Evaluating the performance of a neural network-based multi-criteria recommender system

Mohammed Hassan*

Department of Software Engineering,
Bayero University,
Kano, Nigeria
Email: mhassan.se@buk.edu.ng
*Corresponding author

Mohamed Hamada

Software Engineering Laboratory,
University of Aizu,
Aizuwakamatsu-city, Fukushima, Japan
Email: hamada@u-aizu.ac.jp

Abstract: Frequent use of internet applications and rapid increase in volumes of resources have made it difficult for online users to effectively make decisions on the kinds of information or items to select. Recommender systems (RSs) are intelligent decision-support tools that exploit preferences of users and suggest items that might be interesting to them. RSs are one of the various solutions proposed to address the problems of information overload. Traditionally, RSs use single rating techniques to predict and represent preferences of users for items that are not yet seen. Multi-criteria RSs use multiple ratings to various attributes of items for improving prediction and recommendation accuracy of the systems. However, one major challenge of multi-criteria RSs is the choice of an efficient approach for modelling the criteria ratings. Therefore, this paper aimed at employing artificial neural networks to model the criteria ratings and determine the predictive performance of the systems based on aggregation function approach. Seven evaluation metrics have been used to evaluate and measure the accuracy of the systems. The empirical results of the study have shown that the proposed technique has the highest prediction and recommendation than the corresponding traditional technique.

Keywords: recommender systems; artificial neural networks; ANNs; spacio-temporal data science; prediction accuracy; aggregation function; multi-criteria recommendation.

Biographical notes: Mohammed Hassan obtained his PhD from the University of Aizu in 2018. His research interests are in multi-criteria decision making and learning technology. He obtained his Master's in Computer Science from African University of Science and Technology, Abuja, Nigeria in 2011, and Bachelor's in Mathematics from Bayero University, Kano, Nigeria in 2009.

Mohamed Hamada is currently a Senior Associate Professor working in the Software Engineering Lab, at the University of Aizu, Aizuwakamatsu, Fukushima, Japan. He obtained his PhD from the University of Tsukuba, Japan. His research interest is in artificial intelligence and the study of functional logic languages, natural inspired computation, cloud computing-based e-learning and m-learning systems, multimedia, and web-based and advanced learning technologies. He is also interested in smart devices (such as smartphones and tablets) applications development and innovation, and so on. He is a regular Visiting Professor at Fatih University, Istanbul, Turkey and the African University of Science and Technology, Abuja, Nigeria.

1 Introduction

Recommender systems (RSs) help to provide personalised recommendations of items to online customers for reducing the problems of information overload. They are intelligent systems used by many online shopping sites like http://www.amazon.com, http://www.netflix.com, etc. to market their products to consumers (Hassan and Hamada, 2016a, 2017a). The primary aim of RSs is to provide users with personalised recommendations of online services or items. Besides product recommendations to online users, RSs have recently received considerable attention in various fields of our daily life. For example, in the area of technology-enhanced learning, RSs are used to recommend suitable learning object to online learners (Song et al., 2016; Hassan and Hamada, 2015, 2016b, 2017b). In the tourism domain, RSs are used as a guide to the tourist and for hotel recommendations (Nilashi et al., 2015). Other areas of application of RSs include restaurant guide (Xiang et al., 2007), e-governance in smart-cities (Cortés-Cediel et al., 2017; Reiterer et al., 2015), social network sites (Sun et al., 2015), and so on.

The systems applied some data mining and machine learning techniques to the rating history of users to predict their preferences on items that are not yet seen by the users. Different techniques have been used to design RSs. The four basic techniques used are collaborative filtering (CF), content-based filtering, knowledge-based filtering, and a hybrid-based filtering technique, which combines two or more filtering techniques in different ways (Adomavicius and Kwon, 2015; Hassan and Hamada, 2017c).

Majority of the existing RSs use single rating value to represent the preferences of users on items. The rating can be numerical (e.g., 0 to 5), binary (such as like or dislike), and so on. However, the idea of using the single rating values was considered by several works on RSs as limited, since single rating values cannot represent various attributes of the items (Elahi et al., 2016; Hassan and Hamada, 2017d). Consequently, a multi-criteria recommendation technique was proposed to use multiple ratings assigned to some of the important attributes of items for improving the prediction accuracy of the systems. Although the proposed multi-criteria technique has the potentials to improve the accuracy of the system significantly, the question now is, how can we model the criteria rating

information to determine the final opinion of the user? Along with this development in RSs, however, there is increasing concern over using machine learning algorithms to develop a model that could incorporate the criteria ratings for estimating the final opinions of the users based on their responses to the criteria ratings (Elahi et al., 2016; Hassan and Hamada, 2017c). To date, there has been little work that explored machine learning techniques to model multi-criteria RSs. This paper examines the significance of artificial neural networks (ANNs) in improving the prediction accuracy of multi-criteria CF RSs. The accuracy of the proposed approach has been tested using various values of N for top-N recommendations. A comparative analysis of the two results has been conducted, and a summary of the empirical results of the study has been presented. This current paper has been divided into four sections including this present section. Section 2 gives a brief overview of CF RSs and multi-criteria recommendations. Section 4 presents the experiments and the empirical results of the study. Finally, Section 5 contains the conclusion and potential areas for future research.

2   Background

This section introduces some of the related background and a summary of related works in multi-criteria recommender systems. Section 2.1 gives a general overview of neighborhood-based CF technique. Multi-criteria recommendation technique is summarised in Sections 2.2 and 2.3 presents a panorama of ANNs. Finally, a review of some related works in the domain of multi-criteria recommendation is given in Section 3.

