Research on Personalized Recommendation Based on Deeply Restricted Boltzmann Machine

Guodong Wu [1, 2], Lijing Tu [2], Jingxia Li [2], Fugen Song [1]

[1]Glorious Sun School of Business and Management, Donghua University, Shanghai 200051, China

[2]School of Information and Computer, Anhui Agricultural University, Anhui Hefei 230036, China

Abstract

With the arrival of big data age, the research and application of recommendation system are becoming more and more extensive in order to solve the problem of information overload. Traditional methods such as collaborative filtering recommendations, content-based recommendations, etc., they are based on some shallow linear models; the recommendation results are not very good. Deep learning technology process mass data with its multi-layer network, and automatically extract the characteristics to make learning advantages, provide a more extensive space for the recommendation system. This paper based on the research on Restricted Boltzmann Machine (RBM), and discusses the personalized recommendation problem under (Deep RBM, DBM) model which includes three-layer network structure with a visual layer and two hidden layers.). Experiments on the open data set show that the DBM model based on the three-layer network structure has better recommendation performance than limited Boltzmann machine RBM with two-layer network structure and traditional recommendation method.

Keywords

Restricted Boltzmann Machine, deep learning, Deeply Restricted Boltzmann Machine, Personalized recommendation.

1. Introduction

With the rapid development of the Internet, information on the Internet is more and more, how consumers find their own interested goods in the complex information under the condition of information overload, how merchant quickly display their goods in front of consumers, which is a very important issue. The past search engine can only be applied to situation that consumers have a clear purpose and needs, for some consumers with not clear purpose and the potential needs, how to find their interest, thus recommending goods that they may be interested, The recommendation system emerges as the times require. The recommendation system is widely used in various Web-based services, such as e-commerce (Amazon.com), social networking (Facebook.com), video services (Youtube.com), also achieved some success [1]. Traditional recommendation systems such as collaborative filtering are based on some linear shallow models, whose recommendation performance is not good, and are subject to cold start, sparseness and scalability of data. In recent years, the deep learning technology process massive amounts of data with its multi-layer network, and automatically extract the characteristics to make learning advantages of [2] [3], research and application in the field of personalized recommendation are more and more attract people's attention [4][5]. This paper based on the traditional recommendation method and the limited Boltzmann machine technology, studies and discusses personalized recommendation issues under deep-limited Boltzmann machine (Deep RBM, DBM) model which contains three-layer network structure with one visual layer and two hidden layer. The relevant experimental results show that the DBM with three-layer network structure has better performance compared with the traditional recommendation method and the two-layer structure Boltzmann machine RBM.

2. Related Knowledge

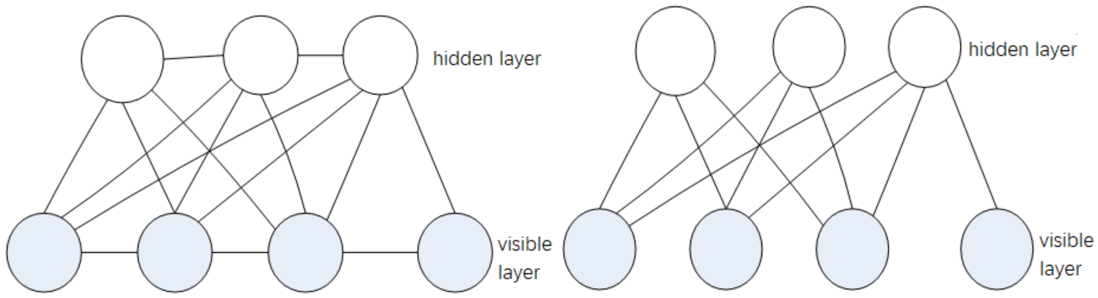
2.1 Personalized Recommendation

The research of personalized recommendation started late until the early nineties of the twentieth century, it is presented as an independent concept[6]. The concept and definition widely accepted recommendation system is given by Resnick and Varian in 1997[7]: "It is to use e-commerce sites to provide customers with product information and advice to help users decide what products should be purchased, and simulate sales staff help customers complete the purchase process." The reference [8] gives the formal definition of the recommendation system: set C be the set of all users, S is the set of all objects that can be recommended, and set the utility function u ( ) to calculate the recommendation R of object user, namely $u: C \times S \rightarrow R$, the main problem of recommendation system is to find the maximum object s of recommendation R, as shown in the formula $\forall c \in C, s^* = \arg\max_{s \in S} u(c,s)$.

