Available online at www.sciencedirect.com
ScienceDirect
Electronic Commerce Research and Applications 5 (2006) 295–304
Electronic Commerce Research and Applications
www.elsevier.com/locate/ecra

Comparison of product bundling strategies on different online shopping behaviors

Tzyy-Ching Yang [a], Hsiangchu Lai [b,*]

[a] Department of Computer Science and Information Management, Providence University, 200 Chung-Chi Road, Taichung 433, Taiwan
[b] Department of Information Management, National Sun Yat-sen University, 70 Lien-Hai Road, Kaohsiung 804, Taiwan

Received 2 May 2005; received in revised form 3 November 2005; accepted 27 April 2006
Available online 12 June 2006

Abstract

Bundling is a very popular sales-promotion tool, in which a critical issue is to decide what products should be sold together in order to improve sales. Traditionally, this decision is based on the order data collected from the points of sale. However, Internet marketing now allows marketers to efficiently collect not only order data but also browsing and shopping-cart data, which provide marketers with information on the consumers' decision-making processes, rather than only the final shopping decisions. The present study aimed to determine the value of this newly available information by comparing the performance of decision-making on product bundling based on three types of data on online shopping behaviors. The results from a field experiment reveal that significantly better decisions are made on the bundling of products when browsing and shopping-cart data are integrated than when only order data or browsing data are used.

Keywords: Online behavior; Product bundling; Shopping cart; Market basket analysis; Association rules

1. Introduction

A better understanding of customers allows better marketing strategies to be designed. The Internet provides marketers with much more data on customers, and consequently has brought marketing management into a new age [41]. In addition to collecting order data through the point of sale (POS), as in the pre-Internet age, on the Internet the entire shopping procedure of any customer can be recorded, including not only what has been ordered, but also what has been clicked or browsed and what has been moved in or moved out from the shopping cart, and the timings of all of these processes.

Various studies have demonstrated the usefulness of data on shopping processes and outcomes [21,22]. The essential difference between order data and other data on online shopping behaviors is that the former indicates the final shopping decisions while the latter provides information on customer behavior whilst making shopping decisions. Moreover, a major difference between browsing and shopping-cart data is that the information embedded in the latter is more closely related to the final shopping decision [5]. Data on customer browsing behavior have been explored in many studies, such as to find path travel patterns [6,12], website traffic, the products being browsed the most, the hot area of a web page, and the customer profiles based on these browsing data [30,31]. However, very few studies have investigated data related to customer shopping carts [15,32].

The usefulness of data on customer online shopping behaviors to learning about customers has resulted in data mining becoming an important and popular technique for exploring customer profiles, shopping behaviors, and other aspects of online shopping, because the associated data sets are often huge. For example, Chen et al. [11] integrated customer behavioral variables, demographic variables, and a transaction database to establish a method for mining changes in customer behavior. Jiao and Zhang [25]

* Corresponding author. Tel.: +886 7 5254 776; fax: +886 7 5254 799.
E-mail addresses: tcyang@pu.edu.tw (T.-C. Yang), hclai@mail.nsysu.edu.tw (H. Lai).

1567-4223/$ - see front matter
doi:10.1016/j.elerap.2006.04.006

attempted to develop an explicit decision support to improve product portfolio identification by applying association-rule mining. Further, Rygielski et al. [44] pointed out that organizations must be aware of the trade-offs associated with choosing suitable data-mining software for use in customer relationship management (CRM).

Several sales promotion tools are used in the marketplace, such as free samples, coupons, premiums, bundle pricing (bundling), and cross-promotions [45]. Bundling is a very common practice that involves combining two or more products or services and selling them at a set price [20,47]. Examples of bundles are opera season tickets (e.g., tickets to various events sold together) and Internet services (e.g., a combination of Web access, e-mail, personalized content, and an Internet search program) [46]. The bundled price is almost always lower than the cost of buying all the products separately. Estelami [18] found that consumers save an average of about 8% when purchasing a bundle of items relative to buying the items individually. The challenge is how to choose the appropriate products to be bundled in order to achieve the expected promotion performance, such as by creating new markets or increasing customer loyalty, sales, or profits [37].

Previous research has investigated product bundling from different viewpoints. Determining the products to be bundled so as to maximize performance is the most critical issue in formulating bundling strategies. Searching for associations in the market basket has been previously used to determining appropriate bundles. Further, Munger and Grewal [36] examined the effects of the bundling format and framing of promotional discounts on perceived quality, price acceptability, perceived value and subsequent purchase intentions. Chakravarti et al. [10] examined the effects of the presentation of a multicomponent product bundle on customer evaluation and choices, as well as the underlying processing effects. Stremersch and Tellis [46] showed that a company that exploits opportunities offered by bundling will enjoy increased market share and profits. However, none of the previous research studies have considered product bundling based on data on online shopping behaviors, which represent newly available data from which customer profiles can be elucidated.

The purpose of this study was to assess the value of data on online shopping behaviors in making decisions on product bundling by examining the performance of product bundling strategies based on different data sets related to online shopping behaviors. The remainder of this paper is organized as follows: Section 2 surveys the related literature on product bundling, market basket analysis, marketing implications of data on online shopping behaviors, and association rules; Section 3 proposes three product bundling strategies based on different types of data on online shopping behaviors; Section 4 compares the performance of these three product bundling strategies based on data collected from a field experiment; and the paper ends with the conclusions drawn in Section 5, which also discusses some of the limitations of the study.

