Novel Perspectives in Collaborative Filtering Recommender Systems

Panagiotis Adamopoulos
Department of Information, Operations and Management Sciences
Leonard N. Stern School of Business, New York University
padamopo@stern.nyu.edu

Supervised by Prof. Alexander Tuzhilin

ABSTRACT

This paper proposes a number of studies in order to move recommender systems beyond the traditional paradigm and the classical perspective of rating prediction accuracy. We contribute to existing helpful but less explored recommendation strategies and also propose new approaches aiming at more useful recommendations for both users and businesses. Working toward this direction, we discuss the studies we have conducted so far and present our future research plans. The overall goal of this research program is to move our focus from even more accurate rating predictions and aims at offering a holistic experience to the users by avoiding the over-specialization and concentration problems and providing the users with non-obvious but high quality personalized recommendations that they will remarkably like. In particular, we propose a concept of *unexpectedness* and specific metrics to measure both unexpectedness and quality of recommendations, as well as different algorithms for generating such recommendations. Then, we introduce a method for generating recommendations, which the users will remarkably like, based on *recommendation opportunities* using higher *weighted percentiles*. Finally, we study the satisfaction of students with Massive Open Online Courses (MOOCs) vis-*à*-vis student retention in order to test the aforementioned approaches on generating utility-based recommendations for online courses that better serve the educational needs of students. Apart from MOOCs, we would like to implement and evaluate the proposed approaches also in a traditional online retail setting. As part of the future work, we would also like to generalize the proposed recommender system algorithms to the domains of personalization, web search, and data mining, using the ensemble learning and exploration/exploitation frameworks, as well as to design online randomized experiments and bucket tests with real users.

WWW 2014, April 7-11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/....

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems - Human Factors; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval - Information filtering, Selection process, Retrieval models; I.2.6 [**Learning**]: Parameter learning

General Terms

Algorithms, Design, Experimentation, Human Factors, Measurement, Performance, Theory

Keywords

Algorithm Design; Unexpectedness; Recommender Systems

1. RESEARCH OBJECTIVES

Over the last two decades, a wide variety of different types of recommender systems (RSs) has been developed and successfully used across several domains [8]. During this time, many researchers have focused mainly on the development and improvement of efficient algorithms for more accurate rating prediction. Although the recommendations of the latest class of systems are significantly more accurate than they used to be a decade ago [9] and the broad social and business acceptance of RSs has already been achieved, there is still a long way to go in terms of satisfaction of the actual needs of the users [13]. This is due, primarily, to the fact that many existing RSs focus on providing even more accurate rather than more useful recommendations. Some of the main problems pertaining to this narrow rating prediction focus of many existing RSs [10] and the ways to broaden the current approaches have been discussed in [16].

Even though the aforementioned rating prediction perspective is the prevailing paradigm in recommender systems, there are other perspectives that try to alleviate the problems pertaining to this narrow accuracy-based focus and have been gaining significant attention in the field of RS [11]. In particular, some of the most recent perspectives maintain that RSs should make the users familiar with the various product categories and the whole product catalog. In addition, recommender systems should provide personalized recommendations from a wide range of items and should also enable the users to find relevant items that might be hard to discover. Also, they should increase user satisfaction and engagement and at the same time offer a superior user experience. Moreover, RSs should also be able to reduce user

search costs, improve the quality of decisions that consumers make, and increase their welfare. Besides, from a business perspective, RSs should increase the sales volume and conversion rates, as well as promote items from the long tail that usually exhibit significantly lower marginal cost and, at the same time, higher marginal profit.

Moving beyond the classical perspective of the rating prediction accuracy, the main objective of this stream of research is to contribute to existing helpful but less explored paradigms of RSs as well as to propose new approaches that will result in more useful recommendations for both users and businesses. Working toward this direction, we discuss the studies we have conducted so far and present our future research plans. In particular, we move our focus from even more accurate rating predictions and aim at offering a holistic experience to the users by avoiding the over-specialization and concentration of recommendations and providing the users with non-obvious but high quality recommendation sets (not restricted to simple lists of individual items) that fairly match their interests and they will remarkably like.

2. RELATED WORK

Other streams of research that improve recommender systems going beyond rating prediction accuracy include work on *human-computer interaction* (HCI) [21], which involves the study and design of the interaction between users and RSs, *explanations* for recommendations [20], which provide transparency into the working of the recommendation process exposing the reasoning and data behind each recommendation. Besides, other approaches pertain to *diversification* [7, 22], which maximizes the variety of items in a recommendation list, *group recommenders* [17], which recommend items for groups of people, rather than individuals, and recommendation *sequences* [15], where sequences of ordered items are recommended instead of single items.

