Analysing Recommender Systems Impact on Users' Choices

Naieme Hazrati
Free University of Bolzano-Bozen
Bolzano, Italy
nhazrati@unibz.it

Mehdi Elahi
Free University of Bolzano-Bozen
Bolzano, Italy
meelahi@unibz.it

Francesco Ricci
Free University of Bolzano-Bozen
Bolzano, Italy
fricci@unibz.it

ABSTRACT

In this paper we introduce a novel model for simulating the choice making procedure of users under the influence of a Recommender System (RS). Our model leverages the knowledge of users' preferences and simulates repeated choices. We investigate the evolution of these simulated choices in the presence of different RSs and analyse their impact on the Gini index, as indicator of choice diversity. Running the simulation we have observed that all the considered RSs increase the awareness of the users about the items while they affect the aggregated choice diversity differently.

KEYWORDS

Recommender systems; Decision Making; Dynamics

1 INTRODUCTION

Recommender systems (RSs) are software tools that serve users with personalized suggestion of items that are predicted to suit their specific needs and constrains [10]. RSs are typically evaluated in terms of their accuracy; how precise they are in suggesting the items that the user actually chose. However, there are other important aspects which have drawn much less attention. One of them is related to how RSs can affect users' choice making process, which can be investigated at the level of a single user or a users' community; in this paper we focus on the latter. Few prior works studied the impact of a RS at the community level by simulating the choices of the users under the influence of the RS [4, 5, 12, 14, 15], and also investigating how different recommendation approaches, with different characteristics and bias, may affect users' choices. While simulating user behaviour is not trivial at all and one is forced to make simplifying assumptions to make the problem tractable, previous studies suffer from limitations that we believe may cause their outcomes to be limited in portraying a realistic picture of the impact of RSs in normal use. Hence, more research is needed. In one of these prior works the authors base their simulation on the knowledge of a data set of ratings and simulated alternative scenarios where the users, in a given period of time, are supposed to rate (with a predicted or true value) certain recommended items instead of those that he rated in reality [12]. For instance, the simulated users rate in a round a number of system recommended items equal to the number of ratings they have actually entered in that round; information that is stored in the rating data set. In that way, by observing the rating data set, the simulation can effectively come close to the real evolution of the user ratings. The cons of this approach is that the simulation cannot be used to predict what the users will rate or choose in the future, which is our main goal.

ImpactRS Workshop, September 2019, Copenhagen, Denmark

Moreover, the authors do not make distinction between recommendation and choice; what is recommended is rated. This unrealistic and deterministic choice model limits the simulation in portraying real world scenarios in which the users do not always accept the recommended items. Moreover, most of the prior researches that have studied the effect of RSs on users' behaviour have analysed the diversity of their choices [4, 5, 12]. While there are some famous and widely used diversity metrics [2, 8, 13], few works have proposed metrics that capture choice diversity evolution and compared alternative RSs with respect to the metrics. One example is [1] where the authors focus on the concentration reinforcement bias of RSs. They propose metrics that can consider the prior popularity of the items, i.e., before the RS influences the users' choices, and measure whether the popularity is reinforced or alleviated after items are recommended. In our study we also aim at better understanding the complex interrelationship between choice diversity, item popularity and their dynamic evolution. With the aim of better modeling choice behaviour when it is influenced by RSs, we propose a novel simulation design, which is inspired by [4, 5], but gives a more realistic and predictive picture of the choice making procedure of a large user community. The most significant differences are a better model of the users' preferences (extracted from a rating data set) and the usage of a real large data set of users and items. In fact [4] instead used a synthetic data set composed by a small number of artificial items and users. Compared with [12] we have used a more realistic choice model where users are not forced to accept recommendations and the rating data set is used only for defining the user preferences. No additional information is extracted from the data set by observing data points related to the simulated period, hence our model can be actually used for predicting the future evolution of a choice/rating data set, starting from an initial observed data set. Hence, addressing these limitation has led to a more realistic simulation and, we believe, a better prediction of the dynamics of the choices of real users. We have considered scenarios where simulated users choose items when recommendations are offered to them by different RSs and we have compared these scenarios with a case when no recommendation is provided. We have studied the global dynamics of the users' simulated choices and analysed the distribution of these choices, as measured by the Gini index. In general, we have observed that while all the considered RSs increase the awareness of the users of items not previously chosen, they have a different impact on the users' choices and hence they result in different Gini index values. This result is in contradiction with prior research that reports increasing choice concentration caused by RSs [4]. Moreover, we discovered that while the Gini indices of the choices produced by two RSs may be similar they can be obtained by recommending items with different popularity.

