Users opinions about Learning Object Recommendations: a case study

Henrique Lemos dos Santos, Cristian Cechinel
Centro de Desenvolvimento Tecnológico (CDTec)
Faculdade de Educação (FaE)
Universidade Federal de Pelotas (UFPel)
Pelotas, RS, Brazil
Email: hldsantos@inf.ufpel.edu.br
Email: contato@cristiancechinel.pro.br

Franco Giustozzi, Ana Casali, Claudia Deco
Fac. de Cs. Exactas, Ingeniera y Agrimensura
Universidad Nacional de Rosario
Rosario, Argentina
Email: fngiustozzi@gmail.com
Email: {acasali, deco}@fceia.unr.edu.ar

***Abstract*—The enormous growth of learning objects on the internet and the availability of preferences of usage by the community of users in the existing learning object repositories have opened the possibility of testing the efficiency of different techniques on recommending learning materials to the users of these communities. In this work, we focus on some particular parameters at the recommendation phase (different similarity algorithms), evaluating the new recommendations not only via offline analysis but also taking into account users feedback. It has been performed a online analysis over a small group of users. The recommendations were presented to these users along with a small inquiry form about each recommendation. Through this study we tried to find out which algorithm performs better from the online analysis and if it is possible to notice a similarity between the results obtained from the offline and online analysis.**

***Keywords*—Learning objects recommendation; Collaborative filtering; Recommendation techniques; Recommendation systems evaluation**

I. Introduction

Learning Objects (LOs) are any digital resources that can be repeatedly used to facilitate the learning process, e.g., LOs are learning units that can be considered the core of Technology Enhanced Learning (TEL), moreover, the LOs constitute a large portion of the open educational resources (OER) available nowadays. They can take different forms and can be reused, remix, updated, combined, separated and referenced. LOs can be used by a student who wants to learn a subject or by a teacher who wants to prepare materials for his/her class. These objects are usually described with metadata like title, description, material type, discipline, etc.

Learning objects are stored in Learning Object Repositories (LOR), which can be organized, for instance, according the subject of its objects. These repositories provide resources to communities of students, educators and other stakeholders to be consumed through different means (e.g. directly from the repository or in other platforms such as LMS - Learning Management System). Existing LORs can differ in several ways, for example, in the specificity of the area, type of materials, metadata standards, etc. [1]. Some repositories also permit the users to register themselves before they navigate through resources, allowing them to rate and comment these objects. This fact turns the repository into a social community based on learning interest. Each repository can contain thousands of different learning objects, reason why is difficult to find relevant materials of interest.

As LORs are naturally organized around communities of interest, such platforms normally rely on the members of these communities to rate and comment the resources so that the higher-rated ones are thus shown in the first places and more visible during the search and retrieval. Therefore, a repository that contains a lot of objects and allows a community of users to rate LOs is a very good environment for a recommender system.

Recommender systems can allow students or professors the freedom to build an unique learning path that suits the students' preferences, abilities and previous knowledge. These systems provide recommendations based on different approaches, among them the best known ones are Collaborative Filtering (CF) and Content-based Filtering. Collaborative filtering techniques focus on the behavior of users towards items -which are to be recommended- rather than on the internal nature of them. These techniques work better when there is a broad user community and each user has already rated a significant number of items. On the other hand, Content-based filtering focuses on the description of the items and the user preference profile [2]. This approach prefers content semantics to social interactions or user behavior. To improve learning object discoverability, the use of recommender systems has been largely investigated [3].

The selection of relevant learning objects for each user is a subject of active research and development in the field of e-learning. While recommendation algorithms are not new, their adaptation and use with learning objects is a field still open [4]. In [5] they have developed a learning object recommendation system prototype that has been used to experiment with 3 different recommendation algorithms based on the user profile.