2.1   CF technique

CF RSs use the wisdom of crowds to predict unknown ratings to users based on the ratings of other online users with similar tastes. This technique stores rating in the form of a $m \times n$ matrix for m number of users and n number of items (Hassan and Hamada, 2016c, 2017d). The CF technique is subdivided into neighbourhood- and model-based techniques, where the former is further subdivided into item-based and user-based approaches. The model-based technique is sometimes called a latent factor model, which build a prediction model to estimate the unknown ratings. This study was conducted based on the neighbourhood-based technique. Therefore, from now on, the acronym CF will be used throughout this paper to refer to the single rating neighbourhood-based RSs. CF use two sets of information: the information about ratings that the active user gave to other items, and the information about ratings given to the potential item by users who have similar opinions with the active user. This can be shown symbolically as follows: let r_{ui} be the rating to item i by user u, and let r_{vi} be the rating to item i by another user v. Then the unknown rating $\widehat{r}_{uj}$ for any new j for u is given as in equation (1), where $\overline{r}_u$ is the average of the ratings of u on other items, $N_i(u)$ is the set of users who previously co-rated i with u and also gave ratings to j, and $sim(u, v)$ is the degree of the similarity between u and v, which can be obtained using any of the well-known similarity metrics like Pearson correlation coefficient in equation (2), with I_{uv} as the set of items rated by both u and v.

$$\widehat{r}_{uj} = \overline{r}_u + \frac{\sum_{v \in N_i(u)} sim(u, v)(r_{vj} - \overline{r}_v)}{\sum_{v \in N_i(u)} |sim(u, v)|} \tag{1}$$

$$sim(u,v) = \frac{\sum_{i \in I_{uv}}(r_{ui} - \overline{r}_u)(r_{vi} - \overline{r}_v)}{\sqrt{\sum_{i \in I_{uv}}(r_{ui} - \overline{r}_u)^2 \sum_{i \in I_{uv}}(r_{vi} - \overline{r}_v)^2}} \qquad (2)$$

2.2 Multi-criteria recommendation

The CF recommenders discussed previously and the content-based recommenders are mostly categorised as traditional RSs as they use only a single rating to recommend items to users. Multi-criteria RSs extend the single rating CF by allowing the user to choose items based on various characteristics of the items. Subsequently, the utility function of CF: $f : u \times i \to r$ has been extended to $f : u \times i \to r_o, r_1, ..., r_n$, to account for all the criteria ratings; where r_o is the overall rating, which in some cases depends on the criteria ratings $r_1, r_2, ..., r_n$.

Modelling the criteria ratings to determine the value of r_o could be achieved using either heuristic or model-based approach (Hassan and Hamada, 2016c). The heuristic-based approach extends the traditional similarity-based CF by using similarities between users based on each criterion to compute r_o, while the model-based approach assumes a relation between r_o and r_i for $i = 1, 2, ..., n$. One of the examples of the model-based approach is the aggregation function method in equation (3), which uses machine learning algorithms or statistical techniques to learn the relationship between the r_o and r_i for $i = 1, 2, ..., n$.

$$r_o = f(r_1, r_2, ..., r_n) \qquad (3)$$

Different machine learning techniques and statistical methods such as a support vector machine (Jannach et al., 2012), multi-linear singular value decomposition (Li et al., 2008), probabilistic method (Sahoo et al., 2012), etc. have bee applied to the aggregation function approach in equation (3) for estimating r_o. In contrast to model-based approach, heuristic-based approach model the similarities between criteria ratings to estimate r_o, hence it is sometimes called similarity-based approach. Some of the techniques applied to heuristic-based method include, a worst-case similarity [see equation (4)], average similarity [see equation (5)], and other statistical methods, where $\widetilde{sim}(u,v)$ is called the overall similarity, which is used to replace $sim(u,v)$ in equation (1) for calculating $\widehat{r}_{uj}$. Note also that each of the $sim_i(u,v)$ is estimated using equation (2).

$$\widehat{sim}(u,v) = \min_{i=0,1,...,n} sim_i(u,v) \qquad (4)$$

$$\widehat{sim}(u,v) = \frac{1}{n+1} \sum_{i=0}^{n} sim_i(u,v) \qquad (5)$$

2.3 Artificial neural networks

In McCulloch and Pitts (1943) published a paper which they introduced the first model of ANNs based on a single-layer neural network (Zhou et al., 2016). ANNs belong to a powerful class of machine learning algorithms that learn complicated patterns from data using collections of simple trainable mathematical functions. While variety of definitions of ANNs have been suggested, we followed the definition given by the inventor of one of the first neuro-computers (Caudill, 1988). He saw it as "a computing system made up of a

number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs".

ANNs mimic the structure of the human brain, which consists of several processing units called neurons. The neurons are made up of dendrites for receiving signals and axon for firing the output signals further (Shiffman, 2012). Human brain comprised of billions of neurons that are interconnected in a specific pattern for helping the regular functioning of organs and the entire human body.

Similarly, ANNs are made up of two or more interconnected layers of neurons that take inputs from the external environment through an input layer and send the results to the external environment through an output layer as an output. In most cases, the input layer is connected to a hidden layer(s) that is/are between the input and output layers. Neurons between layers are connected using a system of weighted connections similar to synapses in the biological human brain.

