Merging implicit and explicit ratings model in recommendation systems

نموذج توحيد التقييم الصريح والضمني في أنظمة التوصية

Dr. Huda Naji Nawaf
Assistant prof. – Information Network
University of Babylon,
College of Information Technology,
Dept. of Information Network,
E-mail:- halmamory@itnet.uobabylon.edu.iq

Intedhar Shakir Nasir
Assistant Lecturer– Computer Science
University of Kerbala,
College of Medicine
Dept. of Community Medicine
E-mail:- intedhar@yahoo.com

Abstract

This study intends to improve the recommendation system by unifying both the implicit and explicit behavior of users. The implicit behavior indicates what users view over time regardless of the rating of what he views (implicit rating), whereas, the explicit behavior in this effort refers to what users rate for an item. The hamming distance is used to create a distance matrix that is converted to similarity matrix for users who are akin in terms of implicit rating. As for explicit ratings, the cosine similarity is used to create similarity matrix for users who are similar in terms of the scale used to rate an item. The proposed method is evaluated using three data sets; Movielens, Hetrec 2011, and Yahoo! Movies. The evaluation of the proposed method constrains with the measures of the related works. Thus, recall, precision, mean absolute error (MAE), and F-measures have been used. The experimental results show that the proposed system has a good performance, particularly Movielens dataset, when compared with the existing works. Our proposed method is free of any complex computations; at the same time it is competitive as comparable and better results are obtained considering other works.

Categories and Subject Descriptors: [Computer-Social Networks]
Key Words: Similarity,Cosine Similarity,Implicit Feedback,Explicit Feedback,Hamming Distance.

الخلاصة

يهدف البحث الى تحسين أنظمة التوصية عن طريق توحيد سلوك المستخدمين الصريح والضمني. يشير السلوك الضمني او المخفي الى ماذا يشاهد المستخدم من الـ Movie بمرور الزمن بغض النظر عن تقييم ما يشاهده , بينما السلوك الصريح يركز على تقييم ما يشاهده. يستخدم hamming distance لبناء مصفوفة المسافات ومن ثم تحول الى مصفوفة التشابهات للمستخدمين الذين يكونون متشابهين من ناحية التقييم الضمني . اما بالنسبة للتقييم الصريح فقد استخدم cosine similarityلخلق مصفوفة التشابه بين المستخدمين من ناحية التقييم الصريح لما يشاهدونه . تم تقييم النظام المقترح باستخدام ثلاث مجموعات من البيانات Movielens, Hetrec 2011 وYahoo! Movies .

تم تحديد مقاييس التقييم طبقا للمقاييس المستخدمة لتقييم الاعمال المتعلقة لأغراض المقارنه. لذا تم استخدام recall ,precision mean absolute error (MAE) وF-measures.

بينت النتائج التجريبية ان أداء النظام المقترح جيد على الأقل فيما يتعلق ببيانات Movielens حين تم مقارنتها مع الاعمال الحالية. يخلو النظام المقترح من أية تعقيدات حسابيه وبنفس الوقت فإن النتائج التي تم الحصول عليها مشابهة واحيانا افضل قياسا بالأعمال الموجودة.

1. INTRODUCTION

Recently, the recommendation system has been widely used by a variety of applications; for example, movies, music, news, books. In these applications, the users offer two types of feedback; implicit and explicit. As user feedback is examined, collaborative filtering is used as it is a common technique for building recommender systems. Explicit feedback can be represented as numeric ratings input to assign the degree of preference for an item. For example, Movielens uses a 1-5 star scale to rate the items, whereas , implicit feedback includes user interests though clicks or purchases.

Collaborative Filtering (CF) is a technique that is used to construct personalized recommendations on the Web, and it relies on past behavior of users. For the past behavior, it is either the history of viewing, purchasing or rating these items. Thus, recommenders infer user preferences depending on implicit and/or explicit feedback. This study will focus on both implicit and explicit feedback as the literature appears lacking in this regard.

[Nathan et al. 2010] unified both explicit and implicit feedback simultaneously in matrix factorization to develop collaborative filtering models, namely the co-rating and the co-ranking models. The co-ranking model provides a highly effective framework for unifying explicit and implicit feedback [1].

