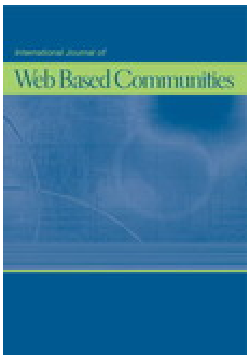


International Journal of Web Based Communities

ISSN online: 1741-8216 - ISSN print: 1477-8394
https://www.inderscience.com/ijwbc

Deep mining of e-commerce consumer behaviour data based on concept hierarchy tree

Yingchun Han

DOI: 10.1504/IJWBC.2024.10061789

Article History:

Received:	30 May 2023
Last revised:	27 July 2023
Accepted:	10 October 2023
Published online:	04 November 2024

Deep mining of e-commerce consumer behaviour data based on concept hierarchy tree

Yingchun Han

College of Marxism,
Heilongjiang Bayi Agricultural University,
Daqing, Heilongjiang, 163319, China
Email: HanYingChun112@163.com

Abstract: In order to solve the problems of low data collection efficiency, high noise, and low accuracy in traditional e-commerce consumer behaviour user data mining methods, a deep mining method for e-commerce consumer behaviour data based on concept hierarchy tree is proposed. Use Python scripting language to collect e-commerce consumer behaviour data from e-commerce platforms, and use Myriad filtering algorithm to remove the interference noise in e-commerce consumer behaviour data. Based on non-interference noise free e-commerce consumer behaviour data, utilising domain expert participation and machine learning algorithms, a concept hierarchical tree based e-commerce consumer behaviour data mining model is established to achieve deep mining of e-commerce consumer behaviour data. Experimental results show that the method proposed in this paper collects e-commerce consumer behaviour data more quickly, effectively removes interference noise contained in e-commerce consumer behaviour data, and can effectively and deeply mine the behavioural preferences of e-commerce consumers, with significant applicability.

Keywords: concept hierarchy tree; e-commerce consumer; behaviour data; deep mining; myriad filtering algorithm.

Reference to this paper should be made as follows: Han, Y. (2024) 'Deep mining of e-commerce consumer behaviour data based on concept hierarchy tree', *Int. J. Web Based Communities*, Vol. 20, Nos. 3/4, pp.323–339.

Biographical notes: Yingchun Han received his Master's degree in Law Science from School of Dongbei University of Finance and Economics in 2010. He is currently an Associate Professor the College of Marxism of Heilongjiang Bayi Agricultural University. Her research interests include ideological and political education, consumer protection, e-commerce regulation, and personal information protection rules.

1 Introduction

With the continuous development of internet technology, e-commerce platforms based on the internet have emerged one after another, bringing infinite convenience to people's lives (Fernández-Bonilla et al., 2022). However, with the increase in operating hours of e-commerce platforms (Cao et al., 2021), user behaviour data on e-commerce platforms has shown explosive growth. The user behaviour data of e-commerce platforms includes various actions such as browsing, searching, and paying by users on the power grid platform. The different granularity and dimensions of user behaviour data can provide reference for analysing user needs (Chen, 2021) and recommending products for users. Compared with offline shopping, e-commerce platforms cannot intuitively obtain users' every move from their behaviour data on the platform. However, based on the behaviour data left by users on the e-commerce platform, they can deeply explore their consumption process on the platform, obtain user preferences, and then recommend satisfactory products based on user preferences (Al-Zyoud, 2020), improving the marketing and service capabilities of the e-commerce platform. In the context of the widespread application of the internet and data processing technology, the problem of information overload on e-commerce platforms has gradually worsened, and a large amount of user behaviour data has been saved in the e-commerce platform database. How to mine useful information from massive user behaviour data to provide a basis for product recommendation and marketing decisions on e-commerce platforms is one of the important issues facing e-commerce platforms. Deep data mining is a data processing technology that includes multiple disciplines such as statistics, artificial intelligence, and mathematics. This technology can mine the association relationships existing in massive big data, improving the level of data processing and application. Deep data mining can not only review and summarise previous data, but also predict future trends based on current data, mine potential unknown relationships within the data, and obtain new knowledge, providing effective means for data applications. Nowadays, many scholars have also studied user data mining methods for e-commerce consumer behaviour, such as (Wagner et al., 2020) collecting consumer behaviour data from e-commerce platforms and using big data mining algorithms to mine consumer behaviour data in different business environments, obtaining consumer behaviour data in different business environments. However, this method did not preprocess consumer behaviour data, resulting in inaccurate final mining results. Riad and Chami (2021), based on historical consumer behaviour data, uses data processing and analysis algorithms to obtain consumer behaviour habits, and improves e-commerce platform services based on consumer behaviour habits. However, the small range of consumer behaviour data collected by this method results in poor final application results. Popp (2022) analyses the impact of the one mile delivery method on consumer consumption feelings on e-commerce platforms based on consumer purchase record data. However, this method has a small scope for mining consumer behaviour data and does not have universality.

The concept hierarchy tree algorithm is one of the big data mining algorithms, which can utilise high-level concept overview data in the tree. The root of the tree is a general description of possible attributes, and the leaves are attribute values of possible attributes. The hierarchical tree method can present feature rules, classification rules, association rules, etc. between data. The concept hierarchy tree is currently widely used in the field of data analysis and processing. Based on the concept hierarchy tree algorithm, this article studies a deep mining method of e-commerce consumer behaviour data based on concept hierarchy tree, providing data support for e-commerce platform marketing planning and product recommendation. The Committed steps studied in this paper are as follows:

1 Collect e-commerce consumer behaviour data from e-commerce platform websites using Python scripting language. After parsing the e-commerce consumer page using Beautiful Soup, you can obtain the labels and elements of the e-commerce consumer.

