The Appliaction of Recommendation Systems

Y.-F. Zhao
Institute of Smart City Systems
Qingdao Academy of Intelligent Industries
Qingdao, China
yfyf.zhao@163.com

H. Gao, Y.-S. Lv and G. Xiong
Qingdao Center for Intelligent Systems and Technology
Institute of Automation, Chinese Academy of Sciences
Qingdao, China
gaohang@casc.ac.cn, {yisheng.lv,gang.xiong}@ia.ac.cn

***Abstract*—This paper focuses on the application of recommendation systems through the perspective of best matching between users and items. After reviewing the traditional recommendation systems and their application areas, this paper then extends recommendation systems to some other areas which may need recommendation systems to solve some key problems. Not only the application areas are extended, but also we try to sublimate recommendation systems from a simple method to a universal methodology.**

Keywords—recommendation; transportation; social manufacturing; knowledge automation

I. INTRODUCTION

Thirty years ago, people could not get their merchandises easily, so they often need to go to big supermarkets; three decades passed, e-commerce brought millions of commodities into Internet, and people could almost get anything they need from Internet. Thirty years ago, people used CDs or tapes which only had 10~20 songs to spend their free time, and they could not easily get many songs for the expensive price; three decades passed, we are overwhelmed by tens of thousands of songs in Internet. Thirty years ago, people spent their weekends in front of TV which had only three channels; three decades passed, innumerable videos and films renewed in Internet continuously, and we could never finish watching them. The Internet did improve life. However, it made people face another problem which was so-called information overload, i.e. excessive information presenting at the same time made users be unable to obtain useful part, and this made the efficiency reduced instead [1]. People have to use a variety of strategies to make choices about what to buy, which song to listen, what film to watch, and even whom to date [2]. Recommendation systems (RS) were born to solve these problems, and has successfully applied into many areas such as e-commerce [3], music, movies [4, 5] and so on.

RS's task is to establish the binary relationship between users and items, then mine the potential and best suitable demand objects according to the existing selection process or similarity relations. Therefore, each user can be personalized recommended. In other words, when applied into an area, RS's essential mission is to find the best matching items for the users [6].

Thanks to the value of RS in the academic research and practical application, many research institutions and colleges (researchers) are devoted into RS, including New York University (Alexander Tuzhilin), the Grouplens in University of Minnesota (Joseph A. Konstan, John Riedl, etc), University of Michigan (Paul Resnick), Carnegie Mellon University (Jaime Callan), Microsoft Research (Ryen W. White) and so on. Besides, some phenomenon in the international academic world show the popularity of RS: 1) an international conference about RS was found by ACM; 2) the number of papers about RS algorithms increased year by year in the top meeting of data mining and machine learning, such as SIGCHI, KDD, SIGIR, WWW, etc.; 3) more and more papers about RS appeared in the higher-order journal of the international data analysis area, such as IEEE Trans. On Knowledge and Data Engineering and ACM Trans. On Information System. However, after investigating the papers about RS in these years, we found that most of them focus on algorithms and methods, methods on data processing, algorithms for similarity computation, strategies about recommendation, and so on. Few papers can be found to introduce the application of RS, and focus on researching areas which are suitable to use RS and how to seek new areas of RS application.

In this paper, we focus on analyzing the application of RS from the perspective of best matching. There are three points of necessity: 1) RS is born to solve the problems about information overload which exist in Internet catholically, and there are many other fields in Internet which will be faced with information overload; 2) besides Internet, there are many other areas with big data, such as financial and medical areas, which include much interference information; 3) erosion by Internet will cause the problem of information overload in many other fields. And throughout the RS algorithms in different areas, their same essential task is to finger out the information of useful and useless, right and wrong, to find the best matching item for the objective user. To our best knowledge, this is the first time to analyze the application of RS from the perspective of best matching, and other possible areas in which RS may also be applied. We try to extend RS from a method to methodology. This is the most important contribution of this paper.

II. RELATED WORK

In the last two decades, RS have been applied successfully into many different areas, and lots of classic methods appeared. In this section, we will give some representative methods and applications about the development of RS, such as emails, netnews, e-commerce and movies.

This work is supported in part by NSFC (Natural Science Foundation of China) projects (71232006, 61233001,61104160, 61203166, 4142055).

