

# Genetic Algorithm Based Similarity Transitivity in Collaborative Filtering

Prof. P. A. Khodke

Dept. of Computer Science & Engg.  
PRMCEAM, Badnera Amravati University,

Miss. Priti B. Rathod

PRMCEAM, Badnera  
Amravati University,

## Abstract:

*Collaborative filtering is the process of filtering for information or patterns using techniques involving collaboration among viewpoints, data sources, etc. Collaborative filtering methods have been applied to many different kinds of data including: sensing and monitoring data. Collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating). In this paper similarity based technique makes the prediction of user with similar taste or similar items. It firstly filter out those inaccurate similarities by clustering methods then set proper intersection threshold value then replace them with the transitivity similarity. This paper proposed genetic algorithm based similarity transitivity which is used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution.*

**Keywords:** *recommenders systems, big data collaborative filtering, data mining, similarity transitivity, map reduce, genetic algorithm*

## 1. Introduction

Information overload means increase in the amount of data. So, it will make users hard to find their preferred items. Collaborative filtering is used to solve the information overload problem by presenting each content to individual users based on their interests. Collaborative filtering is a method which is based on similar factor. It predicts users according to their similar taste or item. But there is a problem when the number of users increases rapidly. This approach is suffering from data sparsity problem which will generate inaccurate neighborhoods for each user.

Collaborative filtering can be classified into similarity based methods and model based methods [15-17, 19]. The similarity based methods predict the user's interests based on the weight which is the combination of rating of the similar users on the same item. However the sparse data will affect the accuracy of similarity based data. It will cause the poor recommendation [10-11] on data which will be generated through these inaccurate similarities [8].

Recently many approaches have been proposed to decrease data sparsity problem [14] that is threshold based similarity transitivity method in collaborative filtering which implement the MapReduce technique to decrease the data sparsity problem that increase the scalability.

Genetic algorithm (GA) [20, 21] is an optimization search function, which can be used to find out the best recommendation of similarity.

## 2. Problem Statement

The problem of the system is to analyse the large number of data results on the public data set and the real-world data set which show that the TST method is much more accurate and provides more diverse recommendations especially on the sparser data set. Moreover, the TST method is developed to be scalable with Map Reduce, which is a programming paradigm that comes with a framework to provide programmers an easy way for parallel and distributed computing. Proposed work basically deals with finding the threshold optimal value theoretically. The genetic algorithm can be used to find the optimal value. Genetic algorithm (GA) is an optimization search function, for finding a best recommendation out of a large variable. GA-based approach is the combination of collaborative filtering. It is one of the most powerful search techniques for finding an optimized solution of a matching item which is recommended by the user.

Genetic Algorithm-based CF scheme is used to improve the accuracy of prediction in a recommendation system. It selectively chooses the top similarity measure values, and codes the user's profile into a chromosome.

### 3.Related Work

D. Goldberg, D. Nichols, B. M. Oki, and D. Terry in 1992 [3] have proposed the Tapestry which is the mail system. It predicts that information filtering can be more effective when humans are involved in the filtering process. Tapestry was designed to support both content-based filtering and collaborative filtering which perform filtering by recording their reactions to documents they read.

S. H. S. Chee, J. Han, and K. Wang in 2001[17] who has extended the work of improving the accuracy of the prediction's, the researcher developed an efficient collaborative filtering method called Rectree [3] which stands for recommendation tree. It addresses scalability problem with divide and conquer approach. By implementing Rectree method it outperforms the well-known collaborative filter CorrCF in both execution time and accuracy.

K. Goldberg, T. Roeder, D. Gupta, and C. Perkins in 2001 introduced Eigentaste [13] which is a collaborative filtering algorithm that uses universal queries to elicit real-valued users rating on a common set of the item and applied principal components analysis (PCA) to the resulting dense subset of the rating matrix. Eigentaste requires  $O(1)$  constant time.

G. Linden, B. Smith, and J. York, in 2003[5] introduced the recommendation algorithm by amazon.com [5] to personalize the online store for each customer. This algorithm use item-to-item collaborative filtering. It produces recommendations in real time, Scales to massive data sets, and generates high quality recommendations. Rather than matching the user to similar customers, item-to-item collaborative filtering matches each of the user's purchased and rated items to similar items, then combines those similar items into a recommendation list.

G. Shani, D. Heckerman, and R. I. Brafman in 2006[6] adopted a static view of the recommendation process and treated it as a prediction problem that is Markov decision processes (MDP).MDPs introduce two benefits: they take into account the long-term effects of each recommendation, and they take into account the expected value of each recommendation. To succeed in practice, an MDP-based Recommender system

must employ a strong initial model and the aim of this paper is concerned with the generation of such a model.