Then use the Breadth-first search strategy to capture e-commerce consumer behaviour data.

2 The e-commerce consumer behaviour data collected from the e-commerce platform contains a large amount of interference noise, and the e-commerce consumer behaviour data has discrete characteristics. The Myriad filtering method is used to filter the e-commerce consumer behaviour data. By finding appropriate adjustment parameter values, the serial number in the window is minimised, and the interference noise of e-commerce consumer behaviour data is removed.

3 Based on non-interference noise free e-commerce consumer behaviour data, a conceptual hierarchy tree of e-commerce consumer behaviour was constructed using domain expert participation and machine learning algorithms, and it was encoded and processed. Using this concept hierarchy tree to establish a preference model for e-commerce consumers, a data mining model for e-commerce consumer behaviour based on the concept hierarchy tree is completed, and the nearest neighbour of consumers on each preference item category is found by calculating similarity, thus achieving deep mining of e-commerce consumer behaviour data.

2 Deep mining of e-commerce consumer behaviour data

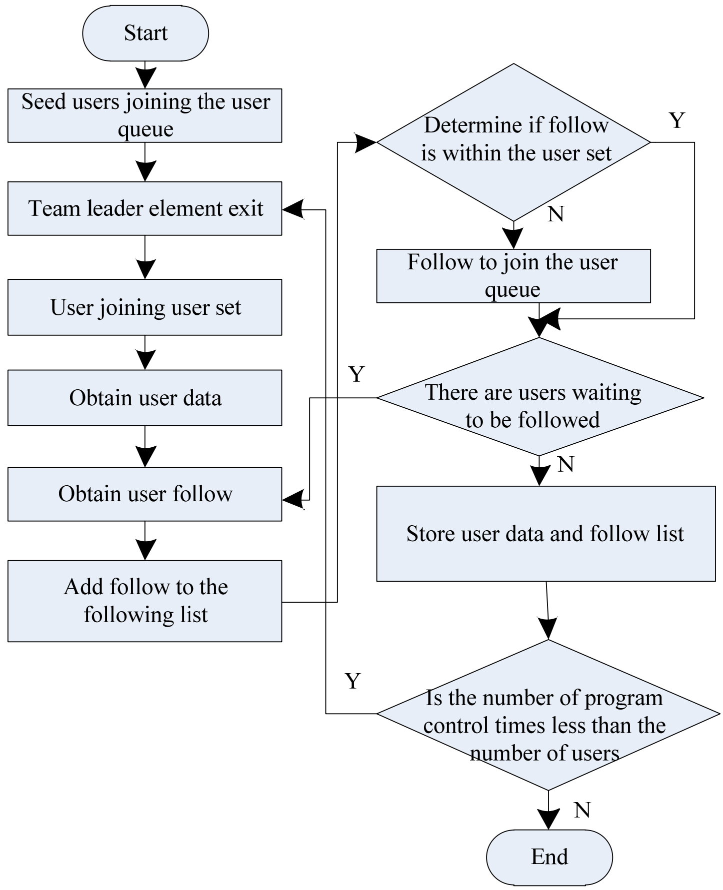
2.1 E-commerce consumer behaviour data collection

Real and complete e-commerce consumer data is the foundation for deep data mining. In order to collect real and complete e-commerce consumer behaviour data (Zhang et al., 2020), Python scripting language is used to collect e-commerce consumer behaviour data from e-commerce platform websites. Currently, e-commerce platforms are divided into web and mobile versions, and collecting e-commerce consumer data includes consumer home pages, details pages, tab pages, etc. The URL addresses of these pages are all associated with the consumer's ID. After parsing the e-commerce consumer page using Beautiful Soup, you can obtain the labels and elements of the e-commerce consumer. Then, using the breadth first search strategy to capture e-commerce consumer behaviour data, the detailed process is shown in Figure 1.

When using the breadth first search strategy to capture e-commerce consumer behaviour data, first select a seed user and add it to the user queue. Then, after the queue leader element in the user queue is removed, the team user is added to the user set. After capturing user data, the following user of the user is captured and added to the following list. Then, it is determined whether the current following user is in the user set, if so, determine whether there are currently users to follow, otherwise add the following users to the user set. If there are several currently pending followers, a list of followers will be obtained. Conversely, user data and their followers will be stored, and the current program control count will be determined to determine whether it has reached the number of users. If it returns to the queue element exit step, if not, the process will end. After repeated iterations of the above steps, complete the collection of e-commerce consumer behaviour data.

After the collection of e-commerce consumer behaviour data is completed, establish user information tables, attention tables, etc. in the MySQL database, store e-commerce consumer behaviour data based on the fields in the tables, and then associate and summarise the e-commerce consumer behaviour data table and product table to complete the collection of e-commerce consumer behaviour data.

Figure 1 Process for capturing e-commerce consumer behaviour data (see online version for colours)



2.2 Data preprocessing based on Myriad filtering

After collecting e-commerce consumer behaviour data from e-commerce platforms (Bao et al., 2021), there is a large amount of interference noise inside, and the e-commerce consumer behaviour data has discrete characteristics. Therefore, it is necessary to filter and preprocess the e-commerce consumer behaviour data (Young et al., 2020) to remove the internal interference noise (Liu and Yan, 2020). Here, the Myriad filtering method is used to filter and process e-commerce consumer behaviour data, providing a data foundation for deep mining (Yang, 2020) of e-commerce consumer behaviour. Let $X = [x_1, x_2, \cdots x_N]^T$ represent the e-commerce consumer behaviour data set containing noise (Khani et al., 2021), and its N is the number of single consumer behaviour data elements, then Myriad filter expression formula is as follows:

$$Y_K = \arg\left\{\min_{\beta}\prod_{i=1}^{M}\left[K^2+\left(x_i-\beta\right)^2\right]\right\} \tag{1}$$

In the above formula, Y_K represents the output result of the filter; M, K represents the filtering window length and linearisation parameters, respectively (Ks et al., 2020); β is the adjustment parameter; x_i represents the elements in the noisy e-commerce consumer behaviour dataset. The Myriad filter is used to remove interference noise from e-commerce consumer behaviour data by finding an appropriate β value to minimise the number of consecutive products in the window (Huang and Zhou, 2020).

