A Markov Decision Process Approach to Analyze Discount & Reputation Trade-offs in E-commerce Systems

Hong Xie, John C.S. Lui
Department of Computer Science & Engineering
The Chinese University of Hong Kong
{hxie,cslui}@cse.cuhk.edu.hk

1. INTRODUCTION

E-commerce systems, e.g., Amazon, eBay and Taobao, are becoming increasingly popular. We consider eBay (or Taobao) like E-commerce systems, where a large number of sellers and buyers transact online. To reflect the trustworthiness of sellers, a reputation system is maintained. In particular, the feedback-based reputation system [4] is the most widely deployed. Sellers of such systems are initialized with a low reputation and they must obtain a sufficiently large number of positive feedbacks from buyers to earn a reputable label. For example, eBay and Taobao use three-level feedbacks, i.e., $\{-1$ (Negative), 0 (Neutral), 1 (Positive)$\}$. Each seller is initialized with a reputation score of zero. A positive (or negative) rating increases (or decreases) the reputation score by one, while a neutral rating does not change the reputation score.

Often, buyers are less willing to buy products from low reputation sellers. It was found that in eBay new sellers need to spend at least seven hundred days (on average) to earn a reputable label [5]. Some sellers resort to "*illegal means*" to increase their reputation, i.e., authors in [6] found that more than eleven thousand sellers in Tabobao have conducted fake transactions. A number of companies, e.g., Lantian, Shuake and Kusha, even provide professional fake transaction services and the per-year fake transaction volume is estimated to be more than six million per company [6]. Fake transactions are illegal, and this motivates us to explore "*legitimate means*" to enhance (new) sellers' reputation.

We propose to enhance sellers' reputation via "*price discounts*". To illustrate, consider the eBay reputation system and that a seller is reputable if and only if her reputation score is no less than 500. A seller can attract 10 transactions per day if she is reputable, otherwise, she can only attract 1 transaction per day. Assume each transaction earns a positive rating of 1. Suppose the price of a product is \$1 and its cost is \$0.8. We have the following cases.

- **No discounts:** *For a new seller (initialized with a reputation score of zero) who does not provide any discount, she needs to spend 500 days to earn a reputable label. The total profit in the first 500 days is* $(1-0.8)\times 1\times 500 = 100$.
- **With discounts:** *A new seller provides a discount of 40% before she earns a reputable label, i.e., the price becomes 0.6, and she does not provide any discount after becoming reputable. Assume this discount increases the transaction volume to 2 per day. She needs to spend* 250 *days to earn a reputable label. The profit in the first 250 days is* $(0.6-0.8)\times 2\times 250 = -100$. *The total profit in the first 500 days is* $(0.6-0.8)\times 2\times 250+(1-0.8)\times 10\times 250 = 400$.

The above cases illustrate: (1) Price discounts can enhance sellers' reputation; (2) Price discounts may lead to profit losses in the short run, but the reputation effect can compensate the profit in subsequent days. Note that in real-world E-commerce systems, the demands (i.e., per-day transaction volumes) are *dynamic*, and buyers may provide *biased* ratings, etc. This paper studies the discount selection problem in such general settings. We develop a mathematical model to capture important factors of an E-commerce system. We formulate a profit maximization framework via a semi-Markov decision process (SMDP) to explore the optimal trade-offs in selecting price discounts. We theoretically characterize the optimal profit and discount.

MAMA 2018 Irvine, California, USA

2. SYSTEM MODEL

2.1 Baseline E-commerce System Model

Consider an E-commerce system like eBay and Taobao where buyers purchase products from online stores operated by sellers, and a feedback-based reputation system is maintained to reflect the trustworthiness of sellers. Sellers set the selling price and advertise the quality of products in their online stores, and finally ship the ordered products to buyers. Let $q \in \mathbb{R}_+$ and $c \in \mathbb{R}_+$ denote the price and overall cost of a product respectively. The overall cost c captures the manufacturing cost, shipment fee, etc. We define the *unit profit* to the seller $u \in \mathbb{R}$ as the price minus the cost, i.e., $u \triangleq q - c$.

