Incorporating Distinct Opinions in Content Recommender System

Conference Paper *in* Lecture Notes in Computer Science · December 2015

DOI: 10.1007/978-3-319-28940-3_9

Grace E. Lee
Nanyang Technological University

Mun Y. Yi
Korea Advanced Institute of Science and Technology

Incorporating Distinct Opinions in Content Recommender System

Grace E. Lee, Keejun Han, and Mun Y. Yi(✉)

Department of Knowledge Service Engineering,
KAIST, Daejeon, Republic of Korea
{leklek0816,keejun.han,munyi}@kaist.ac.kr

Abstract. As the media content industry is growing continuously, the content market has become very competitive. Various strategies such as advertising and Word-of-Mouth (WOM) have been used to draw people's attention. It is hard for users to be completely free of others' influences and thus to some extent their opinions become affected and biased. In the field of recommender systems, prior research on biased opinions has attempted to reduce and isolate the effects of external influences in recommendations. In this paper, we present a new measure to detect opinions that are distinct from the mainstream. This distinctness enables us to reduce biases formed by the majority and thus, to potentially increase the performance of recommendation results. To ensure robustness, we develop four new hybrid methods that are various mixtures of existing collaborative filtering (CF) methods and our new measure of *Distinctness*. In this way, the proposed methods can reflect the majority of opinions while considering distinct user opinions. We evaluate the methods using a real-life rating dataset with 5-fold cross validation. The experimental results clearly show that the proposed models outperform existing CF methods.

Keywords: Distinctness · Bias · Content · Recommender system · Collaborative filtering

1 Introduction

With the advancement of technology, the media content industry has been growing continuously, and an enormous amount of content is now generated on a daily basis. Various strategies such as advertising and Word-of-Mouth (WOM) have been implemented to draw people's attention [1]. Indeed, it has become very common that content providers hire celebrities or well-known bloggers to promote their content and strive to shape mainstream opinions because most users simply follow the majority opinions [2]. In recent years, these promotion strategies have become more unnoticeable and have been used to maneuver more people into biased choices for their purchases without their awareness [3, 4].

In this paper, we attempt to resolve the aforementioned problems by proposing a simple yet novel measure, called *Distinctness*, to estimate how unique a certain rating of a user is from the major trend of ratings. In the field of recommender systems, there have been some prior attempts to exclude the effects of possible biases in systems by

G. Zuccon et al. (Eds.): AIRS 2015, LNCS 9460, pp. 109–120, 2015.
DOI: 10.1007/978-3-319-28940-3_9

detecting users' biased opinions [5, 6]. To the best of our knowledge, however, none of those studies have explored distinct opinions when biases exist, nor did they utilize such opinions for recommender systems.

For instance, The Matrix, a well-known blockbuster movie, has 133, 229 ratings from individual users in our dataset (obtained from MovieLens[1]), the distribution of ratings for the movie is (52, 32, 10, 3, 1) in percentages; the numbers correspond to a rating scale of (5, 4, 3, 2, 1). It should be noted that the rating 5, representing 'strongly like', accounts for more than half of the total ratings and that the majority of ratings, 94 %, are positive ratings (above 3 out of 5). Let us assume that there are user A and user B, each of which rated the movie 5, and user C and user D, each of which rated the movie 1. In accordance with the rating distribution for the movie, it seems that the identical opinions made by users C and D are more distinct than the opinions made by users A and B: they are strongly against the movie while most users liked it. In other words, the two people who gave the movie a rating of 1 express distinct opinions (rating the movie 'strongly dislike'), in contrast to the majority of users, who gave positive ratings. These unique ratings are powerful evidence to explain the characteristics of users. In spite of the potential usefulness of this type of distinctness feature, common CF approaches overlook the feature by assigning the same similarity value for the two user groups because the relations within groups are treated equally [7].