A. Emails

As the increasing amount of emails we received every day, we were often inundated by a huge stream of incoming documents [7, 8, 9]. This not only made reading ineffective, but also it was easy to miss important ones. So mailing lists is needed to handle large volumes of mails, enabling users to subscribe only to those lists of interest to them. And, the earliest RS we can find in literatures occurred in 1992, named Tapestry [10].

Tapestry is an experimental mail system developed by the Xerox Palo Alto Research Center. Different from traditional content-based filtering, the Tapestry system brought a new concept called collaborative filtering (CF), which is the most popular method in RS today. Collaborative filtering just makes use of people collaborating to help another perform filtering by recording their reactions to documents they read, namely, "recommend" to the objective users according to a group of other users. Fig. 1 shows the development of dealing with emails, and collaborative filtering came to the concept of recommendation. And, Tapestry tried to find the best matching emails according to the objective user's interest or demand.

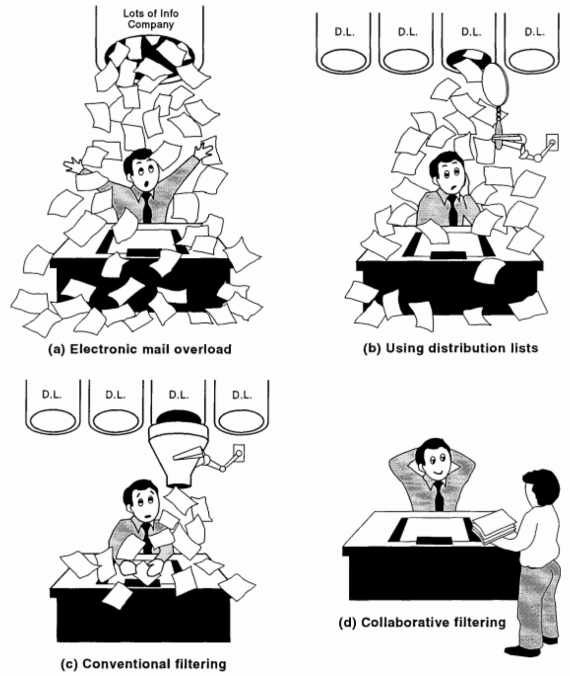

Fig. 1. Development of dealing with emails

Researchers in Google have also spent lots of time in applying RS into emails, and released priority inbox in 2010. And, they claimed that this personalized RS can help users to save 6% time in their paper "The Learning behind Gmail Priority Inbox". No doubt, saving 6% is a big progress in the time of timing being money, because the priority inbox could find the most matching ones according to the user's needs.

B. News

After the successful application of CF in emails, GroupLens enhanced its application in news [11]. GroupLens is a system for CF of netnews, to help people find articles they will like most, i.e. best matching between them, in the huge stream of available articles. As a banner RS in the early 90s last century, GroupLens is well-known as the real starting about user-based collaborative filtering.

GroupLens RS was personalized for news reading, born in 1994. That time, networks formed the interest groups which crossed geographical barriers, and bulletin boards were an important mechanism for that. The writer posted the article in a newsgroup, a public place available to anyone interested in the topic. Then the Usenet netnews system propagated the articles so that an article posted from anywhere in the world was available to everyone else. And over the years netnews had become a principal medium for sharing among computer users.

However, netnews was not completely satisfying for the users. Many users complained that the signal to noise ratio was too low. Except by the vocal few who post responses, writers cannot easily tell whether their comments are valued. Some seem not to care about reader interest, only about their own right to write. Moreover, tastes differ, so that no one article will appeal to all the readers of a newsgroup. Each reader ends up sifting through many news articles to find a few valuable ones. Often, readers find the process too frustrating and stop reading netnews altogether.

GroupLens gives a whole different mechanism, which is based on the hypothesis, i.e. people who gave a high score in the past articles are likely to like it again in the future. News readers display predicted scores and make it easy for users to rate articles after they read them, and GroupLens uses rating servers, called Better Bit Bureaus, gather and disseminate the ratings. The rating servers predict scores based on the heuristic that people who agreed in the past will probably agree again. Users can protect their privacy by entering ratings under pseudonyms, without reducing the effectiveness of the score prediction.

C. E-commerce

The most popular and successful area of RS application is no doubt e-commerce. Amazon.com is a personalized RS e-commerce website which has to be mentioned, and it is called "the king of recommendation systems" [12]. Jeff Bezos, CEO of Amazon.com once said that, "If I have 3 million customers on the Web, I should have 3 million stores on the Web." [13] So Amazon.com focuses on the personalized RS in the large amount of merchandises, and tries their best to help their customers find the best matching merchandises according to their demands.

