Modeling and Analysis of Dynamic Consumer Interest and Purchase Intention Based on Traffic Data

Bingquan Wu[a] ,Hongyan Chen[b*],Yingyi Huang[c]

{[a]190728322@qq.com; [b]1976047985@qq.com;[c]279815010@qq.com}

School of Economics & Management, Quanzhou University of Information Engineering, Quanzhou, China

Abstract: In the era of the Internet, massive user data is becoming a precious resource. By analyzing user data, we can understand their behavior, interests, and preferences, and provide better services and products for users. Therefore, modeling and predicting user interests based on traffic data is crucial for the development of enterprises. By using Bisecting K-Means clustering algorithm,a dynamic interest recognition management model and a model to identify the implicit dynamic interests of users were developed in this study with the aim of analyzing the dynamic interests and purchase intentions of consumers. Analysis result show four types of interest transitions: attention, understanding information, attitude, and purchase intention. And users will exhibit different purchase intention behavior under the influence of these interest states.

Keywords: Traffic data; Dynamic interest; Purchase intention; Modeling analysis

1 Introduction

Under China's new development pattern, the country's economic development has shifted towards a dual-cycle model with the inner cycle as the main driver and the outer cycle as a supplementary force [1]. The significance of bolstering consumption to drive economic development was highlighted in March 2021 with the release of the Outline of the 14th Five-Year Plan and Vision 2035 for the National Economic and Social Development of the People's Republic of China. Novel forms of consumption have emerged in this new economic environment, which involve a departure from the traditional consumer decision-making process and a consequent shift in consumers' cognitive needs and decision-making behaviors.

In the past, consumption began with brand and product awareness, followed by the consideration of whether to make a purchase. However, after conducting Big Data analysis, Alibaba, the world's largest e-commerce company, found that the consumer decision-making chain has become disordered. In this context, online retailers are presented with new opportunities and challenges and must seek out an innovative, sustainable online sales model.

With the development of e-commerce platform and Internet technology, consumers' inclination for online shopping is skyrocketing. In order to attract and satisfy consumers, it's necessary for enterprises to explore and dig out the purchase intention development mechanism, so that enterprises can adjust marketing strategies and get consumers to purchase products.

ICEMME 2023, November 17-19, Beijing, People's Republic of China

DOI 10.4108/eai.17-11-2023.2342730

To investigate the mediating and moderating effects of product involvement and trust toward websites in relation to the effects of the attributes of web advertisements on customers' purchase intentions. Jong utilized the PLS (partial least squares) method to analyze the collected data, and demonstrated that consumers with greater product involvement tended to shop in shopping malls[1]. Li investigates the effect of message grouping, an IDA approach supporting review browsing, onto the consumer's information processing and system usage intention [2]. Leong reveal that consumers' experience is the strongest predictor followed by Facebook usage, hedonic motivation, browsing, age, trust motivation, participation, utilitarian motivation, number of children, monthly income and educational level[3]. The specific aspects of behavioral orientation, that is, utilitarian and hedonic motivations, have a significantly positive impact on user intention to browse products on social media[4-5]. Rosillo specifically investigated the effect of hypothesized variables to online shopping orientation[6]. Perceived product quality, purchase intention and perceived risk are specifically affected by individual cultural dimensions [7]. Hina propose a purchasing behavior analysis model based on Hidden Topic Markov Models (HTMM) [8]. The visual appeal variable had a significant effect on hedonic browsing, the portability variable had no significant effect on perceived risk and has a significant effect on utilitarian browsing, and the hedonic browsing variable has a significant effect on online purchase intentions [9]. The information complexity of online forum pages browsing has an inverted U-shape effect on consumers' purchase behavior [10]. Advertorial attributes influenced the consumers' emotional states via control, and emotional states directly affected the information acceptance and purchase intention [11]. Li Chen builds a Commodity Utility-Behavior Sequence (CUBS) dual-utility model to modify the influence on purchase intention prediction [12]. Consumers will have lower purchase intention when they browse negative comments. The psychological mechanism of this is that consumers are affected by perceived risks, resulting in lower brand identity [13].