The main algorithms of traditional personalized recommendation are: content-based recommendation algorithm, collaborative filtering algorithm. The core idea of content-based recommendation is to find the relevance of goods or content based on the metadata of the recommended goods or content, and then recommend similar items to the user based on the user's historical preference information. Collaborative Filtering (CF), there are two commonly used methods, memory-based (CF) methods and model-based methods (CF). The memory-based method uses user rating data to calculate the similarity between users and goods for recommendation. This is the method used in many business systems because it is easy to implement and has achieved some effect. This method is divided into two types, namely user-based collaborative filtering (User CF) and project-based collaborative filtering (Item CF). The UCF is a recommendation for the target user based on the preference information of the neighbor user. ItemCF is based on the users' scoring data for similar goods to predict the target goods and implement recommendations. Model-based collaborative filtering is to use the existing user preference information as a training sample to train a model that predicts the user's preference, so that the user can re-access the system and can run the model for recommendation. The model-based recommendation method commonly has matrix decomposition algorithm, clustering algorithm, Bayesian algorithm. In addition to the above recommended methods, there are recommendations based on knowledge and association rules, recommendations based on social networks and so on. But these traditional recommendation methods are basically based on some linear shallow models, and there are problems such as sparseness, scalability, cold start and so on, resulting in the recommended performance is not very good [9].

2.2 limited Boltzmann machine - RBM

The limited Boltzmann machine, RBM is a yojan of the BM structure of Boltzmann machine [10], as shown in Figure 1. In RBM, the connection-less topological structure between each unit of visible layer and each unit of hidden layer make the model is relatively simple, the same layer of neurons achieve condition independence, parameter learning is relatively easy. RBM has a visible neuronal layer (visible layer) which is used for receiving the input signal, and a hidden neuronal layer (hidden layer), which can be used as a feature extractor for the input signal. RBM is a probability model where all neurons select a 0 or 1 state with a probability value. But the bias unit is very special; it is always 1, which is mainly used to carry some inherent factors. The limited Boltzmann machine has been successfully applied to solve the problems of regression, dimensionality reduction, and collaborative filtering and image feature extraction and so on[11].



(a) BM model structure (b) RBM model structure

Figure 1 BM and RBM model structure

2.2.1 Calculation of RBM model

Assuming that the number of neurons in the visible layer is n_v, the number of neurons in the hidden layer is n_h, v_i mean state of the neurons in the visible layer, h_j mean the state of the hidden neurons, a_i mean the bias of the neurons in the visible layer, b_j mean the bias of the hidden layer of neurons. $w_{j,i}$ mean the connection weight between the visible and hidden layers. At the same time, mark $\theta=(w,a,b)$

For the specified state (v, h), the energy function of RBM is defined as [12]:

$$E_\theta(v,h)=-\sum_{i=1}^{n_v}a_iv_i-\sum_{j=1}^{n_h}b_jh_j-\sum_{i=1}^{n_v}\sum_{j=1}^{n_h}h_jw_{j,i}v_i$$

This energy formula is used to define the following joint probability distributions:

$$P_\theta(v,h)=\frac{1}{Z_\theta}e^{-E_\theta(v,h)}$$

Among them:

$$Z_\theta=\sum_{v,h}e^{-E_\theta(v,h)}$$

It is called the normalization factor

When there is a joint probability distribution, the edge probability distribution can be defined, namely:

$$P_\theta(h)=\sum_v P_\theta(v,h)=\frac{1}{Z_\theta}\sum_v e^{-E_\theta(v,h)}$$

$$P_\theta(v)=\sum_h P_\theta(v,h)=\frac{1}{Z_\theta}\sum_h e^{-E_\theta(v,h)}$$

Because of the special structure of RBM, it can be seen that neurons and hidden neurons are conditional independent. The above-mentioned joint probability distribution and edge probability distribution with this feature,, it is necessary to know it when the state of the hidden layer is given, the activated probability in visible layer can be seen, or when the state in the visible layer is given, the activated probability of a unit in the hidden layer. As can be seen from the Sigmoid function:

