Winter 12-1-2010

Select Suppliers from Electronic Markets with Incomplete Information

Li-gang Chen

Select Suppliers from Electronic Markets with Incomplete Information

Chen Li-gang, School of Management, Harbin Institute of Technology, Harbin, China.
E-mail: fromsouth@163.com

Abstract

An agent want to buy products from e-market often encounters unknown suppliers, he then must choose between maximizing its expected utility according to the known suppliers and trying to learn more about the unknown suppliers, since this may improve its future rewards. This issue is known as the trade-off between exploitation and exploration. In this research, we study the problem of an agent how to select suppliers from electronic markets with incomplete information. The agent has no knowledge about suppliers, so he needs to learn the information by consuming their product and his object is to maximize total utility.

We consider two different scenarios. The first is an agent selects a single supplier at each time period. By the introduction of Gittins index, we show that by using Gittins index technology, the agent can achieve the optimal solution. The second is an agent can select several suppliers at each time period, we propose four heuristic policies and evaluate them by building up a simulation tool.

Keywords: Business Intelligence; select suppliers; incomplete information; multi-armed bandit problem

1. Introduction

Nowadays, with the development of electronic commerce such as B2B (Business to Business), B2C (Business to customer) and C2C (Customer to Customer), an agent which wants to buy products or raw materials from electronic markets may find there are hundreds of suppliers available. It on the one hand provides many options for the agent, however, on the other hand make the agent much difficult to choose suppliers since the agent has no quality knowledge of each supplier. The agent can learn quality of product after using it, therefore the agent will have prior knowledge of supplier when it has experiences with the supplier. The prior knowledge can be used by the agent to decide which suppliers to choose next round. We consider situations when the agent need to buy products repeated from electronic markets and he prefers product with high quality and the lowest price. Product with high quality and lowest price ensures the high utility of the agent. Since in our context, agent is the representative of buyer, thus we abuse these two words hereafter without explicit explanation.

The issue of uncertainty quality can also happen in traditional market, however, the situation is much more prevail in electronic market since buyers can't see and touch the real product in advance. The information which can help buyers to make decision is pictures and description posted by suppliers. However, since surveillance service is still not good in electronic market, supplier with low quality might provide better pictures and description than supplier with high quality, thus agent could be misled and make wrong choice. Other factors such as larger number of suppliers, unstable markets, delivery time and so on make the uncertainty problem much more pronounced in electronic markets than traditional markets [1, 2]. Consider an agent want to buy fish oil online, he shall use fish oil as keyword and search it on Google, as a result, Google will return a long supplier list with different quality (condition, delivery time and etc). Thus the agent must decide which supplier to choose so as to maximize his utility.

One may argue that buyer can choose supplier with the highest price, since in common sense, high price represents high quality. However, in electronic market, price can't use to identify quality, since unknown supplier may exist, which provides product with high quality, but charge low price. Of course, there may suppliers sell product with low quality and high price. Consider, continue the above example, suppose dailyVita is the largest online supplier of fish oil, but there are other online suppliers sell the same fish oil as dailyVita with discounted price and may also provide better quality product, if quality can be measured as delivery time or customer service. This may due to other suppliers make deal with better delivery company.

In situations where there are a lot of unknown suppliers and unknown suppliers may sell product with high quality and low price, buyers need to employ intelligence policies in order to choose the most profitable deals. We assume that past information can be used as a gauge to measure the quality of supplier. For example, if a buyer makes several deals with supplier A and observes that the mean quality of these deals is negative, thus he might not choose supplier A again due to the unhappy experience. However, on the other hand, if the history experience with supplier A is positive, buyer might choose supplier A again with high probability. Thus the dilemma of the buyer is whether to choose best known suppliers or try unknown suppliers so as to learn the quality of unknown supplier which may improve the future benefit of the buyer. This dilemma is also known as

