A Uncertainty Perspective on Qualitative Preference

Seung-won Hwang and Mu-Woong Lee

POSTECH, Pohang, Korea
{swhwang,sigliel}@postech.ac.kr

Collaborative filtering has been successfully applied for predicting a person's preference on an item, by aggregating community preference on the item. Typically, collaborative filtering systems are based on based on *quantitative* preference modeling, which requires users to express their preferences in absolute numerical ratings. However, quantitative user ratings are known to be biased and inconsistent and also significantly more burdensome to the user than the alternative *qualitative* preference modeling, requiring only to specify relative preferences between the item pair. More specifically, we identify three main components of collaborative filtering– preference representation, aggregation, and similarity computation, and view each component from a qualitative perspective. From this perspective, we build a framework, which collects only qualitative feedbacks from users. Our rating-oblivious framework was empirically validated to have comparable prediction accuracies to an (impractical) upper bound accuracy obtained by collaborative filtering system using ratings.

1 Introduction

Collaborative filtering (CF) [1] has been widely adopted for inferring user preference, by aggregating the community preference. To motivate, we illustrate how such inference typically works in Example 1.

Example 1 (Movie recommendation). Consider a customer renting DVDs from an online rental shop, *e.g.*, netflix.com where recommended items are reported to be responsible for more than 60% of the rental decisions [2]. For accurate recommendation, a CF system infers user rating on the unrated DVD, based on the ratings of other people with similar histories. For instance, some user rated 5 for all the previous Batman movies is likely to rent "Dark knight", if other users who also liked all the previous Batman movies rate highly on "Dark knight".

More specifically, though details may vary, CF systems consist of the following three components:

– **Preference representation:** User preference on an item, *e.g.*, a movie, is typically represented as a numerical rating, *e.g.*, five stars. A user is thus represented as a *rating vector*.
– **Preference similarity:** For inferring a user rating for an item, other users sharing similar rating histories, should be identified. Toward this goal, similarity metrics need to be defined, *e.g.*, correlation between rating vectors.

Dagstuhl Seminar Proceedings 08421
Uncertainty Management in Information Systems
http://drops.dagstuhl.de/opus/volltexte/2009/1932

- **Preference aggregation:** Given a user community of similar preferences identified, CF systems then aggregate their preference, *e.g.*, by averaging community ratings, to compute a single representative community preference.

Prior work focuses on using a quantitative user preference model, *e.g.*, rating vector, for the above components. However, such perspective, by requiring users to provide absolute numerical ratings to represent their preferences, is known to be efficient for computation, but not very user-friendly.

Alternatively, *qualitative model* [3] has been explored in database systems, as a more intuitive alternative to present user preferences or queries in a form of "I liked movie A better than B", rather than giving an absolute numerical rating or score. However, qualitative model also has its drawbacks– While preference similarity and aggregation computation has been studied for qualitative model as well, state-of-the-art algorithms are significantly less efficient, compared to their quantitative counterparts, which partly explains why qualitative model has not been actively explored for CF system. While recently [4] adopted qualitative model for collaborative filtering, this framework builds upon numerical ratings, which defeats our motivation.

This paper aims at combining the strength of the two models, to enable a CF system that is both computation- and user-friendly. In particular, we achieve the goal, by collecting user feedbacks in qualitative forms and computing the similarity and aggregation in quantitative forms. Our extensive evaluation results validate that our proposed system identifies the recommended items with comparable quality to the results from quantitative CF systems, while requiring significantly less cognitive overheads to users.

We summarize our contributions as follows:

- Our proposed framework recommends items based on *qualitative* user feedbacks and thus do not require users to provide absolute numerical ratings.
- For computing similarity and aggregation, we use equivalent quantitative representations for efficiency.
- We validate that our proposed framework has comparable accuracies to a widely adopted implementation of quantitative collaborative filtering.

2 Related Work

This section overviews the related prior research efforts. We first survey existing work on collaborative filtering. We then discuss recent efforts on modeling preference qualitatively and processing technologies that can be used as component technologies for collaborative filtering.

Collaborative filtering: Collaborative filtering approach has been successfully adopted to most real-life recommender systems which can be categorized into the following two categories:

- **Heuristics-based:** In this line of work, each user is represented by her rating vector on data items. For finding similar users, various similarity metrics have been adopted, including correlation coefficient [5], cosine similarity, and Jaccard coefficient [6]. To predict the unknown rating, the ratings from similar users identified will then be aggregated, using various aggregation functions, including weighted average [7–10].

– **Model-based:** Alternatively, each user can be represented as a pre-trained classifier, such as Maximum entropy [11], SVM [12], or linear regression [13] classifiers. A key difference between heuristics- and model-based approaches is that the former refers to other users' ratings at runtime for rating prediction, while the latter refers only to the model, trained in advance for the given user.