2Dynamic interest recognition based on individual consumer browsing behavior

In order to increase the conversion rate of online purchases, both academic and commercial researchers have invested in studying personalized recommendations as an effective technique. However, with the shift in the consumer decision-making chain, identifying and managing the user's real-time dynamic interest machine has become a crucial issue for e-commerce platforms seeking to understand the user's shopping journey and process of making decisions. The present study addresses two main questions: Firstly, whether dynamic interest shifts occur in the shopping process of small consumers, and if so, what form of implicit state the interest output takes; and secondly, whether the user's implicit interest feeds back into the user's next browsing path.

To answer these questions, we developed an implicit dynamic interest identification management model based on household browsing behavior. We collected user traffic data (clicks) on Tmall and Taobao as the experimental research platform. We then sank the data hierarchy to the web function keys of e-commerce platforms as data granularity to classify web pages, and applied the Bisecting K-Means clustering algorithm for interest-state mining.

2.1 Dynamic interest recognition management model

The proposed model accurately describes the user's interest state and predicts future purchase intentions. The Howard-Sheth consumer purchase theory wheel framework (Figure 1) suggests that the user's purchase interest changes as they psychologically perceive a product. The model takes this into account and inputs information in the context of browsing, causing the interest level to change. The user's actual shopping interest is affected accordingly in the process of selecting types of interest states, which to a certain extent affects their browsing behavior and decision-making process.

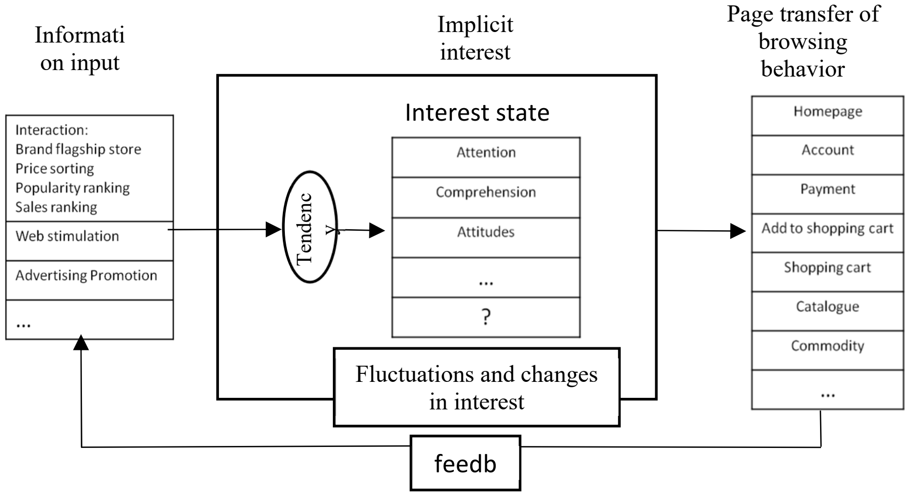

Fig.1: Research framework for real-time interest conversion mechanism based on Howard Sheth theory

2.2 Model construction for identifying users' implicit dynamic interests

During the online shopping process, although users may see similar web content, the e-commerce information system can automatically adjust the content and structure of web pages to provide personalized services if the user's interest changes are identified. To achieve this, we constructed the corresponding implicit interest variables based on click traffic data that reflects the differences in interest generated by users on different web pages. The corresponding web pages were then modeled using Bisecting K-Means clustering modeling.

2.2.1 Quantitative indicators of users' implicit dynamic interests

Users' sessions were extracted from access logs to characterize the behavior exhibited by the user during browsing. We also used three dimensions to represent users' implicit interests after the page classification scheme was established, namely, user selection interest, access time preference, and page interest.

Selection of interest levels

If the user leaves the previous session page q while the corresponding session is in progress, they will be able to choose a larger number of different paths in subsequent browsing sessions. These different types of paths can be represented by n. The selection of pages in i (k=1,2,... ,n) of pages in the scenery degree selection is defined by the Maldives one-step transfer probability, so:

$$P_i = \frac{N_i}{\sum_{i=1}^{n} N_k} \tag{1}$$

N is the user's support for selecting the i-th page.

