# Curated Content Based Recommender System

Praveen Manchale
Asst. Professor, CSE Dept.,
PESIT,
Bangalore, India
praveenm@pes.edu

Mohammed Bilal
Student, CSE Dept.,
PESIT,
Bangalore, India

*Abstract*— **This paper discusses curated content-based recommender system, a type of content-based recommender system, implemented for a discussion forum. Content-based recommender systems are systems that recommend an item to a user based upon a description of the item and a profile of the user's interests. There is a need for a curated system which can study, analyze and understand the user and thus recommend the relevant crafted information. A continually improving collection of questions and answers created, curated, edited, and organized by everyone who uses it. The system has the feature to serve the user with the aid of anonymity.**

**Keywords**—content-based; recommender system; discussion forums

## I. INTRODUCTION

### A. Recommender Systems

Recommender systems or recommendation systems (sometimes replacing "system" with a synonym such as platform or engine) are a subclass of information filtering system that seek to predict the 'rating' or 'preference' that user would give to an item (such as music, books, or movies) or social element (e.g. people or groups) they had not yet considered, using a model built from the characteristics of an item (content-based approaches) or the user's social environment (collaborative filtering approaches)[1].

There is a growing interest in *recommender systems* that suggest music, films, books, and other products and services to users based on examples of their likes and dislikes [2].

A few examples of such systems:

- When viewing a product on Amazon.com, the store will recommend additional items based on a matrix of what other shoppers bought along with the currently selected item [3].

- Pandora Radio takes an initial input of a song or musician and plays music with similar characteristics (based on a series of keywords attributed to the inputted artist or piece of music). The stations created by Pandora can be refined through user feedback (emphasizing or deemphasizing certain characteristics).

- Netflix offers predictions of movies that a user might like to watch based on the user's previous ratings and watching habits (as compared to the behavior of other users), also taking into account the characteristics (such as the genre) of the film.

Possible extensions that can provide better recommendation capabilities include, among others, the improved modeling of users and items, incorporation of the contextual information into the recommendation process, support for multi-criteria ratings, and provision of a more flexible and less intrusive recommendation process[4].

### B. How are we solving the problem?

1   Spool is a collection of questions and answers created, edited, and organized by everyone who uses it. The most important thing is to have each question page become the best possible resource for someone who wants to know about the question.

2   One way you can think of it is as a cache for the research that people do looking things up on the web and asking other people. Eventually, when you see a link to a question page on Spool, your feeling should be: "Oh, great! That's going to have all the information I want about that." It's also a place where new stuff--that no one has written about yet--can get pulled onto the web.

3 Accumulating Knowledge: People use Spool to document the world around them. Over time, the database of knowledge should grow and grow until almost everything that anyone wants to know is available in the system. When knowledge is put into Spool, it is there forever to be shared with anyone in the future who is interested.

4 Reusable: Each question page on Spool is a reusable resource that should help everyone who has the question that the page is about. Answers on question pages don't depend on any context about the asker except for what is specified in the question text and details. There is only one version of each distinct question on the site, so everyone who is interested in or knows about that material is focused on that one place.

5 Collaborative:
Almost any public space on Spool can be edited by anyone who knows how to improve it. This includes the text of questions and the details around them, what topics are attached to which questions, and the summaries of answers. Spool relies on the good faith of everyone using it to make it a high quality resource.

6 Continually Improving
People can write their own answers to questions any time they think a question page could become a better resource with more information added to it. People who read question pages rate the different answers so that the best ones can rise to the top of the page and make it better. And people can comment on each other's answers to help them make those better as well.

7 Organized
People who use Spool keep it organized. Each question has a set of topics attached to it which makes it easier to find questions already on the site. The topics are also used to identify related questions and sometimes give context to a question.

8 Targeted
People can follow topics so that the system can show them questions they are interested in and know about. People can follow individual questions too, which creates a waiting audience for anyone who wants to write an answer to the question. Some people call this "inverse blogging."

9 People
Everything on Spool is tied back to a person. Each question and answer has a revision history associated with it, and each change in the log is associated with the person who made it. People use their real names and pictures on Spool and have a short bio describing who they are; this helps anyone reading things they write to understand why they should believe what is written and take into account the author's perspective. For example, if Michael Jordan gives an answer to a question about basketball that means something really different from someone who has never played the game giving an answer.

