Rating Prediction Operation of Multi-criteria Recommender Systems Based on Feedforward Network

Mohammed Hassan
Software Engineering Lab
University of Aizu
Aizu-Wakamatsu, Japan
d8171104@u-aizu.ac.jp

Mohamed Hamada
Software Engineering Lab
University of Aizu
Aizu-Wakamatsu, Japan
hamada@u-aizu.ac.jp

ABSTRACT

Recommender systems are software systems that have been widely used to recommend items to the user. They have the capacity to support and enhance the quality of decisions people make when finding selecting items online. The most common techniques used by many recommendation systems are collaborative filtering, content-based, knowledge-based and hybrid-based which combines two or more techniques to make predictions and recommendations.

Multi-criteria recommendation technique is a new technique used to recommend items to users based on multiple attributes of items. This technique has been used and proven by researchers in industries and academic institutions to provide more accurate predictions and recommendations than the traditional techniques. What is still not yet clear is the role of some machine learning algorithms such as artificial neural network to improve its prediction accuracy. This paper proposed using feedforward neural network to model user preferences in multi-criteria recommender systems. The operational results of experiments for training and testing the network using two training algorithms and yahoo movie data sets are also presented.

Categories and Subject Descriptors

H.4.2 [**Information Systems and Applications**]: Decision Support; I.2.1 [**Artificial Intelligence**]: Application and Expert Systems

General Terms

Recommender systems, Algorithms, Experimentation

Keywords

Recommender systems, Artificial neural network, Prediction Accuracy

1. INTRODUCTION

A recommender system is an intelligent system that plays an important role in providing suggestions of valuable items to users. Such suggestions take the forms of different processes of decision making, like the kind of movie to watch, a music to listen, items to buy, or an online news to read [17, 12, 16]. Recommender systems are classified based on the technique used to design the system. Traditionally, collaborative filtering, content-based, knowledge-based, and hybrid-based are the major techniques used to describe the name or nature of the system. Therefore, knowing the recommendation techniques is at the heart of our understanding of recommender systems. Those techniques are sometimes called traditional techniques, and are increasingly becoming popular ways of building a system that combat problems of information overload [7, 5].

However, despite their popularity and providing considerable prediction and recommendation accuracies, they suffer from major drawbacks [1, 4, 13] because they work with just a single rating, whereas most of the time the acceptability of the item recommended may depend on several item's attributes [3]. Researchers suggested that if ratings provided to those several characteristics of items will be considered during predictions and recommendation process, it could help to enhance the quality of recommendations since complex opinions of users will be captured from various attributes of the item. Recent developments in this field have led to the existence of new recommendation technique known as multi-criteria recommendation technique [1, 3] that exploits multiple criteria ratings from various item's characteristics to make recommendations. This technique has been used for wide range of recommendation applications such as recommending products to customers [15, 11], hotel recommendation for travel and tourism [8], and so on.

Having considered multi-criteria technique as a resolution to some flaws of traditional techniques, it is also logical to look at various ways of modeling the multiple ratings to enhance the prediction accuracies and recommendation qualities. However, few researchers have been able to advance on some systematic research into improving the accuracy [9]. In addition, no previous research has investigated the effect of using artificial neural network to model user's preferences in order to improve the operations of multi-criteria recommender systems [3]. Therefore, in our quest to compensate for this knowledge gap, this study seeks to shine new light on the best way to use feedforward network through examination of the performance of backpropagation and delta rule algorithms to train the network using multi-criteria rating data set for recommending movies to users based on four

ICAIT '16, Oct. 6 – 8, 2016, Aizu-Wakamatsu, Japan.

attributes of movies. This paper has been divided into four parts including this introduction section. The second part of the paper gives a brief literature review. The experimental methodologies are contained in the third part while the fourth section displays the result and discussions and the final section is concerned about conclusion and presenting future research work.

2. LITERATURE REVIEW

To be able to understand the concept of recommender systems, we introduced some mathematical notations μ, ι, δ, and ψ to represent the set of users, the set of items, a numerical rating, and a utility function respectively. The notation δ (the rating) is the measure of the degree to which a user in the set μ will like and item in ι, while the utility function ψ is a mapping from a $\mu \times \iota$ pair to a real number δ written as $\psi : \mu \times \iota \mapsto \delta$. The value of δ is a real number within a specifically defined interval such as between 1 to 5, 1 to 13, or it can be represented using non-numerical values such as like, don't like, . . ., strongly like, true or false, and so on [14]. Therefore, recommender systems try to predict the value δ of items in ι that have not been seen by the user and recommend those with a high value of δ.

