

# Data Mining and its Application in Recommender Systems based on Tensor Decomposition Method

Abbas Badiheshin\*

Department of Information Technology, Payam Noor University, Tehran 19395 – 3697, I.R of IRAN;

A.Badiheshin@gmail.com

## Abstract

Data mining is the process of arranging and classifying massive data and revealing the related information. Nowadays, data mining is utilized as a highly significant means for managers to understand more precisely the situation of their organization and help with decision making. Utilization of recommender systems is a modern technique to examine the data available in an organization with the use of powerful tools. Challenges involved with the dataset of these systems have intensified the complexities of decision making, such that most effective variables are unknown and the relationship between them is nonlinear and complicated. In such a condition, traditional devices cannot be employed to analyze data and derive knowledge from them; therefore, utilizing matrix analysis and then tensor decomposition has been preferred in recent years compared to other methods. In the present study, time is considered as the third dimension of tensor and an independent factor. The results indicate that using time is effective in reducing estimation error.

**Keywords:** Data Mining, Recommender Systems, Tensor Decomposition

## 1. Introduction

The modern knowledge utilizes various principles and models to analyze physical, biological, economic, and social systems. This attitude started with initial scientific models such as Newton's laws of motion and Maxwell's electromagnetic balance and continued with different applications of the mentioned models in scientific and engineering fields. In most cases; however, the fundamental models are unknown or the target system is highly complicated. In such conditions, the collected data are a valuable resource to obtain initial fundamental models<sup>1</sup>.

Increasing development of computers also creates a large amount of the produced data by different systems. On the other hand, this increasing growth of data collection has facilitated processing, distributing, and manipulating data<sup>2</sup>. Therefore, these data can create some models using the relationship between the variables of the system in the absence of initial fundamental.

Technologies of producing and collecting data have

rapidly developed in recent years. The challenge that organizations are faced with is not to merely collect data, but extracting useful information underlying the collected data is their major challenge<sup>3</sup>. It is in such situations that technology development should sufficiently be utilized in order to make use of this potential knowledge, and data mining is an appropriate response to extract this wealth<sup>4</sup>. Therefore, with development of information technology, decision making systems or in general computer-based systems have gained a high level of significance. In this regard, expert systems play a major role as one of the sections entitled as artificial intelligence. In expert systems, different types of decisions are made with the help of computers. Expert systems are knowledge-based systems, and knowledge is in fact the most important part of them. In these systems, the knowledge is transferred from the expert of any field to computers. Expert systems are widely utilized in different sciences all over the world. Since ever, different expert systems have designed and proposed in fields like industry, control, astronomy, financial decision making, etc<sup>5</sup>.

\* Author for correspondence

Utilization of expert software systems has some advantages including:

- The individual's expertise is transient and mortal. For example, an individual may change his occupation, get sick, and so on, but the computer expertise is permanent.
- Individual's knowledge is not durable, they can have days off, recreational plans, and so on, all of which can affect the individual's natural functioning; however, computers are steady and produce the same output in identical and certain situations.
- The individual's expertise is difficult to transfer. An individual cannot be present in two places at the same time; however, computer expertise is transferable.
- Individual's expertise is expensive. Personnel's wage is more costly than software and hardware.
- High level of functionality, complete and quick response time, acceptable reliability, comprehensibility, flexibility, low risk, durability, and availability of multiple specialties are all among the characteristics of computer expertise.

Data mining is the process of utilizing a computerized methodology that seeks the knowledge hidden in the data by employing different techniques and algorithms<sup>1</sup>. This cooperative process between man and computers<sup>6</sup> ultimately attempts to figure out the meaningful patterns and principles within the data<sup>7</sup>. Data mining considers massive databases as knowledge resources<sup>8</sup>.

Therefore, data mining is the process of discovering different models, summaries and the values originated from a specific dataset<sup>9</sup>.

## 2. Materials and Methods

In data extraction view, the method of providing the users with offers is investigated, in which two algorithms are used depending on circumstances. The first type of algorithm includes memory-based algorithms that use the complete rating matrix to provide recommendations. The second type is model-based algorithms in which the rating matrix is employed to create a model and then the model is used to provide offers. Memory-based algorithms cause better and more precise results compared to model-based algorithms, and when the evaluation matrix changes frequently, they become more appropriate. On the other hand, algorithms need a lot of calculation time, which leads to utilization of approximate model-based algorithms in massive databases.

In data collection view, the method of data collecting

from the users is investigated, and the collected data are generally divided into two categories. The first category includes data that are extracted from the user's behavior (implicit data). Since in some implementations of CF extraction of rating from the users is not simply probable, the data that the users leave while visiting the pages are used to provide offers, like the procedure of visiting the pages and duration of visiting different items on the website. The second category of the data are those retrieved from the user (explicit data) that the users specify while they determine their favorites during purchasing previous items and rating each item whereby he enables the CF system to provide him with more precise offers. In other words, while users are buying their favorite items, they record their comment about the items in the form of feedback (through rating). CF utilizes these data in its subsequent offers and recommends the user with new items according to his favorites. The extracted data for the user are more precise than those derived from his behavior because he expresses his views more precisely<sup>10</sup>.

