ITRO

A JOURNAL FOR INFORMATION TECHNOLOGY, EDUCATION DEVELOPMENT AND TEACHING METHODS OF TECHNICAL AND NATURAL SCIENCES

Issue frequency
Twice a year – electronic and paper issue
Volume 4, Number 1, 2014.

Publisher
University of Novi Sad
Technical Faculty "Mihajlo Pupin" Zrenjanin
Department of Teaching Methods of Science and Education Technology

ISSN 2217-7949

Translator
Erika Tobolka, Ph.D.

Topic areas of the Journal

The Journal issues scientific, review and professional papers encompassing the following areas:

- teaching methods of subjects and educational technology in technical and natural sciences fields in pre-school education and training, elementary and high school, as well as colleges and faculties, and adults' training and education,
- pedagogy, didactics, psychology of learning, organizing of school work, methodology of pedagogical researches,
- papers of home sciences of single educational fields that is teaching subjects directed to bringing up to date the educational contents.

Fields – sections in the Journal

- Information technologies in education development
- General topics important to any teaching methods
- Sections of any teaching methods where papers from natural and technical sciences teaching methods will be published
- Foreign experiences important for teaching methods development
- New issues – professional events of current interests
- Students' papers – special methodic topics

CIP – Каталогизација у публикацији
Библиотека Матице српске, Нови Сад

004:371.3

ITRO [Elektronski izvor]: a journal for information technology, education development and teaching methods of technical and natural sciences / chief and responsible editor Dragana Glušac. – [Online izd.].-Elektronski časopis.- Vol. 1, no. 1 (dec. 2011) - . – Zrenjanin : Technical Faculty “Mihajlo Pupin”, Department of Teaching Methods of Science and Educational Technology, 2011 -

Dostupno i na http://www.tfzr.uns.ac.rs/itro/journal.html
ISSN 2217-7949
COBISS.SR – ID 268534279

CONTENTS

IMPROVED ALGORITHM FOR TAG-BASED COLLABORATIVE FILTERING

UDC: 519.254:[37.018.43:004
Professional Paper

Aleksandar KOTEVSKI, Cveta MARTINOVSKA BANDE
Computer Science Faculty, University "GoceDelcev" – Stip, Republic of Macedonia
aleksandar.kotevski@uklo.edu.mk, cveta.martinovska@ugd.edu.mk

Paper received: 15.10.2014.; Paperaccepted: 3.11.2014.

Abstract - Important aspect in the modern e-learning systems is selecting the most adequate learning materials based on learners' requirements, needs and knowledge goals. Recommender systems based on collaborative filtering contribute to overcoming the information overload in personalized learning environments. That's why there is imminent need of using systems that have the capability to detect the learners' needs and to recommend them the most adequate learning context. In recent years, it is common practice to use tags in the process of filtering the most useful learning materials.Through the tagging, learners can mark or highlight some learning materials and can contribute to organizing and retrieving useful learning materials.

Our previous researches were focused on tag-based collaborative filtering and learning style determination, the factors that affect the tag-based collaborative filtering, in order to suggest useful learning material in adequate format.

In this paper, we propose a new tag-based collaborative algorithm that takes in consideration the factors that affect the tag-based collaborative filtering in order to develop more efficient and accurate algorithm, and suggest the learning materials based on posted tags rating and students rating.

The developed system was implemented at the Faculty of Law – Bitola, and the evaluation results are shown in this paper.

Keywords:tag-based collaborative filtering, algorithm, e-learning, learning materials

I. INTRODUCTION

Undoubtedly, the internet technology has an important rolein the learning systems and, the volume of course-related information available to the learners is rapidly increasing. Searching for useful materials and sources in a large dataset without some tools for context filtering and recommendations leads to inefficient learning process.

Intelligent e-learning systems can improve, modernize and simplify the learning process by using tools for filtering the most adequate learning materials based on users' knowledge level, needs, requirements and interests. In addition, the intelligent e-learning systems are going to motivate and support the learners to achieve the learning goals on efficient and effective way.

Recommender systems in e-learning environments utilize information about learners and learning activities and recommend items such as papers, web pages, courses, lessons and other learning objects. According to Drachsler et al. [1], recommender systems have to meet the pedagogical rules and interests of learners. Because all learners have different characteristics, the effective recommender system in e-learning environments must take in consideration some learners' features like learning goals, knowledge level, learning characteristics, strategies, etc. The main goal of the recommender systems is to make predictions using user ratings and tags available for a given item.

