Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 8836000, 10 pages
https://doi.org/10.1155/2020/8836000

Research Article

The Recommending Agricultural Product Sales Promotion Mode in E-Commerce Using Reinforcement Learning with Contextual Multiarmed Bandit Algorithms

Jyh-Yih Hsu,[1] **Wei-Kuo Tseng**,[2] **Jia-You Hsieh**,[3] **Chao-Jen Chang**,[3] **and Huan Chen**[3]

[1]*Department of Management Information Systems, National Chung Hsing University, Taichung City 40227, Taiwan*
[2]*Cornerstone Center for Academia-Industry Research, National Chung Hsing University, Taichung City 40227, Taiwan*
[3]*Department of Computer Science and Engineering, National Chung Hsing University, Taichung City 40227, Taiwan*

Correspondence should be addressed to Huan Chen; huanchen1107@gmail.com

Received 19 September 2020; Revised 15 November 2020; Accepted 1 December 2020; Published 29 December 2020

Academic Editor: Jason C. Hung

In recent years, sales of agricultural products in Taiwan have been transformed into electronic marketing, and agricultural products with better consumer orientation have been recommended, and farmers' income has been improved through sales websites. In the past, *A*/*B* testing was used to determine the degree of preference for website solutions, which required a large number of tests for evaluation, and could not respond to environmental variables that made it difficult to predict the actual recommendation in advance. Therefore, in this study, the reinforcement learning model combined with different contextual Multiarmed Bandit algorithms can be tested in data sets of different complexity, which can actually perform well in changing products. It is helpful to predict the preferences of the promotion model.

1. Introduction

In recent years, governments of various countries have spared no effort to promote electronic sales of agricultural products. They have not only cooperated with private businesses to set up websites for selling agricultural products [1] but also guided local farmers' associations to establish online shopping malls [2]. This shows that it is important for the electronic sales of agricultural products. The same agricultural products sales websites are also facing the problem of how to market and promote agricultural products websites.

Therefore, in e-commerce websites [3], to increase sales has always been an important issue for website operation. In the field of e-commerce, understanding consumer characteristics and behaviors and to sell recommending products is one of the important goals in e-commerce websites.

In order to predict the will of the consumer, we must rely on the model and find the relevance from different features, such as age, gender, and region. However, because of the user characteristics, we cannot see the correlation with consumer promotion preferences manually, so we must rely on models to find the correlation from the features. In order to improve the traditional solution selection problem, more and more websites use predictive models to solve traditional selection problems. Generally speaking, predictive models are usually implemented using machine learning methods, and there is supervised learning in machine learning, unsupervised learning, semisupervised learning, and reinforcement learning. Among them, reinforcement learning is the situational dobby algorithm that will be used in this article. Reasons for choosing reinforcement learning in this article are as follows:

(1) The e-commerce website will not obtain any data that can train the model before consumers actually come to browse and consume, so it cannot directly use supervised learning

(2) The consumer trend environment of e-commerce websites is constantly changing, and a known data set cannot be used for training and prediction

The situational Multiarmed Bandit algorithm is one of the most commonly used algorithms in reinforcement learning [4]. The contextual Multiarmed Bandit algorithm mainly uses feature vectors to perform calculations. Therefore, when the e-commerce website obtains consumer characteristics, the model can immediately use the contextual Multiarmed Bandit algorithm to obtain the best sales strategy to recommend products and then meet consumers' behavior to enhance purchase intention. The most famous one is the LinUCB algorithm because the LinUCB algorithm is widely known because it has obtained good results in the research using the data set of the Yahoo homepage recommended news in the United States [5]. With the development of the LinUCB algorithm, a variety of situational Multiarmed Bandit algorithms are proposed, each with different operating characteristics.

The goal of this research is to use the contextual Multiarmed Bandit algorithm to perform operations with contextual features, and the context is the environmental feature. The environmental characteristics refer to the characteristics carried by the user and the characteristics of the lever itself, such as the user's gender, age, the product characteristics of the lever itself, the type of product, and the brand. Therefore, the situational Multiarmed Bandit algorithm uses environmental characteristics to do calculations.

In the experimental results, we simulated an agricultural shopping website for an e-commerce website, using instant consumer feedback data to study the recommendation effect of the situational Multiarmed Bandit algorithm on the e-commerce website. We use LinUCB, Hybrid-LinUCB, CoLin, and hLinUCB, and they are the effect of the four situational Multiarmed Bandit algorithms, and then, analyzing and testing the situational Multiarmed Bandit algorithm in the three discount modules of buy one get one free, 10% off, and free shipping. In the simulation, users like to recommend the situation, and the advantages of the UCB algorithm selection method compared with the traditional A/B testing scheme, and the noncontextual Multiarmed Bandit algorithm are proved by experiments.

In summary, there are three main contributions of this work as follows:

(1) In the absence of a known data set for model training and prediction, we propose a reinforcement learning method to train the model based on the consumer trend environment.

(2) The situational Multiarmed Bandit is applied to the e-commerce website. The pull bar and user characteristics are used as feature input, and the model is adjusted with real-time feedback data to enhance consumers' purchase intention.

(3) Four algorithms are applied to e-commerce problems, and then, it is analyzed that the situational Multiarmed Bandit has better performance than traditional A/B testing.

(4) We design multiple sets of different experimental environments to evaluate the effects of different algorithms under different mechanisms. LinUCB has a high degree of recognition of user characteristics and product characteristics, and Hybrid-LinUCB has better results under consideration of effective changes.