2. Literature review

2.1. Product bundling

Product bundling is a pervasive selling strategy in markets, examples of which include sporting and cultural organizations offering season tickets, restaurants providing complete dinners, and retail stores offering discounts to a customer buying more than one product. How to bundle products in order to maximize profits under different market environments has been a popular research issue in the economics field [1,14,16,17,34,38]. However, in this study we focused on the marketing aspects of product bundling. Stremersch and Tellis [46] identified two key dimensions in classifying bundling strategies: focus and form. The focus of bundling can be either the price or the product, while the form of bundling can be none, pure, or mixed. In the paper, they also provide managers with a framework with which to understand and choose bundling strategies.

Janiszewski and Cunha [24] indicated that the impact of a price discount on the perceived attractiveness of a bundle depends on the type of product that is being discounted. They also concluded that reference dependence and product importance independently contribute to the effects of price discounts. Agarwal and Chatterjee [2] examined the decision difficulties experienced by customers when selecting from a menu of bundles. They found that larger bundles make decisions more difficult; more specialized services in the competing bundles increases the decision difficulty for small, but not large, bundles; and that the choice difficulty is greater for bundles that are more similar.

2.2. Market basket analysis

Market basket analysis refers to investigating the composition of the basket of products purchased by a household during a single shopping experience [43]. Retailers have long been interested in learning about the cross-category purchase behavior of their customers [35], since such information makes it easier for the retailer to decide whether to group products by brand or by product type, for example [9]. The market basket choice refers to the decision process in which a consumer selects items from several product categories during the same shopping experience [43].

Therefore, basket analysis can provide the distribution of shoppers' purchases based on different viewpoints, such as the product itself, the product category, the shopper's background, or the average purchases per shopper [26]. Such distribution information will aid decisions on aspects such as planning and designing advertising, sales promotions, store layout, and product placement [8]. In addition, basket data contain important information about the structure of brand preferences both within and across product categories [42], and hence a richer picture of customer behavior and better decisions on the bundling of products may result from identifying the associations between product purchases at the POS [40]. The other way to find such associations is to apply association-rule algorithms to both browsing and shopping-cart data.

2.3. Marketing implications of data on online shopping behaviors

Generally, basket analysis is applied to data on the composition of the basket of purchased products – that is, on the content of the final order. Advances in information technology have radically changed the methods used to collect consumer data [19]. For example, computerized checkouts generate almost immediate feedback about the profitability of brands, product groups, and the effects of marketing activities such as in-store promotions and weekly advertising [26]. In the Internet age, it is possible to collect not only order data at the checkout but also data on the consumer's entire shopping behavior at a website.

Kotler [29] proposed a model of the successive sets involved in the consumer decision-making process, which revealed that consumers' decision-making processes involve five successive sets: total set, awareness set, consideration set, choice set, and final decision. As shown in Fig. 1, the total set includes all brands available to a consumer. However, the customer is likely to be aware of only a subset of these brands (i.e., the awareness set), with only some of these brands (i.e., the consideration set) fulfilling the consumer's initial buying criteria. The collection of more information by the customer will result in only a few brands remaining acceptable, which together form the choice set. Finally, the customer chooses one item from the choice set.

Through the information search process, the products foremost in the consumer's mind will change from the total set to the awareness set, consideration set, and choice set, with the final decision then being made. Whether a product proceeds through each stage and reaches the final choice depends on the information available to the customer and his or her thought processes at each stage of the decision process. If marketers are provided with information about the thought processes that are important at the awareness, consideration, and choice sets to the customers' decision-making processes, they can understand how the final choice results from the successive sets, which should result in better marketing decisions being made.

Collecting data on online shopping behaviors is one way to understand what a customer is interested in at each step and the possible thoughts underlying this process. Traditionally, the only data collected through the POS is the order data, which reflects the final shopping decision rather than the previous sets as outlined in Fig. 1. Although knowing what consumers have bought is useful to marketers, this does not allow them to learn why consumers have made a particular purchase decision and not bought some other item.

In the Internet age, it is relatively easy to collect data related to the decision process such as browsing and shopping-cart data. The browsing data may provide information on the thought processes that lead from the awareness set to the consideration set and the choice set, and the shopping-cart data may provide more information on the thought processes from the consideration set to the choice set, and then to making the final decision. It should be noted that shopping-cart data may differ from the order data, since items in the shopping cart can be changed at any time before the purchase is finalized.