The authors in [Bell, et al. 2007] combined the two forms of feedback via a factored neighborhood model. Their contribution is derives interpolation using weights or ratings for all nearest neighbors simultaneously [2]. This is contrary to the other methods.

For the work in [3], the unique properties of implicit feedback datasets have been assigned. In fact, factor model has been designed for implicit feedback. In addition, the authors have proposed also scalable optimization procedure and applied their algorithm on television show.

2. THE MEASURE OF SIMILARITY

Similarity is measured between two sequences of documents, computer programs, chain letters or behavior of users to find out an evolutionary distance [4].

There are different types of similarity measures and there is no consensus on the “best” similarity measure [5]. The similarity measures for sequential data are the focus here.

The concept of similarity is an important measure in different applications. For instance, the geometric methods are used in studies of congruence and homothety for assessing in mathematics, and trigonometry [6].

2.1 Cosine Similarity

The cosine measure calculates the similarity between two users as the cosine of the angle between their corresponding behavior vectors [7]. Measuring the similarities of proteins pairs is a significant problem in molecular biology.

The cosine similarity, θ, for two vectors of successive data, A and B is: [6]

$$Similarity = \cos(\theta) = \frac{A.B}{\|A\|\,\|B\|} \ldots (1)$$

For comparison between two vectors that represent the behavior of users over time in an online community, such as an online movie recommendation community, Movielens, represents the movies’ ratings over time that have been rated by users.

$$A = \{R_{m1}, R_{m2},, \ldots., R_{mi}\};$$
$$B = \{R_{m1}, R_{m2},, \ldots., R_{mi}\}; i = 1,2, \ldots \ldots, n$$

Where, R_{mi}represents the rating of i-th movie which users viewing over time, and n represents the number of total movies.

2.2 Hamming Distance

The Hamming distance can be defined as the difference between two binary vectors that are equal in size [8]. Each binary sequence represents the views of movies of a user over time. Where 1_i and 0_i mean the ith movie has been viewed and not viewed by user respectively.

However, the similarity matrix can be obtained from hamming distance as follow [9]:

$$S = 1 - D_{hamming} \ldots (2)$$

3. THE PROPOSED MODEL

The proposed model of unifying both the implicit and explicit feedbacks is somewhat different from the others in literature reviews. A primary difference is that it depends on neighborhood domain. Explicit and implicit feedback are represented by the rating of items and what the items are rated by the users. The prediction model for each user is based on two neighborhoods, the first and second related to like-minded neighbors in terms of preferences of items, and in terms of how users rate these items respectively.

This leads to the following prediction model, a matrix $U : (N \times M)$ is computed by the scalar product of the two matrices D and V which represent similarity matrix obtained using cosine similarity and similarity matrix obtained from (1) respectively.

$$U = \prod_{n=1}^{N} \prod_{m=1}^{M} (D_{nm} \cdot V_{nm}) \dots (3)$$

Thus, the prediction model for each user has been modeled by product the n-th rows of both matrices D and V. Each row in modeled matrix U has been sorted descending to obtain k-nearest neighbors for a user. So, the rating prediction formula has been designed as a weighted average as follows:

$$\hat{r_{ui}} = \left(\sum_{k=1}^{K} r_{ki} \cdot w_k \right) \Big/ K \dots (4)$$

Under this model, the rating prediction of i-th item for u-th user has been obtained by weighted average of K nearest neighbors. The weighted average for k-th user w_k should meet the following condition:

$$w_1 > w_2 > \cdots > w_K \dots (5)$$

Where the weight of the first neighbor is greater than that of the second neighbor and so on.

4. EXPERIMENTAL RESULTS

In this section we experimentally evaluate the performance of the proposed system using three popular datasets whose characteristics are shown in Table 1.

Table (1) Descriptions of dataset

Dataset	No. of users	No. of movies
Movielens	943	1682
Hetrec 2011	2113	10197
Yahoo! Movies	7642	11915

The evaluation is a very important step in any recommendation system; otherwise it is not considered a successful system. Thus, the system here is evaluated using the items that have been rated by users and consider it as source condition.

