Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

CLiMF: Collaborative Less-Is-More Filtering

Yue Shi
Delft University of Technology
y.shi@tudelft.nl

Alexandros Karatzoglou
Telefonica Research
alexk@tid.es

Linas Baltrunas
Telefonica Research
linas@tid.es

Martha Larson
Delft University of Technology
m.a.larson@tudelft.nl

Nuria Oliver
Telefonica Research
nuriao@tid.es

Alan Hanjalic
Delft University of Technology
a.hanjalic@tudelft.nl

Abstract

In this paper we tackle the problem of recommendation in the scenarios with binary relevance data, when only a few (k) items are recommended to individual users. Past work on Collaborative Filtering (CF) has either not addressed the ranking problem for binary relevance datasets, or not specifically focused on improving top-k recommendations. To solve the problem we propose a new CF approach, *Collaborative Less-is-More Filtering (CLiMF)*. In *CLiMF* the model parameters are learned by directly maximizing the Mean Reciprocal Rank (MRR), which is a well-known information retrieval metric for capturing the performance of top-k recommendations. We achieve linear computational complexity by introducing a lower bound of the smoothed reciprocal rank metric. Experiments on two social network datasets show that *CLiMF* significantly outperforms a naive baseline and two state-of-the-art CF methods.

1 Introduction

Collaborative Filtering (CF) [Adomavicius and Tuzhilin, 2005] methods are at the core of most recommendation engines in online web-stores and social networks. The main underlying idea behind CF methods is that users that shared common interests in the past would still prefer similar products/items in the future [Resnick *et al.*, 1994]. While a lot of the CF literature has been devoted to recommendation scenarios where explicit user feedback is present (i.e., typically ratings), CF has also shown to be very valuable in scenarios with only implicit feedback data [Hu *et al.*, 2008], e.g., the counts of a user watching a TV show, the counts of a user listening to songs of an artist, etc. These counts can be interpreted as a measure of preference and thus a proxy to explicit feedback.

However, in some scenarios even the "count" information is not available, while only binary relevance data exists, e.g., the friendship between users in a Online Social Network, the follow relationship between users (or between a user and an event, etc.) in Twitter[1] or the dating history in online dating sites [Pizzato *et al.*, 2010], etc. In these scenarios, we use a "1" for a given user-item pair to denote that the user has an interaction (e.g., friendship, follow, "like") with the item, and "0" otherwise. Typically the observed interactions are regarded as positive signals (i.e., indicating relevant items), and although not all items without observed interactions are irrelevant it is safe to assume the vast majority of these items will be irrelevant for the user. In other words, for a given user, the signal "0" indicates an item set containing unobserved items that could be relevant but are most likely irrelevant.

Bayesian Personalized Ranking (BPR) [Rendle *et al.*, 2009] has been recently proposed as a state-of-the-art recommendation algorithm for situations with binary relevance data. The optimization criterion of BPR is essentially based on pair-wise comparisons between relevant and a sample of irrelevant items. This criterion leads to the optimization of the Area Under the Curve (AUC). However, the AUC measure does not reflect well the quality of the recommendation lists, since it is not a top-biased measure [Yue *et al.*, 2007], i.e., the position at which the pairwise comparisons are made is irrelevant to the contribution to the loss: mistakes at the lower ranked positions are penalized as equally as mistakes in higher ranked positions, which is not the desired behavior in a ranked list.

In view of these drawbacks, we propose a new CF approach, *Collaborative Less-is More Filtering (CLiMF)*, that is tailored to recommendation domains where only binary relevance data is available. *CLiMF* models the data by means of directly optimizing the Mean Reciprocal Rank (MRR) [Voorhees, 1999], a well-known evaluation metric in Information Retrieval (IR). Given the analogy between query-document search and user-item recommendation, we can define the Reciprocal Rank (RR) for a given recommendation list of a user, by measuring how early in the list (i.e. how highly ranked) is the first relevant recommended item. The MRR is the average of the RR across all the recommendation lists for individual users. MRR is a particularly important measure of recommendation quality for domains that usually provide users with only few but valuable recommendations (i.e., the *less-is-more* effect [Chen and Karger, 2006]), such

[1] http://twitter.com/

as friends recommendation in social networks where top-3 or top-5 performance is important.

Taking insights from the area of learning to rank and integrating latent factor models from CF, *CLiMF* directly optimizes a lower bound of the smoothed RR for learning the model parameters, i.e., latent factors of users and items, which are then used to generate item recommendations for individual users.