The methods of prediction and recommendation explained in the above paragraph are the mechanisms followed essentially by traditional recommendation techniques. Moreover, a similar approach is followed by multi-criteria recommendation technique with the distinction that it uses multiple values of δ for each $\mu \times \iota$ pairs. In multi-criteria technique, the utility function ψ can be defined using one of the relations in equation 1.

$$\begin{aligned} &\psi : \mu \times \iota \mapsto \delta_0 \times \delta_1 \times \delta_2 \times ... \times \delta_n \\ OR& \\ &\psi : \mu \times \iota \mapsto \delta_1 \times \delta_2 \times \delta_3 \times ... \times \delta_n \end{aligned} \tag{1}$$

It is important to note however, that the two relations in equation 1 are different due to the presence of δ_0 in the first mapping. There are $n + 1$ and n ratings in the relations respectively. The additional rating δ_0 is called the overall rating which is either to be obtained from the user together with other n ratings or its value need to be computed based on the other n values as in equation 2. Detail explanation of multi-criteria recommendation is beyond the scope of this paper, readers can refer to [3, 2, 13] for more information.

$$\delta_0 = f(\delta_1, \delta_2, \delta_3, ..., \delta_n) \tag{2}$$

The technique can work even without taking δ_0 into account so that there is no overall rating, only ratings of other attributes will be used to undertake the operation process. However, evidence observed from many researchers confirmed the efficiency of considering the overall rating than ignoring it [1].

Nevertheless, neural network is one of the powerful classes of machine learning models that can learn a complicated function from a data to solve many optimization problems. A neural network aimed to mimic the functions of biological neurons [6]. It contains sets of connected neurons arranged in a layered style (Figure. 1), where an input layer consist of neurons that receive input from external environment and the output layer neuron receives the weighted sums of the products of input values and their corresponding weights from the previous layer and send its computational result to the outside environment.

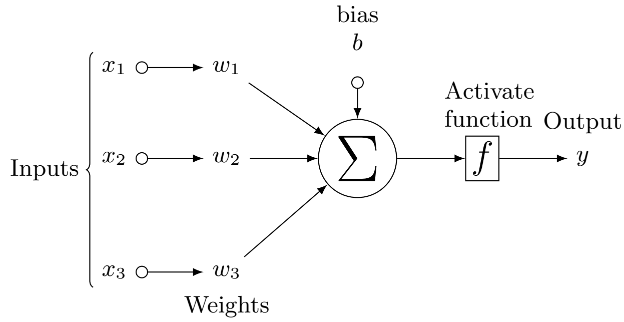

Figure 1: Simple Architecture of Feedforward Network

The features x_1, x_2, and x_3 in Figure. 1 are inputs presented to the input layer, the parameters ω_1, ω_2, and ω_3 are the synaptic weights for links between input and output neurons. $\sum$ is the weighted sum of $\omega_i x_i$ for $0 \leqslant i \leqslant 3$ including the bias b and f is an activation function (nonlinear differentiable function) that estimate the output y of the network. It could also be said that the output y can be written as $f(\sum_{i=0}^{n} \omega_i x_i)$. Feedforward network may contain more than two layers, where hidden layer(s) can be added between the input and output layers. Before the network gives the optimal output, it has to undergo a series of training using recommended algorithms. This is a brief abstract notion of how neural network behaves, details of this process can be found in [6, 19].

3. EXPERIMENT

The experiment was carried using yahoo movie datasets [11] for multi-criteria movie recommendation system, where movies are recommended to users based on four characteristics of movies, namely, action, story, direction, and visual effect of the movie which are represented as c_1, c_2, c_3, and c_4 respectively. In addition to those four criteria, an additional rating c_o called overall rating criterion was used to represent the final user's preference on a movie. The criteria values (ratings) in the dataset were initially presented using 13-fold quantitative scale from A+ to F representing the most and the worst preferences of the user. In the same manner, we changed the rating representation to numerical form (13 to 1 instead of A+ to F), and a total of approximately 63,000 ratings were used for the study. The target of the study was to use feedforward network to learn how to estimate c_o from c_1, c_2, c_3, and c_4. The dataset was divided into training and test data in the ratio of 75:25 for all the two experiments.