Generally,precision,recall and related measures are the most popular metrics for evaluating information retrieval systems.The measures related to the mean absolute error have also been applied.

Precision and recall have been computed from the confusion matrix. Relevant represents only the items with rating scales of 4 or 5; otherwise they have been considered not relevant. The predicted items have been represented in the same manner, where the recommended items are those with rating scales of 4 or 5; otherwise they are not recommended.

Precision represents the likelihood that a prediction item is related, whereas, the recall is defined as the likelihood that an important item will be predicted.

The results are displayed and compared with other recent methods in the following tables. Each table shows the results of the proposed method for one dataset. For instance; Table 1, Table 2, and Table 3 display the results for Movielens, Hetrec 2011, and Yahoo! Movies datasets respectively. It is worth mentioning that the results of the other methods are displayed as published in papers and have not been replicated.

Maybe the other methods share with or differ from the proposed system in terms of the evaluation measures, so the dash cells represent the measures that are not computed for a method in related works as illustrated in Table 1. As explained, the results for the proposed system are calculated for different k, but the sizes of k depends on the size of the dataset and as have been assigned in previous works. Figures 1, 2, and 3 show the precision and recall over different k for the three datasets.

5. DISCUSSIONS

Generally, Table 1 shows substantial improvements of the proposed methods in terms of measures. Table 1 is the only table that displays the comparison with previous works, where the Movielens dataset is widely used in this area. Obviously, the accuracy of F-measures of the proposed method outperforms the others for Movielens dataset, except the first method in terms of MAE. Both Tables 3 and 4 display the results of Hetrec 2011 and Yahoo! Movies, respectively. In the literature, accuracy measures were not found as an evaluation method in rating prediction using both datasets, Hetrec2011 and Yahoo! Movies. The evaluation of the this proposed system using Movielens dataset shows that the likelihood of correctly recommended items from the relevant items is best from that of recommended items, and conversely for the both Hetrec2011 and Yahoo! Movies. Also of note, the mean absolute error increases as k increases, as shown in tables. Generally, the precision and recall increases and decreases as the size of neighbors increases respectively as shown in Figures 1, 2 and 3.

Table (2) Results of F-measures, precision, and recall using Movielens dataset

The Algorithms	Precision	Recall	F_1	F_2	$F_{0.5}$	Accuracy	MAE
Weighted based Slope One [10]	-	74.8%	74.5%	74.7%	74.4%	-	0.7
Weighted based Adjusted Cosine [10]	-	67.9%	67.4%	67.7%	67.2%	-	0.81
(WDE) [4]	71%	28.84%	41.02%	-	-	-	-
The proposed system							
(k=25)	68.5%	74.7%	71.4%	73.3%	69.6%	65.9%	0.78
K=40	68.9%	71.3%	70%	70.8%	69.3%	65%	0.8
(k=50)	68.9%	74%	71.3%	72.9%	69.8%	66%	0.81
k=100	69.8%	69.1%	69.4%	69.2%	69.6%	65%	0.83
K=150	70.6%	60.7%	65.2%	62.4%	68.3%	63%	0.91

Table (3) Results of F-measures, precision, and recall using Hetrec2011 dataset

Size of neighbors	Precision	Recall	F_1	F_2	$F_{0.5}$	Accuracy	MAE
K=40	74.5%	70%	72.1%	70.8%	73.5%	70%	0.75
K=50	74.5%	70.3%	72.3%	71.1%	73.6%	70%	0.75
K=100	75%	68.5%	71.3%	69.2%	73.4%	70%	0.76
K=150	75.2%	66.1%	70.3%	67.7%	73.1%	70%	0.77
K=200	75.4%	64.6%	69.5%	66.5%	72.9%	68.9%	0.79
K=500	76.3%	50.9%	61%	54.5%	69.3%	64%	0.96

Table (4) Results of F-measures, precision, and recall using Yahoo! Movies dataset

Size of neighbors	Precision	Recall	Accuracy	MAE
K=50	84.5%	75.2%	69%	1.1
K=100	84.3%	65.1%	63%	1.2
K=150	84.1%	59.6%	56%	1.31
K=200	84%	55%	57%	1.41
K=250	83.9%	51.9%	55%	1.48

