Context-Aware Strategies For Recommender System On A Professional SNS Platform

Xianghui Zhao, Hui Liu
China Information Technology Security Evaluation Center
Beijing 100085, China
zhaoxh@itsec.gov.cn, liuh@itsec.gov.cn

Wen Tian, Zhaopei Zeng, Lin Ye
Shanghai 30wish Information Security Co., Ltd
Shanghai 200232, China
stellatian@gmail.com, zengzp@30wish.net, yel@30wish.net

***Abstract*—The importance of recommender system on Social Network Services (SNS) platform has been recognized by researchers and practitioners in many disciplines, including e-commerce, information retrieval, social computing, data mining, marketing, etc. While a substantial amount of approaches focus on recommending the most relevant items to users on mainstream SNS platforms, there is still a lack of closer investigation into the context-aware strategies on professional SNS platform whose contextual information varies significantly from generic SNS platforms. Drawing upon existing algorithmic paradigms –content-based methods and collaborative filtering, this paper proposes context-aware strategies that cope with the need to recommend both users and items on a professional SNS platform. A case study has been demonstrated based on such approach and directions for future research have been discussed.**

Keywords-recommender system; social network; collaborative filtering; collective intelligence; professional SNS platform

I. INTRODUCTION

Traditionally, recommender systems deal with applications exerting only two types of entities: users and items, and do not combine them into a context when providing recommendations [1]. By fully utilizing contextual information, context-aware recommender systems can further improve performance accuracy and user satisfaction [2]. It has also become one of the hottest topics in the domain of recommender systems. In addition, the pervasive adoption of social network systems and the use of Web 2.0 technology have generated large amounts of multimedia-enriched data that are enhanced both by explicit user-provided annotations and implicit aggregated feedback describing the personal preferences of each user [3].

However, contextual information varies across Social Network Services (SNS) platforms. For example, generic social networking sites focus on recommending friends who share similar social hobbies such as favoring similar types of photos, music, movies, books, etc. A professional SNS platform usually focuses on certain types of knowledge and there are many operated by enterprises. For instance, Microsoft SharePoint has over 100 million users and offers social networking system capabilities, including the ability to selectively establish relationships between contacts and to grant permission to view selective information posted by users; users can create a profile page that includes their job location, experience, and contact information [4]. The purpose of this kind of SNS platform is to foster knowledge sharing and assist knowledge management among staff members. Unlike generic SNS platforms, the key of professional SNS platforms is to establish connections that could help people locate content, knowledge or experts. Therefore, it is crucial to develop recommender systems with a better awareness of the contextual information on relevant issues.

This paper aims at suggesting several context-aware strategies for recommendation systems on a professional SNS platform. First, algorithmic paradigms of recommendation systems are reviewed, while their limitations and challenges are identified. Second, context-aware strategies are proposed, with the demonstrations from a professional SNS platform. Last, directions for future research are discussed and conclusions are drawn.

II. ALGORITHMIC PARADIGMS OF RECOMMENDER SYSTEMS

Since users provide a rich set of information that can be converted into intelligence, researchers have investigated into two main sources of information that can be harvested for intelligence: the content-based information (e.g., keywords or phrases occurring in the item), the collaborative-based information (e.g., the interactions of users captured by collaborative filtering engine or other agents) [5]. Therefore, two types of algorithmic paradigms are heavily used for developing recommender system on SNS platforms: 1) the content-based methods, and 2) the collaborative filtering.

A. Content-based methods

Content-based methods analyze the content to build a representation usually in the forms of terms or phrases (multiple terms in a row) which appear in the document [5]. In content-based recommendation methods, the utility of item *s* for user *c* is estimated based on the utilities $u(c,s_i)$ assigned by user c to items $s_i \in S$ that are similar to item *s [6]*. For example, in a movie recommendation application, in order to recommend movies to user c, the content-based recommender system tries to understand the commonalities among the movies user c has rated highly in the past (specific actors, directors, genres, subject matter, etc.); then only the movies that have a high degree of similarity to

Published by Atlantis Press, Paris, France.

whatever the user's preferences are would be recommended [6].