$$Sigmoid(x)=\frac{1}{1+e^{-x}}$$

Then:

$$P(h_k=1|v)=Sigmoid(b_k+\sum_{i=1}^{n_v}w_{k,i}v_i)$$

$$P(v_k = 1 | h) = Sigmoid(a_k + \sum_{j=1}^{n_h} w_{j,k} h_j)$$

The update rule of each parameter is:

$$W = W + \eta(\frac{1}{n_s} \Delta W)$$

$$a = a + \eta(\frac{1}{n_s} \Delta a)$$

$$\mathrm{b} = b + \eta(\frac{1}{n_s} \Delta b)$$

η is the learning rate, $\frac{1}{n_s}$ is the total number of samples among them.

2.2.2 learning method of RBM

The purpose of learning RBM network is to make RBM network most likely to fit the input data. At present, the most commonly used learning methods for RBM are Gibbs sampling algorithm, contrast divergence algorithm and PCD sampling algorithm, repectively. This paper uses the fast learning algorithm-Contrastive Divergence (CD) algorithm [13]. The first step of thIS algorithm is initialized w_{ij}, and then take a sample each time (ie, the visual layer is known) traverses the following steps.

(1) The x_i in visible layer calculate the hidden layer x_j, the positive gradient of w_{ij} is:

$positive(W_{ij}) = x_i * x_j$

(2) From the hidden layer x_j to reversely calculate $x_i^{'}$, pay attention x_i at this time is different from with the original x_i, so that the negative gradient of w_{ij} is:

(3) Update weight:

$$W_{ij} = W_{ij} + a * (positive(W_{ij}) - negative(W_{ij}))$$

By cycling training network, keep iterating until it converges ($x_i^{'}$ is very close to x_i) make the probability distribution represented by RBM is consistent with the training data as far as possible.

2.2.3 Learning process of RBM

(1) Initialization

① the data set is divided into train and test. The training set (train) set S is given;

② Training period T, learning rate η, and CD-k algorithm parameters k are given;

③ specify the number of units in visible layer and hidden layer n_v, n_h;

④ initialize the bias vector a, b and the weight matrix W.

(2) Training

FOR iter = 1,2, ..., T DO

{

①calls CDW (k, S, RBM (W, a, b), ΔW, Δa, Δb) and generate ΔW, Δa, Δb;

② Refresh parameters: W, a, b

③ Calculate the cost through the cross entropy cost function

}

Reconstruct test set

3. Deeply Restricted Boltzmann Machine-DBM

The deeply restricted Boltzmann machine DBM is a multi-layer neural network based on the RBM structure of the Boltzmann machine [14]. Generally speaking, the bottom layer of DBM is visual layer which composed of multiple visible neurons, increase the layer number of layers hidden from bottom to top. The DBM used in this paper is composed of three-layer deep learning network with a visual layer and two hidden layers, as shown in Figure 2. The training processes of DBM and RBM are slightly similar, the data set is input from the visible layer in the DBM, and then output from the first hidden layer; for the second hidden layer, The data set processed by the first process is input from the first hidden layer and finally output from the second hidden layer.

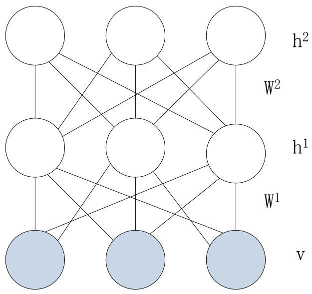


Figure.2 The structure of the deeply restricted Boltzmann machine

3.1 Calculation of DBM model

For this deeply restricted Boltzmann machine with two hidden layers and still unconnected within the layer, if give state, the energy function of DBM is defined as:

$$E_\theta(v,h^1,h^2) = -v^T W^1 h^1 - h^1 W^2 h^2$$

$\theta = \{w^1, w^2\}$ is the parameter of the energy model, which mean connection weight between the first hidden layer and the second hidden layer in the visible l layer.