To reflect the trustworthiness of sellers, a feedback-based reputation system tags each seller with a reputation score $s \in \mathcal{S}$, where $\mathcal{S} \triangleq \{-\hat{S}, \ldots, -1, 0, 1, \ldots, S\}$ and $\hat{S}, S \in \mathbb{N} \cup \{\infty\}$, The reputation score is accessible by all buyers. For example, eBay and Taobao uses $\hat{S} = 0, S = \infty$, in other words $\mathcal{S} = \{0, 1, \ldots, \infty\}$. The higher the reputation score, the more reputable the seller is. When a seller joins an E-commerce system, the reputation system initializes her reputation score as $s = 0$, i.e., a low reputation. Buyers provide feedback ratings to reflect their evaluation on the overall transaction quality (i.e., product quality, trustworthiness of the seller, etc). Each feedback rating is drawn from a discrete rating metric set $\mathcal{M} \triangleq \{-\hat{M}, \ldots, -1, 0, 1, \ldots, M\}$, where $\hat{M}, M \in \mathbb{N}$. For example, $\mathcal{M} = \{-1(\text{Negative}), 0(\text{Neutral}), 1(\text{Positive})\}$ is deployed in eBay. The higher the rating, the more satisfied the buyer is toward that seller. Consider a seller has a reputation score s, her reputation score becomes $s+m$ once she receives a feedback rating $m \in \mathcal{M}$. For example, in eBay $\mathcal{M} = \{-1, 0, 1\}$, a rating of 1 (or -1) increases (or decreases) the reputation score by 1, while a rating of 0 does not change the reputation score. We aim to enhance honest sellers' reputation via price discounts. Thus we focus on sellers who advertise product quality honestly.

2.2 Price Discount Model

To speed up the reputation accumulating process, a seller can set a price discount $a \in \mathcal{A} \triangleq [0,1]$. Precisely, a denotes the discount rate, and the product price under discount a is $q \times (1-a)$. For example, $a = 0.2$ means 20% off and the corresponding price is $0.8q$. Also $a = 0$ captures that a seller does *not* provide any discount. Let $\tilde{u}(a)$ denote the *unit profit* under discount a. Then we have $\tilde{u}(a) \triangleq u - aq, \forall a \in \mathcal{A}$.

Modeling rating behavior under discounts. Human factors like personal preferences or even biases need to be included in our model. Some buyers may provide higher ratings while other may provide lower ratings. Let $R(s,a) \in \mathcal{M}$ denote a rating provided by buyers to the seller who has a reputation score $s \in \mathcal{S}$ and sets a discount $a \in \mathcal{A}$. The rating $R(s,a)$ is a random variable, and we define its cumulative distribution function (CDF) as

$$F_R(m|s,a) \triangleq \mathbb{P}\left[R(s,a) \le m\right], \quad \forall m \in \mathcal{M}, s \in \mathcal{S}, a \in \mathcal{A}.$$

For example, consider $\mathcal{M} = \{-1, 0, 1\}$ and $\mathcal{S} = \{0, 1, \ldots, \infty\}$. Then, one example of $F_R(m|s,a)$ is

$$\begin{cases} F_R(-1|s,a) = [0.1/(1+s)]^{1+a}, \\ F_R(0|s,a) = [0.3/(1+s)]^{1+a}, \qquad F_R(1|s,a) = 1. \end{cases} \tag{1}$$

Definition 1. *Given two random variables X, Y with the same sample space Ω. We say X is larger than Y (written as $X \succeq Y$), iff $\mathbb{P}[X > x] \ge \mathbb{P}[Y > x]$ holds for all $x \in \Omega$.*

Assumption 1. *Given $a \in \mathcal{A}$, $R(s,a) \succeq R(j,a)$ holds for all $s > j$, where $s, j \in \mathcal{S}$. Given $s \in \mathcal{S}$, $R(s,a) \succeq R(s,b)$ holds for all $a > b$, where $a, b \in \mathcal{A}$.*

Assumption 1 captures: (1) The herding effects [1] that buyers give higher ratings to more reputable sellers; (2) The price effect that buyers tend to become more lenient in providing ratings under larger discounts. Equation (1) satisfies Assumption 1.

