Human-centric Computing and Information Sciences

February 2021 | Volume 11

HCIS | Human-centric Computing and Information Sciences

RESEARCH(Research Manuscript) Open Access

Human-centric Computing and Information Sciences (2021) 11:8
DOI: https://doi.org/10.22967/HCIS.2021.11.008
Manuscript received June 13, 2020; accepted January 5, 2021; published February26, 2021

Contextual Collaborative Filtering Recommendation Model Integrated with Drift Characteristics of User Interest

Feipeng Guo[1,2] and Qibei Lu[3,4,*]

Abstract

User interest will drift with the change of context, cognitive psychology, and so on, which leads to inaccurate recommendation. In order to address this issue and the traditional recommendation problems such as cold start and data sparsity, this study proposed a novel contextual collaborative filtering recommendation model. First, the reasons for drift of user interest from the perspective of motivation were analyzed, and this study designed a mechanism based on Maslow's hierarchy of needs to analyze the information category and information behavior corresponding to the hierarchy of users' needs. Then, a novel user interest determination algorithm was proposed based on ontology and hidden Markov. Second, this study introduced the concept of user activity and proposed a user activity computational method integrated with context to solve the cold start and data sparsity problems. Finally, the research proposed a dynamic collaborative filtering recommendation algorithm integrated with user activity to diversify the content of candidate recommendation selectively. By monitoring users' feedback and the learning rules of interest drift, this method can discover drift of user interest and make some adaptation actively. The experimental results showed that this model, which integrates with the drift characteristics of user interest, can effectively improve the adaptability to the drift of user interest, and that it has higher accuracy compared with other recommendation methods.

Keywords

Contextual Recommendation, Maslow's Hierarchy of Needs, User Activity, Interest Drift, Collaborative Filtering Algorithm, Hidden Markov

1. Introduction

Currently, the recommendation service can provide personalized information to users, and this has great advantages in terms of solving the problem of information overload [1]. Nonetheless, dynamic

Manuscript received June 13, 2020; accepted January 5, 2021; published February26, 2021.
Corresponding Author: Qibei Lu (luqibei@zisu.edu.cn)
[1]Modern Business Research Center, Zhejiang Gongshang University, Hangzhou, China
[2]School of Management and E-Business, Zhejiang Gongshang University, Hangzhou, China
[3]School of Cross-Border E-Commerce, Zhejiang International Studies University, Hangzhou, China
[4]School of Science and Technology, Zhejiang International Studies University, Hangzhou, China

contexts have a significant impact on the decision of users in selecting products or services [2]. For example, the user may have different preferences for the same product catalog under a different context, such as book as a birthday present for friends, or as material for improving his/her work ability. Thus, recent studies on contextual information recommendation, which considers context factors such as location, time, and psychological characteristics, have attracted wide attention and become a hot topic. Its core idea is to provide information that is consistent with user interest based on the current contexts. Moreover, the changes of complicated contexts may lead to the drift of user interest.

In addition, the researchers realize that the key of the recommender system is to cognize users' psychological needs, and the goal of recommendation activities is to "recommend a service on demand" [3]. Still, on one hand, a user's complex cognitive process changes user interest, which is hard to grasp. Users have random and jumping interests. The traditional personalized recommendation system lacks learning mechanisms, and this prevents the system from responding flexibly to the drift of user interest. On the other hand, for enterprises, only the user's behavior is observable, but the real reasons that make interests drift are hidden. When user interest has changed, the corresponding personalized recommendation service must be able to make timely adjustments. In fact, the drift of user interest is one of the main factors affecting the performance of recommendation system and hindering its development [4]. Therefore, the main purpose of this research is to master changes of user interest in order to analyze and understand users' needs better. It is only through this that, when user interest drifts, Internet platforms promptly recognize it and make reasonable adjustment to improve the effectiveness of personalized recommendation.

The variations of contexts lead to changes of users' target concept; typically, user interest drifts. The drift phenomenon includes incremental drift with static context change, and radical drift with dynamic context or change of users' requirement level. User interest that are seemingly random but actually regular are called interest evolution. The adaptability of traditional recommendation mechanisms to the change of user interest is different. Contextual collaborative filtering recommendation methods show that context sensitivity affects user's information demands and ultimately influences user's behaviors. It includes pre-filtering and post-filtering recommendation based on complex contexts, and recommendation process based on modeling complex contexts. Nonetheless, most researches have not analyzed the correlation among contexts, psychological perception, and interests. They lack description of the interaction mechanisms among related elements, such as context, user, and product/service. Moreover, drift of user interest has become a key constraint to the service quality of information recommendation.

Therefore, this paper proposes a contextual recommendation model integrated with the theory of Maslow's hierarchy of needs. As for the main contributions of this study, (1) in the aspects of dynamic interest, this paper traces from motivation and psychology to information behavior theory and determines the factors that trigger the drift of user interest. More importantly, these theories are used to model the problem that the computer system is capable of capturing and solving. It provides a new interdisciplinary perspective to the development of research directions. (2) This paper studies two aspects of change with user interest factors, including user subjective perception and user context. Unlike the traditional interest change research, the research object focuses not just on the incremental drift phenomenon of user interest but also on the dramatic drift phenomenon, which is random and not obvious. The theory of hierarchy of needs emphasizes the drift of user interest, which is hierarchical and jumping in nature brought about by the changes of user's subjective cognition. It interprets the classification and level of the main user's demands, providing support for the information service adapting to the change of user interest. (3) By analyzing the characteristics of hierarchy and evolution of user's demands, it proposes user interest level judgment algorithm based on ontology and hidden Markov. Then, after analyzing the dynamics of context, it introduces the user activity calculation method integrated with context, and a novel dynamic collaborative filtering recommendation algorithm integrated with user activity is proposed. The contextual recommendation model optimizes the capabilities of traditional mechanisms to adapt to the change of user needs, so it can adopt the dramatic interest drift under the contextual information services and improve the quality of customer satisfaction.

The rest of this paper is organized as follows. After the introduction part, the paper discusses related work in Section 2. Section 3 proposes a novel contextual collaborative filtering recommendation model based on Maslow’s hierarchy of needs theory. Section 4 evaluates the performance of the proposed model and its algorithms. Finally, Section 5 presents the conclusion and future work.

2. Related Research

2.1 Maslow’s Hierarchy of Needs and Drift Mechanism of User Interest

Maslow's hierarchy of needs theory suggests that people's interests and needs will drift with context [5]. User interest is not static, and is affected by individuals or environment over time. Thus, the recommendation model also requires change when user interest is adjusted, so that it can accurately describe the characteristics of the current user interest. The drift of user interest has two features: incremental drift and radical drift [4, 6]. To overcome the drift problem of user interest, Guo and Lu [7] introduced the time-sensitive function and implemented the system for drift of user interest. Based on the concept of recognition and deformation, Geuens et al. [8] researched the semantic concept drift over time in the knowledge organization system to identify interest change for a given context. Wang et al. [9] proposed the pattern discovery method of interest migration based on hidden Markov model (HMM); the discovery algorithm was derived by incremental interest in migration patterns, and the migration pattern of interest was defined as an association rule.