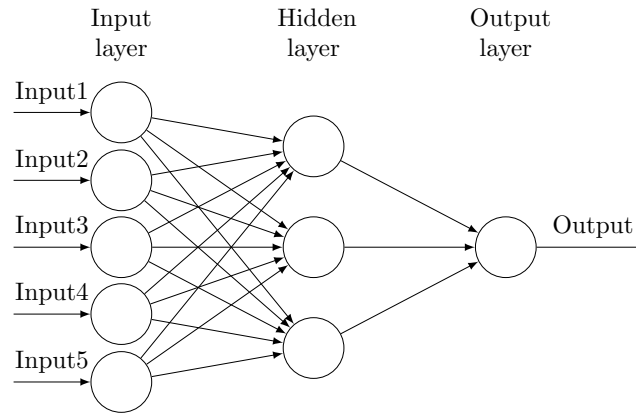**Figure 1**   Example of two-layer neural networks

Figure 1 is an example of multi-layered feed-forward neural network consisting of an input layer, a hidden layer, and an output layer. Each neuron in the hidden is connected to all neurons in the input layer and receives signals from them. Similarly, the neuron in the output layer is connected to the neurons in the hidden layer and receives weighted sum of their computational results.

To explain how neurons receive, manipulate, and send signals to other neurons in the subsequent layer, let x_{ij}^k be the signal sent to neuron j at layer k by neuron i at layer $k-1$, and let ω_{ij} be the synaptic weight between i and j, then the total input received by i is the weighted sum of all the signals sent from neurons of the previous layer multiplied by their corresponding weights. This can be formulated mathematically as $\sum_{i=0}^{n} x_{ij}^k \omega_{ij}$ where $i = 0$ is for the bias neuron, and n is the number of neurons in the previous layer. Further, if we let the z_j^k to represent the calculated weighted sum as shown in equation (6), then the output y_j^k of the neuron j at k will now be determined by applying an activation function f on z_j^k given by equation (7).

$$z_j^k = \sum_{i=0}^{n} x_{ij}^k \omega_{ij} \tag{6}$$

$$y_j^k = f(z_j^k) \tag{7}$$

For the activation function (sometimes called transfer function) on the right-hand side of equation (7), there are different mathematical functions that can perform appropriate mathematical operations on the z_j^k. They include sigmoid function [equation (8)], which forces the value to be between 0 and 1. Others are linear function [equation (9)] (where c is a scalar), the tanh function [equation (10)], and so on (Tong and Mintram, 2010).

$$f(z_j^k) = \frac{1}{1 + e^{-z_j^k}} \tag{8}$$

$$f(z_j^k) = z_j^k + c \tag{9}$$

$$f(z_j^k) = \frac{e^{z_j^k} - e^{-z_j^k}}{e^{z_j^k} + e^{-z_j^k}} \tag{10}$$

Moreover, the output $widehaty$ of the *ANNs* and the required output y from the training data can be used together with an objective function $g(\widehat{y}, y)$ of equations (11) or (12) to determine the current accuracy of the network. The error can be minimised by updating the synaptic weights of the network. The technique used in updating the weights is called the training algorithm. Various algorithms such as backpropagation algorithm, simulated annealing algorithm, genetic algorithm, particle swarm optimisation (PSO) algorithm, etc. can be deployed to train the network.

$$g(\widehat{y}, y) = \sqrt{\frac{1}{2M} \sum_{k=1}^{M} (\widehat{y}_k - y_k)^2} \tag{11}$$

$$g(\widehat{y}, y) = -\frac{1}{2M} \sum_{k=1}^{M} [y_k log\widehat{y}_k - (1 - y_k)log(1 - \widehat{y}_k)] \tag{12}$$

3 Related work

Fuzzy sets are considered to be among the suitable paradigms applied to manage uncertainty problems and for modelling users' behaviours in multi-dimensional decision-making problems. Palanivel and Sivakumar (2010) proposed a fuzzy-based multi-criteria CF, which uses a fuzzy linguistic technique to represent ratings of users systematically, and fuzzy multi-dimensional decision-making technique for ranking the relevant items for users. The proposed approach was developed and experimented with a dataset for multi-criteria music recommender system.

Similarly, Farokhi et al. (2016) proposed a multi-criteria RS for tourism domain using fuzzy-based approach. In their proposed method, they used both item- and user-based CF techniques and applied to them different similarity metrics. Furthermore, they applied fuzzy clustering (C-means) on each technique and analysed their prediction accuracy. The experiments were conducted using a multi-criteria dataset of real-users from a tourism domain (www.tripadvisor.com). In Shambour et al. (2016), an item-based multi-criteria

RS was proposed for enhancing personalised recommendations of items. The approach integrates items' semantic information and the criteria ratings of the items for addressing some of the outstanding limitations of the traditional item-based CF technique. For evaluating the accuracy of the proposed method, they used a Yahoo!Movie dataset for recommending movies to online users. The same approach was followed in Shambour (2016) using a user-based CF technique. Musto et al. (2017) have recently proposed another multi-criteria RS based on CF technique that uses information conveyed from reviews of users for providing a multi-faceted representation of interests of the users. This was done by using opinion mining frameworks and sentiment analysis techniques to derive relevant aspects and sentiment scores of the users' reviews automatically. The experiments were carried out using three multi-criteria datasets from tripAdvisor (http://www.cs. cmu.edu/jiweil/html/hotel-review.html), Yelp (https://www.kaggle.com/c/yelp-recruiting/ data), and Amazon (http://jmcauley.ucsd.edu/data/amazon/links.html). Furthermore, a hybrid multi-criteria RS that uses neuro-fuzzy and ontology techniques was proposed by Kermany and Alizade (2017). The proposed method integrates item-based filtering technique and demographic information for movies recommendations. They suggested applying an adaptive neuro-fuzzy inference system for finding the relationship between criteria ratings and the overall rating. Similar to this approach, Nilashi et al. (2015) come up with a multi-criteria recommendation framework for tourism domain that combines expectation maximisation, adaptive neuro-fuzzy inference system, and principal component analysis techniques.