The remainder of this paper is organized as follows: Section 2 discusses the related work of this research. Section 3 will describe our proposed method. Section 4 shows the effectiveness of the method and compares it with other methods. Finally, we will make a conclusion of this research in Section 5.

2. Related Work

2.1. Traditional Sampling Method A/B Testing. A/B testing is mainly based on a random nature of testing user preferences. First, the two schemes with only one variable difference between them are randomly recommended to users, and the same number of test results must be obtained. It is also necessary to ensure that the user combination characteristics between the schemes are similar to ensure fairness and then determine which scheme is compared to receive user preferences. In the study [6], A/B testing was used to evaluate the block format of the homepage of the website which scheme can achieve a better conversion rate.

2.2. Multiarmed Bandit Algorithm. The Multiarmed Bandit algorithms are mainly to study how to obtain the best total return in the least number of attempts. The core concept that mainly affects the algorithm of the Multiarmed Bandit is how to balance exploration and exploitation. Exploration refers to obtaining feedback results through trial methods, and development refers to the prediction of the algorithm through the feedback results of previous exploration and the number of explorations. That will be the highest return method and uses this method to get the expected better return.

Multiarmed Bandit algorithm is divided into two types. One is the Context-free Multiarmed Bandit algorithm, and the other is the contextual Multiarmed Bandit [6].

2.3. Noncontextual Multiarmed Bandit Algorithm. The epsilon-greedy algorithm was the first proposed by Chirs Watkins et al. [6], which mainly uses the e parameter to affect the probability of exploration and utilization. In each round of selection of the tie bandit, there is a probability of e to randomly select a tie bandit to explore. This strategy is used to avoid the initial selection error of the chance that the best tie bandit cannot be found. As for the exploit lever, in formula (1), the probability of $1 - \varepsilon$ is used to select the lever with the largest average reward $\widetilde{p}$, and the average reward $\widetilde{p}$ is the sum of each reward_i divided by the lever, the number of times k is used.

$$\widetilde{p} = \frac{\sum \text{reward}_i}{k}. \tag{1}$$

The core algorithm of the Thompson sampling method is to use the beta distribution to take into account the concept of exploration and utilization [7]. In formula (2), when the lever is selected ($X\_t = k$), if feedback is obtained, the number of positive feedback α_k is added. If no positive feedback is obtained, then the number of negative feedback β_k will be increased by 1.

$$(\alpha_k, \beta_k) \leftarrow \begin{cases} (\alpha_k, \beta_k), & \text{if } x_t \neq k, \\ (\alpha_k, \beta_k) + (r_t, 1 - r_t), & \text{if } x_t = k. \end{cases} \tag{2}$$

Because the probability density function beta distribution is calculated using (α, β) parameters, the following characteristics are produced:

When α is higher and β is lower, there is a higher probability of achieving higher expectations (α, β). The higher the sum of the two parameters, the more concentrated the probability range of obtaining the expected value.

In order to improve the situation that the greedy algorithm may abandon the potential best pull bar and the inaccuracy of the Thompson sampling method beta distribution, Auer [8] proposed the UCB algorithm. In formula (3), the UCB uses times $T_{j,t}$. The total number of times the tie bandit used is t, and the average reward of the tie bandit j and $\overline{x_j}$ is used to calculate the expected value as show in the following formula:

$$\overline{x}_j(t) + \sqrt{\frac{2 \ln t}{T_{j,t}}}. \tag{3}$$

From the formula (3), it can be known that, in the initial trial stage, the natural logarithm on the right has a considerable initial impact, so every tie bandit will be tried, but when the number of tie bandits $T_{j,t}$ is greater, at that time, the probability of choosing to use the lever will be higher and higher to achieve the goal of increasing the total income.

2.4. Situational Multiarmed Bandit Algorithm. Because the noncontextual Multiarmed Bandit algorithm does not add features to the calculation, it only calculates the profit and the number of attempts. Under such conditions, it is difficult to meet the current complex forecasting needs. Therefore, in 2010, Lihong Li et al. proposed the LinUCB algorithm, which is a situational Multiarmed Bandit algorithm [5].

The LinUCB algorithm here sets the expected return characteristic of each tie bandit a as $x_{t,\alpha}$ and then sets an unknown coefficient θ_a^* for the tie bandit, so the combination is $x_{t,a}^{\top}\theta_a^*$. The expected payoff of the drawbar (expected payoff) $r_{t,a}$ is the feedback of the current drawbar and feature combination, as in the following formula:

$$\mathbf{E}\left[r_{t,a} \mid x_{t,a}\right] = x_{t,a}^T \theta_a^*. \tag{4}$$

Hybrid-LinUCB was also proposed by Lihong Li and others who proposed LinUCB. The main difference from LinUCB is that Hybrid-LinUCB has an additional array of $z_{t,\alpha}$, A_0, and b_0 as a whole array of environmental parameters. The expected value formula (5) is as follows: $z_{t,\alpha}$ refers to the feature dimension array of all users and levers, β is the expected value coefficient of $z_{t,\alpha}$, $x_{t,a}$ is the input feature, and θ_a is the expected value coefficient of the tie bandit.

$$\mathbf{E}\left[r_{t,a} \mid x_{t,a}\right] = z_{t,a}^T \beta^* + x_{t,a}^T \theta_a^*. \tag{5}$$

In [9], the author believes that the traditional situational dobby algorithm ignores the interaction between users, so the relationship array W is added to the algorithm for the feature influence between products. The intention of the algorithm is that if user A likes product C, user B, who has highly similar characteristics, will also like product C. In the CoLin algorithm formula (6), C_t is used to calculate the influence of the relational array, W is the relation matrix of the tie rods, and I is the unit matrix of the number of tie rods multiplied by the number of tie rods. It is used to calculate the characteristics of the tie rods and the user. The specific gravity of the C_t array is affected by the value of α, so the value of α here does not simply represent a parameter for exploration and utilization.