The contributions in this paper can be summarized as:

- We present a new CF approach, *CLiMF*, for MRR optimization for scenarios with binary relevance data. We demonstrate that *CLiMF* outperforms other state-of-the-art approaches with respect to making only a few but relevant recommendations.
- We introduce a lower bound of the smoothed RR measure, significantly reducing the computational complexity of RR optimization, and enabling *CLiMF* to scale for large datasets.

2 CLiMF

In this section, we present the *CLiMF*, Collaborative Less-is-More Filtering, algorithm. We first introduce a smoothed version of Reciprocal Rank by taking insights from the area of learning to rank. Then, we derive a lower bound of the smoothed reciprocal rank, and formulate an objective function for which standard optimization methods can be deployed. Finally, we discuss the properties of the *CLiMF* model.

2.1 Smoothing the Reciprocal Rank

The definition of reciprocal rank of a recommendation list for user i, as defined in information retrieval [Voorhees, 1999], can be expressed as:

$$RR_i = \sum_{j=1}^{N} \frac{Y_{ij}}{R_{ij}} \prod_{k=1}^{N} (1 - Y_{ik} \mathbb{I}(R_{ik} < R_{ij})) \tag{1}$$

in which N is the number of items, Y_{ij} denotes the binary relevance score of item j to user i, i.e., $Y_{ij} = 1$ if item j is relevant to user i, 0 otherwise. $\mathbb{I}(x)$ is an indicator function that is equal to 1, if x is true, otherwise 0. R_{ij} denotes the rank of item j in the recommendation list for user i. Note that the items are ranked in a descending order according to their predicted relevance scores for user i. Clearly, RR_i is dependent on the rankings of relevant items. The rankings of the relevant items change in a non-smooth way as a function of the predicted relevance scores and thus, RR_i is a non-smooth function over the model parameters. The non-smoothness of the RR measure makes it impossible to use standard optimization methods –such as gradient-based methods– to directly optimize RR_i. Inspired by recent developments in the area of learning to rank [Chapelle and Wu, 2010], we derive an approximation of $\mathbb{I}(R_{ik} < R_{ij})$ by using a logistic function:

$$\mathbb{I}(R_{ik} < R_{ij}) \approx g(f_{ik} - f_{ij}) \tag{2}$$

where $g(x) = 1/(1+e^{-x})$, f_{ij} denotes the predictor function that maps the parameters from user i and item j to a predicted relevance score. The predictor function that we use in our model is the basic and widely-used factor model, expressed as:

$$f_{ij} = \langle U_i, V_j \rangle \tag{3}$$

where U_i denotes a d-dimensional latent factor vector for user i, and V_j a d-dimensional latent factor vector for item j. Even though a sophisticated approximation for the item rank was proposed in [Chapelle and Wu, 2010], it has not been deployed in practice. Notice that in the case of RR_i in Eq. (1), only $1/R_{ij}$ is actually in use. We thus propose to directly approximate $1/R_{ij}$ by another logistic function:

$$\frac{1}{R_{ij}} \approx g(f_{ij}) \tag{4}$$

which makes the basic assumption that the lower the item rank, the higher the predicted relevance score, i.e., $1/R_{ij}$ would approach to 1. Substituting Eq. (2) and (4) into Eq. (1), we obtain a smooth version of RR_i:

$$RR_i \approx \sum_{j=1}^{N} Y_{ij} g(f_{ij}) \prod_{k=1}^{N} \big(1 - Y_{ik} g(f_{ik} - f_{ij})\big) \tag{5}$$

Notice that although Eq. (5) is a smooth function with respect to the predicted relevance scores and thus the model parameters U and V, optimizing this function could still be practically intractable, due to its multiplicative nature. For example, the complexity of the gradient of Eq. (5) with respect to V_j (i.e., only for one item) is $O(N^2)$: the computational cost grows quadratically with the number of items N and for most recommender systems N is typically large. In the following, we present a lower bound of an equivalent variant of Eq. (5), for which we derive a computationally tractable optimization procedure.

