The strategy of user experience design driven by big data

Wenjuan Xu
Sanda University, Shanghai, China

Corresponding author and e-mail: Wenjuan Xu, wjxu@sandau.edu.cn

Abstract. In this paper, the algorithm of late factor model is used to calculate the satisfaction degree of the target user to the product to be tested by using the known conditions obtained from big data, so as to design the most ideal user experience. Taking the user experience of app as an example, some known user experience values are hidden, and then the calculated values are calculated by the algorithm. Through the comparison of the two, the minimum relative error between the predicted user experience and the real value is only 0.617%, and the maximum is no more than 2.957%.

1. Introduction
With the rapid development of China's economic level and science and technology, the Internet has gradually become an indispensable part of people's daily life, study and work. And the traces left by everyone in the Internet, the interwoven data network, is big data. In 2011, the 5V features of big data: volume, velocity, variety, value and veracity was proposed by IBM [1]. With these characteristics, big data has gradually attracted more and more attention.

One of the significant advantages of big data is its volume. According to the statistics of Cloud Tweaks [2], the whole world will generate 25 billion bytes of data every day. International Data Center(IDC)[3] also predicts that the global data volume will grow by more than 40% in the next 10 years, and the global data volume will reach 35.2 ZettaByte finally. In addition, big data is the real information generated by users, which reflects the needs, attributes, labels and other issues of real users. So the authenticity and diversity of big data provide are the most reliable and practical analysis materials. Inspired by the various advantages of big data, all companies have become data companies, eager to ride the wave of big data era. big data has become the most practical tool for many industries to analyze market Quotation and user experience(UX) design, and has influenced and reshaped the analysis pattern and design methods of countless industries.

UX design also has a reform in the context of the rapid development of big data. Based on a design concept of user satisfaction, UX design is user-oriented. An excellent UX design needs to conduct an all-round analysis of the user's own situation in multiple dimensions, and collect as much data as possible. At this point, big data comes in handy.

In the algorithms used to process big data, the most widely used is collaborative filtering algorithm at present. It can calculate the schemes that users may be interested in according to the people who have the same interest and experience. Collaborative filtering algorithm can be divided into model based recommendation algorithm and memory based recommendation algorithm. The former is more accurate and effective, but the algorithm is extremely complex and difficult to implement, while the latter is easy to operate, but it is limited by the problem of cold start. Ackley uses collaborative
filtering model-based recommendation to analyze the user experience design of 15690 users of an engineering drawing software [4]. The recommendation scheme obtained by his experimental model has a precision of 98.6%. But because of its complexity, the application range of collaborative filtering model-based recommendation algorithm is narrow, so only one-dimensional user experience design can be analyzed in practical application.

Based on the defects of traditional collaborative filtering algorithm, Salakhutdinov and others proposed to introduce deep learning algorithm into collaborative filtering algorithm, and explained the first two-layer Restricted Boltzmann Machine (RBM) at the same time[5]. However, in the case of big amount of data, the output of RBM model will be relatively broad, which can not accurately locate the needs of users. Sejnowski collected 204386 music listeners' preferences for 4040586 songs in 266 music apps through the algorithm of double-layer restricted Boltzmann machine, and designed a music software recommendation system [6]. When the amount of data is less than 100000, the accuracy rate of the proposed system is 95.6%, but when the amount of data is more than one million, the accuracy rate of the output results is only 45.2%. The data lost its reference value.

Released by China Internet Information Center (CNNIC)[7] the 45th statistical report on the development of Internet in China pointed out that: As of March 2020, the scale of non-internet users in China has reached 496 million, of which 40.2% in urban areas and 59.8% in rural areas. Therefore, non-internet users are still dominated by rural people. It also shows clearly that the lack of skills, limited education and different ages are the main reasons why non internet users is still offline. Therefore, in the ocean of big data, the relevant information of target users may be lost. Therefore, how to grasp the algorithm analysis and design to solve the problem of data omission is also a pivotal step in UX design.

Summarizing the previous problems, it can find that the collaborative filtering algorithm is highly dependent on historical data. However, in general data systems, the preference data is often sparse, so it is necessary to reduce the dimension of the original data. Therefore, this paper will use the LFM to study how to use big data analysis to acquire a better user experience design scheme.