This energy function is used to define the following joint probability distribution:

$$p_\theta(v,h^1,h^2) = \frac{1}{Z(\theta)} \sum_{h^1,h^2} e^{-E_\theta(v,h^1,h^2)}$$

The conditional probability of the visible and hidden layers is available by the activated function:

$$p(h_j^{\ 1} = 1 | v, h^2) = sigmoid(\sum_i W_{ij}^{\ 1} v_i + \sum_m W_{jm}^{\ 2} h_j^{\ 2})$$

$$p(h_m^{\ 2} = 1 | h^1) = sigmoid(\sum_j W_{im}^{\ 2} h_i^1)$$

$$p(v_i = 1 | h^1) = sigmoid(\sum_j W_{ij}^{\ 1} h_j)$$

In order to find the results of maximized likelihood functions, it still like training RBM, so that DBM goes through the same learning process. instead, the training of DBM is equivalent to two RBM training process.

3.2 Learning method of DBM

Because DBM's model structure is similar to RBM, the difference is that DBM has a hidden layer than RBM. Generally speaking, training DBM with three-layer network is equivalent to training twice RBM. Therefore, learning method of DBM can use the same as with RBM, the rapid divergence (CD-k) algorithm is commonly used .

3.3 Learning process of DBM

(1) Initialization

① the data set is divided into train and test. Given training set (train) set S;

② given training period T, learning rate η, and CD-k algorithm parameters k;

③ specify the visible layer and the number of two hidden layer nv, nh1, nh2;

④ initialize the bias vectors b, c1, c2 and the weight matrix W1, W2.

(2) Training

FOR iter = 1,2, ..., T DO

{

① call CD-k (k, S, RBM (W1, b, c1), ΔW1, Δb, Δc1) and generate ΔW1, Δb, Δc1;

② refresh parameters: W1, b, c1;

(2) Call CD-k (k, S ', RBM (W2, c1, c2), ΔW2, Δc1, Δc2) and generate ΔW2, Δc1, Δc2;

④ refresh parameters: W2, c1, c2;

⑤ Calculate the current cost through the cross entropy cost function.

}

(3) Reconstruct test set

4. Personalized Recommendation Based on Deeply Restricted Boltzmann Machine DBM

This paper add a layer of hidden layer based on RBM's double-layer network structure, and the DBM with three-layer network model is used to realize the personalized recommendation. The main steps of the recommendation process are as follows:

(1) Input data preprocessing (user item scoring data cleaning and normalization);

(2) DBM initialization (training cycle, learning rate, visible layer and hidden layer unit number, bias item, weight matrix);

(3) Train DBM (iteratively call CD-k, parameter refresh, cross entropy cost function calculate the current cost);

(4) Reconstruct the test set;

(5) Obtain a new predicted scoring matrix;

(6) According to the score rank, recommend N items for the user (TOP-N recommendation).

5. Experimental Validations and Evaluation

5.1 Related data sets

This experiment uses the MovieLens-100k data set, which contains 943 users' 100,000 score records for 1682 movies, the data set downloaded from the official website also provides sample set for cross-validation. Cross-validation is to obtain a reliable and stable model; the basic idea is to divide the original data into groups in a sense, part as the training set, and the other part as the test set. So not only consider the training error, but also consider the generalization error. The MovieLens-100k data set has randomly divided data sets into training sets (80% of samples) and test sets (20% of samples), these five sets of data can be used, and provides five sets of training sets and test sets, correspondingly input RBM and DBM as the training set and test set respectively, and the five experiment are made to compare performance, respectively.

5.2 Experimental process

First of all, the data set is made preprocessing, the first is record a single source data, the Python language is changed into the user - the film score matrix, the dimension is 943*1682, the line means the different users, the column means different movies. The middle value is the user's rating on the movie, as shown in Figure 3. On this basis, in order to facilitate the input of DBM, the matrix was normalized, as shown in Figure 4.

5.	3.	4.	3.	3.	5.	4.	1.	5.	3.	2.	5.	5.	5.	5.
4.	0.	0.	0.	0.	0.	0.	0.	0.	2.	0.	0.	4.	4.	0.
0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	4.	0.	0.	0.	0.
4.	3.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
4.	0.	0.	0.	0.	0.	2.	4.	4.	0.	0.	4.	2.	5.	3.
0.	0.	0.	5.	0.	0.	5.	5.	5.	4.	3.	5.	0.	0.	0.
0.	0.	0.	0.	0.	0.	3.	0.	0.	0.	3.	0.	0.	0.	0.
0.	0.	0.	0.	0.	5.	4.	0.	0.	0.	0.	0.	0.	0.	0.
4.	0.	0.	4.	0.	0.	4.	0.	4.	0.	4.	5.	3.	0.	0.