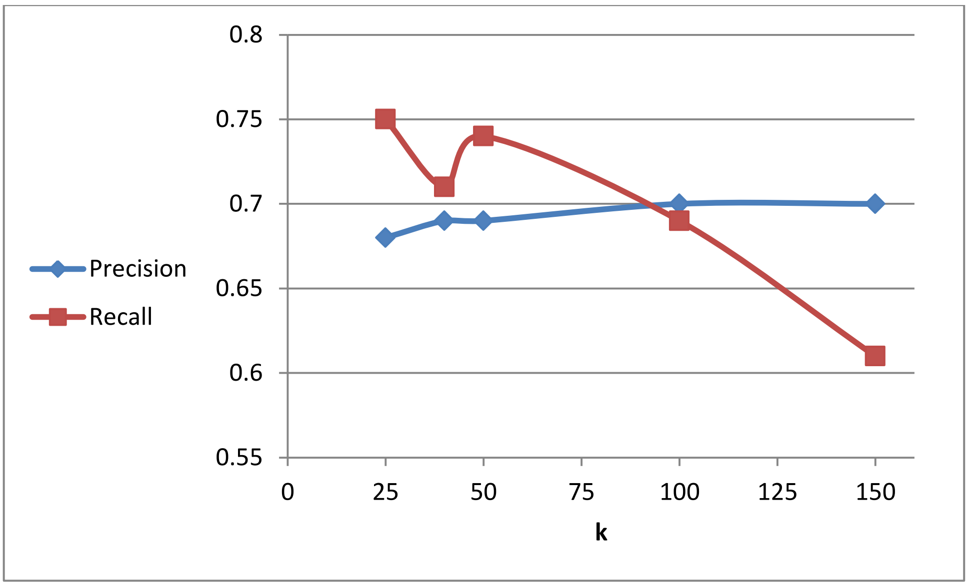

Figure (1) shows the Precision and recall curves over different k for Movielens dataset

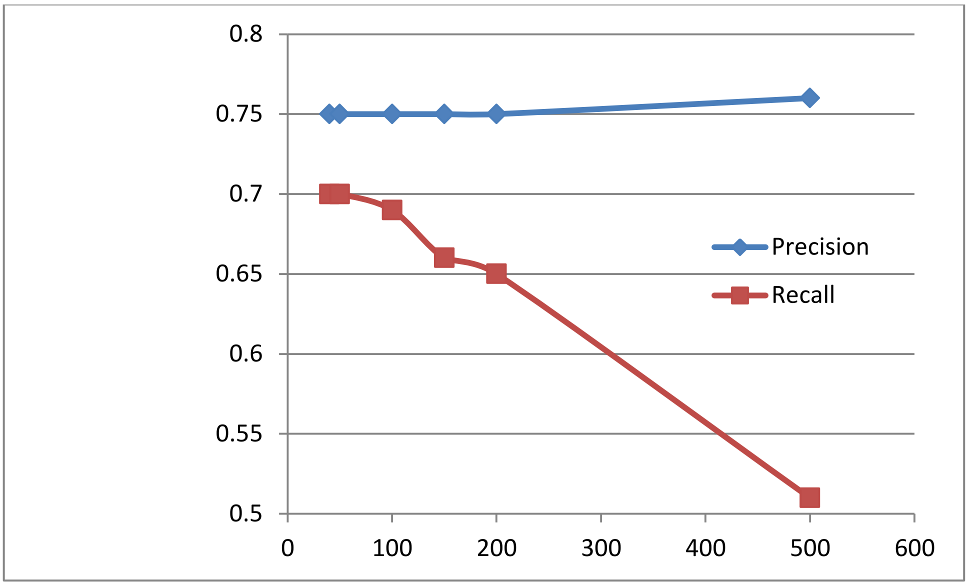

Figure (2) shows the Precision and recall curves over different k for Hetrec2011 dataset