TABLE I. Actual RS Applications

Filed	Systems
	Commercialized recommendation systems
E-Commerce	Amazon.com, eBay, Levis, Ski-europe.com, taobao.com, 360buy.com, dangdang.com
Web page	Fab, Foxtrot, ifWeb, MEMOIR, METIOREW, ProfBuilder, QuIC, Quickstep, R2P, Sitesser, SurfLen
Music	CDNOW, CoCoA, Ringo, Music.Yahoo.com, PANDORA, douban.fm, Ringo, last.fm
Movie	Netfilx.com, Moviefinder.com, MovieLens, Reel.com, hulu.com, movie.douban.com, youku.com, tudou.com
News	GroupLens, PHOAKS, P-Tango, GroupLens, wumii.com, zhiyue.me, newzan.com, trap!t

These days, people have transferred from a world lack of merchandises to a merchandises overload world. In the time with e-commerce, either customers or producers meet big challenges: for customers, it is very difficult to find the just needed merchandises in large amount of merchandises; for producers, how to make their own merchandises outstanding to be seen by the most customers, is also not an easy deal. To solve this problem, Amazon.com proposed a method in 2003 [14], which is the later famous item-based collaborative filtering, used by many other e-commerce companies. This algorithm is based on the hypothesis that users will like the merchandises which is similar to the ones they like before. So the core of the algorithm is to compute the similarity between the merchandises. And Greg Linden, the ex-scientist of Amazon.com, declared in his blog that, when he left Amazon.com, at least 20% of the sales due to the recommendation algorithm, and later the number came to be 35%.

D. Movies

Movie and video websites are also an important area of RS application. RS here can help the users find the movies and videos they are interested in in the vast movies and videos. With Amazon.com, Netflix is the most representative company in RS area.

Netflix is an American provider of on-demand internet streaming media available to viewers in North and South America, Caribbean, and parts of Europe, and of flat rate DVD-by-mail in the United States, where mailed DVDs are sent via Permit Reply Mail. One of the contributions of Netflix makes it a company has to be mentioned when mentioning RS. Netflix held a famous match about RS in 2006, called Netflix Prize. One of the most two important effects of Netflix Prize is that it provides the academic world a large-scale data about user behaviors from a real system, the other is that many perfect algorithms were born in the three years of the match, which lower the forecast error of the system in a significant extent. Besides these, the most important is that, this match enhanced the user experience, and proved that RS is a very good and important tool in the area of movies. And, Netflix announced that 60% of the users could find their own interesting movies through the RS.

Not only Netflix, but also YouTube, the biggest video website in US, face the problem of information overload, and spend lots of time and money in RS researching. In the YouTube paper [15], they show that what they use is also item-based CF, and confirmed that the hits of videos in personalized recommend system are twice that of hot videos.

E. Others

Besides what have been mentioned above, RS have also been used in many other areas successfully, such as personalized music radio, social networks, personalized readings, location based service, personalized advertisement and so on. They are all trying to find the best matching between users and items. Table I shows the areas which RS have been used in, and the systems or companies in the area.

III. THE EXTENSIONS

From the areas which RS have been successfully applied in, we can find that the essential task is to find the most matching items for the objective users [16]. In e-commerce, RS is to find the most needed merchandises for the objective user, i.e. the best matching between customers and merchandises. In the movie and video area, RS is to help the viewers find their favorite movies or videos, i.e. the best matching between movies or videos and viewers. In personalized music radio, RS is to find the music the objective listener like most, i.e. the best matching between listeners and music. In social networks, RS is to find friends who may be suitable for you, i.e. the best matching friends. In personalized readings, RS is to find the articles you may be most interested in, i.e. the best matching between readers and articles. In many other areas, there occurs the same phenomenon; in conclusion, RS is to find the best matching items for the users. Inspired by this conclusion, we may try to apply RS into the areas below, which have never been applied in before.

A. Transportation

This is an area which RS has never been applied in before. RS is to deal with the so-called problem information overload, which is often related with big data. It seems that transportation has no relationship with big data. In fact, there is a great deal of data which is often ignored. Varieties of traffic conditions may be generated every moment. To describe these conditions, we may need flow, delay, stop time and so on. And, there are lots of timing plans used in reality, which could be recorded to mine the other value of them. So transportation is an area filled with big data.