the trade-off between exploitation and exploration. The most pronounced problem of this dilemma is multi-armed bandit problem (MAB) [3, 4]. MAB problems are a class of sequential resource allocation problems concerned with allocating on or more resources among several alternative projects. Such problems are paradigms of making decisions that yield high current rewards, versus making decisions that sacrifice current rewards with the prospect of better future rewards. It can be described as follows: suppose there is a machine with several slots, each slot can generate a certain amount of reward. Rewards are drawn from a certain statistic distribution. A gambler can pull a single slot at each round and receive the reward generated by the slot. Since the gambler has no knowledge such as mean value and variance of each slot, he needs to learn the parameters by pulling them. The problem is what sequence of the slots gambler should pull so as to maximize his total reward. One can easily find that the problem faced by buyer is similar with the gambler if suppliers can be treated as slots.

Rina and Sarit[5] consider the same problem context as described above and apply Gittins Index technology to choose supplier in the condition of incomplete information. They show that how to select supplier using Gittins Index in different settings: consider the probability of the agent to buy at each time period and different sizes of purchases. However in their settings, they don't consider the risk attitude of buyers and the size of suppliers. They assume that at each time period, the agent can only choose a single supplier. Though the assumption is reasonable and enables them to employ Gittins Index, the true situation is that buyer can choose any supplier at each round. Select multiple suppliers at each time period can be modeled as multi-armed bandit problem with multiple plays (MABMP), but the solution of MABMP remains unsolved. In this paper, we extend the work of Rina and Sarit. We consider the risk attitude of buyer as well as the size of supplier buyer can choose.

The paper is organized as follows. In section 2, we discuss related work and, in Section3 we present the formal model. We extend the model and discuss how to consider the risk attitude of buyer in Section 4. Section 5 presents some heuristic algorithm for multiple-supplier choosing problem. Finally, in Section 6, we provide conclusions and suggestions for future extensions.

2. Related Work

The issue of making decision with incomplete information is widely studied. Eric[6] investigates dynamic pricing strategies for maximizing revenue in an internet retail channel by actively learning customers' demand response to price. Wang[7] formulate the optimal pricing problem with a bandit model and characterize the solution by means of stochastic dynamic programming. Cathy and Parijat[8] also study the dynamic pricing issue in e-Services, instead of using standard bandit process they combines annealing algorithm with Bayesian learning to balance the trade-off of exploitation and exploration. Rina and Sarit[5] use Gittins Index technology to select supplier in an environment of incomplete information. Salganicoff and Ungar[9] use Gittins index to select actions which optimally trade-off exploration and exploitation, they combine Gittins indices with decision trees to develop a mapping from state and action to success or failure of that action. Pandey and Olston[10] consider how a search engine should select advertisements to display with search results, they model advertisement placement as a multi-armed bandit problem, their algorithms are based on upper confidence bound[11] algorithm which is mainly to solve non-stochastic multi-armed bandit problem.

The traditional supplier choosing problem is mainly about how to select a supplier among several criteria: quality, delivery reliability, product performance and unit price [12, 13]. Unlike in an uncertainty setting, all criteria are known in advance. However, as we show above, in the environment of e-business, most of the criteria are unknown which make the traditional supplier selection tools ineffective. That's why we need to discuss the problem with incomplete information. Generally, products are classified into three different types: products with a quality ascertained by buyer before a purchase; products with a quality that is learned after the products have been bought; products have a quality that can hardly be learned even after consumption. In this paper, we assume that products' quality can be inferred after consumption.

Since an agents in our context need to decide between choosing best known suppliers or trying other suppliers in order to learn the quality of their products. The technologies for exploitation and exploration trade-off are known as reinforcement learning. The most used technologies are: the dynamic programming approach [14] which is expensive in time and space; the heuristic forward iteration algorithms [15]; the Gittins allocation index [3, 4], also known as dynamic allocation index, which can be used in cases of reward. Gittins index technology is proven to yield the highest expected utility when single supplier is available per round, but when in the background of multi-suppliers per time, Gittins index in not necessary optimal. Thus, we prefer Gittins index when the single supplier situation is discussed, however, we will employ heuristic forward iteration algorithms when the multiple suppliers situation is presented.