Information on the stages closer to the final shopping decision are especially valuable to building a more accurate customer profile. Whilst shopping-cart data is considered to provide more reliable information about why the consumers buy or do not buy certain products, even the browsing data provides useful information on the customers' interest that cannot be found in the shopping-cart or order data. For example, the changes that occur from the awareness set to the consideration set, and then to the choice set might be due to the customer's preferences, budget, available shopping time, and exposure to marketing promotion programs, for example. In addition, the patterns of the

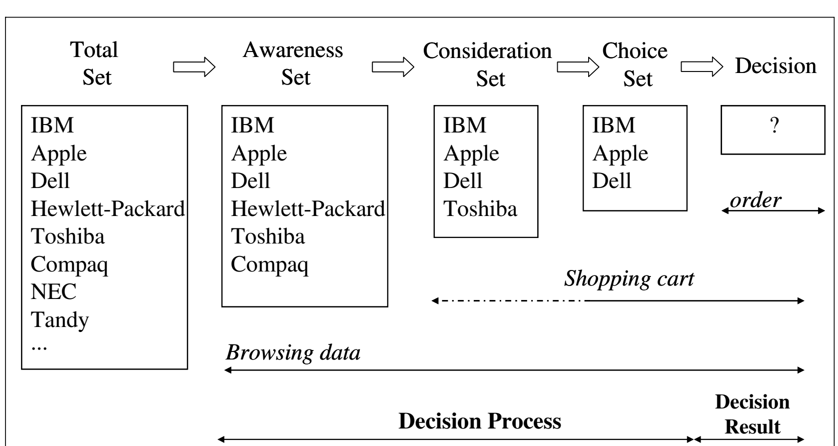

Fig. 1. Relations between successive sets involved in consumers' decision making and their online behavior [modified from [29]].

different successive sets can differ between customers (e.g., a product in a customer's choice set might only be in another customer's consideration set).

2.4. Association rules

The problem of searching for association rules in large databases using data mining has been widely studied [3,4,33]. A very popular application area is to find associations by analyzing the market basket of products purchased in a single shopping experience [13,43]. The goal is to discover buying patterns, such as two or more items that are often bought together. The past research has elucidated techniques for improving the performance of algorithms for discovering association rules in large databases of sales information.

Agrawal et al. [4] first introduced the problem of finding association rules. An association rule can be expressed by the form $X \Rightarrow Y$ when c% of transactions that contain X also contain Y. We call c the confidence of the association rule, where $c\% = \text{Prob}(Y|X)$. The rule $X \Rightarrow Y$ has support **s,** where $s\% = \text{Prob}(X \cap Y)$. An example outcome of association-rule mining is determining that "90% of customers who buy A and B will also buy C" in a transaction database; in this case the confidence of the rule is 90%. Another parameter is the support of an item set, such as $\{A, B, C\}$, which is defined as the percentage of times that the item set is contained in all of the transactions [27]. Therefore, the problem can be defined by generating all association rules that have support greater than the user-specified minimum support [3].

Apriori is the most popular algorithm (Table 1) used to find association rules. This algorithm constructs a candidate set of large $(k-1)$-item sets, counts the number of occurrences of each candidate item set, and then determines large k-item sets based on the minimum support in each iteration [12]. Some algorithms representing revisions of the Apriori have been proposed, such as AprioriTid, AprioriHybrid [3], OCD [33], DHP [39], and SETM [23].

Table 1
Algorithm Apriori and the Apriori–gen function [3]

Algorithm Apriori
(1) $L_1 = \{\text{large } 1 - \text{itemsets}\}$;
(2) for ($k = 2$; $L_{k-1} \neq \varphi$; k++) do begin
(3) C_k = apriori-gen(L_{k-1}); // New candidates
(4) for all transactions $t \in D$ do begin
(5) C_t = subset(C_k,t); //Candidates contained in t
(6) forall candidates $c \in C_t$ do
(7) c.count++;
(8) end
(9) $L_k = \{c \in C_k |$ c.count $\geqslant$ minsup$\}$
(10)end
(11)Answer $= \cup_k L_k$;

Apriori–gen function
(12) insert into C_k
(13) select p.item$_1$, p.item$_2$, ..., p.item$_{k-1}$, q.item$_{k-1}$
from $L_{k-1,p}$, $L_{k-1,q}$
where p.item$_1 = q$.item$_1$,..., p.item$_{k-2} = q$.item$_{k-2}$,p.item$_{k-1} < q$.item$_{k-1}$;
(14) forall itemsets $c \in C_k$ do
(15) forall $(k-1)$ subsets s of c do
(16) if ($s \notin L_{k-1}$) then delete c from C_k;

Notation: (1) L_k: set of large k-item sets. (2) C_k: set of candidate k-item sets. (3) minsup: minimum support.

3. Different product bundling strategies

Based on Fig. 1, we can see there are three different types of data on online shopping behaviors: browsing data, shopping-cart data, and order data. Fig. 1 also displays the relationship between the online shopping behaviors and the successive sets proposed by Kotler [29]. Traditional basket analysis is based on ordered products only; that is, the final shopping decision. Because the preferences underlying the decision process may be hidden, the other types of data on online shopping behaviors such as browsing and shopping-cart data should be considered when determining the customer profile and thereafter the optimal marketing strategy.

Therefore, in addition to searching for associations in the order data, we propose another two bundling strategies based on the other types of data on online shopping behaviors that are available: (1) that based on the browsing data only; and (2) that based on both browsing and shopping-cart data. The general method first requires the collection period of the data set and the required minimum support to be set. The data belonging to the same customer are then merged, and the Apriori algorithm is applied to find associations that fulfill the minimum support. The customer's shopping is merged (rather than using a single shopping record) in order to obtain a more robust result, since a single shopping record may not be representative of the customer behavior. Comparing the performance of these different strategies will reveal if the new proposed strategies are better than the traditional one (based on order data only). In the following sections, we describe the strategies and their advantages and disadvantages.

