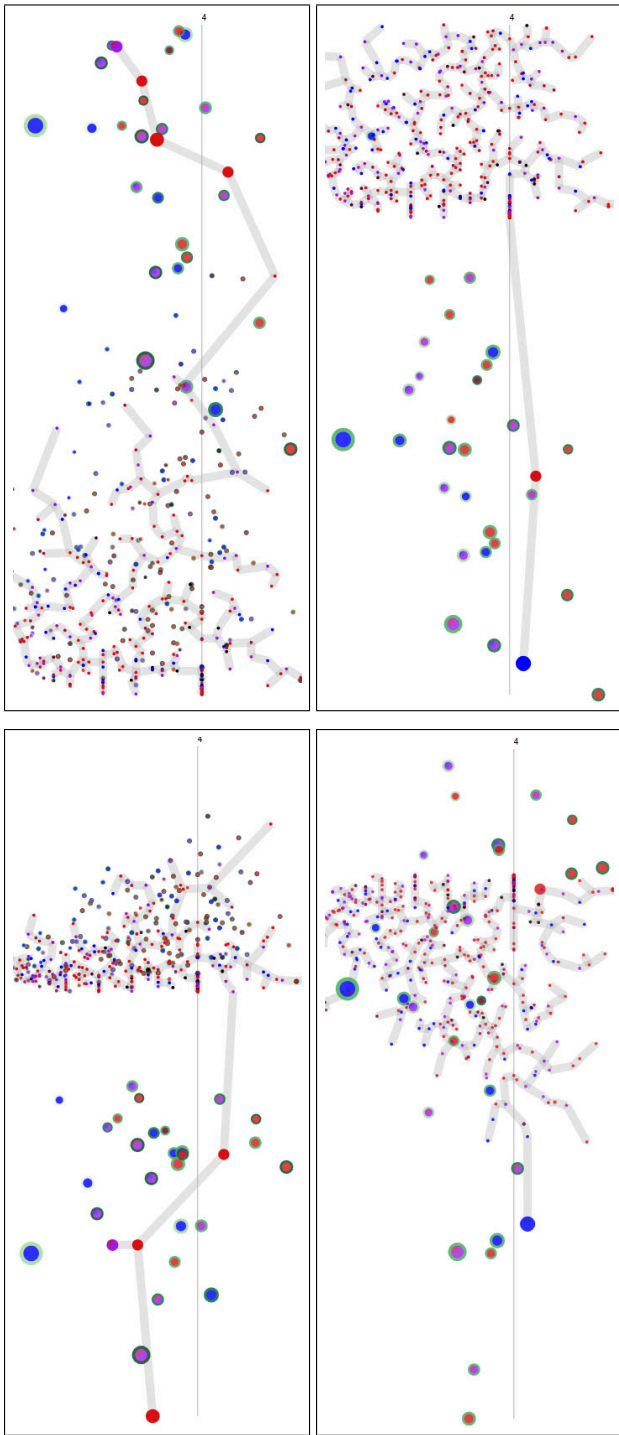
Figure 1: Example results of two users in different SVD dimensions. The images on the left show the results by using the first dimension of SVD spaces. The images on the right use the second and the fifth SVD dimensions. These dimensions can be interactively selected in our prototype system and can be improved with automatic suggestions. The recommended movies are highlighted with the tree paths.
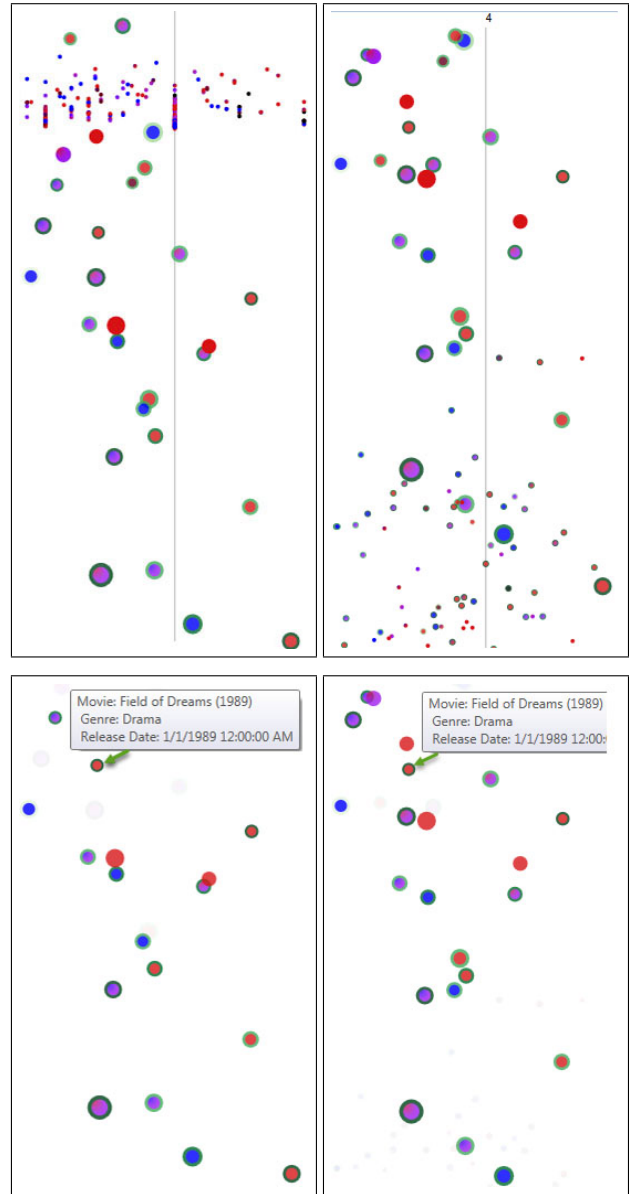

Figure 2: Example results of interaction and rating streams. The two images on the left show the results from two different time frames and the two images on the right show the interaction results. Only portions of the results are shown due to space constraint.