Decision Support Systems xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

From clicking to consideration: A business intelligence approach to estimating consumers' consideration probabilities

Hao Wang, Qiang Wei *, Guoqing Chen

School of Economics and Management, Tsinghua University, Beijing, China

ARTICLE INFO

Available online xxxx

Keywords:
Electronic commerce
Business intelligence
Online marketing
Consideration probability
Collaborative filtering
Latent class model

ABSTRACT

With rapid advances in e-commerce applications and technologies, finding the chance that a product falls into a consumer's consideration set after being inspected (i.e., consideration probability, CP) becomes an important issue of recommendation services and marketing strategies for both academia and practitioners. This paper proposes a novel business intelligence (BI) approach (namely, the two-step estimation approach, TEA) to estimating CPs with a two-step procedure: one is to introduce partial belongings of consumers to the latent classes with both positive and negative preferences (tastes); the other step is to generate CPs based on the degrees of partial belongings in a weighted probability manner. Experiment results from different online shopping scenarios reveal that TEA is effective and outperforms the traditional latent class model.

1. Introduction

Considering the costs of search, consumers are often unable to evaluate all products before making purchase decisions [21,28,47]. Therefore, they tend to adopt a consider-then-choose process in which a consumer first selects a small group of products as a consideration set (also known as choice set or evoked set) and then chooses one of them to purchase [20,39,56,58,59,63]. For example, when shopping online, a consumer first inspects and selects some promising products into a shopping list from recommendations provided by consumer decision support systems (CDSSs, such as search engines or recommender systems), then deeply evaluates these selected products in a comparison matrix (a special type of decision aids that allow consumers to sort products by any attribute in an “products attributes” matrix) to choose the favorite one [16,17,21,63]. The set of products added into the shopping list (comparison matrix) can be viewed as the consideration set which is the output of the first stage (consideration stage) and the input of the second stage (choice stage) [7,21,36]. Compared to the process that directly chooses a product from all available ones, the consider-then-choose process is deemed typical and even more rational [23].

Thus, it becomes a primary focus of attention for e-sellers and e-marketplaces to estimating the probability that a product falls into a consumer's consideration set after being inspected, namely, consideration probability (CP) [15,30,39]. Compared with traditional brick-and-mortar stores where the behavior of inspecting is hard to observe and record, e-marketplaces are able to easily trace consumers' clicking behavior which can be seen as a strong signal of inspecting in online shopping [39]. Therefore, “click” and “inspect” are used interchangeably unless otherwise indicated in this paper. Fig. 1 illustrates the consider-then-choose process.

Essentially, CPs play an important role in predicting the chance that a consumer purchases a product after inspecting it, which is referred to as purchasing conversion rate (PCR) or purchasing probability (PP) and attracts numerous research efforts of academia [7,11,13,18,36,39–41,45,52,54,55]. For e-sellers, PPs can help find targeted consumers and formulate the profit of showing their ads to these targeted consumers [9,18,40,42,54]; and for e-marketplaces, PPs are necessary to effectively rank and recommend sponsored ads for revenue maximization [14,49,62]. From the perspective of the consider-then-choose process, the final choice purchased by a consumer must be 1) selected into the consideration set and 2) chosen from the consideration set. Consequently, the PP is equal to the CP multiplied to the choice probability (ChP) that is defined as the chance that a product is chosen from a consumer's consideration set, i.e., PP = CP × ChP [15,30,39]. The effectiveness of predicting the PP, therefore, greatly relies on the estimation of the CP. The relationship between PP, CP and ChP is illustrated in Fig. 1.

In addition to predicting PPs, CPs can help e-sellers find more targeted consumers which cannot be detected by PPs. Many real business cases indicate that less (more) considering a firm's products may lead to less (more) experiencing its products and, especially, the improved products [22,23]. In other words, consumers with low CPs to a firm's products may never want to experience them, even if the firm's products are greatly improved. The consumers with high

* Corresponding author. Tel.: +86 10 62789824.
E-mail addresses: wangh6.06@sem.tsinghua.edu.cn (H. Wang), weiq@sem.tsinghua.edu.cn (Q. Wei), chengq@sem.tsinghua.edu.cn (G. Chen).

0167-9236/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dss.2012.10.052

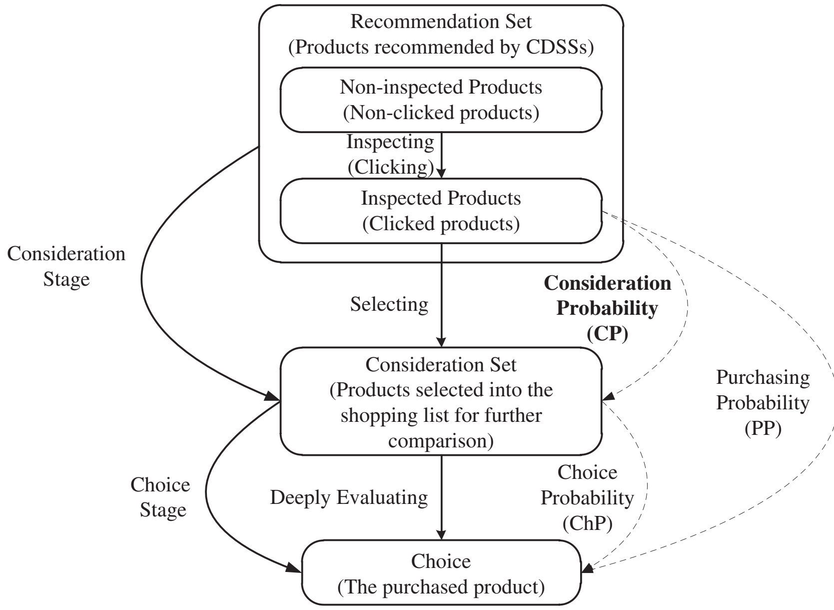

Fig. 1. The consider-then-choose process.