3. PROPOSED APPROACH

In this stream of research, we are moving beyond the classical perspective of the rating prediction accuracy and we are aiming at providing even more useful recommendations for both users and businesses. Working toward this direction, we designed a number of research studies to explore issues related to several new approaches for providing recommendations. In particular, Section 3.1 proposes a concept of *unexpectedness* and specific metrics to measure both unexpectedness and quality of recommendations, as well as different algorithms for generating such recommendations. Then, Section 3.2 introduces a method for generating recommendations based on *recommendation opportunities* using higher *weighted percentiles*. Finally, Section 3.3 examines the satisfaction of students with massive open online courses (MOOCs) vis-*à*-vis student retention in order to generate utility-based recommendations using the aforementioned methods and then experimentally evaluate the corresponding approaches in a real-world application. All these research studies go beyond the classical perspective of RSs and aim at providing the users with non-obvious but high quality recommendation sets that fairly match their interests and they will remarkably like. Moreover, the approaches taken for evaluating the results of each work and the results that have been reached so far are also discussed in the corresponding sections. Thorough discussions of these concepts, implementation details, and experimental results are provided in [1, 2, 3, 4, 5, 6].

3.1 Beyond over-specialization

Sections 3.1.1 and 3.1.2 address the over-specialization and concentration problems in recommender systems focusing on *unexpectedness* and the related concepts of coverage, dispersion, novelty, serendipity, and diversity of recommendation lists.

3.1.1 Expecting the Unexpected

In [3, 4], we propose a concept of unexpected recommendations as recommending those items that significantly depart from the expectations of the users and suggest a method for generating such recommendations, based on the utility theory of economics, as well as specific metrics to measure the unexpectedness of recommendation lists.

In particular, we formally define the concept of *unexpectedness* in recommender systems taking into account the actual *expectations* of the users and discuss how the concept of unexpectedness is differentiated from various related notions, such as novelty, serendipity, and diversity. Following the Greek philosopher Heraclitus, we approach this difficult problem of finding and recommending unexpected items by first capturing the items expected by the users. Toward this direction, we suggest several mechanisms for specifying users' expectations that can be applied across various domains. Such mechanisms include the past transactions performed by the users, knowledge discovery and data mining techniques, and experts' domain knowledge. Besides, we formulate and fully operationalize the notion of unexpectedness and present an algorithm for providing unexpected recommendations of high quality that are hard to discover but fairly match the users' interests, based on the *utility theory* of economics. Moreover, we propose specific performance metrics to measure the unexpectedness of the generated recommendation lists taking into account also the usefulness of individual items.

Using "real-world" data sets, various examples of sets of expected recommendations, and different utility functions and distance metrics, we were able to test the proposed method under a large number of experimental settings including various levels of sparsity, different mechanisms for specifying users' expectations, and different cardinalities of these sets of expectations. The empirical study showed that all the examined variations of the proposed method significantly outperformed in terms of unexpectedness the standard baseline algorithms, including item-based and user-based k-Nearest Neighbors [12], Slope One, and Matrix Factorization [14]. This demonstrates that the proposed method indeed effectively captures the concept of unexpectedness since, in principle, it should do better than unexpectedness-agnostic methods. Furthermore, the proposed method for unexpected recommendations performed at least as well as, and in some cases even better than, the baseline algorithms in terms of the classical accuracy-based measures, such as root-mean-square error (RMSE) and the F-measure, as well as other popular performance measures, such as catalog coverage, aggregate diversity, serendipity, and the Gini coefficient. In addition, we presented a number of actual recommendation examples generated by the proposed method

and the employed baseline approaches and provided insightful qualitative comments.

One of the main premises of the proposed method is that the users' expectations should be explicitly considered in order to provide the users with unexpected recommendations of high quality that are hard to discover but fairly match their interests. Hence, the greatest improvements both in terms of unexpectedness and accuracy vis-à-vis all other approaches were observed in the experiments using the more accurate sets of expectations. Moreover, the use of a utility function of standard form illustrates that the proposed method can be easily implemented in existing recommender systems as a new component that enhances unexpectedness of recommendations, without the need to further modify the current rating prediction procedures.