In 2007 Michael J. Pazzani and Daniel Billsus [15] introduced Content-based recommendation systems. It is used in a variety of domains ranging from recommending web pages, news articles, restaurants, television programs, and items for sale. Content-based recommendation systems share means to describe the items that may be recommended, means to creating a profile of the user that describes the types of items the user likes, and a means of comparing items to the user profile to determine what to recommend.

T. Zhou, J. Ren, M. Medo, and Y. C. Zhang in 2007[20,22] used method to compress bipartite networks. One-mode projection is always less informative than the bipartite representation, a proper weighting method is required to better retain the original information. In this network-based resource-allocation is dynamically a weighting method which can be directly applied in extracting the hidden information of networks, with remarkably better performance than the widely used global ranking method as well as collaborative filtering.

A. Gunawardana, G. Shani in 2009 [1] users a natural accuracy measure for recommendations. As to test recommender systems it present two performance measures that can be estimated, under assumptions, without data ratings that is missing not at random (MNAR). As to achieve optimal test results, the author presented appropriate surrogate objective functions for e-client training on MNAR data. Their main property is to account for all ratings whether observed or missing in the data.

J. Dean and S. Ghemawat in 2008[12] introduced MapReduce model which is implemented for processing and generating large data set. It specifies a map function using the logical record containing a pair of key/value. Then it applies the reduce function that merges all intermediate value.

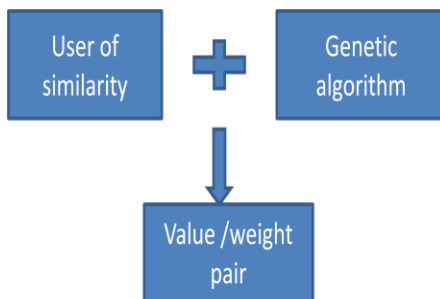
Feng Xie, Zhen Chen, Hongfeng Xu, Xiwei Feng, and Qi Hou in 2013[4] introduced Threshold based Similarity Transitivity (TST) method in collaborative filtering with cloud computing. It filters the inaccurate data using intersection threshold method and then replaces that value with transitivity similarity. The TST method is designed with MapReduce framework which is based on cloud computing platform. The experimental results after demonstrating the TST, shows that it is the tradeoff between quality and quantity of

similarity by calculating appropriate threshold value. It is experimentally known that if data set is sparser then the optimal threshold value is smaller.

Simon Fong, Yvonne Ho, Yang Hang in 2008[18] introduced Genetic Algorithm for Hybrid Modes of Collaborative Filtering in Online Recommenders. In this paper researcher presented a GA-based approach which supports combined modes of collaborative filtering. It shows how the input variables can be coded into GA chromosomes in various modes and how GA can be used by recommenders.

Hema Banati and Shikha Mehta in 2010[7] introduced Genetic and memetic algorithms. It is most successful combinatorial optimization approach. Memetic Algorithms (MAs) are enhanced genetic algorithms which is a combination of local search. Local Search process is solution after every generation which helps in improving the convergence time of MA. In this paper the researcher presents multi-perspective comparative evaluation of memetic and genetic evolutionary algorithms for model based collaborative filtering recommender system.

Jesus Bobadilla, Fernando Ortega, Antonio Hernando, Javier Alcalá in 2011[9] introduces a metric to measure similarity between users, which is used in collaborative filtering processes. The proposed metric is combination of values and weights. Values are calculated for each pair of users based on similarity.



**Fig: metric measure similarity**

ZHU Zhenfang, LIU Peiyu, ZHENG Yan, ZHAO Jing, LI Shaohui and WANG Jinlong in 2011[23] discusses the content-based filtering and collaborative filtering, and proposes a hybrid filtering model using this two methods to overcome their own shortages. In this hybrid filtering method, genetic algorithm is used to generate initial profiles on server-side, and particle swarm Optimization is used to update the profiles with the information from users.

Deepa Anand in 2012 [2] uses feature extraction method to get genetic programming to convert from user-item space to user-feature preference space. In this the feature space is much smaller than the item space. The advantage of using this approach is to reduce sparse high dimensional preference information into a compact and dense low dimensional preference data.

#### 4. Conclusion and Future scope

Similarity based transitivity method is used to filter low-quality similarities and replace them with transitivity similarities to increase similarity quantity. This method is evaluated on dataset MovieLen and AppChina. The performance improves on data sets so similarity transitivity is the well trade off between quantity and quality. Genetic Algorithm-based CF scheme is used to improve the accuracy of prediction in a recommendation system. It is one of the most powerful search techniques for finding an optimized solution of a matching item which is recommended by the user. However, implementation of this work is still under progress and is left as future work.