Modeling demand under discounts. We consider a dynamic demand from buyers and use the transaction's arrival process to model the demand. We define the transaction's arrival process through the inter-arrival time (or waiting time) of transactions. Precisely, let $W(s,a) \in \mathbb{R}_+$ denote the inter-arrival time of transactions to the seller who has a reputation score $s \in \mathcal{S}$ and sets a discount $a \in \mathcal{A}$. For example, $W(0,0)$ measures the amount of time a seller must wait until the next transaction arrives when she has a reputation score $s = 0$ and does not provide any discount. The inter-arrival time $W(s,a)$ is a random variable and we denote its CDF as

$$F_W(w|s,a) \triangleq \mathbb{P}[W(s,a) \le w], \quad \forall w \in \mathbb{R}_+, s \in \mathcal{S}, a \in \mathcal{A}.$$

One example of $F_W(w|s,a)$ is

$$F_W(w|s,a) = 1 - e^{-\lambda(s,a)w}, \tag{2}$$

which means that $W(s,a)$ follows an exponential distribution with a parameter $\lambda(s,a) \in \mathbb{R}_+$. This also models the Poisson arrival of transactions. One example of $\lambda(s,a)$ is

$$\lambda(s,a) = (1+\sqrt{a})/(1+e^{-s}), \qquad \forall s \in \mathcal{S}, a \in \mathcal{A}. \tag{3}$$

Assumption 2. *Given $a \in \mathcal{A}$, $W(j,a) \succeq W(s,a)$ holds for all $s > j$, where $s, j \in \mathcal{S}$. Given $s \in \mathcal{S}$, $W(s,b) \succeq W(s,a)$ holds for all $a > b$, where $a, b \in \mathcal{A}$.*

Assumption 2 captures: (1) The reputation effect that buyers are more willing to transact with reputable sellers; (2) The price effect that buyers are more willing to buy a product under a larger discount. Consider Eq. (2), Assumption 2 means that $\lambda(s,a)$ increases in both s and a. One can easily check that one example of such $\lambda(s,a)$ is derived in Eq. (3).

Assumption 3. *There exists two constants $\epsilon > 0, \delta > 0$ such that $F_W(\delta|s,a) \le 1 - \epsilon$ for all $s \in \mathcal{S}, a \in \mathcal{A}$.*

Assumption 3 states that it is impossible that an infinite number of transactions arrive to an online store within a finite time. Consider Eq. (2), Assumption 3 means that $\lambda(s,a)$ is bounded, e.g., the $\lambda(s,a)$ derived in Eq. (3).

Modeling discount update. Recall that the reputation score influences the demand. We therefore focus on the scenario that a seller updates the discount only after a new transaction arrives, i.e., his reputation score is updated. Under this scenario, we next introduce the formal discount selection models for sellers.

2.3 The Seller's Decision Model

The seller needs to select a discount for each transaction. Thus the decision space for the seller is the discount set $\mathcal{A}$. We consider the full information scenario that $F_R(m|s,a)$ and $F_W(w|s,a)$ are given. We formulate a profit maximization framework via an SMDP to characterize the optimal trade-offs in determining discounts. In practice, $F_R(m|s,a)$ and $F_W(w|s,a)$ are usually unknown, and a seller can infer them from historical transaction data. We leave it as a future work.

We consider a continuous time system with infinite-horizon $t \in [0, \infty)$. Let t_i denote the arrival time of the i-th transaction, where $i \in \mathbb{N}_+$. We say a seller is at state $s \in \mathcal{S}$ if she has a reputation score s. Thus, the state space is $\mathcal{S}$. Decision epochs correspond to the time immediately following an arrival of a transaction. For example, the first decision epoch occurs at t_1. The initial decision epoch does not correspond to any transaction. Without any loss of generality, we index the initial decision epoch with 0, and use $t_0 = 0$ to denote its occurrence time. The seller is the decision maker and the decision to be made at each decision epoch is setting a discount $a \in \mathcal{A}$. We also call a the action. Note that the action set at each decision epoch is the same $\mathcal{A}$. When the seller chooses an action a at state s, she receives a lump sum reward denoted by $k(s,a)$, which can be expressed as $k(s,a) = \tilde{u}(a), \forall s \in \mathcal{S}, a \in \mathcal{A}$. Note that the lump sum reward corresponds to the unit profit earned from the next transaction. Namely, it is delayed to be payed in the next decision epoch.