This approach has its roots in information retrieval [7] and information filtering research [8]. However, it has certain limitations such as the limited content analysis, overspecialization, and new user problem [6]. The limited content analysis refers to two problems: the insufficient automatic feature extraction (especially on multimedia data), and unable to distinguish different items if they are represented by the same set of features. Overspecialization refers to the problem that when the system can only recommend items that score highly against a user's profile, the user is limited to being recommended items that are similar to those already rated [6]. The new user problem happens when the user has to rate a sufficient number of items before a content-based recommender system can really understand the user's preferences and present the user with reliable recommendations [6]. Therefore, relying on content-based methods alone may not get accurate recommendations.

B. Collaborative Filtering

Collaborative filtering systems try to predict the utility of items for a particular user based on the items previously rated by other users [6]. In this way, the user will be recommended items that people with similar tastes and preferences liked in the past [6]. The fundamental assumption of collaborative filtering is that if two users rate several items similarly or have similar behaviors (e.g., buying, watching, listening), then they will rate or act on other items similarly [9].

However, there are also many challenges for collaborative filtering tasks. Collaborative filtering algorithms are required to have the ability to deal with highly sparse data, to scale with the increasing numbers of users and items, to make satisfactory recommendations in a short time, and to cope with other problems like synonymy, shilling attacks, data noise, and privacy protection issues [9].

C. Hybrid Recommenders

To alleviate the problems in each method, hybrid recommenders can be used. There are different approaches to integrate content-based methods and collaborative filtering to form hybrid recommenders. For example, content-based collaborative filtering recommender, content-boosted collaborative filtering recommender, and hybrid collaborative filter combining memory-based and model-based collaborative filtering algorithms [6, 9]. Nevertheless, each method still has certain disadvantages. Thus, it is important to form a set of context-aware strategies that will capture the information features so that better hybrid recommendation methods can be chosen.

III. CONTEXT-AWARE STRATEGY

On professional SNS platforms, the user interaction and the item generated are very contextual. How to choose appropriate recommender system algorithm? In this section, we provide some context-aware strategies that can address the problems in content-based methods or collaborative filtering. To demonstrate how to apply these strategies, a professional SNS platform is chosen as the background case for our study.

The professional SNS platform is initiated by us to support the main function of providing IT security evaluation services. One of the services is to provide information of vulnerability database to the public and related entities. Thus the China National Vulnerability Database of Information Security, known as CNNVD, was founded to execute such duty. CNNVD collects, analyzes, and delivers vulnerability information so that the public and enterprise users could be alerted from information security risks [10].

This professional SNS platform is aimed at involving professional IT experts and those who are interested in the area of Information Security to share and manage knowledge about the vulnerability database. Therefore, recommender systems are very helpful for users to locate experts and items such as vulnerability information, product information, security incident, patch information etc.

To be more aware of the context, we suggest four strategies as follows.

A. Comprehensive understanding of users, items and their relationships

As mentioned above, existing recommendation methods generate ratings that are based on certain aspects of users or items while do not take full advantage of available data. Therefore, the first strategy is to form a comprehensive understanding of users, items and their relationships.

Adomavicius and Tuzhilin [6] have proposed a basic framework to help us build a better understanding of all relevant information, in which they suggest us extend the attribute-based profiles to utilize more advanced profiling techniques.

In our case, the items on the professional SNS platform are heavily interwoven. The users could generate any of them and their links need to be detected. In order to capture the relationships among items and users, two steps can be performed: 1) item relationship structuration, and 2) user behavior structuration.

1) Item relationship structuration

The purpose of item relationship structuration is to capture the linkages and attributes of items so that when algorithms are designed, these linkages and attributes are represented and incorporated.

