Research on Construction of Recommendation System On account of CNN and PMF model

Hao Luo1*, Zhongwu Li1

1School of information, Baoshan University, Yunnan, China

*Corresponding author e-mail: luohao@bsnc.cn

Abstract. As an important branch and direction of machine learning, neural network shows its technical superiority in many aspects. However, due to the shortcomings of existing shallow structure algorithms, it is necessary to strengthen the ability to express effects, so as to learn the essential features of data sets from a large number of unlabeled specimens. On account of this, this paper first analyzes the concept and application of CNN network, then studies the algorithm and process of PMF model, and finally gives a recommendation system on account of CNN and PMF model.

Keywords: Recommendation System, CNN, PMF, Model

1. Introduction

With the iterative progress of intelligent tech represented by neural network, it has been widely used in many fields, and has achieved remarkable results. As an important branch and direction of machine learning, neural network shows its technical superiority in many aspects. In the case of deep layers, the gradient of neural network is easy to disappear, and the more layers of the network, the more serious the effect of the network weight adjustment is, and the performance of the network is reduced and limited. Moreover, due to the lack of shallow structure algorithm, it is difficult to express complex effects in limited specimens and calculation units.

The shortcomings of shallow structure algorithm are mainly reflected in its long model, which needs to rely on manual to extract specimen features, resulting in low efficiency and poor effect. In view of the shortcomings of existing shallow structure algorithms, it is necessary to strengthen the ability to express effects, so as to learn the essential features of data sets from a large number of unlabeled specimens [1]. On the other hand, deep learning model has deep level and strong expression ability, and has the ability to represent large-scale data.

In addition, for the quality of the recommendation system, the lack of quality is mainly caused by the sparseness of user rating data. To deal with these sparsity problems, more information is needed to improve the accuracy of evaluation and prediction. When the rating data is scarce, the document based modeling method improves its accuracy by using additional data as shown in Figure 1. On account of context aware recommendation model, convolution matrix is decomposed and integrated into PMF to improve the effect of score prediction. Therefore, it is of great practical value to study the construction of recommendation system on account of CNN PMF model.
2. The concept and practice process of CNN

2.1. The CNN neuron model

As a multilayer neural network, CNN is composed of a two-dimensional plane composed of several independent neurons, and the inputs of its neurons are connected with the receptive field. Secondly, in the level of local feature extraction, the location relationship between features is closely related to the extraction of local features [2]. In addition, the neurons on the mapping surface share weights, which further reduces the complexity of network parameter selection. In CNN, only the input connections between s-elements are variable, while the input connections of other elements are fixed. The output of an s-element on the $k_i (k_i, n)$ s-plane is represented by $u_{sl} (k_i, n)$, and the output of a c-element on the $k_i c$-plane of this order is represented by $u_{l} (k_i, n)$. In which, $n$ is a two-dimensional coordinate, which represents the location of the receptive field of neurons in the input layer. In the first phase, the area of receptive field is small, and then increases with the increase of $l$.

\[
 u_{sl} (k,n) = r_l (k) \phi \left[ 1 + \sum_{k_{l+1}}^{K_{l+1}} \sum_{v \in A_l} a_l (v, k_{l+1}, k) u_{sl-1} (k_{l+1}, n + v) \right] \\
 \left[ 1 + \frac{r_l (k)}{r_l (k)} b_l (k) u_{sl} (n) \right] - 1
\]  

(1)

Among them, $a_l (v, k_{l+1}, k)$ and $b_l (k)$ represent the connection coefficients of excitatory input and inhibitory input respectively. $r_l (k)$ controls the selectivity of feature extraction, and the larger the value, the worse the tolerance to noise and feature distortion, and it controls the input of each neuron in a single suppressor plane at each s-layer [3]. The higher the value of $r_l (k)$, the greater the excitability in proportion to inhibition, that is, the greater the $r_l$ value, the greater the output. On the other hand, a small $r_l (k)$ value allows the unmatched neurons to excite and only produce a relatively small output; $\phi (x)$ is a nonlinear effect.