Moreover, in addition to using fuzzy approach, research has recently begun to apply some optimisation algorithms such as PSO algorithms for modelling the criteria ratings have started evolving. In Choudhary et al. (2017a), PSO algorithm together with some effective similarity measures has been used for handling natural noise in multi-criteria RSs and for improving their accuracy. Furthermore, PSO algorithm was used in Choudhary et al. (2017b) for learning optimal weights of the similarities between the criteria ratings to estimate the overall rating. Several other studies have explored genetic algorithms (Hwang, 2010; Hassan and Hamada, 2017e) in different ways to model the criteria rating information.

4   Experiment and findings

The experiment was conducted using a Yahoo!Movie dataset for recommending movies to users based on four attributes of the movies. The four attributes used for recommendation of movies are, the action, the story, the direction, and the visual effect of the movies (Lakiotaki et al., 2011). Note that the kinds and number of criteria used in the experiment depend on the design of the original RS where the dataset was collected. Therefore, different datasets collected from different sources might contain additional or entirely different criteria values. The dataset used in this experiment contains ratings for these four attributes of the movies and an additional overall rating. The movies were rated on a scale of 13, from A^+ to F representing the highest and the lowest preferences respectively. To work with the numerical ratings, the dataset was transformed into numbers from 13 to 1 representing A^+ to F respectively. Table 1 contains the original and numerical samples of the experimental dataset used in this study, where the first three rows show the original dataset extracted from Yahoo!Movie website (Lakiotaki et al., 2011), and the last three rows show the numerical representation of the ratings shown in the first three rows. To obtain the numerical ratings,

the original ratings are transformed as follows: $A^+ \rightarrow 13$, $A \rightarrow 12$, $A^- \rightarrow 11$, $B^+ \rightarrow 10$, $B \rightarrow 9$, $B^- \rightarrow 8$, ..., $D^- \rightarrow 2$, $F \rightarrow 1$. Column one of the table contains the user ID, which in this case we took ratings of a user whose ID is 101. The movie ID in the second column represents different movies that the user gives ratings to them. Column three to column seven are the ratings of the user to the movies based on the four criteria (r_1 to r_4) and an overall rating. After cleaning the dataset to remove unlabelled and incomplete entries, the dataset contains a total of 62,156 ratings to 976 different movies from 6,078 different users.

Table 1 Extracted dataset: orignal and numerical data

User ID	Movie ID	Direction (r_1)	Action (r_2)	Story (r_3)	Visual (r_4)	Overall (r_o)
101	1	A^+	C	C^-	B^-	C^-
	3	B	B^+	B^+	A^-	B
	5	B^-	A^-	B^+	A^-	A^-
101	1	13	6	5	8	5
	3	9	10	10	11	9
	5	8	11	10	11	11

Considering the size of the dataset, where the number of users is more than the number of movies, an item-based CF technique was used so that depending on the neighbourhood size; each user could have a neighbour. Using the item-based CF was mainly to prevent cases of 'no neighbours' in a particular neighbourhood. We used a neighbourhood size $k = 100$ and the ranking accuracy was evaluated based on top-N recommendation with different values of N.

Though our explanation of the ANN in Section 2.3 was based on multi-layered neural network, the neural network used to conduct this experiment was a one-layer network with five input neurons for the four criteria and a bias, and a single output neuron, which was trained using a delta rule technique. The experimental parameters such as the learning rate, the maximum number of iterations, and the target error were selected appropriately to obtain the empirical result of the study. We performed several test-runs to choose these parameters. We selected a learning rate $\delta = 0.01$, the target error $\varepsilon = 0.1$, the maximum number of iterations (epoch = 100), and the stopping condition depends on ε or epoch. Furthermore, since we used CF technique, it became necessary also to specify the neighbourhood size, the value of N for top-N recommendation, and various evaluation metrics employed (Jannach et al., 2013). Ten experiments were conducted with varying values of N (5, 10, 15, ..., 50).

In implementing the single rating and the proposed multi-criteria RS, we used Java programming language and followed the paradigms outlined for building traditional RS (Jannach et al., 2013) and multi-criteria RS using aggregation function approach (Adomavicius and Kwon, 2007). The summary of the four steps followed to model the proposed MCRSs using aggregation function approach is presented in Figure 2. The pseudocode of concept presented in the figure is also shown in Algorithm 4. The figure and the pseudocode give an overview of how we integrate the traditional single rating CF with ANNs to provide top-N recommendations. However, to get the general overview for a better understanding of the mechanisms of the aggregation function, and detailed discussions of previous works, the reader could also refer to Adomavicius and Kwon (2007, 2015) respectively.

Moreover, to evaluate the prediction and recommendation accuracy of the proposed system, the following evaluation metrics have been implemented to measure and

compare the accuracy of the proposed technique with the accuracy of the traditional neighbourhood-based CF.

1   RMSE: root mean square error is a standard deviation of the prediction errors. For a prediction $\widehat{y}_k$ and the actual value y_k for the kth feature from the dataset, RMSE is computed using equation (11).

2   MAE: mean average error is also a measure of the deviation of prediction errors, which is considered as the most natural measure of prediction errors (Willmott and Matsuura, 2005). Equation (13) shows how to compute the prediction errors using MAE.

$$MAE = \frac{1}{M} \sum_{k=1}^{M} |\widehat{y}_k - y_k| \tag{13}$$

3   Precision: it is an accuracy metric that measures the relevancy of the recommendations made from the top-N recommended items.