$$C_{t+1} \leftarrow \left(W^T \otimes I\right) A_{t+1}^{-1} (W \otimes I). \tag{6}$$

hLinUCB was also proposed by Qingyun Wu et al. [10]. In the hLinUCB algorithm, it is assumed that the hidden characteristics of consumers affect the value of θ. In the research of Koren et al. [11], it was confirmed that, it can be achieved through matrix factorization. Obtain the hidden interaction parameters between features. In the actual situation, in a large part of the situation, it is impossible to obtain all the characteristic data, so the author believes that the influence of hidden characteristics can be obtained by ridge regression. The author adds v_{a_t} and θ_u^v to formula (7) to calculate hidden features; v_{a_t} represents the hidden feature, θ_u^v represents the θ value of the hidden feature, $r_{a_t,u}$ is the feedback value, and η_t is the environmental noise.

$$r_{a_t,u} = \left(x_{a_t}, v_{a_t}\right)^T \left(\theta_u^x, \theta_u^v\right) + \eta_t. \tag{7}$$

In addition, another feature of hLinUCB is that the algorithm adds a new random initial array to predict the θ parameters of hidden features can get a fast convergence effect.

2.5. Research on User Preferences of E-Commerce. In the past, many researchers used various methods to predict user preferences on promotion models. Wan et al. proposed a matrix decomposition framework with nested features to model preferences and price sensitivity simultaneously, which can be used to obtain economic insights into consumer behavior and provide personalized promotions [12]. Ling et al. proposed a combined deep learning method FC-LSTM, aiming at multiple online promotion channels, using the characteristics of interactive communication between customers and promotion channels to estimate users' purchase intentions. The result proves that the deep learning method proposed in the paper does improve the accuracy and $f1$ score [13]. Vanderveld et al. use the customer relationship management system for analyzing every aspect of the relationship between each customer and our platform based on the life cycle value. It can quickly iterate new products and find the

current buyer frequency in the best inventory [14]. Cai et al. coded each state to establish a Markov model as a decision, using deep deterministic policy gradient, gated recurrent unit model, greedy myopic, and linear UCB methods, through ϵ-Greedy, ϵ-First strategy, UCB1 strategy, and Exp3 strategy, to compare the reward of each time step to show different performance and integration guarantees [15]. Broden et al. proposed the Thompson Sampling Bandit Policy, using Multiarm Bandit within bandit ensemble for e-commerce recommendations, which can coordinate the collection of basic recommendation algorithms for e-commerce and various behavior-based and attribute-based predictions. The problem was found that the context turns into a Multiarmed Bandit, using precision, recall, and normalized discounted cumulative gain as an evaluation indicator [16].

3. Materials and Methods

3.1. Planning Stage. When performing a Multiarmed Bandit algorithm, the most necessary thing is to first define the data features we will need to use and then obtain the data. We simulate different numbers of users and user characteristics according to the determined data features and randomly define the user's preference sales plan, agricultural product preference, and user characteristics. These preference sales plan preference features include time characteristics in order to satisfy the characteristics of e-commerce websites with great changes.

In the Multiarmed Bandit algorithm, we must first define the roles of the user and the lever. In LinUCB, for example, the lever maintains its own feature pattern for recording feedback or other parameters, such as A and B arrays in LinUCB. Therefore, it should be noted here that although in the Multiarmed Bandit algorithm, the lever can be added and changed; the increase of the lever will increase the calculation time and memory consumption. Hence, the role defined as a tie rod should be a fixed and identifiable role, such as a product in an e-commerce website or the subject matter itself in financial investment, rather than selecting items that will continue to increase as a tie rod, such as consumers or users.

In the literature [5], the article was used as a lever, but in the literature [9], the user was used as a lever. The difference between the two articles is that the former article is due to many users but the number of articles is fixed. And in the latter case, due to the experimental environment, there are few users but a considerable number of articles [9]. So, in the Multiarmed Bandit algorithm, the definition of the lever role must also be considered in accordance with the environment.

The following lists the features of the lever role in the Multiarmed Bandit algorithm:

(1) Noninfinitely new data, for example, products will not be added infinitely to the website

(2) Recognizable data, for example, can correspond to a certain drawbar and maintain the drawbar array continuously

Then, the three-stage experimental process is shown in the Figure 1. In the first stage, we use the data set that simulates the browsing of users of agricultural shopping websites to evaluate the Multiarmed Bandit algorithm. These simulated levers are defined as products and then use the data set to apply to the Multiarmed Bandit algorithm to obtain the best choice of product solution and further obtain the reward value to compare the performance of the Multiarmed Bandit algorithm. Therefore, because we use a data set that simulates the user's browsing, we emphasize the sensitivity of the algorithm in using the data set features.

The second stage will use the optimal setting parameters obtained in the first stage to apply to each algorithm. Using simulated agricultural products shopping websites to browse the web and purchase action programs will generate different feedback scores due to different actions. In addition, it is possible to modify the product promotion strategy by simulating the promotion strategy suggested using the Multiarmed Bandit algorithm under the limited commodities by the merchants of the agricultural shopping website and even compare the total and average revenue.

The third stage uses the better-performing algorithm obtained in the second stage to test the preferences of simulated consumers. The predicted benefits of the three discount modules are buy one get one free, 10% off, and free shipping. By analyzing the recommended discount module of the algorithm and simulating the change of user preferences, it can prove the advantages of the situational Multiarmed Bandit algorithm compared with the traditional A/B testing method.

