Implicit Feedback Recommender System Based on Matrix Factorization

Na Jiang[1]

Zhaotong University, China
`27805044@qq.com`

Abstract. With the development of the internet age, information overload problem is imminent. At now, almost of recommended models use the explicit feedback. But lots of implicit feedback data are missing. The paper explores the area of recommendation based on large-scale implicit feedback, Where only positive feedback is available. Further, the paper carried on the empirical research on the Implicit Feedback Recommendation Model. By maximized the probability of the user's choices, IFR mean the progress task into optimization problems In the way, the experiment results confirm the superiority of the model. However, the model is insufficient about online research and a lack of details.

Key words: Recommendation system, matrix decomposition, implicit feedback.

1 Introductions

With the evolution of the times the pace continues to accelerate, the amount of information overload has become a problem can not be ignored in order to avoid the user to spend a lot of time browsing useless information, recommendation system came into being. The recommendation system recommends the user's interest and merchandise to the user by observing the user's interest characteristics and selection behavior. In recent years, the research on the recommendation system has been deepening, and the huge data has put forward new challenges to the recommendation system.

Core resources to support the recommendation system is the user's historical behavior data, including explicit feedback and implicit feedback. At present the classic explicit feedback systems such as collaborative filtering (CF) [1], most of the explicit feedback based collaborative filtering recommendation system is based on user ratings and trust information to improve the accuracy of recommendation [2]. But this will miss a lot of implicit feedback data. Implicit feedback data is more common than additional inputs required for explicit feedback data, and its collection costs are low and do not affect the user experience. Aiming at the shortcomings of the explicit feedback data recommendation system, a recommendation system based on implicit feedback data has appeared. However, implicit feedback data can express the user's positive feedback, but less able

to express the user's negative feedback, so how to solve the problem of lack of negative feedback is very important.

At present, the recommendation system based on implicit feedback data is mainly faced with the following challenges:

– Imbalance. In implicit feedback data, there is usually only positive feedback and lack of negative feedback. Not as explicit feedback data can directly reflect the tendency of the user's likes and dislikes, implicit data includes the "selected" and "unselected" categories. Although "selected" may indicate a positive tendency, "unselected" can not directly represent a user negative tendency. Because "unselected" not only includes the user does not really interested in the product, but also includes the user interested but not yet found the product. This is only the lack of positive examples set a negative example, to set the model adds difficulty.
– Noise. Explicit rating opposite, in the implicit feedback data, the user can produce a lot of noise for various misuse.
– Large scale. In the actual scene of the recommended model, we will face large-scale data, which requires the model has sufficient efficient performance and excellent scalability, in order to be able to handle vast amounts of data.

In this paper, the model proposed in the above problem is used to model the implicit feedback data directly by converting the recommended task into the probability of maximizing the probability of user selection behavior. This will be able to "unselected" the information fully utilized, but also to avoid the introduction of negative cases and noise, so as to ensure the improvement of recommended quality.

2 Related Work

2.1 Negative Feedback

Traditional recommendation systems are difficult to apply directly to implicit feedback data because the data only contains positive feedback from the user and lacks negative feedback. In [3] Pan et al. novelly propose One Class Collaborative Filtering (OCCF), which is generally summarized as Unbalanced Class Problem (UCP) [4]. The main way to deal with this problem is the introduction of negative feedback, and the introduction of negative feedback in the following ways:

– The particularity of the application environment, and addition of a negative artificial rules [5]. For example, be forwarded in the other two are not forwarded between microblogging can be considered negative samples, as users browse them but there is no forwarding, which may indicate a lack of interest, but there are users microblogging attention but It is not forwarded, and this method relies on domain knowledge, can not be extended.
– Label from an unknown sample as negative samples in a random sample [6]. This method generally assumes that most of the unknown label samples are negative, so most of the randomly sampled samples are negative.

– The samples are unknown label as negative samples, but set a lower weight, the weight of the sample is negative reflecting the confidence level of the sample [7]. But the disadvantage of this approach is that it can not guarantee "true".

In fact, these three methods are trying to add negative samples in the experiment but can not guarantee the authenticity of negative cases. And the introduction of negative samples will increase the burden of training, larger scale also affected the efficiency of the experiment. The model proposed in this paper is based on the idea of probability generation model, by maximizing the observed probability of user feedback to directly select the user to choose the trend of modeling, no negative feedback can be trained, and therefore applicable to a variety of implicit feedback recommendations Scenes.

2.2 Collaborative Filtering

With the development of the times, matrix decomposition has gradually replaced the traditional user-based or product collaborative filtering algorithm, has become a research system in the field of research hot spots. Neighbor-based collaborative filtering algorithms need to calculate the user-user or product-product similarity, and if large amounts of data are encountered, a large amount of computation is required, which means a lot of overhead. Probabilistic Matrix Factorization (PMF) [8] is mapped to the product or the user low-dimensional space, the feature vectors to estimate the potential user or by fitting product rating information, in order to reconstruct the scoring matrix. The use of dimensionality reduction technology to deal with high-dimensional sparse data, fully embodies the advantages of matrix decomposition.