One of the most importation problems in transportation is traffic signal control. Due to the rapid development of social economic, vehicle number raced up and up to an unprecedented level. And various problems ensued, such as the deterioration in the quality of environment, the decrease of travel safety and so on. Researchers noted that it would be not advisable to alleviate traffic congestion by changing the traffic facilities only, so they turn their attention to time, i.e. traffic signal control.

There are many researches about traffic signal control, and the traditional control strategies have been confirmed to have some effect [17, 18]. We have to admit that it still need to be improved. Because of not only embodying in the vehicles and pedestrians, roads, signal lights, but also including the weather environment, legal policy, the influence of social economy and ecological resources, actual transportation systems are so complicated that it is almost impossible to be modeled accurately. In the traditional methods, relatively accurate models tend to be large amount of calculation, complex process, and difficult to achieve. It often chooses to sacrifice some accuracy to seek a balance point which can ensure certain speediness and real time.

So we may try to apply RS to transportation system. The goal of our RS is to recommend the most appropriate timing plans to the intersection according to different traffic demands. The new method won't rely on the accurate mathematical models of transportation system, and what we need is input and output of control system. This is similar to the idea of model-free adaptive control [12], which is significant for solving complex system problems. Another advantage of the new method is that RS could improve themselves by enriching the timing plan library after every recommendation activity. It looks like a person who has the ability of learning. And, the key point is to collect transportation data.

B. Social Manufacturing

Social manufacturing is a completely new area proposed by Fei-Yue Wang in 2012 [19]. The key to the success of the social manufacturing is to integrate effectively social demands and production capacity in real time. So it must make use of crowdsourcing to make customers participate fully in the whole life cycle of production processes. Quirky is one of the most successful social manufacturing company in the world, and Fig. 2 is the flow chart of Quirky.

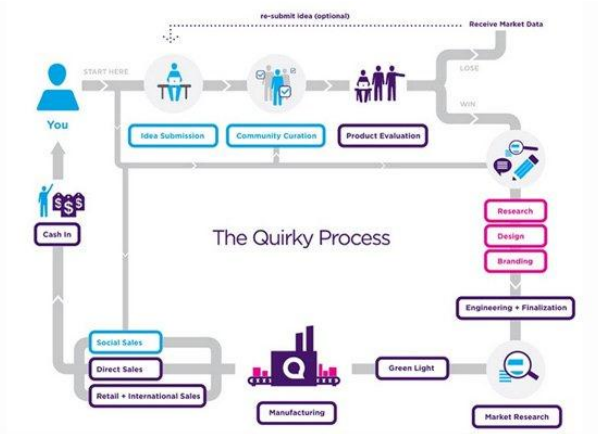

Fig. 2. Flow chart of Quirky

There is a problem. In the process, there are many links, and each link has a demand, which need a best matching customer who has the production capacity to satisfy; for customers, everyone wants to do the thing which they are good at as well. How to find the best matching between the demands of the links and the customers, is the core of making full use of crowdsourcing effectively. So RS is needed now. We may build a database which records the links of the process, the customers, and the relations between the former two. Here relations may be the extent of interesting, how good at the links the customer are, and so on, which may be transferred to ratings in traditional RS. After the database built, RS may solve the problems in the social manufacturing.

C. Knowledge Automation

Applying RS into knowledge automation seems not a simple method, but more likely methodology research. Last two centuries are the time of industry, and industry automation changed people's life. Today, the emerge and development of new medium like twitter, facebook, QQ, weibo, weixin and so on, bring us to the time of information overload. There are many values in the information, and noise we cannot easily eliminate. The useful information knowledge is what we need when we make decision. So in the time of knowledge, of course, knowledge automation is needed, to help us to a new life way. Fig. 3 shows knowledge automation in the narrow sense.

