INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

*Corresponding author.

animaabhi@gmail.com

Implementation of Sequential Pattern Neural Classifier in E-Commerce Data Behavioral Characteristic Extraction

P Anima[1]*, A S Aneeshkumar[2]

1 Research Scholar, Department of Computer Science, AJK College of Arts and Science, Coimbatore, 641105, Tamil Nadu, India
2 Research Supervisor & Head, Research Department of Computer Science & Applications, AJK College of Arts and Science, Coimbatore, 641105, Tamil Nadu, India

Abstract

Objectives: To present a framework for sequential pattern analysis using deep learning with behavioral characteristic extraction. This work intends to address the two major problems of accuracy and false positives predictions due to higher self-similarity in historical clicks. **Methods:** It implements sequence-aware recommenders for product recommendation using hybrid historical sequential pattern recommendation system (HSPRec) and hybrid sequential pattern based neural (HSPN) algorithm to take advantage of this crucial attribute. Data is gathered from Amazon, Flipkart, and other e-commerce sites. The simulation is carried out in Matlab. **Findings:** The proposed deep learning model provides 98% accuracy across 46 epochs, which is at least 8% higher, compared to existing works. The false positives in proposed solution are at least 6% lower compared to existing works. **Novelty**: The accuracy attained in recognizing consumer sequential patterns using HSPN is 95% and 98%, respectively. The computational effectiveness and viability of the neural network algorithm have been shown via our testing of the method.

Keywords: Historical Click; Purchase Data; E-Commerce; Sequential Pattern; Initial Clicks And Purchases

1 Introduction

This research concentrate on the employment of supervised learning to analyse clickstream data in the context of marketing use cases. A comprehensive review of relevant studies is also presented in this section. These studies are relevant to develop models to enhance product recommendations. The existing recommendation systems LiuRec09, ChoiRec12, SuChenRec15, and HPCRec18[1–3] employ mining algorithms with certain sequences. The LiuRec09 method groups users with comparable clickstream sequence data into clusters, and then selects Top-N neighbours from the cluster to a target user who belongs to segmentation-based collaborative filtering and association rule mining. HSPRec calculates a user's rating for a product as the proportion of all the users who purchased the item.

In the field of e-commerce today, data created by customers during the buying process may be utilised as a tool for data analysis, and the association rule algorithm in data mining technology is frequently employed [4]. The advantages are maximised by the widespread use of data mining methods [5]. Additionally, [6] offered a fresh method for customers to push relevant adverts when looking at products on e-commerce platforms.

According to [7], financial specialists may examine a client's savings, loans, and daily consumption bills to determine their capital situation and level of consumption capacity, allowing them to offer the appropriate financial products to the consumer.

The majority of Zhao's research on big data mining, feature selection, clustering, and granular computing, which can be used to process big data from social networks using text semantic processing, clustering, and other mining techniques [8]. Wan et al. use particular scenarios to demonstrate the usefulness of the suggested approach. A new feature selection strategy is suggested for high-dimensional data mining [9].

Data mining is the process of extracting meaningful information from imperfect data using a variety of approaches, including archives and expert knowledge [10]. In light of the difficulties inherent in mining such massive amounts of data, granular computing has emerged as a novel strategy for tackling today's complex issues. The Spark platform was created for the use of -means and FP in parallel computation [11]. Massive amounts of information related to thermal energy may be retrieved. A computational model based on the results of big data mining technology is developed [12] to examine several approaches to analyzing time series and data in big data.

2 Methodology

The experimental dataset is used to predict consumer spending information in 2021. This dataset consists information about the past online shopping habits of 9,000 individuals over the whole range of products from June to December 2021.

The following describes the design of the e-commerce data analysis and prediction method:

- First, deleting all selected e-commerce data and remove missing values. The distribution of the data as a whole is then characterised in order to determine the distribution of user behaviour for better data analysis and prediction.
- Users' browsing histories are grouped by date and time. Another category is used to categorise consumer behaviour.
- Select the features that can reflect the data to create the SPHRec model and initialize the parameters of SPHRec model, including learning rate, the number of base learners, thresholds, and others.
- Train the SPHRec model using the training data, then use random search to fine-tune the model's parameters. The training of the model is ended when it reaches the maximum number of iterations or the optimal parameters, and the best model is then produced.
- Predict the data using the best model that was created during training, and then output the results. As a result, e-commerce data analysis and prediction are accomplished.

2.1 Sequential Pattern

Sequential patterns are a series of ordered things (events) that happen in relation to time. The angular bracket (<>), designates a sequential pattern, and each item set comprises sets of items. Each item contained in parentheses () and separated by commas denotes a group of things that were purchased at the same time. For instance, the e-commerce sequential pattern < (Rice, Oil), (Rice, Oil, Spices), (oil), (Vegetables, Salt)> means that the customer first bought Rice and Oil, then Rice, Oil, and Salt on their second purchase, then Rice on their third purchase, and finally Spices and vegetables on their fourth purchase.