Access time preference

(1)Duration of page visit

If the user spends a relatively long time on a given page, this indicates a relatively

strong interest in the page, i.e:

$$Time(q) = T_i - T_q \tag{2}$$

Time(q) is the browsing time for users on page q; Ti is start time for page i.

(2)Relative view time rate

For Time(q)，cases where the user views other pages in the same session are typically ignored. The preferences generated on other pages can be expressed in terms of the rate of time the user spends browsing through the page, which can be expressed as follows:

$$Timeratio(i) = \frac{Time(i)}{\sum_{i=0}^{m} Time(i)} \tag{3}$$

m is the browsing time spent for users in a certain session.

Page interest level

(1) Page hit rate

The user's preference for giving attention to a certain page can be expressed as:

$$Clickratio(i) = \frac{visit(i)}{m} \tag{4}$$

where m denotes the total number of pages viewed by the user in the session and visit(i) indicates the number of visits by the user for page type i during the session.

(2)Session access depth

Generally, when the user browses through web pages in the chronological order, a corresponding tag value is generated. This tag value represents the depth of the user's access to the page during the session. For example, when the user opens a shopping application, the corresponding tag value of the page is 2. When the app automatically jumps to the corresponding product A page, its tag value becomes 3. When it jumps to the page of product

B, its tag value becomes 4. As the user browses more pages, the tag value increases accordingly. However, when the user returns to a previously opened page, the session depth returns to its initial state. This suggests that the user may have developed a different state of interest during the second visit to the page than during the first visit. Therefore, the user's session depth can be represented by the page's tag value.

$$Sessiondepth(i) = tabId(i) \tag{5}$$

According to the above formula, if the variable value gradually increases, then the user browses the page for a long time and the corresponding variable value can be regarded as a variable of interest in the session.

2.2.2 Identification clustering modeling of implicit dynamic interest states

Bisecting K-Means Clustering algorithm

The K-Means clustering algorithm focuses on finding points in space as centers, then clusters, integrates, and groups them together close to the sample. After the clustering is complete, an iterative algorithm updates the centroids of the clusters to optimize the clustering results .

Suppose that the data set X = {x1,x2,...,xN} of each of the N browsing pages of dimension V ,xN}∈R×V, then each page xi contains v interest variables. At this point, the clustering algorithm can divide the N pages into K clusters around the cluster center C={c1,c2,... ,ck} and non-empty disjoint clusters S={s1,s2,... ,sK} until the objective function Wk converges, i.e:

$$W_k = W(S,C) = \sum_{K=1}^{K} \sum_{i \notin S_K} d(x_i, c_k) \tag{6}$$

where Wk denotes the sum of the errors of all pages in K clusters from each cluster center, CK denotes the corresponding cluster Sk, and $d(x_i, c_k)$ denotes the distance between the pages xi and CK . When operating clustering algorithms, Wk is typically used as evaluation indicator. Clustering quality is higher when Wk is smaller.

Clustering evaluation with profile coefficients

The contour coefficient is an important metric reflecting the validity of clustering, which can reveal whether a data set presents a consistent state within a cluster. For a particular page xi , the contour coefficient can be expressed as follows:

$$S(x_i) = \frac{b(x_i) - a(x_i)}{\max\{a(x_i), b(x_i)\}} \tag{7}$$

where a(xi) is the average of the calculated distance between xi and all other pages within this cluster, which generally reflects cohesion within the cluster; b(xi) is the separation between clusters that xi can be used to quantify. The contour coefficient is in the interval of [-1,1]. When the contour coefficient S is close to -1, the distance between xi and clusters other than the given cluster is relatively close; xi is assigned to the wrong cluster. When S tends toward 0, there is no natural cluster of xi in the set X and the pages belong to a random distribution state. When S is close to 1, xi is far from other clusters, close to the cluster in which it is presently located, and is within a relatively independent cluster. For a given specified K-cluster clustering of a collection of page data X, the average contour coefficient can be expressed as follows:

$$S_k = \frac{1}{N}\sum_{i=1}^{N} S_i \tag{8}$$

where N is the total number of pages in the traffic dataset and Si can be applied in the validity analysis of the clusters.