Collaborative filtering is a wildly used approach to recommend adequate items to users based on the assumption that similar minded people will have similar taste, requirements, needs or behaviors. According to Huizhi et al. [2], the collaborative filtering can help users organize, share and retrieve information in an easy and quick way.

With the increased use of the collaborative tagging systems, tags become useful information to enhance and optimize the algorithms for recommender systems. These systems can support learners by recommender learning resources and tags too. Collaborative tagging is a mechanism for describing items in large on-line collections. In other words, collaborative filtering approaches predict the rating of items for a specific user based on the ratings and tags from other users with similar interests. The same holds for tag suggestions. Tagging has recently become very popular and useful. At the same time it's an effective way of classifying items and categorizing them in groups that contain items with similar characteristics. According to Wartena et al. [3], tags are assigned by users to describe and find back items. Based on Musto et al. [4] the use of tags, keywords freely chosen by users for annotating resources, offers a new way for organizing and retrieving web resources that closely reflect the users' mental model and also allow the use of evolving vocabularies.

Because different learners may set different tags for the same learning material and the learning material can be tagged with multiple tags, the learner profile should be profiled not only by the tags and used learning materials, but also by the relationship between the tags and tagged learning materials from the learner [5].

In our previous researches [6,7,8], we have implemented an intelligent e-learning system that was used in the educational process at the Faculty of Law in Bitola. It includes adaptation rules and ontology for knowledge representation and supports the learners by recommending learning materials, online learning activities based on their learning style, used tags, knowledge level and the browsing history of other students with similar characteristics. In other words, the system uses tag-based collaborative filtering in order to recommend the most adequate learning materials to the students. The students can add tags for the learning materials by using an interface and simply by entering one or more tags separated by commas in the free-text input text field. In our another research [9], we have identified the factors and parameters that impact a tag based collaborative filtering used for recommending the most adequate learning materials. In that content, we have identified the following factors: students rating, tags rating and learning materials rating.

In the scope of this paper, we review several tag-based collaborative filtering algorithms and propose a new algorithm that takes in consideration the factors that affect the tag-based collaborative filtering in order to develop more efficient and accurate algorithm. Our approach determinates the similar profile with logged student, selects the adequate learning materials and forces the more important learning materials – materials that have tags with high rating set by the students with high rating.

II. RELATED WORKS

Liang et al. [2] proposed a tag-based collaborative filtering approach for recommending personalized items to the users. Based on the distinctive three dimensional relationships among the users, tags and items, they proposed a new similarity measuring method which generates the neighborhood of users with similar tagging behavior instead of similar implicit ratings. Based on experimental result, the authors show that by using the tagging information, the proposed approach outperforms the standard user and item based collaborative filtering approaches. Carmagnola et al. [10] proposed a framework for improvingrecommender systems through exploiting the users tagging activity. They stress social annotation as a new and powerful kind of feedback and as a way to infer knowledge about users. Also, they investigated the role of tags in the definition of the user model and the impact of the tags on the accuracy of the recommendations. Yue et al.[11]proposed a novel algorithm for tag-based collaborative filtering, which exploits user-contributed tags that are common to multiple domains in order to establish the cross-domain links necessary for successful cross-domain collaborative filtering. The authors introduced a constraint involving tag-based similarities between pairs of users and pairs of items across domains. By using two publicly available collaborative filtering data sets as different domains, the authors experimentally demonstrated that the new algorithm substantially outperforms other state-of-the-art single domain collaborative filtering and cross-domain collaborative filtering approaches.Rong et al.[12] proposed a collaborative approach for expanding tag neighbors and investigate the spectral clustering algorithm to filter out noisy tag neighbors in order to get appropriate recommendation for the users. Based on the preliminary experiments that have been conducted on MovieLens dataset to compare the proposed approach with the traditional collaborative filtering recommendation approach and native tag neighbors expansion approach in terms of precision, the result demonstrates that the proposed approach could considerably improve the performance of the recommendations. Wartena et al. [3] focused on generating tag-based profiles for the users and then recommended new learning materials based on the generated profile. Also they introduced topic aware recommendation algorithm - first detect different interests in the user's profile and then generate recommendations for each of these interests. The authors in [13] present a tag recommender system which extends the collaborative filtering with a content-based approach able to extract tags directly from the textual content of HTML pages. Results of their experiments carried out on a large dataset gathered from Bibsonomy, where's shown that the use of content-based techniques improves the predictive accuracy of the tag recommender.