3.1.2 Probabilistic Neighborhoods

In [5], we propose a novel method for recommending items based on probabilistic neighborhood selection in collaborative filtering (CF) models. In particular, we illustrate the practical implementation of the proposed method in the context of memory-based collaborative filtering systems adapting and improve the standard user-based k-nearest neighbors (k-NN) approach. In the proposed variation of the classical k-NN collaborative filtering method, the neighborhood selection is based on an underlying probability distribution, instead of just the k neighbors with the highest similarity level to the target user. For the probabilistic neighborhood selection (k-PN), we use an efficient method for weighted sampling of k neighbors without replacement that also takes into consideration the similarity levels between the target user and all the candidate neighbors. The key intuition for this *probabilistic nearest neighbors* collaborative filtering method is two-fold. First, using the neighborhood with the most similar users to estimate unknown ratings and recommend candidate items, the generated recommendation lists usually consist of known items with which the users are already familiar. Second, because of the multi-dimensionality of users' tastes, there are many items that the target user may like and are unknown to the k most similar users to her/him. Thus, we propose the use of probabilistic neighborhood selection in order to alleviate the aforementioned problems and move beyond the limited focus of rating prediction accuracy.

To investigate this claim, we conducted an empirical study and we tested the proposed method under a large number of experimental settings. In detail, we used a large number of probability distributions from different families of distributions with various location and shape parameters, in order to compare the proposed probabilistic method for neighborhood selection against the standard collaborative filtering approach in terms of popular evaluation metrics for item prediction accuracy, utility-based ranking, coverage, diversity, unexpectedness, and dispersion of recommendations. We also test whether the proposed k-PN method alleviates the common problems of over-specialization and concentration of recommendations in terms of a new proposed metric that measures the mobility of recommendations.

The experimental results illustrate that the proposed approach indeed generates recommendations that are orthogonal to the classical CF method. We also demonstrated that the proposed method performs at least as well as, and in some cases even better than, both the standard user-based k-nearest neighbors and the k-furthest neighbors approaches (k-FN) [19, 18] in terms of popular item prediction accuracy and utility-based ranking measures, such as the F-measure and the normalized cumulative discounted gain, across various experimental settings. These results are also in accordance with the ensemble learning theory that we employ in the neighborhood-based collaborative filtering framework showing the suboptimality of the k-NN algorithm also in terms of predictive accuracy. Besides, we showed that the performance improvement is not achieved at the expense of some other popular performance measures that go beyond the rating prediction accuracy, such as catalog coverage, aggregate diversity, and recommendation dispersion and mobility. Finally, we identified a particular implementation of the k-PN method that performs consistently well across various experimental settings.

The proposed method can be further extended and modified in order to sample k neighbors from the x nearest candidates, instead of all the available users. Also, apart from the user-based and item-based k-NN collaborative filtering approaches, other popular methods that can be easily extended with the use of probabilistic neighborhood selection (k-PN), in order to allow us to generate both accurate and novel recommendations, include Matrix Factorization approaches.

3.2 Recommendation Opportunities

In [6], under a definition of a *recommendation opportunity* as how much a user *could* realistically like an item, we aim at recommending items that the users will remarkably like. Moving beyond the standard perspective of rating prediction accuracy and exploring such recommendation opportunities can increase user satisfaction and engagement and offer a superior user experience through the discovery of items that the users will really like.

In particular, we illustrate the practical implementation of the proposed approach presenting a certain variation of the classical user-based k-NN collaborative filtering method in which the estimation of an unknown rating of the target user for an item is based not on the weighted averages of the k nearest neighbors but on the *weighted percentile* of the ratings of these k neighbors. For the estimation of the weighted percentile of the distribution of the ratings in the neighborhood of the target user, an efficient method is used that does not increase the computational complexity of the classical k-NN method. The key intuition behind the weighted percentile method, instead of using weighted averages, is that high percentiles (such as in the 70% to 90% range) constitute more realistic estimates of how much a targeted user *could* possibly like the candidate item. As a consequence, the proposed approach not only provides more useful for the users recommendations of items that they remarkably like but also has the potential to let us better identify and serve any specific niches of the market.

To support this claim, we conducted an empirical study and showed that the proposed percentile method outperforms by a wide margin, across various experimental settings, the standard user-based CF approach in terms of item prediction accuracy measures, such as precision, recall, and the F-measure, and also utility-based ranking metrics, such as normalized cumulative discounted gain and mean reciprocal rank. Finally, we demonstrated that this performance improvement is not achieved at the expense of some other popular performance measures, such as catalog coverage, ag-

gregate diversity, and the Gini coefficient. This illustrates that our proposed weighted percentile method for *recommendation opportunities* performs at least as well as, or even significantly better than, the classical user-based collaborative filtering method in terms of these important measures, in most of the experiments.

Nevertheless, apart from the user-based and item-based k-NN CF approaches, other popular RSs methods that can be easily extended, with the use of quantile regression, in order to allow us both to build models that predict high percentiles and to evaluate them with regard to the goal of predicting percentiles of estimated ratings, include content-based methods and Matrix Factorization.

