Int. J. Nonlinear Anal. Appl. In Press, 1–14
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2023.32461.4828

Analysis of user behavior using hierarchical classification and fuzzy logic

Mohammad Hossein Hosseini, Azadeh al-Sadat Khodaei*

Department of Computer Engineering, Safadasht Branch, Islamic Azad University, Safadasht, Tehran, Iran

(Communicated by Majid Eshaghi Gordji)

Abstract

In today's age, the Internet has become the most important source of information and its supply for many goods and services due to its ease of use, wide range and high speed. The unique characteristics of the Internet and its complete superiority over other markets have led many organizations to expand their services and products in the Internet markets. Therefore, the Internet environment has become a very competitive environment for organizations. In the process of buying and checking user and consumer behavior, goods or services will be exchanged, and these goods or services can be information, physical or virtual products, and even emotions. Many kinds of research have been conducted in the analysis of users' behavior, but these researches were not highly accurate, therefore, in this research, an attempt was made to provide a solution for recommending items based on users' past behaviors to improve the accuracy of past methods. This proposed method can eliminate the noise in the data by using fuzzy logic and hierarchical analysis which is used for the phase of selecting the effective features and also works on more effective data to increase the accuracy. The tree proposed in this research has the lowest possible height and this causes the computational overhead to be greatly reduced. The proposed method was compared with several competing methods and the results indicate the superiority of the proposed method over the compared methods.

Keywords: user behavior, fuzzy logic, hierarchical analysis, intelligent recommender, modeling
2020 MSC: 03B52

1 Introduction

A correct understanding of customer behavior can lead to many advantages, which can be used to increase sales and more acceptance of products. What is exchanged in most social networks may not be physical, but it is equally or even more important. Many conclusions can be drawn from the study of user behavior in social networks. Also, understanding user behavior can help attract more users and use their information in social networks [1] The most important thing is to be able to find out what item a user is most interested in and this can be done by using The behavior of the user in the past is understood, that is, by analyzing the behavior of a user in the past, it is possible to find out which item the corresponding user pays more attention to and which item is more accepted by him. The user has already rated the items for the user, so the need for such an intelligent rating is strongly felt here.

Generally, a user is known based on his past and what he has done in the past. One of the ways of user modeling can be the manual extraction of heterogeneous characteristics of the user, which in this case may be due to the diversity

*Corresponding author
Email addresses: hosseim40@gmail.com (Mohammad Hossein Hosseini), aakhodaei@safaiau.ac.ir (Azadeh al-Sadat Khodaei)

Received: October 2023 *Accepted:* December 2023

of human instincts. Manually extracting user behavior is a challenging task. Therefore, it is generally difficult to understand user behavior, which is inherently heterogeneous and flexible. These days, for many organizations and companies, it is necessary to build a model for users based on the information they have and make suggestions based on this behavior model that the user is likely to use. A user may view, buy or tag items, view video clips of items, leave comments for products, search for keywords. It can show part of user behavior and these aspects can be useful. These actions can provide the user with very effective and accurate behavioral models. According to these behavioral models, suggestions can be made to users. Of course, these behavioral models can be used for other purposes as well, but we're focusing on product offerings here. otherwise we can analyze user behavior [10]. Many researches have been done in the analysis of user behavior, but the accuracy of these methods can be improved with innovative methods. Therefore, in this research, an attempt was made to provide a solution for recommending items, based on the past behavior of users, for this purpose, a fully improved decision tree was used [15]. The decision tree used in this research is an improved example of a decision tree that tries to improve this tree so that it has the lowest possible height and at the same time has a higher accuracy in decision making.

The second part is dedicated to the review of the background of the research done in this field, and the third part describes the proposed method, and in the fourth part, the proposed method is evaluated with several similar methods, and the fifth part includes the detailed analysis of the results, and in the sixth part, Conclusions and suggestions for future research are dedicated.

2 Research literature

In this section, we first define the basic concepts of user behavior analysis. Then, the background of empirical research is examined.

2.1 Theoretical background

In this section, we try to explain the basic concepts necessary to understand the proposed method.

2.1.1 Decision tree

The decision tree structure in machine learning is a predictive model that maps the facts observed about a phenomenon to conclusions about the target value of that phenomenon. The machine learning technique to infer a decision tree from the data is called decision tree learning, which is one of the most common data mining methods [15].

Each internal node corresponds to a variable and each arc to a child represents a possible value for that variable. A leaf node represents the predicted value of the target variable by having variable values represented by a path from the root of the tree to that leaf node. A decision tree represents a structure where the leaves represent the categories and the branches represent the seasonal combinations of attributes that result in these categories. Learning a tree can be done by separating a source set into subsets based on an attribute value test. This process is repeated recursively in each subset resulting from the separation. The regression operation is complete when further separation is no longer useful or a classification can be applied to all samples in the resulting subset.

Decision trees are able to produce human-understandable descriptions of the relationships in a data set and can be used for classification and prediction tasks. This technique has been widely used in various fields such as disease diagnosis, classification of plants and customer marketing strategies [21]. The main objectives of the decision tree are as follows:

1. Classify the input data as correctly as possible.
2. Generalize the knowledge learned from the training data in such a way that they classify the unseen data with the highest possible accuracy.
3. If new training data is added, the decision tree can be easily expanded (have an incremental feature).
4. The resulting tree structure should be as simple as possible [15].