However, matrix decomposition does not apply to implicit feedback data, since implicit feedback data are not explicitly given to the probability matrix decomposition model fit. In order to apply to the matrix decomposition implicit feedback stream, 0-1 matrix decomposition [9] have been proposed, but the introduction of such a matrix decomposition as negative cases, the addition of unwanted noise, affecting the recommendation result. In [10], He et al. proposed a Bayesian-based personalized ranking, by maximizing the positive example came in front of the probability of a negative example to enhance the effect is recommended. But this method can not avoid the introduction of negative cases, the noise on the recommended effect of a great impact.

2.3 Optimization algorithm of matrix decomposition

In recent years, the field of recommendation systems mentioned more matrix decomposition model, the idea is to score the user's behavior with a scoring matrix to R, where R_{ij} is the user i rating to the product j, but the user can not for all products So many of the elements in this score matrix are empty.Matrix decomposition model by the method of dimension reduction of the scoring matrix completion, and the traditional Singular Value Decomposition (SVD) [11]

inconvenience to handle a large number of missing items, in order to solve this problem, we have the following optimization algorithm:

– Alternative Least Squares (ALS): For matrix $R_{m \times n}$, ALS aimed at finding two low-dimensional matrix $X_{m \times k}$ and $Y_{n \times k}$ to approximate $R_{m \times n}$, can be expressed as:

$$R_{m \times n} \approx X_{m \times k} Y_{n \times k}^T \tag{1}$$

Wherein $R_{m \times n}$ represents the user of the product scoring matrix, $X_{m \times k}$ represents the user's preference matrices of potential features, $Y_{n \times k}$ represents a matrix comprising the product of the potential features. Through alternately fixing a matrix (such as X), another matrix optimization (such as Y), the algorithm can ensure that the final convergence. Scalability algorithm can be improved by ALS, while in parallelism to optimize traditional matrix decomposition algorithm.

– Stochastic Gradient Descent (SGD): is a common method for minimizing the risk function, loss of function, and is effective in solving SGD matrix decomposition optimization [12], and is suitable for large-scale computing problems, so as to Optimization Algorithm of Matrix Decomposition. For the matrix $R_{m \times n}$, each random from the SGD finding a pair (i, j) and updates the corresponding U_i and V_j, respectively, which is a method for updating U_i and V_j partial derivative as a gradient direction, Using the iterative strategy, a certain step is advanced along the negative direction of the current point along the objective function until it moves to the minimum point. Step length control by the learning rate α.

There is also another problem, when a user-item rating matrix is very sparse and they will appear over-fitting (overfitting) problem. A solution to this problem is to regularization (regularization). Regularization is the addition of the user factor vector and the product factor vector to the objective function.The introduction of regularization factor can solve the problem of over-fitting [13].

3 System Model

On the basis of the algorithm mentioned above, this paper proposes an recommender based on implicit feedback (FRI). It assumes that the user's choice behavior is determined by the user's "selectivity" of the product.

it assumes that the user i and j in potential product K-dimensional feature space is represented as a set of vectors $U_i = (U_{i1}, U_{i2}, \ldots, U_{ik})$, $V_j = (V_{j1}, V_{j2}, \ldots, V_{jk})$, user and product features collaborative determine the behavior of the user's choice, so that A_{ij} represents the preference of product j from user i, which is presented in Equation 2.

$$A_{ij} = \sum_{k=1}^{K} U_{ik} V_{jk} \tag{2}$$

But the absolute tendency itself does not fully reflect the user's choice of behavior, so to compare to make sense. Assuming that the user i's preference for product j is 20, user i is likely not to select the product if the user i's preference for most products is higher than this value (the average selection is 30) If user i is less likely to choose the majority of products than this value (the average choice is 10), user i is likely to choose this product. Suppose M is the number of products, so that δ_{ij} represents the user i to select the relative degree of tendency product j, δ_{ij} represents the larger the more the user i may select a product j, and vice versa δ_{ij} closer to 0 indicates that the user i can not select the product j. The relative resistance can be defined in Equation 3.

$$\delta_{ij} = \frac{A_{ij}}{\overline{A_i}} = \frac{A_{ij}}{\frac{1}{M}\sum_{h=1}^{M} A_{ih}} \tag{3}$$

Wherein $\overline{A_i}$ represents user i's average preference of M products. Furthermore, user selection probability Pr_{ij} is determined by δ_{ij}. To make the data easier processing, sigmoid function is used to normalize δ_{ij} to (0,1) as shown in Equation 4.

