The Application of Visualization Techniques in Recommendation Systems

Dong Qian, Cheng Yang, Chen Li
Science and Technology Department
Communication University of China
Beijing, China

Abstract— The personalized recommendation system has been widely and maturely applied to various domains from social network to items recommendation such as videos, music, movies, books and online shopping. In the meantime, the information visualization technology based on big data has a substantial development. Considering the common based on big data, this paper discussed the connections of graph visualization and recommendation system from a new perspective. We propose three combinative points: Firstly, the Visual Analytics of social network can be combined with the recommendation computing; secondly, the graph drawing algorithms can be combined with the terminal presentation of recommended results. On the account of the evolitional nature of recommendation, the dynamic graphs would be the key area; thirdly, user interaction with the recommendation system can use the graph interaction strategies to achieve a more accurate and understandable result. In this paper, we discuss the visual techniques which can applied to recommendation system from a more general perspective. We also discuss some research challenges in the future.

Keywords- Recommendation Systems; Dynamic graph; Graph Visualization; Interaction methods.

I. INTRODUCTION

The visualization techniques reveal complex and multidimensional data decently. Thus help people extract hidden useful information in the data and make right decisions. The main goal of personalized recommendation systems is finding interested users or items from a vast amount of data set and recommending them to the current user. The Automated Collaborative Filtering (ACF) is a recommended technology which have been used widely. However, most of the exiting ACF- recommendation systems are remain black-box for users. As a result, users and the recommendation systems are separated from each other, which would affect the results of the recommendation. How to make users understand the recommendation and even involved in the process? The visualization technology can be used to solve this problem.

At present, visualization technology has applied to some recommendation applications. SmallWorlds [1] recommends interested items to users based on the Facebook API. They show users the recommendation process by a visual interactive interface. A user study indicates that visualization technology improve the accuracy of results and increase users’ satisfaction of the recommendation system. Peerchooser [2] allows users manipulate their recommended data at varying levels of granularity to reflect their current needs by an interactive interface.

Currently, there is no systematic study about the visualization technology which can be applied to the recommendation systems. This paper proposes three research points of recommendation systems combined with visualization from a general perspective, and illustrates these points in details.

The rest parts of this paper is organized as follows. Section II discusses some definitions and classifications of visualization technology are given. Section III shows the visual analytics combined with recommendation. Section IV described graph layout algorithms which can be used in the recommendation interface. We elaborate user interaction of visualization and its links with the recommendation systems in Section V. Finally, conclusions and future challenges are discussed.

II. BASIC DEFINITIONS AND CLASSIFICATIONS

A. Definitions

Graph is an important notion in visualization. In graph theory, a graph is a representation of a set of points (nodes/vertices) connected by lines (arcs/edges). Generally, a graph can be represented as $G = (V, E)$, $V$ as vertices and $E$ as edges. In computer science, a graph is a data structure of a set of nodes together with a set of edges of these nodes. Moreover, some node or edge value can also associate to each node or edge of a graph data structure, which means a lot to depict the real world data.

In the most common sense, graphs are often sorted by directed graph and undirected graph. A directed graph has a set of vertices and edges and each edge can be denoted as $e = (v_1, v_2)$. It differs from an undirected graph, in the latter is defined in terms of unordered pairs of vertices. Generally, directed graph which has weighted edges is named network.

However, the real world data structures are usually complex with non-trivial topological features, which cannot be depicted as a simple graph. As a result, a notion of complex network is proposed to represent a graph with non-trivial topological features, which has been applied to many research areas such as physics, biology, telecommunications, sociology, epidemiology, and others.
As a network can be seen as a complex graph, which makes no difference in recommendation systems, we would not deliberately distinguish these two notions in the rest of this paper.

B. Classifications

Graphs can be classified in various ways. From the structure perspective, graphs can be classified as trees, generic graphs and compound graphs. A tree is an undirected graph which connected without simple cycles. And a rooted tree is often hierarchical. The degree of a vertex is represented by its length to the root vertex. A compound graph is defined as a graph and a rooted tree which uses same vertices.

As graphs may evolve over time, it can also be classified as static graphs and dynamic graphs. Static graph visualization focuses on producing a better layout while the dynamic graph visualization mainly research on representing the evolution of relationships between entities in readable, scalable, and effective diagrams [3].

III. VISUAL ANALYTICS AND RECOMMENDATION COMPUTING

Visual Analytics is the science of analytical reasoning which supported by visual interfaces [4]. Visual analytics involves human interactivities and help users to get an insight into data as to make more effective decisions. [5].

Generally, the set of users and items can be depicted as nodes in a graph, and their relationships can be depicted as edges. We have studied the techniques of visual analytics and concluded three parts of them which can be applied in recommendation data process – graph preprocessing, topological structure analysis and trend tracking. We elaborate the three parts respectively in the next of this section.