2.2 Contextual Recommendation Method

Contextual recommendation takes context information into consideration in the recommendation process for modeling and forecasting user preference. It defines the “context” as other categories of data. The use of context information in the recommendation system can be traced back to the research of Herlocker et al. [10]. They assumed that the context of users’ task was introduced into the recommendation algorithm at a particular application, which can bring a better recommendation result. Palmisano et al. [11] found that taking context factors into consideration could discover more purchase patterns by recording user’s purchase history data. It could also provide better prediction of potential user’s purchase demand and stimulate user purchase. Thus, it integrated and applied the context factors in information recommendation to distinguish users more accurately and provide more appropriate information resources to users. Adomavicius et al. [12] put forward a three-dimensional contextual recommendation space that expanded the problem model proposed by Adomavicius and Tuzhilin [13].

3. Contextual Collaborative Filtering Recommendation Model Based on Maslow’s Hierarchy of Needs Theory

The incremental drift of user interest is easy to grasp because it can be modeled by observing user behavior data with time information. Nonetheless, users whose interests change have certain jumping characteristics that usually present sharp preferences. It often occurs due to the change in user subjective perception; thus, we should construct the hierarchy and jumping model to deal with such radical drift characteristics. Maslow's hierarchy of needs theory based on behavior motivation mainly elaborates on the classification and hierarchy of people’s needs and the jumping rules among each level. Therefore, a contextual collaborative filtering recommendation model based on Maslow’s hierarchy of needs

(CRMM) is presented in this paper, whose originality are listed below.

First, ontology of hierarchical information category (OHIC) and decision model of hierarchical information behavior (DHIB) are designed based on Maslow's hierarchy of needs in order to determine the level of user interest. Second, the concept of user activity is introduced, and its calculation method that incorporates with context-ACUC (activity calculation method integrated with user context) is presented to solve the cold start and data sparsity problems. On this basis, dynamic collaborative filtering recommendation algorithm incorporating user active (DCFUA) is proposed. The algorithm integrates activity and context into supervision and inspection of user interest drift and explains the classification and hierarchy of user's main demands and rules of user's jumping among these levels. Finally, this paper completes the recommendation by using an improved dynamic collaborative filtering algorithm. It models a significant drift in the behavior of user interest and optimizes the ability to adapt to the radical drift of user's needs in the traditional recommendation mechanism. This model also realizes recommendation content diversity, monitors user interest in evolution and deals with it, and provides high-quality personalized information services that can adapt to radical interest drift.

3.1 User Interest Hierarchical Judgment Algorithm based on Ontology and HMM

Different types of information can meet different levels of user interest. Consistency of information behavior and daily behavior can ensure the applicability of the hierarchy of needs theory to analyze informational behavior. After OHIC and DHIB constructed, UIHOH (user interest hierarchical decision algorithm based on OHIC and DHIB) was proposed to analyze the hierarchical structure of user information behaviors. The mechanism of UIHOH is shown in Fig. 1.

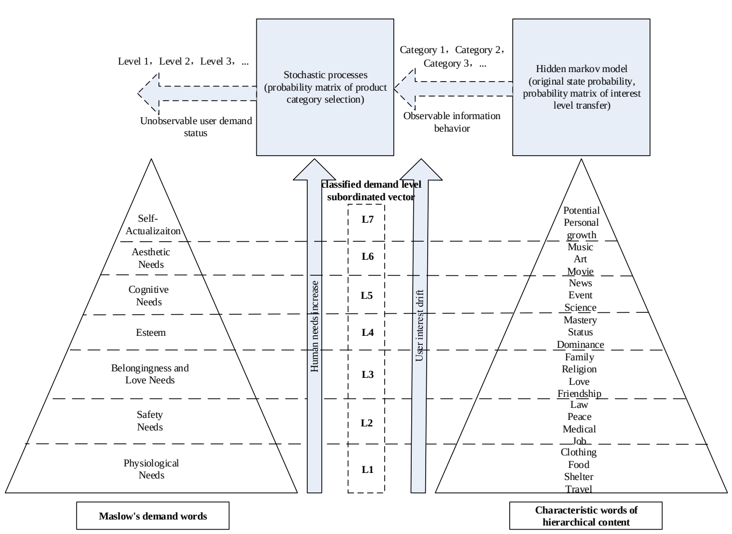

Fig. 1. Framework and mechanism of UIHOH.

OHIC uses ontology to determine the extent to which each category of information content satisfies each level of interests. It includes many category directories, such as digital, cloth, maternal, food, cosmetics, sports, entertainment, etc., eigenvectors of each product category, all interest levels and their characteristic word lists, and subordinated vector wherein each product category belongs to each interest level, e.g., the food category belongs to the first layer, and the maternal category belongs to the third layer. DHIB uses hidden Markov to model the process wherein users jump among different levels of interests. $\mathrm{Levstat}$ is utilized to express hidden state collections of interest levels. $\mathrm{Catobsv}$ is the N

observation state set of product category. Π is the probability of initial interest state. A is the transition probability matrix of interest level, whose element is expressed as A_{ij}. $A_{ij} = P(X_t = j|X_{t-1} = i)$ is the probability of interest state in the *i*-th layer at *t*–1 moment, and it will jump to the *j*-th layer at *t* moment. B is the probability matrix of product category selection, whose element is expressed as B_{ij}. $B_{ij} = P(Y_t = j|X_t = i)$ is the probability when they are in the state of the *i*-th interest layer and the *j*-th information category is selected. $Levstat$, $Catobsv$, A and B are used as input parameters. DHIB uses the Viterbi algorithm [14], whose output is the shift sequence of interest level, with the sequence having the highest probability. The detailed information is described below.

$$\begin{cases} Levstat = \{Lev_1 \cdots Lev_i \cdots Lev_7\} \\ Catobsv = \{Cat_1 \cdots Cat_i \cdots Cat_n\} \\ \Pi = (\pi_i), A = (A_{ij}), B = (B_{ij}) \\ Input = \langle C_1, C_2 \cdots C_n \rangle \\ Output = \langle L_1, L_2 \cdots L_n \rangle \end{cases} \tag{1}$$

Algorithm 1. User interest hierarchical decision algorithm based on OHIC and DHIB (UIHOH)

Input: information behavior corresponding to categories of information sequence.

Output: transition sequence of interest state corresponding to hidden hierarchy of needs.

Step 1: needs hierarchy determination for the information category.

Step 1.1: building the content of each interest level. According to the connotation of the hierarchy of needs theory, it maps to the content of each information category.

