International Journal of Recent Trends in Engineering & Research
ISSN (ONLINE) : 2455 – 1457
IMPACT FACTOR : 4.101

Recommendation system Based On Cosine Similarity Algorithm

Christi pereira[1], Sridhar Iyer[2], Chinmay A. Raut[3]
[1,2,3] *Computer Engineering, Universal college of engineering,*

Abstract—Recommender system recommends the object based upon the similarity measures. Similarity between this objects can help in organizing similar kind of objects. Similarity can also be seen as the numerical distance between multiple objects that represented as value between the range of 0 (not similar at all) and 1 (completely similar).Similarity are highly subjective in nature and dependent on the domain and application. , system build a model from a user's past activities as well as similar decisions made by other users; then use that model to predict items(or ratings for items) that user may have an interest in. In this model, a user rates a set of items based on which we find the similarities between the users who has nearly same ratings for a set of items, similarity is calculated based on cosine similarity method. After finding the similar object, we recommend relevant item sets to the users who are similar with the users who had already rated the items which we had recommended. This feature is extremely beneficial for the users as well as the website because an item that seems excellent to one person may seem dull for the other to buy. It basically tries to find the users which are similar to the current user behavior.

Keywords— Recommendation system; Similarity measures; cosine similarity;

I. INTRODUCTION

With increasing dynamic data on the Web has created a need for recommender system to help in selecting intersecting data. Recommender Systems assist users to Travers through large product assortments, making decisions in e-commerce scenario and overcome information overload. Most prominent example is the book recommendation service of Amazon. In daily life, people rely on recommendations by spoken words, reference letters, and news reports from news media, general surveys from other people. Recommender systems assist and augment this natural social process to help people sort through available restaurants, books, WebPages, movies, music, articles, jokes, grocery products, and to find the most interesting information and valuable items for them. As a whole the goal of any recommendation system is to present users with a highly relevant set of items. This project is aims in designing and implementation of a recommendation engine. Recommender systems typically produce a list of recommendations Recommendation system is a specific type of *information filtering* system that tries to present information of items (such as movies, music, news) that are likely of user interest. Recommender systems help users navigating through large product assortments, in making decisions in an e-commerce background and overcome information overload over web. Users should not have access to the precise kind of information they were looking out for.

A user would obviously prefer a website who recommends him something that may be useful to him over a website that simply requires the user to navigate or dig deep into the site to find the products that the user needs. Technology has dramatically reduced the barriers to publishing and distributing information. Recommender System is designed to provide meaningful references to a group of users with a common interest. Many algorithms and techniques are used to understand the traditional and modern approaches. One of the most promising such technologies is cosine similarity. Cosine similarity works by building a database of preferences for items by users. A new user, say Bill, is matched against the database to discover neighbors, which are other users who have historically had

similar taste to Bill. Items that the neighbors like are then recommended to Bill, as he will probably also like them.

In this paper, we first calculate similarity measures logically and illustrate recommender system based cosine similarity measure and similarity is calculated based on user stated rating for each object and then similar kind of object are recommend to similar type of user.

II. LITERATRE SURVEY

"*Towards the Next Generation of Recommender Systems: A survey of the state of the art and possible extensions*" (Gediminas Adomavicius and Alexander Tuzhilin, 2005) researchers have describes the current generation of recommendation methods like content-based, collaborative, and hybrid recommendation approaches [1]. In this paper, various limitations of various recommender systems have been reviewed and discussed possible extensions that can provide better recommendation capabilities. Possible extensions like, improved modeling for users and items, incorporation of the contextual information into the recommendation process, support for scalability and trustworthiness and privacy need to be solved. "Implementation of item and content based Collaborative Filtering Techniques based on ratings average for recommender systems" (Rohini Nair and Kavita Kelkar, 2013) here researchers have stated that system suggests items to purchase according to the users interest. Almost every applications based on e commerce are working on the concept of recommendation system. The survey is done on many recommender systems to shows that a lot of work is being carried out in this area and the project proposes a mixture of various techniques for recommendation systems. All the basic techniques are included based on item based approach and content based which are the basic building blocks for a recommender systems. In this paper both algorithms are implemented and their results are presented and compared [2]. In this paper two algorithms on Item based and Content based collaborative filtering techniques were successfully implemented on MySQL/PhP.