4   Recall: it measures the ability of the system in recommending only the relevant items.

5   F_1 score. it is a measure of accuracy that combines both precisions and recalls to compute the score. It is usually interpreted as the weighted average of the precision and recall [see equation (14)].

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{14}$$

6   Fraction of concordant pairs (FCP). It is a ranking evaluation metric that looks at the fractions of all the pairs that the system ranked them in a correct order.

7   Area under the curve of the receiver-operating characteristic (ROC) written as ROCAU.

The above metrics are commonly known and used in most of the research in RSs. We therefore explained them briefly. From metric number 4 to number 7, they usually provide a value between 0 and 1, between 0% to 100%, and the higher the value, the better the accuracy. However, in the case of RMSE and MAE, the smaller value signifies, the better prediction accuracy, and their values are not bounded within an interval.

Figure 2   A sketch of the approach of the proposed method (see online version for colours)
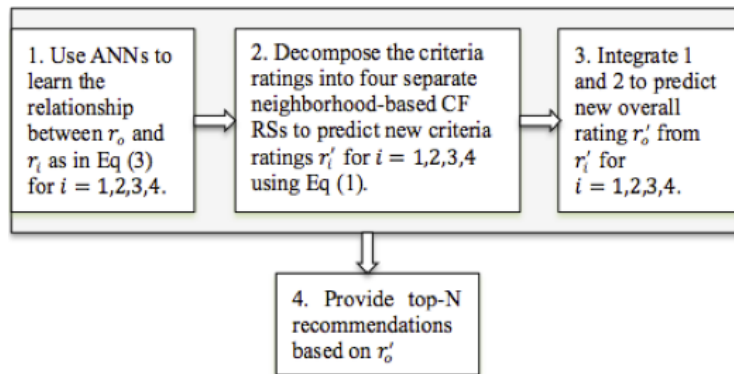
Algorithm 1   Pseudocode of the proposed approach

    Input: *Multi-criteria dataset*
    Output: *List of top-N recommended items*
1: **Begin algorithm**
2:   **Phase 1**: *training the ANN*
3:   **Step 1**: *From the dataset, from the Item- and User-ID*
4:   **Step 2**: *Each critirion rating is an input given to an input neuron*
5.   **Step 3**: The sample outputs are the overall rating r_o
6.   **Step 4**: *Lean the relationship shown in equation (3)*
6:   **Phase 2**: *Predicting single criterion rating*
7:   **Step 1**: *Decompose the criteria ratings*
8:   **Process:** *Each component is of the form:*
9:   *User_ID x Item_ID ➔ r_i*
10:   **Step 2**: *Develop 4 KNN_based Single rating RSs*
11:   **Process 1**: *for i from 1 to 4*
12:   *KNNi Learn to predict r_i*
13:   **Phase 3**: *Integrate ANN and the 4 KNN*
14:   **Process 1**: *Use the Learn ANN from Phase 1 and the trained of step 2 in phase 2 and integrate them*
15:   **Phase 3**: *Recommendation*
16:   **Step 1**: *Generate the List of recommendations*
17:   **Step 2**: *Select the first N-items as the Top-N recommended set to the active user*
18: **End algorithm**

To analyse the results obtained from our experiment, Table 2 presents the results of the experiment based on the neighbourhood size of 100 and top-10 recommendations. Although there are several possible explanations for this result, the single most striking observation to emerge from the data comparison was the significant differences between the two RSs, which shows that the proposed ANN-based MCRS performs better than the single rating RS. The last column of the table shows the percentage improvements, obtained by calculating the difference between the two RSs divided by the performance of single rating CFRSs. For instance, in terms of RMSE, the single rating CF has 2.960, and the proposed MCRS has 2.153, then the percentage improvement is $\frac{2.153-2.960}{2.960} = -27.26\%$ The negative sign indicates a decrease in RMSE. The same procedure was followed to compute the percentage improvements on the remaining evaluation metrics. The possible explanations for these results are:

- The decrease in MAE and RMSE show that the predictions of the proposed MCRS are much closer to the actual ratings from the dataset than the predictions of the traditional CF.

- The improvements in the classification accuracy (precision, recall, and F_1) show the increase in accuracy of the system to make correct decisions on whether the item recommended is relevant for the recommendation or not.

- In terms of ordering the recommended items based on their degree of relevancies, the proposed ANN-based MCRS also has the high ability to provide a good ranking of items that could match the same way the user might have ordered them.

The experiments were repeated several times by changing the value of N to see whether there will be an instance where the proposed MCRS might fail to outperform the corresponding single rating technique. Fortunately, The change in N did not cause the accuracy of the proposed system to be lower than that of the single rating technique. In fact, as expected, for all values of N used, the proposed approach outperformed the single rating technique. For example, as we can see in Figure 3, the precisions of the proposed MCRS have never been close to that of single rating technique. The figure also shows that the precision curve of the proposed technique is more resistant to the change of N than the

curve of the single rating technique. This indicates the consistency and resistivity of the proposed MCRS towards providing good recommendations and choosing the experimental parameters respectively.