$$Pr_{ij} = \frac{\delta_{ij}}{1 + \delta_{ij}} \tag{4}$$

The objective of the model is given by set O. For maximizing the posterior probability of the potential features matrix U and V, which can be described in Equation 5.

$$P(U,V|O) \propto \prod_{(i,j)\in O} \frac{1}{1 + \delta_{ij}^{-1}} \prod_{i,k} N(U_{ik}|0,\sigma^2) \prod_{j,k} N(V_{jk}|0,\sigma^2) \tag{5}$$

As shown in Equation 6, the optimization target equivalent is available by logarithmic inversion of Equation 5.

$$argmin_{U,V} L := \sum_{(i,j)\in O} \ln(1 + \delta_{ij}^{-1}) + \lambda(\|U\|_F^2 + \|V\|_F^2) \tag{6}$$

Wherein $\|U\|_F^2$ represents 2-norm, while $\lambda = \frac{1}{2\sigma^2}$ is adjusted parameter for controlling parameter complexity to prevent over-fitting.

Thus, based on the recommended model into model implicit feedback data for an optimization problem, wherein the matrix U,V is obtained according to the potential given set O, when the prediction of the user i and a given candidate product set $j_1, j_2, \ldots, j_k$, according to Equation 2, the recommendation list can be generated by A_{ij}.

4 Experiment

4.1 Experimental Dataset

The experimental data is from Douban, which is a leading community site that offers audio book recommendation, city activities, group exchanges under the topic line features a variety of services. This experiment will be regarded as a song book or product, all the information collected to obtain implicit feedback stream sort by time. 383 which grab the user's information from watercress online, including the historical behavior of their selection of books and music, the two products. Select the behavior of several books have $373, 648, 346, 242$ are selected musical acts. And extracted from the tab feature information products, respectively, song information (including artist information, song type), book information (including the type of book, author information) as a product feature.

4.2 Experimental Design

Comprehensive evaluation experiment using the F-Measure and mean average precision (MAP) as evaluation index [14]. Based on these evaluation indexes, the proposed FIR is compared with the following two algorithms.

1. User-based CF: The algorithm first by finding the target user set of users with similar interests, then find this collection a user-friendly, and the target user never heard of recommended items to the target user [15].
2. Item-based CF: The algorithm to calculate the similarity between the items, the user generates a recommendation list based on the similarity and the user behavior history items [16].

4.3 Experimental results and analysis

First, it should verify the effectiveness of FRI. As shown in Table 1, FRI advantage in the MAP evaluation index is relatively large, indicating that FRI improve the recommendation accuracy rate in top-K on outstanding performance. On the F-Measure evaluation index, FRI than the other two algorithms is also an advantage, proved FRI indeed more secure on the accuracy of the recommended list.

Table 1. Performance comparison.

Algorithm	MAP	F-Measure (%)
user-based CF	0.131	55.8
item-based CF	0.135	56.08
FRI	0.145	70.35

Furthermore, the affection of K (i.e. the dimension of potential feature) and λ (i.e. regularization) should be discussed. As shown in Fig. 1, the MAP and

F-Measure is not affected by K considerably, wherein the feature dimension is between 10 and 20, and $\lambda = 10^{-5}$.
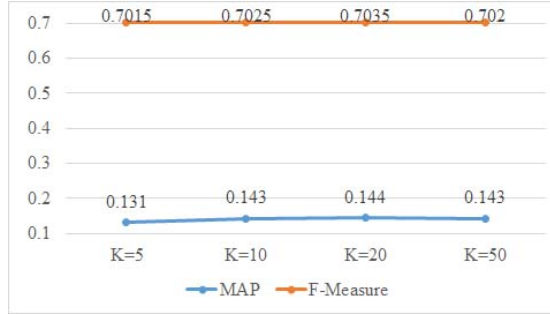

Fig. 1. K affection on MAP and F-Measure

As shown in Fig., the regularization parameter has a little effect on MAP and F-Measure, but still non-ignorable, wherein $K = 10$. When $\lambda = 0$, the data potential features in data matrix is still small. This illustrates and compares traditional matrix decomposition, FRI in terms of preventing over-fitting advantage [20] .This is because the model optimizes the relative proportions between the selected product and the general product, rather than fitting a value. The optimal regularization parameter is 10^{-3} $10-5$.
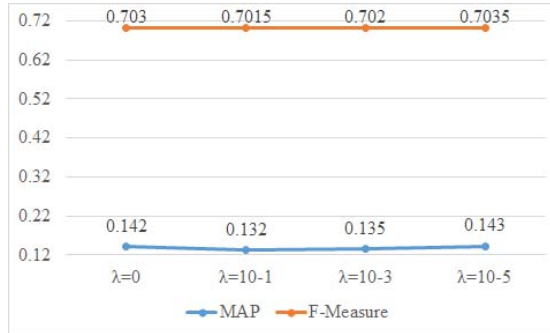