Item based considers other users aspects also into its predictions while content based is confined to its own available information. On implementing item based approach proved to be more efficient as compared to the content based approach, however depending on a user's need either of the recommendations.

The over-specialization problem in item based needs to be overcome. "Evaluating the Performance of Similarity Measures Used in Document Clustering and Information Retrieval" (R.Subhashini and V.Jawahar Senthil Kumar, 2010) researcher has evaluated that Cosine Similarity measure is particularly better for text Documents [3]**.** It is not suitable for calculating distance measures.

III. TRADITIONAL SIMILARITY ALGORITHM BASED ON COSINE

A. User Item Rating(UIR)

The prediction of the target user's rating for the target items are observed based on the users' ratings. And the user-item rating are keep in database as centralize. Each user is represented by item-rating pairs, and can be summarized in a item-user table, which contains the ratings Rik that have been provided by the ith user for the item kth, the table as following:

TABLE I. ITEM-USER RATINGS TABLE

[1] User [2] items	[3] User 1	[4] User 2	[5] User 3	[6] …….	[7] User n
[8] Item1	[9] R11	[10] R12	[11] R13	[12] ……	[13] R1n
[14] Item2	[15] R21	[16] R22	[17] R23	[18] …….	[19] R2n
[20] …….	[21] ……	[22] ……	[23] ……	[24] ……	[25] ……
[26] Item m	[27] Rm1	[28] Rm2	[29] Rm3	[30] ……	[31] Rmn

Where Rik denotes the score of item k rated by an active user i. If user i has not rated item k, then Rik =0. The symbol m denotes the total number of users in system, and n denotes the total number of items in database.

B. Calculating cosine similarity

A popular measure metric of similarity between two vectors of n dimensions *is the cosine similarity metric*[5]. Cosine similarity is used in many applications, such as text mining and information retrieval [6, 7]. When documents are represented as term vectors, then the similarity between two documents corresponds to the correlation between the vectors. This is quantified as the cosine similarity of the angle between vectors, that is, the so-called cosine similarity. Cosine similarity is one of the most popular similarity measure applied to text documents, such as in numerous information retrieval applications [9] and clustering too [8]. Cosine measures are stated, the angle between two vectors of ratings as the target item t and the remaining item r(1).

$$sim(t,r) = \frac{\sum_{i=1}^{m} R_{it} R_{ir}}{\sqrt{\sum_{i=1}^{m} R_{it}^{2} \sum_{i=1}^{m} R_{ir}^{2}}} \quad \ldots (1)$$

Where R_{it} is the rating of the target item t by user i, R_{ir} is the rating of the remaining item r by user i, and m is the number of all rating users to the item t and item r[4].

C . Generating Recommendations

After calculating similarity based on rating of individual items, then we can calculate the weighted average of neighbors' ratings [4], weighted by their similarity to the target item (2). The rating of the target user u to the target item t is as following:

$$P_{ut} = \frac{\sum_{i=1}^{c} R_{ui} \times sim(t,i)}{\sum_{i=1}^{c} sim(t,i)} \quad \ldots (2)$$

Where Rui is the rating of the target user u to the neighbor item i, sim (t, i) is the similarity of the target item t and the neighbor it user i for all the co-rated items, and m is the number of all rating users to the item t and item r [4].

D. Performance Measurement.

Several metrics have been proposed for assessing the accuracy of cosine measure method. They are divided into two main categories: statistical accuracy metrics and decision-support accuracy metrics[4].

Statistical accuracy metrics estimate the accuracy of a prediction algorithm by comparing the numerical standard deviation of the predicted ratings from the actual user ratings. Most frequently used are mean absolute error (MAE), root mean squared error (RMSE) and correlation between ratings and predictions [4]. As statistical accuracy measure, mean absolute error is used.