III. Proposed Approach

The proposed algorithm for tag-based collaborative filtering is a part of a larger adaptive e-learning system. Except recommendation, the system delivers the learning materials in format adequate to the learners’ learning style.

All of the learners can describe learning materials with a set of tags, whereby the system creates a complex network of learners, learning materials and tags. To understand the main idea of the proposed algorithm, we need to consider that network as a three-dimensional relation: learner – learning material – tag. That structure allows determination oflearners that set tags for specific learning material. In that manner,we can define the following sets:

S = {S_1, S_2, ... S_n}: set of learners (in our case students)
L = {L_1, L_2, ... L_n}: set of learning materials
T = {T_1, T_2, ...T_n}: set of tags posted from the students S for learning materials L

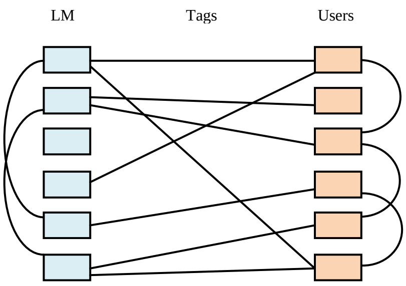

Figure 1.Conceptual model for collaborative tagging system

Additionally, learners and tags have their own rating. The learning material becomes important if it has been tagged with important tags (tags with high rating)from important learners (learners with high rating). For instance, one learning material could be tagged with important tags by important learner. Then, the tagged learning material can be considered as an important learning material and suggest it to the logged learner. The same holds for the learners and tags.

Undoubtedly, learners have different knowledge level and have different interest. To achieve greater efficiency in the educational process and in the process of recommendation, we group learners in few groups and subgroups named as virtual learning group. There, virtual learning group is a set of learners with the same knowledge level, the same learning interests and use the same course. For instance, learner A and learner B belongs to the same virtual learning group only if they have the same knowledge level and they have the same learning interests.

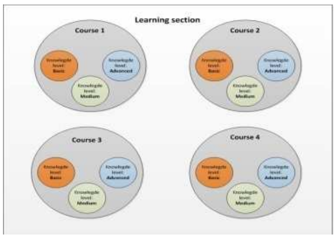

Figure 2. Virtual learning groups

The main idea of our paper is to suggest the most relevant learning materials to the learners using tag-based collaborative filtering, but also to take in consideration and the learners and learning materials rating.With other words, the suggested algorithm will force the more important learning materials – materials that have tags with high rating posted by learners with high rating.

To generate the suggested list, the system needs to complete the following steps: determinate similar learners, select the most adequate learning materials and order the selected learning material by their rating.

The main goal of the first step is to determinatesimilar learners with the logged learner and to generate a set of the top N most similar learners, ordered by their rating. In this step, system first selects all learners that belong to the same virtual group with the logged student (all students with the same knowledge level and learning interests) and students with higher knowledge level but with the same knowledge level. For instance, if logged user A has basic knowledge level and his learning interest is PHP programming language, then the algorithm will select all students with the same properties (basic knowledge level and interest in PHP programming language) and students with medium or advanced knowledge level that are interested for PHP programming language too.

Once the most similar learners are identified, the second step is to select the most adequate learning materials in order to be recommended to the logged learner. In that manner, the algorithm takes in consideration all materials which have been tagged from the similar learners generated in the previous step but not used from the logged learner. Important aspect in this step is the learning materials rating because the algorithm will force the learning materials with higher rating.