3.2. Simulation Data. The generation of simulation data must produce the characteristics of multiple users and multiple products, which are used to compare the performance of the multiarmed bandit algorithms between the number of users and the number of different products.

Feature standardization is mainly used to avoid the uneven impact of feature parameters on the array [17]. For example, assuming that the user's feature browsing time is morning, afternoon, evening, and early morning, we can divide it into 0.01, 0.33, 0.66, and 0.99. The following parameters are between 0 and 1.

The feature parameters used in this article are common consumer feature data. From the literature, we can see that some operators use the number of consumer web views, stay time, clicks, page scrolling, and mouse movement trajectory data as a consumer's group characteristics. And then use algorithms to push discount modules to consumers according to the feature data to improve performance. In our experiment, we selected the following user characteristics data and the current environment state, and we organized the following characteristics as shown in Table 1.

The user features are the data filled in by the user, and the time feature represents the user's shopping habits, such as buying in the morning or afternoon, buying things on Mondays, and the month representing the festive period that affects shopping.

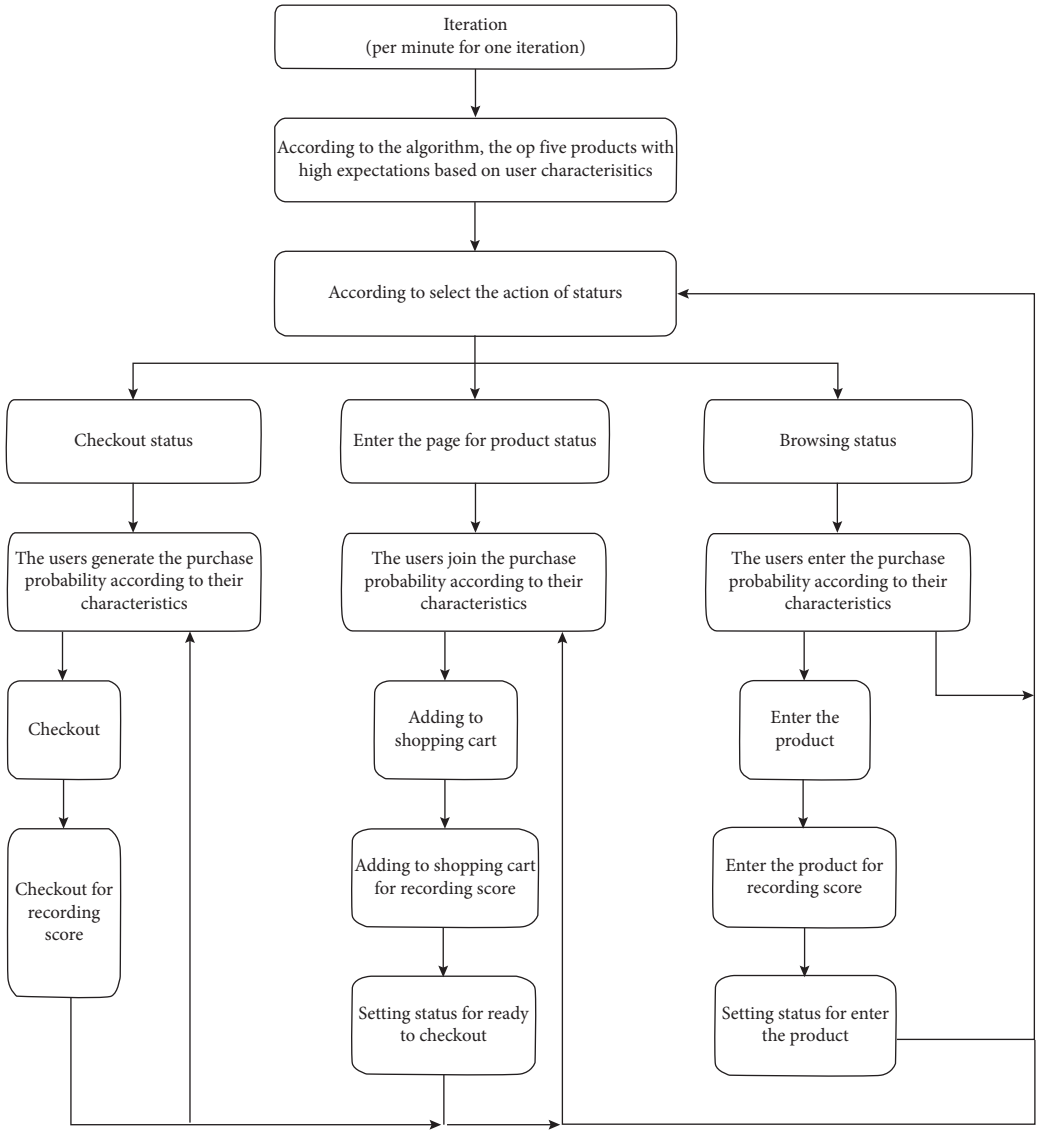

Figure 1: Planning the experiment process.

Table 1: Features collected by users.