2.1.2 Naïve Bayes classification

The Bayesian algorithm is a classification method based on the application of Bayes theorem, which includes the strong assumption that all predictors are independent of each other. In simpler terms, it is assumed that the presence of a feature in a class is independent of the presence of any other feature in the same class. For example, if a customer has interests such as sports, healthy eating, technology news, etc., he may be considered a smart customer. Although all these characteristics are interdependent, they independently contribute to the probability of the customer being intelligent [12].

2.1.3 SVM classification

Support Vector Machine, abbreviated as SVM, is a supervised machine learning algorithm that separates data samples represented as points in space using a line or hyperplane. This separation is such that the data points that are on the same side of the line are similar to each other and are placed in the same group. New data samples will be placed in one of the existing categories after being added to the same space. This algorithm is also used in hierarchical classification in this article [11].

2.1.4 Fuzzy logic

Fuzzy logic is a form of multi-valued logic in which the logical value of variables can be any real number between 0 and 1 and themselves. The word fuzzy means imprecise, unclear and ambiguous. This logic is used to implement the concept of partial correctness, so that the degree of correctness can be any value between completely true and completely false. The term fuzzy logic first appeared in the scene of new computing after the formulation of the theory of fuzzy sets by Lotfi [20].

2.1.5 Hierarchical analysis

Analytical Hierarchy Process or AHP is one of the decision making methods. The AHP method, which is also called the hierarchical analysis process, is one of the famous and widely used methods in multi-criteria decision-making, the purpose of which is to calculate the weight of each of the criteria and features used under a hierarchical model. is. In this model, first, pairwise comparisons are formed and given to the experts to express their opinions based on the two-by-two comparison of criteria. The main goal of the Hierarchical Analysis Process method is to select the best option based on different criteria through pairwise comparison. This technique is also used to weight the criteria. Because the increase in the number of elements of each cluster makes pairwise comparison difficult, therefore, decision criteria are usually divided into sub-criteria. The reason for the hierarchical study of this method is that one should first start with the goals and strategies of the organization at the top of the pyramid and by expanding them, determine the criteria to reach the bottom of the pyramid [6]. This method is one of the most widely used methods for ranking and determining the importance of factors, which prioritizes each of the criteria using a pairwise comparison of options.

2.2 The background of experimental research

Recommender systems are systems that try to provide suggestions using data mining and analysis of available data. In recent years, various resources and methods have been presented to analyze user behavior and use it in recommender systems. Rodoula and Tsiotsou argued the need to adjust personal recommender systems in learning enhancement technology for a specific characteristic of learning instead of using recommender systems for other contexts. In fact, the purpose of this article is to provide a suitable method for building a personal recommender system for lifelong learning. In this article, memory-based methods such as: collaborative filtering, which consists of context-based fans and hybrid fans, are used. In collaborative filtering, items that similar users have used in the past are recommended. The content-based fan method recommends options similar to those previously preferred by users. In the combined technique, both techniques are used to provide more accurate recommendations [18]. In a research, Jing sun et al proposed an analysis to find experts for a specific topic in online learning communities and focused on ranking users in a specific field in online communities and then finding expert users based on these ratings. It does this based on two hypotheses. First, a user posts or discusses more documents (eg, blogs, discussion text) about a topic or, again, is posted by other influential users [17]. In an article, the development of a recommender system as a reading tool was discussed based on a knowledge engineering method. In this article, expert knowledge is obtained from interviews with two experienced English teachers in an old high school. In addition, characteristics and tastes were obtained to determine readers from students and experts. It consists of two types of storage networks. One is a network to

categorize selected articles and the other is to determine readers based on their taste for English reading. The study has successfully developed an expert system to recommend English reading based on the opinions and knowledge range of expert English teachers. In [3] the authors state that the traditional emphasis on user similarity may be exaggerated. This paper proves that traditional factors play an important role in guiding recommendations and especially the trustworthiness of users should be considered as an important point. For this purpose, two computational models of trust are presented and it is shown how they can be used to combine with standard joint filtering frameworks in different ways. In fact, the purpose of this paper is to improve collaborative filtering. Danesh Qalichkhani et al investigated business intelligence in organizational structure, which they did through mediating variables [2]. With the spread and general use of the Internet, new information and communication technologies have greatly affected business and its methods. To the extent that traditional business methods have given way to new methods based on electronic business models. E-business models develop business processes through web-based information systems. In various web applications, the required knowledge and information are extracted from the massive amount of data using data mining techniques. Business intelligence is a process in which raw data are converted into business and management information using data mining. Vahidah et al, focusing on electronic business and data mining, expressed the challenges and benefits of business data mining. Also, this article presents a category of data mining applications used in e-business [13]. By searching in the field of improving recommender systems, we found that research has been done in this field and in many cases, it has brought challenges in the field of improving the algorithm, and an example of these challenges is the speed and accuracy of the proposed output. In this article, a recommender system based on a decision tree is presented, which first selects the effective features by using the rough fuzzy method, and then classification is done using the decision tree, which is created here with a hierarchical method.

3 Suggested method

In this section, the steps of the proposed method are fully explained. Here, the goal is to reach a solution that can be used to provide suitable suggestions that are close to the individual's desires, which is actually the understanding and analysis of different user behaviors. Using the method described here, we intend to provide a solution based on a hierarchical classification.