Unlike the aforementioned CF approaches, we measure the relations of users while considering the degree of their differences from the majority. By doing so, we are able to identify distinct opinions that can be potentially highlighted to generate more accurate recommendation results. In our experiment, it is shown that, compared with the existing CF approaches, an approach that exploits the *Distinctness* feature can improve the accuracy for recommender systems. In addition, we introduce three hybrid collaborative filtering methods combined with *Distinctness*; the results show unanimous increases compared to the baselines. The experiment was conducted on a real-life movie dataset, MovieLens, which is well-known for containing reliable data that can verify the performance of recommender systems [8–10]. By choosing the common dataset, we expect that our work can be easily reproducible.

The remainder of the paper is followed by related work on bias in recommender systems in Sect. 2. We then present the details of the proposed measure, *Distinctness*, and four variations of the CF methods utilizing *Distinctness* in Sect. 3. In Sect. 4, we describe experiments and results. Finally, the conclusion and future work are given in Sect. 5.

2 Related Work

In this section, we discuss some of the state-of-the-art recommendation models dealing with possible influences within interactions between users and recommender systems, and examine the limitations of these systems.

[1] http://movielens.org.

When users interact with a recommender system, it is possible that a variety of biases are involved in the interaction. User tendencies for rating and item selection are identified as biases in [11] and one work [12] pointed out user background and personal interest as potential sources of biases. In addition to the biases from individual users, context information influences opinions and responses. Regarding context information, many approaches have been studied. Context information such as weather, time, and companions does matter in recommender systems that have a multidimensional perspective since it affects the way users react [14] and thus its influences correspond to biases. The interactive methods that are used to detect context information and adapt to changes of information are able to reduce possible biases [15].

Designs of recommender systems also lead to biases in interactions with users. For example, different kinds of interfaces of recommender systems are able to intentionally instruct users in certain ways of expressing opinions and lead users' behaviors without their perceptions [9, 16]. Visually effective readability drives presentation and temporal biases in online reviews and comments [9]. Different types of rating scales also guide users toward certain ways of interacting with recommender systems and a 1-to-5 star scale allows users to express extreme feelings with great ease [16].

In this paper, we specifically look into influences coming from outside of users and recommender systems. In terms of popularity biases, several models are suggested for improvement of recommendations. A recent work [10] analyzed the causes and effects of popularity bias and proposed an algorithm to weaken the phenomenon in which only popular items are frequently recommended to users regardless of users' preferences. In [17], it was found that a function to penalize popular items in item-based collaborative filtering was able to decrease the chances of popular items appearing as recommendations.

Further, WOM both offline and online has an impact on users' overall acceptance, purchases, and opinions. By continuously appearing on a front page, early written online reviews lead to sequential biases and finally influence others' opinions [6]. Also, a greater volume of WOM results in a higher box office performance in the movie industry and directly connects to revenue in the field [2].

The aforementioned research has mainly attempted to identify types of biases and to exclude the effects of biases in the system. However, in this paper we explore implicit and unbiased opinions and propose an efficient measure to build sophisticated relationships among users. To the best of our knowledge, unbiased and distinct opinions have never been exploited in recommender systems, though use of such opinions can potentially be effective in bolstering relations among users.

3 Proposed Methods

In this section, we will first introduce the concepts of popularity and entropy. Based on those concepts, we first describe a new measure called *Distinctness* to estimate distinct similarities between users and then move on to explain four CF methods based upon *Distinctness*.

Popularity. In content industries including movies and books, opinions about content are highly susceptible to advertising and WOM [1, 2]. This situation eventually leads to the Matthew Effect: "the rich get richer and the poor get poorer" [10]. As a result, once items become widespread and establish certain reputations, their images are accumulative and fossilized [5]. Therefore, when a major trend of opinion for an item is formed by the public, *Distinctness* of unique opinions increases. The total number of ratings can indicate the *Popularity* of a certain movie. The advantage of *Popularity* is that it is straightforward and easy to compute, but the possibility of prefix bias – popular items having enough ratings usually defeat unpopular items – is a disadvantage [18]. However, in this study, to lessen the influence of possible weakness, we take advantage of a property of the logarithm function, which can transform an exponential-like curve to a linear-like curve by compressing large values.