Parameter type	Feature	Description of data
User feature	Sex	Male or female or unfilled
User feature	Year of birth (age)	\A.D.
User feature	Occupation	Occupation type, ex: agriculture, animal husbandry, technology, military, and public education.
User feature	Income	Using level distance, ex: less than 30,000 yuan, 30,000 to 60,000
User feature	Marriage	Ex: married and unmarried
User feature	Education level	Ex: elementary school, middle school, and university
Time variable	Current time	Ex: Morning, afternoon, evening, and early morning
Time variable	Month of date	Ex: January and February
Time variable	Day of the week	Ex: Monday and Tuesday

The lever selected in the situational Multiarmed Bandit algorithm defined in this paper is a promotion scheme for agricultural products. In the Multiarmed Bandit algorithm, not only the characteristic of the user is included in the calculation but also the characteristics of the product itself.

On the agricultural product sales website that we used to test our simulation in this article, we will provide merchants with instructions for filling in the characteristics of agricultural products. Some important characteristics are shown in Table 2.

Table 2: Characteristics of agricultural products on agricultural product sales websites.

Parameter type	Feature	Description of data
Product	Types of agricultural products	Ex: Chinese cabbage, cabbage, and green.
Product	Counties and cities of agricultural products	Ex: Taichung, Taipei, and Kaohsiung.
Product	Special attributes of agricultural products	Ex: organic label, production and sales resume, safe fruits and vegetables, and label of Jiyuan Garden.
Promotion mode	Main promotion mode	Ex: immediate product discount, free shipping, and full discount.

After we define the user features and the agricultural features, then formalize the user features, and use the user's features to calculate the degree of preference for the product, calculate the probability of purchasing the product, and finally record.

The data set is divided into three data sets, with 800 people to 5 products on 30 days, 800 people on 30 days to 10 products, and 30 days 800 people to 20 products. The number of people fixed at 800 is the daily number of visitors to the simulated site, and the change in the number of products is the complexity of the simulation environment selection. The reason why the products are divided into 5, 10, and 20 is also due to the increase in the calculation and quantity of products between the two algorithms, Hybrid-LinUCB and CoLin. Therefore, it must be limited to 20 products in order to complete the experiment. The main reason for dividing 30 days is that date features are also included in user parameters, so the performance sections of the algorithm are listed daily for comparison.

3.3. Stage 1: Algorithm Parameters and Efficiency. In this paper, the first phase of the experiment will compare the different performance between the different situational multiarm machine calculations. The simulated agro-shopping site user browsing data set is used to test the algorithm's learning ability for the dataset and the reward is recorded, and the effect between the number of different products and the different algorithm parameters on different multiarmed bandit algorithms is tested.

In the collection of experimental results, at least 30 repeated experiments will be taken in each experimental stage to make an average value, because we have added the concept of sampling to the use of simulated agricultural shopping website user browsing data sets. Therefore, in each update of the contextual Multiarmed Bandit algorithm, different feedbacks will be generated due to different sample records, which have a chance of affecting factors. So, an average value must be obtained to review the performance to obtain an objective argument.

3.4. Stage 2: Simulation of Browsing Experiments on Agricultural Product Sales. In terms of the previous theory, we use a simulated agricultural shopping website user browsing data set to test the efficiency of the algorithm to select the best lever, but this is far from the actual user browsing consumption. Because the feedback of the real-life algorithm will directly affect the user's next consumption, and we also need to verify the effectiveness of the algorithm for maximizing the promotion plan and website revenue; we will use the simulated agricultural shopping website environment to achieve this research aims.

Simulated products will randomly produce 5, 10, and 20 different agricultural products with different prices and product characteristics. In this product, the promotion plan aims to provide the simulated agricultural shopping site users with the opportunity to purchase and update the product promotion plan every day. Simulated users purchase products at certain time intervals to perform operations, click products or add shopping carts, and check out, and there is a purchase limit, a total of 24 hours a day, simulated agricultural shopping site 30 days of data, and calculate the total daily income. And total feedback score changes to compare the efficiency of the algorithm.

In addition, this side not only simulates the consumer action of the agricultural shopping website but also adds the decision-making simulation of the agricultural business. There will have some feature argricultural products in a year, such as strawberries in spring, watermelons in summer, and pears in autumn. Therefore, in this simulation experiment, we will trigger the update of the promotion strategy event at a fixed time every day to simulate the characteristics of seasonal replacement of agricultural products.

The algorithm will be based on the user features of the previous day to determine the promotion strategy to be used in the day's products, and the promotion strategy has a corresponding degree of preferential, the higher the degree of preferential, the lower the business score, the lower the level of concessions, the higher the business, but the relative consumer purchase rate will also decline. This side simulates the complex action of consumers and merchants of agricultural shopping website, mainly to be closer to the reality of agricultural product selling sites will encounter seasonal replacement of products, so as to compare the different situational Multiarmed Bandit algorithms, the more details for each action type, score, and description are shown in Table 3.

3.5. Stage 3: Comparison of Discount Recommendation Methods. In the second stage, we can obtain a better algorithm in the simulation of agricultural shopping sites, and then, we apply the algorithm into the discount module, in which we will have three simulation experiments. First of all, we define buy one-to-one, 10% discount, and free shipping as three preferential modules.

Table 3: Simulated user website action score.

Action type	Score	Description
Click on the product	1	Write a score when you click to enter the product page
Adding to shopping cart	Product price multiplied by coefficient	When adding a product into the shopping cart, write a score
Shopping cart checkout	Commodity price multiplied by coefficient	Write the product price into a score when the shopping cart is checked out

In the third phase of experiment one, we assume that consumers have a higher willingness to buy goods with a lower average unit price after total shopping cart plus freight and that consumers have the highest return on buying one-to-one preferential module merchandise, but pay more at a time. Then, there is the 10% discount group merchandise merchants slightly less for free shipping, the highest merchant income, but to bear the cost of freight. This experiment compares the situational Multiarmed algorithm in three preferential modes with the average unit price orientation of users recommended to reflect the situation.