Table 2 Prediction results

Mertics	Models		Differences (%)
	Single CF	MCRSs	
RMSE	2.960	2.153	−27.27%
MAE	2.076	1.521	−26.73%
Precision	0.893	0.916	2.58%
Recall	0.771	0.850	10.25%
F_1	0.828	0.882	6.57%
FCP	0.720	0.839	16.53%
ROCAU	0.633	0.744	17.20%

Figure 3 Curve of the precisions at top N (see online version for colours)
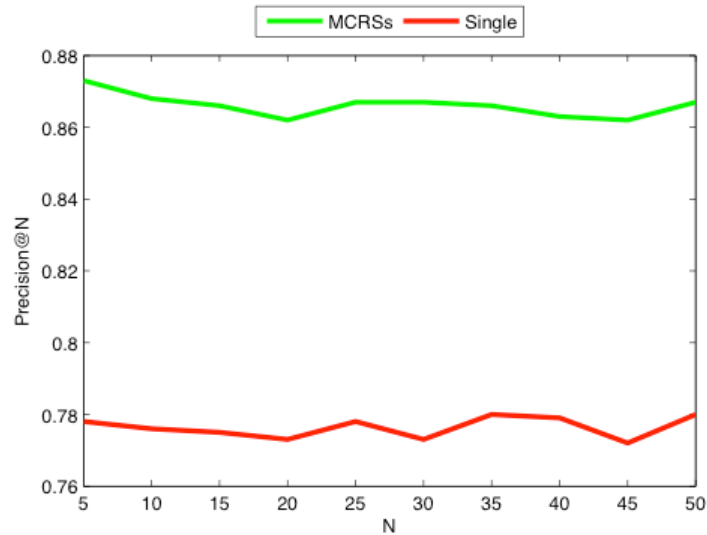
Furthermore, to test the correlation and the monotonic relationships between the predicted ratings and the corresponding actual ratings from the data set, Figure 4 shows some of the main characteristics of the predictions of the two techniques with respect to the actual values. The figure also illustrates that the predictions of the proposed ANN-based MCRSs are closer to the actual values than that of the single rating RSs. It can be seen from the figure that the predicted ratings of the MCRSs reported significantly more correlations with the actual ratings than those of the single technique. Moreover, it can be observed from the figure that in several occasions, the predictions of the single rating technique have either be much greater than the actual rating or much smaller than the actual ratings. Overall, the evidence presented in Table 2 and the two figures provide important insights into exploring ANNs into multi-criteria recommendation problems.

Figure 4 Graph of the actual and predicted ratings (see online version for colours)
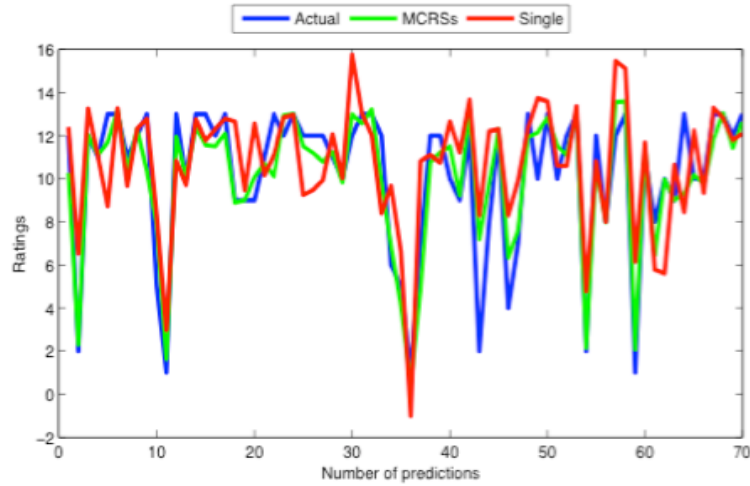
Finally, it is important to compare the performance of the proposed method with other existing methods. Several methods such as clustering- and regression-based methods (Nilashi et al., 2015), hybridisation of multi-criteria techniques (Nilashi et al., 2014), probabilistic latent semantics methods (Zhang et al., 2009), etc. have been successfully applied to multi-criteria user modelling. Recently, Zheng (2017) proposed another novel approach called 'criteria chain', for better prediction of users' preferences by using a combination of criteria ratings in the way of contextual situation. In his experiment, Zheng (2017) evaluated the accuracy of his proposed approach using two public datasets from different recommendation domain, the YahooMovie and TripAdvisor dataset for recommending hotels to users. The results are analysed for each of the datasets used. Based on the experiment, an improvement of 19.0% in terms of MAE has been recorded. Furthermore, far back 2009 when research in multi-criteria recommendation was in its infancy stage, Zhang et al. (2009) proposed a probabilistic latent semantic technique for modelling the criteria rating in multi-criteria recommendations. Their experiment yields a slightly smaller improvement than the criteria chain (Zheng, 2017). An improvement of 3.6% was observed. Together, our proposed ANN-based method has by far provide higher accuracy than these results. That is an improvement in accuracy of approximately 26.7% and 27.3% in MAE and RMSE respectively is much better than the 19.0% and 3.6% of Zheng (2017) and Zhang et al. (2009) respectively. The utility-based approach proposed by Lakiotaki et al. (2011) that worked based on preference disaggregation techniques provides good prediction accuracy. However, as the system was implemented using utility algorithm for modelling preferences of users by applying additive utility function, the performance of the model would be inefficient when there are few ratings for the active user. Though no experimented results presented in Teng and Lee (2007) which uses data query techniques to estimate preferences of users in multi-criteria recommendation problem since the recommendation problem was no longer recognised as an optimisation, no attention given to feature weights for efficient prediction of users' preferences. Furthermore, Hwang (2010) proposed a genetic algorithm-based model for feature weighing. Nonetheless, genetic algorithm-based model is subject to the following limitations: no guarantee of

finding global minima, it takes longer time for convergence, and difficulties in generating the fitness function could make the provide incomprehensible predictions (Hassan and Hamada, 2017e). Our proposed ANN-based has high fault tolerance which does not prevent the system to learn users' behaviour and predict the overall rating even if the criteria ratings are incomplete. This is due to the ability of neural networks to provide outputs even with incomplete inputs (Mathioulakis et al., 2018; Zhang, 2018; Hassan and Hamada, 2017c).