For instance, if we consider only two item sets, then the sequence of events will look like this: <(Rice), (Vegetables, oil)>. Furthermore, an item may appear more than once in one event (itemset), but it may also appear more than once in other events (itemsets) within the same sequential pattern. Thus, the length of a sequence is defined as the number of times each item appears in the sequence. For examples, the 4-event sequence of the length 8 is< (Rice, Oil), (Rice, Oil, Salt), (Oil), (Vegetables, Salt)>.

2.2 Database in Sequential order

A collection of sequences {sq1, sq2,...,sqn} that are grouped according to time to make up a sequence database. One way to express a sequence database is as a tuple, where SID: stands for the sequence identity and sequence-item sets that refer to the sets in the item wrapped within the parentheses ( ). Let's look at Table 1, historical regular purchase data from an online store. It includes a timestamp to indicate the moment the transaction took place, a Customer ID to represent a customer, and a Purchased Item to represent a group of things that consumers have purchased.

Table 1 contains the daily sequential database built from historical data, where SID stands for sequence identity. Table 2 shows that SID(01) indicates that customer (01) initially purchased Rice and Oil together, followed by Spices, Salt, Vegetables,

Table 1. Historical dataClickstream

Customer ID	Purchased Item	Time stamp
01	Rice, Oil	10, Aug 2021 00:34:43
02	Spices	12, Aug 20212:12:37
01	Rice, Oil, Salt	19, Aug 202110:36:54
02	Salt, Vegetables	21, Aug 202109:28:13
01	Oil	24, Aug 202104:53:42
01	Spices, vegetables	27, Aug 202102:40:44

and Spices in second transaction, and Oil in third. Spices and veggies were purchased together most recently.

Table 2. Daily sequential database created from historical purchase

SID	Sequences
01	<(Rice, Oil), (Rice, Oil, Spices), (Vegetables, Salt)
02	<(Rice),(Vegetables, Salt)>

2.3 Sequential Pattern Mining

Hybrid Sequential Pattern Neural algorithm (HSPN) may be used to extract common sequences, repeating patterns, from an input historical sequential E-commerce database. These associations can then be utilised to assess customer purchasing behaviour. To put it another way, it is a method for extracting sequential patterns is greater than a minimum support level. Formally, The problem of sequential patterns is to find the set of all frequent sequences Sq in the given sequence database (Sqd) of items (iE) at the given minimum support where k, a set of sequential records (called sequences) representing a sequential database D, a minimum support threshold, and a set of U unique items or events E = {e1, e2,..., eK} shown in Figure 1.

SID	Sequences Sq
1	<(SA),(SB),(SF SG),(SC),(S D)>
2	<(SB),(SG),(SD)>
3	<(S B),(SF),(SG),(SA, SB)>
4	<(S F),(SA, S B),(SC),(SD)>
5	<(SA),(S B, SC),(SG),(S F),(S D,E)>

Fig 1. Sequential Clickstream Database

2.4 E-commerce Data Types

A list of the products a user has clicked on and/or purchased during a given time period makes up historical data for e-commerce. Table 3 contains a portion of historical data from an e-commerce database, with the schema Uid, Click (c), Clickstart (Cs), Clickend (Ce), Purchase (P), and Purchase time (Pt). Uid stands for User identity, Click(c) for a group of items a user has clicked, and ClickstartClickstart (Cs) and Clickend (Ce), for the timestamps at which the user began and ended their clicks respectively. Additionally, purchase contains a list of items a user has purchased, and purchase time shows the timestamp for the transaction.

Table 3. Ecommerce Clickstream data

Uid	Click(c)	Clickstart(Cs)	Clickend(Ce)	Purchase(P)	Purchasetime(Pt)
1	1,2,3	2021-02-05 10:25:15	2021-02-08 10:25:19	1, 2	2021-02-17 10:44:55
2	1, 4	2021-02-13 4:01:11	2021-02-17 11:50:19	3	2021-02-23 10:06:37
3	5	2021-02-23 11:00:05	2021-02-23 11:15:20	1, 4	2021-02-26 10:06:37
4	1,5	2021-02-26 9:45:11	2021-02-26 10:20:15	1, 2,5, 6	2021-02-05 15:55:32
5	3,5	2021-02-27 16:30:25	2021-02-29 16:40:42	5	2021-02-19 04:04:34

Clickstream (Cs) data from an e-commerce site reveals the visitors' journey across the site.A session is a collection of online E-commerce store data, that a user accesses during a single visit. In an e-commerce setting, click stream data is a collection

of sessions. Raw page requests and the information they are associated with (such as timestamps, IP addresses, URLs, status, amounts of data transmitted, referrers, user agents, and occasionally cookie data) that is stored in web server logs can be used to create click stream data.