To obtain the appropriate value of β, assuming that the value of M is 5, analyse formula (1) and it can be rewritten as:

$$f(\beta)=\left[K^2+\left(x_1-\beta\right)^2\right]\cdot\left[K^2+\left(x_2-\beta\right)^2\right]\cdots\left[K^2+\left(x_5-\beta\right)^2\right] \tag{2}$$

In the above formula, β represents a function of numerical values. From this formula, it can be seen that the value of parameter β is equal to $2M$, and the larger the window value, the greater the value of parameter β.

Using the linear operation method (Higham and Mary, 2020), calculate the minimum value of formula (2), and first define the intermediate variable as follows:

$$f_i(\beta)=K^2+\left(x_i-\beta\right)^2 \tag{3}$$

In the above formula, $f_i(\beta)$represents the i intermediate variable.

Substitute formula (3) into formula (2), which can be rewritten as:

$$f(\beta)=\prod_{i=1}^{5}f_i(\beta) \tag{4}$$

Taking the derivative on both sides of the equal sign in formula (4) yields:

$$f'(\beta)=f_1'f_2'f_3f_4f_5+f_1f_2'f_3f_4f_5+\cdots+f_1f_2f_3f_4f_5' \tag{5}$$

Among them,

$$f_i'(\beta)=-2\left(x_i-\beta\right) \tag{6}$$

If the value of formula (6) is 0, rewrite formula (5) as:

$$\frac{x_1-\beta}{1+\left(x_1-\beta\right)^2/K^2}+\frac{x_2-\beta}{1+\left(x_2-\beta\right)/K^2}+\cdots+\frac{x_5-\beta}{1+\left(x_5-\beta\right)^2/K^2}=0 \tag{7}$$

In the above formula, when $K\to\infty$ is used, the Myriad filter is transformed into a linear mean filtering form and the calculation formula for β is as follows:

$$\beta=\frac{x_1+x_2+\cdots+x_M}{M} \tag{8}$$

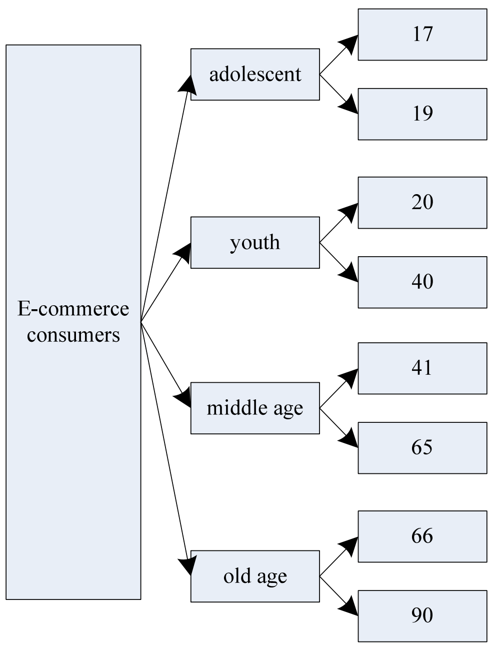
By substituting the results of formula (8) into formula (1), the interference noise in e-commerce consumer behaviour data can be removed.

2.3 E-commerce consumer behaviour data mining based on concept hierarchy tree

Based on e-commerce consumer behaviour data without interference noise, a concept hierarchy tree algorithm is used to establish an e-commerce consumer behaviour data mining model, and the results of e-commerce consumer behaviour data mining are output through this model.

The conceptual hierarchy of e-commerce consumer behaviour data is a partially ordered set, which can be represented by (q, <), where q is a limited concept set, and < is a partially ordered set on q, which is described in the form of a tree or linked list (Shao et al., 2021). The nodes of the concept hierarchy tree represent a concept, while the branches represent partial order relationships. Taking the age group data of e-commerce consumers as an example, the concept hierarchy tree is established, as shown in Figure 2.