Qualitative preference modeling/processing: Qualitative preference modeling has gained attention lately, as it is more intuitive to user to formulate or elicit their preferences qualitatively.

Qualitative preferences are typically represented as *partial orders*, which generalizes quantitative models mapping each object into numerical scores which generates *total orders*.

For similarity metrics, the most widely adopted metrics are Spearman [14] and Kendall-tau [15] distance. Unlike quantitative similarity metrics, measuring the score difference, these two metrics do not consider scores at all and only consider the rank difference of each object in the two orders– The Spearman footrule distance is the sum of the absolute rank difference in the two orders, while the Kendall tau distance counts the number of pairwise disagreements between two lists. However, these metrics, though not using numerical scores, assume preferences can be represented as *total orders* of data objects and cannot apply to *partial orders*, as we will discuss further in Section 3.

For aggregating similar preferences, many aggregation algorithms have been studied for generating the optimal combination, minimizing the distance with all the input preferences, based on Spearman and Kendall-tau distances discussed above. Such optimal aggregation, however, is studied to be NP-hard [16]. While an approximation with quality guarantee was studied [17] and later extended to aggregate bucket orders in [18], this extension cannot support arbitrary partial orders.

As overviewed above, the computation of similarity and aggregation for qualitative preferences are yet to be fully studied and current findings suggest that such computation is computationally intractable even in limited problem settings. This paper thus explores using the equivalent quantitative representations instead, where the similarity and aggregation computation is well-studied and efficient. The closest work to our paper is [4], which similarly explores to adopt qualitative modeling for recommendation, though this work still requires users to rate items numerically unlike ours.

3 Preliminaries

In this section, we first overview how a typical quantitative CF system implements the three main components for recommendation– preference representation, similarity, and aggregation. (Section 3.1). We then discuss how qualitative modeling generalizes the task (Section 3.2).

3.1 Quantitative CF

To illustrate a typical CF system implementation, this section describes how Open source toolkits CoFE[1] implements three major CF tasks.

Preference Representation CoFE assumes we can obtain absolute numerical ratings on data items, *i.e.*, quantitative modeling. Each user is thus represented as a d-dimensional numerical vector of ratings on d data items, where some items are yet to be rated, denoted by $r_{u,o}$ for the rating of user u on object o.

Preference Similarity For the missing rating, CoFE predicts the numerical rating, based on the ratings from the set N of "neighboring" users similar tastes. To identify such set, CoFE quantifies the similarity between two users v and w, using the Pearson correlation coefficient metrics below:

$$\omega_{v,w} = \frac{\sum_{i=1}^{d} (r_{v,i} - \overline{r_v}) \times (r_{w,i} - \overline{r_w})}{\sigma_v \times \sigma_w} \tag{1}$$

Note that $\overline{r_v}$ and σ_v represent the average and standard deviation of the ratings by user v respectively.

Preference Aggregation The rating vectors of the users in N are then aggregated to predict the missing ratings. For such aggregation, CoFE uses a weighted average, weighting the ratings from the users with more similar tastes. (We discard users with negative correlations.):

$$r_{u,o} = \overline{r_u} + \frac{\sum_{i \in N} (r_{i,o} - \overline{r_i}) \times \omega_{u,i}}{\sum_{i=1}^{n} \omega_{u,i}} \tag{2}$$

3.2 Qualitative CF

We now move on to discuss how qualitative modeling can generalize the above problem, and how such generalization complicates the problem.

Preference Representation As overviewed in Section 2, recent research efforts point out that it is non-trivial for users to represent their preferences adequately with absolute numerical scores. In contrast, qualitative representation enables to state preference in more intuitive form of comparing whether user "likes A better than B". For effectively collecting qualitative user preferences to

[1] http://eecs.oregonstate.edu/iis/CoFE/

identify highly relevant results with minimal user intervention, many *elicitation schemes* [19–25] were studied.

We represent the qualitative preferences collected as a $m-by-m$ pair order matrix C, where each entry C_{ij} indicates whether user prefers item i over item j. Specifically, we explore two representations for the qualitative preferences of user u.