Figure 1 shows how item relationship structuration is conducted on the SNS platform chosen for this study. There are 16 items in the figure. The lines in the figure indicate connections among items with the type of relationships identified for each connection. There are massive amount of these items, if their relationships have not been analyzed, it will be difficult to formulate models or represent the items.

Published by Atlantis Press, Paris, France.

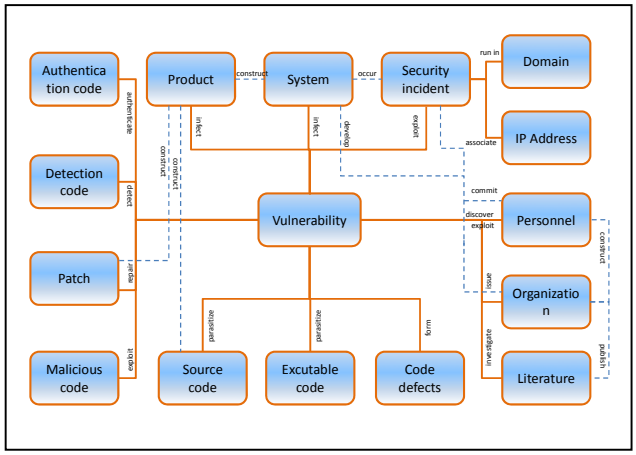

Figure 1. Example of item relationship structuration

2) User behavior structuration

User behavior structuration consists of three steps: categorizing user behavior, identifying how to represent the behavior, and collecting user preferences.

TABLE I. USER BEHAVIOR AND PREFERENCES

User Behavior Category	Representation Vector	Preference Collected
Rating	Rating on a scale of 1 to 5 or 10	Ratings can provide more accurate user preferences to the vulnerabilites.
Voting	Binary preference, the value is 0 or 1	Voting can provide more accurate user preferences to the vulnerabilities.
Forwarding	Binary preference, the value is 0 or 1	Forwarding can be used to predict preferences of users. Preferences of those being forwarded can also be inferred(inaccurate).
Following	Binary preference, the value is 0 or 1	Following can provide more accurate user preferences to the vulnerabilities.
Tagging	Some tags need word analysis before obtaining preference.	Tags can be viewed as an approximate representation of the user's understanding and emotion (like or dislike) in the corresponding domain.
Comment	Paragraph, need text analysis before obtaining preference	Comment can provide the user's emotion (like or dislike).
Click stream	The user's preference can be inferred Click stream toward the vulnerability.	Click stream can be viewed as an approximate representation of the user's attention which reflect one's preference.
Time spent per page	Time information, needs to be analyzed before obtaining preference.	Time spent on the pages can be viewed as an approximate representation of the user's attention or preference, but need to be denoised.
Downloading	Binary preference, the value is 0 or 1	Downloading can prove the user's interest in the project or associate projects.

To formulate better recommendation, user behavior must be analyzed so that patterns can be found. Table 1 provides an example to how to collect user behavior on the SNS platform. First, user behavior is categorized into 9 aspects: rating, voting, forwarding, tagging, comment, click stream, time spent per page, downloading. Then representation vector is assigned and users' preferences are collected.

B. Pay attention to new users and inactive users

New users and inactive users seem not to have sufficient behavior data. On the professional SNS platform, many users just browse the information rather than conduct any "behavior" that has been categorized in Table 1. Direct feedback is hard to obtained, and they have shown little preferences. As a result, new users and inactive users should be paid attention to. To make more accurate recommendations, their behavior can be categorized differently than normal users.