3.3 Massive Open Online Learning

In [2], we study the satisfaction of students with *Massive Open Online Courses* (MOOCs) vis-*à*-vis student retention in order to use the aforementioned approaches and generate utility-based recommendations for online courses that better serve the educational needs of students.

In particular, MOOCs have remarkably expanded during the last years offering a rapidly growing number of courses across various platforms. However, the very high drop-out rates indicate that much more should be done in order to satisfy the actual educational needs of the students. Tackling this important problem, we employ the *Grounded Theory Method* (GTM) on quantitative data, a less frequently applied paradigm. In particular, we present a novel analysis using a real-world data set with user-generated online reviews, where we both identify the different *student*, *course*, *platform*, and *university* characteristics that affect student retention and study their relative effect. An important aspect of the conducted analysis is the integration of state-of-the-art *econometric, text mining, opinion mining, and machine learning techniques*, building both explanatory and predictive models, toward a more complete and in-depth analysis of the information captured by user-generated content. Extending the methodological ideas of GTM to explanatory quantitative analysis and predictive models and going beyond descriptive statistics of coded verbal data, we both contribute to the related literature discovering new rich findings and provide actionable insights with implications for both MOOCs and education. The use of radically different approaches and fresh interdisciplinary perspectives in this paper offers the potential to inspire future research and open up new streams of scientific inquiry.

The proposed approach and the corresponding findings have several implications for the universities and platforms offering online courses and can be used for real-time detection of dissatisfied students as well as the design of better and more engaging courses that will increase retention rates. In detail, the results suggest that the course characteristics (e.g. estimated difficulty, workload, duration, whether there is automated grading, etc.) are important determinants of students' satisfaction and suggest useful guidelines for course design. For instance, MOOCs in general should have a specific instructor-based timetable, but for the most difficult courses students should be allowed to follow their own pace. Also, the findings suggest that there is room for improvement in the current form of certifications which should be redesigned in order to become more useful for the students and better motivate them to successfully complete the corresponding course. Similarly, better technical solutions are needed for automatically providing feedback and evaluating the assignments of the students. Moreover, our results also illustrate that, in addition to discussion forums, better mechanisms or complementary technologies, such as wikis, are still needed in order to successfully advice, assist, connect, and motivate the students. In addition, our results show that open textbooks, in contrast to paid textbooks, have a positive effect. Finally, apart from the implications in education and the methodological ideas of GTM, we also discuss the managerial implications of the proposed analytical approach in other domains, such as social media and online commerce.

4. CONCLUSION AND FUTURE WORK

Successfully completing the aforementioned work will help the recommender systems field move further beyond the perspective of rating prediction accuracy. Following the proposed stream of research that both contributes to existing helpful but less explored paradigms for recommender systems and proposes new valuable approaches and perspectives, we discussed the studies we have conducted so far and also presented in detail some of our future research plans. In particular, the studies discussed in Section 3 move our focus from even more accurate rating predictions and aim at offering a holistic experience to the users by avoiding the over-specialization and concentration of recommendations and providing the users with non-obvious but high quality personalized recommendations that fairly match their interests and they will remarkably like.

In detail, Section 3.1 proposes a concept of *unexpectedness* and specific metrics to measure all unexpectedness, concentration, and quality of recommendations, as well as different algorithms for generating such recommendations, Section 3.2 introduces a method for generating recommendations that the users will remarkably like based on *recommendation opportunities* using higher *weighted percentiles*, and Section 3.3 examines the satisfaction of students with massive open online courses (MOOCs) vis-*à*-vis student retention in order to generate utility-based recommendations using the aforementioned methods and then experimentally evaluate the corresponding approaches in a real-world application.

Adhering to our main research objective, we work toward the direction of providing more useful recommendations for both users and businesses. Avoiding obvious and expected recommendations while maintaining high predictive accuracy levels, we can alleviate the common problems of over-specialization and concentration bias that often characterize the CF algorithms. Moreover, building such a recommender system, we also have the potential to further increase user satisfaction and engagement and offer a superior experience to the users. Besides, the proposed approach exhibits a potential positive economic impact based on (i) the direct effect of increased sales and enhanced customer loyalty through offering more useful for the users recommendations from a wider range of items, enabling them to find relevant items that are harder to discover, and making the users familiar with the whole product catalog, and (ii) the indirect effect of recommending items from the long tail and not focusing mostly on bestsellers that usually exhibit higher marginal costs and lower profit margins because of acquisition costs and licenses as well as increased competition.