Step 1.2: $Need_i$ is the degree representing the information category belonging to i-th layer needs; it then calculates the subordinated vector of interest level of each information category.

$$Need_i = \frac{\sum_{j=1}^{Terms(i)} Weight(j)}{\sum_{k=1}^{7} \sum_{j=1}^{Terms(k)} Weight(j)} \tag{2}$$

$Terms$ represent the intersection between the feature words of this category and the feature words describing a certain interest level. $Terms(i)$ represent the number of intersected elements in i-th layer, and j is the sequence number of feature vocabularies. $Weight(j)$ is the weight of the feature word, with the value using word frequency.

Step 1.3: tool used to construct the ontology and develop a program. Input category C_i, and then obtain the output with subordinated vector $V(C_i)$ of hierarchy of needs.

Step 2: determination of needs level corresponding to information behavior.

Step 2.1: B_{ij} is calculated by the following formula:

$$B_{ij} = \frac{Need_i(C_j) * T_j}{\sum_{k=1}^{n} Need_i(C_k) * T_k} \tag{3}$$

T_j is the visit times of a certain product category, $Need_i(C_j)$ is the subordinated degree of the j-th information for the i-th layer, and n is the number of all product categories.

Step 2.2: the calculation process of the Viterbi algorithm. Partial probability $\delta_i(i)$ is the maximum probability of state i at time t. By using the definition, the probability of all states of the nodes can be obtained recursively. The calculation formula is as follows where B_{ikl} is the probability value of selecting the k-th information category in the interest state of Level i at time l:

$$\begin{cases} \delta_l(i) = \pi(i) * B_{ikl} \\ \delta_t(i) = \max_j(\delta_{t-1}(j) * A_{jt} * B_{jkt}) \end{cases} \tag{4}$$

Step 2.3: after the termination state is obtained, the entire state transition sequence can be obtained through stepwise backtracking. If each state gives back pointer ϕ, $arg\ max$ is used to calculate maximum index j of the value of $\delta_{t-1}(j) * A_{ji}$.

$$\phi_t(i) = arg\ max_j(\delta_{t-1}(j) * A_{ji}) \tag{5}$$

3.2 User Activity Calculation Method Integrated with Context

3.2.1 Concept of user activity

The E-commerce platform has a large volume of user data and trade data, which leads to the data sparsity problem in the “user-product” rating matrix. According to market research, though some users do not buy the product, these users cannot be said to have no interest in this product. How to work out the “user-product” score without purchase relationship is the key to solving the data sparsity problem. Therefore, user activity is introduced to score for “user-product.” User activity is divided into two categories according to whether the user has purchased a product before.

Definition 1 (User activity 1). It refers to the activity degree of user access to the product category. The expression of user activity on the category includes the time spent on the same interest category, the time interval between leaving a category and returning back to the category, and the frequency of category visits sorted by interest level.

$Interest\_Cat(u,i)$ is defined as the activity. $ST(u,i)$ pertains to the time spent on the same interest category. $IT(u,i)$ denotes the time interval between leaving a category and returning back to the category, and $FR(u,i)$ is defined as the frequency of category visits sorted by interest level.

The calculation formula of user activity is as follows:

$$Interest\_Cat(u,i) = \int\big(ST(u,i), IT(u,i), FR(u,i)\big) = \frac{ST(u,i)}{\overline{ST(\iota)}} * W_{ST} + \frac{IT(u,i)}{\overline{IT(\iota)}} * W_{IT} + \frac{FR(u,i)}{\overline{FR(\iota)}} * W_{FR} \tag{6}$$

$\overline{ST(\iota)}$ represents the average time of users staying on the same interest category, $\overline{IT(\iota)}$ is the average time interval between leaving a category and returning back to the category, $\overline{FR(\iota)}$ represents the average frequency of category visits sorted by interest level, W denotes each weight, u is the user, and i represents the category.

Definition 2 (User activity 2). It refers to the activity degree of user who bought certain products. The expression of user activity for product generally includes the purchase frequency, last purchase time, and geographical factors.

$Interest\_Buy(u,i)$ is defined as the activity of user u for product i, and r_{ij} is the relative importance weight of geographical factors D_i. FR indicates the purchase frequency during observation period [0,T], and t represents the last purchase time. The calculation formula of user activity is as follows:

$$Interest\_Buy = r_{ij}(FR \,/\, (T - t)) \tag{7}$$

Purchase frequency is used to measure the frequency distribution of purchasing products within a certain period of time. The unit of time is usually month. When calculating the frequency, the days, weeks, and months should be considered in order to ensure improved accuracy of the calculation of user activity. It should also decrease the interference of sudden events, such as COVID-19. The calculation formula is as follows:

$$FR(u,i) = (D_n + W_n + P_n) \,/\, C \tag{8}$$

C is a constant, representing the number of days in a period when a user activity is evaluated. D_n denotes the activity days of users within C, W_n represents the activity weeks of users within C, and P_n is the activity period of ten days within C.

In summary, this paper uses all or part of parameters to represent user activity; meanwhile, we can set different weights for different users and products.

$$Interest(u,i) = \begin{cases} Interest\_Cat(u,i), activity\ of\ category \\ N * Interest\_Buy(u,i), activity\ of\ purchase \end{cases} \tag{9}$$

(1) N equals $\overline{Interest(u)}$, which is defined as the average activity of user u in the "user-product" matrix, with the product having an order record.

(2) N equals $\overline{Interest(\iota)}$, which is defined as the average activity of product i in a "user-product" matrix; the product has no order record and no user visiting record.

For the user with purchasing record, Definition 2 is used to optimize product rating. Meanwhile, for the user who has no order record and no user visiting record, Definition 1 is used to optimize product rating.

3.2.2 User activity calculation method integrated with context

User consumption behaviors, consumption habits, and cognitive habits have significant regional difference. The degree of economic development in eastern seaboard areas is often higher than that of Midwest, and online shopping population coverage and penetration rate also have a different situation. Moreover, the logistics industry, which is closely related to the online shopping industry, is also more developed in the eastern seaboard than that of Midwest. Therefore, a number of contexts are considered, not just focusing on the last purchase interval time but also considering past purchase times of online shopping individuals. The most important factor is that the geographical factor is integrated into the heuristic algorithm, taking it as the weight of a user's activity indicators. The geographical factor implicitly reflects the regional economy of the online individuals, the relevance between the regional economy and online user consumption structure, and the relevance between the regional economy and online user consumption behavior. Additionally, the traditional collaborative filtering algorithm usually takes questionnaires to allow users to participate directly in scoring. It will increase the difficulty and cost for the E-commerce platform, which has more than 100,000 or 1,000,000 users. In addition, the direct scoring algorithm may also cause subjective bias, which cannot truly reflect the degree of user preferences for products and affect the recommendation results. It is the key point to reflect correctly the user interest in the product in solving the accuracy deviation of collaborative filtering recommendation results. Therefore, the concept of user activity integrated with the context is introduced to calculate the activity, and the details are described as follows.