3. Formal model with single Supplier

Consider an agent facing a market of N suppliers who sell a given item, the quality of the item sold by supplier i is u_i, the standard deviation is σ_i and its price is p_i, the agent need to buy the item regularly at discrete time period. At each time period, the agent must choose one supplier among the N suppliers, he has no knowledge about the mean quality and deviation about each supplier, but he knows the price of each supplier, besides, the agent maintains a history of n_i length of previous interactions with supplier i. Then the average quality and standard deviation of supplier i is $\overline{x}_i$ and $\hat{s}_i$ respectively. The utility of the agent buying an item from supplier i with quality x_i and price p_i is $x_i - p_i$.

3.1 Two supplier scenario and Gittins Index

We first introduce the simplest situation in which the agent only facing two suppliers S (safe) and R (random). The agent has full information about safe supplier and its quality X_S, while alternative supplier R is unknown to the agent and its quality X_R whose distribution law f on $\mathbb{R}$ also unknown to the agent. We assume that the agent has a prior on f. In a dynamic environment, the agent may learn about the distribution law f of the random quality X_R by buying product from R and observing realizations $X_R(t)$ of X_R at time period t and update his prior on f after t. The agent selects a $Y(t) \in \{S, R\}$ at each time period t. Then the total utility of the agent can be formulated in a stochastic discounted inter-temporal matter:

$$U(Y(\cdot)) = \sum_{0}^{+\infty} \rho^t H(\Phi(Y(t), X(t+1))) \qquad (1)$$

Where $\rho \in [0,1]$ is the discount rate, $\Phi(S, X(t+1)) = X_S$, $\Phi(R, X(t+1)) = X_R(t)$. $H(\cdot)$ is the utility function. Since we assume that agent's utility can be expressed by the subtraction of quality and price, thus when supplier R is selected, $H(\cdot)$ is $X_R(t) - p_R$.

The above problem is a standard approach of classical bandit problem. It's one armed bandit problem because the state of arm S is stationary. The state of the other arm R $\hat{f}_t$ is a Markov process whose transition subject to the Bayesian updating with respect to the observation of $X_R(t)$. Gittins index is an optimal strategy for this bandit problem. At a certain stage, the agent need to compute index of each arm and the optimal strategy is to select the arm with the higher index. The state of the selected arm evolves according to a given rule, but the states of other arms remain unchanged. Gittins index is defined as follows:

$$GI(v) = \sup_{\tau>0} \frac{E[\sum_{t=0}^{\tau-1} \rho^t \Phi(v_t) \mid v_0 = v]}{E[\sum_{t=0}^{\tau-1} \rho^t \mid v_0 = v]} \qquad (2)$$

While v_0 is the initial state of a given arm, $\Phi(\cdot)$ is reward function of state v_t. The index is the maximal average reward over a stopping time of the arm. The index value also makes the option of continuation or retirement no differences. As in our context, since the index of supplier S is constant, then given $GI(S)$, $GI(R,v)$ and the state $v_t = \hat{f}_t$ at stage t, the optimal strategy is to select the supplier with the higher index: supplier S if $GI(S) > GI(R, \hat{f}_t)$ and supplier R else.

The calculation of Gittins index by definition (2) is difficult. Fortunately, there is a convenient approach to compute Gittins index. Let $g(\overline{x}, \hat{s}, n)$ denotes the index value of an arm with a history of n length, average value of $\overline{x}$ and standard deviation $\hat{s}$, Gittins proved that [16]

