Non-Personalized Movie Recommendation by Maximum k-Coverage

Nicollas Silva, Diego Carvalho, Adriano C. M. Pereira,
Fernando Mourão, Leonardo Rocha

Responses to Reviewers' Comments

We thank the reviewers for all comments that helped to improve considerably our manuscript. Next, we present our responses to specific comments of the two reviewers.

Reviewer #1's Comments

<u>Comment #1:</u> **Concerning the definition of the problem, in the letter submitted with the paper, the authors distinguish the "ramp-up" and "cold start" problems. I think I can summarize their explanation as follows: the ramp-up problem is when you have no ratings at all while, in the cold start problem there are some, although few ratings. As the authors acknowledge, the problems are highly related. In the recommender systems literature, as far as I know, the "ramp-up problem" term is not used but the term "new user problem" is. See, for instance, the following survey, which is cited in your paper:**
Bobadilla, J., Ortega, F., Hernando, a., & Gutiérrez, a. (2013).
Recommender Systems Survey. Knowledge-Based Systems, 46, 109–132.
https://doi.org/10.1016/j.knosys.2013.03.012

I still think this needs to be more deeply discussed in the paper. As I mentioned in my previous review, a mathematical formalization of the problem would probably be very helpful. In any case, I will make a suggestion. If you take the perspective of:
A. M. Rashid, I. Albert, D. Cosley, S. K. Lam, S. M. McNee, J. A. Konstan, and J. Riedl, "Getting to know you: learning new user preferences in recommender systems," in Proceedings of the 7th international conference on Intelligent user interfaces, pp. 127–134, ACM, 2002.

which is cited in the paper, I think you could present your work as supporting a variant of the strategies for item selection: minimize user effort and maximize recommendation accuracy, focusing on niche users.

A. Originally, the term Cold-start was used to denote scenarios of recommending items that no one has yet rated [1]. Recently, however, several works refer to Cold-start as the problem of recommending items to target users with few ratings [2, 3, 4, 5, 6]. In order to better explain the problem addressed in our work, we have added a more detailed explanation of the problem in Sections 1 (Introduction) and 2 (Related Work), respectively.

In Introduction, we add: *"This problem is known in the literature in two ways: (1) Cold-Start problem; And (2) Ramp-up problem. The Cold-Start problem is related to generating recommendations for new users, whose consumption history is small and* *little relevant [9, 10]. On the other hand, the Ramp-up problem is even more complicated, since it is related to first-time users, for whom there is still no information in the system [11]."*

In Related Work, we add: *"The literature in Recommender Systems defines this problem as: (1) Cold-Start problem; and (2) Ramp-up problem. The Ramp-up problem is commonly deemed as a variation of the Cold-Start problem [10]. Despite being closely related, both problems should be addressed differently. Whereas the Cold-Start problem deals with users with small consumption histories (i.e., inactive or new users), in the Ramp-up there is no consumption information about the users (i.e., first-time users). For e-commerce systems, any information is better than none and, for this reason, Ramp-up is a major challenge for Recommender Systems."*

[1] Robin Burke. 2002. Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction 12, 4 (2002), 331–370.
[2] Iman Barjasteh, Rana Forsati, Farzan Masrour, Abdol-Hossein Esfahanian, and Hayder Radha. 2015. Cold-start item and user recommendation with decoupled completion and transduction. In Proceedings of the 9th ACM Conference on Recommender Systems. ACM, 91–98.
[3] Antonio Hernando, Jesús Bobadilla, Fernando Ortega, and Abraham Gutiérrez. 2017. A probabilistic model for recommending to new cold-start non-registered users. Information Sciences 376 (2017), 216–232.
[4] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the cold start problem in recommender systems. Expert Systems with Applications 41, 4 (2014), 2065–2073.
[5] Suvash Sedhain, Scott Sanner, Darius Braziunas, Lexing Xie, and Jordan Christensen. 2014. Social collaborative filtering for cold-start recommendations. In Proceedings of the 8th ACM Conference on Recommender systems. ACM, 345–348.
[6] Chirayu Wongchokprasitti, Jaakko Peltonen, Tuukka Ruotsalo, Payel Bandy-opadhyay, Giulio Jacucci, and Peter Brusilovsky. 2015. User model in a box: Cross-system user model transfer for resolving cold start problems. In International Conference on User Modeling, Adaptation, and Personalization. Springer, 289–301.
[7] Konstan, J. A., Riedl, J., Borchers, A. and Herlocker, J. L.: 1998, 'Recommender Systems: A GroupLens Perspective.' In: Recommender Systems: Papers from the 1998 Workshop (AAAI Technical Report WS-98-08). Menlo Park, CA: AAAI Press, pp. 60^64
[8] Nguyen, An-Te, Nathalie Denos, and Catherine Berrut. "Improving new user recommendations with rule-based induction on cold user data." Proceedings of the 2007 ACM conference on Recommender systems. ACM, 2007.
[9] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE transactions on knowledge and data engineering, vol. 17, no. 6, pp. 734–749, 2005.
[10] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," Knowledge-Based Systems, vol. 46, pp. 109–132, 2013.
[11] R. Martin, M. Walid, W. Robert, and Z. Thomas, Recommendation Systems in Software Engineering. Springer, 2014.