2 SIMULATION OF USERS' CHOICES

We denote with U the set of users and I the set of items. The preferences of the users for the items are assumed to be stored in a $|U| \times |I|$ matrix $\hat{R}$, where $\hat{r}_{uj}$ indicates the predicted rating of item j for user u; this can be obtained, for instance, by predicting missing ratings in a given (sparse) rating matrix R. Let P be the $|U| \times |I|$ choice matrix where an element of this matrix p_{uj} is 1 if the user u has chosen the item j, and $p_{uj} = 0$ otherwise. The matrix P is derived from the matrix R; we assume that rated and chosen items coincide. With p_u we will denote the u-th row vector of the matrix P and with $P_u = \{j : p_{uj} = 1\}$ the set of items that user u has chosen (the column indices of the entries in p_u that are equal to 1).

The choices of the users are time stamped: t_{uj} is the time when the user u chose item j. Assume that t_0 is a selected time point; we denote with P^0, the "initial" choice matrix, formed by all the choices p_{uj}, s.t. $t_{uj} \leq t_0$. The simulation procedure starts from this initial knowledge and is aimed at simulating users' choices made after this time point. We will therefore consider successive time intervals, after one or more months from the time point t_0. So, for instance $]t_0, t_1]$ denotes the time interval spanning from t_0 (excluded) to t_1 (included), and t_1 is a time point one month after t_0. The simulation iterates on these time intervals to identify $\hat{P}^l$, that is the matrix of the simulated choices in $]t_{l-1}, t_l]$, and the aggregated simulated choices $Q^l = P^0 + \hat{P}^1 + \cdots \hat{P}^l$. In principle, one could compare the real choices made in a time interval [12], i.e., for instance P^1, with the predicted ones $\hat{P}^1$, but in this short paper we will not discuss this aspect, and we will only focus on the analysis of choice diversity observed in $\hat{P}^l$ and Q^l, by using the Gini metric [11].

2.1 Awareness and Choice Model

We assume that users are not aware of the entire catalogue of the items and can only choose items in a set called *Awareness Set*, A_u^l; it contains the items that the user u can choose in l-th time interval, $]t_{l-1}, t_l]$. An item j is added to or removed from the awareness set A_u^l in the following cases:

- if user u chooses item j in that time interval, then j is removed from the u's awareness set (because we do not want to simulate multiple choices of the same item);
- if the item j is recommended to user u at that time interval, then j is added to the awareness set A_u^l;
- if j is among the top-k most popular chosen items in the previous time intervals, then j is included in the awareness set. The entering of top-k most popular chosen items to all users' awareness sets is due to the assumption that the users are aware of the top popular items. It is indeed similar to the real world cases were the most popular items are usually known by the users.

During a time interval a user is given the chance to make some choices (for items). We assume that a user decides to choose an item (only once) according to a probabilistic model. The utility of the item j for the user u is assumed to be equal to the estimated rating of user u for item j: $v_{uj} = \hat{r}_{uj}$. The user u chooses an item j among those in the awareness set A_u^l, with the following probability:

$$p(u \ chooses \ j) = \frac{v_{uj}}{\sum_{k \in A_u^l} v_{uk}} \tag{1}$$

2.2 Recommendations

The following three RSs are considered and implemented:

- α_1: is a neighborhood-based CF RS that computes the cosine similarity between the 0/1 choices' vector of a target user u, q_u^l, and the choice vector of the other users to find the nearest neighbors. The most popular item among the choices of the nearest neighbor users is recommended to the target user.
- α_2: is similar to α_1, but it penalizes the score of the recommended items multiplying it with the inverse of their popularity.
- α_3: is a Factor Model (FM) RS which generates a recommendation following the approach proposed in [6].