$$g(\overline{x}, \hat{s}, n) = \overline{x} + sg(0,1,n) \qquad (3)$$

In (3), $g(0,1,n)$ is the standard Gittins index with mean value 0, standard deviation 1, and history length n. Gittins calculated $g(0,1,n)$ given different combinations of discount rate and n. This show that multiplying the standard index value by the deviation of the arm's reward and adding the average reward of the arm forms the index value of the arm. It can be observed that as an arm's average rewards increases, its index value increases too. The standard deviation and the history length also play important roles in the index calculation. Since the standard Gittins index is only significant when n is small (n <10), when n is getting larger $g(0,1,n)$ becoming small drastically. Thus $\hat{s}g(0,1,n)$ shows that the contribution of the standard deviation of an arm to the index value decreases greatly when its history length increases. It can be explained as follows: when the experience to an arm is low (history length is small) it is better to select arm with highly risky (higher standard deviation), because the risky arm might generate high rewards in the future. However, as the experience of an arm is long enough, then the average rewards will take dominance in the calculation of the index value. The theory behind it is that the law of larger numbers ensures that when history length is long enough the average rewards is extremely close to its mean rewards.

Proposition 1. Consider two agents with common prior belief f_0, and one agent is more risk averse than the other. If at the beginning, the more risk-averse agent selects supplier R based on f_0, then the less risk-averse agent will select R too.

And as long as the more risk-averse agent selects supplier R, so does the less risk-averse agent.

The proposition is proved by Chancelier[17], it implies that the agents can be ranked by their degree of risk aversion. Based on proposition 1, we have direct corollary 1 and corollary 2.

Corollary 1. If the less risk-averse agent selects supplier S based on f_0, so does the more risk-averse agent, and as long as the less risk-averse agent selects supplier S, so does the more risk-averse agent.

Corollary 2. The number of times agent select supplier R is a decreasing function of the degree of its risk-aversion.

Proposition 2. An agent selects the safe supplier if and only if the index value of random supplier R less than the index value of safe supplier S, and once an agent selects the safe supplier, he would stick to the safe supplier forever.

Proof. The first assertion is the major result of optimal strategies for bandit problems. That is choose the higher index at every stage. If the agent selects the safe supplier, thus the state of the random supplier remains fixed. Then the index value of supplier S will always larger than supplier R at the future stages, therefore, the agent shall select safe supplier all the time.

3.2 Multiple Suppliers Scenario

Problem Statement: an agent wants to buy an item from N suppliers repeatedly, but he has no knowledge about the quality of each suppliers. The agent's object is to maximize his utility.

Proposition 3. Given the price p_i of supplier i and remains as a constant, then the calculation of Gittins index is as follows:

$$g(\bar{x}_i, \hat{s}_i, n_i) = \bar{x}_i + s_i g(0,1,n_i) - p_i \qquad (4)$$

Proof. Since the price p_i is a constant over time, it will not influence $\hat{s}_i$ on the whole. As we assume above, the utility of the agent can be expressed as the subtraction of quality and price, then the price p_i just decreases the agent's utility from select supplier i. Actually, the Gittins index can be divided into two parts: exploitation and exploration where $\bar{x}_i$ is the part of exploitation and $\hat{s}_i g(0,1,n_i)$ is the part of exploration. The constant p_i doesn't influence the exploration part, let the exploitation part be $\bar{x}_i$ - p_i, then the structure of Gittins index is maintained. ■

The proposition makes sure that the agent only pays more for higher quality. This does accord with economic sense.

Based on the proposition 3, the optimal strategy for the agent is straight. The steps are as follows:

Step 1: at every stage t, compute Gittins index of every supplier $g(\bar{x}_i, \hat{s}_i, n_i)$ according to (4).

Step 2: select supplier j with the largest Gittins index value, that is

$$j = \arg\max_i g(\bar{x}_i, \hat{s}_i, n_i)$$

Step 3: choose supplier j and, buy product from j, observe the quality of the product and update parameters as follows:

(1) $\bar{x}_j = \dfrac{n_j \bar{x}_j + x_j(n_j+1)}{n_j+1}$

(2) $\hat{s}_j = \sqrt{\dfrac{1}{n_j}\sum_{k=1}^{n_j+1}\left(x_j(k) - \bar{x}_j\right)^2}$

(3) $n_j = n_j + 1$

Step 4: move on to the next time period and repeat step 1 to step 3.