Despite the possible benefits that collaborative filtering algorithms adoption could provide for the field of recommender systems in TEL, there is still a lack of studies reporting results obtained specifically from the use of CF in large data samples. This could be the consequence of a major lack of sharable learning objects datasets that would be useful in order to generalize the results [6] or also consequence of the ratings sparsity normally found on these repositories. [7] stated that this lack of evaluative data about several LOs is mainly caused by the disproportion among the LORs growth (regarding number of resources) and the capacity of the community to evaluate the resources. The authors, then, proposed the automatic extraction of information quality about the LOs via Artificial Neural

Networks models using the provided metadata by the two LORs considered on the study. As exposed by [8] experiments where teachers and learners provide some feedback about recommended resources are also useful and valuable, specially if conducted on large datasets that allow real interactions between users. Also, [9] states that few researchers have tested and validated their recommendation systems on real-life data.

The enormous growth of learning objects on the internet and the availability of preferences of usage by the community of users in the existing learning object repositories have opened the possibility of testing the efficiency of CF algorithms on recommending learning materials to the users of these communities. [10] evaluated recommendations of LOs generated by different well known memory-based CF algorithms using two databases (with implicit and explicit ratings) gathered from the popular MERLOT repository. The results obtained in this study highlight the fact that these two datasets represent very different information about the preferences of the users and thus, the recommendations generated through the use of them were also highly different.

The MERLOT[1] repository is a well-known OER provider (community-based) [11] that contains thousands of learning objects (from 9 different major disciplines) and congregates not only students but also teachers and experts who are gathered into peer committees in order to review the submitted resources. [12] stated that this kind of evaluation may be a good solution to the OER selection issue, although [13] points out that this method dramatically slows down the process of quality assessment. Nevertheless, the data available from MERLOT are useful to researches focused on solving the selection issue, either through recommendation systems or search engines, for instance.

In [14] the authors evaluate a pre-processsing method through clustering for future use of CF algorithms. For that they also use a large dataset collected from MERLOT. The results of a quantitative and offline analysis point out that clustering learning objects before the use of collaborative filtering techniques can improve the recommendations performance.

The evaluation of general purpose recommendation systems is an established study field since several works have been developed along the years. Firstly, these works focused on metrics such as precision, accuracy and similar ones, but in the last years the real user opinion, gathered explicitly or not, has been pointed as the most reliable metric. For instance, previous work of [15] developed a framework to evaluate a recommender system on an user-centric approach. They analyzed not only the final result of the recommendation process but also all parts of the user's interaction with the designed system. They divided the user experience intro three components: process (e.g. perceived effort, difficulty), system (e.g. perceived system effectiveness) and outcome (e.g. choice satisfaction) and found several behavioral correlations among them. [16] conducted a study where it is proposed an evaluation framework which consists in pairing two recommender systems which simultaneously compete to give the best recommendations to the same user at the same time. The authors discussed several aspects of the evaluation process such as the recommended item presentation policy and the evaluation feedback, where the proposed framework made use of an inferring preference method although the authors stated that it is preferable to choose a method where these preferences are directly asked to the user.

In this work, following [14] we focus on comparing also an users clustering approach with the two others (traditional CF and LOs clustering), evaluating a different similarity algorithm used in the recommendation phase and evaluating the recommendations not only via traditional error metrics (offline analysis) but also taking into account users feedback. In fact, the new approach that presented the best performance at the offline analysis was chosen in order to generate the recommendations that were experimented later against recommendations made with the traditional CF.

We perform a online analysis over a small group of users, where each user was provided with one or more recommendation (the maximum of four, two provided by the traditional CF and two provided by one of the clustering methods). The recommendations were presented to the user along with a small inquiry form about each recommendation. In the form we asked about the recommendation relevance, the resource quality and also the difficulty level of the object. Therefore, during the present study we intend to answer the following questions:

- From the users feedback, which algorithm performs better?
- Is it possible to notice a similarity between the results obtained from the offline analysis and the ones obtained from the users feedback?

The paper is organized as follows. Section II describes the dataset and the techniques used to cluster users and learning objects along with the methods to generate recommendations. Section III presents the offline evaluation results comparing the recommendation methods. On Section IV a online analysis from real users feedback is exposed. Finally, on Section V conclusions are presented.