A. Graph preprocessing

The data source of recommendation systems is usually an unwieldy mass of data, which makes it impossible to visualize all the data in a single graph. Even if there are enough screen spaces for visualizing, it remains overwhelming to users. Hence, we can use graph preprocessing techniques to make data reduction so as to simplify the subsequent process. Graph filtering and graph clustering are two main methods of graph preprocessing.

Graph filtering is a set of algorithms which select some nodes/edges and remove them from the graph. The selection can be stochastic or deterministic [6]. Stochastic filtering selects nodes and edges at random while deterministic filtering selecting based on some attributes or topologic values. There are more details in [7] and [8] (see Figure 1).

Graph clustering is the process of merging nodes and edges in small groups or classes to reducing the size of the original graph. The clustering based on graph structural information is called structure-based clustering. The clustering which uses semantic data of a graph is named content-based clustering [9]. However, these two techniques can be combined together to produce a better reduction.

In recommendation systems, the users set and the items set can be assigned to two layers, and establishing connections between them. The preprocessing can be done in each layer separately. For instance, the users can be clustered into communities based on similarity, and the items can be aggregated according to their categories, such as movie genres, music types, etc. After the clustering within layers is completed, edge bundles can be applied to the links between the two layers. Thus the vast data set can be simplified.

B. Topological structure analysis

As the recommendation data source displayed in a graph, not only the semantic structure, but especially topological structure of the dataset can also be clearly studied.

Collaborative filtering is a conventional recommendation method, which makes automatic filtering about the interest of current user based on collecting the items’ ratings from many users. Collaborative filtering algorithms can compute user-user similarity [10] or item-item similarity [11] to make recommendation [12]. However, for a new user, the collaborative filtering faces a cold-start problem – there is no history item rating.

We can use a new way to tangle the cold-start problem by analyzing the topological structure of social network. A famous phenomenon in social network analysis named “birds of a feather flock together” [13], which means that users tend to make friends with similar people. As a result, we can measure user similarity by using the topology of social network [14]. For example, SFviz [15] uses cosine similarity value to calculate the user similarity.

The process to provide item recommendation to a cold-start user can be summarized as below (see Figure 2).

If a new user is registered without history item ratings in the dataset, other information from his profile is used to calculate the topological similarity and find similar people with history item ratings. Then the recommendation system recommends the similar users’ items to the new user.
C. Trend tracking

T Takaffoli M [16] classified the works on evolution of dynamic networks into three categories: microscopic, macroscopic and mesoscopic approaches. The microscopic approaches study the evolution at the level of nodes and edges, such as preferential attachment phenomenon [17] or the nodes/edges creations [18]. The macroscopic perspectives focus on the evolution of the substantive characteristics of networks, for instance, the evolution of nodes distribution, community, and nodes correlation of networks [19], or the shrinking diameter of social networks [20]. The mesoscopic approaches predict the trend of networks based on an intermediate structure of the networks.

In recommendation systems, the users’ behavior and preferences will change over time, which result in the graph evolution. The definition of dynamic graphs has been given in Section II. Therefore, the feature tracking technology of dynamic graphs can be used to track the users’ behavior and preference changes. Then the updated recommendation results can be produced and accuracy of the whole system would be improved. We illustrate the strategy from following two aspects.

- The importance of user nodes is measured by centrality [21] or ranking [22]. The edge amounts between user-user nodes can also be used to calculate the importance of user node. Then, various weights assigned to different users according to the importance. If other important users preferences change in the community of current user, the preferences of current user may suffer the same changes in the future. In this way, the system can predict the preferences of the user even before itself.

- The manipulations of adding or deleting links with a single node can be mapped to the item clustering. Thus we can speculate the users’ attitude towards some item categories and update the recommendation accordingly. This method should take into account the user interactions, which will be discussed in Section V.

IV. GRAPH LAYOUT AND RECOMMENDATION INTERFACE

How to present the recommendation results to the users is a critical issue. We will not expound the specific user interface here, but discuss the terminal graph layout problem. The graph layout problem is also a core research area of graph visualization.

A. Aesthetic criterions

Generally, there are different aesthetic criterions for layouts of static graphs and dynamic graphs respectively.

The criterions of static graphs include minimizing the edge crossings, avoiding the node overlaps, minimizing the total drawing area, maximizing symmetries [23].

Dynamic graphs evolve over time. This brings an additional aesthetic criterion which named “preserving the mental map” [24], which also called “dynamic stability” [25]. Misue K [24] proposed three dynamic models: the orthogonal order, the proximity, and the topology models.

B. Graph layout algorithms

Force-directed layout is a typically and widely used static layout algorithm. The Kamada-Kawai model [26] and the Fruchterman-Reingold model [27] are the two most classical force-directed models. The Kamada-Kawai algorithm can produce a layout from an random initial position, the Fruchterman-Reingold force-directed algorithm is more quickly but more sensitive to the initial layout. Hence, some researchers use Kamada-Kawai algorithm in conjunction with Fruchterman-Reingold algorithm to produce a fast and aesthetic layout [28].