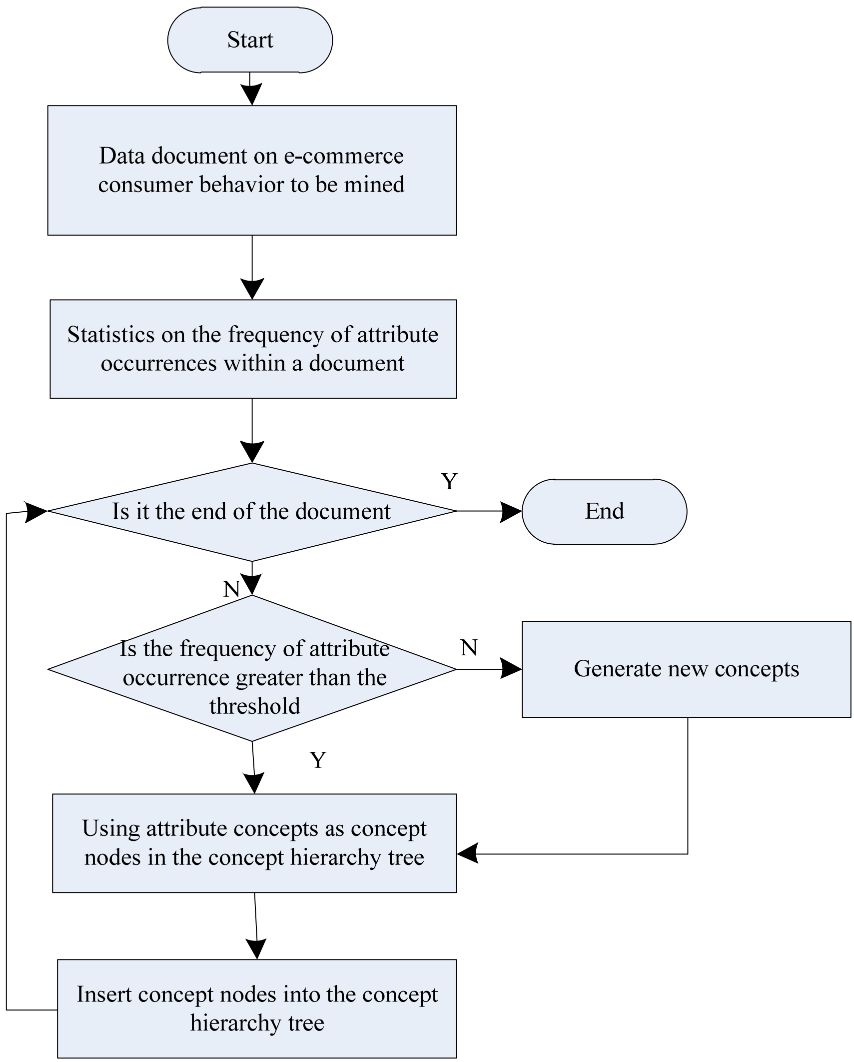
Figure 2 Age concept hierarchy tree of e-commerce consumers (see online version for colours)



After the establishment of the e-commerce consumer concept hierarchy tree, a bottom-up e-commerce consumer concept hierarchy tree is generated using domain expert participation and machine learning algorithms (Richens et al., 2020). The generation process is shown in Figure 3.

Whether the front position is the end of the document, if so, the process will end. If not, the frequency of attribute occurrences in the document will be determined to be greater than the threshold. If so, the concept hierarchy tree corresponding to the attribute will be used as a concept node. If so, a new concept will be generated, and the new attribute concept will be used as a concept node in the concept hierarchy tree. Then, the concept node will be inserted into the concept hierarchy tree, and the step of determining whether it is the end of the document will be returned, after repeated iterations of the above steps, a hierarchical tree of e-commerce consumer concepts is generated.

Figure 3 Generation process of e-commerce consumer concept hierarchy tree (see online version for colours)



After the generation of the concept hierarchy tree for e-commerce consumers, the concept hierarchy tree is encoded. This article combines sequential encoding and binary encoding to form a depth first traversal and assigns sequential encoding to the encoding form. This encoding form encodes a node as a, and then when traversing to its first child node, the child node is encoded as b, the above nodes are encoded as a/b, and when traversing to the next node as c, encode it as a/c. According to the above coding form, the pseudocode of e-commerce consumer concept hierarchy tree is as follows:

```
Input: E-commerce consumer concept hierarchy tree
Initialisation: Using depth first traversal to establish a binary linked list of e-commerce
consumer concept hierarchy trees
i = 1;
If (root !=null)
{
    Print(“root->data”,IntToString(i));
  P:=root;
    Do
          If (“root->RChild->data”, “root->parent->code”+”/root->code +”/”+ IntToString(++i)”);
          P:=root->RChild;
        }
    Until(P==null);
    DO
    If (P->LChild!=null)
      {
        Print(“root->LChild->data”, “root->parent->code”+”/root->code +”/”+ IntToString(++i)”)
           P:=root->RChild;
        }
Until(P==null);
        Output: Conceptual hierarchy tree all node encoding sets
```

According to the above program, the hierarchical tree nodes of e-commerce consumer behaviour concepts in Figure 4 can be encoded in the form of Table 1.

Figure 4 E-commerce consumer behaviour hierarchy tree nodes (see online version for colours)

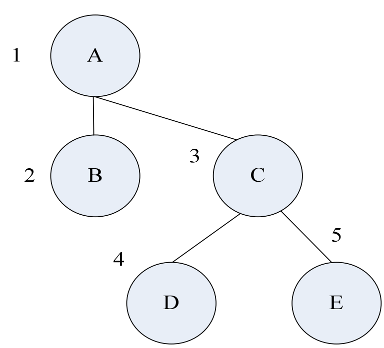


Table 1 Conceptual hierarchy tree node encoding

Concept node	*Code*
A	1
C	1/2
E	1/2/3
D	1/2/4
B	1/5

After completing the coding of the e-commerce consumer concept hierarchy tree nodes through the above steps, a preference model for e-commerce consumer c_i is established, with the expression formula as follows:

$$v_i = \left(v_{i1}, v_{i2}, \cdots, v_{i|D|}\right) \tag{9}$$

In the above formula, v_i represents the preference model of e-commerce consumer c_i, and $|D|$ represents the total number of nodes in the concept hierarchy tree; v_{ik} represents the evaluation of c_i e-commerce consumer on nodes within the conceptual hierarchy, and $k \in D$.

Let Q represent the total number of ratings given by e-commerce consumer c_i for each e-commerce item, and its calculation formula is as follows:

$$Q = \sum_{k=1}^{|D|} v_{ik} \tag{10}$$

Distribute the results of formula (10) evenly to the items that e-commerce consumers like, and then evenly distribute the scores of each item to its sub items (Yang et al., 2020). Use this method to obtain the initial value of the category that e-commerce consumer c_i likes, and the calculation formula for this initial value is as follows:

$$t\left(d_{jk}\right) = Q \times \frac{1}{\left|f\left(b_j\right)\right| \cdot \left|R_i\right|} \tag{11}$$

In the above formula, b_j represents the preferences of e-commerce consumers; d_{jk} represents the type of project; $t(d_{jk})$ represents the initial value of the category of e-commerce consumer c_i preferences; E represents the type of preference items that R_i commerce consumers belong to; $f(b_j)$ represents the total number of project categories.