- **Boolean perspective:** The most widely adopted representation is using Boolean value for C^u_{ij} to indicate whether user prefers i over j is true (represented by value 1) or false (represented by 0).

$$C^u_{ij} = \begin{array}{l} 1, \text{ if i is preferred over j} \\ 0, \text{ if j is preferred} \\ -, \text{ if unrated} \end{array} \quad (3)$$

- **Uncertainty matrix:** Alternatively, we can extend the representation to use C^u_{ij} to indicate a fuzzy value, indicating the *degree* of how certain it is that user prefers i over j. The higher value indicates higher certainty.
  More specifically, for each pairwise user elicitation, *e.g.*, user prefers i over j is true, we take it as an *evidence* toward C^u_{ij} and thus increment its value by 1. In addition, for every pairwise elicitation on (i, j), we increment not only C^u_{ij} but also all the other entries affected by transitivity combined with the past elicitation– For instance, if user already expressed that she prefers j over k before, user elicitation of preferring i over j is an evidence for both C^u_{ij} and C^u_{jk} (by transitivity), both of which we increment by 1.

A quantitative modeling is a special case of this representation where rated items generate a total. In a clear contrast, our model is more general to support any arbitrary partial orders.

Similarity Similarity metrics between two *total orders*, such as Kendall-tau and Spearman, are well studied. However, these metrics, relying on the rank of each object in the entire data set, cannot be defined in arbitrary partial orders. To address this problem, recent efforts extended two partial orders into two sets of all possible total orders, then compared the expected distance between the two sets in [26]. However, such enumeration is proved to be expensive, *i.e.*, *#P-hard*.

In this work, we keep qualitative user preferences in quantitative forms, as a pairwise comparison matrix C (Eq. 3), and use quantitative similarity metrics (Eq. 1), to identify neighboring users. While this naive representation of pairwise comparison matrix extends the problem size quadratically, our empirical results validate this qualitative modeling achieves comparable accuracies to CF system (serving as accuracy upper bound) without exploiting numerical user rating as CF does. We leave a more compact representation for the given qualitative preferences as future work.

Aggregation We now discuss how we aggregate pairwise comparison matrix C_n of all neighboring users $n \in N$, to predict C_u matrix for the given user u.

As overviewed in Section 2, finding the optimal aggregation of qualitative total orders has been proven to be NP-hard [16]. Though approximation for bucket orders has been studied [27], it is not applicable for aggregating arbitrary qualitative partial orders.

Similarly to the similarity computation, in this work, we explore aggregating preference matrices using quantitative aggregation function (Eq. 2) and leave more effective schemes as future work.

4 Experiments

This section reports our experiment setting then our evaluation results comparing quantitative and qualitative CF systems.

For dataset, we used MovieLens rating dataset, collected by the GroupLens Research Project [7]. This set consists of user ratings on movies, rated in five scales $\{1, 2, 3, 4, 5\}$.

For accuracy metrics, we use *kendall-tau* rank metric, which compares how pairwise comparisons in the predicted user preference are *concordant* with those in the "ground-truth ordering". More specifically,

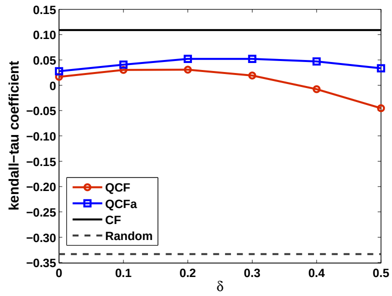
$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \tag{4}$$

where n_c is the number of concordant pairs, and n_d is the number of discordant pairs in the data set.

Among the 100,000 ratings by 943 users on 1682 items in MovieLens rating dataset, we selected heavily rated 100 items with the largest number of ratings. For each prediction, we aggregated the ratings of 100 neighboring users. From this dataset, we set aside the ratings from randomly chosen 5% of the users as ground-truth and compute τ compared to the predicted preferences for such users. We then compare the accuracy τ of our proposed frameworks with that of quantitative CF (using CoFE engine). Note, we compute τ for each user and report the average of all users.

More specifically, we implement recommendation engines QCF and QCFa using Boolean and uncertain representations respectively. As the MovieLens dataset are rated in five discrete scales, we discretize pairwise comparisons, to consider two entries with value difference less than δ as ties.

We stress that from the numerical ratings, our framework only extracts pairwise comparisons. Our goal is thus, unlike class CF where users have to specify the strength of preference in absolute numerical values, to predict such strength from pairwise comparisons, to predict closely to CF, without referring to ratings as CF does. We thus use the accuracy of CF to serve as an (impractical) upper bound while using that of random prediction as a lower bound. Our empirical results in Figure 4 indicate that, our qualitative CF engines, both QCF and QCFa, are comparable to upper bound.



5 Conclusion

This paper presents the problem of collaborative filtering from a qualitative perspective. In particular, we study how to adopt a qualitative model on the preference representation, aggregation, and similarity computation. Our extensive evaluation results report the prediction accuracy of our qualitative implementation, compared against an open-source quantitative collaborative filtering implementation.