In the third phase of experiment II, we will simulate consumers plus their own preferences and assume the correlation between user characteristics and preferences, to determine whether the algorithm can correctly identify the relationship between features and preferences, and compared with the traditional A/B testing method and the UCB algorithm of the Non-multiarmed bandit algorithm, whether it can save the number of times and achieve the advantages of learning with the user's preferences.

4. Results and Discussion

4.1. Arrangement of Parameters. After our experiments, we can know that the α constant in LinUCB, Hybrid-LinUCB, CoLin, and hLinUCB is related to the characteristics of the data set. The constant α does not only have to be larger or smaller but does also have setting which was based on the current environment. Therefore, it must be noted that, in each performance of the algorithm, the current sample will determine the difference in user characteristics, and there will be a certain degree of uncertainty in the simulation test.

Finally, based on the above test, we can sort out the α constants that perform better in the browsing data set of 800 people and 20 product simulated users among LinUCB, Hybrid-LinUCB, CoLin, and hLinUCB, as follows in Table 4.

We will use the above parameters in the second stage of the simulation experiment to obtain the performance of the algorithm in the experiment that simulates user browsing.

Table 4: Sorting out the α constants that the algorithm performs better in the 20 product simulated user browsing data sets.

Algorithm	A
LinUCB	0.05
Hybrid-LinUCB	0.8
CoLin	0.4
hLinUCB	0.2

4.2. Simulate In-Service User Browserling. As for the calculation of the feedback score, here we add a feedback parameter to write the feedback score. For example, when a user enters the product's inner page, the feedback parameter is updated to the situational Multiarmed Bandit algorithm. Furthermore, for the score feedback added to the shopping cart and shopping cart checkout, we formulate a formula equation (8) for the score feedback based on the principle of "the higher the profit, the higher the score".

$$\text{reward} = (C' - C) \cdot q \cdot \beta. \tag{8}$$

C' is the sales amount, C is the cost of sales, q is the number of products, and β is a constant to affect the size, and because the initial array of the situational Multiarmed Bandit algorithm is 0 to 1.

In this experiment, we added a mechanism that simulates the daily update of agricultural products shopping website merchants. Generally speaking, the total product categories on the sales website have small changes, so the products recommended by the contextual Multiarmed Bandit algorithm are recommended for the products; for example, there are a total of A to Z products; we follow the user's characteristics and product characteristics entering into the algorithm, and we can obtain the products with the highest expected value and recommend them to consumers. If the consumer clicks to enter the product or it adds to the shopping cart, the algorithm will be updated.

Therefore, our method of updating the discount module is to first list the combination of the product and the discount module as a product and then only take "the same product, different discount modules" as the discount that will be used for the product that day module. Then, the sum of the expected value of each consumer feature and each discount module of each product browsed on the previous day is obtained, and then, the discount module with the highest expected value of each product is taken as the discount module combination of the product of the day.

$$p_{t,s} = \sum_{i=0}^{u} \alpha_{t,s}. \tag{9}$$

4.3. The Influence of Random Users and Products. In order to accurately test the performance, we first do 30 experiments from 5 products and go to the extreme value to get the average value to see the experimental results. Here, we use the converged value to compare the performance. Converged means that the average daily income does not increase due to the increase in the number of days, so we take the last 7 days of the 30-day experimental data as the converged performance as shown in Figures 2-3 and Table 5.

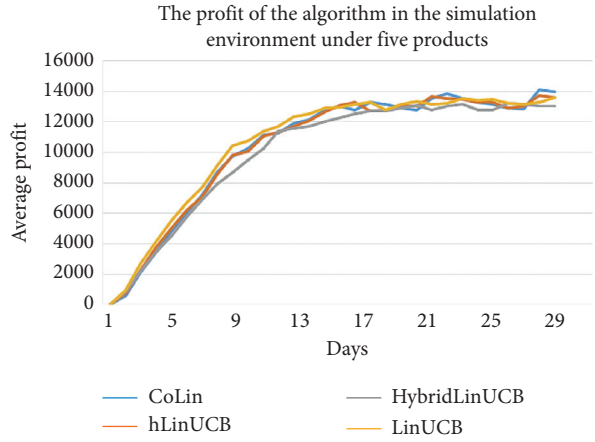

Figure 2: Algorithmic revenue performance under 5 products.

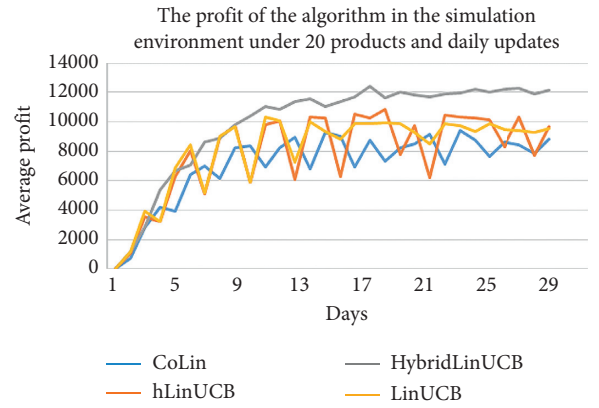

Figure 4: The performance of algorithm revenue updated daily under 20 products.