Assign the initial score of project category d_{jk} to its upper level project categories. If P_0 and P_r respectively represent the top and bottom nodes of the conceptual hierarchy tree, then the node path from P_0 to P_r can be represented by $(p_0, p_1, \cdots, p_r)$, and $P_r = d_{jk}$, then the score calculation formula for the node P_t of each level in path $(p_0, p_1, \cdots, p_r)$ is as follows:

$$\sum_{t=0}^{r} Q\left(p_t\right) = t\left(d_{jk}\right) \tag{12}$$

$$Q\left(p_t\right) = Q\left(p_{t+1}\right) \times \frac{1}{b\left(p_{i+1}\right) + 1} \tag{13}$$

In the above formula, $Q(p_t)$ represents the score obtained by node P_t of each level in path $(p_0, p_1, \cdots, p_r)$; $b(p_t)$ represents the number of sibling nodes in node P_t.

Add the result of formula (13) to the vector component of the i preference inner level node P_t in the preference model of e-commerce consumer c_i. By iterating repeatedly, the preference model of e-commerce consumer c_i can be obtained. By using this model, the behavioural preferences of e-commerce consumers can be mined.

In order to mine more accurate behavioural preferences of e-commerce consumers (Silva et al., 2020), we searched for the nearest neighbours of e-commerce consumers in each preference category by calculating similarity, making the deeply mined results of e-commerce consumer behaviour more accurate. The detailed process is as follows:

Step 1 Divide a subset of project categories

Based on the node attributes in the conceptual hierarchy tree of e-commerce consumer behaviour data, the direct subtree of its root contains attributes of item types as independent relationships. The categories of e-commerce consumer items are divided into:

$$D = \left(Root, D_1, D_2, \cdots, D_w\right) \tag{14}$$

In the above formula, D represents the set of types of e-commerce consumer projects; *Root* represents the root node of the conceptual tree of e-commerce consumer behaviour data; w represents the total number of elements within the category set of e-commerce consumer projects.

Step 2 Find neighbours with preferred species

For categories of items that e-commerce consumers pay less attention to or have never paid attention to, they are considered as categories that e-commerce consumers are not interested in. When the result of dividing the number of items in the subset of visited categories by the total number of visited items is less than 10%, it is determined that the subset of item categories is an item category that users do not prefer. After determining the type of preference items for e-commerce consumers, based on this type of item, calculate the Pearson correlation coefficient between e-commerce consumers and each subset of preference items. The expression formula is as follows:

$$\begin{aligned} sim_h\left(c_i, c_j\right) = &\left[\sum_{k=1}^{|D_h|}\left(v_{ik} - \overline{v_j}\right)\cdot\left(v_{jk} - \overline{v_j}\right)\right] \\ &\times \frac{1}{\sqrt{\sum_{k=1}^{|D_h|}\left(v_{ik} - \overline{v_j}\right)^2 \cdot \sum_{k=1}^{|D_h|}\left(v_{jk} - \overline{v_j}\right)}} \end{aligned} \tag{15}$$

In the above formula, v_{ik} and v_{jk} respectively represent the ratings of e-commerce consumers c_i and c_j on the types of projects; D_h is a subset of project types; $\overline{v_j}$ represents the average score of the project category; $sim_h(c_i, c_j)$ represents the Pearson correlation coefficient values of e-commerce consumers c_i and c_j on each subset of preference items.

Use formula (15) to repeatedly calculate the Pearson correlation coefficient value between e-commerce consumers, and select the top K users who are most similar to user preferences as a subset of project types. This set is the result of deep mining of e-commerce consumer behaviour data.

E-commerce consumer behaviour data mining based on concept hierarchy trees can present a large amount of data in a visual manner, and more intuitively understand the relationships and structures between data. The concept hierarchy tree can divide data into different levels and categories, facilitating the classification and organisation of consumer behaviour data, thereby better understanding and analysing user behaviour. In addition, by analysing the relationships between various levels in the conceptual hierarchy tree, potential patterns and trends can be discovered, thereby predicting user behaviour, formulating marketing strategies, and providing personalised recommendations and services. Finally, the concept hierarchy tree can also quickly and accurately query and retrieve consumer behaviour data, improving work efficiency. In summary, e-commerce consumer behaviour data mining based on concept hierarchy trees can optimise operational strategies, improve market competitiveness, and help enterprises better utilise and understand consumer behaviour data.

3 Experimental analysis

Taking a certain e-commerce platform as the experimental object, it is a major 3C online shopping professional platform in the B2C market. The platform takes products, prices, and services as its long-term development strategy. After development, its market share ranks among the top e-commerce network platforms. This e-commerce platform has a large number of consumer users, covering various age groups. In order to more accurately recommend products to users, this e-commerce platform applies the method in this article to mine its e-commerce consumer behaviour data and obtain user shopping preferences.

Using the method described in this article, consumer behaviour data on the e-commerce platform was collected. A certain e-commerce consumer was used as the experimental object, and consumer behaviour data was collected within 12 days. The results are shown in Table 2.