2. Principle

LFM is a method of dimension reduction analysis through matrix factorization. The decomposed matrix represents the hidden features of users and products. In this paper, the user experience satisfaction is used to represent the advantages and disadvantages of UX design, and the best UX design is obtained through LFM algorithm and big data. Assuming that the user experience satisfaction matrix is R, there are m users and n products. In order to find k hidden classes, it is necessary to find two matrices P and Q, so that the product of these two matrices is approximately equal to R, that is, the user experience satisfaction matrix R is decomposed into two low dimensional matrices and multiplied by each other.

\[ R_{m \times k} = P_{m \times k}^T \cdot Q_{k \times n} \approx R \]  

As can be seen from Figure 1, the user experience satisfaction matrix R is decomposed into user characteristic matrix P and product feature matrix Q. In R matrix, \( a_{ij} \) represents the user's satisfaction with the product; \( x_{ij} \) in the P matrix represents the user's attention to the feature class; and in the Q matrix, \( y_{ij} \) represents the weight of the product in the feature class.
Figure 1. User experience satisfaction matrix R is decomposed into user characteristic matrix P and product feature matrix Q.

Conversely, the satisfaction matrix of user experience R is obtained by multiplying the user characteristic matrix P by the product feature matrix Q. Among them, the elements in the user experience satisfaction matrix R are the dot product of the relevant eigenvectors in the matrix P and the matrix Q, which reflects the degree of correspondence between the user characteristics and the product features. Therefore, the bigger the $a_{ij}$ number in R, the greater the user satisfaction with the items. The hidden factors that affect users' satisfaction with products can be mined out, such as product popularity, product sales or downloads, price, beauty and other factors. By finding these hidden factors, users and products can be associated, and then the predicted satisfaction of user experience can be determined according to the correlation between the user experience to be analyzed and these hidden factors.

For the products that have been experienced by users, there will be corresponding satisfaction, but for the products that have not been used or to be launched, the user satisfaction is unknown. Therefore, most items of the user experience satisfaction matrix are empty, which is a sparse matrix. At this point, it is necessary to infer the user's satisfaction with the products that have not been used according to the user's satisfaction with the existing products.

A scoring matrix R of m x n can be approximated by the product matrix $R'$ of two small matrices $P_{mxk}$ and $Q_{kxn}$.

$$\mathbf{R}_{ui} = \mathbf{P}_u^T \cdot \mathbf{Q}_i = \sum_{k=1}^{K} \mathbf{P}_{uk} \cdot \mathbf{Q}_{ki}$$

(2)

It is found that the product $R'$ of $P_{mxk}$ and $Q_{kxn}$ is no longer sparse, and the items not previously in $R'$ can also be calculated by the product of $P$ and $Q$. Thus, a predictive user satisfaction matrix is obtained. As shown in Figure 2, if the values of the predicted user satisfaction matrix $R'$ and the
original user satisfaction matrix R are approximate at known positions, then their values are also considered to be approximate in predictive position.

![Figure 2. Known user satisfaction matrix R and prediction matrix R'.](image)

The predicted score matrix R' obtained by matrix decomposition may have errors in the known scoring terms with the original scoring matrix R. In order to eliminate the error, the square loss function is selected and the regularization term is added to prevent over fitting.

\[
C = \sum_{(u,i) \in R_0} \left( R_{ui} - R_{ui}' \right)^2 + \text{Reg} = \sum_{(u,i) \in R_0} \left( R_{ui} - P_u^T \cdot Q_i \right)^2 + \lambda \sum_u \| P_u \|^2 + \lambda \sum_i \| Q_i \|^2
\]  

(3)

\[
\lambda \sum_u \| P_u \|^2 + \lambda \sum_i \| Q_i \|^2
\]

Among them, \( \lambda \) is a regularization term, \( \lambda \) is obtained by cross validation generally.