#### 5. REFERENCES

- [1] A. Gunawardana, G. Shani, A survey of accuracy evaluation metrics of recommendation tasks, *The Journal of Machine Learning Research*, vol. 10(2009),2935-2962.
- [2] Deepa Anand, Feature Extraction for Collaborative Filtering: A Genetic Programming Approach, *IJCSI International Journal of Computer Science Issues*, Vol. 9, Issue 5, No 1, September 2012.
- [3] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 12(1992), 61-70.
- [4] Feng Xie, Zhen Chen\_, Hongfeng Xu, Xiwei Feng, and Qi Hou. TST: Threshold based Similarity transitivity method in collaborative filtering with cloud computing. *TSINGHUA SCIENCE AND TECHNOLOGY* ISSN 1007-0214 11/11 Volume 18, Number 3(2013), 318-327.
- [5] G. Linden, B. Smith, and J. York, Amazon. Com recommendations: Item-to-item collaborative filtering, *IEEE Internet Computing*, no. 1(2003), 76-80.
- [6] G. Shani, D. Heckerman, and R. I. Brafman, an MDP based recommender system, *Journal of Machine Learning Research*, vol. 6, no. 2(2006), 1265-1295
- [7] Hema Banati and Shikha Mehta, A MULTI-PERSPECTIVE EVALUATION OF MA AND GA FOR COLLABORATIVE FILTERING RECOMMENDER SYSTEM, Vol.2, No.5, and October 2010:
- [8] H. Steck, Training and testing of recommender systems on data missing not at random, in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, Washington DC, USA, (2010), 713-722.

- [9] Jesus Bobadilla, Fernando Ortega, Antonio Hernando, Javier Alcalá, Improving collaborative filtering recommender system results and performance using genetic algorithms, *Knowledge-Based Systems* 24 (2011),1310–1316.
- [10] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. Riedl, Evaluating collaborative filtering recommender systems, *ACM Transactions on Information Systems (TOIS)*, vol.22, no. 1(2004), 5-53.
- [11] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, Collaborative filtering recommender systems, *The Adaptive Web*. Heidelberg: Springer Berlin, (2007), 291-324
- [12] J. Dean, and S. Ghemawat, Map Reduce: simplified data processing on large clusters, *Communications of the ACM*, vol. 51, no. 1(2008), 107-113.
- [13] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, Eigentaste: A constant time collaborative filtering algorithm, *Information Retrieval*, vol. 4, no. 2(2001), pp. 133-151
- [14] Manos Papagelis, Dimitris Plexousakis, Themistoklis Kutsuras, Alleviating the Sparsity Problem of Collaborative Filtering Using Trust Inferences.
- [15] Michael J. Pazzani and Daniel Billsus. Content-based recommendation system. *The adaptive web*. Heidelberg: spring berlin (2007), 325-341.
- [16] M. Van Setten, M. Veenstra, A. Nijholt, and B. van Dijk, Prediction strategies in a TV recommender system method and experiments. in *Proceedings of the Second IADIS International Conference Algarve, Portugal( 2003)*, 203-210
- [17] S. H. S. Chee, J. Han, and K. Wang, Rectree: An efficient collaborative filtering method, *Data Warehousing and Knowledge Discovery*, Springer Berlin Heidelberg (2001), pp. 141-151.
- [18] Simon Fong, Yvonne Ho, Yang Hang. Using Genetic Algorithm for Hybrid Modes of Collaborative Filtering in Online Recommenders (2008).
- [19] T. Hofmann, Latent semantic models for collaborative filtering, *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1(2004), 89-115.
- [20] T. Zhou, J. Ren, M. Medo, and Y. C. Zhang, Bipartite network projection and personal recommendation, *Physical Review E*, vol. 76, no. 4(2007), 046-115.
- [21] Yvonne Ho, Simon Fong, Zhuang Yan. Using Genetic Algorithm for Hybrid Modes of Collaborative Filtering in Online Recommenders.
- [22] Z. Chen, F. Y. Han, J. W. Cao, X. Jiang, and S. Chen, Cloud computing-based forensic analysis for collaborative network security management system, *Tsinghua Science and Technology*, vol. 18, no. 1(2013),40-50.
- [23] ZHU Zhenfang1, LIU Peiyu, ZHENG Yan, ZHAO Jing, LI Shaohui and WANG Jinlong, Hybrid filtering model based on particle swarm optimization and genetic algorithm, *International Journal of the Physical Sciences* Vol. 6(14) (2011),3518-3523