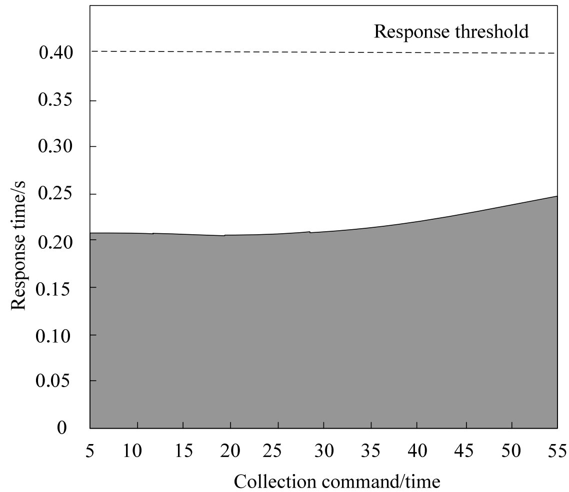
Table 2 Collection results of e-commerce consumer behaviour data

Time/day	*Login to e-commerce platform/time*	*Retrieve products/time*	*Place an order/time*
1	1	15	1
2	1	9	1
3	2	8	0
4	4	22	3
5	3	17	2
6	1	9	2
7	0	0	0
8	3	28	5
9	4	19	4
10	5	15	1
11	4	9	1
12	2	10	0

According to Table 2, the method proposed in this article can effectively collect the number of times e-commerce consumers log in to e-commerce platforms, search for products, and place orders for products. This can provide a good data foundation for further in-depth exploration of e-commerce consumer behaviour.

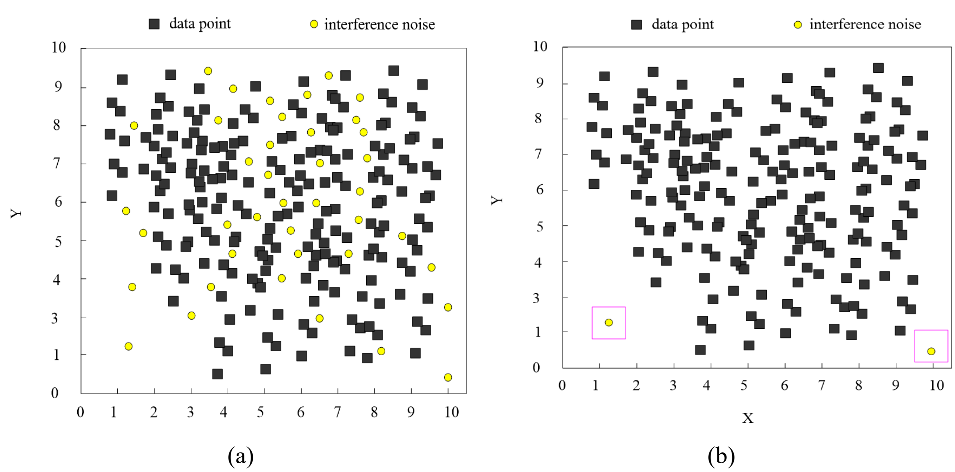
Test the responsiveness of the method used in this article to collect e-commerce consumer behaviour data. Under different requests for collecting e-commerce consumer behaviour data, the method collects the response time of e-commerce consumers and sets the response time threshold to 0.40s. The test results are shown in Figure 5.

Figure 5 Response curve of e-commerce consumer behaviour data collection



Analysing Figure 5, it can be seen that the response curve of this method in collecting e-commerce consumer behaviour data increases with the number of commands collected, but the increase is extremely low. When the number of commands to collect e-commerce consumer behaviour data is 55, the command response value is about 0.25s, which is lower than the preset command collection response threshold. The above results indicate that the method proposed in this article responds quickly when collecting e-commerce consumer behaviour data, and can effectively collect e-commerce consumer behaviour data.

Figure 6 Results of removing interference noise from e-commerce consumer behaviour data, (a) before noise removal (b) after removing noise (see online version for colours)



To verify the ability of the method proposed in this article to remove interference noise from e-commerce consumer behaviour data, 200 e-commerce consumer behaviour data were used as experimental objects. The spatial representation of e-commerce consumer behaviour data and perception noise was used, and the interference noise contained within it was removed using the method proposed in this article. The results are shown in Figure 6.

By analysing Figure 6, it can be seen that using the method proposed in this article to eliminate the interference noise contained in e-commerce consumer behaviour data can effectively remove a large amount of interference noise. Only residual noise exists in the lower left and right corners of the space, but the interference noise is slightly far from the e-commerce consumer behaviour data point and does not affect it. In summary, the method proposed in this paper has a strong ability to remove interference noise from e-commerce consumer behaviour data, which indirectly indicates that the method has a strong ability to deeply mine e-commerce consumer behaviour data.

Validate the method in this article to deeply explore the behavioural ability of e-commerce consumers. Using 12 e-commerce consumer behaviour data as the experimental object, we deeply mine their shopping behaviour preferences. The results are shown in Table 3.

Table 3 Deep mining results of e-commerce consumer behaviour data

User code	*Product preferences*	*Shopping time preference*	*Colour preference*	*Price preference*
1	Home life products	09:00–15:00	blue	100–300
2	Local specialties	13:00–21:00	yellow	10–200
3	Beauty products	11:00–13:00	pink	20–500
4	Women's clothing	10:00–18:00	white	300–600
5	Women's clothing	18:00–24:00	grey	100–300
6	Maternal and infant products	13:00–17:00	red	100–400
7	Sport products	15:00–21:00	yellow	50–600
8	health food	21:00–24:00	blue	300–700
9	fruit	11:00–14:00	yellow	10–100
10	home decoration	10:00–16:00	blue	300–1,000
11	Men's clothing	18:00–23:00	blue	200–800
12	Electronic digital	13:00–19:00	black	1,000–8,000

By analysing Table 3, it can be seen that using this method to deeply mine e-commerce consumer behaviour data can effectively mine the product preferences of e-commerce consumers, the time period when they like to log in to e-commerce platforms for shopping, the colour of the products they like, and the level of consumption ability. Based on the above preferences of e-commerce consumers, it is possible to recommend categories that meet their consumption level and preferences, in order to increase the revenue of e-commerce consumption platforms. The above results indicate that the method proposed in this article can effectively and deeply mine e-commerce consumer behaviour data, obtaining a large amount of detailed consumer information, and has significant applicability.