Entropy. *Entropy* of an item's ratings indicates the distribution of ratings. For instance, when ratings are evenly distributed on a 1-to-5 star rating scale, the value of *Entropy* is greater. In contrast, when the majority of ratings are positive ratings, 4 or 5, the *Entropy* value is smaller. As the *Entropy* value decreases, the meanings of distinct ratings increase in terms of *Distinctness*. *Entropy*, however, does not imply the total number of ratings for items. Two rating distributions, (1, 0, 0, 0, 5) and (100, 0, 0, 0, 500), have the same *Entropy* values, although the former distribution has fewer ratings. In other words, *Entropy* alone never enables us to completely represent the concept of *Distinctness* of ratings due to the total number of ratings. *Entropy* of item p's rating distribution is calculated as follows:

$$Entropy(p) = \sum_i P(i|p) \log_2 P(i|p) \tag{1}$$

where $P(i|p)$ denotes the relative frequency of rating i in an item p. In a 1-to-5 star rating scale, i can be 1, 2, 3, 4, or 5.

Distinctness. Both *Popularity* and *Entropy* are correlated with *Distinctness*. We estimate *Distinctness*, D_{pi} for each rating i in an item p, using Bayes' theorem [13] and as follows:

$$D_{pi} = \log N_p \times \frac{1}{Entropy(p)} \times \left(1 - \frac{N_{pi}}{N_p}\right) \tag{2}$$

$$\text{where } N_p = \sum_i N_{pi}$$

where N_p is the total number of given ratings to item p and N_{pi} is the number of rating i in an item p.

$$D_p = \sum_i D_{pi} \tag{3}$$

The total *Distinctness* value for item p, denoted by D_p, is derived from the sum of D_{pi}. N_p is the total number of given ratings to item p and N_{pi} is the number of rating i in an item p. In order to apply Bayes' theorem in *Distinctness*, *Popularity* and *Entropy* have less correlation [8, 18]. Unlike the previous studies applying the concepts of *Popularity* and *Entropy* [8, 10, 18], this research exploits the idea of *Entropy* in reverse. Skewed distributions of ratings represented in smaller values of *Entropy* play a crucial role in *Distinctness*.

Distinctness-Based Collaborative Filtering (DISTINCT). All the ratings i of every item p can be represented by *Distinctness* values, D_{pi}. We use the *Distinctness* values to calculate *Distinctness*-based user-user similarity, S_D as follows:

$$S_D(u,v) = \sum\nolimits_{p \in \mathrm{P}} \min\{D_{pr_{u,p}},\, D_{pr_{v,p}}\} \tag{4}$$

where P is a set of items rated by both users, u and v, and p is each item included in P. $r_{u,p}$ and $r_{v,p}$ are user u's and v's ratings for p, respectively. Based on the similarity, S_D, the predicted rating of an unknown item i for the target user u is computed as follows:

$$\widehat{r_{u,i_D}} = \overline{r_u} + \frac{\sum\nolimits_{v \in K} S_D(u,v) * (r_{v,i} - \overline{r_v})}{\sum\nolimits_{v \in K} S_D(u,v)} \tag{5}$$

where $r_{v,i}$ is user v's rating for item i, $\overline{r_u}$ and $\overline{r_v}$ are user u's and v' average ratings, and K is a set of u's neighbors who rated the target item i, and at the same time satisfied with a parameter k from 0.1 to 1.0 which is detailed in Sect. 4.3.

Further, we present three hybrid methods combined with *Distinctness* and conventional rating-based collaborative filtering in recommender systems. Generally, in the conventional collaborative filtering, which directly uses ratings, user-user similarity using Pearson Correlation Coefficient is derived as follows:

$$S_R(u,v) = \frac{\sum\nolimits_{p \in P} (r'_{u,p} - \overline{r'_u})(r'_{v,p} - \overline{r'_v})}{\sqrt{\sum\nolimits_{p \in P} (r'_{u,p} - \overline{r'_u})^2} \sqrt{\sum\nolimits_{p \in P} (r'_{v,p} - \overline{r'_v})^2}} \tag{6}$$

The predicted ratings are computed in the same way as shown in Eq. 5 except S_R replaces S_D. In the first two hybrid algorithms, the *Distinctness*-based CF (DISTINCT) and the conventional CF are linearly combined together.