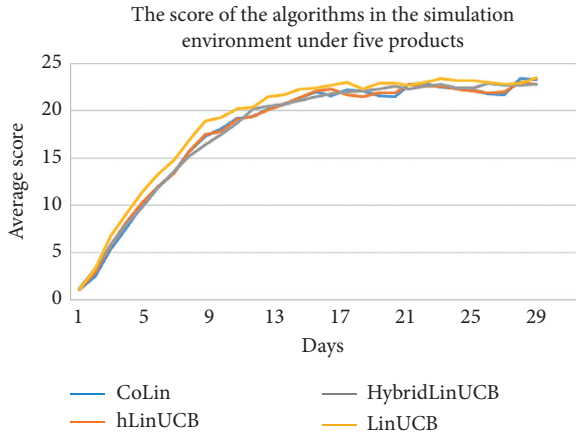

Figure 3: Algorithm score performance under 5 products.

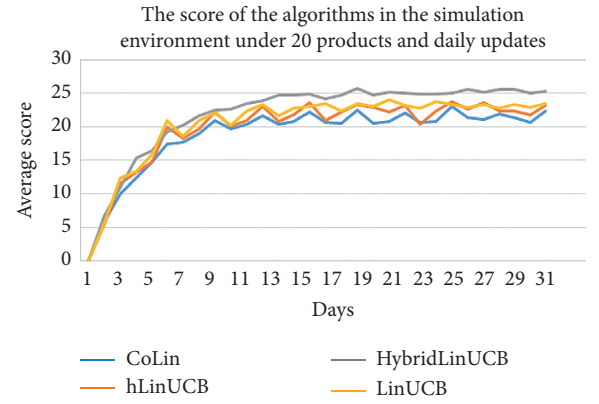

Figure 5: The performance of the algorithm scores updated daily under 20 products.

Table 5: The average performance of the algorithm revenue and scores in the last 7 days under 5 products.

Algorithm	Income	Score
LinUCB	13384.55	**23.133**
Hybrid-LinUCB	13036.29	22.6591
hLinUCB	13326.1	22.3913
CoLin	**13417.69**	22.481

Table 6: The average performance of the last 7 days of the algorithm updated daily under 20 products.

Algorithm	Income	Score
LinUCB	9509.5	23.1692
Hybrid-LinUCB	**12104.6**	**25.3376**
hLinUCB	9515.7	22.7945
CoLin	8502.6	21.6711

4.4. Comparison of Promotion Plan. In general consumer websites, promotion schemes are an important part of selling products. Because each consumer has a different degree of adaptation to promotion schemes, how to choose products and promotion schemes is also a key issue. So here we follow the daily update mechanism mentioned earlier to simulate the impact of the daily update of the promotion plan on the benefits of each algorithm as shown in Figures 4-5 and Table 6.

4.5. Comparative Advantages of Preferential Recommendation. In this phase of the experiment, we assume that male users have a higher choice of 80% for buy one get one free, 50% for other schemes, simulate male users' reaction actions when browsing schemes, and are in line with traditional A/B Testing methods. Comparing the UCB algorithm of the noncontextual Multiarmed Bandit algorithm, the following results can be obtained as shown in Figure 6.

From the above experimental data, it can be proved that using the situational Multiarmed Bandit algorithm can save the number of experiments and can automatically select the best plan. Then we continue to assume the following conditions as shown in Table 7.

Such a setting environment is mainly to test the reaction ability of the situational Multiarmed Bandit algorithm when the user's preference orientation changes. Then, the results are obtained as shown in Figure 7.

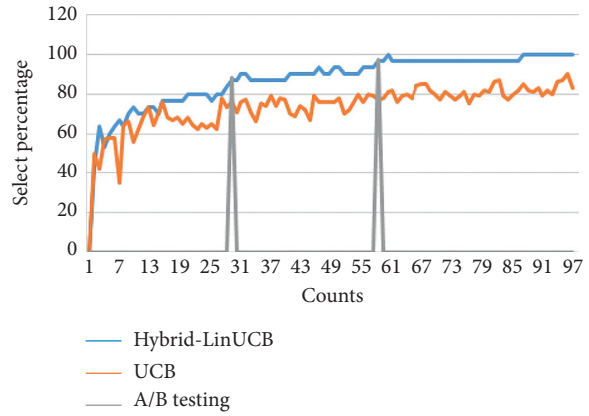

Figure 6: Hybrid-LinUCB, UCB, and A/B testing have 80% selection probability, number of attempts, and recommendation probability.

Table 7: User preferences and frequency change settings.

	Testing for 100 times ago (%)	After 100 testing (%)
Probability of buy one and get one free	80	50
Probability of free shipping	50	80

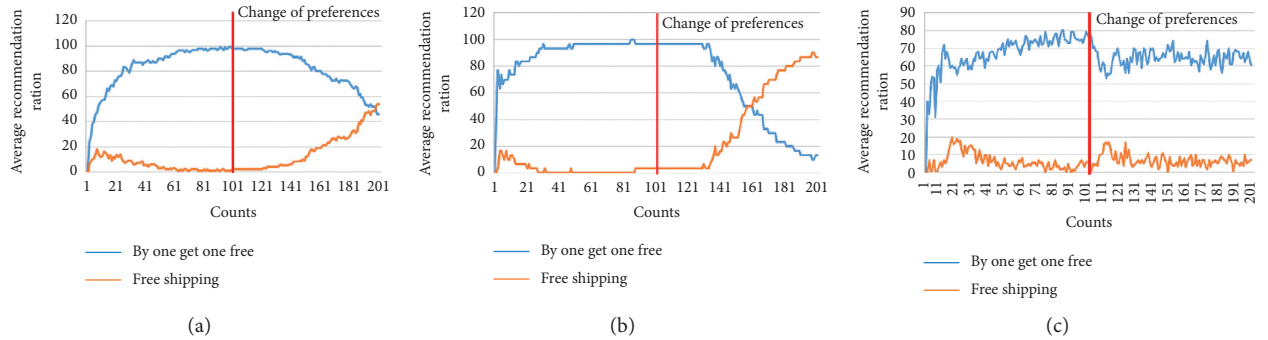

Figure 7: (a) Probability of recommendation of *A*/*B* testing changes in user preferences. (b) Recommendation probability of Hybrid-LinUCB algorithm in user preferences. (c) Probability of recommendation of UCB algorithm in changing user preferences.