3 Experimental design and data description

The experiment was divided into three stages: Preliminary, formal, and final comparison. The click traffic data obtained in the experiment was gathered over 14 sessions and 1,773 pages of access logs to generate a total of seven purchase behaviors. After the corresponding page classification process, we found that the set of pages U contained n=14 page categories. Among these 14 pages, product T was the most viewed when users were browsing, accounting for 33.1%, followed by product C. Product B was a good indicator of users' preference for the product. In-depth interviews were conducted to verify that users' perceived concerns about factors such as price and sales can be represented by mouse click information as well as page navigation information when compared to objective browsing behavioral click traffic, which fully demonstrates that computerized interactive implicit ratings can map users' inherent dynamic interest shifts. A detailed description of page category statistics can be found in Table 1. Table 2 provides a detailed description of implicit interest variable statistics.

Table 1 Page Category Statistical Description

Class name	Home page	Account	Payment purchase	Add to shopping cart	Shopping Cart	Product	Evaluation	Brand	Price	Popularity	Sales volume	Product attributes	catalogue	other
frequency	138	96	7	30	52	170	11	142	17	5	4	588	438	74
Frequency(%)	7.78	5.43	0.41	1.69	2.93	0.63	0.62	8.01	0.96	0.29	0.24	33.1	24.7	4.19

Table 2 indicates that the mean value of 1 is 12.27 s,but at faster browsing speeds, making the mean value of 2, 3, and 4 are 0.72%, 27.68% (<30%), and 28.21%, respectively. These values represent the proportion of pages viewed by the user in the session.

Table 2 Statistical Description of Implicit Interest Variables

Serial No	variable	mean value	standard deviation	minimum	median	Maximum
1	Page Duration	12.27	45.33	0.00	3.00	1492.00
2	Relative page browsing time	0.72	3.43	0.00	0.09	100.00

	rate(%)					
3	Click-through rate of the page(%)	27.68	18.75	0.28	26.20	100.00
4	Session access depth(page)	28.21	25.71	2.01	22.01	102.00

4 Model solution and analysis of results

We constructed a model system using the four interest variables mentioned above, which were derived from the users' implicit dynamic interests. However, we found that the K-class status of purchase interests lacked a certain degree of accuracy. To quantify users' implicit interests, the statistical software R was used to perform the corresponding clustering. Once the quantification was complete, the K-values were optimized using contour coefficients (Figure 2) to identify the user's optimal interest state. Typically, k-values are set between 2 and 8.

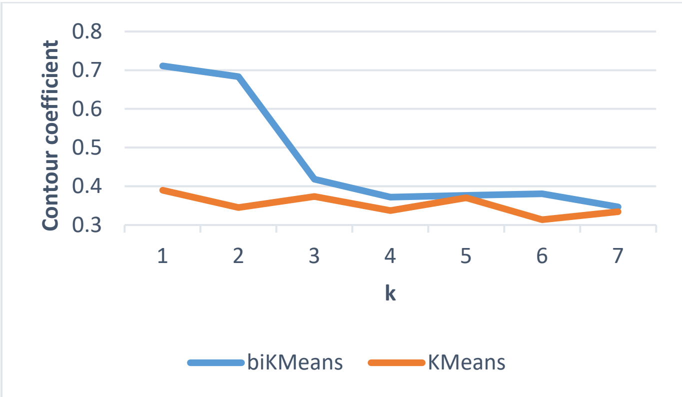

Fig.2 Relationship between contour coefficient and K value

As depicted in Figure 2, the contour coefficients reach their maxima when K is 2. As K increases by 1, the corresponding contour coefficients gradually decrease, and the curve shows a slow decline when K reaches 5, indicating that the characteristics of the new cluster are closer to those of the previous cluster. The contour coefficients show obvious changes when K=4, which can be considered the critical point. The changes in contour coefficients between K=1 and K=4 can be divided into four different types of dynamic interest (Table 3).