5   Conclusions and future works

In recent years, RSs have been recognised as software tools capable of solving problems of information overload. Current development in RSs research has come up with several techniques for improving the prediction accuracy of the RSs (Adomavicius and Kwon, 2012). Multi-criteria recommendation is a new technique that accounts for several characteristic/features of the items when estimating the preferences of users on items. The efficiency of MCRSs has been tested, and they are proved to provide high prediction and recommendation accuracy than single rating-based techniques. However, the key central question on MCRS is how to efficiently model the criteria ratings for accurate predictions of users' preferences. Adomavicius and Kwon (2015) challenge the RSs research community to explore some sophisticated machine learning algorithms such as ANN for modelling the MCRS using an aggregation function approach. The present study was designed to evaluate the effectiveness of ANNs in modelling multi-criteria recommendation problems. Two different models were designed: one of them uses single rating CF, and the other one integrates four single rating CF recommenders and ANN for developing the proposed MCRS. The study was conducted to evaluate the performance of the two models and compared their accuracy using seven evaluating metrics. The current findings support the relevance of ANN in improving the accuracy of MCRS.

However, this research has thrown up many questions in need of further investigation. For instance, as we used only one layer ANN, what about using multi-layered ANNs? And which training algorithms can provide better accuracy? Also what about using model-based single rating RSs such as a matrix factorisation SVD-based RSs? Further research is needed to account for the various ANNs architectures, the training algorithms for the ANNs, and different kinds of single rating RSs. Another possible area of future research would be to introduce deep learning technique into modelling the preferences of users in MCRS. This could be achieved by extracting a large volume of dataset and use it to train the networks to learn the relationships between criteria ratings information and the overall rating.

References

Adomavicius, G. and Kwon, Y. (2007) 'New recommendation techniques for multi-criteria rating systems', *IEEE Intelligent Systems*, Vol. 22, No. 3.

Adomavicius, G. and Kwon, Y. (2012) 'Improving aggregate recommendation diversity using ranking-based techniques', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 24, No. 5, pp.896–911.

Adomavicius, G. and Kwon, Y. (2015) 'Multi-criteria recommender systems', in *Recommender Systems Handbook*, pp.847–880, Springer.

Caudill, M. (1988) 'Neural networks primer, part 3', *AI Expert*, Vol. 3, No. 6, pp.53–59.

Choudhary, P., Kant, V. and Dwivedi, P. (2017a) 'Handling natural noise in multi-criteria recommender system utilizing effective similarity measure and particle swarm optimization', *Procedia Computer Science*, Vol. 115, pp.853–862.

Choudhary, P., Kant, V. and Dwivedi, P. (2017b) 'A particle swarm optimization approach to multi-criteria recommender system utilizing effective similarity measures', in *Proceedings of the 9th International Conference on Machine Learning and Computing*, ACM, pp.81–85.

Cortés-Cediel, M.E., Cantador, I. and Gil, O. (2017) *Recommender Systems for e-Governance in Smart Cities: State-of-the-Art and Research Opportunities*.

Elahi, M., Ricci, F. and Rubens, N. (2016) 'A survey of active learning in collaborative filtering recommender systems', *Computer Science Review*, Vol. 20, pp.29–50.

Farokhi, N., Vahid, M., Nilashi, M. and Ibrahim, O. (2016) 'A multi-criteria recommender system for tourism using fuzzy approach', *Journal of Soft Computing and Decision Support Systems*, Vol. 3, No. 4, pp.19–29.

Hamada, M. and Hassan, M. (2015) 'Learning system and analysis of learning style for African and Asian students', in *2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, IEEE, pp.83–87.

Hassan, M. and Hamada, M. (2016a) 'Recommending learning peers for collaborative learning through social network sites', in *2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, IEEE, pp.60–63.

Hamada, M. and Hassan, M. (2016b) 'A multimedia learning environment for information theory', in *2016 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, IEEE, pp.55–61.

Hamada, M. and Hassan, M. (2016c) 'Performance comparison of featured neural network trained with backpropagation and delta rule techniques for movie rating prediction in multi-criteria recommender systems', *Informatica*, Vol. 40, No. 4, p.409.

Hassan, M. and Hamada, M. (2017a) 'A neural networks approach for improving the accuracy of multi-criteria recommender systems', *Applied Sciences*, Vol. 7, No. 9, p.868.

Hamada, M. and Hassan, M. (2017b) 'An enhanced learning style index: implementation and integration into an intelligent and adaptive e-learning system', *Eurasia Journal of Mathematics, Science and Technology Education*, Vol. 13, No. 8, pp.4449–4470.

Hassan, M. and Hamada, M. (2017c) 'Performance comparison of feed-forward neural networks trained with different learning algorithms for recommender systems', *Computation*, Vol. 5, No. 3, p.40.

Hassan, M. and Hamada, M. (2017d) 'Smart media-based context-aware recommender systems for learning: a conceptual framework', in *2017 16th International Conference on Information Technology Based Higher Education and Training (ITHET)*, IEEE, pp.1–4.

Hassan, M. and Hamada, M. (2017e) 'Improving prediction accuracy of multi-criteria recommender systems using adaptive genetic algorithm', in *Proceedings of the SAI Intelligent Systems Conference (IntelliSys)*, IEEE, pp.326–330.

Hwang, C-S. (2010) 'Genetic algorithms for feature weighting in multi-criteria recommender systems', in *Journal of Convergence Information Technology*, Citeseer.

Jannach, D., Gedikli, F., Karakaya, Z., Juwig, O. et al. (2012) *Recommending Hotels Based on Multi-dimensional Customer Ratings*, na.