Analysis of click streams reveals how online shoppers utilise and browse an e-commerce website. In an e-commerce setting, click streams in online E-commerce stores give data on how customers locate the site, what goods they see, and what things they purchase. This information is crucial to determining the success of marketing and merchandising efforts. To increase the efficacy of recommendations in online retailers, it is essential to analyse the information included in click stream data. Table 5 contains an example of e-commerce stores click stream data.

Table 4. Click Stream

Session ID	Timestamp Ts	Item ID	Category ID
*****Gf4J7	2021-09-22T21:25:10+00:00	8454876943	2
*****Gf4J7	2021-09-24T23:25:10+00:00	858958985	3
*****MM5K7	2021-09-24T22:25:14+03:03	86755732	5
*****DV5Z7	2021-08-24T22:25:14+05:07	863748388	4
*****ZN5S7	2021-08-24T20:38:12+00:00	8737885398	3

The session ID (*****Gf4J7 to ***** ZN5S7) is to identify the user, the timestamp used to identify the item visited, the item ID used to identify the item the user visited, and the category ID used to identify the category to which the item belongs (for example, rice is a member of the grocery category).

Rule generation; Frequently occurring sequences are represented by the rule Uclick→ Upurchase, where the left-hand side of the rule refers to a collection of clicked items and the right-hand side refers to a set of recommended products to buy. Confidence in the rule is defined as in equation to confirm the validity of the sequential rule for recommendation.

$$\text{confidence}\left(\text{Uclick} \rightarrow \text{Upurchase}\right) = \frac{\text{Support}\left(\text{UclickUUpurchase}\right)}{\text{Support(Uclick)}}$$

3 Results and Discussion

The neural network is a type of operating model that has many nodes that are coupled to one another. An artificial neural network's connections between its neurons can have different strengths, which can be changed. The processing and storing of information is therefore a function of the human brain based on this property. Similar to this, the fundamental properties of the human brain may be reflected and its capacity for learning can be abstracted by simulating the neurons of the human brain with a variety of simple electrical components. The neuron's input may be written as (a1, a2, a3, a4,........an), and its output can be written as follows: Hw,b(a)=f(wt,a). The sigmoid function or the tanh function is typically used as the activation function for the function f. The neural network's cost function for a single training sample is as follows, suppose there are m training samples in the output layer. E-commerce generates a massive quantity of data that needs to be handled in order for neural networks to be effective. The neural network performs better the more data it processes. On the other hand, performance will stagnate at a certain point for typical machine learning algorithms, even with more data.
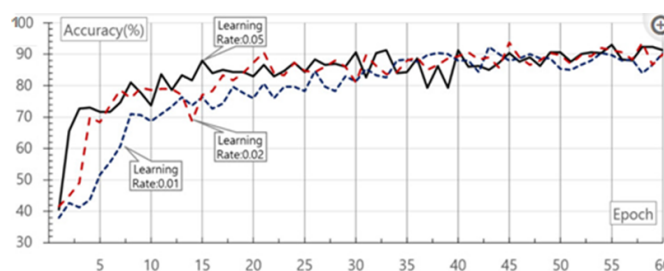

Fig 2. Result of accuracy learning rate

The evaluation's findings demonstrate that accuracy rises as learning counts increase. Although, accuracy also increases as learning rate rises, there is no difference when learning 45 epochs or more. The best accuracy is 95% after 55 epochs of learning

at a rate of 0.02.

$$Recall = \frac{TP}{TP + FN}$$

The terms TP and FN stand for true positives and false negatives. The neural network's recall performance was evaluated independently at a learning rate of 0.02. The evaluation of recall and evaluation of learning and evaluation data were extracted at random. The average of the numbers measured repeatedly three times is the outcome. Consequently, the recall was 98.00% across 46 epochs.

The results indicate that when there are many calculations, the time complexity increases, but this does not necessarily mean that the accuracy and performance are high. Of course, the accuracy of the recommendation condition is the most critical factor to consider when evaluating the best methods because it ensures that the customer receives the best advice and will accept the recommendation (maximize acceptance). According to that, the most of these techniques are seen to be effective, but the one provided may be the best since it combines complexity reduction with increased performance and judgement accuracy. Finally, the historical sequential recommendation of HSPRecis used the purchase of frequency matrix and frequently occurring sequences of purchase data to create sequential rules, used sequential rules to enhance the user-item matrix, and applied sequential rules to neural filtering, found better results than choiRec12 and HPCRec18. We have demonstrated the effectiveness of our enhanced neural network method through the aforementioned test exercise. The computation process took less time, was more accurate, and had a little offset from the original result. The algorithm that was optimized worked well. Figure 2 summarizes the findings and demonstrates our HSPN method's increased modulation recognition ability.