Linearly Combined Similarities CF (LCS). The two similarities from each CF method are joined together in the first hybrid CF as follows:

$$a \times S_D(u,v) + (1 - \alpha) \times S_R(u,v) \rightarrow S'(u,v) \tag{7}$$

where $S_D(u,v)$ and $S_R(u,v)$ are similarity values from Eqs. 4 and 6, and α is simply set as 0.5 to balance the two similarity scores. The new combined similarity value S' is used to compute the predicted rating, $\widehat{r_{u,i}}$ as follows:

$$\widehat{r_{u,i}} = \overline{r_u} + \frac{\sum_{v \in K} S'(u,v) * (r_{v,i} - \overline{r_v})}{\sum_{v \in K} S'(u,v)} \quad (8)$$

Linearly Combined Ratings CF (LCR). The second hybrid method linearly combines two predicted ratings coming from the different CF strategies based on Eqs. 4 and 6. The combination of two predicted ratings of an unknown item i for the target user u is then calculated as:

$$a \times \widehat{r_{u,i}}_D + (1 - \alpha) \times \widehat{r_{u,i}}_R \rightarrow \overline{r_{u,i'}} \quad (9)$$

where $r'_{u,i}$ is the final predicted rating as a result of the second hybrid CF.

Distinctness Weighted CF (DWCF). In the last hybrid CF method, the *Distinctness* values are applied to S_R as weighting parameters to control the contributions of rating-based relationships as follows:

$$\frac{\sum_{p \in P} \min\{D_{pr_{u,p}}, D_{pr_{v,p}}\}}{\sum_{p \in P} D_p} \times S_R(u,v) \rightarrow S'(u,v) \quad (10)$$

In DWCF, $S'(u,v)$ from Eq. 10 is used to compute the final predicted rating in Eq. 8.

4 Experiment

4.1 Experiment Setup

To evaluate our approach, we carry out experiments on the MovieLens dataset, which consists of more than 10 million ratings given by 70,250 active users for approximately 10,000 movies. This dataset is commonly used to evaluate recommendation tasks, and thus, we expect that our work can be easily reproducible. The ratings are on a 1-to-5 star scale. In order to focus on the users having common rating behaviors, we randomly chose 1,000 users who rated individual movies between 1,000 and 5,000 times. For users having given a myriad of ratings, the rating behaviors show stricter standards in rating [12], and in the case of users who have rated movies only a few rating times, the cold start problem occurs [8], so such types of users are excluded from this study. We then follow 5-fold cross validation by categorizing 80 % of the data as training data which generates recommendations and the rest of the data as test data to evaluate the recommendation results.

To compare the performance of the proposed methods, Pearson Correlation Coefficient CF (PEARSON) and Cosine Similarity CF (COSINE) are used as baseline models. All experiments are evaluated for two types of accuracies. Mean Absolute Error (MAE) and Root-Mean-Square Error (RMSE) are used to evaluate the prediction accuracy and Normalized Discounted Cumulative Gain (NDCG) is used to assess the ranking accuracy. Due to the limits of space, details of the measures are referred to [7, 19].

4.2 Experiment Results

In this section, we analyze the performances of our approaches, which adopt *Distinctness*: DISTINCT, LCS, LCR, and DWCF compared with the baselines. Figure 1(a) shows MAE values from the different methods; and Fig. 1(b) displays RMSE values. Note that lower values indicate better performances for those metrics and all the improvements are significant with a confidence level of 0.01. From the two graphs, the baselines show lower performances than the methods using the *Distinctness* feature. DISTINCT using the *Distinctness* feature alone outperforms the COSINE and PEARSON in 5.5 and 4.56 %, respectively. This is due to the fact that the method contains not only the *Distinctness* feature that detects the unique characteristics of the users, but also the popularity feature that reflects the major trend of ratings. Furthermore, we can see that adopting the hybrid approaches boosts the overall performance. DWCF especially presents the highest performance among our approaches, as it indeed outperforms the baselines in 9.32 and 9.28 %, respectively.