3.1. Strategy based on order data only

Before the rise of the Internet it was only possible for order data to be collected efficiently, and hence it was very common for traditional marketers to make bundling decisions based on these data only. In order to improve the robustness of product bundling, we suggest searching for associations based on all orders of a customer. First, all of the order data belonging to the same customer are merged, and then the algorithm is applied to the reorganized data set to determine the product associations. The process is summarized in Table 2.

The weakness of the strategy in Table 2 is that the result will not be robust if insufficient order data are available. Further, the order data directly imply only the purchase decision rather than the underlying decision process, and therefore cannot be applied to one-time shopping such as that of large electric appliances. In addition, a customer's

Table 2
Product bundling strategy based on order data

Step 1: Let marketers set the period of analysis (*P*) and the minimum support (*S*)
Step 2: Select data collected during *P*
Step 3: Merge the order data belonging to the same customer
Step 4: Apply an association-rule algorithm to determine the large item sets whose support is larger than *S*

final order may also depend on factors such as his or her budget, preferences, available shopping time, and the other available substitute products. For example, a customer may browse a product on the Internet but not buy it, instead buying the product later at a physical store; in this case the Internet order data do not reflect the customer's interest. However, the advantage of this bundling strategy is that it is easier to collect more comprehensive information about the customer's shopping behavior.

3.2. Strategy based on browsing data only

The second proposed product bundling strategy is based on the browsing data so as to overcome the possibility of the order data not reflecting the hidden preferences of customers. As discussed in Section 2, browsing data may include more information about the shopping process, in terms of the customer's consideration set or choice set (but not what is actually bought by the customer). Hence, customers are interested in these products but may still choose not to buy them due to reasons such as high price or low product quality, or a low customer demand. Bundling might allow these products to reach the customer's final stage of the shopping process. Table 3 lists the process of determining product bundling based on browsing data.

The weakness of the strategy in Table 3 is that the browsing behavior might be influenced by the design of the website, such as what is contained within the hot area and how the hyperlinks are arranged. That is, a product may be browsed the most due to its place in the website layout rather than the customer's preferences. One way to reduce this interference is to delete data from the data set whose browsing time is less than a minimum threshold, as in Table 3. One advantage of this strategy is that the collected data set will be much larger, and hence will be more applicable to data-mining techniques.

Table 3
Product bundling strategy based on browsing data

Step 1: Let marketers set the period of analysis (*P*), minimum support (*S*), and minimum browsing time (*T*)
Step 2: Select browsing data collected during *P*. If the browsing time of the data is less than *T*, the data are removed
Step 3: Merge the browsing data belonging to the same customer
Step 4: Apply an association-rule algorithm to determine the large item sets whose support is larger than *S*

Table 4
Product bundling strategy based on both browsing and shopping-cart data

Step 1: Let marketers set the period of analysis (*P*), minimum support (*S*), and minimum browsing time (*T*)
Step 2: Select browsing data collected during *P*. If the browsing time of the data is not greater than *T* or the page is not linked to the shopping cart, the data are removed
Step 3: Merge the browsing data belonging to the same customer
Step 4: Apply an association-rule algorithm to determine the large item sets whose support is larger than *S*
Step 5: Select shopping-cart data collected during *P*, and then merge the data belonging to the same customer
Step 6: Calculate the support based on the shopping-cart data for each of the large item sets found in Step 4, and choose the large item sets whose support is larger than *S*

3.3. Strategy based on both browsing and shopping-cart data

The above problems of the order data only reflecting the final purchase preferences of a few consumers and the browsing data being easily influenced by the design of the websites are overcome by using a strategy based on both browsing and shopping-cart data (Table 4). This strategy involves finding the large item sets based on the browsing data first, and then examining whether the support of each of these large item sets exceeds the minimum support requirement based on the shopping cart.

The advantage of this integrated method is that it can reflect the behavior of more consumers and the hidden preferences of the customers by using browsing data instead of order data to find the large item sets. Moreover, it also can avoid the influence of the website layout of product catalogs by considering the shopping-cart data as well. In addition, the associations resulting from the strategy are supported both by the browsing behavior and by the items placed in the shopping cart, and therefore they more accurately reflect the customers' potential shopping preferences.

4. Experimental design

4.1. Data collection

For the purpose of examining and comparing the performance of the three product bundling strategies, in the present study the data on the online shopping behaviors of customers were collected from the website of a publisher specializing in information technology and electronic commerce books. The site Web server was internet information server (IIS) running on Windows NT, with SQL Server as its database. ASP was adopted to develop the online membership functions of the publisher's website and dynamic shopping homepages in order to collect customers' backgrounds, shopping-cart data, and order data. In addition, the log files were implemented in the W3C Extension format, and several cookie items were created to record all the browsing behavior of every member. The customers' browsing data could be extracted from the log files.

Before the field experiment, the online shopping behaviors of customers were recorded by the aforementioned technologies for six months, during which the publisher published 136 books in 14 categories. A total of 1500 customers joined up as members, among whom 77.5% were male, 68.4% were educated at least to university level, and 53.7% were students (the remaining customers were all employed).