There is an important issue to use the Gittins index in the above strategy. The standard Gittins index $g(0,1,n_i)$ requires a history length n_i at least larger than 2, i.e. the agent must buy product from each supplier at least two times first. If there are too many suppliers in the market, then the agent must spend a lot of opportunities in order to meet the requirement of 2 history length. In this situation, the agent faces double bandit processes. At each time period, the agent needs to select supplier from the suppliers with history length larger than 2 or select new supplier with no history. In order to cope with the new supplier issue, the agent must balance the trade-off select old suppliers or new suppliers. Though there is not optimal strategy for the new supplier issue, the agent can employ a heuristic to solve it: at the initial stages, when all suppliers are new, the agent needs to explore new suppliers aggressively, as the agent has a lot of old suppliers, he might want to look at the old suppliers more often, thus he can employ a greedy algorithm to decide which group of supplier to select. The process is: set up a probability w, with probability w choose supplier from old supplier group and with probability 1- w from new group. When the group of old supplier is big enough, then the agent needs to gradually increase w so as to make sure the agent can take advantage of exploitation sufficiently for the sake of utility maximization.

4 Choosing K-Suppliers Problem

Section 3 discusses the issue of how to select a single supplier per stage and provide optimal strategies by using Gittins index technology. Choose a single supplier per stage is suitable for small businesses, but as for big corporation, for the sake of diminish risk, they might want to buy product from different supplier, or at each time period, the agent want to select k-suppliers from all the supplier pool.

Problem statement: an agent wants to buy items from N suppliers repeatedly, at every stage, he

shall select k-suppliers, but he has no knowledge about the quality of each suppliers. The agent's object is to maximize his utility.

The problem of selecting k-suppliers can be modeled as multi-armed bandit problem with multiple plays, i.e. the gambler can pull multiple slots at each time period. Though there are a lot concerns about the MABMP problem, the optimal solution is still not ready. In this section, we provide several heuristic algorithms for it and by constructing a simulation platform we evaluate these algorithms in different settings.

The first algorithm is based on Gittins index. Since Gittins index is optimal for single supplier scenario, we believe the index value is a good heuristic information for each supplier. The process is as follows:

Step 1. Select each supplier two times by any order;

Step 2. Compute Gittins index for every supplier according to equation (4);

Step 3. Select k suppliers with the k-highest Gittins index;

Step 4. Buy product from these k suppliers, observe their quality and update parameters as follows:

For j=1 to k

(1) $\overline{x}_j = \dfrac{n_j \overline{x}_j + x_j(n_j+1)}{n_j+1}$

(2) $\hat{s}_j = \sqrt{\dfrac{1}{n_j}\sum_{k=1}^{n_j+1}\left(x_j(k)-\overline{x}_j\right)^2}$

(3) $n_j = n_j + 1$

End for

Step5. move on to the next time period and repeat step 3 and 4.

The second heuristic algorithm is straightforward. We call it uniform play and empirical best (UPEB). The process is as follows:

Step 1. Fix up a positive integer M ;

Step 2. Select each supplier M times at any order;

Step 3. Compute the utility of each supplier

Step 4. Always select the k-highest utility suppliers for the future time period.

The third heuristic algorithm is interval estimation strategy (IES). IES strategy choose suppliers by estimate their upper quality bound and choose the k-highest upper quality bound suppliers at every time period. The process is as follows:

Step 1. Set up a quantile a ;

Step 2. Select each supplier two times at any order.