Definition 3 (Geographical factors and their relative importance weights). The geographical location of online users i and j is named D_i and D_j, respectively. w_{ik} and w_{jk} represent the indicators that affect the resident's consumption characteristics in D_i and D_j, respectively. Per capita GDP (w_{i1} or w_{j1}), per capita disposable income (w_{i2} or w_{j2}), per capita expenditure on consumption (w_{i3} or w_{j3}), and online shopping penetration rate (w_{i4} or w_{j4}) are considered the index to measure regional differences. The relative importance weights of geographical location D_i are further defined as r_{ij}.

$$r_{ij} = \sum_{r=1}^{4}(w_{ir} \,/\, w_{ir} + w_{jr}) \,/\, 4 \quad (i, j = 1, \cdots, n) \tag{10}$$

Definition 4 (Buying patterns of online shopping individual). $X = (r_{ij}, FR, t, T)$ is represented as individual buying patterns where r_{ij} means the relative importance weight of geographical location D_i , FR means the purchasing frequency in observational time $[0,T]$, and t means the last purchasing time. The threshold $Para_{active}$ is used to divide the user activity $Interest(u, i)$.

Inference 1. A higher value of $Interest(u, i)$ means higher relative importance weight of the geographical factor, higher historical purchasing frequency of user, and shorter purchasing interval of last purchase. Then all those mean the possibility that the user is more likely to buy in the future and is more active; the opposite means that the user is less active.

Assumption 1. There is a positive correlation between the possibility that the user is likely to buy in the future and the frequency of historical purchasing frequency. There is a negative correlation between the possibility that the user is likely to buy in the future and his/her last purchasing interval.

Assumption 2. There is a positive correlation between the purchasing possibilities in the future and geographical factor, which means that the higher the per capital GDP, per capita disposable income, per capita expenditure on consumption, and online shopping penetration rate, the higher the possibility that the user will buy in the future.

Assumption 3. When $(u,i) > Para_{active}$, which means that active probability in time T is bigger than threshold $Para_{active}$, then the user is active. When $Interest(u,i) \leq Para_{active}$, which means that active probability in time T is smaller than threshold $Para_{active}$, then the user is inactive.

Assumption 4. To test the judgment quality of the user activity calculation method, user purchase data are divided into the estimation phase and validation phase. When a user's buying behavior occurs at once in the validation period, the user is defined as active; otherwise, the user is defined as inactive. This way, users can be divided into four types, a, b, c and d, by comparing the predicted results with the actual results.

Inference 2. Users classified correctly as active are denoted as "a," which means that $Interest(u,i) > Para_{active}$, and at least one purchase behavior happens in the validation phase. Users who are active but incorrectly classified as inactive are denoted as "b"; this means that $nterest(u,i) \leq Para_{active}$, and at least one purchase behavior happens in the validation phase. Users who are inactive but incorrectly classified as active are denoted as "c," which means that $Interest(u,i) > Para_{active}$, and user purchase behavior does not occur in the validation phase. Users correctly classified as inactive are denoted as "d," which means that In$terest(u,i) \leq Para_{active}$, and that user purchase behavior does not occur in the validation phase.

The aim of the ACUC algorithm is to find an optimal value to divide the active and inactive users correctly; $(a+d)/(a+b+c+d)$ has maximum value, so we can determine individual activity under this threshold.

Algorithm 2. Activity Calculation Method Integrated with User Context (ACUC)

Input: X is an unknown category in online user samples, $X \in \{D_1 \cup D_2\}$, D_1 is the dataset in the estimated phase, and D_2 is the dataset for the validation phase. Its purchase modes are $(\mathrm{r}_{ij}, FR, t, T)$, where r_{ij} is defined as the relative importance weight for the user's geographical location D_i , FR indicates the purchase frequency in period $[0,T]$, and t $(0 < t \leq T)$ indicates the last purchase time.

Output: $Para_{active}$, user activity $Interest(u,i)$.

Step 1: variable initialization, $Para_{active} = 0.0$, $\mathrm{k} \in [0,+\infty]$, $Max = 0$, $Sum = 0$.

Step 2: find the optimal threshold: $Para_{active}$.

for m=1: length(D1)
k= D1 (m)
for j=1: length(D1)
//user number is correctly classified as active and inactive in statistics.
if(k>=D1(j))&&(D2(j)>=1) || (k<D1(j))&&(D2(j)==0)
Sum=Sum+1
end
end
if(Sum>=Max)
Max=Sum
//The value of maximum optimal ratio that meets the proper division for active and inactive user "$\mathrm{Para}_{\mathrm{active}}$"
$Para_{active} = k$
end
Sum=0
end

Step 3: Find user activity, and let $\mathrm{p(X_i)} = 1$, $\mathrm{p(X_i)} = 0$, represent active users and inactive users respectively.

for i=1: length (D1+D2)

If (no purchase record but have access records) then
$Interest(u,i) = Interest\_Cat(u,i)$
If (exit purchase record) then
If((T-t)! =0) then
$Interest(u,i) = N * Interest\_Buy(u,i)$
If($Interest(u,i) > Para_{active} \| (T-t) = 0$) $p(X_i) = 1$ *(current active state)*;
Then p(X_i)=0(current inactive state)
End
Step 4: standardization activity.

$$Norm(Interest(u,i)) = \frac{\frac{w-\bar{A}}{\sigma_A} - \min_{v\in A}(\frac{v-\bar{A}}{\sigma_A})}{\max_{v\in A}(\frac{v-\bar{A}}{\sigma_A}) - \min_{v\in A}(\frac{v-\bar{A}}{\sigma_A})} \tag{11}$$

If the activity value does not fall on interval [0, 1], this may have greater impact on the accuracy of the final personalized recommendation algorithm. Thus, first, we used the standardized *z-score* method to process the user activity and eventually let it match the standard normal distribution. Second, we did the normalization process for the set of user activity that meets the standard normal distribution and made standardized data belonging to [0, 1], where w is the original user activity, $\bar{A}$ is the mean value of the original property A, and σ_A is a standard deviation of the original value of property A.

The value $Interest(u,i)$ may have a situation wherein the denominator is $zero$ $(T-t=0)$. It means that the last purchase behavior occurs at the end of the estimated phase. Generally, it assumes that the user is most likely to shop in the validation phase. Therefore, a large value is given to replace $r_{ij}(x/(T-t))$ in order to ensure that its value is much larger than $Para_{active}$. After the activity is calculated by the ACUC algorithm, judgment should be made for the interest state of the target user. At the same time, the algorithm can use the threshold to determine whether the drift of user interest occurred with the context. For example, when user activity is not greater than the threshold, interests in the category or products probably change, and we should consider whether to change the recommendation strategy.