The method used in this research is a combined method that can greatly increase the accuracy of the method. Here, ID3 decision tree, Naïve Bayes algorithm and SVM are used in the hierarchical classification stage. Using this proposed solution, a weight is considered for each of these algorithms, and to calculate and predict the value calculated by each of these algorithms, the weight of that algorithm is multiplied and finally the real result is obtained. which will be more accurate because it has used the benefits of all these algorithms. The proposed solution that is expressed in this research includes the following steps:

1- Data preprocessing operations

2- Transferring data within the right limits

3- Selecting the effective features using rough fuzzy sets

4- Dividing data for training and testing

5- Hierarchical classification with three algorithms of support vector, Bayesian and decision tree and calculation of weights

6- Using weights to predict and test the method

The general flowchart of the proposed method can be seen in Figure 1. In the following, the work steps are described in detail.

3.1 Data preprocessing

In this stage, data preparation is done and the data is prepared for analysis. At this stage, missing data or columns of data that are unique and do not extract any knowledge are deleted, that is, data such as the national code, in fact, no knowledge can be extracted from it, so they must be deleted. The action takes place in this part. Some records may not have a value for some of its attributes, in which case two things can be done:

1- Delete the entire record

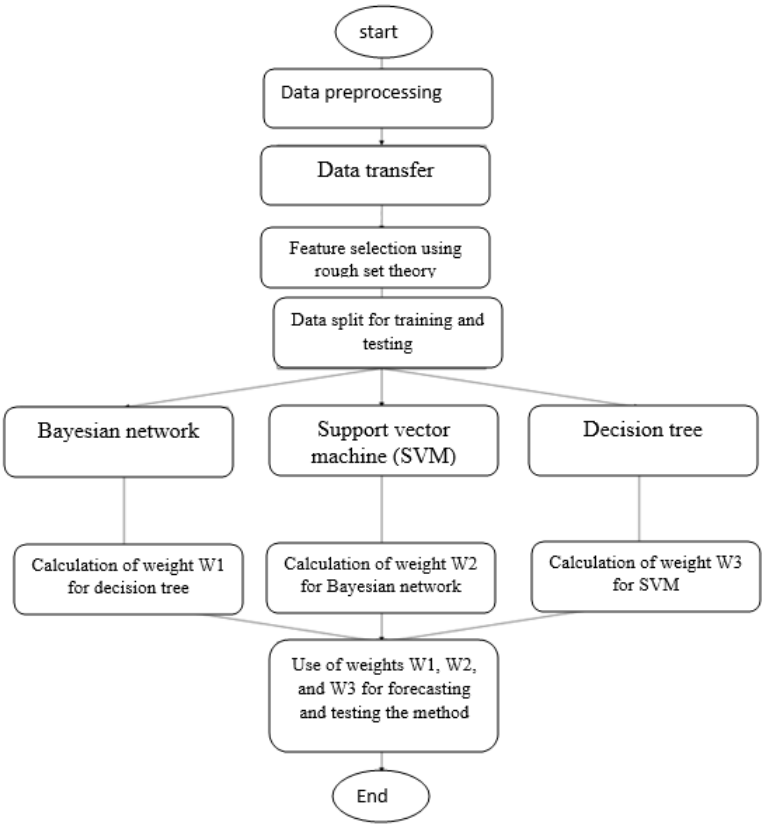

chart 1: Flowchart of the proposed method

2- Predicting the missing amount

By using prediction, incorrect knowledge may be entered into the data set because these prediction data have an accuracy factor and at this stage the prediction may not have been done correctly, which is why in this research, the act of deleting the entire record is used.

3.2 Data normalization

This step, which completes the previous step, must have the data in the correct range or range, which means that for example, if there is an age field, a person whose age range is between 55 and 70 should be very old in the system. It should be considered that this part is automatically completed from the dataset.

3.3 Selecting the effective feature by using rough fuzzy set

Raff fuzzy set transformation is a tool that can be used for ambiguity and uncertainty conditions, which is used in various fields such as decision analysis, decision support systems. A rough number has a lower limit (L), an upper limit (U) and a middle limit, which is known as the rough boundary distance. Rough numbers are used in issues where a set of experts are involved, which can somehow cause ambiguity and uncertainty. In this research, using the Raff set method integrated with the AHP method, which is a decision-making method, the parameters are weighted and the effective parameters are identified [16]. To use rough numbers in the AHP method (rough AHP), we do the following.

1. In the first stage, pairwise comparisons are checked from the point of view of incompatibility rate, and if this rate is less than 0.1, the pairwise comparison is considered compatible, and if it is greater, it is known as incompatible, and in this case, it should be These numbers are corrected for pairwise comparisons.
2. Creating rough numbers from experts' numbers using the relations mentioned in the theory.
3. Calculation of distance weights of criteria using the geometric mean method

After these steps, the best features are selected.

3.4 Hierarchical classification

In order to train, the data were randomly divided into two parts: 70% (training data) and 30% (test data). Due to the fact that decision trees are able to create decision-making conditions with high accuracy, a decision tree is made in this part. The features that are considered as inputs to provide suggestions and behavior of past users and new users and identified as influential parameters at this stage should be considered as tree nodes, but each of these features In which node they are placed or at what level of the tree they are placed is very important. In this section, a decision tree based on the ID3 algorithm is used, but many improvements have been made on it, and its weak points, which include the high height of the tree and processing and memory overhead, have been covered. The tree described in this research has the shortest possible height, which means that it has a very low processing and memory overhead, and we have tried to increase its accuracy as much as possible and not decrease it.

If a collection has only one class, it is called homogeneous, and if it contains several classes, it is called heterogeneous. There are different ways to measure homogeneity or not, among which we used entropy. Entropy is able to obtain the amount of disorder in a set, which can be calculated using the formula.