## II. RELATED WORK

The primary goal of this web-based service is to provide an intuitive, user-friendly forum to ask questions and get them answered as well as answer various questions along with participating in other threads by voting others' answers. There is a platform for students to get their classroom doubts cleared as well as learn extensively the subjects of their interests.

To create a unique and one of a kind web based question-answering system. The traditional means of asking questions involve lot of manual effort and searching with the likes of Yahoo! Answers and Stack Overflow. Keeping steady with the pace of developments in hardware at client side as well as server side, we created a new means of student body discussion forum online. There is an eco-system dictating the trends in education and especially computer science.

The user's question is mapped to topics using Tags. A Graph based model approach (content based) is followed [5, 6, and 7]. The task of recommending tags based on content analysis of the item or of metadata about the item was chosen for the ECML PKDD 2009 Discovery Challenge task (http://www.kde.cs.uni-kassel.de/ws/dc09/online) [8, 9]. Tagging as the user types a question, one of the features of Spool is an active area of research.

## III. PRODUCT FUNCTIONS

The internal functions performed by the product:

• Database of the users and the information about the user ,data posted including the questions and answers are stored

• Recommendation System – Depending on the usage pattern of the user, the system recommends related questions based on intense and nifty algorithms based on word count for the corresponding tags.

- Forum Discussion – Discussions can take place and the same can happen anonymously. If a user prefers not to disclose his information, he can stay anonymous and share his views on the thread under discussion.

- A user can follow his favorite questions and users thus he can track them all the time. Say, I am to appreciate a certain user who generally takes part in interesting questions/discussion, I can follow him and I will get his updates on my feed. So this way a user can always customize his feed with the information which would interest him. Follow the users/threads which interest him.

- Searching – this is one of the most vital and nifty functions of this system. OpenSearch is a tool which will study your usage patterns, analyze it and curate your search results. And while the user is typing out the search query, he will get instant recommended search results giving aid to a faster usage for the user. It also makes use of crafty data mining techniques to pop up the relevant questions that are already asked and hence avoiding the possibility of duplicated questions.

- Filtration of information based on the users need is done and recommend the questions needed. The respective records are handled and only those questions show up (among the recommended ones) which will interest the user at that instant of time and hence saving enormous amount of time and strain.

The external functions

- User logins via Email id, account is created
- The logged in users can use the following services.
1. Post ,edit and comment
2. like or dislike
3. Follow users
4. Notification

## IV. PRODUCT DESCRIPTION

Social bookmarking tools become more and more popular nowadays and tagging is used to organize information and allow users to recall or search the resources. Users need to type the tags whenever they post a resource, so that a good tag recommendation system can ease the process of finding some useful and relevant keywords for users. Researchers have made lots of relevant work for recommendation system. Tagging has become a popular way to organize the information and help users to find other users with similar interests and useful information within a given category. Tags posted by a user are not only relevant to the content of the bookmark but also to the certain user.

A primary concern in personalized recommender systems is to present users with instrumental means for navigating the resources that are most relevant to their real information needs. In social annotation systems, tags serve as an intermediate metadata between users and resources conveying the users' navigational preference, therefore the key challenge in social annotation recommender systems are to accurately capture user preferences through tags and make use it for personalized recommendation. A typical social annotation system has three kinds of entities: user, resource and tag. The user prefers some resources and annotates with some words.

Systems implementing a content-based recommendation approach analyze a set of documents and/or descriptions of items previously rated by a user, and build a model or profile of user interests based on the features of the objects rated by that user. The profile is a structured representation of user interests, adopted to recommend new interesting items. The recommendation process basically consists in matching up the attributes of the user profile against the attributes of a content object. The result is a relevance judgment that represents the user's level of interest in that object. If a profile accurately reflects user preferences, it is of tremendous advantage for the effectiveness of an information access process. For instance, it could be used to filter search results by deciding whether a user is interested in a specific Web page or not and, in the negative case, preventing it from being displayed.