Fig. 2. λ affection on MAP and F-Measure

5 Conclusions

In this paper, the empirical model based on implicit feedback data is studied empirically. Firstly, the implicit feedback data stream is explained, and then the potential feature implicit feedback recommendation model based on matrix

decomposition is given. The probability of the proposed problem is modeled as the optimization problem by maximizing the probability of the user's selection behavior. The model is verified by the relevant data set, and the superiority of the model is verified by comparison with other recommended models. The lack of this paper is that it does not implement its online recommendation, and this model only applies to the implicit feedback data, in fact, implicit feedback data and display feedback data in reality may exist at the same time, then how to integrate these two Resources to be more accurate recommendations to be further studied. Also in contrast to the experiment, the contrast algorithm selection is still inadequate, more contrast the new algorithm can be able to reflect the performance of FRI. In the future, I will be applied according to the characteristics of online recommendation to our model, the model for online recommendation, and I will try to add a network into reality recommendation scene, there are still many factors that influence the choice of users to achieve more Refine modeling.

References

[1] B. Yang, Y. Lei, J. Liu, and W. Li, "Social collaborative filtering by trust," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 8, pp. 1633–1647, 2017.

[2] R. He and J. McAuley, "Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering," in *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016, pp. 507–517.

[3] R. Pan, Y. Zhou, B. Cao, N. N. Liu, R. Lukose, M. Scholz, and Q. Yang, "One-class collaborative filtering," in *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. IEEE, 2008, pp. 502–511.

[4] L. Duan, M. Xie, T. Bai, and J. Wang, "A new support vector data description method for machinery fault diagnosis with unbalanced datasets," *Expert Systems with Applications*, vol. 64, pp. 239–246, 2016.

[5] Y. Zhang, M. Chen, D. Huang, D. Wu, and Y. Li, "idoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization," *Future Generation Computer Systems*, vol. 66, pp. 30–35, 2017.

[6] Y. Zhang, "Grorec: a group-centric intelligent recommender system integrating social, mobile and big data technologies," *IEEE Transactions on Services Computing*, vol. 9, no. 5, pp. 786–795, 2016.

[7] Y. Zhang, D. Zhang, M. M. Hassan, A. Alamri, and L. Peng, "Cadre: Cloud-assisted drug recommendation service for online pharmacies," *Mobile Networks and Applications*, vol. 20, no. 3, pp. 348–355, 2015.

[8] M. Fraccaro, U. Paquet, and O. Winther, "Indexable probabilistic matrix factorization for maximum inner product search." in *AAAI*, 2016, pp. 1554–1560.

[9] X. He, H. Zhang, M.-Y. Kan, and T.-S. Chua, "Fast matrix factorization for online recommendation with implicit feedback," in *Proceedings of the*

39th International ACM SIGIR conference on Research and Development in Information Retrieval.   ACM, 2016, pp. 549–558.

[10] R. He and J. McAuley, "Vbpr: Visual bayesian personalized ranking from implicit feedback." in *AAAI*, 2016, pp. 144–150.

[11] Y. Zhang, M. Chen, S. Mao, L. Hu, and V. C. Leung, "Cap: Community activity prediction based on big data analysis," *IEEE Network*, vol. 28, no. 4, pp. 52–57, 2014.

[12] C. De Sa, M. Feldman, C. Ré, and K. Olukotun, "Understanding and optimizing asynchronous low-precision stochastic gradient descent," in *Proceedings of the 44th Annual International Symposium on Computer Architecture.*   ACM, 2017, pp. 561–574.

[13] Y. Zhang, Z. Tu, and Q. Wang, "Temporec: Temporal-topic based recommender for social network services," *Mobile Networks and Applications*, Apr 2017. [Online]. Available: https://doi.org/10.1007/s11036-017-0864-3

[14] J. Song, Y. Zhang, K. Duan, M. S. Hossain, and S. M. M. Rahman, "Tola: Topic-oriented learning assistance based on cyber-physical system and big data," *Future Generation Computer Systems*, vol. 75, pp. 200–205, 2017.

[15] C. M. Rodrigues, S. Rathi, and G. Patil, "An efficient system using item & user-based cf techniques to improve recommendation," in *Next Generation Computing Technologies (NGCT), 2016 2nd International Conference on.* IEEE, 2016, pp. 569–574.

[16] D. Li, C. Chen, Q. Lv, L. Shang, Y. Zhao, T. Lu, and N. Gu, "An algorithm for efficient privacy-preserving item-based collaborative filtering," *Future Generation Computer Systems*, vol. 55, pp. 311–320, 2016.