Within the first step, we use BM25, also known as Okapi BM25. Manning et al. [14] defined it is a non-binary probabilistic model used in information retrieval. The system takes into consideration a set of tags of each learner and make two analogies, comparing the tags of the logged learner with a query, and the set of tags of each similar profile as a document. It means that we performed calculation of learners profile similarity based on the BM25 model and thus we generate a set with all the similar profiles to the logged learner. The BM25-based similarity model is taken from the calculation of the Retrieval Status Value of a document (RSV_d) of a collection of a given query [14]:

$$RSV_d = \sum_{t \varepsilon q} IDF * \frac{(t_1+1)tf_{td}}{k_1\left((1-b)+b*\left(\frac{L_d}{L_{ave}}\right)\right)+tf_{td}} * \frac{(k_3+1)tf_{td}}{k_3+tf_{td}}$$

RSV_d represents the similarity score between the logged learner (the terms of the query q) and one similar learner (the terms of the document d) from the same virtual group. This similarity is calculated as a sum over every tag t posted by the logged student. The similar learner n is represented as a set of tags with their frequencies. L_d is the sum of the frequencies of each tag of the similar learner n. L_{ave} is the average of the L_d of every similar learner. The term tf_{tn} is the frequency of the tag t into the set of tags of the similar learner n, tf_{tq} represents the frequency of the tag t into the query - the set of tags of the logged user.

After calculating the similarity between the logged learner and each similar learner (learners from the same virtual learning group or learners with higher knowledge level but with the same learning interests), we choose the top N similar learners with the highest rating.

Within in the second step, the system uses cosine-based similarity to calculate the similarity between two learning materials – learning materials tagged from the logged learner and learning materials tagged from the similar learners. Then, the system will select top N materials with highest rating. To get the more reliable results for calculating the similarity between learning material *a* and learning material *b*, we need to isolate the students who have set tags to both of these items and then to apply a similarity computation technique to determine the similarity between learning material *a* and learning material *b*.

We use cosine-based similarity to calculate the similarity between two learning materials. In this case, the learning materials are thought of as two vectors in the m dimensional user space[15]. The similarity between the materials is measured by computing the cosine of the angle between these two vectors, based on following calculations:

$$Similarity\ (a,b) = \cos(\vec{a},\vec{b}) = \frac{\vec{a} - \vec{b}}{\|\vec{a}\|_2 * \|\vec{b}\|_2}$$

Because the learners rating and the learning materials rating have an impact on the process of determining the relevant learning materials, we need to calculate them.

A. Student rating

In order to calculate the learners rating, the system uses two coefficients: knowledge level coefficient (C_{kl}) and student activity coefficient (C_{sa}).

Total student rating C_{kl} can be calculated as an average value of the two coefficients:

$$C_{kl} = \sum\left(\frac{Pn}{Nt} * K_{ln}\right)$$

P_n is a score from the test of knowledge level $K_{ln}$and N_t is the maximum number of test points.

The student activity coefficient (C_{sa}) can be calculated as:

$$C_{sa} = \frac{Tsu}{Tt}$$

$T_{su}$is number of total tags posted from the student S, while T_t is total number of tags posted from the other learners for learning materials tagged by learner S.

Finally, learner ratingS_{rat} can be calculating as:

$$S_{rat} = \frac{Ckl+Csa}{2}$$

B. Learning material rating

Average material rating *(LM_r)* can be calculated as an average value of two coefficients: average rating posted from the learners *(R_{av})* and learners' average rating that post rating to learning material (R_{sav}):

$$LM_r = \frac{Rav+Rsav}{2}$$

The Figure 3 shows the diagram of proposed approached.

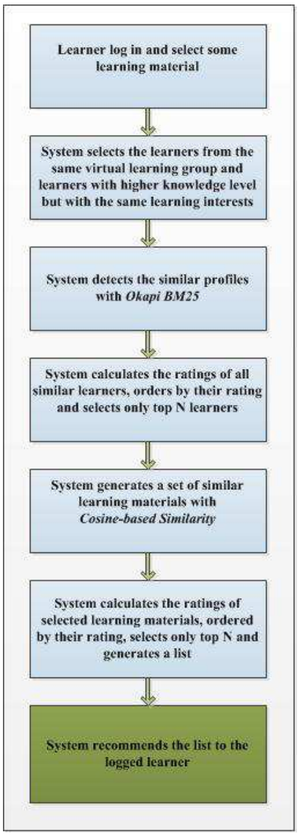

Figure 3. Process of learning materials recommendation

The implemented system uses following logics in the process of learning materials recommendation:

