CBR-Recommendation System on Massive Contents Processing Using Optimized MFNN Algorithm

Rui Li[1, a], Jianyang Li[2, b], Benkun Zhu[2, c]

[1]*Department of Information Engineering, Anhui Communications Technical College*
Hefei, 230051, China
[2]*School of Computer and Information, Hefei University of Technology*
Hefei, 230009, China
[a]*email: Liruilary@gmail.com,* [b]*email: lijianyang@sina.com,* [c]*email: zbk@zjc.edu.cn*

Abstract

Though recommendation systems have been widely used for websites to generate new recommendations based on like-minded users' preferences, IEEE Internet Computing points out that current system can not meet the real large-scale e-commerce demands, and has some weakness such as low precision and slow reaction. Huge personalized data are the key to successfully give a new recommendation, but they are difficultly dealt with for they are massive with high dimensional; addressing such problems, the paper suggests to use multi-layer feed-forward neural networks (MFNN) system based on case intelligence to partition massive personalized data into the most similar groups. The subsequent experiment indicates that our system model is constructive and understandable, and our algorithm can decrease the complexity of ANN algorithm, for which the system performance can be guaranteed.

Keywords:CBR-Recommendation System; Optimized MFNN Algorithm; Automatic Retrieval; Massive Contents

Introduction

E-commerce is increasing quickly and has been reshaping the world trades, and how to acquire the consumers' needs and improve their satisfaction, is emerging from the both sides' needs -the suppliers and customs. How to make a successful recommendation becomes the essential task for recommendation system, which can provide consumer with the information and advice of products, simulate a sales to help our consumers through the purchase process. As well known, acquiring the personalized knowledge is the key process for recommender to achieve success, which must identify the specific needs of each consumer on the basis of the consumer preference [1].

The system uses data mining and other artificial intelligence technology [2], analyzes the collected data, obtain the behavior and generate interests, where web-mining emerges in response to the e-commerce need and has gradually grown up to a complete technical system. Both IR and IE are the two essential steps to explore personalized data from websites, and the system construction must involve in many technologies, like database technology, information technology, statistics, ANN and machine learning, to exploit potential information or patterns useful [3]. In order to achieve those multiplex tasks in such complex environments and complicated actions, the paper [4] has described that case-intelligence recommendation system, using a variety of data mining technology, which can be used for acquiring effective personalized knowledge.

Plentiful personalized data is the key to meet the individual needs for recommendation system, for the recommender is a data priority- the more accumulation of data, and the higher accuracy the recommender can perform. But the real users' behavior drilled from websites can accumulate up to millions or even billions, the processing of massive users' data is the greatest challenge- they are huge with high dimensional and involves the system performance sharply. That is why the statistics report from ACM points out such dissatisfactory performances in the current recommender, which has discussed in the paper [5], this paper proposes a new method by using MFNN algorithm to solve such problems.

MFNN Algorithm

Great achievement has been acquired for CBR in the field of knowledge lack, which is the simulator of human analogy learning, and becomes the foundation of the case intelligent decision techniques, for which case is the integrated representation of the human sense, logics and creativity. ANN has the natural relationship with CBR, and they can complement each other perfectly.

Case Intelligence with MFNN

MFNN consists of one input layer, one or more hidden layers, one output layer, where each neuron in every layers is a processor to be used to process simple information. It has proved that 3-layer MFNN can realize any given function for approximate accuracy, thus it can be used to solve the nonlinear classification problem. The case library in the CBR system can be viewed as a CSP, therefore CS-ANN model, such as Schema model, Hopfield model, Boltzmann and Harmony theory can be employed to construct the case library[6,7].

Case retrieval is the key process of the CBR intelligent system. Currently the main way for case-matching is the k-nearest neighbor algorithm, but it can not reflect the relationship between the cases and their attributes, neither can it shows the preference of the customers; especially for the large-scale case library, the retrieving time is unacceptable.

Facing such problems, several successful theories have been put forward to integrate ANN into the CBR system and covering all the application aspects of it

with the ANN components. Theoretically, in the symbolic description model-based CBR system, rules can be elicited by ANN method; and in the quantitative description model-based CBR system, due to the system's flexibility, many mathematical approaches and optimization techniques can be employed in the definition and analysis of similarity measurement or case adaptation criteria.