There were 24,316 browsing sessions during the period of online data collection. Removing data without login information resulted in 2836 valid sessions belonging to 1472 members. There were 197 orders, involving 459 books bought by 168 members. The shopping-cart data comprised 719 valid records belonging to 447 members.

4.2. Operation of product bundling strategies

The following sections discuss the product bundling decisions that resulted from the three strategies based on the collected data on online shopping behaviors.

4.2.1. Strategy based on order data only

Merging the order data belonging to the same customer revealed that 95 customers purchased more than one book. We set the minimum support as 5(%), indicating that the chance of any pair of products being purchased by the same customer was greater than 5%. We then applied the Apriori algorithm in Table 1 to find product associations; the top-ten product associations with higher support are listed in Table 5, which indicates, for example, that 17.9% of the customers purchased both book L001 and book L06.

4.2.2. Strategy based on browsing data only

Any records associated with a browsing time of less than 5 s were first removed, which left 1134 sessions. Merging the browsing data belonging to the same customer resulted in 517 valid browsing sessions. Again, the Apriori algorithm was applied to find the product association with the minimum support set as 5(%) (i.e., the chance of any pair of products being browsed by the same customer was greater than 5%). The top-ten product associations are summarized in Table 6, which indicates, for example, that 18% of the customers browsed both book L001 and book S07.

Table 5
Product associations based on order data

Product item 1	Product item 2	Support (%)
L001	L06	17.9
B15	B16	13.7
B18	B21	10.5
L001	T001	10.5
L06	T001	9.5
B15	B18	9.5
B19	B21	9.5
B15	B21	8.4
B09	B15	8.4
B14	B15	7.4

Table 6
Product associations based on browsing data

Product item 1	Product item 2	Support (%)
L001	S07	18.0
L06	S07	16.8
B21	S07	11.6
L001	L06	10.8
S07	T002	8.9
S07	T001	7.9
B20	S07	7.9
B18	S07	7.4
N26	S07	7.4
L002	S07	7.4

4.2.3. Strategy based on both browsing and shopping-cart data

First, the Apriori algorithm was used to find product associations based on browsing data with the minimum support set as 5(%). In this case, the data cleaning process deleted not only the records associated with a browsing time of less than 5 s but also those records relating to books that did not link to shopping cart. After data cleaning, 232 valid browsing sessions were used to find product associations. We then checked whether or not either the product of each two-product association based on the browsing data appeared in the shopping-cart data. Table 7 presents the results in the order of the probability of both of the two products in a product association appearing in the shopping cart simultaneously. Taking the first pair as example, the chance of both B18 and B2 being browsed by the same customer is 8.2%, while the chance of both being placed in this customer's shopping cart is 6.9%.

4.2.4. Selection of the product bundling strategies

Table 8 summarizes the product associations extracted by applying the above three strategies. After discussions with the manager of the publishing company, we decided

Table 7
Product associations based on both browsing and shopping-cart data

Based on browsing data			Probability of product associations appearing in a consumer's shopping cart		
Product item 1	Product item 2	Support (%)	None (%)	One product (%)	Both products (%)
B18	B21	8.2	78.4	14.7	6.9
B19	B21	5.6	86.2	8.7	5.0
L001	L06	15.9	86.7	10.1	3.2
B20	B21	11.6	83.9	13.3	2.8
B18	B20	7.8	80.7	17.0	2.3
B21	B22	6.9	83.9	13.8	2.3
L001	L002	6.9	89.0	10.1	0.9
B21	SB01	5.6	85.3	13.8	0.9
B19	B20	5.6	89.0	10.1	0.9
L002	L303	8.2	94.0	5.5	0.5

Table 8
Summary of association rules

Priority	Based on order data only	Based on browsing data only	Based on both browsing & shopping-cart data
1	(L001, L06)[1]	(L001, S07)	(B18, B21)
2	(**B15, B16**)	(**L06, S07**)	(**B19, B21**)
3	(B18, B21) or (L001, T001)[1]	(B21, S07)	(L001, L06)[1]
4	(B18, B21) or (L001, T001)[1]	(L001, L06)[1]	(B20, B21)
5	(L06, T001)[1] or (B15, B18) or (B19, B21)	(S07, T002)[2]	(B18, B20) or (B21, B22)
6	(L06, T001)[1] or(**B15, B18**) or (B19, B21)	(**S07, T001**) or (B20, S07)	(B18, B20) or(**B21, B22**)
7	(L06, T001)[1] or (B15, B18) or (B19, B21)	(S07, T001) or (B20, S07)	(L001, L002) or (B21, SB01) or (B19, B20)
8	(B15, B21) or (B09, B15)	(B20, S07) or (N26, S07) or (L002, S07)	(L001, L002) or (B21, SB01) or (B19, B20)
9	(B15, B21) or (B09, B15)	(B20, S07) or (N26, S07) or (L002, S07)	(L001, L002) or (B21, SB01) or (B19, B20)
10	(B14, B15)	(B20, S07) or (N26, S07) or (L002, S07)	(L002, L303)[1]

Notes. 1. L001, L002, L06, L303, L306, and T001 had been promoted just prior to the experiment, so these rules were removed from the candidate list. 2. T002 was sold out, so this rule was also removed.