For each student S do
If the student S belongs to the same virtual learning group with the logged student Sa
If the student S has rating >rating_limit
For each similar student to the logged student Sa (based on Okapi BM25) do
For each tag on a learning material Lm of similar student Sa do
For each commonly tagged learning material Lmk that Sa has with S do
If the learning material Lmk has rating >rating_limit_material
For each similar tag i
Calculate the $sim(T_{Sa\,->Lmk[i]}, T_{Sa->Lm[i]})$
Add $sim(T_{s1->LMk[i]}, T_{Blmk->k[i]}) + 1$ *to student_temp_similarity*
Add_material_to_finall_list()
End for
End if
End for
End for
End for
End if
End if
End for

IV. Results

The system was implemented at the Faculty of Law in Bitola. It was used in the period of 6 months by the students from the undergraduate studies. The survey was conducted on total 110 students, divided into twovirtual learning groups: Computer Technology and Constitutional Law, which contains three sub virtual groups, based on student knowledge level: basic, medium and advanced knowledge level. Table I shows the number of students in virtual learning groups and sub-groups.

TABLE I. Number of Students in Virtual Learning Groups

		Knowledge level		
		Basic	*Medium*	*Advanced*
Courses	*Computer Technology*	22	24	12
	Constitutional law	18	21	13

We compared the results from our previously research and the current research. In our previously research, we were using simple collaborative filtering for learning materials recommendation, but we didn't take in consideration any additional factors that affect the collaborative filtering process.In the current research, we use BM25 probabilistic model for determination of similar students with the logged student and cosine-based similarity for selecting the most adequate learning materials. Additionally, in the scope of this paper we have taken in consideration learners rating and learning materials ratingin order to check their impact on the process of determining the most relevant learning materials. Table II shows results differences.

TABLE II. Comparation of the Results

Activity	*The old research*	*The current research*
Number of learning units	91	148
Number of students	110	110
Number of tags	739	1345
Average students rating (1-5)	/	3.89
Average learning materials rating (1-5)	/	4.12
Used learning materials from the suggested list (%)	74.6	83.8

According to the results in TableII, it's clear that the approach that takes in consideration learning materials and students rating bring to more effective process of tagging, but also and more valuable and useful recommendation. The Figure 4 shows the graph presentation of the results differences.

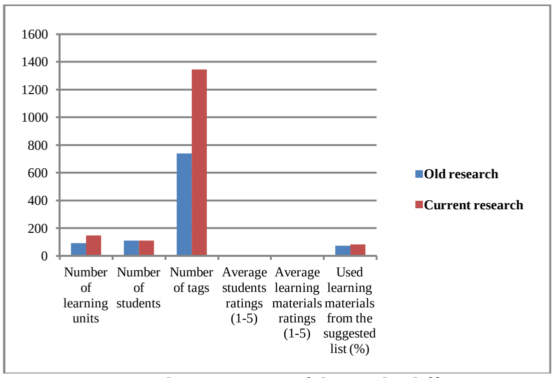

Figure 4: Graph presentation of the results differences

An e-survey was conducted as a last part of this research. The survey was conducted to the students after using the system and their answers are shown inTable III.

TABLE III. RESULTS FROM THE E-SURVERY

#	Question	Answers (1 - StronglyDisagree 5- Strongly Agree)				
		1	2	3	4	5
1.	I found useful using learning materials with adequate complexity to my knowledge level and prior knowledge	5	9	22	25	49
2.	I found useful recommendations for learning materials	1	7	17	39	50
3.	The recommend learning materials are adequate to my needs	2	8	12	37	51
4.	The system was user-friendly and easy for using	3	8	14	29	56
5.	The approach of suggesting learning materials is useful and helpful in the educational process	4	11	26	30	39
6.	Learning materials with high rating are more valuable and helpful for me	7	15	22	30	36
7.	I want to use the learning materials which were used from the student with high rating	8	11	24	29	38

According to the answers, the students are satisfied with the quality of recommendation for the next learning materials, and they agree that the system is user-friendly and easy for using. Furthermore, they confirmed that the learning materials rating and student rating has important impact in the process of learning material recommendation.

V. CONCLUSION

The system's ability to select the most adequate learning content and deliver it in the adequate format to the users is very important aspect in adaptive e-learning systems. The main goal of our system is to recommend the most appropriate materials to the students based on the tags they set for the learning materials. Additionally, we have taken in consideration students rating and learning material rating in the process of collaborative filtering.