Two feedforward networks were developed using object oriented programming techniques in java [10] with learning capacities in delta rule and backpropagation. The Adaline network consists of input and output layer as in Figure 1, with input layer containing four neurons and a bias for passing the data to the output layer. A linear activation function f was used in the output neuron to process the weighted sum $(\sum_{i=1}^{5} x_i \omega_i)$ of the inputs x_i received from input layer. Furthermore, in addition to the two layers in the Adaline, a network containing additional hidden layer with the same number of neurons as the input layers was used for backpropagation training with additional activation function g (sigmoid function) that receives the weighted sum from the input layer and send the result of its computation to the output neuron. For measuring the training and test error, a mean square error ($MSE = \frac{1}{2N}\sum_{j=1}^{N}(y_j - o_j)^2$) for real output o_j and the estimated output y_j, were used to compute the errors. Pearson correlation coefficient ($Corr = \frac{\sum(y_j-\overline{y})(o_j-\overline{o})}{\sqrt{\sum(y_j-\overline{y})^2}\sqrt{\sum(o_j-\overline{o})^2}}$) were also used as a metric for measuring the relative relationship between the real and estimated output for the test data.

Finally, in order to make the training faster and to avoid the chances of getting stuck in local optima, the input data were normalized to real numbers within the interval (0, 1] instead of between 1 and 13 inclusive.

4. RESULT AND DISCUSSION

In each of the two algorithms, neurons weights ω_i were initially generated at random and the network compute the outputs and the corresponding errors (as $\frac{1}{2}(y_j - o_j)^2$). Iteratively, the algorithms search for a set of weights ω_i $i = 1, 2, ..., 5$ that minimize the error. The adaptive linear neuron (Adaline) network trained using delta rule shows a quick convergence within a few number of iterations (less than 10 iterations) with a very performance. On the other hand, backpropagation algorithm prolongs the learning process where a large number of training cycles (epochs) have been used to monitor its performance and the result is presented in Figure 2. This figure shows the average MSE for the various number of training cycles. It is apparent from the figure that the number of training cycles is inversely proportional to the errors observed. It shows that the convergence can only be attained at a very high number of iterations. However, for the purpose of comparison, the number of training cycle was set 10,000 cycles (epoch = 10,000), the training error and correlation between the real and estimated output of the test set for the two algorithms are shown in Table.1.

Table 1: **Performance Statistics**

Algorithm	Number of Iterations	Average Training MSE ($\times 10^{-3}$)	Percentage correctness
Adaline	10	5.34	94.4%
BPA	10,000	7.30	90.0%

5. CONCLUSION AND FUTURE WORK

This study was carried out to investigate the relative performance of single layer and multilayer feedforward network trained using delta rule and backpropagation algorithm respectively. The performance of each model was measured using MSE for the training and the percentage of the correct predictions were evaluated on the test data using Pearson correlation coefficient. From Figure.2 and Table.1, it can be seen that backpropagation algorithm has a greater demand for longer training cycles to converge. Moreover, the results indicate that the one layer network trained using adaptive linear neuron algorithm is more efficient than the two layer network which supports the traditional belief that single layer network produces less error than multilayered network [18]. Up to our last experiment with epochs of 10,000, backpropagation did not completely show the final convergence, further investigation is recommended to estimate the approximate epochs required by the algorithm to converge and to know whether it will produce a better result than Adaline. The study confirmed the usefulness of training a neural network model with features of inputs obtained for predicting the user preferences on items based several characteristics of items in multi-criteria recommender systems.