II. Data description and techniques

Data Description

In order to evaluate the algorithms used at the recommendation stage, we used the same Merlot dataset presented on [14], which is an updated version of a dataset earlier collected by [17]. This dataset includes data from 3659 users and 4968 LOs, and the total number of LOs comments is 9910.

In the clustering stage, we also used LOs and users meta-data, such as object description and title, and user's disciplines. Then, in the recommendation stage, a group of 9910 tuples presented as <user id; object id; rating> was provided to the recommendation engine. Description and title are textual fields and the rating can be seen as a value that represents how much satisfied is a user after reading or watching (depending on

[1] https://www.merlot.org

material type) the LO. At Merlot, the rating range varies from 1 to 5, in a simple Likert scale.

However, the data contained at the dataset is richer than that, having others LOs attributes (e. g. material type, language, reviews, etc.) and users metadata (e. g. affiliations, member type, etc.). Furthermore, there are also other possibilities of relationship between users and LOs at Merlot, such as personal collections.

Generating clusters and recommendations

Our approach to recommendation is based on applying collaborative filtering to clusters of learning objects and users instead of recommending across all available objects. In order to do so, we first perform a content-based clustering (using TF-IDF and k-means algorithm) of objects and users and then, generate recommendations within each cluster. We analyze the performance of these recommendations when the number of clusters and parameters of the recommender algorithm is changed. The implementations used are those available in the Apache Mahout environment[2], version 0.7. The same environment was used to generate the offline evaluation. Also, a traditional CF recommender engine (without any content filtering) was evaluated in order to be compared with the two new methods.

To generate LOs clusters, we have chosen to represent them as a bag-of-words textual file, where the content was their description and title. Then, a TF-IDF algorithm converted each textual file into a n-dimensional weighted vector, where each position represents a single word and the value represents the word weight. This technique ideally discards stop-words by making their values closer to zero, whereas relevant words receive higher values. Moreover, resources that contain several simultaneous words tend to be converted into similar vectors. The same approach was performed to cluster the users, except from the fact that a user bag-of-words contained all the disciplines in which the user was assigned in. In this case, a direct categorization was not chosen, mainly because MERLOT users are usually attached to more than one discipline, thus, if the users were simply grouped into nine clusters (the number of primary disciplines at MERLOT), we would have lost significantly information derived from another sub-disciplines. After applying TF-IDF, the traditional k-means algorithm was chosen to group LOs, in the first scenario, and users, in the second scenario, according to their similarity. The k parameter, which indicates the number of desired clusters, was varied between 2 to 9 since higher values than 9 led to high losses on the user-space coverage.

At the recommendation phase, a user-based collaborative filtering engine was used and its parameters neighborhood size and similarity algorithm were varied between 2 to 20 and LogLikelihood Ratio to Euclidean Distance, respectively. This memory-based method basically uses the rating registry to calculate the similarity among the users (according to the similarity algorithm chosen) and then creates neighborhoods of similar individuals for each user. The recommendations are then generated within the neighborhood, since users who agreed in the past, tend to agree again in the future - basic principle of collaborative filtering.

III. OFFLINE ANALYSIS: MAE ERROR AND COVERAGE

Objectives and methodology

We ran an offline evaluation of our new proposed methods and a traditional CF recommender. As stated by [18], offline analysis of recommender systems has the benefits of being quick even when different algorithms and parameters are tested. But it suffers from a natural weakness that is the impossibility of evaluating the appropriateness of a recommended item since no real users ratings about the item are available.

On a training-test approach, the mean average error (MAE) between predicted rating (generated by the recommender) and real user rating was measured considering all three methods: LOs clustering, users clustering and traditional CF. Each dataset was split in 90% for training and the remaining for test. It is important to note that when considering the clustering methods, a dataset is a cluster. Thus, this measurement needed to be repeated for each cluster from the k clusters generated, since there were eight different values of k. The quickness of the offline analysis was essential in order to test several combinations of parameters. Moreover, each dataset was evaluated 50 times and then the average error was calculated.