Classic MFNN Algorithms

MFNN has a clear structural layer, and can be a digraph as input a vector x, then through the networks to get an output vector y, which can be used as a feedforward networks to process the mapping x to y of a converter. MFNN has many classic models and algorithms, such as back-propagation network, radial basis function network, simulated annealing algorithm and their ameliorated algorithms.

MFNN has a lot of various improved algorithms, and has achieved many significant results as we have seen. But all of these changes make the networks become more complex (or performance function is more complex, such as linear becomes nonlinear, etc.; or structures become more complex), which hopes to increase the complexity of the network structure to improve network learning speed as return.

After investigating the behavior of MFNN for case retrieval, the weaknesses such as having lower speed and local extreme value, are inherent in those algorithms. For example, RBF is a good similarity detector, but it can hardly deal with huge user data directly. However, many of those weaknesses are resident in current algorithms, and can not be conquered to achieve satisfactory level, especially for such complex data.

System Construction and Processing

Our domain algorithm is a constructive method of MFNN, which based on the geometrical representation MP model to build three-layer networks by means of its own structure of input data.

Domain Algorithm

Each input vectors x of an n-dimension can be projected to a certain hyper-sphere S^n of an expanding (n+l)-dimensional space, and a "certain spherical domain" is corresponded with a neural weight and threshold function, where the transformation $T : D \to S^n$, $T(x) = (x, \sqrt{R^2 - |x|^2})$ should be used to achieve the projection, thus all points of sample D are projected to S^n [8].

Assume the input samples set $K = \{x^1, x^2, \cdots, x^m\}$ can be classified as subset of so many with r classes $K = \{x^1, x^2, \cdots, x^r\}$; then, a group of sphere

domains can be used to cover the samples set K:

(1) Calculate the center of all the samples, and then cover from this point, which is the nearest sample point $a^i \in K^j$ from the center;

(2) Calculate the domain $C(a^i)$, which centers from a^i. Suppose $C(a^i) \cap K^j = D^i, i = 1,2,\cdots, \quad D_0 = \varnothing$, $d^1(i) = \max\{\langle a^i, x \rangle\}$, $d^2(i) = \min\{\langle a^i, x \rangle | \langle a^i, x \rangle > d^1(i)\}$. Then compute the radii of sphere covering domain: $d(i) = (d^1(i) + d^2(i))/2$;

(3) Calculate barycenter b of D_i if D_{i-1} is the subset D_i, suppose $a^{i+1} = b$, $i++$, return step (2) until the number of samples covered are not more than the number of the samples. Delete all the points covered by C_k, $K^{jr} = C_k \cap K^j$, $K^m = K^j / K^{jr}$, $K^j = K^m$, $r++$. Then calculate another covering.

After executing all these steps, a group of sphere domains $\{C_1, C_2, ..., C_p\}$ can be gained, and can be proved that the samples in the same covering domain must have a high degree of similarity. In our recommendation system, the personalized user data can be partitioned into several "domains", where they are the most similar groups.

System construction

By this way of vectors dimension expansion and space projection, these domains can be used as the input users' vectors to the MFNN for case retrieval. Thus, the next two steps are built for our recommendation system.

Firstly, the construction of MFNN can be described as follows:

The first layer: assume total amounts of P neurons $A^1, A^2 ... A^p$, A^i represents the neuron of covering $C(i)$. $W(1) = (a^1), \theta(1) = (\theta_1)$

The second layer: select the same amounts of neurons as the first step $B^1, B^2, \cdots, B^p$.

$$W_j^i(2) = \begin{cases} -1, j < i \\ 1, j = i \\ 0, j > i \end{cases}, \theta_i(2) = i - 1, \begin{matrix} (i = 1,2,\cdots, p) \\ (j = 1,2,\cdots, p) \end{matrix}$$

The last layer: select total amounts of T neurons $C^1, C^2, \cdots, C^T$, T is the total classification type of samples.

$$W_j^i(3) = \begin{cases} 1, (j-i)\bmod T = 0 \\ 0, others \end{cases}, \theta_i(2) = i-1, (i = 1,2,\cdots,T), (j = 1,2,\cdots,p)$$

The network adds a hidden layer to make the neuron weights increase linearly and decrease the complexity of the algorithm, in contrast, the original three-layer neural networks makes the neuron weights increase with index.