Jannach, D., Lerche, L., Gedikli, F. and Bonnin, G. (2013) 'What recommenders recommend – an analysis of accuracy, popularity, and sales diversity effects', in *International Conference on User Modeling, Adaptation, and Personalization*, pp.25–37, Springer.

Kermany, N.R. and Alizade, S.H. (2017) 'A hybrid multi-criteria recommender system using ontology and neuro-fuzzy techniques', *Electronic Commerce Research and Applications*, Vol. 21, pp.50–64.

Lakiotaki, K., Matsatsinis, N.F. and Tsoukias, A. (2011) 'Multi-criteria user modeling in recommender systems', *IEEE Intelligent Systems*, Vol. 26, No. 2, pp.64–76.

Li, Q., Wang, C. and Geng, G. (2008) 'Improving personalized services in mobile commerce by a novel multi-criteria rating approach', in *Proceedings of the 17th International Conference on World Wide Web*, ACM, pp.1235–1236.

Mathioulakis, E., Panaras, G. and Belessiotis, V. (2018) 'Artificial neural networks for the performance prediction of heat pump hot water heaters', *International Journal of Sustainable Energy*, Vol. 37, No. 2, pp.173–192.

McCulloch, W.S. and Pitts, W. (1943) 'A logical calculus of the ideas immanent in nervous activity', *The Bulletin of Mathematical Biophysics*, Vol. 5, No. 4, pp.115–133.

Musto, C., de Gemmis, M., Semeraro, G. and Lops, P. (2017) 'A multi-criteria recommender system exploiting aspect-based sentiment analysis of users' reviews', in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, ACM, pp.321–325.

Nilashi, M., Ibrahim, O.b. and Ithnin, N. (2014) 'Hybrid recommendation approaches for multi-criteria collaborative filtering', *Expert Systems with Applications*, Vol. 41, No. 8, pp.3879–3900.

Nilashi, M., Ibrahim, O.b., Ithnin, N. and Sarmin, N.H. (2015) 'A multi-criteria collaborative filtering recommender system for the tourism domain using expectation maximization (em) and pca–anfis', *Electronic Commerce Research and Applications*, Vol. 14, No. 6, pp.542–562.

Nilashi, M., Jannach, D., Ibrahim, O.b. and Ithnin, N. (2015) 'Clustering-and regression-based multi-criteria collaborative filtering with incremental updates', *Information Sciences*, Vol. 293, pp.235–250.

Palanivel, K. and Sivakumar, R. (2010) 'Fuzzy multi-criteria decision-making approachfor collaborative recommender systems', *International Journal of Computer Theory and Engineering*, Vol. 2, No. 1, p.57.

Reiterer, S., Felfernig, A., Jeran, M., Stettinger, M., Wundara, M. and Eixelsberger, W. (2015) 'A wiki-based environment for constraint-based recommender systems applied in the e-government domain', in *UMAP Workshops*.

Sahoo, N., Krishnan, R., Duncan, G. and Callan, J. (2012) 'Research note – the halo effect in multicomponent ratings and its implications for recommender systems: the case of Yahoo!Movies', *Information Systems Research*, Vol. 23, No. 1, pp.231–246.

Shambour, Q., Hourani, M. and Fraihat, S. (2016) 'An item-based multi-criteria collaborative filtering algorithm for personalized recommender systems', *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 8, pp.275–279.

Shambour, Q. (2016) 'A user-based multi-criteria recommendation approach for personalized recommendations', *International Journal of Computer Science and Information Security*, Vol. 14, No. 12, p.657.

Shiffman, D. (2012) *The Nature of Code: Simulating Natural Systems with Processing*, Daniel Shiffman.

Song, L., Tekin, C. and van der Schaar, M. (2016) 'Online learning in large-scale contextual recommender systems', *IEEE Transactions on Services Computing*, Vol. 9, No. 3, pp.433–445.

Sun, Z., Han, L., Huang, W., Wang, X., Zeng, X., Wang, M. and Yan, H. (2015) 'Recommender systems based on social networks', *Journal of Systems and Software*, Vol. 99, pp.109–119.

Teng, W-G. and Lee, H-H. (2007) 'Collaborative recommendation with multi-criteria ratings', *Journal of Computers*, Vol. 17, No. 4, pp.69–78.

Tong, D.L. and Mintram, R. (2010) 'Genetic algorithm-neural network (gann): a study of neural network activation functions and depth of genetic algorithm search applied to feature selection', *International Journal of Machine Learning and Cybernetics*, Vol. 1, Nos. 1–4, pp.75–87.

Willmott, C.J. and Matsuura, K. (2005) 'Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance', *Climate Research*, Vol. 30, No. 1, pp.79–82.

Xiang, Z., Kim, S-E., Hu, C. and Fesenmaier, D.R. (2007) 'Language representation of restaurants: implications for developing online recommender systems', *International Journal of Hospitality Management*, Vol. 26, No. 4, pp.1005–1018.

Zhang, Y., Zhuang, Y., Wu, J. and Zhang, L. (2009) 'Applying probabilistic latent semantic analysis to multi-criteria recommender system', *Ai Communications*, Vol. 22, No. 2, pp.97–107.

Zhang, Z. (2018) 'Artificial neural network', in *Multivariate Time Series Analysis in Climate and Environmental Research*, pp.1–35, Springer.

Zheng, Y. (2017) 'Criteria chains: a novel multi-criteria recommendation approach', in *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, ACM, pp.29–33.

Zhou, T., Gao, S., Wang, J., Chu, C., Todo, Y. and Tang, Z. (2016) 'Financial time series prediction using a dendritic neuron model', *Knowledge-Based Systems*, Vol. 105, pp.214–224.