C. Identifying sparse and asymmetric data

In traditional recommendation methods, the amount of users and items are usually at the same order of magnitude. However, on the professional SNS platform, there is a substantial amount of sparse and asymmetric data. In the case of CNNVD, the numbers of product, system, personnel, organization are growing at a steady pace, whereas the numbers of vulnerability, patch, codes etc. are growing far more rapidly. Thus, the amount of items exceeds the amount of users to an extremely great extent. This will bring challenge to the dealing with sparse and asymmetric data. Adomavicius and Tuzhilin [6] summarized the disadvantage of sparse data and suggested a few ways to deal with the sparsity problem. But problems still remain as the item data and user data are in such an asymmetric position.

D. Be aware of dynamic and changing user preferences

On SNS platforms, the users' preferences do not stay unchanged, since the newly generated contents are constantly driving users' interests. When developing algorithms, it is common to assume that the users choose items based on their own interests. However, when recommendation systems are applied, the popped up items will boost users' interests. Unlike on generic SNS platforms, on the professional SNS platform, users have dynamic and changing preferences because the items are very time sensitive. Once a piece of vulnerability information has been detected and solved, the users' attention will shift away. Therefore, this type of dynamic must be considered in the algorithm development to main the level of accuracy of the recommendation.

IV. DISCUSSION AND CONCLUSION

As contextual information on SNS platforms becoming critical in improving the accuracy of recommendation systems, it is very important to develop context-aware strategies. This study suggests four context-aware strategies that are considered useful to cope with the challenges from a professional SNS platform. When developing hybrid recommenders, the attributes of users and items, the

activeness of users, the sparsity and asymmetricity of data, and the dynamic of user preferences should be considered.

As for future research, how to manage the effect of recommender systems and make adjustment to the algorithm are potential directions. In general, the effect of recommender system is two-folded: it could help SNS users discover new items and thus increase knowledge diversity, and it could also reinforce the popularity of already-popular items [11]. Knowledge diversity and knowledge sharing matter on professional SNS platforms. The context-aware strategies proposed study may shed light on how to improve the recommender systems on this type of platforms, and consequently stimulate knowledge discovery and knowledge sharing among SNS users.

ACKNOWLEDGMENT

This work is supported by the project of the State Key Program of National Natural Science Foundation of China (No. 90818021). Meantime, it is supported by a grant from the National High Technology Research and Development Program of China (863 Program) (No.2012AA012903). The authors are also thankful to the anonymous reviews for their constructive comments.

REFERENCES

[1] Adomavicius, G. and A. Tuzhilin, *Context-aware recommender systems*, in *Recommender Systems Handbook*, F. Ricci, et al., Editors. 2011, Springer US: New York. p. 217-253.

[2] Wang, L., X. Wang, and Y. Zhang, *Context-aware recommender systems* . Journal of Software, 2012. **23**(1).

[3] Konstas, I., V. Stathopoulos, and J.M. Jose, *On social networks and collaborative recommendation*, in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*2009, ACM: Boston, MA, USA. p. 195-202.

[4] Anderson, S. and K. Mohan, *Social Networking in Knowledge Management.* IT Professional, 2011. **13**(4): p. 24-28.

[5] Alag, S. and R. MacManus, *Collective intelligence in action*2009: Manning.

[6] Adomavicius, G. and A. Tuzhilin, *Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions.* IEEE Trans. on Knowl. and Data Eng., 2005. **17**(6): p. 734-749.

[7] Baeza-Yates, R. and B. Ribeiro-Neto, *Modern information retrieval*. Vol. 463. 1999: ACM press New York.

[8] Belkin, N.J. and W.B. Croft, *Information filtering and information retrieval: two sides of the same coin?* Communications of the ACM, 1992. **35**(12): p. 29-38.

[9] Su, X. and T.M. Khoshgoftaar, *A survey of collaborative filtering techniques.* Advances in Artificial Intelligence, 2009. **2009**: p. 4.

[10] CNNVD. *China National Vulnerability Database of Information Security*. 2012; Available from: http://www.cnnvd.org.cn/ .

[11] Fleder, D. and K. Hosanagar, *Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity*. Management science, 2009. **55**(5): p. 697-712.