Time has never been considered in recommender systems as a significant dimension. The present study; however, deals with a dynamic recommender system that not only includes items and users but also focuses on time as an important dimension. In fact, the data have three dimensions, with the third one being time. Therefore, the collected data are stored in a three dimensional arrangement which is called tensor<sup>11</sup>. Now the major challenge is how to work with this three dimensional object in order to create a recommender system. This modeling enables us to provide different individuals with different offers at different times. However, the tools of working with tensors are in fact generalizations of matrix decompositions for the three dimensional status.

As was determined, most of the proposed algorithms in matrix decomposition for recommender system are two-dimensional solution. In the present study, a three-dimensional tensor of user-item-time was utilized, in which the following decomposition can be used likewise the two-dimensional status, which is calculated according to HoSVD algorithm<sup>12</sup>.

In the present study, a tensor  $R$  that has three dimensions of user-item-time is intended to be decomposed into the three matrices of hidden features of user ( $U$ ), item ( $I$ ), and time ( $T$ ). As was referred to, it is calculated based on HoSVD algorithm. In fact, the problem is like the following formula:

$$R = (U, I, T) S \quad (1)$$

Since the results of calculating the matrices of the user's characteristics, item and time cannot be equal to the main tensor R due to the loss of a large part of the dataset, the result is approximate and should be presented in a way that the difference between the approximate and real values can be minimum.

$$R = (U, I, T) S \tag{2}$$

The above formula is the main problem, where U is the matrix of user vectors, I is the matrix of item vectors and T is the matrix of time vectors. In fact, S will be a tensor which is called core. Matrices of user, item and time are orthogonal ones. The number of the rows of the decomposed matrices of user, item and time is identical to the number of users, items, and time, and the number of its columns is different. The best number of columns of the matrices that is indicated with k in the results is obtained by calculating different values. In fact, the value of k is the amount of the final choice that has the minimum error while calculating the approximate amount of R. Here, the user and the item are assumed separated from each other, and assuming time as an independent factor is the feature of this approach.

The offer to user *u* in item *I* and at time *t* through VoSVD decomposition is determined as follow:

$$r_{uit} = \sum_m \sum_n \sum_l (u_{um} i_{nl} t_{tl}) S_{uit} \tag{3}$$

In the above formula, if UmT is the m<sup>th</sup> row of matrix U, In T is the n<sup>th</sup> row of matrix I, and TIT is the i<sup>th</sup> row of matrix T, then the approximate value of tensor can be simulated as follow:

$$r_{uit} = (U_u^T, I_i^T, T_t^T) S \tag{4}$$

However, in the response of decomposing HoSVD it is likely that the number of offers increases in which no data are available because like the two-dimensional status, we do not intend to come up with very massive components reconstructed by this algorithm. Therefore, the number of the columns of the decomposed matrices should be controlled in a way that the components do not become very massive. In tensor situation; therefore, the problem of HoSCD decomposition can be replaced with the problem of regulated minimization, and the appropriate response can be obtained from this problem.

$$\sum_{(u,i,t) \in k} [(r_{uit} - \bar{r}_{uit}) + \lambda(\|U_u\|^2 + \|I_i\|^2 + \|T_t\|^2)]^2 \tag{5}$$

The second part of the above formula is aimed at preventing the sudden leap and the amount of  $\lambda$  is

between 0 and 1. It should be noted that the number of the elements with score is k, which is 1% and much fewer than the total tensor elements.

### 3. Conclusion and Suggestions

The proposed algorithm, which is derived from HoSVD algorithm, has been carried out for different K values on different Movie Lens databases with various dimensions. In this database, users attributed different points, from 1 to 5, to several different videos on different occasions. Films that are not rated by the users are specified with 0 values. The type of database value is nominal. The first database, used for assessment, is a tensor with a volume of 100\*100\*7, 100 members, 100 films and seven different intervals. The number of points in this tensor is 503 and the density of the tensor is 71%. The second database is a tensor with a volume of 200\*200\*7, 200 users, 200 films and seven various intervals. Each interval covers a month. This database has been attributed 1945 points with a density of 69%. If we don't consider time as a dimension, the density of the database is multiplied 7 times and the problem can be solved easier; however, time, as an independent dimension, heightens the precision of the prediction. To analyze matrix decomposition, SVD method and to evaluate and compare two methods, RMSE method, with the following formula, has been used.