Secondly, the system framework that we proposed based on our MFNN algorithm (figure omitted for page limited, which has described in our paper [5]), mainly constructs by three parts: input module, recommendation methods and output module, where our MFNN algorithm is added in as retrieval process to be evaluated directly with huge data. Our case recommender has such characteristic advantage- excellent flexibility, which leads to a regenerate process for case retrieval.

Experiments and Analysis

The data of "forest cover type" in our experiment is downloaded from UCI repository, and is designed to validate our MFNN algorithm, whose main information is described as follows: Number of instances (observations) 581,012, Number of Attribute: 54; Number of Class: 7.

Each record represents the user personalized data collected from the websites, which is regarded as a user behavior vector with 54 Attributes, and users' data library accumulates to 581,012 users' sessions. Then, the normal Macro-averaging is used to calculate all classes' means F-score.

As table 1 shows, we can find the data are spare matrix with high dimensional and huge records, and the system efficiency is significantly enhanced. Our experimental results also indicates that recommendation system runs in two stages, the first spends on training the users' personalized cases, which costs too many time, but can be run in the backstage; and the second spends on recommending the proper case just in a few time. Thus, ours can manipulate massive personalized data effectively, and improve the performance of e-commerce recommendation.

Table 1 system performance

User Cases	F-score(%)	T-partion(s)	Time(ms)
10,000	79.1	14.207	14.81
15,000	81.7	33.225	34.339
20,000	82.9	49.209	50.391
30,000	81.3	68.316	69.577
40,000	82.4	86.05	87.349

50,000	81.6	103.06	104.4
100,000	83.2	272.31	273.75

Conclusion

Personalized data involves a process of gathering and storing information about site visitors, they are the key assets for analyzing current and past user interactive behavior, and delivering the right content to each visitor; but they are massive along with high dimension, and can be hardly manipulated. The sequent investigations indicate that our recommender has clear system structure, feasible component combination, easy integration and construction. Our experimental results suggest that our MFNN algorithm is suitable for the large-scale and high dimensional data processing, which can guarantee the better performance for CBR-recommendation system.

Acknowledgement

References

[1] Wei Chu，Seung-Taek. Park, "Personalized Recommendation on Dynamic Content Using Predictive Bilinear Models" [C], WWW2009, pp691-700

[2] Bach, K., Althoff, K.-D., Newo, R., Stahl, A. "A Case-Based Reasoning Approach for Providing Machine Diagnosis from Service Reports" [C]. ICCBR 2011. LNCS, (6880), pp. 363-377

[3] Zurina Saaya, Markus Schaal, Maurice Coyle, Peter Briggs, and Barry Smyth. "Exploiting Extended Search Sessions for Recommending Search Experiences in the Social Web" [C]. ICCBR 2012. LNAI, (7466), pp. 369-383

[4] Jianyang Li, Xiaoping Liu. "Personalized Recommendation System on Massive Content Processing Using Improved MFNN" [C]. Springer's LNCS 7529 (2012), pp183-190

[5] Jianyang Li, Xiaoping Liu, Rui Li. "Optimized RBF for CBR-Recommendation System" [J]. AMM 214 (2012), pp568-572

[6] Debarun Kar, Sutanu Chakraborti, and Balaraman Ravindran. "Feature Weighting and Confidence Based Prediction for Case Based Reasoning

Systems" [C]. LNAI, (7466), pp. 211-225

[7] Zhiwei Ni，Jianyang Li，Fenggang Li，Shanlin Yang. "Survey of Case Decision Techniques and Case Decision Support System" [J]. Chinese Computer science,2009,36(11),pp18-24

[8] ZHANG Ling. "The relationship between Kernel Functions Based SVM and Three-layer Feedforward Neural Networks" [J]. Chinese J. Computer, 25(7): 696-700, 2002.