\[
 \phi (x) = \begin{cases} 
 x, & x \geq 0 \\
 0, & x < 0 
\end{cases}
\]  

(2)

The action effect of s-element can be divided into excitatory effect and inhibitory action effect. The former increases the membrane potential, while the latter acts as a shunt:
2.2. The practice process of CNN network

The practice process of CNN network mainly includes two phases: forward propagation phase and backward propagation phase. The former takes a specimen from the specimen set and inputs the specimen into the network [4]. Secondly, the actual output is obtained, and the information is transformed step by step, and finally transmitted to the output layer to realize the normal operation of CNN after practice. In addition, in the second phase, the weight matrix is adjusted according to the method of minimizing the error between the actual output and the corresponding ideal output. The error measure of the network controlled by precision needs to be defined with respect to the whole specimen set, as shown in the following formula:

$$E_p = \frac{1}{2} \sum_{j=1}^{m} (y_{pj} - a_{pj})^2$$  \hspace{1cm} (4)

In the specific process of CNN network practice, it first needs to select the practice group, and then randomly select a certain number of specimens from the specimen set as the practice group. Secondly, the weights and thresholds are set to random values, and the precision control parameters are initialized. In addition, an input mode from the practice group is added to the network and its target output vector is given [5]. On account of the intermediate layer output vector, the actual output vector of the network is calculated. The elements in the output vector are compared with those in the target vector, and the output error term is calculated:

$$\delta_j = h_j(1-h_j) \sum_{k=0}^{M-1} \delta_k W_{jk}$$  \hspace{1cm} (5)

At the end of the practice, the weights and thresholds are saved in the file, and the weights are stable to form a classifier. During the practice, the weights and thresholds are directly exported from the file for practice without initialization.

3. The concept and process of PMF

3.1. Conceptual model of PMF

As a collaborative filtering recommendation algorithm on account of matrix decomposition, PMF model assumes that the system has $m$ users, $R$ is a system scoring matrix with $m$ rows and $n$ columns, and $R_{i,j}$ represents the score of user $i$ on item $j$. $R$ contains the user's hobby characteristics, and also contains the hobby characteristics of this item. Let $U$ be a matrix of $d$ rows and $m$ columns, and each column represents the eigenvector of users; $V$ is a matrix of $d$ rows and $n$ columns, and each column represents the eigenvector of an item [6]. Let $f$ be the prediction effect, which takes the user characteristics and item characteristics as the input value, and the output value as the predicted score of the user for the item. The difference between the score data and the predicted score follows normal distribution:

$$p(R | U, V, \sigma^2) = \prod_{i=1}^{m} \prod_{j=1}^{n} \left[ \mathcal{N}(R_{i,j} | f(U_i^T V_j), \sigma^2) \right]^{I_{i,j}}$$  \hspace{1cm} (6)

In which, $\mathcal{N}(R_{i,j} | U_i^T V_j, \sigma^2)$ is a normal distribution with mean value of $U_i^T V_j$ and variance of $\sigma^2$, and $I_{i,j}$ is an indicator effect. When the user has scored $j$ on item $i$, then $I_{i,j}$ is equal to 1,
otherwise \( I_{i,j} \) is equal to 0. Assume that the errors of the scores are independent of each other, so it could write them in the form of multiplication. Similarly, assume that the user's eigenvector and the item's eigenvector obey the normal distribution with mean value 0, variance \( \sigma_u^2 \) and mean value 0, variance \( \sigma_v^2 \), and finally:

\[
p(U, V | R) = \frac{p(R | U, V) p(U) p(V)}{p(R)}
\]

(7)

3.2. Gradient descent method

The idea of gradient descent method is that the value of effect decreases the fastest along the opposite direction of partial derivative of multivariate effect. After the objective effect is obtained, the gradient descent method can be used to continuously adjust the parameters \( U \) and \( V \). The partial derivatives of \( U \) and \( V \) are obtained respectively:

\[
-\frac{\partial E}{\partial U_i} = \sum_{j=1}^{M} I_{i,j} ((R_{i,j} - p(U_i, V_j)V_j) - \lambda_u U_i)
\]