Also, an analysis of how each different algorithm impacted on user-space coverage was made. We calculated how many users stood with no recommendation for each value of k for each combination of similarity algorithm and proposal (LOs clustering, users clustering and traditional CF).

Results and discussion

Figures 1 and 2 show how each k value (including k=1 which is the entire database without any clustering) behaved regarding their recommendations accuracy for the two scenarios. On the LOs clustering approach (1), both Euclidean and Loglikelihood presented some cases that performed better than the pure CF approach. Mainly we highlight the Euclideans' 6, 7 and 8 clusters case and the 6 and 8 clusters on Loglikelihood. When comparing the two similarities between themselves, the Euclidean distance seems better by a narrowly margin. On the other hand, when it comes to users clustering situation (shown in 2), all different values of k performed worse (both Euclidean and Loglikelihood) than the no clustering approach.

We also performed an ANOVA [19] test to verify whether or not there were significant difference regarding the MAE error among different values of k. The test confirmed significant differences at the 95% confidence level for the two approaches, indicating that the means observed in the boxplots above can be considered in order to choose a best k value and also that the user clustering approach is definitely worse than the traditional CF (k=1).

Another important measure to evaluate recommender systems is the coverage. As stated by [20], coverage is related to the degree to which resources can be recommended to all potential users and the percentage of items that are effectively

[2] https://mahout.apache.org/

recommended to a user. In this work, we measured only an user-space coverage, which is the number of users to whom at least one LO were recommended. As our new approaches consisted in splitting the entire dataset into k parts, the coverage of a recommender that runs over smaller datasets is expected to be also smaller. Figure 3 shows this natural behavior: as k grows, the coverage decreases. More than that, it becomes more evident that the user clustering approach is the worst of all three, since its coverage was also worse along all k values. The general loss on the coverage is significantly for the k values that performed better on MAE error analysis when they are compared to pure CF. For instance, considering LOs clustering and Euclidean Distance similarity, k=6 presented a 10% loss on coverage in relation to the k=1 case. However, this case with k=6 obtained one of the smallest MAE errors of all experiment and the loss on the coverage is also smaller than the other case with good MAE error (k=8).

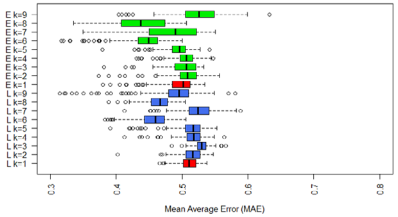

Fig. 1. Boxplots for MAE errors on LOs clustering case to Euclidean Distance (the green and red upper ones) and LogLikelihood (blue and red remaining)

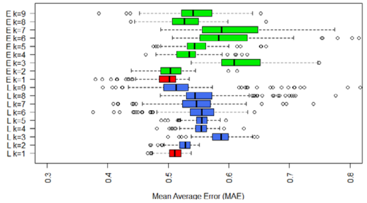

Fig. 2. Boxplots for MAE errors on users clustering case to Euclidean Distance (the green and red upper ones) and LogLikelihood (blue and red remaining)

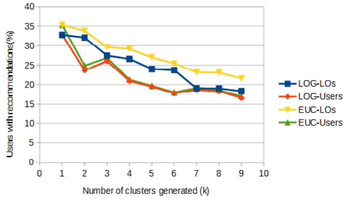

Fig. 3. General user-space coverage among k values

IV. Online Analysis: Real Users Feedback

Objectives and methodology

When considering recommendation systems, the user satisfaction is not always well predicted by offline measures [21], such as MAE and other error metrics, this lack of precision can be more expressive when the recommender is built on a e-learning environment where different learning goals and contexts can change the user perception about a recommended item. In order to analyze and compare the results of our proposed method of recommending respect to the traditional CF, we asked 18 volunteers from different countries to rate some LOs at the MERLOT repository. From their contribution, we gathered 108 total ratings and joined them with our existing collected data to generate recommendations based on the traditional CF engine and on our proposed clustering methods. Specifically, taking into account the results obtained in the offline analysis (III), we have chosen a LOs clustering preprocessing approach with k=6 to be compared to the traditional CF, since this value presented one of the best performances overall and also caused less damage to the user-space coverage, when compared to LOs clustering with k=8. For both two recommendation engines, in the recommendation phase, we used the Euclidean Similarity and a neighborhood size of 10.