Entropy calculates the amount of disorder, and the lower the class diversity, the lower the entropy, and the higher the diversity or disorder, the higher the entropy. Entropy can be maximum 1 and minimum 0.

$$\text{Entropy} = \sum_{j} -p_j \log_2 p_j. \tag{3.1}$$

In this proposed method, entropy is used to select the root node at each level. The more entropy a feature is, it means that by changing that feature, the result may change with a greater effect. As a result, these features with more entropy should be placed in the higher levels of the tree, and the lower we go in the tree, the node should be be placed with less influence. In order for us to be able to calculate the degree of influence of a feature, we must subtract the entropy of the presence of this parameter from the entropy of its absence in order to find out the degree of influence of a feature, for which the following formula is used. That is, the entropy of the set A from the whole set S is calculated in this way, we calculate the total entropy or S and subtract it from the entropy of the set A, and in this case the weight of the set A is obtained, which is the weight of influence or the degree of influence:

$$G(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values } (A)} \text{Entropy}(v). \tag{3.2}$$

After the tree is created, the rules obtained from the tree analysis should be extracted. The process starts from the root and continues until reaching each of the leaf nodes. The rules that are extracted from this tree are called associative rules and decisions are made using it and predictions are made. In this way, it is possible to extract decisions in order to provide suggestions and understand the behavior of users. Support vector machine algorithm and Bayes algorithm are among the algorithms that are based on learning. In the next stage of this research, we are trying to use these algorithms to improve the tree so that the continuity between the features is maintained and as a result, according to the continuity between the users' behavior and the suggestions that can be presented, more accurate predictions can be made. Did. In this research, SVM algorithm and Bayesian network have been used because these methods have very high analysis power and also due to the accuracy of these two methods, these two methods have been used along with the proposed method until a solution can be provided. It has a much higher accuracy than normal and can achieve better results.

Analytical Hierarchy Process

Saaty [14] developed a strong and helpful tool for managing qualitative and quantitative multi-criteria elements involving in decision-making behavior. This model is called Analytical Hierarchy Process (AHP) and is based on a hierarchical structure. This procedure occupied an assortment of options in the decision and capable to apply sensitivity analysis on the subsequent criteria and benchmarks. In addition, it makes judgments and calculations easy because of paired comparisons. Moreover, it demonstrates the compatibility and incompatibility decisions which is the recompense of multi criteria decision making [8]. Analytical Hierarchy Process is one of the most inclusive system is considered to make decisions with multiple criteria because this method gives to formulate the problem as a hierarchical and believe a mixture of quantitative and qualitative criteria as well. The first step is to create a hierarchy of the problem. The second step is to give a nominal value to each level of the hierarchy and create a matrix of pairwise comparison judgment [7].

2 Steps to Conduct AHP

At the first stage, the issue and goal of decision making brought hierarchically into the scene of the related decision elements. Decision making elements are decision indicators and decision choices. The group established a hierarchy according to Figure 1 which should reflect the understudy problem.

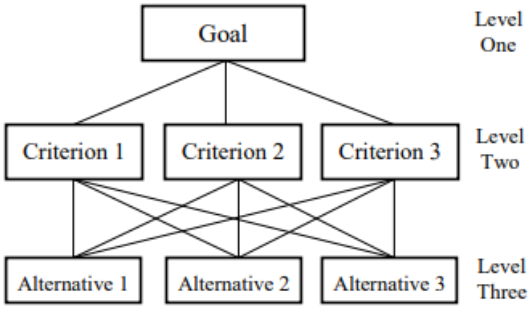

Figure 1: Sample Hierarchical Tree

In second step and in order to conduct pair comparison, a questionnaire should be designed and distributed among the respondents (can be managers, experts, users and etc.) to collect their opinion. It is noteworthy that each decision maker entered their desired amount for each member and then individual judgments (of each respondents) have been converted into group judgments (for each one of the pair comparison) using their geometrical average. The scale ranges from one to nine where one implies that the two elements are the same or are equally important. On the other hand, number nine implies that one element is extremely more important than the other one in a pairwise matrix. The pairwise scale and the importance value attributed to each number are illustrated in the Table 1, 2 shows the sample of the questionnaire.

Table 1: Relative scale for paired comparison

Intensity of importance		Description
Equal importance	1	Both activities equally contribute to the objective.
Moderate importance	3	Weak or slight importance over another – Experience and judgment slightly favor one activity over another
Strong importance	5	Greater or more essential importance when compared with another – Experience and judgment strongly favor one activity over another.
Very strong importance	7	Very high or demonstrated importance – An activity is favored very strongly over another; its dominance is demonstrated in practice.
Extreme importance	9	Extremely high importance – The evidence favors one activity over another with the highest level of certainty

Source: Adapted from Saaty [14] and Granemann & Figueiredo (2013)

Table 2: Sample AHP Questionnaire How important are the following security criteria in comparison

Factor	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Factor
Privacy	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Rellabillity
Privacy	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Validation
Privacy	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Verification
Privacy	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Integrity
Privacy	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Confidentially
Privacy	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Availability

The data analyze procedure involves the following steps. First the pairwise comparison matrix which is called matrix A is extracted from the data collected from the interviews. The principal right eigenvector of the matrix A is computed as 'w'.