The user can find questions, post new questions, answer questions, comment and vote (up or down) on questions. A super-fast auto-complete search-box at the top doubles as the method for entering new questions. Questions can be assigned to topics. One can write a post (an informative statement, rather like an orphaned answer or blog post), follow topics and users. Anonymous users (Users without profiles) can only comment. One can follow questions, topics or other users. Rankings or Points are assigned to users. The domain expert will obviously have a higher ranking. Tags are assigned after the question is broken into parts i.e. basically the question is understood by the system and assigned to topics (if a match is found).

## V.   TESTING

We tested the modules separately

- Login or register user.
- Comment or vote for questions and answers
- Asking or Answering questions
- Recommendation system

### A.   Login or register user

We created more than 50 user accounts with same name and different names with different email addresses. User can upload his or her picture into spool profile and user image will be displayed whenever use involved in different actions in spool. If user tries to login with wrong password a warning message will be displayed indicating that "invalid combination".

Registered users activity will be recorded in the user database. We have tested all users profile with their activity it indicates correct information about users. Now a day's people wanted to use system as anonymous user. In this case user can do all activities without registering. We tested using different computers anonymously performing different activities in spool.

### Comment or vote for questions and answers

All registered user can vote for the threads. We posted more than 50 questions in the spool and vote for the each questions anonymously as well as with login. All test cases are working properly. We tested performance of this module based on its efficiency and time efficiency. Results of testing are,

- Efficiency is tested on 100 comments. Efficiency of this module is 84.5 % (with integration of normalization code).
- Performance is tested on around 50 comments and it took average of 5 seconds for single comment.

### B.   Asking or answering questions.

Register user or anonymous user can ask or answer the questions in the spool. We posted more than 100 questions and each question has related answer followed by comments from different users. We tested the performance of this module by asking more number of questions from single user profile and anonymously.

### C.   Recommendation System

This module is tested with different topics and different key-words. Each question has its own recommendation and related questions. This module is tested based on performance. 80% of the recommendation is related to that question itself. A sometimes question with no tags which has recommendation is 60% valid questions.

## VI. RESULTS

|  | Yahoo! Answers | Stack Overflow | Spool |
|---|---|---|---|
| **Description** | Yahoo! Answers is a place where people ask and answer questions on any topic. | This is a collaboratively edited question and answer site for professional and enthusiast programmers. | Spool is in a way is a hybrid of Yahoo! Answers and Stack overflow. Though the focus is on students and professionals in IT and Education, it is open to all. |
| **Who uses the site?** | Everybody | Developers, programmers, experts and enthusiasts. | Everybody |
| **Ratings for question** | mark question as interesting | up or down | up or down |
| **Ratings for comments** | up | up or down | Like/dislike |
| **Related questions** | No | No | Yes. |
| **Email notifications** | No | Yes | Yes |
| **Follow topics** | No | No | Yes. |
| **Follow members** | No | No | Yes. Both people and questions are important. |
| **Content license / Terms** | Yahoo copyright | Open data: CC-by-SA license (free to Remix, to reuse with attribution link) | Free. |

## VII.   FUTURE ENHANCEMENTS

To design a better system, we will need to work on some of the key aspects which we believe will make Spool a very powerful tool for students to use.

*A.       Accuracy*

The accuracy and consistency with which the questions are recommended will be quite a vital part of Spool. For each user, if we study the patterns of usage more clinically, it will be possible to be so accurate and personalized.

*B.   Scalability*

The community can get only powerful with more number of users and more questions. If the database is flooded with questions and answers then it also becomes easier for the recommendation system to recommend the questions more precisely.

*C.   Security*

The users need not be logged in to be able to take part in the discussions which happen on Spool. But we can't do justice to the recommendation system without having the user to login. As the recommender works with user data and does its job. Hence if we increase the security levels to that extent that the user would not hesitate to input his credentials and login.

## VIII.  POSITIVES

The adoption of the content-based recommendation paradigm has several advantages when compared to the others.

### A.   User Independence

Content-based recommenders exploit solely ratings provided by the active user to build her own profile. Instead, collaborative filtering methods need ratings from other users in order to find the "nearest neighbors" of the active user, i.e., users that have similar tastes since they rated the same items similarly. Then, only the items that are most liked by the neighbors of the active user will be recommended.