3.3 Dynamic Collaborative Filtering Recommendation Algorithm Integrated with User Activity

In actual applications, the traditional collaborative filtering recommendation algorithm based on the two-dimensional "user-product" model does not have good accuracy of similarity among users. The main reason is that contexts affect users' behaviors. Therefore, it needs to construct a three-dimensional activity matrix based on "user-product-context."

Assumption 5. Context c_n and context c_m are two different contexts. When the user has no major difference in product preferences under these two contexts, then c_n and c_m are similar.

When user's context attributes change, recommendation is not always needed. Thus, how to give dynamic recommendation service with the change of context is the innovation of this study.

Algorithm 3. Dynamic collaborative filtering recommendation algorithm integrated with user activity (DCFUA).

Input: $Para_{active}$, user activity $Interest(u,i)$.
Output: Set_{Rec}, recommendation products set.
Step 1: Calculate the context similarity by using user activity to predict whether user interests drifts or not.
Step 1.1: Assume user u, with the context changing from c_n to c_m. Under context c_n, the "user-product activity" matrix is $Matrix_{c_n}(u,p)$. Under context c_m, the "user-product activity" matrix is $Matrix_{c_m}(u,p)$. Threshold $Para_{c_i}$ is used to judge whether the context is similar.
Step 1.2: Calculate context similarity $Sim(c_n,c_m)$ between c_n and c_m.
The calculation formula of $Sim(c_n,c_m)$ on product i is as follows.

$$Sim(c_n, c_m, i) = \frac{\sum_{u=1}^{n}(r_{(u,i,c_n)} - \bar{r}_{i,c_n})(r_{(u,i,c_m)} - \bar{r}_{i,c_m})}{\sqrt{\sum_{u=1}^{n}(r_{(u,i,c_n)} - \bar{r}_{i,c_n})^2}\sqrt{\sum_{u=1}^{n}(r_{(u,i,c_m)} - \bar{r}_{i,c_m})^2}} \quad (12)$$

c_n and c_m are two kinds of contexts, and u represents the user collection with activity for product i in the two contexts. $r_{(u,i,c_n)}$ represents the activity of user u for product i in context c_n and $r_{(u,i,c_m)}$ denotes the activity of user u for product i in context c_m. $\bar{r}_{i,c_n}$ represents the average activity for product i in context c_n and $\bar{r}_{i,c_m}$ is the average activity for product i in context c_m.

Step 1.3: Comparing c_n and c_m with the threshold set.

When similarity $Sim(c_n, c_m) > Para_{c_i}$, there is no need to change the recommendation strategy for users. When similarity $Sim(c_n, c_m) \le Para_{c_i}$, and c_n and c_m are judged as a different context, then user's context has changed, and user interest requires drift monitoring and processing.

Step 2: Monitoring and processing the drift problem of user interest.

Step 2.1: Under context c_m, the activity matrix $Matrix_{c_m}(u,p)$ is used to conduct collaborative filtering recommendation for users.

The user activity values in this matrix come from the context matrix that is most similar to context c_m and the recommendation products set is finally obtained as Set_{Cand} based on the recommendation priority in descending order. Some of the products are selected into set Set_{Rec}.

Step 2.2: Using OHIC to calculate the hierarchical membership of all information in recommendation candidate content Set_{Cand} and the set to be recommended, Set_{Rec}, which is denoted as $V(C_i)$.

Step 2.3: By using DHIB, information category sequences corresponding to information behaviors on time window $Para_{Win}$ are obtained as input, with the jumping sequences of the level of interests subsequently obtained.

Step 2.4: Processing method of user's incremental interest drift.

According to the output, the amount of user's browsing information category Cat_k and total browsing number Inf_k are counted. The ratio between Cat_k and Inf_k is called diversity, denoted as Div_k. All levels of diversity Div compose $\langle \cdots Div \cdots \rangle$, which is called the vector of the diversity, representing the degree of diversification of users when they meet each interest level. The hierarchy of needs which has a bigger diversification level of demand, can have more information so as to increase the products category.

Step 2.5: Processing method of user's radical interest drift.

A monitor is used to observe user activity $Interest(u,i)$ on windows $Para_{Win}$, it shows that the user interest changed greatly. Counting the ratio named $A(C_i)$ that the information of category C_i in $Para_{Win}$ is accepted and calculating its average acceptance rate $\bar{A}$, if $A(C_i) < \bar{A}$, then $C_i \in Set_{Low}$, if $A(C_i) \ge \bar{A}$, then $C_i \in Set_{High}$. Calculating the vector center of $V(C_i)$ in Set_{Low} and Set_{High} and marking them as $V_c(Set_{Low})$ and $V_c(Set_{High})$, respectively, i means the category number in set Set, and n means the category quantity in set Set.

$$V_C(Set) = \frac{1}{n} * \sum_{i=1}^{n} V(C_i) \quad (13)$$

Step 2.6: Similarly, calculating the cosine similarity mean of $V_c(Set_{Low})$ and $V_c(Set_{High})$, denoted as $L(Set_{Low})$ and $L(Set_{High})$, respectively, j means the category serial number in set Set, and m means the category quantity in set Set.

$$L(Set) = \frac{1}{m} * \sum_{j=1}^{m} \frac{V(C_j) \cdot V_c(Set)}{\|V(C_j)\| \cdot \|V_c(Set)\|} \quad (14)$$

If $L(Set_{Low}) > Para_{Low}$, and $L(Set_{High}) > Para_{High}$, it means all kinds of information in two sets, Set_{Low} and Set_{High}, can meet similar demand levels. In other words, user interest presents significant hierarchy. Based on this, it can be concluded that changes in context will bring about the jump between levels of user demand, and finally lead to changes in user interest. In that case, we can take action to deal with it.

Step 2.7: Calculating the cosine similarity degree of $V(C_i)$ and $V_c(Set_{High})$ of all information in Set_{Cand}, if the value is bigger than $Para_{Evn}$, then put the information into Set_{Rec}.Calculating the cosine similarity degree of $V(C_i)$ and $V_c(Set_{Low})$ of all information in Set_{Rec}, if the value is bigger than $Para_{Evn}$, then delete the information from Set_{Rec}. All the processes above make the information corresponding to the level of interests more consistent with the next possible jumping level of interests, and then Set_{Rec} is recommended to users.

As to the reasons for calculating the changes of activity after detecting the changes of context, first, it can enhance the calculation efficiency by detecting the changes in context. Second, it can reduce the calculation complexity of "user-product" scoring matrix by computing the change in activity. Giving a recommendation of similar neighbor set and calculating after monitoring context changes ensure high efficiency.