Note that the inter-arrival (or waiting) time of decision epochs is $W(s,a)$, which has a CDF $F_W(w|s,a)$. Let $p(j|s,a)$, where $s, j \in \mathcal{S}, a \in \mathcal{A}$, denote the state transition probability

$$\begin{aligned} p(j|s,a) &\triangleq \mathbb{P}[\text{next state is } j | \text{current state } s, \text{ discount } a] \\ &= F_R(j-s|s,a) - F_R(j-s-1|s,a). \end{aligned}$$

Namely, $p(j|s,a)$ models the dynamics of the reputation score.

Setting price discounts may lead to some profit losses at the present decision epoch, but it can speed up the reputation score accumulation process, which may improve sellers' profit in subsequent decision epochs. To quantify the optimal discount and reputation trade-off, we use an *expected infinite-horizon discounted profit* for the seller. Precisely, we consider a continuous-time discounting rate $\alpha \in \mathbb{R}_+$ and define the expected infinite-horizon discounted profit as

$$v^\pi(s) \triangleq \mathbb{E}\left[\sum_{i=0}^{\infty} e^{-\alpha t_{i+1}} k(s_i, a_i) \middle| s_0 = s, \pi \right], \quad \forall s \in \mathcal{S},$$

where s_i, a_i denote the reputation score and discount at the i-th decision epoch, and π denotes a policy [2], which prescribes a discount for each transaction (or decision epoch). We also call $v^\pi(s)$ the *long-term profit*. For example, the long term profit for a new seller is $v^\pi(0)$. One interpretation of the discounting rate α is inflation from economic perspectives. The discounting rate α also reflects the willingness of a seller to trade discounts for reputation. Increasing α means that the seller cares less about the future profit (or is more keen about the present profit). In other words, she is less willing to trade discounts for reputation.

We define a stationary and deterministic (SD) policy as $\pi = (d)^{\infty}$, where $d : \mathcal{S} \to \mathcal{A}$ denotes a Markovian deterministic decision rule, which maps each state to a price discount.

Problem 1. *Given $F_R(m|s,a)$, $F_W(w|s,a)$ and s_0, select price discounts to maximize the long term profit. Formally,*

$$\underset{\pi \in \Pi}{\text{maximize}} \qquad v^{\pi}(s_0)$$

where Π denotes a set of all possible SD policies.

Problem 1 optimizes the long term profit over a special class of policies, i.e., SD policies, because SD policies suffice to attain the global maximum long term profit.

3. OPTIMAL PROFIT AND DISCOUNTS

It is mathematically intractable to derive the closed-form expression for the maximum long-term profit denoted by $v^*(s)$. In the following theorem, we identify a monotone property of $v^*(s)$.

Theorem 1. *For all $s \geq j$, where $s, j \in \mathcal{S}$, $v^*(s) \geq v^*(j)$ holds. Furthermore, $v^*(s)$ is non-increasing in α.*

Due to page limit, we present proofs to theorems in our technical report [3]. Theorem 1 states that the seller can earn more profit if her reputation score increases or the inflation decreases. In other words, sellers always have incentive to increase their reputation s-cores. Note that these monotone properties serve as an important building block for us to characterize the optimal discount.

Definition 2. *For each reputation score $s \in \mathcal{S}$, we define the associated action-dependent long term profit as*

$$Q(s,a) \triangleq \phi(s,a)V(s,a), \qquad \forall s \in \mathcal{S}, a \in \mathcal{A},$$

where $\phi(s,a) = \int_0^\infty e^{-\alpha w} dF_W(w|s,a)$ and $V(s,a) = k(s,a) + \sum_{j \in \mathcal{S}} p(j|s,a) v^*(j)$.

Given that a seller has a reputation score s, the $Q(s,a)$ gives the maximum long term profit she can earn by setting a discount a. The optimal discount $d^*(s)$ satisfies $d^*(s) \in \arg\max_{a \in \mathcal{A}} Q(s,a)$. We refer reads to our technical report [3] for more derivation details.

Theorem 2. *Suppose $a \in \mathcal{A}_s$, where $\mathcal{A}_s$ is defined as*

$$\mathcal{A}_s \triangleq \{a | Q(s,a) > 0, a \in \mathcal{A}\}, \qquad \forall s \in \mathcal{S}.$$

For all $j > \ell \geq s$, where $j, \ell, s \in \mathcal{S}$, $Q(j,a) \geq Q(\ell,a)$ holds.