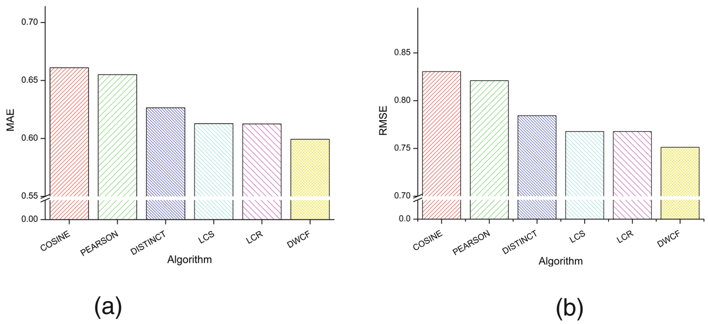

Fig. 1. Performance evaluation on (a) MAE and (b) RMSE

By looking at the NDCG scores, we also can observe which method consistently performs well in the higher ranking predicted accuracy. In detail, Fig. 2(a) represents NDCG score at N while N varies from 1 to 10; and it shows that distinctness methods including DWCF return higher NDCG score at every N compared to the baselines. Figure 2(b) shows that this tendency still resides while N increases from 10 to 40. A cursory look at the results is that utilizing the *Distinctness* feature reaps benefits in recommendations as the algorithms including the feature show better performances than the baselines. Especially, in both graphs, DWCF presents the highest performances in most cases of N compared to all the methods, implying that using the *Distinctness* feature as a weighting parameter mixed with the conventional CF methods promises the best recommendation results. Our proposed method DWCF outperforms baselines by approximately 1.48 % and 1.64 % on average for NDCG at 5 and 10, respectively. Note that the improvements are significant with a confidence level of 0.01.

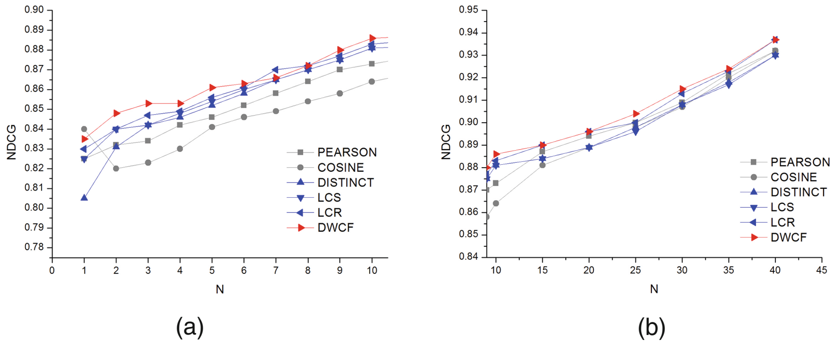

Fig. 2. Performance evaluation on NDCG at N varying from (a) 1 to 10 and (b) 10 to 40

Our assumption for *Distinctness* is that the relationship exhibited in a unique trend is stronger than the relationship in major trend. By focusing on the information on the minor rating on users, *Distinctness* enables us to successfully predict the minor ratings of users, although they have been hardly exploited for rating the movies. To verify the assumption and understand why the performance of the certain algorithm surpasses the other methods, we looked into how accurate the predicted ratings are in each rating scale. Let us denote HIT (is a counting measure) if an error value, the difference between a predicted rating and a target rating, lies in the range from zero to the MAE. HIT can be regarded as the number of correct predictions since the predicted rating values are quite close to the target answers within the setup error range; HIT Ratio is calculated that the number of the ratings is divided by the number of HITs. The higher the HIT Ratio is, the more precise prediction the method generates.