5. Conclusions

In this research, we can know that the LinUCB algorithm is highly recognizable for the linear relationship between user characteristics and product characteristics and can be used in most cases. The Hybrid-LinUCB algorithm has common environmental characteristics, and it has better results for changing products and can avoid the problem of cold start of expected value. If it is for agricultural sales websites that often change products now, the Hybrid-LinUCB algorithm is the best choice. Moreover, from the experiment on contextual Multiarmed Bandit algorithm for recommending preferential modes, it shows that when users have their own preferential modules, the algorithm can predict the user's preferential module preferences through user characteristic data and compared with traditional A/B testing and non-contextual Multiarmed Bandit algorithm; it has the advantages of faster and automatic acquisition of prediction results and changes with the environment.

In terms of research limitations, it is difficult to simulate the complex factors of agricultural shopping websites with different complexity every time from different algorithms, which can effectively have significant parameters.

In future works, it is possible for establishing questionnaires from the website, collecting relevant characteristics from customer data, using exploratory factor analysis, or confirmatory factor analysis to obtain significant characteristics. Then, integrating several types of LinUCB algorithms with agricultural sales websites and browsed and consumed by real consumers can be compared in practice for improving product sales performance.

Data Availability

The raw research data are are provided in Supplementary Materials.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by grants of Taiwan's Ministry of Science and Technology: The Program for

Formulation, Maintenance and Operation of Innovative Business Models Integrating Smart Manufacturing and Information System under grant numbers MOST 109-2425-H-005-001 and MOST-109-2221-E-005-047.

Supplementary Materials

The supplementary data file includes the raw research data. (*Supplementary Materials*)

References

[1] A. G. Abishek, M. Bharathwaj, and L. Bhagyalakshmi, "Agriculture marketing using web and mobile based technologies," in *Proceedings of the 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, pp. 41–44, IEEE, Chennai, India, July 2016.

[2] R. Robina-Ramírez, A. Chamorro-Mera, and L. Moreno-Luna, "Organic and online attributes for buying and selling agricultural products in the e-marketplace in Spain," *Electronic Commerce Research and Applications*, vol. 42, Article ID 100992, 2020.

[3] T. Oliveira, M. Alhinho, P. Rita, and G. Dhillon, "Modelling and testing consumer trust dimensions in e-commerce," *Computers in Human Behavior*, vol. 71, pp. 153–164, 2017.

[4] A. Chamorro-Mera and L. Moreno-Luna, "Organic and online attributes for buying and selling agricultural products in the e-marketplace in Spain," *Electronic Commerce Research and Applications*, vol. 42, Article ID 100992, 2020.

[5] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, pp. 661–670, Raleigh, NC, USA, April 2010.

[6] C. J. C. H. Watkins, "Learning from delayed rewards," *Robotics and Autonomous Systems*, vol. 15, no. 4, 1989.

[7] D. J. Russo, B. VanRoy, A. Kazerouni, I. Osband, and Z. &Wen, "A tutorial on Thompson sampling," *Foundations and Trends in Machine Learning*, vol. 11, no. 1, pp. 1–96, 2018.

[8] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 23, pp. 235–256, 2002.

[9] Q. Y. Wang, H. Z. Gu, Q. Q. Gu, and H. N. Wang, "Contextual bandits in a collaborative environment," in *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 529–538, Pisa, Italy, November 2016.

[10] H. Wang, Q. Wu, and H. Wang, "Learning hidden features for contextual bandits," *International Conference on Information and Knowledge Management, Proceedings*, vol. 24–28, pp. 1633–1642, 2016.

[11] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[12] M. Wan, D. Wang, M. Goldman, and M. Taddy, "Modeling consumer preferences and price sensitivities from large-scale grocery shopping transaction logs," in *Proceedings of the 26th International Conference on World Wide Web*, pp. 1103–1112, Perth, Australia, April 2017.

[13] C. Ling, T. Zhang, and Y. Chen, "Customer purchase intent prediction under online multi-channel promotion: a feature-combined deep learning framework," *IEEE Access*, vol. 7, pp. 112963–112976, 2019.

[14] A. Vanderveld, A. Pandey, A. Han, and R. Parekh, "An engagement-based customer lifetime value system for e-commerce," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 293–302, New York, NY, USA, August 2016.

[15] Q. Cai, A. Filos-Ratsikas, P. Tang, and Y. Zhang, "Reinforcement mechanism design for e-commerce," in *Proceedings of the 2018 World Wide Web Conference*, pp. 1339–1348, Lyon, France, April 2018.

[16] B. Brodén, M. Hammar, B. J. Nilsson, and D. Paraschakis, "Ensemble recommendations via thompson sampling: an experimental study within e-commerce," in *Proceedings of the 23rd International Conference on Intelligent User Interfaces*, pp. 19–29, Tokyo Japan, March 2018.

[17] W. Chu and S. T. &Park, "Personalized recommendation on dynamic content using predictive bilinear models. WWW'09," in *Proceedings of the 18th International World Wide Web Conference*, pp. 691–700, New York, NY, USA, April 2009.