1.	0.6	0.8	0.6	0.6	1.	0.8	0.2	1.	0.6	0.4	1.	1.	1.
0.8	0.	0.	0.	0.	0.	0.	0.	0.	0.4	0.	0.	0.8	0.8
0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.8	0.	0.	0.
0.8	0.6	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
0.8	0.	0.	0.	0.	0.	0.4	0.8	0.8	0.	0.	0.8	0.4	1.
0.	0.	0.	1.	0.	0.	1.	1.	1.	0.8	0.6	1.	0.	0.
0.	0.	0.	0.	0.	0.	0.6	0.	0.	0.	0.6	0.	0.	0.
0.	0.	0.	0.	0.	1.	0.8	0.	0.	0.	0.	0.	0.	0.
0.8	0.	0.	0.8	0.	0.	0.8	0.	0.8	0.	0.8	1.	0.6	0.

Figure 3 User-Movie Score Matrix   Figure 4 Normalized User-Movie Score Matrix

The normalized user - the film score matrix input the DBM model, and after the learning and training of visible and hidden layer, a predictive scoring matrix is obtained, as shown in Fig.5, This predictive scoring matrix is used to achieve the project recommendation for the target user.

[0.94144985	0.66933228	0.85647064	0.63396069	0.66001183	0.94381703]
[0.86673473	0.0547389	0.30160626	0.03467637	0.13439187	0.10878264]
[0.03453112	0.53479726	0.83468347	0.23512283	0.03466927	0.63475454]
[0.02180735	0.23028435	0.57094535	0.38311334	0.63505102	0.35668143]
[0.81713519	0.6693366	0.04456853	0.19734683	0.06396843	0.04298768]
[0.86825985	0.34259877	0.26768973	0.44356493	0.01657345	0.39846512]
[0.39787613	0.0693366	0.05637471	0.93606939	0.01860013	0.04702638]
[0.18067965	0.38725967	0.62893507	0.12731896	0.74659924	0.27459263]
[0.34144985	0.84769952	0.20872654	0.13396038	0.38357699	0.79567246]
[0.85945141	0.19568325	0.15758664	0.8339607	0.47624588	0.48783465]

Figure 5 Predictive scoring matrix based on DBM model

5.3 Analysis and evaluation of experimental results

The cross entropy cost function [15] is a way to measure the predicted and actual values of the artificial neural network. If the gradient of the training parameter slowly reduce, then the value of the cross entropy cost function changes slowly, if the cross entropy cost function changes greatly from at the beginning of the training, proving that the learning is faster. Cross entropy cost function expression formula:

$$C = -\frac{1}{\mathrm{n}}\sum_{x}[y\ln a + (1-y)\ln(1-a)]$$

C means the cost, x means the sample, y means the actual value, a means the output value, and n mean the total number of samples. The influence that the number of hidden neurons and training times on the cross entropy cost function, it is necessary to find the relative optimized number of hidden neurons and training for the cross entropy cost function through experiment.

(1) Testing the influence of the number of hidden neurons on the cross entropy cost function, and find the number of neurons in hidden layer for RBM relative optimization. The parameters discussed in Figure. 6 are carried out under the condition of 100 times of training, it can be seen from the results that the number of hidden neurons has little effect on the cross entropy cost function. But considering from the experimental results, take the lowest value here, and the number of neurons in the RBM hidden layer is 70 in the following experiment.

(2) Testing the effect of RBM training number on the cross entropy cost function, and find the relative optimized training times for RBM. As can be seen from Fig. 7, as the number of training increases, the cost of the cross entropy cost function is lower, and when the number of training times is 600, the curve is slightly gentle, so training times of RBM take 600 times.

(3) The experiments in Figure 5-1 show that the number of hidden neurons has little effect on the cross entropy cost function, and some research have shown that the number of neurons in the second hidden layer is twice the number of neurons in the first hidden layer, the prediction effect is better. Therefore, when the number of hidden neurons in RBM is set to 70, the number of two hidden layer neurons in DBM should be set to 70 and 140, respectively.

