Annales Univ. Sci. Budapest., Sect. Comp. 42 (2014) 83–91

INSTABILITY OF MATRIX FACTORIZATION USED IN RECOMMENDER SYSTEMS

Antal Buza (Dunaújváros, Hungary)
Piroska B. Kis (Dunaújváros, Hungary)

Dedicated to András Benczúr on the occasion of his 70th birthday

Communicated by Péter Racskó

(Received June 1, 2014; accepted July 1, 2014)

Abstract. The factorization of the matrices is not unique. The several different results of the factorization of a matrix would be seem as reliable and consequently usable in the recommender systems if ones were stable. It means that the deducted conclusions are mainly the same without reference to differences in the product of the factor matrices. In this paper we discuss the interpretation of the stability of matrix factorization and give a few definitions. We ascertain that the matrix factorization is unstable.

1. Overview

Recommender systems are such systems which based on previously collected observations try to determine the parameters of previously not observed cases.

1.1. Recommender systems

Recommender systems are more and more widely used for selling of various products and services such as tourism, movie, book and so on. Recommender

Key words and phrases: Factorization of matrices, stability, recommender system.
2010 Mathematics Subject Classification: 15A23, 49K40

systems observe the user preferences over items and based on the experiences ones attempt to predict the relations between users and items. The goal of a recommender system is to help the user in selection items from a large number of choices. Recommender systems attempt to discover the needs of the users even in the cases when no user activities in the past are available. Using this approach, the user first provides ratings of some items, and after it the system recommends other items based on ratings similar users have already provided.

Regarding to the recommender systems, the most cited examples are for entertainment products such as movies. Many customers/users will see less or more number of movies and each user will appreciate each movie. Thus the level of satisfaction with particular movies might be seen as observations in the past. Based on past user behavior, the ratings will be used to identify new user-items associations. This approach of content filtering is termed *collaborative filtering* (CF).

For example, let us consider the opinions of the customers on movies. Table 1 shows the ratings. The rows correspond to the customers, while the columns to the items respectively. The jth entry of the ith row of the table contains the opinion/reflection of the ith customer regarding to the jth movie. For example, *customer2* sets the value on 16 scores regarding for *movie3*. There may

	movie1	movie2	movie3	movie4
customer1	16			
customer2		17	16	
customer3		15	10	
customer4				18
customer5		7	6	
⋮	⋮	⋮	⋮	⋮

Table 1. Table of the customers' evaluations.

be empty places in the table representing the missing data. We wish to predict the evaluation of a customer referring to a not seen movie yet. In other words, we have to complete the table as precisely as possible. Personalized recommendations are especially important in markets where the variety of the choices is large and the taste of the customer is also essential. Numerous recommender systems deliver automatically generated personalized recommendation to their customers. The recommendation system compares the user's ratings to other users' ratings and finds the most similar users and then recommends similar items to similar users.

The data contained in Table 1 are entries of a sparse matrix. Therefore we have to find the missing entries of a matrix.

1.2. Methods for finding the recommendations

There are several methods for determining of the missing values. One of them is the *k-nearest neighbors methods*. Its principle is as follows: If the ratings of two users are close enough to each other based on some criterion of similarity in the cases of the known values/entries then theirs ratings are close together or at least similar for the unknown values. Suppose that we have a missing number/value in a given row. At first we will find similar row or rows which have the same or similar numbers/values to the given row in the filled positions and then we will take the missing value from the similar row. Problem may arise if we did not find any row which was similar enough or the rows were very sparse. Besides, two or more similar rows found may result very different values for the missing value.

Many researchers recommend *matrix factorization* to compute the missing value, for example, Koren et al. in [4]. However, the paper contains only allusions to the problems about the factoring the user-item rating matrix. The high portion of missing values caused by the sparseness in the user-item rating matrix may cause difficulties.

Karimi et al. investigated the recommender systems and turned out that the matrix factorization is very suitable for applying active learning in the recommender systems [3].

Forbes and Zu described a simple, content-boosted matrix factorization algorithm for collaborative filtering [2]. The usefulness of the method was proved via experiments. Non-negative matrix factorization methods were published by Liu and Shen [6], and Lee and Sheung [5]. However, the difficulties caused by the missing entries of the matrix were not discussed in the papers.

As we showed in [1] the matrix factorization is generally solvable in the sense that the product of the factor matrices reproduces the known values of the original matrix. Because the matrix factorization needs a high number of operations so several factorizing algorithms do not try for precise reproducing of known entries of the original matrix. These algorithms aim to produce the approximate values of the known entries. Our considerations are valid for these algorithms, too.