(8)

\[
-\frac{\partial E}{\partial V_j} = \sum_{i=1}^{N} I_{i,j} ((R_{i,j} - p(U_i, V_j))U_i) - \lambda_v V_j
\]

(9)

3.3. PMF process initialization and system evaluation

First, at the initialization level, the setting of initial values of variables will affect the rate of model learning. The initialization method is to use random values as the initial values. In order to avoid the influence of the randomness of initial values on the stability performance of the system, initialization is required:

\[
U_{i,j}, V_{i,j} = \sqrt{\frac{R-a}{d}} + n(r)
\]

(10)

In which, \( a \) is the minimum value of all scoring data, \( d \) is the dimension of feature vector, and \( n(r) \) is a very small random value in a certain range. Secondly, at the level of system evaluation, RMSE is used to evaluate the prediction accuracy of the system:

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{m} \sum_{j=1}^{n} I_{i,j}(R_{i,j} - f(U_i, V_j))^2}{\sum_{i=1}^{m} \sum_{j=1}^{n} I_{i,j}}}
\]

(11)

4. Recommendation system on account of CNN PMF model

4.1. Advantages of CNN recommendation system on account of PMF model

By integrating CNN into PMF, the ConvMF can capture the context information of documents and further improve the accuracy of score prediction. The results of extensive evaluation of data sets show that the character of the fusion model is much better than the most advanced model in the case of less data, and the model can successfully capture the subtle differences of words in documents. Figure 2 below shows the probability graph model of ConvMF which integrates PMF model and CNN model.
4.2. Recommendation system constructed by CNN and PMF fusion model

ConvMF of probability matrix decomposition model and CNN model uses the detailed framework of project description, and uses the document potential vector obtained from CNN model as the mean of Gaussian distribution of project variables, and plays an important role in the fusion process of CNN and PMF models. Secondly, convMF achieves significant improvement on all datasets. ConvMF is well integrated into PMF for recommendation tasks and can utilize rating information.

5. Conclusion

In summary, due to the deficiency of shallow structure algorithm, it is necessary to strengthen the ability to express effects, so as to learn the essential features of data sets from a large number of unlabeled specimens. On account of the context aware recommendation model, the convolution matrix is decomposed and integrated into PMF, which can effectively improve the effect of score prediction. On account of the analysis of the concept and application of CNN, this paper studies the CNN neural model and practice process. Through the research of PMF model algorithm model and process, the PMF process initialization and system evaluation are analyzed. Through the analysis of the recommendation system constructed by CNN fusion PMF model, the advantages of the recommendation system constructed by CNN fusion PMF model are studied.

References

[1] Chen Yong, Zhao Hangsheng. Design and implementation of fast acquisition scheme for large frequency difference long address code spread spectrum system [J]. Acta electronica Sinica, 2017,29 (9): 2187-2190.

[2] Dai Qingyan, Zhu Zhongjie. Image segmentation on account of super pixel and improved iterative graph cut algorithm [J]. Computer Engineering, 2016 (42): 220-226.

[3] He Zhiming. Overview of image segmentation [J]. Shandong industrial tech, 2016 (22): 226-226.

[4] Luo Jufeng, Wang Xiang, Fu Yaoxian. An improved PMF-FFT acquisition method for resource constrained applications [J]. Computer research and development, 2010,47 (2): 144-150.

[5] Zhou Feiyan, Jin Linpeng, Dong Jun. Review of convolutional neural networks [J]. Acta Sinica Sinica, 40 (6): 1229-1251.

[6] Zhu Jiawei, Pan Zhouxian, Chen Shi. Application and Prospect of virtual reality tech in medical field [J]. Basic medicine and clinic, 2018, 38 (3): 422-425.