6 Acknowledgement

This work was supported by the Engineering Research Center of Excellence Program of Korea Ministry of Education, Science and Technology (MEST) / Korea Science and Engineering Foundation (KOSEF), grant number R11-2008-007-03003-0 and Korean Research Foundation Grant (MOEHRD, Basic Promotion Fund; KRF-2007- 331-D00377).

References

1. David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Commun. ACM*, 35(12):61–70, 1992.
2. Netflix consumer press kit. In *http://www.netflix.com/MediaCenter/*.
3. W. Kiessling. Foundations of preferences in database systems. In *VLDB*, 2002.
4. N. Liu et. al. Eigenrank: A ranking-oriented approach to collaborative filtering. In *SIGIR*, 2008.
5. Robert S. Pindyck and Daniel L. Rubinfeld. *Econometric Models and Economic Forecasts*. MacGraw-Hill, New York, fourth edition, 1991.
6. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, first edition, 2005.
7. GroupLens Research. [online] `http://www.grouplens.org/`.
8. John S. Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI-98)*, 1998.

9. Jonathan L. Herlocker, Joseph A. Konstan, Al Borchers, and John Riedl. An algorithmic framework for performing collaborative filtering. In *SIGIR '99: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 230–237, New York, NY, USA, 1999. ACM.
10. Bharath Kumar Mohan, Benjamin J. Keller, and Naren Ramakrishnan. Scouts, promoters, and connectors: the roles of ratings in nearest neighbor collaborative filtering. In *EC '06: Proceedings of the 7th ACM conference on Electronic commerce*, pages 250–259, New York, NY, USA, 2006. ACM.
11. John Browning and David J. Miller. A maximum entropy approach for collaborative filtering. *J. VLSI Signal Process. Syst.*, 37(2-3):199–209, 2004.
12. Zhonghang Xia, Yulin Dong, and Guangming Xing. Support vector machines for collaborative filtering. In *ACM-SE 44: Proceedings of the 44th annual Southeast regional conference*, pages 169–174, New York, NY, USA, 2006. ACM.
13. Slobodan Vucetic and Zoran Obradovic. Collaborative filtering using a regression-based approach. *Knowl. Inf. Syst.*, 7(1):1–22, 2005.
14. Jerome L. Myers and Arnold D. Well. *Research Design and Statistical Analysis*. Lawrence Erlbaum Assoc., Inc., second edition, 2002.
15. Neil J. Salkind, editor. *Encyclopedia of Measurement and Statistics*. Thousand Oaks, Calif. : SAGE, 2007.
16. Cynthia Dwork, Ravi Kumar, Moni Naor, and D. Sivakumar. Rank aggregation methods for the web. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 613–622, New York, NY, USA, 2001. ACM.
17. Nir Ailon, Moses Charikar, and Alantha Newman. Aggregating inconsistent information: ranking and clustering. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 684–693, New York, NY, USA, 2005. ACM.
18. Nir Ailon.
19. Navneet Bhushan and Kanwal Rai. *Strategic Decision Making: Applying the Analytic Hierarchy Process*. Springer-Verlag, London, UK, 2004.
20. Craig Boutilier. Toward a logic for qualitative decision theory. In *In Proceedings of the KR'94*, pages 75–86. Morgan Kaufmann, 1994.
21. Craig Boutilier, Ronen I. Brafman, Carmel Domshlak, Holger H. Hoos, and David Poole. Cp-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research*, 21:2004, 2004.
22. Robin D. Burke, Kristian J. Hammond, and Benjamin C. Young. Knowledge-based navigation of complex information spaces. In *In Proceedings of the 13th National Conference on Artificial Intelligence*, pages 462–468. AAAI Press, 1996.
23. Robin D. Burke, Kristian J. Hammond, and Benjamin C. Young. The findme approach to assisted browsing. *IEEE Expert: Intelligent Systems and Their Applications*, 12(4):32–40, 1997.
24. Jon Doyle, Yoav Shoham, and Michael P. Wellman. A logic of relative desire (preliminary report). In *ISMIS '91: Proceedings of the 6th International Symposium on Methodologies for Intelligent Systems*, pages 16–31, London, UK, 1991. Springer-Verlag.
25. Thomas L. Saaty. *Fundamentals of Decision Making and Priority Theory*. RWS Publications, Pittsburgh, PA, USA, 2001.
26. Ronald Fagin, Ravi Kumar, Mohammad Mahdian, D. Sivakumar, and Erik Vee. Comparing partial rankings. *SIAM J. Discret. Math.*, 20(3):628–648, 2006.

27. Ronald Fagin, Ravi Kumar, Mohammad Mahdian, D. Sivakumar, and Erik Vee. Comparing and aggregating rankings with ties. In *PODS '04: Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 47–58, New York, NY, USA, 2004. ACM.