The application of the matrix factorization based on the consideration that there is a relationship between the entries of the original matrix and this relationship can be described just by the matrix factorization. Remark that the multiplication of matrices is a special operation. There are various methods to derive a result matrix from different matrices. In paper [1] we demonstrate a few cases where the multiplication of matrices describes the real relationship.

2. Definitions of stability for matrix factorization

As we showed in paper [1] the factorization of matrices is not unique, moreover it may cause substantially inconsistent conclusions. It has been suggested in spite of the products of the factor matrices of the same matrix are different, but when the values of the observed entries have stability property in some sense, the result is regarded as trusty.

Let us denote the original matrix by $\mathcal{R}$, while the entry of $\mathcal{R}$ in the ith row and in the jth column is denoted by $\mathcal{R}_{i,j}$. Let $\mathcal{F}$ denote a concrete factorization. The product of the factor matrices yielded by factorization $\mathcal{F}$ of matrix $\mathcal{R}$ is called *result matrix*. The result matrix is denoted by $\mathcal{F}(\mathcal{R})$.

It can be shown for the matrix $\mathcal{R}$ there exist several number of factorizations $\mathcal{F}$, consequently there are several result matrices $\mathcal{F}(\mathcal{R})$ which are typically (but not definitely) different from each other.

We introduce different definitions for the stability of the matrix factorization.

- We can observe the stability of an entry which is unknown in the original matrix $\mathcal{R}$.
- We can observe the stability of $\mathcal{F}(\mathcal{R})$.
- We can define stability for perturbation of the original matrix $\mathcal{R}$. That is what differences are caused in matrix $\mathcal{F}(\mathcal{R})$ by the perturbation of known entries of $\mathcal{R}$.

Definition 2.1. *A fixed* ***entry*** $\mathcal{R}_{i,j}$ ***of matrix*** $\mathcal{R}$ ***is stable from the point of view of matrix factorization****, if the relative deviation of the entries* $\mathcal{F}(\mathcal{R})_{i,j}$ *is less than the threshold, that is, it can be given* $\epsilon_{i,j} > 0$ *such that*

$$\frac{\sigma(\mathcal{F}(\mathcal{R})_{i,j})}{|\overline{\mathcal{F}(\mathcal{R})_{i,j}}|} \leq \epsilon_{i,j},$$

where σ *is the deviation, the overline symbolizes the average of the entries* $\mathcal{F}(\mathcal{R})_{i,j}$. *The smallest suitable* $\epsilon_{i,j}$ *can be considered the* ***stability measure of element*** $\mathcal{R}_{i,j}$.

Definition 2.2. *We assume that* ***matrix*** $\mathcal{R}$ ***is stable from the point of view of matrix factorization****, if the stability measure of the unknown entries in matrix* $\mathcal{R}$ *is under a suitable threshold, that is there exists* $\epsilon > 0$ *such that*

$$\max_{\textit{unknown element of } \mathcal{R}} \epsilon_{i,j} \leq \epsilon.$$

*The smallest suitable ϵ can be considered as the **stability measure of matrix** $\mathcal{R}$.*

Definition 2.3. *We assume that **matrix $\mathcal{R}$ can be factorized stably from the point of view of perturbation**, if a small perturbation of the known entries of matrix $\mathcal{R}$ causes a small perturbation of the entries of matrix $\mathcal{F}(\mathcal{R})$. That is if for every $\epsilon > 0$ there is a $\delta > 0$ so that whenever*

$$|\mathcal{R}_{i,j} - \mathcal{R}'_{i,j}| \leq \delta \quad \textit{then} \quad |\mathcal{F}(\mathcal{R})_{i,j} - \mathcal{F}(\mathcal{R}')_{i,j}| \leq \epsilon$$

*for any known $\mathcal{R}_{i,j}$. In the above formula $\mathcal{R}'$ symbolizes the perturbation of $\mathcal{R}$. The smallest suitable ϵ can be considered as the **factorization stability measure of matrix $\mathcal{R}$ belonging to the perturbation** δ.*

Apparently we can give different other definitions for the stability and stability measurement.

3. The matrix factorization is unstable

Let $\{\mathcal{F}(\mathcal{R})\}$ denote the set of the result matrices $\mathcal{F}(\mathcal{R})$ produced by the process of factorization of matrix $\mathcal{R}$.