CPs but low PPs, however, would like to experience the improvement of the products, although they choose some other products currently. Therefore, these consumers with high CPs but low PPs to a firm's products are still targeted consumers, since they may switch to the firm's products when they experience its improvements [22,23].

Although the roots of related studies about consideration sets and CPs can be traced back to the extensive work in consumer behavior and marketing [27,44,50], there are few studies aiming to directly estimate CPs. One possible reason is that consideration sets are hard to observe without enough technical support from information systems [3,39]. For example, empirical studies usually use surveys to collect the data about consideration sets [10,48], which seemed to be neither efficient nor effective [39]. Another reason is that studies on consumer decision making usually treat consideration sets to be latent, not observable, to explain the consumer's purchasing behavior [43,59]. With the rapid advances in e-commerce technologies and applications nowadays, consumers' online behaviors, such as browsing, clicking, comparing, selecting and purchasing, can be recorded more effectively and efficiently, which makes consideration sets relatively observable [4,34,39]. For example, it is regarded as a more effective method to use the products clicked by a consumer as an estimation of his or her consideration set than to survey the consumer after purchase [39]. Moreover, with the help of more e-commerce tools, such as the shopping list and comparison matrix, the products added to the shopping list (comparison matrix) for further comparison can be seen as a more appropriate representation of consideration sets [7,21,36].

More observable consideration sets and detailed historical data about consumers' online behavior provide an opportunity for estimating CPs. In this paper, we focus on a general and representative problem: given the products that a consumer has inspected along with the products that have fallen into his or her consideration set, what are the CPs of other non-inspected products to this consumer if they are inspected, i.e., the probabilities that other non-inspected products are selected into the consideration set after being inspected by this consumer? In answering this question, this paper presents a novel two-step approach, in which CPs are effectively estimated. The paper is organized as follows. The problem is defined in Section 2. Related studies and their limitations are discussed in Section 3. The proposed approach is presented in Section 4. Section 5 illustrates experimental results as well as the analysis. The conclusion is provided in the last section.

2. Problem definition

Formally, the research question is stated as follows. A consumer, $c \in C$ (C is the set of all consumers), wants to select several products as his or her consideration set from recommended products, S. Let S_c ($\bar{S}_c$) denote the set of inspected (non-inspected) products for consumer c, where $\bar{S}_c = S - S_c$. Let a_{cs} be a binary variable with $a_{cs}=1$ ($a_{cs}=0$) denoting that product s is (is not) in the consideration set of consumer c after being inspected, where the values of a_{cs} are supposedly known for s in S_c and unknown for s in $\bar{S}_c$. Then the research question is to estimate $\Pr(a_{cs}=1)$, $\forall s \in \bar{S}_c$ based on the historical data about all consumers' inspected products and their consideration sets.

For example, suppose that a consumer $c \in C$ wants to select several laptop computers into the comparison matrix for further evaluation at an e-marketplace. The recommendations provided by CDSSs are 4 different computers, i.e., $S=\{s_1,s_2,s_3,s_4\}$. At the time of t_0, c has inspected no computer (i.e., $S_c=\{\}$, $\bar{S}_c=\{s_1,s_2,s_3,s_4\}$ and all a_{cs}'s are unknown). That is, the task is to estimate all computers' CPs (i.e., $\Pr(a_{cs}=1)$, $\forall s \in \{s_1,s_2,s_3,s_4\}$). Suppose that at the time of t_1, c inspects computer s_1 and adds it into the comparison matrix (i.e., $S_c=\{s_1\}$, $\bar{S}_c=\{s_2,s_3,s_4\}$, $a_{cs_1}=1$ and the values of a_{cs} are unknown $\forall s \in \{s_2,s_3,s_4\}$). Then, what needs to be done is to estimate all non-inspected computers' CPs (i.e., $\Pr(a_{cs}=1)$, $\forall s \in \{s_2,s_3,s_4\}$). If at the time of t_2, c inspects product s_3 but does not add it into the comparison matrix (i.e., $S_c=\{s_1,s_3\}$, $\bar{S}_c=\{s_2,s_4\}$, $a_{cs_1}=1$, $a_{cs_3}=0$ and the values of a_{cs} are unknown $\forall s \in \{s_2,s_4\}$), then $\Pr(a_{cs}=1)$ needs to be estimated for the remaining products, i.e., $\forall s \in \{s_2,s_4\}$. If at the time of t_3, c inspects product s_2 and adds it into the comparison matrix (i.e., $S_c=\{s_1,s_2,s_3\}$, $\bar{S}_c=\{s_4\}$, $a_{cs_1}=1$, $a_{cs_2}=1$, $a_{cs_3}=0$, and the value of a_{cs} is unknown $\forall s \in \{s_4\}$), $\Pr(a_{cs}=1)$ needs to be estimated $\forall s \in \{s_4\}$. This process is illustrated in Table 1.

Table 1
The search process of consumer c.

Time	Non-inspected products	Inspected products	Consideration set
t_0	s_1, s_2, s_3, s_4		
t_1	s_2, s_3, s_4	s_1	s_1
t_2	s_2, s_4	s_1, s_3	s_1
t_3	s_4	s_1, s_2, s_3	s_1, s_2
…	…	…	…

Please cite this article as: H. Wang, et al., From clicking to consideration: A business intelligence approach to estimating consumers' consideration probabilities, Decision Support Systems (2012), http://dx.doi.org/10.1016/j.dss.2012.10.052