Formally, if n is the number of ratings in an item set, then MAE is well-defined as the average absolute difference between the n pairs. Assume that p1, p2, p3... pn is the predicted users' ratings, and the corresponding real ratings data set of users is q1, q2, q3... qn(3). See the MAE definition as following:

$$MAE = \frac{\sum_{i=1}^{n} | p_i - q_i |}{n} \quad \ldots (3)$$

The lesser the MAE, the more precise the predictions would be, allowing user for better recommendations to be achieved. MAE has been computed for different prediction algorithms.

III. CONCULSION AND FUTURE SCOPE

Recommender systems can assistance users in finding interesting items and they can be widely used in our life with the development of e-commerce. Many recommendation systems employ the cosine similarity method, which has been proved to be one of the most successful techniques in similarity measures systems in recent years. With the continuing increase of customers and products in e-commerce systems, the time consuming search of the target customer in the total customer space resulted in the failure of ensuring the requirement of recommender system. At the same time, it suffer in quality when the number of the records in the user increase. Sparsity of source data set is the major reason causing the poor quality. The quality of rating given by user may affect the similarity score. To verify the user stated rating for each items must be verified by some means to get accurate results.

REFERENCES

I. Gediminas Adomavicius and Alexander Tuzhilin "Toward the Next Generation of Recommender Systems: A survey of te State-of-the-Art and Possible Extensions", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 6, June 2005.
II. Rohini Nair and Kavita Kelkar "Implementation of Item and Content based Collaborative Filtering Techniques based on Ratings Average for Recommender Systems", International Journal of Computer Applications (0975-8887), Volume 65- No.24, March 2013.
III. R.Subhashini and V.Jawahar Senthil Kumar "Evaluating the Performance of Similarity Measures Used in Document Clustering and Information Retrieval", First International Conference on Integrated Intelligent Computing, 2010
IV. SongJie Gong "A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering", JOURNAL OF SOFTWARE, VOL. 5, NO. 7, JULY 2010.
V. A. Karnik, S. Goswami. And R. Guha, "Detecting Obfuscated Viruses Using Cosine Similarity Analysis," Proceedings the First Asia International Conference on Modelling & Simulation (AMS'07), 2007.
VI. M. Steinbach, G. Karypis, and V. Kumar, "A comparison of document clustering techniques," KDD workshop on text mining, 2000.
VII. http://www.stanford.edu/class/cs276/handouts/lecture13-vector-classify.ppt.
VIII. R. B. Yates and B. R. Neto. Modern Information Retrieval. ADDISONWESLEY, New York, 1999.

IX. Y. Zhao and G. Karypis. Evaluation of hierarchical clustering algorithms for document datasets. In Proceedings of the International Conference on Information and Knowledge Management, 2002.

X. F. Gregory Ashby and Daniel M. Ennis, "Similarity_measures", http://www.scholarpedia.org/article/Similarity_measures, 2005.

XI. Shyam Boriah, Varun Chandola and Vipin Kumar, " Similarity Measures for Categorical Data: A Comparative Evaluation", SIAM.

XII. Sameh Al-Natour, Izak Benbasat and Ronald T. Cenfetelli , " The Role of Similarity in e-Commerce Interactions: The Case of Online Shopping Assistants", Association for Information Systems AIS Electronic Library (AISeL), 2005.

XIII. HengSong Tan and HongWu Ye, " A Collaborative Filtering Recommendation Algorithm Based On Item Classification", Pacific-Asia Conference on Circuits,Communications and System,2009.

XIV. Anna Huang, "Similarity Measures for Text Document Clustering", NZCSRSC 2008, April 2008, Christchurch, New Zealand.

XV. BadrulSarwar, George Karypis, Joseph Konstan, John Riedl "Analysis of Recommendation Algorithms for E-Commerce", GroupLens Research Group/Army HPC Research Center, Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455.

XVI. Lokesh Sahu and Mr. Biju R. Mohan " An Improved K-means Algorithm Using Modified Cosine Distance Measure for Document Clustering Using Mahout with Hadoop".

XVII. Saprativa Bhattacharjee, Anirban Das, Ujjwal Bhattacharya, Swapan K. Parui and Sudipta Roy" Sentiment Analysis using Cosine Similarity Measure", IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS),2015