Step 3. For $i = 1$ to N

Compte the upper bound of each supplier as (5)

$$upper_{i,\alpha} = \overline{x}_i + \hat{s}_i \frac{student_{\alpha/2}(1)}{\sqrt{2}} - p_i \qquad (5)$$

End for ($student_{\alpha/2}(1)$ denotes the student distribution function with freedom 1 and quantile $a/2$)

Step 4. Select the k-highest $upper_{i,\alpha}$ suppliers

Step 5. Buy product from the suppliers at step 4, observe their quality and update related parameters:

For j=1 to k

$$n_j = n_j + 1$$

$$\overline{x}_j = \frac{n_j - 1}{n_j}\overline{x}_j + \frac{1}{n_j}x_j(n_j)$$

$$\hat{s}_j = \sqrt{\frac{1}{n_j - 1}\left(\sum_{i=1}^{n_j}\left(x_j(i)-\overline{x}_j\right)^2\right)}$$

$$upper_{j,\alpha} = \overline{x}_j + \hat{\sigma}_j \frac{student_{\alpha/2}(n_j-1)}{\sqrt{n_j}} - p_j$$

End for

Step 6. Repeat step 4 and step 5.

The reason we use student distribution to compute the upper bound is that, when we have no knowledge about the mean value and variance, the normal distribution is the most common used distribution to represent it.

The last heuristic policy is called stepwise T-checked policy (STP), this policy also uses the student distribution, and the process is as follows:

Step 1. Set up a quantile a , a positive integer B

Step 2. Select all suppliers two times at any order;

Step 3. Compute the utility of every supplier based on the two samples.

Step 4. Select the k-highest utility suppliers

For the B lowest utility suppliers in the k-highest suppliers i

For other suppliers not in the k-highest suppliers j

Let stat= $\dfrac{\overline{x}_j - \overline{x}_i}{\hat{s}_j}\sqrt{n_j}$

If stat< $-student_{\alpha}(n_j - 1)$

Then substitute supplier i with supplier j

End for

End for

Step 5. Buy product from the revised k-suppliers, and then update their mean quality and standard deviation.

Step 6. Repeat step 4 and step 5.

In order to test the performance of these four heuristic strategies, we build up a simulation platform with different settings. In our simulation, we defined several agents which behave

according to the strategies above. The quality of the products produced by each supplier is derived from a normal distribution, but the details of this distribution are unknown to the agents. The mean quality and price of each supplier is drawn randomly from the interval [20,140] in all runs, and the standard deviation is drawn randomly from the interval [40,100].

In our first simulation, the size of suppliers is 40, and the agent can choose 10 suppliers simultaneously at every time period. We let the horizon be 200, and run the simulation 10000 times. The result is presented in table 1.

Table 1 simulation results with 40 suppliers and 10 available per period with 200 horizons

Agent	UPEB 1	UPEB 2	UPEB 3	UPEB 4	UPEB5
Total utility	77201	77542	75626	67801	58352
Agent	GIH	IEP	STP1	STP2	STP3
Total utility	90347	91467	91291	91552	91422

UPEB1, UPEB2, UPEB3, UPEB4 and UPEB5 are the agents of UPEB with the value of M 5,8,10 and15 respectively. STP1, STP2 and STP3 are the agents of STP policy with the value of B 2, 3 and 4 respectively. GIH is the policy based on Gittins index. The result shows that all UPEB-based policies are strictly worse than other policies. Among the five UPEB policies, UPEB2 whose sample time is suitable achieve the best utility. Since small sample times lead to under-exploration, large sample times lead to over-exploration. Indeed, Gittins index is good heuristic information, though it doesn't achieve the best value. The outputs of IEP and STP are very close, and STP2 is the best strategy over all. The reason is similar with UPEB.

We extend the total horizon to 300 periods. We find that the STP2 is the best strategy also. The total utility of all agents is double than 200 periods, that's because during the first 200 periods, the agents spend a lot of periods on exploration.

Let the size of supplier is 60 and other settings are the same as the first simulation, the result is showed in table 2.

Table 2 simulation results with 60 suppliers and 10 available per period with 200 horizons

Agent	UPEB1	UPEB2	UPEB3	UPEB4	UPEB5
Total utility	96137	92145	86898	71328	53122
Agent	GIH	IEP	STP1	STP2	STP3
Total utility	110079	113984	111271	111306	110890

The results are similar with table1, but the best strategy is IEP. The reason might be as the size of supplier increases, the upper bound policy can find more high quality suppliers with low price.