2.2 Lower Bound of Smooth Reciprocal Rank

Suppose that the number of relevant items for user i in the given data collection is n_i^+. Given the monotonicity of the logarithm function, the model parameters that maximize Eq. (5) are equivalent to the parameters that maximize $\ln(\frac{1}{n_i^+} RR_i)$. Specifically, we have:

$$\begin{aligned} U_i, V &= \arg\max_{U_i, V} \{RR_i\} = \arg\max_{U_i, V} \{\ln(\frac{1}{n_i^+} RR_i)\} \\ &= \arg\max_{U_i, V} \{\ln\Big(\sum_{j=1}^{N} \frac{Y_{ij}}{n_i^+} g(f_{ij}) \prod_{k=1}^{N} \big(1 - Y_{ik} g(f_{ik} - f_{ij})\big)\Big)\} \end{aligned} \tag{6}$$

Based on Jensen's inequality and the concavity of the logarithm function, we derive the lower bound of $\ln(\frac{1}{n_i^+} RR_i)$ as below:

$$L(U_i, V) = \sum_{j=1}^{N} Y_{ij} \big[\ln g(f_{ij}) + \sum_{k=1}^{N} \ln\big(1 - Y_{ik} g(f_{ik} - f_{ij})\big)\big] \tag{7}$$

We can take a close look at the two terms within the first summation. The maximization of the first term contributes to

learning latent factors that promote relevant items. However, given one relevant item, e.g., item j, maximizing the second term turns to learning latent factors of all the other items (e.g., item k) in order to degrade their relevance scores. In sum, the two effects come together to promote and scatter the relevant items at the same time. In other words, ***CLiMF*** will lead to a recommendation where some but not all relevant items are at the very top of the recommendation list for a user. We notice that this behavior of ***CLiMF*** corresponds to the analysis of MRR optimization for a search result list [Wang and Zhu, 2010], i.e., optimizing MRR results in diversifying ranked documents.

Taking into account the regularization terms that usually serve to control the complexity of the model (i.e. in order to avoid overfitting), and all the M users in the given data collection, we obtain the objective function of ***CLiMF***:

$$F(U,V) = \sum_{i=1}^{M}\sum_{j=1}^{N} Y_{ij}\big[\ln g(U_i^T V_j) + \sum_{k=1}^{N}\ln\big(1 - Y_{ik} g(U_i^T V_k - U_i^T V_j)\big)\big] - \frac{\lambda}{2}(\|U\|^2 + \|V\|^2) \quad (8)$$

in which λ denotes the regularization coefficient, and $\|U\|$ denotes the Frobenius norm of U. Note that the lower bound $F(U,V)$ is much less complex than the original objective function in Eq. (5), and standard optimization methods, e.g., gradient ascend, can be used to learn the optimal model parameters U and V.

2.3 Optimization

We use stochastic gradient ascend to maximize the objective function in Eq. (8). By taking the deriviatives of eq. (8) with respect to U_i and V_j.

Note that optimizing (8) involves a sum over the average number of relevant items $\tilde{n}$ for each user since usually we have $\tilde{n} << S$, the complexity is $O(dS)$ even in the case that $\tilde{n}$ is large, i.e., being linear to the number of non-zeros (i.e. relevant observations in the data). In sum, our analysis shows that ***CLiMF*** is suitable for large scale use cases.

3 Experimental Evaluation

In this section we present a series of experiments to evaluate ***CLiMF***. We first describe the datasets used in the experiments and the setup.

3.1 Experimental Setup

Datasets

We conduct experiments using two social network datasets from Epinions[2] and Tuenti[3]. The Epinions dataset is publicly available[4], and contains trust relationship between 49288 users. The Epinions dataset represents a directed social network, i.e., if user i is a trustee of user j, user j is not necessary a trustee of user i. For the purpose of our experiments, we exclude from the dataset the users who have less than 25 trustees. The second dataset collected from Tuenti, one of the largest social networks in Spain, represents an undirected social network, containing friendship between 50K users. Similar to the Epinions dataset, we also exclude the users with less than 25 friends. Note that in these two datasets, friends or trustees are regarded as "items" of users. The task is to generate friend or trustee recommendations to individual users. Statistics on the two datasets used in our experiments are summarized in Table 1.

Table 1: Statistics of the datasets

Dataset	Epinions	Tuenti
Num. non-zeros	346035	798158
Num. users	4718	11392
Num. friends/trustees	49288	50000
Sparseness	99.85%	99.86%
Avg. friends/trustees per user	73.34	70.06

Experimental Protocol and Evaluation Metrics

We separate each dataset into a training set and a test set under various conditions of user profiles. The condition of "Given 5" denotes that for each user we randomly selected 5 out of her trustees/friends to form the training set, and use the remaining trustees/friends to form the test set. We repeat the experiment 5 times for each of the different conditions of each dataset, and the performances reported are averaged across 5 runs.