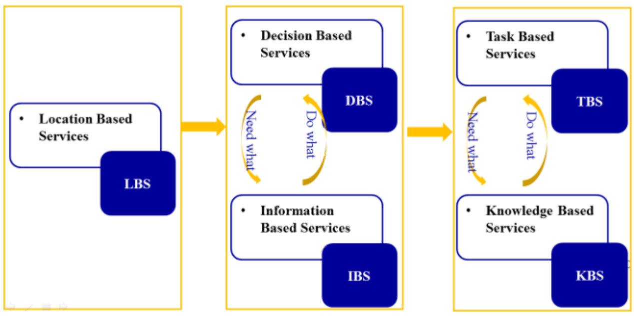

Fig. 3. Knowledge Automation in the narrow sense

In the narrow sense, knowledge automation is to solve the problem of information overload. The information overload in knowledge automation relates to the fragmentation of information which could not be applied, while information overload in traditional RS is mainly about the overmuch choices. Although there are some differences between knowledge automation and traditional RS, RS is born to solve the problem of so-call information overload. So we could try to borrow the RS technics to solve the problems in the knowledge automation. RS here could build the relationship between decisions and information. This is different from traditional RS, which is just recommending the items for the users. Here RS is bidirectional. When information comes, what decisions we should make, and when we must make a decision, what information we need, i.e. recommend decisions for the information, and recommend the information for the decisions. Different areas have different knowledge, and the information and the decisions are all different. So when applying RS into knowledge automation, it is more likely methodology research.

Before the formation of the methodology, more practice in different areas are needed.

Regardless of the specific areas, we could first do some research from the perspective of abstraction. Different from users and items in traditional RS, what affect a decision may not be single information, and single information may cause several decisions. So we may apply permutation and combination into RS, to find the best matching between combinations of information and decisions. The key point here is the representations of information and decisions, and how to build the relationship between the former two. What's more, the representation here should be convenient to use permutation and combination in math.

IV. CONCLUSION

This paper focuses on the application of RS through the perspective of best matching. We first give an introduction of RS, to declare that the necessary of a survey about the application of RS. Then we retrospect the development history of RS through some classic method about the different area. Finally, the extensions may be the future work.

REFERENCES

[1] J. O'Donovan and B. Smyth, "Trust in recommender systems," Proceedings of the 10th International Conference on Intelligent User Interfaces, ACM, 2005.

[2] Book, recommendation systems

[3] M. Levene, An introduction to search engines and web navigation, John Wiley & Sons, 2011.

[4] http://cdn-0.nflximg.com/us/pdf/Consumer_Press_Kit.pdf

[5] J. Davidson, et al., The YouTube video recommendation system. *Proceedings of the fourth ACM conference on Recommender systems*, ACM, 2010.

[6] P. Resnick, and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56-58, Mar. 1997.

[7] Denning, P.J. Electronic junk. Commun. ACM 25, 3 (Mar. 1982), 163-165.

[8] Palme, J. You have 134 unread mail! Do you want to read them now? In Proceedings IFIP WG 6.5 Working Conference on Computer-based document Ser vices (Nottingham, England May 1984), pp. 175-184.

[9] Terry, D.B. 7 steps to a better mail system. Message Handling Systems and Application Layer Communication Protocols, P. Schicker and E. Stefferud, Eds., North Holland, 1991, pp. 23- 33.

[10] Goldberg, D., et al. (1992). "Using collaborative filtering to weave an information tapestry." Communications of the ACM **35**(12): 61-70.

[11] Resnick, P., et al. (1994). GroupLens: an open architecture for collaborative filtering of netnews. Proceedings of the 1994 ACM conference on Computer supported cooperative work, ACM.

[12] A Guide to Recommender Systems.

[13] Schafer, J. B., et al. (2001). E-commerce recommendation applications. Applications of Data Mining to Electronic Commerce, Springer**:** 115-153.

[14] Linden, G., et al. (2003). "Amazon. com recommendations: Item-to-item collaborative filtering." Internet Computing, IEEE **7**(1): 76-80.

[15] Davidson, J., et al. (2010). The YouTube video recommendation system. Proceedings of the fourth ACM conference on Recommender systems, ACM.

[16] Resnick, P. and H. R. Varian (1997). "Recommender systems." Communications of the ACM **40**(3): 56-58.

[17] C. Chen, F.-H. Zhu, and Y.-F. Ai, "A survey of urban traffic signal control for agent recommendation system," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*, Anchorage, AK, 2012, pp. 327-333.

[18] F. V. Webster. (1958). Traffic signal settings. *Great Britain Department of Scientific A*. [Online]. 39. Available: www.bl.uk/services/document/lps.html

[19] From Social Computing to Social Manufacturing: TheComing Industrial Revolution and New Frontier in Cyber-Physical-Social Space