4. Experiment and Analysis

4.1 Description of Dataset and Evaluation Index

The data used in this paper were collected from a B2C platform including mobile commerce data (Table 1). The empirical research time started in November 2018 and ended in October 2019. The dataset included users' purchase transaction records, merchandise comments records, merchandise browse log such as click, purchase, add to cart, favorites, etc. Users who registered and successfully purchased goods in the fourth quarter of 2018 were extracted from the database, and 890 samples were finally obtained. This dataset contained users' personal information, consumption records, consumption statistics information such as number of repeat purchases, last purchase time, average number of monthly consumption amount and number of purchases in the sub-period, etc., user access log such as merchandise category clicks, etc. The first two quarters within a period were regarded as estimation phase, and the other two quarters, the validation phase. The first time users purchased was regarded as starting point 0 in the entire observation period. Therefore, the time of repeat purchases was counted from the occurrence of second consumption. The last time of purchase was counted from the first purchase to the last purchase.

Table 1. The description of dataset

Data sources	Time period	Number of users	Data content	Region
A B2C platform	From November 2018 to October 2019	890	Users' purchase transaction records, merchandise comments records, merchandise browse logs, etc.	Shanghai, Sichuan, Henan, Zhejiang, Beijing, Guangdong, Shandong

The evaluation index of the recommendation method is also an important research topic. Some common evaluation indices include recommendation accuracy, product coverage, user satisfaction, product diversity and novelty, etc. In addition, recall rate and accuracy rate are often used as the evaluation index of performance in the recommendation system. All experimental data were proportionally divided into disjoint training set and test set. The training set was used to construct the user interest model, and then products were selected from the test set by using user interest mode to the target users.

Accuracy of recommendation is usually measured by mean absolute error (MAE) or root mean square error (RMSE). Assuming N_u is the test set for evaluation user u, R_{ui} is the actual score of user u for product i, and R'_{ui} is the prediction score. The formulas are as follows:

$$\mathrm{MAE} = \frac{\sum_{(u,i)\in N_u}|R'_{ui}-R_{ui}|}{|N_u|} \tag{15}$$

$$\mathrm{RMSE} = \frac{\sqrt{\sum(R'_{ui}-R_{ui})^2}}{\sqrt{|N_u|}} \tag{16}$$

For the Top-N problem, it means recommending user N items that are most likely to be bought by the user. It is mainly based on the user's past behavior records to analyze and establish the user interest model. It can measure the accuracy of recommendation by predicting the precision and rate of cover recall.

N is the number of users to predict in the dataset, and $|hit_u|$ is the number of products that user u is predicted to buy from the brand list. pre_u is the number of intersections between product or brand prediction list of customer u and actual product or brand purchase list of user u. The recall rate is calculated as follows:

$$\text{Precision} = \frac{\sum_N |hit_u|}{\sum_N |pre_u|} \tag{17}$$

$$\text{Recall} = \frac{\sum_M |hit_u|}{\sum_M |Buy_u|} \tag{18}$$

M is the number of users who actually have a deal. Buy_u is the number of products or brands that user u actually purchased, hit_u is the number of intersections between the actually purchased list of products or brands and predicted list of products or brands for user u. $F_1 - Score$ is used to calculate the precision and recall rate.

$$\text{F}_1 - \text{Score} = \frac{2*P*R}{P+R} \tag{19}$$

4.2 Experimental Results Analysis of User Interest Hierarchy

First, we marked users' needs. Low-level needs include physiological level (NL_1), security layer (NL_2), love and belonging layer (NL_3), and respect layer (NL_4). High-level needs include cognitive layer (NL_5), aesthetic layer (NL_6), and self-actualization layer (NL_7). It abstracts three behavior sequences simulating user different interest levels to monitor three user browsing information behaviors dynamically. Then, it derives the user's practical jumping path and trend based on OHIC to test whether the model can reversely get the output that is consistent with the simulative jumping trend. As shown in Fig. 2, three groups of behavior sequence, ICV_l, ICV_2, and ICV_3, are the input. As shown in Fig. 3, three groups of interest level shift sequence, ILT_l, ILT_2, and ILT_3, are the output. The results show that the interest transfer curve calculated by the user interest level decision algorithm based on the ontology and hidden Markov is consistent with the actual interest transfer trend. Thus, UIHOH is effective.

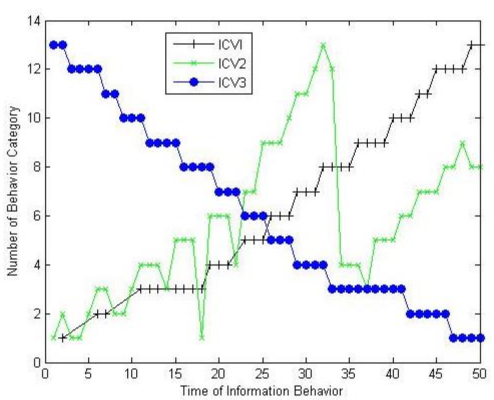

Fig. 2. Users' behavior classification sequences.

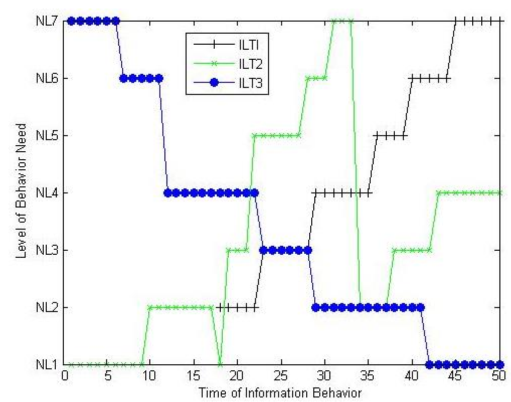

Fig. 3. Users' needs level sequences

Table 2 takes $User_1$ and $User_2$ as examples; with the growth of their age, income, and education level, the interest of $user_1$ shifts from clothing, accessories to clothing and literary. Moreover, they will pay more attention to maternal and child products when they get married. With the growth of age and income and change of geography, the interest of $User_2$ shifts from clothing and food to digital and literary products.

By using the calculation method UIHOH, the interest drift mode of $User_3$, $User_4$, $User_5$, and $User_6$ is described in Table 3.