Table 9
Nine bundling strategies selected for the field experiment

Product bundling strategies	Discounted bundling strategies
Based on order data only	(B15, B16), (B15, B18), (B15, B16, B18)
Based on browsing data only	(L06, S07), (S07, T001), (L06, S07, T001)
Based on both browsing and shopping-cart data	(B19, B21), (B21, B22), (B19, B21, B22)

to choose two sets of product associations for each strategy to serve as the product bundling strategies in the experiment. In order to reduce any interference, the selected product associations extracted from different online data should have the same ranking. In addition, books that had been promoted recently or sold out were removed from the candidate list. Finally, the product associations in boldface font in Table 8 were selected as the targets of product bundles in the experiment. The selected product bundles are all ranked at 2nd and 5th for the three different bundling strategies.

We found that one book appeared in both of the two chosen bundles for each of the different strategies. We decided to add a new bundling that included three products comprising the union of both product bundles selected from each strategy. Finally, the nine bundling strategies listed in Table 9 were selected for inclusion in a field experiment on the website of the publisher.

4.3. Implementation

After the product bundling had been decided, a field experiment was implemented. We collaborated with the publisher to provide three different pricing bundles on the

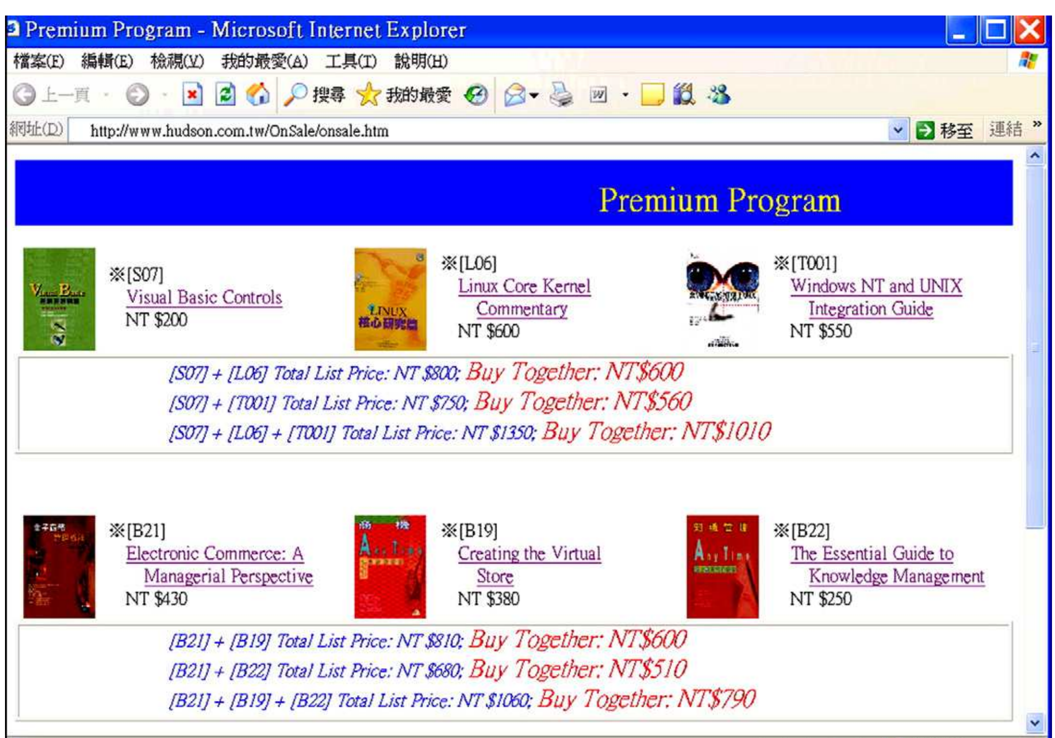

Fig. 2. Webpage showing the pricing of product bundles.

publisher's website (Fig. 2) that offered discounts of approximately 25%. This represents a type of field experiment because the experiment was embedded in the normal functioning of the publisher's website, and customers did not know that it formed part of an experiment. All of the online shopping behaviors were recorded.

5. Data analysis

Before the field experiment, the publisher had provided bundling strategies for 6 months that were based on the manager's expertise. During this period there were means of 2.7 customers, 2.8 orders, and nine books sold per month. Nine bundling strategies were promoted for 1 month in our field experiment, during which there were 22 customers, 33 orders, and 78 books sold, which indicates a clear improvement in the performance.

We chose the number of purchased books to be the measurement variable; that is, more books being purchased due to the discounted bundling strategies indicates a better performance. The mean and standard deviation values for the number of purchased books for each method are summarized in Table 10. Two-way ANOVA was used to determine whether the performance of applying both browsing and shopping-cart data to deciding product bundling was significantly better than that of the other two strategies that applied one type of data only. The results are summarized in Table 11.