In the scope of this paper we proposed a tag-based collaborative filtering algorithm that takes in consideration the factors that affect the tag-based collaborative filtering in order to develop more efficient and accurate algorithm. Our approach determinates the similar profile with logged student, selects the adequate learning materials and forces the more important learning materials – materials that have tags with high rating set by students with high rating.The system calculates the rating of the learning materials and students first. Then, the system determinate the similar profiles to the logged learner based on the BM25 probabilistic model. Second, by using cosine-based similarity the system calculates the similarity between two learning materials – learning materials for which the logged learner has set tags and learning materials for which the similar learner has set tags. Then, the system will select top N materials with the highest rating.

After a period of using the system, we have compared the results obtained from the student's activities and we can conclude that the proposed algorithm for tag-based collaborative filtering that takes in consideration ratings of students and learning is more efficient that a standard collaborative filtering. It can be concludes based on the highest percentage of accepted items from the suggested list in the current research versus the percentage in the preview research.

The future researches could be focused on including lists with synonyms for the tags and cold star problem. in tag-based collaborative filtering process in order to be recommend more adequate learning materials.

REFERENCES

[1] Drachsler, H., Hummel, H., & Koper, R. (2008). Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model. International Journal of Learning Technology 3 (4), 404-423

[2] Lecture Notes in Computer Science Volume 5589, 2009, pp 666-673 Tag Based Collaborative Filtering for Recommender Systems Huizhi Liang, Yue Xu, Yuefeng Li, Richi Nayak

[3] Christian Wartena, Martin Wibbels, Improving Tag-Based Recommendation by Topic Diversification, Advances in Information Retrieval Lecture, Notes in Computer Science Volume 6611, 2011, pp 43-54

[4] Cataldo Musto, Fedelucio Narducci, Pasquale Lops, Marco de Gemmis, Combining Collaborative and Content-Based Techniques for Tag Recommendation, E-Commerce and Web Technologies, Lecture Notes in Business Information Processing Volume 61, 2010, pp 13-23

[5] A.Kotevski, C.Martinovska, R.Kotevska (2013) - Learning style determination in e-learning system, International conference of young scientists – Plovdiv

[6] A.Kotevski, R.Kotevska, Virtual learning group in intelligent e-learining systems, The 2nd International Virtual Conference 2013, (ICTIC 2013), Slovakia

[7] A.Kotevski, Gj.Mikarovski, Intelligent learning system for High education, ICEST 2012

[8] A.Kotevski, C.Martinovska Bande, Recommending audio and video materials based on tag-based collaborative filtering, 11th International Conference on Informatics and Information Technologies, CIIT 2014

[9] Aleksandar Kotevski, Cveta Martinovska Bande and Gjorgi Mikarovski, Factors that affect the tag-based collaborative filtering, XLIX International Scientific Conference on Information, Communication and Energy Systems and Technologies, ICEST 2014 (in printing)

[10] Francesca Carmagnola, Federica Cena, Luca Console, Omar Cortassa, Cristina Gena, Anna Goy, Ilaria Torre, Andrea Toso, Fabiana Vernero, Tag-based user modeling for social multi-device adaptive guides, User Modeling and User-Adapted Interaction November 2008, Volume 18, Issue 5, pp 497-538

[11] Yue Shi, Martha Larson, Alan Hanjalic, Tags as Bridges between Domains: Improving Recommendation with Tag-Induced Cross-Domain Collaborative Filtering, User Modeling, Adaption and Personalization Lecture Notes in Computer Science Volume 6787, 2011, pp 305-316

[12] Rong Pan, Guandong Xu, Peter Dolog, Improving Recommendations in Tag-Based Systems with Spectral Clustering of Tag Neighbors, Computer Science and Convergence Lecture Notes in Electrical Engineering Volume 114, 2012, pp 355-364

[13] Huizhi Liang, Yue Xu, Yuefeng Li, Richi Nayak, Collaborative Filtering Recommender Systems Using Tag Information, 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology

[14] Manning, C., Raghavan, P. and Schutze, H. 2008, Introduction to Information Retrieval, Cambridge University Press

[15] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, Item-Based Collaborative Filtering Recommendation Algorithms, GroupLens Research Group/Army HPC Research Center