The main evaluation metric utilized to measure the recommendation performance is MRR, the measure that is optimized in our model. In addition, we also measure the performance by precision at top-ranked items, using precision at top-5 (P@5), which reflects the ratio of the number of relevant items in the top-5 recommended items. In order to emphasize the value of "less-is-more" recommendations, we also use the measure of 1-call at top-ranked items [Chen and Karger, 2006]. Specifically, 1-call at top-5 recommendations (1-call@5) reflects the ratio of test users who have at least one relevant item in their top-5 recommendation lists.

Finally, as revealed in recent studies from different recommender domains, popular items could highly dominate the recommendation performance [Cremonesi *et al.*, 2010; Shi *et al.*, 2011; Steck, 2011]. We also notice this effect in our experiments, namely, recommending the most popular friends or trustees (i.e., those have the most friends or trusters) could already result in a high performance. For this reason we consider the top three most popular items as being irrelevant in order to reduce the influence from the most trivial recommendations [Cremonesi *et al.*, 2010; Shi *et al.*, 2011].

3.2 Performance Comparison

We compare the performance of ***CLiMF*** with three baselines, PopRec, iMF and BPR, which are described below:

[2]http://www.epinions.com

[3]http://www.tuenti.com

[4]http://www.trustlet.org/wiki/Downloaded_Epinions_dataset

Table 2: Performance comparison of *CLiMF* and baselines on the Epinions dataset

	Given 5			Given 10			Given 15			Given 20		
	MRR	P@5	1-call@5	MRR	P@5	1-call@5	MRR	P@5	1-call@5	MRR	P@5	1-call@5
PopRec	0.142	0.035	0.166	0.127	0.032	0.134	0.117	0.032	0.136	0.131	0.048	0.210
iMF	0.154	0.059	0.225	0.143	0.059	0.236	0.155	0.063	0.231	0.153	0.059	0.226
BPR-MF	0.241	0.148	0.532	0.167	0.072	0.334	0.177	0.098	0.380	0.216	0.096	0.422
CLiMF	**0.292**	**0.216**	**0.676**	**0.233**	**0.092**	**0.392**	**0.248**	**0.127**	**0.496**	**0.239**	**0.110**	**0.448**

Table 3: Performance comparison of *CLiMF* and baselines on the Tuenti dataset

	Given 5			Given 10			Given 15			Given 20		
	MRR	P@5	1-call@5	MRR	P@5	1-call@5	MRR	P@5	1-call@5	MRR	P@5	1-call@5
PopRec	0.096	0.029	0.138	0.074	0.017	0.080	0.074	0.019	0.088	0.074	0.019	0.086
iMF	0.064	0.020	0.090	0.065	0.017	0.076	0.065	0.021	0.098	0.076	0.023	0.108
BPR-MF	0.096	0.030	0.142	0.075	0.025	0.116	0.075	0.020	0.090	0.076	0.021	0.106
CLiMF	**0.100**	**0.039**	**0.190**	**0.077**	**0.027**	**0.124**	**0.077**	**0.022**	**0.104**	**0.083**	**0.024**	**0.116**

- **PopRec**. A naive baseline that recommends a user to be a friend or trustee in terms of her popularity, i.e., the number of friends or trusters of each user.
- **iMF**: A state-of-the-art matrix factorization technique for implicit feedback data by Hu et al. [Hu *et al.*, 2008].
- **BPR-MF** and implemented in MyMediaLite [Gantner *et al.*, 2011]. Bayesian personalized ranking (BPR) represents the state-of-the-art optimization framework of CF for binary relevance data [Rendle *et al.*, 2009].

The recommendation performances of *CLiMF* and the baseline approaches on the Epinions and the Tuenti datasets are shown in Table 2 and Table 3, respectively.