Table 10
Means and standard deviations of purchasing for the discounted bundling strategies

Strategy	Mean	Standard deviation
Based on order data	0.818	1.296
Based on browsing data	0.591	1.008
Based on both browsing and shopping-cart data	2.136	2.765

Table 11
Results of two-way ANOVA for the discounted bundling strategies

Source	Sum of squares	DF	Mean square	F	p
Users	47.15	21	2.25	0.55	0.929
Methods	30.64	2	15.32	3.78	0.031*
Error	170.03	42	4.05		
Total	247.82	65			

In the experiment, every customer was allowed to purchase any product bundles extracted from the three strategies. Therefore, the sources of variation included both the preferences differing between users and the different strategies. The results indicated that the purchasing of the discounted bundles was not significantly affected by the users' preferences ($F = 0.55$, $p = 0.929$) but was significantly affected by the three product bundling strategies ($F = 3.78$, $p = 0.031$).

We next used Scheffe's multiple-comparison test to estimate the 95% simultaneous confidence intervals of the different strategies in order to compare the means for each pair of strategies. Table 12 indicates that the strategy based on both browsing and shopping-cart data performed significantly better than the strategy based on order data only and the strategy based on browsing data only. Note that the performance did not differ significantly between the strategy based on order data only and the strategy based on browsing data only.

6. Conclusions

The Internet makes it easy to collect both browsing and shopping-cart data, in addition to traditionally collected order data. This study evaluated the value of this newly available information by comparing the performance of making decisions on product bundling based on different types of data on online shopping behaviors. A field experiment revealed that the usefulness in making decisions on product bundling based on both browsing and shopping-cart data was significantly higher than that when using order data only or browsing data only.

This study was subject to some limitations. The bookstore chosen for performing the field experiment specializes in selling books on technology and electronic commerce, and sells a relatively small number of books. Further, a book essentially represents a type of one-time purchase and the field experiment was implemented for only 1 month, resulting in a small total number of orders. Many users browsed the website without logging in, and hence their browsing and shopping-cart data could not be collected. Finally, the customers' online behavior would be impacted by several other factors such as the website design, the popularity of the publisher, and the maturity of electronic commerce.

There are some promising directions for future research – including several possible benefits of bundling

Table 12
Simultaneous 95% confidence intervals for the different product bundling strategies

Confidence interval (i–j)	Strategy$_j$		
Strategy$_i$	Based on order data	Based on browsing data	Based on both browsing and shopping-cart data
Based on order data		0.227 ± 1.089	-1.318 ± 1.089*
Based on browsing data	-0.227 ± 1.089		-1.545 ± 1.089*
Based on both browsing and shopping-cart data	1.318 ± 1.089*	1.545 ± 1.089*	

such as promoting complimentary or new products – which were not considered in our research. How bundle pricing influences sales performance is another promising issue for future research, as is determining better data-mining techniques. Finally, a major challenge for Internet retailers is the high percentage of consumers who abandon their virtual shopping carts without finalizing their orders [28], and hence another worthwhile research topic is how best to interact with potential buyers so as to maximize the probability of their shopping carts being finalized as orders [7].

Acknowledgement

Hsiangchu Lai thanks the Taiwan National Science Council for the funding provided under Grant No. NSC89-2416-H-110-029.

References

[1] W.J. Adams, J.L. Yellen, Commodity bundling and the burden of monopoly, The Quarterly Journal of Economics 90 (2) (1976) 475–498.

[2] M.K. Agarwal, S. Chatterjee, Complexity, uniqueness, and similarity in between-bundle choice, Journal of Product and Brand Management 12 (6) (2003) 358–376.

[3] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proceedings of the 20th VLDB Conference, Santiago, Chile, 1994.

[4] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, 1993.

[5] J. Balla, G. Desai, J. Fenner, Understanding e-commerce, Inform 13 (8) (1999) 28–30.

[6] R. Barrett, P.P. Maglio, D.C. Kellem, Autonomous interface agents, in: Proceedings of the Conference on Human Factors in Computer Systems, Atlanta, GA, 1997.

[7] R. Bayan, How business owners can prevent online shopping failures, Link-up 17 (4) (2000) 32–33.

[8] T.J. Blischok, Every transaction tells a story, Chain Store Age Executive 71 (3) (1995) 50–62.

[9] R.R. Burke, Virtual shopping: breakthrough in marketing research, Journal of Product Innovation Management 13 (6) (1996) 558–559.

[10] D. Chakravarti, R. Krish, P. Paul, J. Srivastava, Partitioned presentation of multicomponent bundle prices: evaluation, choice and underlying processing effects, Journal of Consumer Psychology 12 (3) (2002) 215–229.

[11] M.-C. Chen, A.-L. Chiu, H.-H. Chang, Mining changes in customer behavior in retail marketing, Expert Systems with Applications 28 (4) (2005) 773–781.

[12] M.-S. Chen, J. Han, P.S. Yu, Data mining: an overview from a database perspective, IEEE Transactions on Knowledge and Data Engineering 8 (6) (1996) 866–883.

[13] Y.-L. Chen, K. Tang, R.-J. Shen, Y.-H. Hu, Market basket analysis in a multiple store environment, Decision Support Systems 40 (2) (2005) 339–354.

[14] R. Coase, J.L. Yellen, The problem of social cost, Journal of Law and Economics 3 (1) (1960) 1–44.

[15] G. Dalton, S. Gallagher, Online data's fine line, InformationWeek 27 (1999) 18–20.