Table 1 shows a partial rating distribution for randomly chosen 40 users and HIT and HIT Ratio performances. In this random set, users frequently used 3 and 4 when rating the movies compared to 1, 2, and 5. The common ratings such as 3 and 4 have more opportunities to be predicted and obtain higher HITs. Nevertheless, if a method has higher HIT Ratio values for minor ratings like 1, 2, and 5, the method incorporates the information buried in ratings data including the distinct feature in efficient ways for recommender systems.

Table 1 demonstrates HIT and HIT Ratio of the methods for each rating scale as well as the rating distribution. In Table 1, a [a] indicate that the value is significantly higher within each method. From the table we can see that PEARSON has noticeable HIT accuracy when the target ratings to predict are 3 and 4 rather than 1 and 5. On the contrary, when the answer ratings are 1 and 5, the algorithm DISTINCT which uses

Table 1. Partial rating distribution and HIT and HIT Ratio performances

Rating scale		1	2	3	4	5	Total
# of ratings		68	244	639	745	285	1981
COSINE	HIT	1	62	466	329	50	908
	HIT Ratio	0.014	0.254	0.729	0.441	0.175	1.615
PEARSON	HIT	1	63	498	540	44	1146
	HIT Ratio	0.014	0.258	0.779[a]	0.724[a]	0.154	1.931
DISTINCT	HIT	48	62	480	540	64	1194
	HIT Ratio	0.706[a]	0.254	0.752	0.725	0.225[a]	2.665
LCS	HIT	23	62	498	555	64	1202
	HIT Ratio	0.338	0.254	0.779	0.745	0.224	2.341
LCR	HIT	22	66	498	555	73	1214
	HIT Ratio	0.324	0.270	0.779	0.745	0.256	2.374
DWCF	HIT	42	183	498	555	72	1350
	HIT Ratio	0.618[a]	0.750[a]	0.779[a]	0.745[a]	0.253[a]	3.145

only the *Distinctness* values has far higher HIT numbers than the baseline models (the HIT for 5 is reasonably better than PEARSON and COSINE).

DWCF is the combination of DISTINCT and one of the baseline models, PEARSON, since it exploits the *Distinctness* feature as weights on the conventional CF method. The performances of DWCF for 1 and 5 are as good as DISTINCT and for 2, 3, and 4 are more superb than PEARSON. This implies that DWCF mutually adopts both strengths from DISTINCT and PEARSON. A clear summarization of the above statements is given in Fig. 3 as it indicates how much each method generates HIT prediction for target ratings. In Fig. 3, the methods COSINE and PEARSON fail to predict the target rating 1 and 2, while DWCF yields the much better predictions in the target ratings. Again, as DWCF is the hybrid technique based upon PEARSON, it also shows a robust performance in predicting the major ratings such as 3 and 4 like PEARSON does.

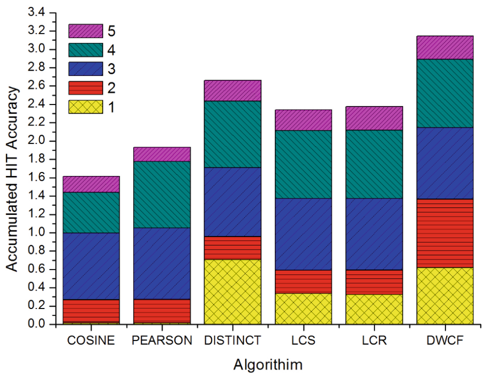

Fig. 3. Accumulated HIT accuracy graph

4.3 Impact of Parameter for Selecting Neighbors

Previously we introduced a parameter *k* in order to choose the size of similar neighbors for a certain user. It has been believed that the parameter plays a vital role to generate precise recommendations [11]. If the value of *k* is too small, some of the neighbors' ratings that match with the current user might be completely overlooked. On the other hand, if the value of *k* is too large, the ratings of users who have strongly the opposite tendency against the user might be potentially regarded for prediction. To evaluate the impact of the parameter, we observe the MAE, RMSE, NDCG at 5 and NDCG at 10 results by slowly increasing the parameter from 0.1 to 1.0. Note that the value of 1.0 indicates that all users are used as neighbors. Meanwhile, the value of 0.1 indicates that, after sorting all users in a descending order of each similarity method: COSINE, PEARSON, and DWCF, top 10 % of the users are chosen to be a set of neighbors with the target user.