(4) Testing effect of DBM training times on the cross entropy cost function. As shown in Fig. 8, it can be seen that the cross entropy cost when the number of training times of DBM is 120, it is less than 600 times for RBM training, so the number of training times of DBM is 120 times.

(5) The accuracy based on of DBM personalized recommendation use root mean square error (RMSE) as the evaluation criterion.

$$RMSE = \sqrt{\frac{\sum r_{ij} \in R_{test} (\hat{r_{ij}} - r_{ij})^2}{|R_{test}|}}$$

R_{test} is the test set, $\hat{r}_{ij}$ is the prediction score. (DBM), the restricted Boltzmann machine (RBM) and DEEPLY restricted Boltzmann machine mentioned by this article (DBM), user-based collaborative filtering (UCF), project-based collaborative filtering (ICF) are compared, the experiment results are shown in Figure 9. It can be seen that DBM works better than other recommendations method in five sets of data.

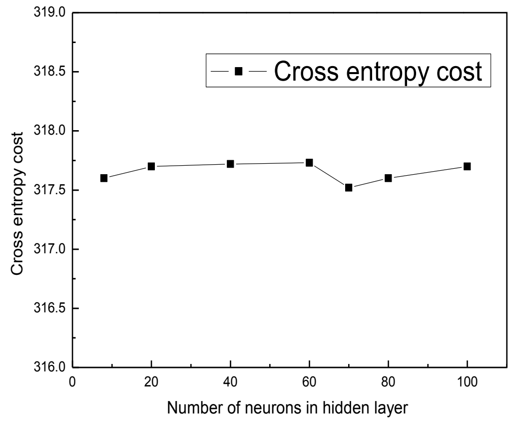


Figure 6 relationships between the number of hidden layers of RBM and the cross entropy cost function