Theorem 2 states that given the same discount $a \in \mathcal{A}_s$, the seller can earn more profit if she has a higher reputation score. We formulate the following problem to further study the optimal discount.

Problem 2. *Given s, select a to maximize $\ln Q(s,a)$:*

$$\underset{a \in \mathcal{A}}{\text{maximize}} \qquad \ln Q(s,a) = \ln \phi(s,a) + \ln V(s,a)$$

In Problem 2, we maximize the log function of the action-dependent long term profit. This treatment does not change the optimal discount and will facilitate the analysis.

Theorem 3. *Suppose $F_W(w|s,a)$ is strictly concave with respect to a and $F_R(m|s,a)$ is convex with respect to a. Problem 2 has a unique optimal solution.*

Theorem 3 derives sufficient conditions to guarantee the uniqueness of the optimal discount for each given s. This uniqueness enables us to further characterize the optimal discount via *comparative statics*. When the optimal discount is unique, it is algorithmically easy to locate it. For example, Eq. (1) satisfies the condition on $F_R(m|s,a)$.

Corollary 1. *Suppose $F_W(w|s,a)$ satisfies Eq. (2). If $\lambda(s,a)$ is strictly concave in a and $F_R(m|s,a)$ is convex in a, there exist a unique optimal discount for each reputation score s.*

Corollary 1 states that given the Poisson arrival of transactions, if the transaction's arrival rate $\lambda(s,a)$ has a diminishing return in the discount a, then the optimal discount is unique for each reputation score. For example, Eq. (3) satisfies the condition on $\lambda(s,a)$.

In order to apply comparative statics to further characterize the optimal discount, we define the following notation.

Definition 3. *We define the hazard function of $Q(s,a)$ with respect to the discount a as*

$$h(s,a) \triangleq -\frac{\partial Q(s,a)}{Q(s,a)} / \partial a = -\frac{\partial Q(s,a)}{\partial a} \frac{1}{Q(s,a)}, \quad \forall s \in \mathcal{S}, a \in \mathcal{A}_s.$$

The hazard function $h(s,a)$ measures the proportional reduction in the discount-dependent long-term profit (i.e., $-\partial Q(s,a)/Q(s,a)$) with respect to the marginal change in the price discounts (i.e., ∂a).

Theorem 4. *Suppose the conditions in Theorem 3 hold. If $h(s,a)$ is non-decreasing in α, the unique optimal discount $d^*(s)$ is non-increasing in α. If $h(s,a)$ is non-decreasing in s, the unique optimal discount $d^*(s)$ is non-increasing in s.*

Theorem 4 states sufficient conditions under which the unique discount is non-increasing in the discounting rate α and non-increasing in reputation score s. One interpretation is that the seller sets smaller discounts when the inflation increases or she is more keen about the present profit. More reputable sellers set smaller discounts.

4. CONCLUSION AND FUTURE WORK

In this paper, we develop a mathematical model to capture important factors of an E-commerce system. We formulate a profit maximization framework via a semi- Markov decision process (S-MDP) to explore the optimal trade-offs in selecting price discounts. We theoretically characterize the optimal profit and discount. Our work has number of future directions. In practice, the discount-dependent demands (i.e., buyers preferences over discounts) are unknown. How to perform online inference to determine the optimal discount? There may be multiple sellers selling similar product. How to capture the competition among them? How to capture competition among sellers? Seller updates discount when no transaction arrives for a while. How to extend our model to capture it?

5. REFERENCES

[1] L. Muchnik, S. Aral, and S. J. Taylor. Social influence bias: A randomized experiment. *Science*, 341(6146):647–651, 2013.

[2] M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[3] T. report. *http://www.cse.cuhk.edu.hk/%7Ehxie/MAMATR.pdf.*

[4] P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman. Reputation systems. *CACM*, 43(12):45–48, Dec. 2000.

[5] H. Xie and J. C. S. Lui. Modeling ebay-like reputation systems: Analysis, characterization and insurance mechanism design. *Performance Evaluation*, 91:132–149, 2015.

[6] H. Xu, D. Liu, H. Wang, and A. Stavrou. E-commerce reputation manipulation: The emergence of reputation-escalation-as-a-service. In *Proc. of WWW*, 2015.