Later, we generated for each volunteer, at most 2 recommendations using the traditional CF approach and 2 using the LOs clustering method. Afterwards, webpages containing the recommendations and a brief questionnaire were sent to the users via email. Three Likert scale questions were presented to the users. The content of these questions is listed below.

- Q1 - How relevant do you consider this recommendation?
- Q2 - How difficult do you consider the content of this resource?
- Q3 - Which is your rating for Material Quality?

Results and discussion

We were able to generate 49 recommendations where, 26 were produced by the traditional CF engine and 23 by the LOs clustering combined with CF. Henceforward, these 26 recommendations will be called general recommendations and the 23 remaining will be treated as cluster recommendations. An initial analysis showed that 50% of the general recommendations were evaluated with a rating greater or equals to 4 in Likert scale whereas only 34% of the cluster recommendations were also highly evaluated. The ratings distribution over each case is presented on Figure 4 and shows that cluster recommendations concentrated the ratings among the lower values 3 and 2, while the general ones have almost 50% concentrated on a higher value equal to 4.

A question about the relevance of each recommendation was also asked to the users. Figure 5 denotes a similar behavior to the rating distribution with the general recommendations again concentrating their distribution over higher values of relevance while the values regarding cluster recommendations are mostly 1, 2 and 3. Figure 6 exposes how the difficulty of each recommended LO was evaluated according to each type of recommendation. In this case, the cluster recommendations presented higher values of difficulty.

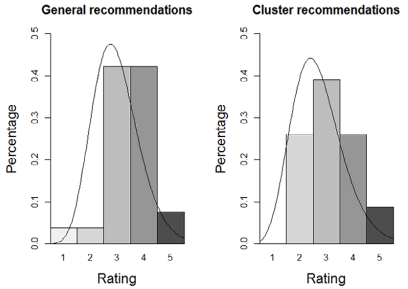

Fig.4. Histogram for the rating distribution for general recommendations (left side) and cluster recommendations (right side)

We consider the three numeric answers presented above plus an error, which was calculated as the absolute value of the difference between the real user rating and the rating predicted by the recommender. Table I resumes the obtained results showing a better overall performance coming from the traditional CF recommender. The only parameter that have better value on cluster recommendations is the error. This can indicate that the LOs clustering recommender is a good rating forecaster but apparently not a good recommender.

However, we ran a Mann-Whitney-Wilcoxon Test [22] to identify if there were a significant difference between these three quality parameters of general recommendations and the relevance of cluster recommendations. The results showed that it is not possible to determine, with a confidence level of 95%, that rating, relevance or difficulty are significantly different among the recommendation types tested.

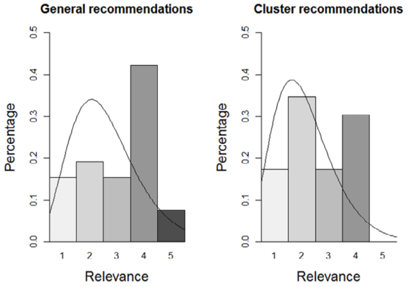

Fig. 5. Histogram for the relevance valuation distribution for general recommendations (left side) and cluster recommendations (right side)