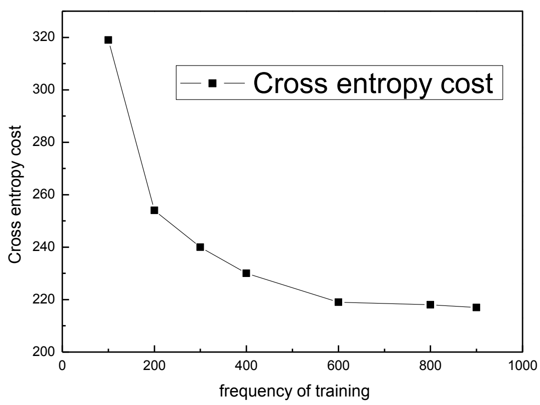


Figure.7 influence of RBM training times on cross entropy cost

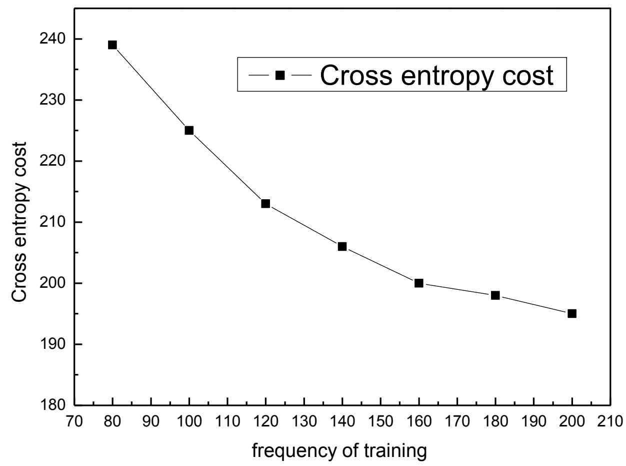


Figure.8 influence of DBM training times on cross entropy cost

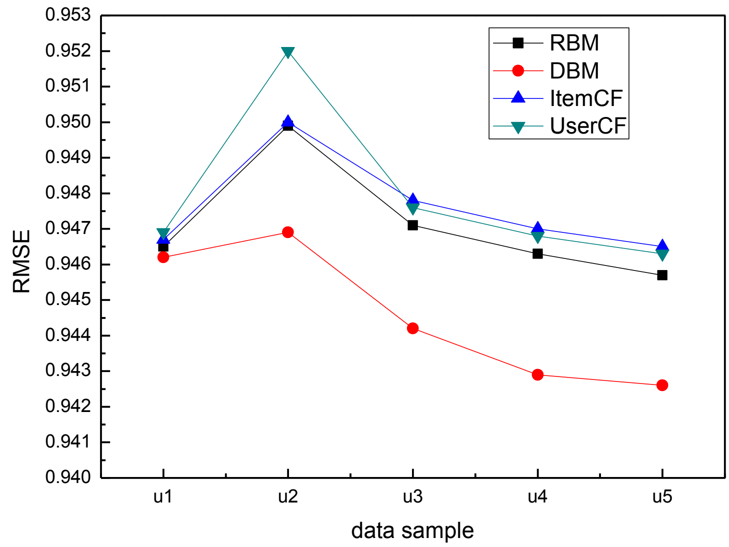


Figure.9 RMSE comparison of several recommendation methods

6. Conclusion

The personalized recommendation issues based on deeply restricted Boltzmann machines are discussed in this paper, DBM with three layers of network structure have been built, three layers of DBM is equivalent to the superposition of two RBM compared with RBM with two layer of structure, which has a hidden layer for iterative training. Although the three-layer DBM has added some algorithm complexity, but computer technology is rapidly developing today, this complexity is not an important problem at all. Through the experiment of the moviels-100k data set, it shows that the three layer DBM personalized recommendation has better recommendation performance than RBM, UserCF, ItemCF and other methods.

Acknowledgements

Fund Project: Science and Technology Key Project of Anhui Province (1501031082).

References

[1] Recommendation System Practice [M]. Beijing: Posts & Telecom Press, 2012:59-60.
[2] Geo_rey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with the networks networks. Science, 313 (5786): 504-507, 2006.
[3] Yoshua Bengio. Learning deep architectures for ai. Machine Learning, 2 (1): 1-127, 2009.
[4] Ruslan Salakhutdinov, Andriy Mnih, and Geo_rey Hinton. Restricted boltzmann machines for conduct filtering. In Proceedings of the 24th international conference on Machine learning, pages 791-798. ACM, 2007.
[5] Jian Wei, Jianhua He, Kai Chen, etc. Collaborative filtering and deep learning based recommendation system for cold start items. Expert Systems with Applications, 2017, (69): 29-39.
[6] Li Chun, Zhu Jiamin, Ye Jian, Zhou Jiaying. Overview of Personalized Service Research [J], Application Research of Computers, 2009, 26(11):4001-4005.
[7] Xie X Q, Song C C, Zhang Z Q. A service discovery method based on referral network and ant-colony algorithm[J]. Chinese Journal of Computers, 2010, 33(11): 2093-2103.
[8] Adomavieiu G,Tuzhilina. Toward the next generation of recommender systems:A survery of the state of the art and possible extensions. IEEE Trans, on knowledge and Data Engineering,2005,17(6):734-749.
[9] Jiang W, Yang L. Research of improved recommendation algorithm based on collaborative filtering and content prediction[C]//Proc of the 11th International Conference on Computer Science & Education. 2016:598-602.
[10] Zhang Chunxia, Ji Nannan, Wang Guanwei. Restricted Boltzmann Machine [J]. Journal of Engineering Mathematics, 2015(32), 2: 159-173.
[11] Li Tong, Research on Feature Extraction of Multi-Spectral Remote Sensing Image and Collaborative Interpretation [DB], CMFD, 2015.
[12] Zhang Yanxia. Deep Learning Model Based on the Restricted Boltzmann Machine and Its Application [D], Chengdu: University of Electronic Science and Technology of China, 2016.
[13] Asela Gunawardana,Christopher Meek.Tied boltzmann machines for cold start recommendation[J]. ACM international conference on recommender systems, 2008, 10: 23-25.
[14] Lu Ping, Chen Zhifeng, Shi Lianmin. Comparison of learning methods of RBM [J]. Computer Era, 2014(11):10-13.
[15] You Bo, Li Zhongjie, Huang Ling, Hand gesture recognition based on improved BP neural network [J/OL]. CAAI Transactions on Intelligent Systems: 1-7. (2017-07-03). http://kns.cnki.net/kcms/detail/23.1538.TP.20170703.1601.008.html.