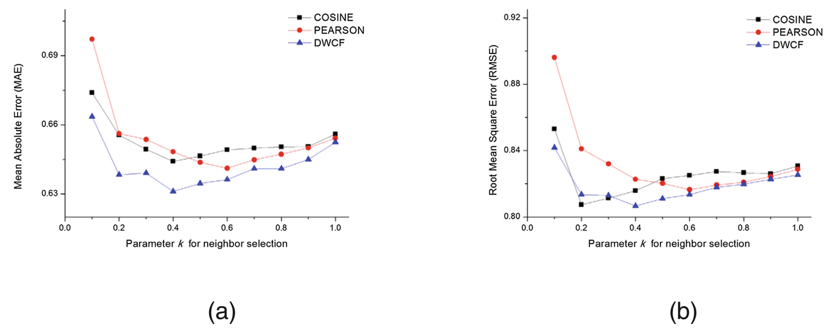

Fig. 4. Parameter *k* graph on (a) MAE and (b) RMSE

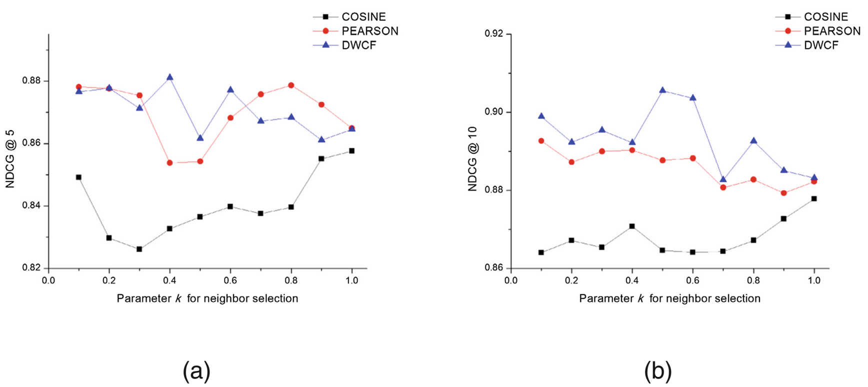

Fig. 5. Parameter *k* graph on NDCG at (a) 5 and (b) 10

Figures 4 and 5 present the effect of varying the parameter *k*. In specific, the MAE performance for each *k* is shown in Fig. 4(a), the RMSE performance is given in Fig. 4 (b), and the NDCG performance is given in Fig. 5. In terms of DWCF, we observe that the overall performance for each metric peaks approximately between 0.4 and 0.5, gradually decreasing afterwards. On the other hand, PEARSON and COSINE show lower performances than DWCF except for few cases. Although the performance varies depending on the value of k, we still can observe that the best value for DWCF presents the best performance compared to the best cases of the compared methods. Another observation is that each method peaks at different k and thus, we set the parameter to its best performing value in accordance with the method chosen for our experiment.

5 Conclusion and Future Work

The goal of this paper is to suggest a new collaborative filtering method for content recommender systems. To achieve this goal, we have presented a novel measure, *Distinctness*, to estimate unique and distinct ratings that do not follow major trends. Using *Distinctness*, we have developed four CF approaches: DISTINCT, LCS, LCR, and DWCF. Following the proposed approaches, similarity and predicted ratings have been computed while considering the degree of *Distinctness*. Throughout our experiment, we have showed that our models effectively utilize data and clearly outperform comparable models. Especially, by exploiting the concept of HIT and HIT Ratio, we have detailed an investigation of superior results from DWCF.

Our study needs further work. We will have to apply classification of users on the basis of analyzing rating patterns. We assume that there might be different types of users in terms of ways of reacting to external influences according to rating times and experiences. In our experiment, we have chosen only common users who have a certain range of rating times, to avoid cases of the cold-start problem and unusual rating behaviors. Additionally, in order to focus on discovering the distinctness information and validating this new feature for the first time, we have used easily computable algorithms as the baselines. Based on improvements from the distinctness feature proposed in this study, we believe that it is also worthwhile to incorporate sophisticated algorithms like Matrix Factorization approach into this feature in future. Despite the need for the further work, the proposed methods show the promising potential of *Distinctness* to improve the overall performance of recommendation results.