<u>Comment #2:</u> Concerning the section on related work:

- in the related work, the authors categorize papers in three classes. Isn't there a

taxonomy of methods that can be reused?

- the argument against questionnaires is not very convincing, as you may simply ask: list 5 items you like and 5 times you don’t

- the argument in favor of non-personalized RS is a bit contradictory with the motivation of this work (namely “generalization capability” and “good performance”).

A. The taxonomy used in the Related Work comes from the main works related to the Ramp-up problem, as well as from the surveys of Recommender Systems, as mentioned in Section 2 [1,2,3,4,5,17]. On the other hand, the main argument for using non-personalized strategies is related to the problem addressed from a point of view of e-commerce markets. Often, in the context of e-commerce markets, users are only interested in making purchases, without providing personal or demographic information. For instance, Amazon's system does not ask users to tell about 5 items they like and 5 they do not like. In general, the key worldwide market players systems have chosen to use non-customized strategies that can attract multiple first-time user profiles. The main reason for this choice lies in the characteristics of these strategies, as discussed in Section 2 of the article (Related Work):
“*Simplicity, generalization capability, domain independence and good performance are characteristics that make non-personalized RSs the main strategy to address the ramp-up problem in practical scenarios*.”

In other words, these strategies can deal with this scenario by ignoring the target user's past, and taking into account the overall context of the system (i.e., generalizing information from users and/or items).

[1] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, “Recommender systems survey,” Knowledge-Based Systems,
vol. 46, pp. 109–132, 2013.
[2] C. He, D. Parra, and K. Verbert, “Interactive recommender systems: A survey of the state of the art and future
research challenges and opportunities,” Expert Systems with Applications, vol. 56, pp. 9–27, 2016.
[3] 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, vol. 107, no. 10, pp. 4511–4515,
2010./
[4] S. Loh, F. Lorenzi, R. Granada, D. Lichtnow, L. K. Wives, and J. P. M. de Oliveira, “Identifying similar users by
their scientific publications to reduce cold start in recommender systems.,” in WEBIST, vol. 9, pp. 593–600, 2009.
[5] J. B. Schafer, J. Konstan, and J. Riedl, “Recommender systems in e-commerce,” in Proceedings of the 1st ACM
conference on Electronic commerce, pp. 158–166, ACM, 1999.
[6] S. A. Puthiya Parambath, N. Usunier, and Y. Grandvalet, “A coverage-based approach to recommendation diversity
on similarity graph,” in Proceedings of the 10th ACM Conference on Recommender Systems, pp. 15–22, ACM, 2016.

<u>Comment #3:</u> The explanation of the method could be significantly improved. First of all, the Maximum Coverage method should be explained when it is introduced, even if in an informal way. Furthermore, there are several smaller issues, including:

- in page 4 it is stated that "The problem is to determined the subset F_k^* of size k…" It's a subset of which set?

- Algorithm 1 is confusing, especially line 4

- the caption of fig.1 doesn't seem appropriate.

An additional issue concerning the method is the computational complexity, which is cubic. This should not be claimed as a scalable strategy. In fact, execution time results should be included.

A. We agree with this reviewer´s suggestion and have updated the definitions of the Maximum k-Coverage problem to make it easier for readers to understand it. In addition, a formal definition of the problem has been included, as can be seen at the beginning of the Section 3 of the article. The caption of Figure 1 was updated to better describe the figure, as requested. In addition, Table III was added in Section 5.3 (Application on Real Scenarios), showing the execution time of each strategy used. This analysis shows that Maximum k-Coverage is scalable and can be used in real scenarios.

<u>Comment #4:</u> Formalize all the measures correctly and motive why they are used. Concerning the description of the evaluation measures, accuracy is defined in a way which is, according to Bobadilla, J., Ortega, F., Hernando, a., & Gutiérrez, a. (2013) a set recommendation metric. Then, the explanation of diversity is not clear as well.

A. We adopt the formalism used in Bobadilla, J., Ortega, F., Hernando, a., & Gutiérrez, a. (2013), as requested and also described each strategy motivation. These changes can be seen in Section 4.3 of the article (*Quality Requeriments*).

<u>Comment #5:</u> Motivate the new measures adequately and contextualize it in the literature The F-measure is a well established measure in information retrieval. It's not clear why a new F-measure is proposed here. It also seems related to the "balanced strategies" in A. M. Rashid, I. Albert, D. Cosley, S. K. Lam, S. M. McNee, J. A. Konstan, and J. Riedl (2002).