Table 3 Cluster centers of four types of user interests

Dynamic interests	k-value	Implicit Dynamic Interest Variables			
		Page Duration(s)	Relative page browsing time rate(%)	Page Click-through rate	Session access depth
cluster 1	1	5.274380	0.5805211	50.57809	16.92206
cluster 2	2	7.042511	0.2328112	19.05582	64.23078
cluster 3	3	11.558825	0.6666170	17.15117	12.02802
cluster 4	4	155.5406	8.1338871	21.006	19.62163

Purchase intention

Purchase intention is a predictive indicator used to forecast future purchases. In the context of this study, "online purchase intention" refers to using one's own or someone else's electronic shopping cart to make a purchase during a corresponding session. For instance, based on the four interest states of the user, more frequent addition of a product to the shopping cart or the site's "favorites" indicates stronger purchase intention. Additionally, the user's browsing time on the e-commerce website is positively associated with their purchase intention. Longer browsing time indicates stronger purchase intention. Based on the above, the use of shopping cart behavior can be divided into two cases: Browsing the shopping cart or adding products to the shopping cart. It is believed that users generate a corresponding purchase intention interest when they spend a long time browsing the site and frequently utilize the shopping cart.

To examine the relationship between the frequency of shopping cart usage and actual purchases, the number of times "Add to Cart", "Browse Cart", and "Check Out" pages appear in a session can be mechanically counted by cluster. More highly correlated clusters in this case indicate stronger purchase intention. Validation results are detailed in Table 4, and the estimated frequency mean of the "Add to Cart & Favorite" page is shown in Figure 3.

Table 4 Correlation test between the number of shopping cart usage behaviors and actual purchase times

Number of shopping cart usage behaviors	Actual number of purchases
1th Cluster	/
2th Cluster	-0.002
3th Cluster	-0.081
4th Cluster	0.679**

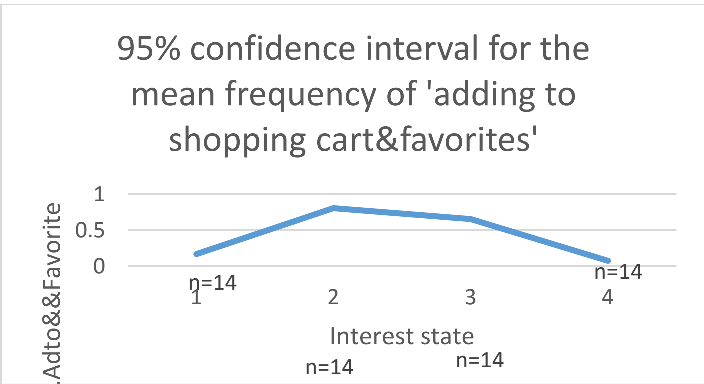

Fig.3 Frequency Mean Estimation of "Add Shopping Cart" Page

(*a*) Attitude states

"Attitude" is defined here as the tendency to develop a preference for an item, which changes over time. When a user adds an item to their shopping cart, it indicates that the user has developed a preference for the item. The frequency of different interests in the 14 sessions was used accordingly to determine the preference developed by the user for said item during their browsing session. As shown in Figure 3, when the user is in Cluster 2 & 3 states, the preference for the product is relatively high. Table 3 shows that when users are in Cluster 3 cluster interest state, their average browsing time is longer comparable to the second cluster. The user appears to have a certain understanding of the product before developing a corresponding interest and purchase intention.

(*b*) Status of concern

As shown in Table 3, the page click-through rate is highest in Cluster 1 when the visual module is viewed for a relatively short period of time, indicating that users remain in the information search phase at this point. The depth of the user's visit is relatively shallow in Cluster 1.

(*c*)State of understanding

(*c*) The “state of understanding” in this context refers mainly to the user's knowledge of the product's brand and information about product itself. The depth of access deepens during the user's page-browsing process as they gather information about merchandise. Understanding of knowledge about goods determines the sales of the goods. If the user is thoroughly informed of a product before making a purchase decision, their purchase intention will be stronger than if they had lacked information. As shown in Table 3, the user has the highest session access depth in the second cluster state, at which point they are in the “understanding stage” of gaining information about products.