Loss function:

\[
L(P, Q) = \sum_{(u,i) \in R_0} \left( R_{ui} - P_u^T \cdot Q_i \right)^2 + \lambda \sum_u \| P_u \|^2 + \lambda \sum_i \| Q_i \|^2
\]

(4)

For each \( P_u \), the partial derivative is obtained as follows:

\[
\frac{\partial L}{\partial P_u} = \frac{\partial}{\partial P_u} \left[ \sum_{u,j} \left( R_{u,j} - P_u^T \cdot Q_j \right)^2 + \lambda \| P_u \|^2 \right] = \sum_i 2 \left( P_u^T \cdot Q_i - R_{ui} \right) Q_i + 2 \lambda P_u
\]

(5)

The modified solution of user characteristic matrix P is obtained by gradient descent iteration:

\[
P_u = P_u - \alpha \cdot \frac{\partial L}{\partial P_u} = P_u - \alpha \cdot \left[ \sum_{i} 2 \left( P_u^T \cdot Q_i - R_{ui} \right) Q_i + 2 \lambda P_u \right]
\]

(6)

Similarly, the solution of the modified product characteristic matrix Q is obtained:

\[
Q_i = Q_i - \alpha \cdot \frac{\partial L}{\partial Q_i} = Q_i - \alpha \left[ \sum_u 2 \left( P_u^T \cdot Q_i - R_{ui} \right) P_u + 2 \lambda Q_i \right]
\]

(7)

3. Experimental design

In this paper, the app score data set to carry out the experiment. The satisfaction matrix R of the app is decomposed into user characteristic matrix P and app feature matrix Q. The features of APP valued by users and the weight factors of APP in this feature are popularity, price, beauty, practicability and memory. Assign values to these factors. The bigger the value, the higher the degree of representation.
The assignment of characteristic attributes is shown in table 1.

**Table 1.** The features of APP valued by users and the weight factor of APP in this feature.

| assignment | popularity | price | beauty | practicability | memory | satisfaction |
|------------|------------|-------|--------|----------------|--------|--------------|
| 1          | widely known | Expensive | Wrong | useless | big | dissatisfied |
| 2          | famous | ordinary | ordinary | Low | Relatively big | ordinary |
| 3          | relatively unknown | Low-priced | Good | ordinary | ordinary | satisfied |
| 4          | not in vogue | Free | Excellent | High | Small | very satisfied |

Ten common users of five apps were selected and their satisfaction with the app was investigated. The known R matrix, user characteristic matrix P and app feature matrix Q were obtained as shown in Figure 3.

![Figure 3](image)

**Figure 3.** The known R matrix, user characteristic matrix P and app feature matrix Q.

By erasing some data of R matrix randomly, a rare matrix R1 is obtained, as shown in Table 3. Then, the complementary R' matrix is obtained by LFM, as shown in Table 4.

**Table 2.** The rare matrix R1 obtained by randomly erasing some data for real matrix R.

| 3   | 1   | 0   | 0   | 1   |
|-----|-----|-----|-----|-----|
| 0   | 2   | 2   | 3   | 4   |
| 3   | 0   | 2   | 0   | 1   |
| 1   | 4   | 2   | 3   | 1   |
| 2   | 1   | 4   | 0   | 0   |
| 0   | 1   | 4   | 0   | 3   |
| 3   | 2   | 0   | 2   | 1   |
| 2   | 4   | 4   | 0   | 3   |
| 2   | 1   | 2   | 0   | 2   |
| 2   | 0   | 2   | 0   | 4   |
Table 3. The complement matrix R obtained by LFM algorithm.

|        |        |        |        |        |
|--------|--------|--------|--------|--------|
| 2.9837003 | 0.99058921 | 3.12897163 | 1.22010161 | 1.01225612 |
| 1.17877794 | 1.96375517 | 2.02879585 | 3.01795811 | 3.95788238 |
| 2.92549222 | 0.94397232 | 2.07529467 | 1.46335648 | 0.98091455 |
| 0.9459312  | 3.96893731 | 2.07046355 | 2.94976199 | 1.02894816 |
| 2.22881019 | 1.08990179 | 3.74960771 | 3.07263729 | 2.3798948  |
| 3.2171912  | 1.01296121 | 3.97610682 | 2.99611181 | 3.02102659 |
| 2.9616376  | 1.96203842 | 3.93448692 | 2.04299671 | 0.98352403 |
| 2.07750894 | 4.02781819 | 3.90446004 | 3.63911314 | 3.00115344 |
| 1.74161394 | 0.90501043 | 2.28828467 | 2.10455294 | 1.96774298 |
| 1.96751484 | 1.66735931 | 2.02255954 | 3.47437167 | 3.99126694 |

4. Discussion and analysis

By comparing the R obtained by LFM method with the original matrix R of users' satisfaction with the app, it can be found that this algorithm has high accuracy in predicting scarce data. In Figure 4, according to the cluster bar chart of 10 users, it can be seen that the satisfaction degree of the actual user experience is similar to the calculated value.
This paper mainly studies the design strategy of user experience driven by big data, which can be used as a reference for various elements when developing new products in various industries. Here, we can compare the calculated value of each app with the real value, and analyze the average absolute relative deviation according to the calculation results. The results are shown in table 4. The results show that the prediction result of app3 is good, and the relative error is only 0.617%. The prediction results of other apps are similar, and the highest relative error is no more than 2.957%.