If $a_{ik} \cdot a_{kj} = a^{ij}$ is not confirmed for all k, j, and i the Eigenvector method is selected [5]. If the matrix is incompatible and in case of incomplete consistency, pair comparisons matrix cannot be used normalizing column to

get Wi. For a positive and reversed matrix, Eigenvector technique can be used which in it:

$$e^T = (1, 1, \cdots, 1)$$

$$W = \lim_{k \to \infty} \frac{A^k \cdot e}{e^T \cdot A^k \cdot e}.$$

To reach a convergence among the set of answers in to successive repetition of this process, calculation should be repeated several times in order to take a decision when facing an incompatible matrix. Then, the following formula is applied to transform the raw data into meaningful absolute values and normalized weight $w = (w1, w2, w3, \ldots, wn)$:

$$Aw = \lambda_{\max} w, \quad \lambda_{\max} \geq n$$

$$\lambda_{\max} = \frac{\sum ajwj - n}{w1}$$

$$a = \{a_{ij}\} \text{ with } a_{ij} = 1/a_{ij}$$

A: pair wise comparison
w: normalized weight vector
λ max: maximum eigen value of matrix A
a_{ij}: numerical comparison between the values i and j

In the next step, in order to validate the results of the AHP, the consistency ratio (CR) is calculated using the formula, $CR = CI/RI$ in which the consistency index (CI) is, in turn, measured through the following formula:

$$CI = \frac{\lambda_{\max} - n}{n - 1}$$

The value of RI is related to the dimension of the matrix and will be extracted from Table 3. It should be noted that consistency ratio lower than 0.10 verifies that the results of comparison are acceptable.

Table 3: The value of Random Consistency Index, Source: Golden and Wang [4]

Dimension	RI
1	0
2	0
3	0.5799
4	0.8921
5	1.1159
6	1.2358
7	1.3322
8	1.3952
9	1.4537
10	1.4882

Considering that this method is also an expertise-oriented technique and the sample size should be less than 10 people [9], therefore, in this section, the opinions of the same 12 people selected from the previous stage are again used. Its calculations using Expert choose software, during three stages of pairwise comparisons, normalization, weighting and final ranking, calculation of compatibility rate in judgments, calculation of vectors of local priorities and finally determination of final priorities, market share development criteria based on strategies Entering international markets were prioritized.

4 Findings

In this section, the proposed method is evaluated and compared with similar methods.

4.1 Data set

In this article, the research data of Wang et al. [19] have been used to review and compare the proposed method with different behavioral set models. In this research, we intend to evaluate the proposed method and compare its performance with other behavioral models.

The implementation done here is done by Python language version 2.7, in this direction, a series of libraries such as tensorflow-Gpu have also been used. Considering the use of the tensorflow library on the GPU, the research requires a graphics card, so the system on which the test is conducted must have high graphics specifications. The system used in this research has the following specifications:

- Has 16 GB of RAM
- 5th generation Corei7 processor
- NVIDIA 900-52401-0020-000 GRID K1 graphics card with 16 GB capacity
- Windows 10

In this article, different subsets of different products are used for the dataset [19]. In general, 5 subsets have been used, which have the characteristics of times_stamp, item_id, user_id and cate_id. More complete information about the number of features in the used dataset can be seen in Table 4.

Table 4: Data set statistics

Data collection	**users**	**items**	**Groupings**	**samples**
Electro.	192403	63001	801	1689188
Clothing.	39387	23033	484	278677

The Taobao dataset data is from a very famous Chinese store that has made the data open. This data set includes several different data sets that can be very useful in this research. In this thesis, the types of user behaviors collected from a Chinese store in an e-commerce business website are used as a multi-behavior dataset.

Table 5: Behavior set information

Multi-behavior dataset	
30358	users
19.8	Average size
49859	Brands
447878	items
4704	categories
64388	Coupons
109665	stores
111859	Queries
247313	Records

This site is a store where the ready and multi-behavior dataset has been made available for research purposes and can be accessed. In this data set, three types of behavioral groups are used: item behavioral group, search behavioral group, and coupon receiving behavioral group. done. The attributes considered for an item are item_id, shop_id, brand_id and category_id. The third group, that is, the behavior group receiving the coupon, also uses a series of special features, including coupon_type, coupon_id and shop_id. In this research, only data were used in which users have at least three types of behavioral data. In Table 2, you can see more detailed information about the data set.

4.2 The methods compared with the proposed method

In this section, the methods that are considered as competing methods and are compared with the proposed method are:

- CNN+Pooling: This method is based on convolution network. In this method, it has tried to predict behaviors using neural networks.

- Bi-LSTM: This method presented by Zhang is able to perform well by compressing the past information of users, and with this, the speed of its operation has also increased.
- BPR-MF: Bayesian ranking method, which is a framework for ranking provided by Rendell. In this method, it has been tried to use the relationship between the product and the user, and these pairs have also been used.

In this article, we considered the area under the AUC curve as an evaluation parameter. Of course, in the following, accuracy and precision parameters are also used for evaluation. AUC is the area under the ROC chart, the higher this number is, the more efficient the method is. ROC chart is a method to check the efficiency of categories.

4.3 Evaluation results

In this section, different models are trained for a type of user behavior and the results are compared. In reviewing the results, the Amazon dataset is used to calculate the AUC. Table 6 shows the results of these calculations.

Table 6: Comparison of AUC on Amazon dataset

Model	Electro.	Cloth.
BPR	0.7982	0.7061
Bi-LSTM	0.8757	0.7869
CNN+Pooling	0.8804	0.7786
Fuzzy hierarchy	0.8921	0.7905

According to Table 6, it can be seen that the proposed method has a better performance than the other methods investigated in this research, and it is therefore more satisfactory.