If the item j is recommended to the user u, by a RS, then the utility v_{uj} is boosted by a constant value δ, i.e., $v_{uj} = v_{uj} + \delta$, before the choice simulation takes place. In this way the recommended item becomes marginally more likely to be chosen by the user, compared with an item with the same (estimated) utility.

2.3 Simulation Procedure

We assume that in each time interval $]t_{l-1}, t_l]$, $l = 1, \ldots L$, the users make z_l choices, and we simulate two types of scenarios:

- Scenario s_0: there is no recommendation;
- Scenario s_1, s_2 and s_3: before a user chooses an item, one of the recommenders α_1, α_2 or α_3 recommends an item to her.

So, suppose that time is t_0, we have observed P^0 choices, then we generate $\hat{P}^1$, the 0/1 matrix of size $|U| \times |I|$ that contains the simulated choices made in the interval $]t_0, t_1]$. In order to do that z_1 users are sampled (with replacement) and inserted in a list. The probability that a user u is selected is proportional to the number of choices that she has made until t_0.

Then, for each user in that list, if we consider a scenario where a RS is active, a recommendation j^* for the user is generated and the utilities of all the items in the awareness set of this user are computed; the recommended item has a boost in utility and the other items have the standard utility. Thereafter, user u chooses an item j_c according to the choice model in Eq.1, this choice is inserted into the choice matrix $\hat{P}^1$ and removed from the awareness set A_u^1. The simulation continues to the next iteration to create and fill $\hat{P}^2$ with the choices in $]t_1, t_2]$, while the RSs use $Q^1 = P^0 + \hat{P}^1$ in order to generate recommendations, and so on so forth. Eventually, after L steps we have generated $\hat{P}^L$ which contains the predicted choices of the users in $]t_{L-1}, t_L]$. Moreover, the matrix Q^L contains all the simulated choices till the L-th step.

3 EVALUATION AND RESULTS

We have used the *MoiveLens* 100*K* data set which contains 100,000 ratings provided by 943 users for 1,682 items. The ratings span on an 8 month period. We used this data set in order to form the $|U| \times |I|$ sparse rating matrix R and adopted Factorization Machine [9] in order to predict the missing ratings and generate $\hat{R}$. It is worth noting that an alternative utility prediction method may affect as well the dynamics of the user choices [7]. This aspect will be analysed in a future work.

We formed the initial choice matrix P^0 by considering all the ratings/choices with time-stamp within the first 4 months: 496 users and 1467 items. We then simulated the choices of the users month-by-month, for 4 successive months. In our simulation, the δ parameter that determines the boosting in utility for a recommended item, is set to 0.3 for all RSs.

We have observed the evolution of the Gini index [3] in the different recommendation scenarios as the simulation proceeds, i.e., in the four successive time intervals: $Q^1, \ldots, Q^4$. Initially, the Gini index is the same for all the scenarios (0.61), but over the time, in the scenarios s_0, s_1 and s_3 it increases a bit (0.63/0.64), without differing too much. This indicates that the recommenders α_1 and α_3 do not generate bigger concentration of the choices in comparison to the scenario where no RS is influencing the users' choices. Conversely, as expected, we have observed a substantial decrease of Gini index in scenario s_2 (0.35 at the end of the fourth month of simulation, Q^4). This means that the recommender α_2, which penalises the popular items, can actually improve choice diversity.