Three main observations can be drawn from the results: First, the proposed *CLiMF* model *significantly* outperforms the three baselines in terms of MRR across all the conditions and the two datasets. Note that in our experiments, the statistical significance is measured based on the results from individual test users, according to a Wilcoxon signed rank significance test with p¡0.01. This result corroborates that *CLiMF* achieves the goal that was designed for and optimizes the value of the reciprocal rank for the recommendations to the individual users. Second, *CLiMF* also achieves a *significant* improvement over the baselines in terms of P@5 and 1-call@5 across all the conditions and the two datasets. The improvement of P@5 indicates that by optimizing MRR, *CLiMF* also improve the quality of recommendations among the top-ranked items. In addition, the improvement of 1-call@5 supports that *CLiMF* particularly contributes to providing valuable recommendations at the top-k positions, i.e., raising the chance that users would receive at least one relevant recommendation among just a few top-ranked items. Compared to BPR, where AUC is optimized, *CLiMF* succeeds in enhancing the top-ranked performance by optimizing MRR, the top-biased metric. As can be also seen from the results, iMF performs worse than both BPR and *CLiMF* in all the conditions of the Epinions dataset and in most of the conditions of the Tuenti dataset. The reason might be that iMF is particularly designed for implicit feedback datasets with the "count" information, while it may not be suitable for the scenarios with only binary relevance data. Third, relatively large improvements are attained in terms of most of the metrics when users have a low number of known friends/trustees, i.e., the case of "Given 5". This result suggests that *CLiMF's* key mechanism of scattering relevant items could be particularly beneficial for recommendation scenarios under very high data sparseness.

4 Future work

Future work involves a few interesting directions. First, we would like to extend our *CLiMF* model to suit domains with explicit feedback data, e.g., ratings. Second, it is also interesting to experimentally investigate the impact of *CLiMF* on the recommendation diversity, by exploiting external information resources, such as the categories of items. Third, we are also interested in investigating recommendation models that optimize other evaluation measures, and in exploring the impact of optimizing different measures on various aspects of recommendation performance [Wang and Zhu, 2010].

5 Acknowledgements

This work is funded as part of a Marie Curie Intra European Fellowship for Career Development (IEF) award (CARS, PIEF-GA-2010-273739) held by Alexandros Karatzoglou.

References

[Adomavicius and Tuzhilin, 2005] Gediminas Adomavicius and Alexander 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*, 17(6):734–749, 2005.

[Chapelle and Wu, 2010] Olivier Chapelle and Mingrui Wu. Gradient descent optimization of smoothed information retrieval metrics. *Inf. Retr.*, 13:216–235, June 2010.

[Chen and Karger, 2006] Harr Chen and David R. Karger. Less is more: probabilistic models for retrieving fewer relevant documents. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '06, pages 429–436, New York, NY, USA, 2006. ACM.

[Cremonesi *et al.*, 2010] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 39–46, New York, NY, USA, 2010. ACM.

[Gantner *et al.*, 2011] Zeno Gantner, Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. Mymedialite: a free recommender system library. In *Proceedings of the fifth ACM conference on Recommender systems*, RecSys '11, pages 305–308, New York, NY, USA, 2011. ACM.

[Hu *et al.*, 2008] Yifan Hu, Yehuda Koren, and Chris Volinsky. Collaborative filtering for implicit feedback datasets. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pages 263–272, 2008.

[Pizzato *et al.*, 2010] Luiz Pizzato, Tomek Rej, Thomas Chung, Irena Koprinska, and Judy Kay. Recon: a reciprocal recommender for online dating. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 207–214, New York, NY, USA, 2010. ACM.

[Rendle *et al.*, 2009] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Schmidt-Thie Lars. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, pages 452–461, Arlington, Virginia, United States, 2009. AUAI Press.

[Resnick *et al.*, 1994] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, CSCW '94, pages 175–186, New York, NY, USA, 1994. ACM.

[Shi *et al.*, 2011] Yue Shi, Pavel Serdyukov, Alan Hanjalic, and Martha Larson. Personalized landmark recommendation based on geotags from photo sharing sites. In *Proceedings of the fifth international conference on weblogs and social media*, ICWSM '11, pages 622–625. AAAI, 2011.

[Steck, 2011] Harald Steck. Item popularity and recommendation accuracy. In *Proceedings of the fifth ACM conference on Recommender systems*, RecSys '11, pages 125–132, New York, NY, USA, 2011. ACM.

[Voorhees, 1999] Ellen M. Voorhees. The trec-8 question answering track report. In *TREC-8*, 1999.

[Wang and Zhu, 2010] Jun Wang and Jianhan Zhu. On statistical analysis and optimization of information retrieval effectiveness metrics. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 226–233, New York, NY, USA, 2010. ACM.

[Yue *et al.*, 2007] Yisong Yue, Thomas Finley, Filip Radlinski, and Thorsten Joachims. A support vector method for optimizing average precision. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '07, pages 271–278, New York, NY, USA, 2007. ACM.