The three simulations show that the agent should employ IEP or STP strategy since they can yield the best utility. When the size of supplier is relatively small, the agent should favor STP strategy, and when the size of supplier is relatively large, the agent should favor IEP strategy.

5. Conclusion

In this research, we discuss the issue of select suppliers from e-market with incomplete information. We show that the problem can be modeled by the multi-armed bandit problem. We first consider the simplest scenario where there are only two suppliers in the market and the agent has full information of safe supplier, then we show how to solve it by the technology of Gittins index and the impact of risk-aversion. Besides, we show that Gittins index is also optimal when there are several suppliers.

We also consider the situation when an agent can choose multiple suppliers simultaneously at each time period. We provide four heuristic policies to solve the trade-off of exploitation and exploration issue of multiple suppliers' selection problem. In order to test the performance of different policy, we develop a simulation tool and shows that the agent should favor STP policy when the size of supplier is relatively small and IEP policy otherwise.

References

[1] Westland J C. Transaction risk in electronic commerce[J]. Decision Support System. 2002, 1(33): 87-103.

[2] Strader T J, Shaw M J. Characteristics of electronic markets[J]. Decision Support System. 1997, 3(21): 185-198.

[3] Whittle P. Multi-Armed Bandits and the Gittins Index[J]. Journal of the Royal Statistical Society. Series B (Methodological). 1980, 42(2): 143-149.

[4] Gittins J C. Bandit Processes and Dynamic Allocation Indices[J]. Journal of the Royal Statistical Society. 1979, 41(2): 148-177.

[5]. Azoulay-Schwartz R, Kraus S, Wilkenfeld J. Exploitation vs. exploration: Choosing a supplier in an environment of incomplete information[J]. Decision Support Systems. 2004, 38(1): 1-18.

[6]. Cope E. Bayesian Strategies for Dynamic Pricing in E-Commerce[J]. Naval Research Logistics. 2007, 54(3): 255-281.

[7]. Wang X. Dynamic Pricing with a Poisson Bandit Model[J]. Sequential Analysis. 2007, 26(4): 355-365.
[8]. Xia C H, Dube P. Dynamic Pricing in e-Services under Demand Uncertainty[J]. Production and Operations Management. 2007, 16(6): 701-712.
[9]. Salgonicoff M, Ungar L H. Active exploration and learning in real-valued spaces using multi-armed bandit allocation indices[C]. Proceedings of the 12th International Conference on Machine Learning. San Francisco, CA: 1995: 480-487.
[10]. Pandey S, Olston C. Handling advertisements of unknown quality in search advertising[J]. Advances in Neural Information Processing Systems. 2007, 19: 1065-1072.
[11]. Auer P, Cesa-Bianchi N, Fischer P. Finite-time analysis of the multiarmed bandit problem[J]. Machine Learning. 2002, 47(23): 235-256.
[12]. Tracey M, Tan C L. Empirical analysis of supplier selection and involvement, customer satisfaction, and firm performance[J]. Supply Chain Management: An International Journal. 2001, 6(4): 174-188.
[13]. Verma R, Pullman M E. An analysis of the supplier selection process[J]. Omega. 1998, 26(6): 739-750.
[14]. Berry D A, Fristedt B. Bandit Problems: Sequential Allocation of Experiments[M]. London, UK: Chapman and Hall, 1985.
[15]. Sutton R S, Barto A G. Reinforcement Learning:An Introduction[M]. Massachusetts, US: The MIT Press, 2005.
[16]. Gittins J C. Multiarmed Bandit Allocation Indices[M]. New York: Wiley, 1989.
[17]. de Palma A, Picard N. Route Choice Behaviour with Risk-Averse Users[J]. Spatial Dynamics Networks and Modelling. 2006: 139-179.