Table 2. Description of dataset

User	Gender	Region	Marital Status	Income (yuan)	Age (yr)	Education	Topic of interest 1	Topic of interest 2	Topic of interest 3
$User_1$	Female	East of China	No	3,000	25	Bachelor	Clothing	Accessories	Entertainment
			No	6,000	27	**Master**	Clothing	**Literary**	Cosmetology
			Yes	**10,000**	**29**	**Master**	**Maternal and child products**	**Literary**	Clothing
$User_2$	Male	Northwest of China	No	3,000	25	Bachelor	Clothing	**Literary**	Entertainment
		East of China	No	**6,000**	27	Bachelor	Literary	Food	Digital products
		East of China	**Yes**	**9,000**	**29**	Bachelor	**Digital products**	**Literary**	Maternal and child products

Table 3. Instances for drift mode of user interest

User	Description of users' interests drift mode
$User_3$	digital products, maternal and child products, food → digital products, food, literary
$User_4$	clothing, accessories, entertainment → clothing, accessories, digital products → clothing, digital products, cosmetology
$User_5$	clothing, cosmetology, food → clothing, accessories, cosmetology, clothing, digital products, cosmetology → clothing, accessories, digital products
$User_6$	cosmetology, literary form, food → clothing, literary form, food → clothing, maternal and child products, literary form → clothing, maternal and child products, cosmetology → clothing, maternal and child products, food

4.3 Experimental Results Analysis of ACUC

Unlike Beijing, Shanghai, Guangzhou, and other first-tier cities, the physical stories' coverage of medium- and high-grade brands in three-four tier cities is limited. Nonetheless, the number of affluent consumers in these three-four tier cities is increasing; in order to get better-quality goods, they tend to be more interested in online shopping. Thus, the four geographical importance weights, per capita GDP, per capita disposable income, per capita expenditure on consumption, and online shopping penetration rate (including mobile phones), are used in this paper as shown in Table 4.

Table 4. Economy and consumption statistics of seven provinces

	Region index						
	Shanghai	Sichuan	Henan	Zhejiang	Beijing	Guangdong	Shandong
Per capita GDP (Yuan)	157,300	55,774	56,388	107,624	164,000	94,172	70,653
Per capita disposable income (Yuan)	69,442	24,703	23,903	49,899	67,756	39,014	31,597
Per capita expenditure on consumption (Yuan)	45,605	19,711	16,332	32,026	43,038	28,995	20,427
Online shopping penetration rate, including mobile phone (%)	85.1	74.6	80.7	89.8	86.9	97.1	76.2

Source: China Statistical Yearbook 2019, National Economic and Social Statistics Bulletin.

Table 5. Results of relative importance weight of the seven provinces

Region	Sichuan	Henan	Zhejiang	Beijing	Guangdong	Shandong
Shanghai	0.6769	0.6824	0.5624	0.5012	0.5860	0.6489

Sichuan	0.5080	0.3767	0.3241	0.3996	0.4663
Henan		0.3697	0.3183	0.3922	0.4583
Zhejiang			0.4388	0.5250	0.5919
Beijing				0.5849	0.6480
Guangdong					0.5677

As seen from Tables 4 and 5, the formation of regional culture in various regions is closely related to the economic development situation and Internet penetration. The user activity threshold is 0.3 as calculated by ACUC, and individual activity is determined according to the threshold. Threshold calculation is shown in Table 6. The correct division ratio of “active” is the percentage derived by dividing the predicted number of active users by the actual number of active users. The correct division ratio of “inactive” is the percentage derived by dividing the predicted number of inactive users by the actual number of inactive users. The incorrect division ratio of “active” is the percentage derived by dividing the predicted number of inactive users by the actual number of active users. The incorrect division ratio of “inactive” is the percentage derived by dividing the predicted number of active users by the actual number of inactive users. The division has the highest correct rate when the threshold value is 0.3 as calculated by the ACUC algorithm.

Table 6. Calculation of activity threshold

	ACUC: 0.3	**Heuristic algorithm: 21**	**BG/NBD: P(Active)=0.5**
Correct division ratio of “inactive”	87.31	84.67	85.17
Correct division ratio of “active”	73.26	60.37	50.46
Correct division ratio	80.29	72.52	67.82
Incorrect division ratio of “inactive”	12.25	15.78	14.24
Incorrect division ratio of “active”	22.72	36.95	46.14
Incorrect division ratio	17.49	26.37	30.19

4.4 Experimental Results Analysis of Dynamic Collaborative Filtering Recommendation

This study proposed the DCFUA to verify the effectiveness of improved personalized information services for adapting to the drift of user interest. To reflect the effectiveness of the model, this study used indices, such as MAE, average visited time of information category, and acceptance rate, to measure the quality of service. The experimental results are shown in Fig. 4. CBR represents content-based recommendation, CFR means collaborative filtering-based recommendation, and DCFUA is the recommendation algorithm proposed in this paper.

In addition, this study used three novel recommendation methods cited from reference [1-2, 7] to do a comparison. The results listed in Table 7 show that the prediction accuracy of the proposed algorithm is higher than others. The comparison uses RMSE index, and the training set and test set are divided into different proportions, such as 80%–20%, 90%–10%, 70%–30%, and 50%–50%.

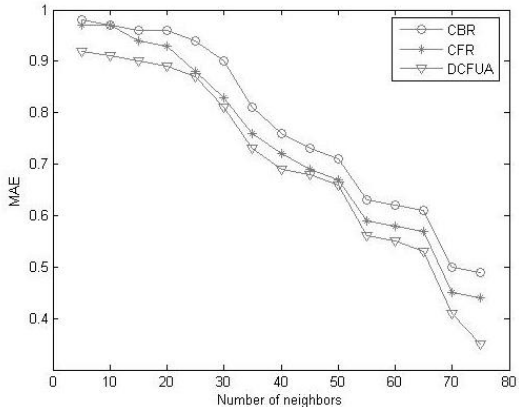

Fig. 4. MAE value comparison of three algorithms.

Table 7. Comparison results of different methods

Algorithm	RMSE index			
	80%–20%	90%–10%	70%–30%	50%–50%
C-CB [7]	0.52	0.55	0.60	0.65
CS-UCF [1]	0.51	0.54	0.57	0.61
CICC [2]	0.43	0.45	0.47	0.52
DCFUA	0.35	0.36	0.37	0.41

We compared the average number of merchandise categories where the user stays and the degree of user's acceptance of the recommendation information, and Fig. 5 shows the results. With the increase in log number, the number of categories that users are interested in overall is in an upward trend, which means the recommendation accuracy gradually increases. The analysis of users' accept information ratio is shown in Fig. 6. Original values in ten experiment sets represent the original accepted recommendation ratio. Boosted values represent the accept recommendation ratio after applying this algorithm. The value is the percentage of users entering the product details page view with recommended links, and they also have transaction records occurring on the same day.

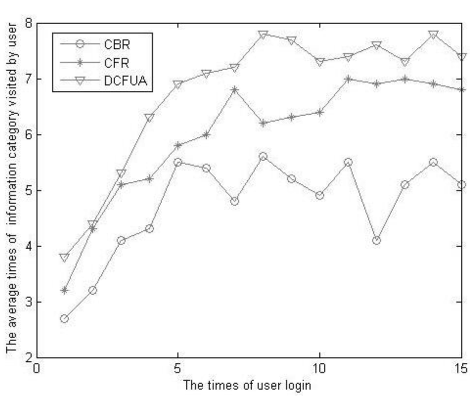

Fig. 5. Three sets of algorithms corresponding average receivable value of recommendation categories.