$$RMSE = \sqrt{\frac{\sum_{u,i,t} (r_{u,i,t} - \hat{r}_{u,i,t})^2}{n}} \tag{6}$$

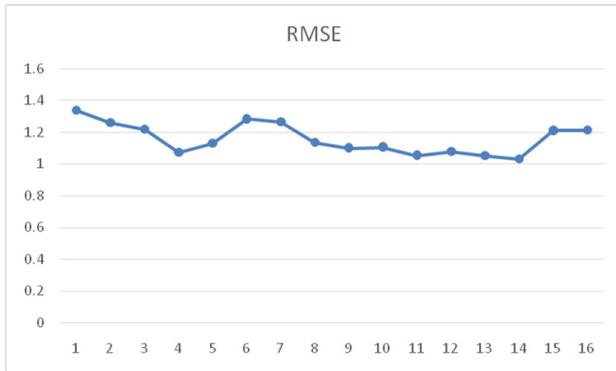
In Table 1 comparison done for two data between matrix analysis with user-product dimensions and three-dimensional tensor of user-item-time with hidden attributes 1-6

**Table 1.** RMSE value in the database with the k = 1-6

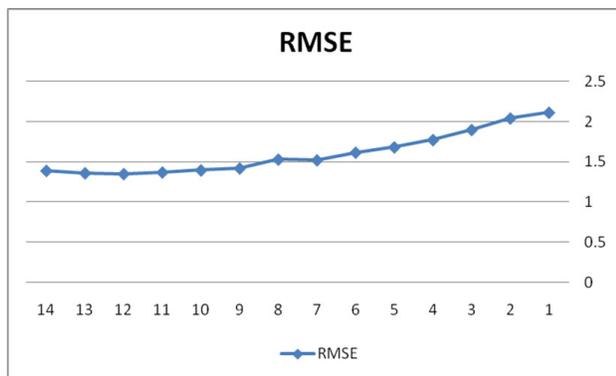
K	TF (7*200*200)	MF (200*200)	TF (7*100*100)	MF (100*100)
1	11.2	56.2	34.1	41.2
2	4.2	56.2	26.1	24.2
3	90.1	29.2	22.1	10.2
4	77.1	19.2	7.1	0.2
5	69.1	12.2	13.1	91.1
6	11.2	7.2	28.1	83.1

RMSE value decreases for low hidden features and after k = 5 this value is rising, but in 6 decreases up to k = 14 again. After k = 14, value is rising again. According to this table, k = 14 is the optimum value for analysis. The

initial decrease, up to  $k = 4$ , maybe due to the over fitting issue the accuracy of which is open to research and study in Figure 1 and Figure 2.



**Figure 1.** RMSE values in tensor decomposition with different amounts of hidden attributes in the dimensions  $100 \times 100 \times 7$ .



**Figure 2.** RMSE values in tensor decomposition dimensions  $7 \times 200 \times 200$ .

Nowadays, were surrounded by a large number of data that are not so valuable for the managers since they are not known; therefore, they will be neglected. However, if these apparently undervalued data are stored purposefully and used in data mining, they result in a large body of knowledge and can be useful in managerial decisions. Time is a datum that has never been used as

an independent component in recommender systems. In the present study; however, time was examined as an independent factor, and the network was investigated in user-item-time form. In this method, SVD algorithm was compared in identical circumstances by utilizing the structure change of HoSVD algorithm for the recommender system and tensor decomposition through optimization method. As the results indicated, using time as an independent factor, despite of its difficulty, involved fewer errors. For future applications, it is suggested that a criterion be used to divide time into intervals so that errors decrease as far as possible.

## 4. References

1. Kantardzic M. Data Mining: Concepts, Models, Methods, and Algorithms. John Wiley; 2003.
2. Fayyad U, Uthwrsusamy R. Data Mining and Knowledge Discovery in Databases. Communication of ACM; 1996.
3. Lee SJ, Siau K. A Review of Data Mining Techniques. Industrial Management and Data Systems; 2001. p. 41–6.
4. Kudyba S, Hoptroff R. Data Mining and Business Intelligence: A Guide to Productivity. IGP; 2001.
5. Durkin J. Expert Systems: Design and Development. New York: Prentice Hall; 1994.
6. Berson A, Smith S, Thearling K. Building Data Mining Applications for CRM. McGraw Hill; 2000.
7. Berry M, Linoff G. Mastering Data Mining. John Wiley Hoboken, NJ; 2000.
8. Larose DT. Discovering Knowledge in Data. John Wiley; 2005.
9. Hand D, Mannila H, Smith P. Principles of Data Mining. Prentice Hall; 2005.
10. Xu W, Liu X, Gong Y. Document clustering based on non-negative matrix factorization. Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA; 2003. p. 267–73.
11. Kolda T, Bader B. Tensor Decompositions and Applications. SIAM Rev. 2009; 51:455–500.
12. Kruskal J. Rank, decomposition, and uniqueness for 3-way and N-way arrays. In R. Coppiand S. Bolasco (Eds.), Multi-way data analysis; 1989. p. 7–18.