### B.   Transparency

Explanations on how the recommender system works can be provided by explicitly listing content features or descriptions that caused an item to occur in the list of recommendations. Those features are indicators to consult in order to decide whether to trust a recommendation. Conversely, collaborative systems are black boxes since the only explanation for an item recommendation is that unknown users with similar tastes liked that item.

### C.   New Item

Content-based recommenders are capable of recommending items not yet rated by any user. As a consequence, they do not suffer from the first-rater problem, which affects collaborative recommenders which rely solely on users' preferences to make recommendations. Therefore, until the new item is rated by a substantial number of users, the system would not be able to recommend it.

## VIII. CONCLUSIONS

Content-based recommendation systems recommend an item to a user based upon a description of the item and a profile of the user's interests . While a user profile may be entered by the user, it is commonly learned from feedback the user provides on items. A variety of learning algorithms have been adapted to learning user profiles, and the choice of learning algorithm depends upon the representation of content. We created a unique and one of a kind web based question-answering system. The traditional means of asking questions involve lot of manual effort and searching with the likes of Yahoo! Answers and Stack Overflow. With the recommendations for user to read more about the topic and related discussions and also enabling the system to serve the user correspondingly with the aid anonymity, we have improved existing recommender systems for discussion forum. This can be of great use in colleges and schools for classrooms to discuss about topics described by the teachers. Such a nifty system with crafted user attention is missing today. Possible extensions that can provide better recommendation capabilities include, among others, the improved modeling of users and items, incorporation of the contextual information into the recommendation process, support for multi-criteria ratings, and provision of a more flexible and less intrusive recommendation process. We have developed a less intrusive recommendation process with improved modeling of users and questions. In addition we take the text of the posts into account to improve the quality of recommendations further, which is a future work suggested in [10].

## REFERENCES

[1] Francesco Ricci and Lior Rokach and Bracha Shapira, "Introduction to Recommender Systems Handbook," Recommender Systems Handbook, Springer, 2011, pp. 1-35

[2] P. Resnik and H. R. Varian, "Introduction," (to the special section on recommender systems). *Communications of* the Association for Computing Machinery, 40(3):56–59, 1997.

[3] Linden,et al.,"Collaborative recommendations using item-to-item similarity mappings," United States Patent 6,266,649, July 24, 2001

[4] Gediminas Adomavicius and Alexander Tuzhilin, "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," Knowledge and Data Engineering, IEEE Transactions on (Volume:17 , Issue: 6 ), June 2005

[5] Jaschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G," Tag Recommendations in Folksonomies," In: Kok, J.N., Koronacki, J., De Mantaras, R.L., Matwin, S., Mladeni, D., Jaschke, R., Skowron, A. (eds.) Proceedings of the 11th European conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2007), pp. 506–514. Springer-Verlag, Berlin, Heidelberg (2007)

[6] Zan Huang, Wingyan Chung, Thian-Huat Ong, Hsinchun Chen, "A Graph-based Recommender System for Digital Library," *JCDL'02,* July 13-17, 2002, Portland, Oregon, USA.

[7] Joseph A. Konstan, John Riedl, "Recommender systems: from algorithms to user experience," Published online: 10 March 2012,Springer Science+Business Media B.V. 2012

[8] Mrosek, J., Bussman, S., Albers, H., Posdziech, K., Hengefeld, B., Opperman, N., Robert, S., Spira, G, "Content- and graph-based tag recommendation: two variations," In: Proceedings of the ECML-PKDD Discovery Challenge Workshop, pp. 189–199. Bled, Slovenia, Europe, September 7th (2009)

[9] Michael J. Pazzani and Daniel Billsus, "Content-Based Recommendation Systems," In "The Adaptive Web" : P. Brusilovsky, A. Kobsa, and W. Nejdl (Eds.), LNCS 4321, pp. 325 – 341, 2007. © Springer-Verlag Berlin Heidelberg 2007

[10] Fabian Abel, Ig Ibert Bittencourt, Nicola Henze, Daniel Krause, Julita Vassileva , "A Rule-Based Recommender System for Online Discussion Forums," Adaptive Hypermedia and Adaptive Web-Based Systems Lecture Notes in Computer Science Volume 5149, 2008, pp 12-21