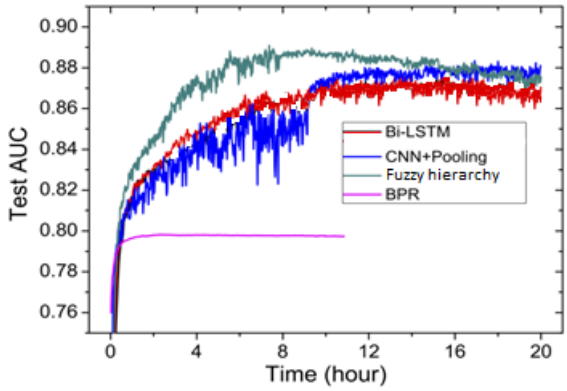

Figure 2: Area under Electro's ROC or AUC plot from the Amazon dataset

In Figure 2, it can be seen that the proposed method has a higher coverage power than the methods investigated in this research. This shows the comprehensiveness and proper performance of the proposed method. In the second stage of evaluation, we intend to check the proposed method and the compared methods on multi-behavior models. These models are:

- One-to-one model: It is said that models are trained for one type of behavior and are used to predict the same type of behavior that we have used so far. We called this model one2one.
- All-to-one model: In this model, different behaviors are used for training, but only one type of behavior is considered for prediction. Therefore, in this research, it can be called all2one.
- All-to-all model: In this type of model, different behaviors are used for training, and for all types of behavior, we call this model all2all.

The continuation of the three different models presented has been examined, and the results obtained from the implementation can be seen in Table 4. According to the results obtained from this table, it is possible to conclude that the all2one model has performed much stronger and better than the one2one model.

Table 7: AUC results on different behavioral sets

Model	Item	Query	Coupon
Bi-LSTM	0.6779	0.6019	0.8500
CNN+Pooling	0.6762	0.6100	0.8611
Fuzzy hierarchy -one2one	0.6785	0.6132	0.8601
Fuzzy hierarchy -all2one	0.6825	0.6297	0.8725
Fuzzy hierarchy -all2all	0.6759	0.6199	0.8587

In this part, it was tried to compare the important parameters of data mining for Bi-LSTM, CNN+Pooling and the proposed Fuzzy hierarchy algorithm, which include MSE and parameters expressing prediction accuracy such as True Positive, False Positive, True Negative and False It becomes negative. The mean square error or MSE is calculated using the following formula and shows the amount of prediction error compared to the actual value:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(r_i - p_i)^2 \tag{4.1}$$

In evaluating the accuracy of the models, they have been compared with each other, whose formula is as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2}$$

that the positive correct criterion (TP) means that if the answer obtained from the prediction is p and the actual value is p, then 1 unit is added to TP, so in the data set used for the test, this value is collected will be and the false positive criterion (FP), that is, if the answer obtained from the prediction is p and the actual value is n, then 1 unit is added to FP, so this value is collected in the data set used for testing. to be It means that in general, the answer obtained from the prediction is different from the actual value. The correct negative criterion (TN) means that if the answer obtained from the prediction is n and the actual value is n, then 1 unit is added to TN, so this value is collected in the data set used for testing. The difference between this part and TP is that in TP we count the positive answers and here we count the negative answers. The false negative criterion (FN) means that if the answer obtained from the prediction is n and the actual value is p, then 1 unit is added to Fn, so this value is collected in the data set used for testing. It means that in general, the answer obtained from the prediction is different from the actual value.

Table 8: Prediction accuracy and error

MSC	Accuracy	Model
20%	80%	Fuzzy hierarchy
21.11%	78.88%	Cnn+Pooling
26.11%	73.88%	Bi+LSTM

The difference between this part and the previous part is that in the previous part we considered negative answers that were mistakenly considered positive, and here we consider positive answers that were mistakenly considered negative.

The results of percentage accuracy of predictions and error of predictions are shown in Table 8.

According to the received results, it can be clearly seen that the proposed algorithm has the highest accuracy with 80% accuracy and the lowest error rate with 20% error. According to the obtained accuracies, it can be seen that the proposed method has performed much better than other methods, and in the meantime, the Cnn+Pooling algorithm has also performed better than the Bi-LSTM algorithm. In the following, the graphs obtained from the program have been examined. The first graph is the graph related to the percentage of correct prediction among the test data, which can be seen in Figure 3. The percentage of wrong prediction can be seen in the graph presented in Figure 4.

According to Figure 4, it can be seen that the prediction error is almost equal and close in the investigated algorithms, but it is less in the proposed algorithm. Figure 5 compares the average error of the evaluated models. It can be seen that the average error rate (MSE) in the proposed method is lower than all other methods.