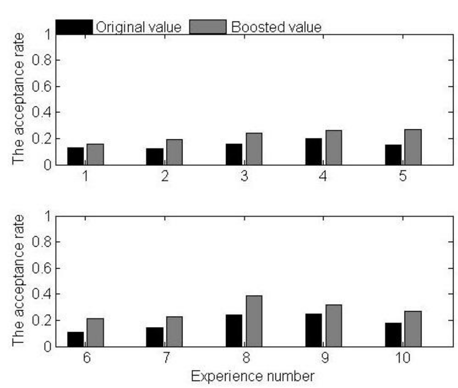

Fig. 6. Comparison of users' receiving information proportion.

First, for the evaluation of recommendation results in the E-commerce platform, 100 users with interest drift were extracted from a total of 890 data sets. Second, different numbers of data sets without the interest drift were extracted from the remaining 780 data sets. Finally, the two data sets above were constructed for comparison. The comparison indices were the prediction rate and recall rate of the recommendation system. The results are shown in Table 8; the experimental results indicate that, after integrating those three algorithms, OHIC, DHIB, and DCFUA, into the recommendation system, the recommendation prediction rate and the recall rate of CRMM have improved; thus suggesting that the model is useful in the practical recommendation system.

To summarize the experiment above, it can be concluded that the contextual information recommendation mechanism based on Maslow's hierarchy of needs is effective and accurate in adapting to the drift problem of current user interest, such as interest migration and interest evolution problems.

Table 8. Comparison of prediction rate and recall rate in recommendation system

Number of users		Prediction rate (%)		Recall rate (%)	
Number of concept drift users	Number of users without concept drift	CRMM is not integrated	CRMM is integrated	CRMM is not integrated	CRMM is integrated
100	100	0.31	0.48	0.21	0.29
100	230	0.39	0.59	0.29	0.38
100	400	0.36	0.56	0.25	0.37
100	580	0.40	0.60	0.30	0.37
100	620	0.41	0.61	0.31	0.38
100	700	0.49	0.70	0.35	0.46
100	790	0.53	0.75	0.40	0.48

5. Conclusion and Future Work

Based on the analysis of current problems in the recommendation model, the contextual information recommendation model integrated with drift characteristics of user interest was proposed in this study. From the perspectives of psychological perception, demand hierarchy, and internal and external context, this study analyzed the reasons for change of user interest, and then designed the interest change capture mechanism to discover the change of user interest in time. Finally, based on the drift model of user interest, a high-quality dynamic contextual collaborative filtering recommendation service was completed. The main conclusions are as follows. (1) Based on Maslow's hierarchy of needs, this study designed mechanisms of information category behavior and interest behavior corresponding to the hierarchy of needs, and it used ontology and hidden Markov to judge the level where the interest belongs. (2) In order to solve the new user cold start and data sparsity problems in the recommendation process, the concept of user activity and the corresponding algorithm were introduced, and, at the same time, contextual user interest model for monitoring interest drift was built. (3) According to the drift model of user interest, a dynamic collaborative filtering recommendation algorithm integrated with user activity was proposed to determine interest drift trend and complete dynamic adaptive recommendation.

In future studies, we will focus on how to use E-commerce transactions comments, social networking, and other factors that affect the contextual recommendation service.

Author's Contributions

Conceptualization, Guo F, Lu Q. Writing—original draft, review, editing, Guo F, Lu Q.

Funding

This work was supported by the Philosophy and Social Science Planning Project of Zhejiang Province, China (No. 21NDJC017Z); the National Natural Science Foundation of China (No. 71802180); the Humanity and Social Science Project of Ministry of Education of China (No. 18YJC870007); the Basic and Public Welfare Research Project of Zhejiang Province, China (No. LGJ21G010001, LGF19G020002); the National Key R&D Program of China (No. 2018YFF0213102); Key R&D Project in Zhejiang Province of China (No. 2021C03143); and the 2019 special plan for improving the scientific research of BoDa young teachers (No. BD2019B1).

Competing Interests

The authors declare that they have no competing interests.

References

[1] L. Xiao, Q. Lu, and F. Guo, "Mobile personalized recommendation model based on privacy concerns and context analysis for the sustainable development of M-commerce," *Sustainability*, vol. 12, no. 7, article no 3036, 2020. https://doi.org/10.3390/su12073036

[2] Q. Lu and F. Guo, "Personalized information recommendation model based on context contribution and item correlation," *Measurement*, vol. 142, pp. 30-39, 2019.

[3] C. Yin, S. Ding, and J. Wang, "Mobile marketing recommendation method based on user location feedback," *Human-centric computing and information sciences*, vol. 9, article no. 14, 2019. https://doi.org/10.1186/s13673-019-0177-6

[4] P. M. Goncalves Jr, S. G. de Carvalho Santos, R. S. Barros, and D. C. Vieira, "A comparative study on concept drift detectors," *Expert Systems with Applications*, vol. 41, no. 18, pp. 8144-8156, 2014.

[5] A. Noltemeyer, K. Bush, J. Patton, and D. Bergen, "The relationship among deficiency needs and growth needs: an empirical investigation of Maslow's theory," *Children and Youth Services Review*, vol. 34, no. 9, pp. 1862-1867, 2012.

[6] L. Xiao, F. P. Guo, and Q. B. Lu, "Mobile personalized service recommender model based on sentiment analysis and privacy concern," *Mobile Information Systems*, vol. 2018, article no. 8071251, 2018. https://doi.org/10.1155/2018/8071251

[7] F. Guo and Q. Lu, "A novel contextual information recommendation model and its application in e-commerce customer satisfaction management," *Discrete Dynamics in Nature and Society*, vol. 2015, article no. 691781, 2015. https://doi.org/10.1155/2015/691781

[8] S. Geuens, K. Coussement, and K. W. de Bock, "A framework for configuring collaborative filtering-based recommendations derived from purchase data," *European Journal of Operational Research*, vol. 265, no. 1, pp. 208-218, 2018.

[9] S. Wang, W. Gao, J. T. Li, and T. J. Huang, "Mining interest navigation patterns based on Hidden Markov model," *Chinese Journal of Computers*, vol. 24, no. 2, pp. 152-157, 2001.

[10] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 5-53, 2004.

[11] C. Palmisano, A. Tuzhilin, and M. Gorgoglione, "Using context to improve predictive modeling of customers in personalization applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 11, pp. 1535-1549, 2008.

[12] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin, "Incorporating contextual information in recommender systems using a multidimensional approach," *ACM Transactions on Information Systems (TOIS)*, vol. 23, no. 1, pp. 103-145, 2005.

[13] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734-749, 2005.

[14] A. Hayashi, K. Iwata, and N. Suematsu, "Marginalized Viterbi algorithm for hierarchical hidden Markov models," *Pattern Recognition*, vol. 46, no. 12, pp. 3452-3459, 2013.