**Table 4.** Comparison of actual and calculated user satisfaction of five apps.

| APP  | The average of actual user experience satisfaction | The average of calculated user experience satisfaction | relative error |
|------|-----------------------------------------------|-----------------------------------------------|----------------|
| APP1 | 2.2                                           | 2.222817                                       | 1.037%         |
| APP2 | 1.8                                           | 1.853234                                       | 2.957%         |
| APP3 | 2.9                                           | 2.917903                                       | 0.617%         |
| APP4 | 2.4                                           | 2.598096                                       | 8.254%         |
| APP5 | 2.2                                           | 2.232461                                       | 1.476%         |
From table 4 and figure 5, it can be found that the complementation value of R 'matrix derived from the algorithm is basically consistent with the original R matrix, indicating that this algorithm is suitable for predicting the target users and UX satisfaction of unsold and unlisted products.

It can be seen that the algorithm has a certain reference value for the analysis of UX design, but there are still some gaps. For example, the average satisfaction of the calculated value of app4 is about 2.598, but the actual satisfaction is 2.4. The difference between the calculated value and the actual value is quite big. Through the analysis, it can be concluded that the difference may be caused by the following reasons:

1. The implied factors are not comprehensive enough. In this paper, only five factors are used to realize the association between users and products. There are many objective factors that affect UX. It is unrealistic to make P and Q matrices for all factors that may affect user satisfaction. We can only use big data to conduct a lot of analysis to approximate the real value.

2. In the prediction, the parameter assignment is not accurate enough. For each feature assignment, the more detailed the better, it is convenient for the algorithm to give more accurate prediction value. The calculated value of app2 of user 10 in Fig. 5 is 1.667, which is the most different from the actual value 1, which may be caused by inaccurate parameter allocation.

For the research strategy of using big data to design UX, the following suggestions can be summarized:

1. As much as possible to analyze the target customers' attention to various product factors, and put more energy and capital into the factors that users attach importance to, in order to obtain the products with high customer satisfaction and make the target customers get better UX.

2. In order to make the algorithm universal, manufacturers can survey some more personalized users to obtain some relatively rare data. For example, Li Guangming and Fang Jingli [8], in the design of the movie recommendation system under the spark platform, scored about 20 million pieces of 30106 different kinds of films from 138 491 users, screened out the audiences whose scores were distributed among others, and investigated and analyzed them, so as to fill in the missing algorithm statistical materials.

5. Conclusions
This paper analyzes the problem of big data driven UX design by implicit factor dimensionality reduction method. Taking app products as an example, the theoretical value of vacancy data is calculated by randomly removing the known user data, and the result is almost the same as the practical value.

It shows that the algorithm can accurately estimate the value of incomplete data through a big amount of data, which is of great significance to the strategy research of UX design.
References

[1] IBM. The Four V’s of Big Data[EB/OL]. Big data analysis, 2019(12):11-13.
[2] CloudTweaks. The statistics of big data. Big data, 2019(10):34-36.
[3] IDC. The increments of global data quantity. Big data statistics, 2020(05):13-20.
[4] Ackley. Collaborative filtering of engineering drawing software based on model recommendation. Collaborative recommendation, 2017(04):42-45.
[5] Salakhutdinov, R, MNICH A, HINTON G. Restricted Boltzmann Machines for Collaborative Filtering[C]. Proceedings of the 24th International Conference on Machine Learning, 2007(09):791-801.
[6] Sejnowski. Analysis of music recommendation based on Algorithm of bilevel constrained Boltzmann machine. Boltzmann machine, 2018(09):10-17.
[7] China Internet Information Center (CNNIC). Statistical report on the development of Internet in China. Statistical report, 2020(04):13-20.
[8] Li Guangming Fang Jingli. Design of movie recommendation system based on spark platform. Recommendation system, 2019(7):36-46.