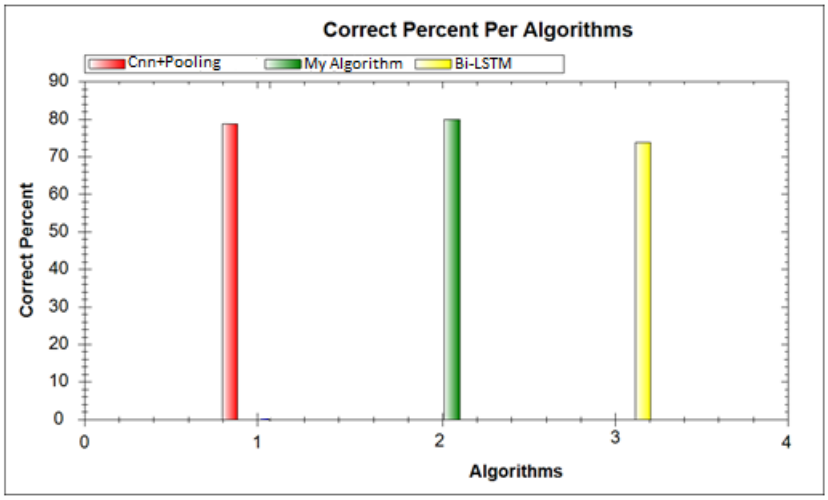

Figure 3: Percentage of correct prediction among test data

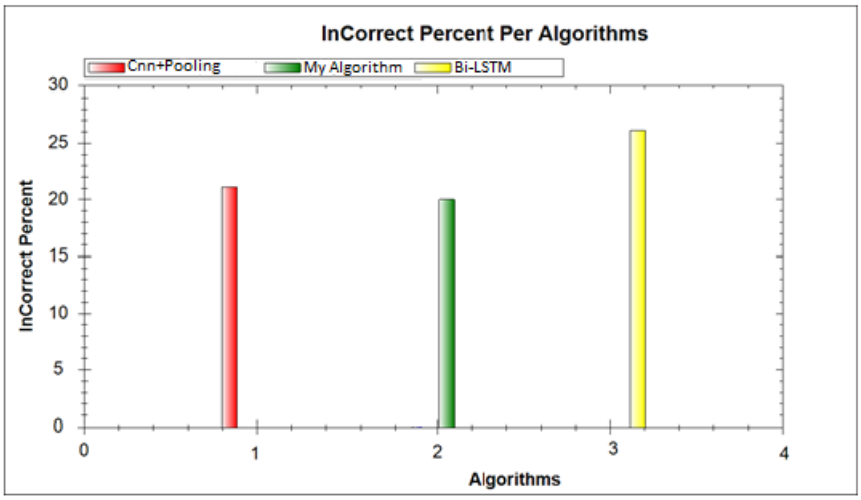

Figure 4: Percentage of wrong prediction

5 Analysis of the results

In the first stage of the evaluations, different models were trained for a type of user behavior. The obtained results show that the proposed method has a better performance than the other investigated methods in a type of behavior and is therefore more satisfactory. In the next step, we evaluated the proposed method and the compared methods on multi-behavior. The results showed that the all2one model has performed much stronger and better than the one2one model because it is able to act on several types of behavior instead of one type of behavior and therefore has a better performance. One of the interesting results is that the all2all model has a worse performance than all2one, even though it examines more behaviors. This result is due to the fact that it is not possible to check all the behavior models and make predictions for all the other behavior models according to those trained behaviors. From a logical point of view, this result can also be guessed. Because in this case, a poor prediction is made and as a result, the accuracy is greatly reduced. In the evaluation, it was tried to compare the important parameters of data mining for Bi-LSTM, CNN+Pooling and the proposed fuzzy hierarchy algorithm, which includes the mean MSE error and prediction accuracy. According to the received results, it can be clearly seen that the proposed algorithm has the highest accuracy with 80% accuracy and the lowest error rate with 20% error. The proposed method has performed much better than other methods, and in the meantime, the Cnn+Pooling algorithm has also performed better than the Bi-LSTM algorithm. Our proposed method has more correct prediction accuracy than other investigated algorithms, i.e. Cnn+Pooling network and Bi-LSTM algorithm. According to the percentage of wrong prediction, it can be

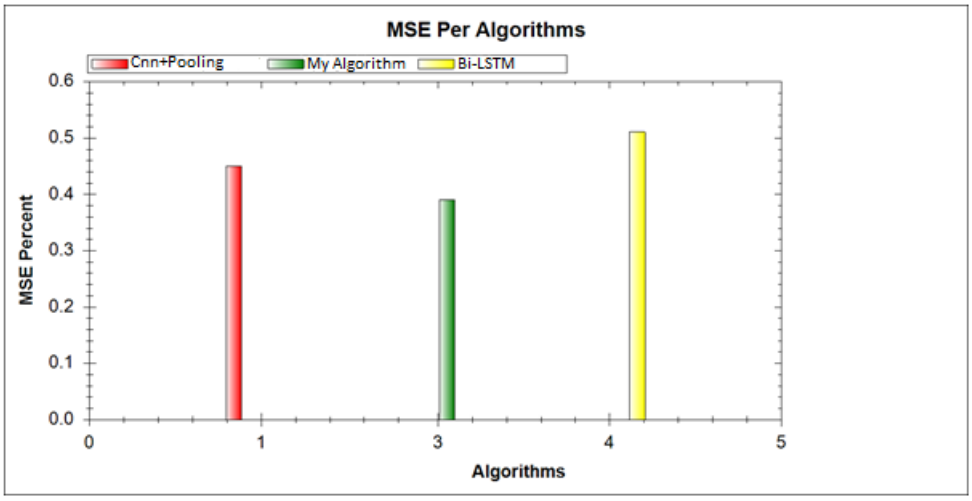

Figure 5: Prediction error comparison

understood that the proposed method has a lower value than the rest of the methods for the same reason that was said earlier regarding the percentage of correct prediction, that is, it has less wrong prediction, so the proposed method is better than the rest of the methods. has acted The prediction error is almost equal and close in the studied algorithms, but it is less in the proposed algorithm because the higher the accuracy rate, the lower the false rate, and this indicates the proper performance of the proposed method. In the results, it can be seen that the mean error rate (MSE) in the proposed method is lower than all other methods, and the Cnn+Pooling algorithm has a lower error rate than the Bi-LSTM algorithm. This error rate not only checks the wrongness, but also calculates how far the predicted answer is from the actual answer, in which case it is observed that the proposed algorithm behaves better than other algorithms. If you pay attention to this diagram, you can understand this because the Bi-LSTM algorithm has more wrong predictions than the Cnn+Pooling algorithm, but because the distance was not considered there, Bi-LSTM is worse than Cnn+Pooling. but it can be seen in Figure 5 that the Bi-LSTM algorithm has a lower MSE than the Cnn+Pooling algorithm, which means it has a lower error rate. In general, in the science of data mining, MSE is very important and has a lot of credibility.