In the recommender systems the method is as follows: starting with matrix $\mathcal{R}$ – using the matrix factorization – a single matrix $\mathcal{F}(\mathcal{R})$ is produced and the values of the unknown entries of $\mathcal{R}$ are determined as the computed entries of $\mathcal{F}(\mathcal{R})$. The aim is to find the answer for a given question by applying a recommender system. For determining the answer for this question we use the entries of $\mathcal{F}(\mathcal{R})$. Fixing a concrete question, suppose that the set of answers is set $\mathcal{A}$. Matrix $\mathcal{R}$ has unknown entries. Let us define the *expansion* of matrix $\mathcal{R}$ as follows:

Definition 3.1. *The **expansion of matrix** $\mathcal{R}$ contains all known entries of $\mathcal{R}$ and some originally unknown entries of $\mathcal{R}$. Let $\mathcal{R}^+$ denote the expansion of $\mathcal{R}$.*

For any element a of the answer set $\mathcal{A}$ there exists an expansion of matrix $\mathcal{R}$ which results the answer a for our observed question. Let $\mathcal{R}^{+a}$ denote the expansion of $\mathcal{R}$ which results the answer $a \in \mathcal{A}$. In general for a given $\mathcal{R}$ and a given a there are infinitely many different expansion-matrices $\mathcal{R}^{+a}$.

Example 3.1. *The initial matrix belonging to the ratings shown in Table 1 is as follows*

$$\mathcal{R} = \begin{pmatrix} 16 & . & . & . \\ . & 17 & 16 & . \\ . & 15 & 10 & . \\ . & . & . & 18 \\ . & 7 & 6 & . \end{pmatrix}$$

There are several factorizations of matrix $\mathcal{R}$, three examples are as follows

$$\begin{pmatrix} 1 & 1 & 4 \\ 0 & 2 & 3 \\ 2 & 1 & 2 \\ 2 & 2 & 2 \\ \frac{1}{2} & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 & 3 \\ 0 & 1 & 2 & 3 \\ 4 & 5 & 4 & 3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{16} & 23 & 18 & 18 \\ 12 & \boldsymbol{17} & \boldsymbol{16} & 15 \\ 8 & \boldsymbol{15} & \boldsymbol{10} & 15 \\ 8 & 16 & 12 & \boldsymbol{18} \\ 4 & \boldsymbol{7} & \boldsymbol{6} & 7.5 \end{pmatrix}$$

$$\begin{pmatrix} 6 & 5 \\ 3 & 2 \\ 1 & 3 \\ 8 & 10 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 4 & 1 \\ 2 & 4 & 2 & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{16} & 38 & 34 & 11 \\ 7 & \boldsymbol{17} & \boldsymbol{16} & 5 \\ 7 & \boldsymbol{15} & \boldsymbol{10} & 4 \\ 28 & 64 & 52 & \boldsymbol{18} \\ 3 & \boldsymbol{7} & \boldsymbol{6} & 2 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 5 \\ 3 & 2 \\ 1 & 3 \\ 8 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 3 & 4 & 1 \\ 2 & 4 & 2 & 5 \end{pmatrix} = \begin{pmatrix} \boldsymbol{16} & 26 & 18 & 27 \\ 13 & \boldsymbol{17} & \boldsymbol{16} & 13 \\ 9 & \boldsymbol{15} & \boldsymbol{10} & 16 \\ 28 & 32 & 36 & \boldsymbol{18} \\ 5 & \boldsymbol{7} & \boldsymbol{6} & 6 \end{pmatrix}$$

When the observed question is "Should we suggest for customer5 the film4?" then the set of the possible answers is $\mathcal{A} = \{yes, no, not\_known\}$. Some expansions of $\mathcal{R}$ which support these answers:

$$\mathcal{R}^{+yes} = \begin{pmatrix} 16 & . & . & . \\ . & 17 & 16 & . \\ . & 15 & 10 & . \\ . & . & . & 18 \\ 4 & 7 & 6 & \boldsymbol{7.5} \end{pmatrix}$$

Because the observed entry of the matrix (line 5, column 4) is 7.5 and this is the maximum value of its line.

$$\mathcal{R}^{+no} = \begin{pmatrix} 16 & . & . & . \\ . & 17 & 16 & . \\ . & 15 & 10 & . \\ . & . & . & 18 \\ 3 & 7 & 6 & \boldsymbol{2} \end{pmatrix}$$

Because the observed entry of the matrix (line 5, column 4) is 2 and this is the minimum value of its line.

$$\mathcal{R}^{+not\_known} = \begin{pmatrix} 16 & . & . & . \\ . & 17 & 16 & . \\ . & 15 & 10 & . \\ . & . & . & 18 \\ 5 & 7 & 6 & \boldsymbol{6} \end{pmatrix}$$

Because the observed entry of the matrix (line 5, column 4) is 6 and this is neither maximum nor minimum value of its line.
As we can see the above three matrices $\mathcal{R}^{+yes}$, $\mathcal{R}^{+no}$, and $\mathcal{R}^{+not\_known}$ belong to the result matrices of the factorizations, hence depending on which factorization was found from the possible ones, each possible proposal can be supported, in spite of the fact that there exists only one right proposal.