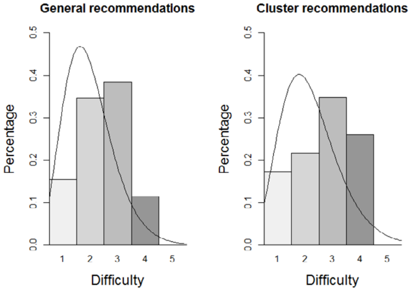

Fig. 6. Histogram for the difficulty rates distribution for general recommendations (left side) and cluster recommendations (right side)

TABLE I. AVERAGE OF THE THREE ANSWERS AND A CALCULATED ERROR

Type	Q3 - Rating	Error	Q1 - Relevance	Q2 - Difficulty
General	3,5	1,3	3,2	2,5
Cluster	3,2	1,1	2,6	2,7

V. CONCLUSIONS

The present work has evaluated two new recommendation engines (both based on CF) proposals and a traditional CF method with datasets extracted from Merlot. We tested two different ways of clustering (pre-processing) and two similarity algorithms. The offline analysis covered not only the quality of generated recommendations but also how many users were able to receive a recommended LO. From that analysis, we extracted the better configuration in order to be evaluated, along with the traditional CF, by real users in the onlineanalysis. During this research we were able to answer the questions previously set.

On the one hand, from the online analysis performed with user feedback, apparently, the traditional CF presented the better means, for all questions asked (rating, relevance and difficulty) but specially for the relevance. However, a Mann-Whitney-Wilcoxon Test of medians proved that there were no significant differences between our new proposal performance and the traditional CF engine.

On the other hand, it is not possible to notice a similarity between the results obtained from the offline analysis and the ones obtained from the users feedback While the offline analysis presented significant differences in favor of our LOs clustering proposal, the users feedback ashowed a performance slightly better by the traditional CF algorithm, however, with no significant difference as it was proved by a test of medians.

The results gathered from the users feedback were not enough definitive to reprove the LOs clustering approach when compared to the traditional CF. Thus, future research is needed towards an ultimate conclusion, specially considering a LOR different than Merlot. Also, we plan on redoing the recommendations for this experiment but with improvements taking into account the evaluations already made, which leads to an idea of a critiquing-based recommender system [23].

ACKNOWLEDGEMENT

This work was supported by CYTED (Ibero-American Programme for Science, Technology and Development) as part of project "RIURE - Ibero-American Network for the Usability of Learning Repositories", code 513RT0471 (www.riure.net) and by FAPERGS (Research SupportFoundation of Rio Grande do Sul – Brazil) through Edital 02/2014 - Programa Pesquisador Gaúcho – PqG (number 2296-2551/14-8).

REFERENCES

[1] R. McGreal, "A typology of learning object repositories," in Handbook on Information Technologies for Education and Training, ser. International Handbooks on Information Systems, H. Adelsberger, Kinshuk,J. Pawlowski, and D. Sampson, Eds. Springer Berlin Heidelberg, 2008, pp. 5–28.

[2] A. Casali, V. Gerling, C. Deco, and C. Bender, "A recommender system for learning objects personalized retrieval," in Handbook on Educational Recommender Systems and Technologies: Practices and Challenges, O. Santos and J. Boticario, Eds. IGI Global, 2012, pp. 182–210.

[3] N. Manouselis, H. Drachsler, K. Verbert, and O. C. Santos, Recommender Systems for Technology Enhanced Learning: Research Trends and Applications. Springer Publishing Company, Incorporated, 2014.

[4] O. Santos and J. Boticario, Handbook on Educational Recommender Systems and Technologies: Practices and Challenges. IGI Global, 2012. [Online]. Available: https://adenu.ia.uned.es/web/en/projects/ersat

[5] X. Ochoa and G. Carrillo, "Recomendación de objetos de aprendizaje basado en el perfil del usuario y la información de atención contextualizada," Conferencias LACLO, vol. 4, no. 1, 2013.

[6] H. Drachsler, T. Bogers, R. Vuorikari, K. Verbert, E. Duval,N. Manouselis, G. Beham, S. Lindstaedt, H. Stern, M. Friedrich, and M. Wolpers, "Issues and considerations regarding sharable data sets for recommender systems in technology enhanced learning," Procedia Computer Science, vol. 1, no. 2, pp. 2849 – 2858, 2010, proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010).