[16] R.E. Dansby, C. Conrad, Commodity bundling, The American Economic Review 74 (2) (1984) 377–381.

[17] H. Demsetz, The cost of transacting, Quarterly Journal of Economics 82 (1968) 33–53.

[18] H. Estelami, Consumer savings in complementary product bundles, Journal of Marketing Theory and Practice 7 (3) (1999) 107–114.

[19] J.L. Gogan, The web's impact on selling techniques: historical perspective and early observations, International Journal of Electronic Commerce 1 (2) (1997) 89–108.

[20] J.P. Guiltinan, The price bundling of services: a normative framework, Journal of Marketing 51 (2) (1987) 74–85.

[21] P.J. Haynes, M.M. Helms, A.R. Casavant, Creating a value-added customer database: improving marketing management decisions, Marketing Intelligence & Planning 10 (1) (1992) 16–21.

[22] D.L. Hoffman, T.P. Novak, Marketing in hypermedia computer-mediated environments: conceptual foundations, Journal of Marketing 60 (3) (1996) 50–68.

[23] M. Houtsma, A. Swami, Set-oriented mining for association rules in relational databases, in: Proceedings of the 1995 International Conference on Data Engineering, Taipei, Taiwan, 1995.

[24] C. Janiszewski, M. Cunha Jr., The influence of price discount framing on the evaluation of a product bundle, Journal of Consumer Research 30 (4) (2004) 534–546.

[25] J. Jiao, Y. Zhang, Product portfolio identification based on association rule mining, Computer-Aided Design 37 (2) (2005) 149–172.

[26] C.-R. Julander, Basket analysis: a new way of analysing scanner data, International Journal of Retail & Distribution Management 20 (7) (1992) 10–18.

[27] M. Kitsuregawa, M. Toyoda, I. Pramudiono, Web community mining and web log mining: commodity cluster based execution, in: Proceedings of the 13th Australasian Conference on Database Technologies, Melbourne, Victoria, Australia, 2002.

[28] G. Koloszyc, Abandoned 'shopping carts' pose major challenge for internet retailers, Stores 81 (7) (1999) 41–44.

[29] P. Kotler, Marketing Management – Analysis, Planning, Implementation, and Control, Prentice-Hall, NJ, 1997.

[30] H. Lai, An object-oriented architecture for intelligent virtual receptionist, International Journal of Electronic Commerce 4 (3) (2000) 69–86.

[31] H. Lai, T.-C. Yang, A system architecture for intelligent browsing on the web, Decision Support Systems 28 (3) (2000) 219–239.

[32] H. Lai, T.-C. Yang, Applying shopping cart data to web customers clustering, in: Proceedings of the Western Decision Sciences Institute 30th Annual Meeting (WDSI 2001), Vancouver, Canada, 2001.

[33] H. Mannila, H. Toivonen, A.I. Verkamo, Efficient algorithms for discovering association rules, in: Proc. AAAI Workshop Knowledge Discovery in Databases (KDD'94), 1994.

[34] R.P. McAfee, J. McMillan, M.D. Whinston, Multiproduct monopoly, commodity bundling, and correlation of values, The Quarterly Journal of Economics 104 (2) (1989) 371–383.

[35] A. Mild, T. Reutterer, An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data, Journal of Retailing and Consumer Services 10 (3) (2003) 123–133.

[36] J.L. Munger, D. Grewal, The effects of alternative price promotional methods on consumers' product evaluations and purchase intentions, Journal of Product and Brand Management 10 (3) (2001) 185–197.

[37] A. Ovans, Make a bundle bundling, Harvard Business Review 75 (6) (1997) 18–20.

[38] T.R. Palfrey, Bundling decisions by a multiproduct monopolist with incomplete information, Econometrica 51 (2) (1983) 463–483.

[39] J.S. Park, M.-S. Chen, P.S. Yu., An effective hash-based algorithm for mining association rules, in: Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, San Jose, CA, 1995.

[40] P.R. Peacock, Data mining in marketing: Part I, Marketing Management 6 (4) (1998) 8–18.

[41] J. Reedy, S. Schullo, K. Zimmerman, Electronic Marketing: Integrating Electronic Resources into the Marketing Process, The Dryden Press, Harcourt College Publishers, TX, 2003.

[42] G.J. Russell, W.A. Kamakura, Modeling multiple category brand preference with household basket data, Journal of Retailing 73 (4) (1997) 439–461.
[43] G.J. Russell, A. Petersen, Analysis of cross category dependence in market basket selection, Journal of Retailing 76 (3) (2000) 367–392.
[44] C. Rygielski, J.-C. Wang, D.C. Yen, Data mining techniques for customer relationship management, Technology in Society 24 (4) (2002) 483–502.
[45] J. Strauss, A.I. El-Ansary, R. Frost, E-Marketing, Prentice Hall, NJ, 2003.
[46] S. Stremersch, G.J. Tellis, Strategic bundling of products and prices: a new synthesis for marketing, Journal of Marketing 66 (1) (2002) 55–72.
[47] M.S. Yadav, K.B. Monroe, How buyers perceive savings in a bundle price: an examination of a bundle's transaction value, Journal of Marketing Research 30 (3) (1993) 350–358.