It is evident that when we factorize any expansion of the original matrix then its result matrix is also the result matrix of the factorization of the original matrix. It would be summarized as follows: for any $\mathcal{R}$ and for any $a \in \mathcal{A}$

$$\{\mathcal{F}(\mathcal{R}^{+a})\} \subset \{\mathcal{F}(\mathcal{R})\},$$

and in general

$$\{\mathcal{F}(\mathcal{R}^{+})\} \subset \{\mathcal{F}(\mathcal{R})\}.$$

Because the number of elements of set $\{\mathcal{F}(\mathcal{R}^{+a})\}$ is infinite and $\{\mathcal{F}(\mathcal{R}^{+a})\} \subset \{\mathcal{F}(\mathcal{R})\}$ it means that during the factorization of the original matrix the procedures produce infinite number of those matrices which contain any value in the observed position. So we can state that **the matrix factorization is unstable**. It is true for all of the above given definitions of the stability.

Remark. In the datamining frequently applied method is that one part of the known facts is used for learning and the other part of known facts is used to test the model. To decide about the conformance of the result matrix produced by the matrix factorization this technique is unusable. In the case of the matrix factorization (when we take into account only a part of known data at processing the algorithm, the further part of data is used to check the correctness of results) this means the deletion of certain elements of the original matrix $\mathcal{R}$. It can be symbolized by $\mathcal{R}^-$. We observe the values of those entries of matrix $\mathcal{F}(\mathcal{R}^-)$ which were left out from $\mathcal{R}$. When the applied matrix factorization for these elements (which are used for control) proves to be proper, then this correspondence has no influence on the unknown elements because – as we showed above – the set $\mathcal{F}(\mathcal{R}^-)$ contains infinite number of matrices which could have any values in the observed position. Consequently from the fact that the applied factorization produces good values in known positions does not follow that the values in the unknown positions are good or false.

4. Conclusion

Although there are infinitely many different factorizations of a matrix, in practice during the processing of a concrete matrix we can produce only finite number of factorizations of course. These factorizations – specifically when we use the same factorizing algorithm – seem to be stable, so one may have the impression that from the result matrix one can derive reliable conclusions. However it is not true because – as we showed above – during the factorization a number of result matrices with essentially different properties would come off. Based only on careful mathematical conditions, there is no way to find out which factorization is "good" and which ones are "wrong" from the point of view of recommender systems. The faulty decision will obviously cause damage and problem moreover it will badly shake the confidence in the application.

So we can confirm the statement published in [1]: Application of matrix factorization in recommender systems would be dependably usable if we found rationale for factor-matrices and the multiplication of the factor-matrices. According to our opinion, the right explanation of the factor-matrices based on merely matrix $\mathcal{R}$ is impossible. In the case of some matrices, finding of the correct explanation is possible only if we knew the practical environment.

In general the matrix factorization is unstable. We think it is possible that there are special cases when the matrix factorization is stable. It is a difficult problem the existence of the stable cases and how can we determine it by the analysis of matrix $\mathcal{R}$.

References

[1] **Buza, A. and P.B. Kis,** Pitfall of the matrix factorization for recommender systems (submitted)

[2] **Forbes, P. and M. Zhu,** Content-boosted matrix factorization for recommender systems: Experiments with recipe recommendation, *RecSys'11*, 2011.

[3] **Karimi, R., C. Freudenthaler, A. Nanopoulos, and L. Schmidt-Thieme,** Comparing prediction models for active learning in recommender systems, `http://www.ismll.uni-hildesheim.de/pub/pdfs/karimi-gfkl.pdf`, 2010.

[4] **Koren, Y., R. Bell, and C. Volinsky,** Matrix factorization techniques for recommender systems, *IEEE, Computer*, (2009), 42-49.

[5] **Lee, D.D. and H.S. Seung,** Algorithms for non-negative matrix factorization, *Advances in Neural Information Processing Systems 13, Proc. of the 2000 Conference¡* MIT Press, 2000, 556-562.

[6] **Liu, Y. and C. Shen,** Orthogonal nonnegative matrix factorization for multi-type relational clustering, *International Journal of Computer and Information Technology,* **2** (2) (2013), 215-221.

[7] **Takács, G., I. Pilászy, B. Németh and D. Tikk,** Investigation of various matrix factorization methods for large recommender systems, *KDD netfix workshop*, 2008.

Antal Buza
College of Dunaújváros
Dunaújváros, Hungary
`antalbuza@yahoo.com`

Piroska B. Kis
College of Dunaújváros
Dunaújváros, Hungary
`pbkism@yahoo.com`