We were a bit surprised to note that Gini indices in the scenarios s_1 and s_3 where very similar, notwithstanding the differences in the RSs. Hence, we analysed the Gini index computed on the choices made by users only within every individual time interval, i.e., $\hat{P}^1, \ldots, \hat{P}^4$. We found that the Gini values of the scenario s_3 were substantially larger than Gini values in the scenario s_1. For instance, in the last month, the simulated choices in s_1 had a Gini index of 0.57, while in scenario s_3 it was 0.67. This basically means that the recommendations of α_3, in each month, causes a larger concentration of choices in comparison to α_1. To understand the reason of this apparent contradiction, we computed the average popularity of the chosen items within the 4 months of simulation. Item popularity of an item, in a month, is counting the number of times the item was chosen in the previous months. We found that in scenario s_3, every month, the simulated user choices are more concentrated on a narrow range of items, however, these items are indeed less popular, compared to those made in scenario s_1, and therefore, overall the (aggregated) Gini indices in the two scenarios can still become similar. This shows that the same level of diversity in the user choices can be achieved by distributing more uniformly the choices on a wider range of items or on a narrower range of less popular items. This is an unexpected outcome and it shows the highly complex nature of the human choices and the dynamics originated from recommendations based on different RSs.

We have also analysed the impact of the RSs on the awareness set of users. In all scenarios, the presence of RSs has substantially increased the average size of this set. Awareness increase is a natural consequence of recommending items which is observed in our simulation.

4 CONCLUSION AND FUTURE WORK

In this short paper we have presented a generic model for simulating the collective choice behaviour of a population of users that are influenced by a recommender system. We have compared the effect of alternative recommender systems on choice diversity measured by Gini index.

Interestingly, we discovered that the Gini index may vary under the influence of alternative recommenders but we have also discovered that a more detailed analysis of the distribution and the types of choices is in order. In a future work, we will analyse the dynamic evolution of other quality metrics related to choice concentration and quality, as was done in [12], and we will run our analysis on more data sets.

Our ultimate goal is to be able to predict with good precision aggregated measures of choice diversity in operational recommender systems and therefore help the recommender's owner to choose the recommendation technology that better fits his business goals.

REFERENCES

[1] Panagiotis Adamopoulos, Alexander Tuzhilin, and Peter Mountanos. 2015. Measuring the concentration reinforcement bias of recommender systems. *rN (i)* 1 (2015), 2.

[2] Pablo Castells, Neil J Hurley, and Saul Vargas. 2015. Novelty and diversity in recommender systems. In *Recommender Systems Handbook*. Springer, 881–918.

[3] Robert Dorfman. 1979. A formula for the Gini coefficient. *The review of economics and statistics* (1979), 146–149.

[4] Daniel Fleder and Kartik Hosanagar. 2009. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management science* 55, 5 (2009), 697–712.

[5] Daniel M Fleder and Kartik Hosanagar. 2007. Recommender systems and their impact on sales diversity. In *Proceedings of the 8th ACM conference on Electronic commerce*. ACM, 192–199.

[6] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE International Conference on Data Mining*. Ieee, 263–272.

[7] Dietmar Jannach, Lukas Lerche, Fatih Gedikli, and Geoffray Bonnin. 2013. What recommenders recommend–an analysis of accuracy, popularity, and sales diversity effects. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 25–37.

[8] Matevž Kunaver and Tomaž Požrl. 2017. Diversity in recommender systems–A survey. *Knowledge-Based Systems* 123 (2017), 154–162.

[9] Steffen Rendle. 2012. Factorization machines with libfm. *ACM Transactions on Intelligent Systems and Technology (TIST)* 3, 3 (2012), 57.

[10] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender systems: introduction and challenges. In *Recommender systems handbook*. Springer, 1–34.

[11] Amartya Sen. 1976. Poverty: an ordinal approach to measurement. *Econometrica: Journal of the Econometric Society* (1976), 219–231.

[12] Zoltán Szlávik, Wojtek Kowalczyk, and Martijn Schut. 2011. Diversity measurement of recommender systems under different user choice models. In *Fifth International AAAI Conference on Weblogs and Social Media*.

[13] Saúl Vargas, Linas Baltrunas, Alexandros Karatzoglou, and Pablo Castells. 2014. Coverage, redundancy and size-awareness in genre diversity for recommender systems. In *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, 209–216.

[14] Ting Wang and Dashun Wang. 2014. Why Amazon's ratings might mislead you: The story of herding effects. *Big data* 2, 4 (2014), 196–204.

[15] Ting Wang, Dashun Wang, and Fei Wang. 2014. Quantifying herding effects in crowd wisdom. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1087–1096.