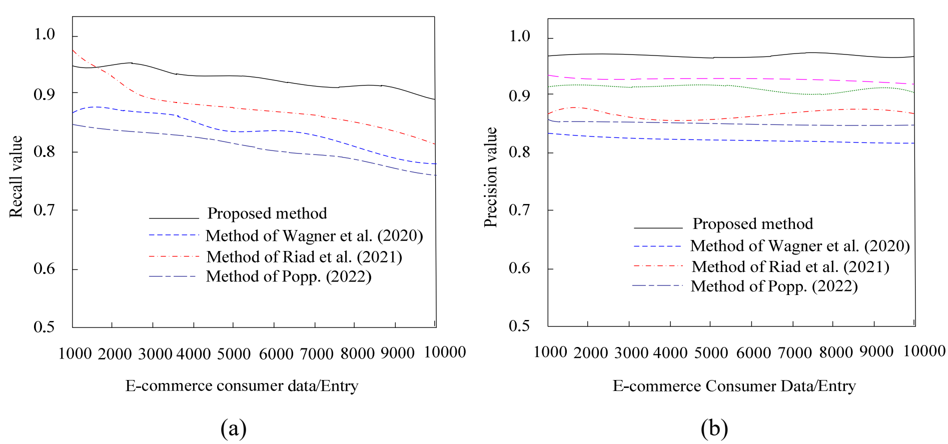
The Pearson correlation coefficient is used as an indicator to measure the user behaviour data of deep mining e-commerce platforms in this article's method. Twelve users from the above experiments were used as experimental subjects, and the Pearson coefficient values of their user preference deep mining results were calculated. The results are shown in Table 4.

Table 4 Pearson coefficient of deep mining results of e-commerce consumer behaviour

User code	*Product preferences*	*Shopping time preference*	*Colour preference*	*Price preference*
1	0.91	0.93	0.91	0.95
2	0.93	0.92	0.95	0.94
3	0.94	0.96	0.93	0.91
4	0.91	0.95	0.91	0.93
5	0.95	0.94	0.93	0.96
6	0.93	0.91	0.96	0.95
7	0.91	0.91	0.95	0.94
8	0.93	0.92	0.94	0.93
9	0.95	0.93	0.93	0.92
10	0.92	0.95	0.91	0.96
11	0.94	0.93	0.93	0.92
12	0.92	0.92	0.96	0.93

According to Table 4, after mining e-commerce consumer behaviour data using our method, the Pearson coefficient values of consumer behaviour preferences are all higher than 0.9, indicating that our method is more accurate in mining e-commerce consumer behaviour data.

Figure 7 Precision test results of e-commerce consumer project classification, (a) recall rate (b) accuracy (see online version for colours)



Using recall and precision as measurement indicators, the classification of e-commerce consumer item categories is a key step in mining e-commerce consumer behaviour data. Testing the accuracy of the method used in this article for classifying e-commerce consumer item categories. To make the experimental results more comprehensive, method of Wagner et al. (2020), method of Riad and Chami (2021), and method of Popp (2022) conduct the experiment, and the experimental results are shown in Figure 7.

Analysing Figure 7(a), it can be seen that when the four methods are used to deeply mine e-commerce consumer behaviour data, the recall value when dividing consumer item categories decreases with the increase of e-commerce consumer data volume. Before the number of e-commerce consumer behaviour data reached 2,000, the recall value of Method of Riad and Chami (2021) was higher than that of this method. However, as the number of e-commerce consumer behaviour data gradually increased, the recall value of this method was lower than that of this method. Among the four methods, when 10,000 consumer data were caught, the recall value of this method was the highest. From Figure 7(b), it can be seen that the precision numerical curve of the method in 4 shows a stable state as the amount of e-commerce consumer data increases, with the method in this paper having the highest precision value. The above results indicate that the method proposed in this paper has a high accuracy in classifying the types of e-commerce consumer behaviour data, and its deep mining effect is good.

To verify the application results of the method in this article, three e-commerce consumers were selected as the experimental subjects. After the application of the method in this article, the year-on-year consumption increase ratio of these three e-commerce consumers in different product categories was calculated, and the results are shown in Table 5.

Table 5 Year-on-year consumption growth ratio of e-commerce consumers in different product categories (%)

E-commerce platform product categories	*Consumer 1*	*Consumer 2*	*Consumer 3*
Women's clothing	10.8	10.4	18.2
Beauty makeup	5.2	8.2	10.5
Department store	3.6	5.6	11.4
Wash and care	4.2	4.7	5.9
Food	15.9	4.6	19.4
Menswear	2.3	4.1	6.8
Jewellery	3.8	6.5	8.4
Footwear	4.1	5.2	2.5
Mother and baby	16.6	3.8	6.8
Luggage and bags	7.1	1.1	6.4
Digital appliances	0.9	2.3	3.7

According to Table 5, after the application of the method in this article, the consumption amounts of the three consumers on the e-commerce platform under different product categories have all increased compared to the same period, with the maximum increase being 19.4% and the minimum increase being 0.9%. The above results indicate that after applying the method proposed in this article to provide product recommendations for e-commerce consumers, the consumption amount of consumers in different product categories has been effectively increased. The above results indicate that the method proposed in this paper has good practical application effects and can be widely applied in the field of e-commerce.