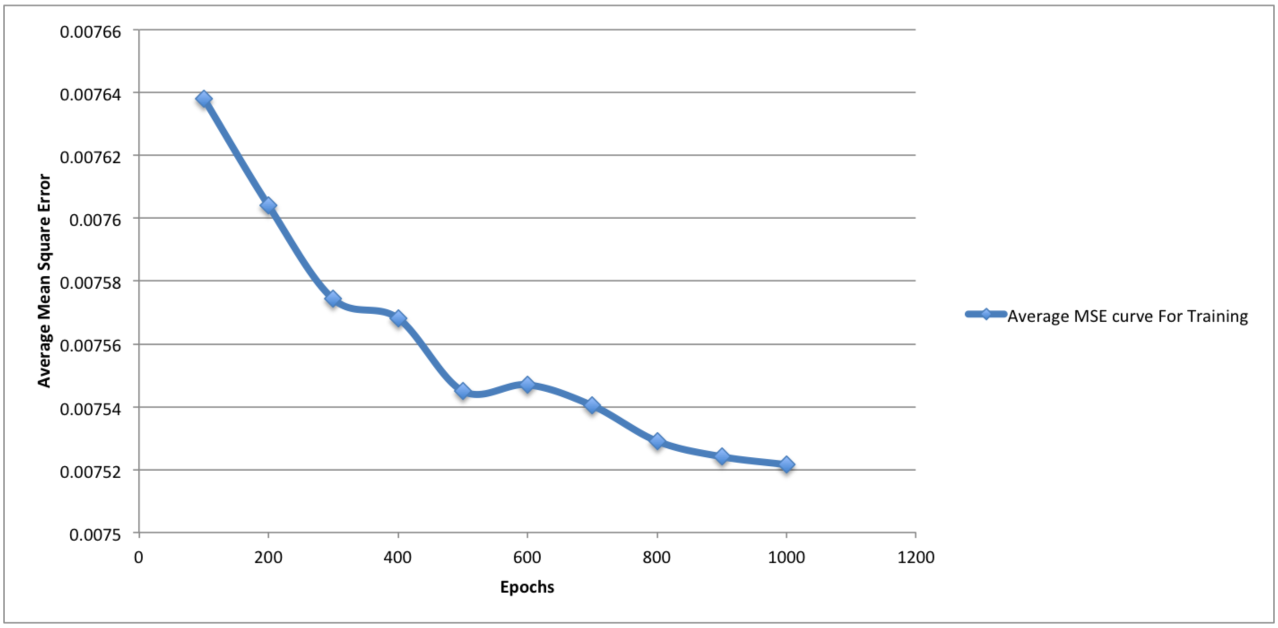

Figure 2: Average training MSE for Backpropagation

REFERENCES

[1] G. Adomavicius and Y. Kwon. New recommendation techniques for multicriteria rating systems. *IEEE Intelligent Systems*, 22(3):48–55, 2007.

[2] G. Adomavicius, N. Manouselis, and Y. Kwon. Multi-criteria recommender systems. In *Recommender systems handbook*, pages 769–803. Springer, 2011.

[3] G. Adomavicius, N. Manouselis, and Y. Kwon. Multi-criteria recommender systems. In *Recommender systems handbook*, pages 854–887. Springer, 2015.

[4] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749, 2005.

[5] N. Farokhi, M. Vahid, M. Nilashi, and O. Ibrahim. A multi-criteria recommender system for tourism using fuzzy approach. *Journal of Soft Computing and Decision Support Systems*, 3(4):19–29, 2016.

[6] D. Graupe. *Principles of artificial neural networks*, volume 7. World Scientific, 2013.

[7] M. Hassan and M. Hamada. Recommending learning peers for collaborative learning through social network sites. *IEEE ISMS, Intelligent Systems, Modeling and Simulation*, 2016.

[8] D. Jannach, F. Gedikli, Z. Karakaya, and O. Juwig. *Recommending hotels based on multi-dimensional customer ratings.* na, 2012.

[9] D. Jannach, Z. Karakaya, and F. Gedikli. Accuracy improvements for multi-criteria recommender systems. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 674–689. ACM, 2012.

[10] S. Kendal. *Object Oriented Programming using Java.* Bookboon, 2009.

[11] K. Lakiotaki, N. F. Matsatsinis, and A. Tsoukias. Multicriteria user modeling in recommender systems. *IEEE Intelligent Systems*, 26(2):64–76, 2011.

[12] T. Mahmood and F. Ricci. Improving recommender systems with adaptive conversational strategies. In *Proceedings of the 20th ACM conference on Hypertext and hypermedia*, pages 73–82. ACM, 2009.

[13] N. Manouselis and C. Costopoulou. Analysis and classification of multi-criteria recommender systems. *World Wide Web*, 10(4):415–441, 2007.

[14] X. Ning, C. Desrosiers, and G. Karypis. A comprehensive survey of neighborhood-based recommendation methods. In *Recommender systems handbook*, pages 37–76. Springer, 2015.

[15] K. Palanivel and R. Sivakumar. A study on collaborative recommender system using fuzzy-multicriteria approaches. *International Journal of Business Information Systems*, 7(4):419–439, 2011.

[16] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim. A literature review and classification of recommender systems research. *Expert Systems with Applications*, 39(11):10059–10072, 2012.

[17] F. Ricci, L. Rokach, and B. Shapira. Recommender systems: introduction and challenges. In *Recommender Systems Handbook*, pages 1–34. Springer, 2015.

[18] A. M. Souza and F. M. Soares. *Neural network programming with Java.* Packt Publishing Ltd, 2016.

[19] P. D. Wasserman. *Advanced methods in neural computing.* John Wiley & Sons, Inc., 1993.