6Concluding

In this study, we developed a dynamic interest identification and management model based on user browsing behavior, in the context of Big Data, which can be used to analyze consumers' dynamic interests and purchase intentions. This model provides a new perspective for e-commerce website engineers to analyze consumption and interest data. Our results may be helpful in guiding new strategies for e-commerce platforms and may provide useful data support for e-commerce companies to formulate marketing strategies. The findings of this study may also contribute to a better understanding of consumer behavior and assist e-commerce companies in improving their market competitiveness.

Acknowledgments. This research was supported by the Natural Science Foundation of Fujian Province (China) under Grant(2021J01326; 2023J011803; 2023J011806), the Social Science Planning Project of Fujian Province (China) under Grant FJ2019B099.

References

[1] Kim J U, Kim W J, Park S C. Consumer perceptions on web advertisements and motivation factors to purchase in the online shopping[J]. Computers in Human Behavior, 2010, 26(5): 1208–1222.

[2] Li, Jia; Li, Xinmiao; Yen, David C.; Zhang, Pengzhu. Impact of online review grouping on consumers' system usage behavior: A system restrictiveness perspective[J]. Journal of Global Information Management,2016,24(4): 45-66.

[3] LY Leong，NI Jaafar, S Ainin. Understanding facebook commerce (f-commerce) actual purchase from an artificial neural network perspective[J]. Journal of Electronic Commerce Research,2018,19(1): 75-103.
[4] Mikalef, P.a; Giannakos, M.b;Pateli, A.a. Shopping and word-of-mouth intentions on social media[J]. Journal of Theoretical and Applied Electronic Commerce Research,2013,8(1):17-34.
[5] Pramono Hari Adi;;Faizal Wihuda;;Wiwiek Rabiatul Adawiyah. The Role of Social Media Browsing Intention for Behavioral Outcomes of Young Consumers[J]. Trziste/Market,2017,29(1):39-57.
[6] Rosillo-Díaz; E.;Blanco-Encomienda; F.J.;Crespo-Almendros, E. A cross-cultural analysis of perceived product quality, perceived risk and purchase intention in e-commerce platforms[J]. Journal of Enterprise Information Management,2019,33(1):139-160.
[7] Mio Hotoda; Gendo Kumoi; Masayuki Goto. A Study on Customer Purchase Behavior Analysis Based on Hidden Topic Markov Models[J]. Industrial Engineering & Management Systems,2021,20(1):48-60.
[8] Hina Yaqub Bhatti; Madiha Bint E Riaz; Shazia Nauman; Muhammad Ashfaq. Browsing or buying: A serial mediation analysis of consumer's online purchase intentions in times of COVID-19 pandemic[J]. Frontiers in psychology,2022,13: 1008983.
[9] Lu, Xianghua; He, Shu; Lian, Shaohua; Ba, Sulin; Wu, Junjie. Is user-generated content always helpful? The effects of online forum browsing on consumers' travel purchase decisions[J]. Decision Support Systems,2020,137(0):11336
[10] Chang, Hsin Hsin; Wong, Kit Hong; Chu, Tsun Wei. Online advertorial attributions on consumer responses: materialism as a moderator[J]. Online Information Review,2018,42(5): 697-717.
[11] Li Chen; Hui Zhu. Behavior prediction based on a Commodity Utility-Behavior Sequence model[J].2022,9: 100314.
[12] Sun Liyuxi. The Impact of Taobao’s Negative Comments on Consumer Willingness[J]. SHS Web of Conferences,2023,155: 01015.
[13] Peng Minjing; Liu Xinglin; Luo Ximing; Zhu Mingliang; Zhang Xianyong; Deng Xiangming; Wu Mingfen. Recognizing intentions of E-commerce consumers based on ant colony optimization simulation[J]. Journal of Intelligent & Fuzzy Systems,2017,33(5): 2687-2697.