4 Conclusions

This article studies a deep mining method for e-commerce consumer behaviour data based on concept hierarchy trees. By establishing a concept hierarchy tree for e-commerce consumer behaviour data, the behavioural preferences of e-commerce consumers are mined. E-commerce platforms can recommend related products based on their behavioural preferences. After multi-perspective validation of the method in this article, it has been found that the method can quickly and effectively collect e-commerce consumer behaviour data, effectively remove the interference noise contained within it, and accurately mine e-commerce consumer behaviour data. By applying this method to provide product recommendations for e-commerce consumers, the consumption amount of consumers in different product categories has been effectively increased. Therefore, this method can provide users with more accurate and personalised product recommendations by mining the behavioural preferences of e-commerce consumers. However, the establishment of the concept hierarchy tree in this method requires manual participation and definition of the relationships between concepts, which may lead to the introduction of subjectivity and subjective bias. In future work, recommendation algorithms can be further improved to improve the accuracy and user experience of recommendations.

Acknowledgements

This work was supported by "Heilongjiang Bayi Agricultural University Talent Support Plan" Research on the legal balance mechanism between the use of big data on e-commerce platforms and the protection of consumer personal information (project: RRCPY202013).

References

Al-Zyoud, M.F. (2020) 'The impact of gamification on consumer loyalty, electronic word-of mouth sharing and purchase behavior', *Journal of Public Affairs*, Vol. 20, No. 3, pp.1–9.

Bao, X., Gu, T., Chang, L., Xu, Z. and Li, L. (2021) 'Knowledge-based interactive postmining of user-preferred co-location patterns using ontologies', *IEEE Transactions on Cybernetics*, Vol. 25, No. 99, pp.1–14.

Cao, X., Deng, M. and Li, H. (2021) 'How does e-commerce city pilot improve green total factor productivity? Evidence from 230 cities in China', *Journal of Environmental Management*, Vol. 289, No. 7, p.112520.

Chen, J.L. (2021) 'Real time prediction of cross border e-commerce flash sale performance based on neural network', *Advances in Applied Mathematics*, Vol. 10, No. 7, pp.2494–2499.

Fernández-Bonilla, F., Gijón, C. and Vega, B. (2022) 'E-commerce in spain: determining factors and the importance of the e-trust', *Telecommunications Policy*, Vol. 46, No. 1, p.102280.

Higham, N.J. and Mary, T. (2020) 'Sharper probabilistic backward error analysis for basic linear algebra kernels with random data', *SIAM Journal on Scientific Computing*, Vol. 42, No. 5, pp.A3427–A3446.

Huang, M. and Zhou, M. (2020) 'Linear quadratic mean field games: asymptotic solvability and relation to the fixed point approach', *IEEE Transactions on Automatic Control*, Vol. 65, No. 4, pp.1397–1412.

Khani, S., Haghighat, F., Panchabikesan, K. and Ashouri, M. (2021) 'Extracting energy-related knowledge from mining occupants' behavioral data in residential buildings', *Journal of Building Engineering*, Vol. 39, No. 6, p.102319.

Ks, A., Gang, D.B. and Zg, C. (2020) 'Cauchy noise removal by nonlinear diffusion equations', *Computers & Mathematics with Applications*, Vol. 80, No. 9, pp.2090–2103.

Liu, Y. and Yan, Z. (2020) 'Application of a cascading filter implemented using morphological filtering and time–frequency peak filtering for seismic signal enhancement', *Geophysical Prospecting*, Vol. 68, No. 6, pp.1727–1741.

Popp, B. (2022) 'Last-mile delivery methods in e-commerce: does perceived sustainability matter for consumer acceptance and usage?', *Sustainability*, Vol. 13, No. 39, p.14.

Riad, E. and Chami, R.A. (2021) 'Consumer behavior decision making prior to shopping online 'why e-commerce is not working for Lebanese'', Vol. 25, No. 7, p.576.

Richens, J.G., Lee, C.M. and Johri, S. (2020) 'Improving the accuracy of medical diagnosis with causal machine learning', *Nature Communications*, Vol. 11, No. 1, p.3923.

Shao, T., Zhang, H., Cheng, K., Zhang, K. and Bie, L. (2021) 'The hierarchical task network planning method based on Monte Carlo tree search', *Knowledge-Based Systems*, Vol. 225, No. 4, p.107067.

Silva, A.A., Tavares, M.W., Carrasquilla, A., Misságia, R. and Ceia, M. (2020) 'Petrofacies classification using machine learning algorithms', *Geophysics*, Vol. 85, No. 4, pp.1–71.

Wagner, G., Schrammklein, H., Steinmann, S. and Woodside, A.G. (2020) 'Online retailing across e-channels and e-channel touchpoints: empirical studies of consumer behavior in the multichannel e-commerce environment', Vol. 21, No. 11, pp.176–189.

Yang, H. (2020) 'Feature mining method of equipment support data based on attribute classification', *Ordnance Material Science and Engineering*, Vol. 43, No. 6, pp.124–128.

Yang, Z., Gang, A. and Bajwa, W.U. (2020) 'Adversary-resilient distributed and decentralized statistical inference and machine learning: an overview of recent advances under the byzantine threat model', *IEEE Signal Processing Magazine*, Vol. 37, No. 3, pp.146–159.

Young, S.I., Girod, B. and Taubman, D. (2020) 'Gaussian lifting for fast bilateral and nonlocal means filtering', *IEEE Transactions on Image Processing*, Vol. 29, No. 1, pp.6082–6095.

Zhang, P., Tang, Y., Hu, J.F. and Zheng, W.S. (2020) 'Fast collective activity recognition under weak supervision', *IEEE Transactions on Image Processing*, Vol. 29, No. 1, pp.29–43.