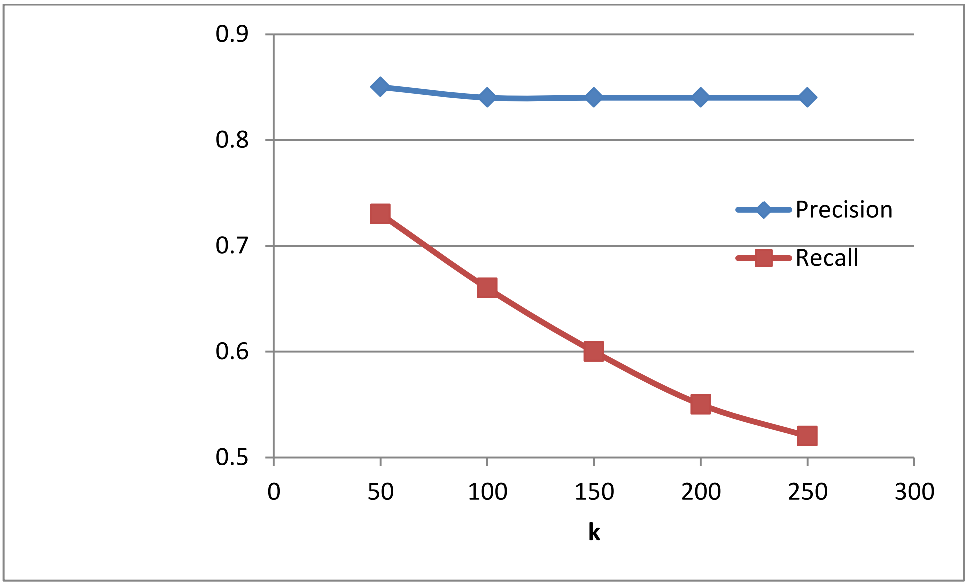

Figure (3) shows the Precision and recall curves over different k for Yahoo! Movies dataset

6. CONCLUSION

This study suggested item and rating- based recommendation system that unifies implicit and explicit user behavior. The system is characterized by its simplicity, but the performance of the proposed system is comparable or sometimes better than related works, particularly Movielens dataset. To sum up, in the proposed system, the probability of correctly recommended items from all related or relevant items (recall) is better than from that of recommended items (precision) for small dataset and, conversely, for larger dataset.

References

[1] Nathan Liu, Evan Xiang, Min Zhao and Qiang Yang. 2010. Unifying explicit and implicit feedback for collaborative filtering. Proceedings of the 19th ACM International Conference on Information and Knowledge Management. Toronto, ON, Canada, 1445-1448.

[2]Robert Bell and Yehuda Koren. 2007. Scalable Collaborative Filtering with Jointly Derived Neighborhood Interpolation Weights. Proceedings of the Seventh IEEE International Conference on Data Mining, IEEE Computer Society. 43-52.

[3] Yifan Hu, Yehuda Koren and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. ICDM '08 Proceedings of the 2008 Eighth IEEE International Conference on Data Mining. 263-272

[4] Ming Li, Xin Chen, Xin Li, Bin Ma, and Paul M.B. Vit́any. 2004. The Similarity Metric. IEEE Transactions on Information Theory. 50. (Aug. 2004), 1-13.

[5] Vikas Thada and Vivek Jaglan. 2013. Comparison of Jaccard, Dice, Cosine Similarity Coefficient to Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm. International Journal of Innovations in Engineering and Technology, 2, 1 (Aug. 2013), 202-205.

[6] K.P.N.V. Satya sree and J. V. R. Murthy. 2012. Clustering Based on Cosine Similarity Measure. International Journal of Engineering Science and Advanced Technology, 2, 3, (May-June), 508-512.

[7] Lan Huang, David Milne, Eibe Frank and Ian H. August 2012. Witten, Learning a Concept-based Document Similarity Measure. Journal of the American Society for Information Science and Technology, 63, 8, August 2012.

[8] SHIMODAIRA, H. 2014. Similarity and recommender systems. Learning and Data Note2.

[9] Zeshui Xu. 2014. Hesitant fuzzy set theory. Springer Publishing Company, Incorporated ©2014.

[10] Min Gao, Yunqing Fu , Yixiong Chen , Feng Jiang. 2012. User-Weight Model for Item-based Recommendation Systems. Journal of software , 7,9, September 2012. 2133-2140.