As part of the future work, we would like to further optimize and streamline the proposed methods as well as to integrate them with related existing approaches in the fields of web search and data mining. Moreover, we would like to implement and evaluate the proposed approaches in a traditional on-line retail setting as well as in a platform for MOOCs. Besides, we would like to conduct a series of live controlled experiments with human subjects in order to study the on-line user behavior, examine and actively adjust the trade-off between exploration (e.g. unexpectedness, serendipity, diversity, etc.) and exploitation (e.g. accuracy) of recommender systems, and better evaluate the proposed perspectives in a user-centric framework for top-N reocmmendations. In addition, we plan to investigate how the proposed recommendation approaches and perspectives can be effectively combined with traditional approaches in hybrid recommender systems aiming at accurate and non-obvious recommendation lists that the users will remarkably like.

5. REFERENCES

[1] P. Adamopoulos. Beyond Rating Prediction Accuracy: On New Perspectives in Recommender Systems. In *Proceedings of the seventh ACM conference on Recommender systems*, RecSys '13, New York, NY, USA, Oct. 2013. ACM.

[2] P. Adamopoulos. What Makes a Great MOOC? An Interdisciplinary Analysis of Student Retention in Online Courses. In *Proceedings of the 34th International Conference on Information Systems*, ICIS '13, 2013.

[3] P. Adamopoulos and A. Tuzhilin. On Unexpectedness in Recommender Systems: Or How to Expect the Unexpected. In *DiveRS 2011 - ACM RecSys 2011 Workshop on Novelty and Diversity in Recommender Systems*, RecSys 2011. ACM, Oct. 2011.

[4] P. Adamopoulos and A. Tuzhilin. On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected. *Working Paper: CBA-13-01, New York University*, 2013. http://ssrn.com/abstract=2282999.

[5] P. Adamopoulos and A. Tuzhilin. Probabilistic Neighborhood Selection in Collaborative Filtering Systems. *Working Paper: CBA-13-04, New York University*, 2013. http://hdl.handle.net/2451/31988.

[6] P. Adamopoulos and A. Tuzhilin. Recommendation Opportunities: Improving Item Prediction Using Weighted Percentile Methods in Collaborative Filtering Systems. In *Proceedings of the seventh ACM conference on Recommender systems*, RecSys '13, New York, NY, USA, Oct. 2013. ACM.

[7] G. Adomavicius and Y. Kwon. Improving aggregate recommendation diversity using ranking-based techniques. *Knowledge and Data Engineering, IEEE Transactions on*, 24(5):896 –911, may 2012.

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

[9] R. M. Bell, J. Bennett, Y. Koren, and C. Volinsky. The million dollar programming prize. *IEEE Spectr.*, 46(5):28–33, 2009.

[10] P. Cremonesi, F. Garzotto, S. Negro, A. V. Papadopoulos, and R. Turrin. Looking for "good" recommendations: a comparative evaluation of recommender systems. In *INTERACT '11*, 2011.

[11] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich. *Recommender systems: an introduction*. Cambridge University Press, 2010.

[12] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. Grouplens: applying collaborative filtering to usenet news. *Comm. of the ACM*, 40(3), 1997.

[13] J. A. Konstan and J. T. Riedl. Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22:101–123, 2012.

[14] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8), 2009.

[15] J. Masthoff. Group recommender systems: Combining individual models. In *Recommender Systems Handbook*, pages 677–702. Springer, 2011.

[16] S. M. McNee, J. Riedl, and J. A. Konstan. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *Proceedings of CHI '06*, pages 1097–1101. ACM, 2006.

[17] M. O'Connor, D. Cosley, J. A. Konstan, and J. Riedl. Polylens: a recommender system for groups of users. In *Proceedings of ECSCW '01*. Kluwer Academic Publishers, 2001.

[18] A. Said, B. Fields, B. J. Jain, and S. Albayrak. User-centric evaluation of a k-furthest neighbor collaborative filtering recommender algorithm. In *Proceedings of the ACM 2013 conference on Computer Supported Cooperative Work, ACM (New York, NY, USA, 2013)*, 2013.

[19] A. Said, B. J. Jain, B. Kille, and S. Albayrak. Increasing diversity through furthest neighbor-based recommendation. In *Proceedings of the WSDM'12 Workshop on Diversity in Document Retrieval (DDR'12)*, 2012.

[20] N. Tintarev and J. Masthoff. Designing and evaluating explanations for recommender systems. In *Recommender Systems Handbook*, pages 479–510. Springer, 2011.

[21] K.-H. Yoo, U. Gretzel, and M. Zanker. *Persuasive Recommender Systems: Conceptual Background and Implications*. Springer, 2013.

[22] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, and Y.-C. Zhang. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107(10):4511–4515, 2010.