However, the Kamada-Kawai model and Fruchterman-Reingold model only performs well for lightweight graphs, but it may not handle larger and denser graphs decently. On account of this problem, Harel [29] proposed an algorithm which uses multiple scales to minimize the time of convergence. Hachul [30] developed an algorithm of multiple levels which can achieve the same speed of the algorithms of single level. Andersen [31] classified the graph edges into local edges and global edges to enhance the force-directed method. Chan [32] applied the Fruchterman-Reingold model on different layers.

The layout algorithms of dynamic graphs can be divided into two categories: online layout and offline layout. When computing each layout of the time sequence, the online dynamic layout considers just the previous graphs in the sequence while the offline dynamic layout taking all the graphs into account.

The offline layout offers the future information of the graph layout. Obviously, the offline model is not suitable for the recommendation systems as we having no information about the future graphs.

There is more research on the online layout model [33] [34]. S.C. North [25] developed methods for drawing online directed acyclic graphs. C. Gorg [35] proposed algorithms for orthogonal and hierarchical graphs. Frishman Y researched on the dynamic clustered graphs [33] and proposed an online algorithm based on force-direct layout techniques and the whole graph structure[28]. Moreover, Frishman Y also computed layouts of online graphs by GPU. They applied a variable degree for the move of nodes [36]. Brandes [37] proposed an algorithm which used Bayesian to generate online dynamic graphs. Recently, Feng K C [38] formulated the layout of time-varying graph as an optimization problem.

V. USER INTERACTION WITH RECOMMENDATION INTERFACE

Interaction is a main part of visualization technology. Applying interaction techniques in the recommendation system can not only help users understand the recommendation process, it also can help the recommendation systems understand the users’ needs in turn. Yi J S [39] summarized the interaction techniques as selecting, exploring, reconfiguring, encoding, abstracting, filtering, and connecting. Considering
the interaction techniques can be used in the recommendation systems, we conclude several principle techniques into two categories: graph exploration and data manipulation.

A. Graph exploration

1) Zooming and panning: Zooming techniques can be two forms: geometric zooming and semantic zooming. Geometric zooming simply enlarges the graph. But semantic zooming would change the information displayed and add more details to the enlarged area. In this sense, it is more difficult to decide what and where to add rather than the operation itself.

However, zooming techniques may create additional cognitive load for users due to the temporal separation of views.

2) Focus + context: Compared to zooming and panning, focus + context will allow users to focus on details as well as keeping the context displayed. The relative positions of nodes and edges is critical for a comprehensible focus + context keeping the context displayed. The relative positions of nodes and edges will allow users to focus on details as well as keeping the context displayed. The relative positions of nodes and edges is critical for a comprehensible focus + context visualization [38].

There are several strategies for focus + context displaying: distortion to select the scale of items within a given interest value, removal of the non-focused items from the current screen, and wide angle lens of the focal area displaying.

B. Data manipulation

Data manipulation allows users to interact with graph data (nodes and edges) according to their current interests.

1) Single data editing: Users can delete or insert nodes and edges by the visual interface. These editing actions will affect the topological structure of the graph and change layout. However, the layout style should be maintained and the topological change should be trivial.

2) Clustering data editing: From a coarser granularity, users can add or remove the data clusters. Archambault D. [40] presents different views of the graph structure (see Figure 3).

Figure 3. Figure 3 Example of clustering data editing, a to b add a node clustering while c to d delete a node clustering. From [40].

Furthermore, it is more convenient for users to manage their items according to changed tastes.

VI. CONCLUSIONS AND FUTURE CHALLENGES

Visualization technology can assist the recommendation systems to achieve a more accurate result and improve the satisfaction of users. In this paper, we discussed the connective research points of recommendation systems and visualization technology from three aspects: visual analytics, graph layouts and user interactions.

However, there still are some unsolved problems and open challenges in this crossing field. We proposed three research directions here.

1) Reduction of the time complexity: Recommendation system needs to update data fast to support the real-time interaction with user. The optimization of current algorithms need more research to minimize the time complexity.

In addition, the parallel computing of Graphics Processing Unit (GPU) can be used to improve the efficiency of visualization algorithms. We can compute the complex algorithms on the GPU in a parallel way. Each pixel has a kernel program runs on GPU. [36] Frishman Y. [36] introduced the static force-directed layout on the GPU, and the dynamic layout implementation is discussed in [28].

However, the main challenge is how to map a unstructured data set to a structured graph.

2) Real-time interaction with data: Users need to get feedback rightly after an interactive action. Thus, the major challenges are to synchronize the data processing and visualization interactions effectively, and seamlessly connect the recommendation process and the visualization interface.

3) Innovation of user interaction in the multidisciplinary field: How to produce a better recommendation result and optimize the interactive interface can be concluded into the problem of how to performing better human-computer interaction. The understanding of graph involves the cognitive science, sociology, psychology and other interdisciplinary fields. In the future, research on the multidisciplinary field will be an open challenge.

REFERENCES