Acknowledgements. This work was supported by the Industrial Strategic Technology Development Program, 10052955, Experiential Knowledge Platform Development Research for the Acquisition and Utilization of Field Expert Knowledge, funded by the Ministry of Trade, Industry & Energy (MI, Korea).

References

1. Gopinath, S., Chintagunta, P.K., Venkataraman, S.: Blogs, advertising, and local-market movie box office performance. Manage. Sci. **59**(12), 2635–2654 (2013)
2. Duan, W., Gu, B., Whinston, A.B.: The dynamics of online word-of-mouth and product sales-an empirical investigation of the movie industry. J. Retail. **84**(2), 233–242 (2008)

3. Gal-Or, E., Geylani, T., Yildirim, T.P.: The impact of advertising on media bias. J. Mark. Res. **49**(1), 92–99 (2012)
4. Faber, R.J., O'Guinn, T.C.: Effect of media advertising and other sources on movie selection. Journalism Mass Commun. **61**(2), 371–377 (1984)
5. Liu, Y.: Word of mouth for movies: its dynamics and impact on box office revenue. J. Mark. **70**(3), 74–89 (2006)
6. Kapoor, G.: Sequential bias in online product reviews. J. Organ. Comput. Electron. Commer. **19**(2), 85–95 (2009)
7. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. **22**(1), 5–53 (2004)
8. Rashid, A.M., Albert, I., Cosley, D., Lam, S.K., McNee, S.M., Konstan, J.A., Riedl, J.: 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)
9. Sparling, E.I, Sen, S.: Rating: How difficult is it? In: Proceedings of the Fifth ACM Conference on Recommender Systems, pp. 149–156 (2011)
10. Zhao, X., Chen, W., Niu, Z.: Opinion-based collaborative filtering to solve popularity bias in recommender systems. In: Decker, H., Lhotská, L., Link, S., Basl, J., Tjoa, A.M. (eds.) DEXA 2013, Part II. LNCS, vol. 8056, pp. 426–433. Springer, Heidelberg (2013)
11. Koren, Y.: Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 426–434 (2008)
12. Tan, C., Chi, E.H., Huffaker, D., Kossinets, G., Smola, A.J.: Instant foodie: predicting expert ratings from grassroots. In: Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management, pp. 1127–1136 (2013)
13. Stoimenova, E.: Methodology in robust and nonparametric statistics. J. Appl. Stat. **40**(12), 2773–2777 (2013)
14. Oku, K., Nakajima, S., Miyazaki, J., Uemura, S.: Context-aware SVM for context-dependent information recommendation. In: Proceedings of the 7th International Conference on Mobile Data Management, p. 109 (2006)
15. Hariri, N., Mobasher, M., Burke, R.: Context adaptation in interactive recommender systems. In: Proceedings of the 8th ACM Conference on Recommender Systems, pp. 41–48 (2014)
16. Yeniterzi, R., Callan, J.: Analyzing bias in CQA-based expert finding test sets. In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 967–970 (2014)
17. Lai, S., Liu, Y., Gu, H., Xu, L., Liu, K., Xiang, S., Zhao, J., Diao, R., Xiang, L., Li, H., Wang, D.: Hybrid recommendation models for binary user preference prediction problem. In: JMLR: Workshop and Conference Proceedings 18, pp. 137–151 (2012)
18. Rashid, A.M., Karypis, G., Riedl, J.: Learning preferences of new users in recommender systems: an information theoretic approach. ACM SIGKDD Explor. Newsl. **10**(2), 90–100 (2008)
19. Yilmaz, E., Kanoulas, E., Aslam, J.A.: A simple and efficient sampling method for estimating AP and NDCG. In: Proceedings 31st SIGIR, pp. 603–610 (2008)