4 Conclusion

In this study, we introduced the HSPN model, which is a deep learning model derived from the Neural Network and it is suggested for the historical sequential pattern recommendation system modulation categorization. In comparison to the previous suggested strategy, the new model demonstrated an increase in the percentage of modulation recognition for low S/N values and decrease the time needed for the learning phase of the neural network. In light of this, we suggest creating an HSPN based method on our fundamental theory of using the historical sequential pattern as a source of knowledge.
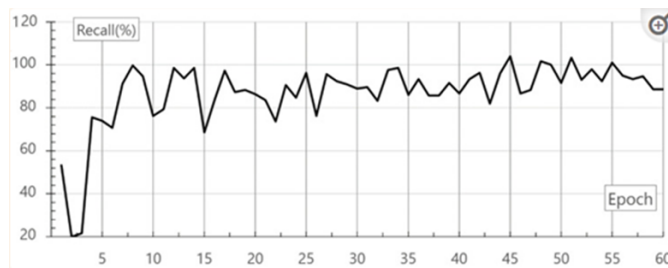

Fig 3. Result of recall evaluation

This study examined the e-commerce journey of e-commerce items using the neural network method. We tested an improved neural network approach and discovered that while there was a modest increase in calculation time approximately 45 epochs at a time the accuracy of the results remained over 98% throughout. The computed outcome was lower than learning at a rate of 0.02, and the improved neural network algorithm's time to recall performance was also shorter. In conclusion, our use of the neural network method proved practical and can guide us as we investigate the future of goods e-commerce.

A sequential pattern-based suggestion produced better result. Thus, potential future works include: (a) finding more possible way of integrating sequential pattern to collaborative filtering. (b)finding the most practical method for incorporating a sequential pattern from online data into a user-item matrix.

Acknowledgments

References

1) Kim YS, Hwangbo H, Lee HJ, Lee WS. Sequence aware recommenders for fashion E-commerce. *Electronic Commerce Research*. 2022. Available from: https://doi.org/10.1007/s10660-022-09627-8.
2) Ehikioya SA, Zeng J. Mining web content usage patterns of electronic commerce transactions for enhanced customer services. *Engineering Reports*. 2021;3(11). Available from: https://doi.org/10.1002/eng2.12411.
3) Koehn D, Lessmann S, Schaal M. Predicting online shopping behaviour from clickstream data using deep learning. *Expert Systems with Applications*. 2020;150:113342. Available from: https://doi.org/10.1016/j.eswa.2020.113342.
4) Li F, Li Y. Big Data Mining Method of E-Commerce Consumption Pattern Based on Mobile Platform. 2022. Available from: https://doi.org/10.1155/2022/3991135.
5) Zuo J. Analysis of e-commerce characteristics. *Wireless Communications and Mobile Computing*. 2021;2021. Available from: https://doi.org/10.1155/2021/6628068.
6) Lv X, Li M. Application and Research of the Intelligent Management System Based on Internet of Things Technology in the Era of Big Data. *Mobile Information Systems*. 2021;2021:1–6. Available from: https://doi.org/10.1155/2021/6515792.
7) Chen S. Analysis of Customization Strategy for E-Commerce Operation Based on Big Data. *Wireless Communications and Mobile Computing*. 2021;2021:1–11. Available from: https://doi.org/10.1155/2021/6626480.
8) Zhao Z, Wang H. The analysis of internet commercial judicial based on big data alliance and mining service process model. *Complexity*. 2021;2021. Available from: https://doi.org/10.1155/2021/2897823.
9) Wan YJJ, Yu ZCC. Exploring the Development of E-commerce for Fresh agricultural Products under the Background of E-commerce Law. *The Journal of Internet Electronic Commerce Resarch*. 2020;20(6):135–163. Available from: https://doi.org/10.37272/JIECR.2020.12.20.6.135.
10) Liu S, Zhang W. Application of the Fuzzy Neural Network Algorithm in the Exploration of the Agricultural Products E-Commerce Path. *Intelligent Automation & Soft Computing*. 2020;26(3):569–575. Available from: https://doi.org/10.32604/iasc.2020.013935.
11) Maslak N, Lei Z, Xu L. Analysis of agricultural trade in China based on the theory of factor endowment. *Agricultural and Resource Economics: International Scientific E-Journal*. 2020;6(1):50–61. Available from: https://doi.org/10.22004/ag.econ.302968.
12) Meiling H. Big Data Mining and Analysis of Agricultural Products Based on e-Commerce Platform. 2022. Available from: https://doi.org/10.1155/2022/1730934.