[7] C. Cechinel, S. da Silva Camargo, M.-A. Sicilia, and S. Sánchez-Alonso, "Mining models for automated quality assessment of learning objects," Journal of Universal Computer Science, vol. 22, no. 1, pp. 94–113, jan 2016.

[8] K. Verbert, H. Drachsler, N. Manouselis, M. Wolpers, R. Vuorikari, and E. Duval, "Dataset-driven research for improving recommender systems for learning," in Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ser. LAK '11. New York, NY, USA: ACM, 2011, pp. 44–53.

[9] N. Manouselis, H. Drachsler, R. Vuorikari, H. Hummel, and R. Koper, "Recommender systems in technology enhanced learning," in Recommender Systems Handbook, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer US, 2011, pp. 387–415.

[10] C. Cechinel, M.-A. Sicilia, S. Sánchez-Alonso, and E. García-Barriocanal, "Evaluating collaborative filtering recommendations inside large learning object repositories," Information Processing & Management, vol. 49, no. 1, pp. 34 – 50, 2013.

[11] J. Hylen,´ "Open educational resources: Opportunities and challenges," Proceedings of Open Education, pp. 49–63, 2006.

[12] K. Larsen and S. Vincent-Lancrin, "The impact of ict on tertiary education: advances and promises," 2005.

[13] S. Downes, "Models for sustainable open educational resources," Interdisciplinary Journal of Knowledge and Learning Objects, vol. 3, pp. 29–44, 2007.

[14] H. L. dos Santos, C. Cechinel, R. M. Araujo, and M.-Á. Sicilia, "Clustering learning objects for improving their recommendation via collaborative filtering algorithms," in Metadata and Semantics Research, ser. Communications in Computer and Information Science, E. Garoufallou, R. J. Hartley, and P. Gaitanou, Eds. Springer International Publishing, 2015, vol. 544, pp. 183–194.

[15] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, and C. Newell, "Explaining the user experience of recommender systems," User Modeling and User-Adapted Interaction, vol. 22, no. 4-5, pp. 441– 504, Oct. 2012.

[16] C. Hayes, P. Massa, P. Avesani, and P. Cunningham, "An online evaluation framework for recommender systems," in AH2002 Workshop on Recommendation and Personalization in E-Commerce, 2002, pp. 50–59.

[17] M.-Á. Sicilia, E. García-Barriocanal, S. Sánchez-Alonso, and C. Cechinel, "Exploring user-based recommender results in large learning object repositories: the case of MERLOT," Procedia Computer Science, vol. 1, no. 2, pp. 2859 – 2864, 2010, proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010).

[18] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," ACM Trans. Inf. Syst., vol. 22, no. 1, pp. 5–53, Jan. 2004.

[19] J. Chambers and T. Hastie, "Analysis of variance and designed experiments," in Statistical Models in S, J. Chambers and T. Hastie, Eds. Wadsworth & Brooks/Cole, 1992, ch. 5, pp. 34–47.

[20] M. Ge, C. Delgado-Battenfeld, and D. Jannach, "Beyond accuracy: Evaluating recommender systems by coverage and serendipity," in Proceedings of the Fourth ACM Conference on Recommender Systems. New York, NY, USA: ACM, 2010, pp. 257–260.

[21] S. M. McNee, I. Albert, D. Cosley, P. Gopalkrishnan, S. K. Lam, A. M. Rashid, J. A. Konstan, and J. Riedl, "On the recommending of citations for research papers," in Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work. New York, NY, USA: ACM, 2002, pp. 116–125.

[22] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," in The Annals of Mathematical Statistics, vol. 18, no. 1, 1947, pp. 50–60.

[23] L. Chen and P. Pu, "Critiquing-based recommenders: survey and emerging trends," User Modeling and User-Adapted Interaction, vol. 22, no. 1, pp. 125–150, 2011.