A. The use of the F-measure metric is commonly used in the literature to make a trade-off between precision and recall. This metric consists of a harmonic mean between these two other metrics. The new F-measure proposed is just a harmonic mean between two distinct metrics of diversity and accuracy, as can be seen in Equation (4) of Section 4.3. We do not intend to use such metrics in practice, but rather to only evaluate the performance of the recommendations against these two quality requirements. High values indicate that the recommender is able to present a useful and diverse set of items to the users.

<u>Comment #6:</u> Explain research questions more carefully the research questions investigated in each part of the study should be explained more carefully in the introduction of section 5.

A. We added in the beginning of Section 5 the main objectives of each of our following analyzes. We made clear the need to compare the items presented by each strategy, the performance of the evaluated algorithms, and, finally, the practical use of these approaches. With these analyzes we have been able to answer the main questions raised in the Introduction:
"(1) Considering a product catalog, which items have the highest potential to turn a given first-time user into a returning one?" and *"(2) How to recommend available items in order to retain the maximum number of first-time users?"*

We answer (1) and (2) when we show that presenting relevant items to as many different users as possible is the best way to satisfy first-time users.

<u>Comment #7</u>: Compare the proposed method with a method that recommends items for the masses with a small number of random items.

A. To better define the performance of the Maximum k-Coverage strategy, we have implemented and analyzed the performance of a random approach, as suggested. We implemented the Random Popularity strategy, which is defined by this work as: "Random Popularity - *aims to recommend k random items within the group of items that are rated as popular. The popular items group is defined as the items present at the head of the popularity distribution, which are a percentage of items in the domain [1]. This strategy is not used in practice, but is used to compare whether proposed strategies are effective in selecting potentially relevant items, or whether only a random selection would be sufficient."*

The application of this approach has allowed us to observe that the Maximum k-Coverage strategy is able to find not only several items, but several relevant items. This observation is clear when we use the metric of new F-measure proposed by this work, as shown in Figure 6. The high level of diversity obtained by non-relevant items is not able to satisfy users.

[1] C. Anderson, The long tail: Why the future of business is selling less of more. Hyperion, 2006.

<u>Comment #8:</u> Extend the experimental setup by testing more algorithms or analyzing the robustness of the method by manipulating the dataset used.

A. This question was answered in comment # 7.

<u>Comment #9:</u> Include comparison in terms of execution time.

A. This analysis was included in Section 5.3 (Application on Real Scenarios), as answered in comment # 3.

__Comment #10:__ **It is not clear why the scenarios of 5, 10 and 20 are realistic and the others aren't.**

A. In this work, we consider the scenarios recommending 5, 10, and 20 items as more realistic scenarios, since the major e-commerce systems are limited to user´s screen space. In general, real e-commerce players, like Amazon.com and Netflix, have a maximum of 20 items at a time for each user.

Reviewer #2's Comments

__Comment #1:__ **The paper presents a non-personalized approach for movie recommendation based on Maximum Coverage. The approach is interesting and the results shows competitive results with respect to those based on popularity and others. It would have been important to compare (or justify why it is not compared) with the approaches described in Section 2.1.**

A. The response to this issue is covered in the answer for Comment #2 of Reviewer 1.

__Comment #2:__ **Also, in many graphics the approach performs a little worse than others (like Fig 4 and some of Fig 7). The reasons for this should be further explained and justified.**

A. The performance of each strategy is related to its strengths and weaknesses. In general, the TopRated and Popularity strategies present more items that are relevant, therefore they achieve high levels of accuracy, as can be seen in Figure 4 of Section 5.2. As stated in the first paragraph of Section 5.2, Popularity and TopRated are able to present relevant items to users, since the studied scenarios are related to mass consumption. On the other hand, the other strategies aim to diversify the presented items, obtaining more diversity, as shown in Figure 5. This result is related to the strategies that aim to select random (Random Popularity) or even items that interest different users (Maximum k -Coverage). This explanation has been added in the second paragraph of Section 5.2. For this same reason, Maximum k-Coverage achieves a high performance in the new F-measure metric, as shown in Figure 6 of Section 5.2.

__Comment #3:__ **More importantly, in many cases that differences in some metrics are small. The statistical significance of differences should be included in the analysis of results.**

A. In this work, we use the kolmogorov test for non-normal distributions, and as detailed in Table II of Section 5.2, the Maximum k-Coverage strategy presents statistical gains related to the other strategies.

__Comment #4:__ **Finally, conclusions should be more compelling a proper literature review of strategies designed for dealing with such issue.**

A. The conclusion of the article (Section 6) was improved to attend this request.

Comment #5: **Regarding presentation, there are several things to correct. Section 2.1 is a unique subsection in Sec 2, the text should be split in at least 2 subsections.**

A. Fixed.

Comment #6: **In the end of Sec 1 the organization of the work should be included.**

A. Fixed.

Comment #7: **The quality of Fig 1 is poor.**

A. Fixed.

Comment #8: **There are a number of errors to correct, for example:**

- In last paragraph of sec 2.1: “consider aims”

- Sec 5, 1st paragraph: “we perform” --> “we performed”

A. Fixed.

--//--