6 Conclusion

Due to the growth of data mining and the process of data review, today various algorithms have been presented for data analysis, therefore, in this article, a model for using data and analyzing them has been presented, which is able to cope with the overhead of memory and overhead. Low processing. In order to reduce the overhead of training as well as prediction, hierarchical analysis and fuzzy logic are used in this proposed method, which leads to a better and more effective selection of features, and as a result, reduces computational overhead and increases accuracy. Here we examined different models such as one2one, all2one and all2all and it was observed that all2one performs much better than the rest of the models in which different behaviors are taught and a prediction is made for each type of behavior. will be The proposed method was also compared with CNN+Pooling and Bi-LSTM methods and it was observed that the proposed method can work with high accuracy and much more suitable performance. In future work, deep learning hybrid models can be used to train the model and evaluate its performance.

References

[1] P. Covington, J. Adams, and E. Sargin, *deep neural networks for YouTube recommendations*, Proc. 10th ACM Conf. Recomm. Syst., 2016, pp. 191–198.

[2] R. Danesh Qalich Khani, M. Hakak, and A.A. Farhani, *A model for measuring the direct and indirect impact of business intelligence with the partial mediating role of empowerment on organizational agility (Case study: Tehran engineering system organization and Atka organization)*, Int. Res. Conf. Sci. Engin., 2016. [In Persian]

[3] Gh. El-Hadad, D. Shawky, and A. Badawi, *Adaptive learning guidance system (ALGS)*, arXiv preprint arXiv: 1911.06812 (2019).

[4] B.L. Golden and Q. Wang, *An alternative measure of consistency*, B. L. Golden, A. Wasil & P.T. Harker (eds.), Analytic Hierarchy Process: Applications and Studies, 1990, pp. 68–81.

[5] N. Jalaliyoon, N.A. Bakar, and H. Taherdoost, *Accomplishment of critical success factor in organization; using analytic hierarchy process*, Int. Jo. Acad. Res. Manag. **1** (2012), no. 1, 1–9.

[6] E.L. José, *AHP-express: A simplified version of the analytical hierarchy process method*, MethodsX **7** (2020), 1–11.

[7] M. Krejnus, J. Stofkova, K.R. Stofkova, and V. Binasova, *The use of the DEA method for measuring the efficiency of electronic public administration as part of the digitization of the economy and society*, Appl. Sci. **13** (2023), no. 6, 36–72.

[8] M.C. Lee, H.W. Wang, and H.Y. Wang, *A method of performance evaluation by using the analytic network process and balanced scorecard*, Int. Conf. Converg. Inf. Technol., IEEE, 2007, pp. 235–240.

[9] L.C. Lin and G.P. Sharp, *Quantitative and qualitative indices for the plant layout evaluation problem*, Eur. J. Oper. Res. **116** (2000), no. 1, 100–117.

[10] J. McAuley, Ch. Targett, Q. Shi, and A. van den Hengel, *Image-based recommendations on styles and substitutes*, Proc. 38th Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval, 2015.

[11] L. Ren and W. Wang, *An SVM-based collaborative filtering approach for Top-N web services recommendation*, Future Gen. Comput. Syst. **78** (2018), no. 4, 531–543.

[12] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, *BPR: Bayesian personalized ranking from implicit feedback*, Proc. 25th Conf. Uncertain. Artiff. Intell., 2009, 452–461.

[13] V. Rokhi Nesab, Z. Ebrahimi and A. Rajaei, *the role of data mining in electronic business*, Nat. Comput. Conf. Inf. Technol. Artif. Intell. Appl., Ahvaz, 2016.

[14] T. L. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, New York, 1980.

[15] H. Sharma and S. Kumar, *A survey on decision tree algorithms of classification in data mining*, Int. J. Sci. Res. **5** (2016), no. 4, 2094–2097.

[16] M. Stanković, P. Gladović and V. Popović, *Determining the importance of the criteria of traffic accessibility using fuzzy AHP and rough AHP method*, Decis. Mak.: Appl. Manag. Engin. **2** (2019), no. 1, 86–104.

[17] J. Sun, H. Zhao, S. Mu, and Z. Li, *Purchasing behavior analysis based on customer's data portrait model*, IEEE 43rd Ann. Comput. Software Appl. Conf., 2019.

[18] R.H. Tsiotsou, *Value creation in tourism through active tourist engagement: A framework for online reviews*, Women's Voices Tourism Res. **12** (2021), no. 1, 1–14.

[19] J. Wang, P. Huang, H. Zhao, Z. Zhang, B. Zhao, and D.L. Lee, *Billion-scale commodity embedding for e-commerce recommendation in Alibaba*, Proc. 24th ACM SIGKDD Int. Conf. Knowledge Discov. Data Min., 2018.

[20] A.L. Zadeh, *Fuzzy logic*, Computer **21** (1998), no. 4, 83–93.

[21] G. Zhou, J. Zhao, T. He, and W. Wu, *An empirical study of topic-sensitive probabilistic model for expert finding in question answer communities*, Knowledge-Based Syst. **66** (2014), no. 4, 136–145.