International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:8, No:5, 2014

Associated Map and Inter-Purchase Time Model for Multiple-Category Products

Ching-I Chen

***Abstract*—**The continued rise of e-commerce is the main driver of the rapid growth of global online purchase. Consumers can nearly buy everything they want at one occasion through online shopping. The purchase behavior models which focus on single product category are insufficient to describe online shopping behavior. Therefore, analysis of multi-category purchase gets more and more popular. For example, market basket analysis explores customers' buying tendency of the association between product categories. The information derived from market basket analysis facilitates to make cross-selling strategies and product recommendation system.

To detect the association between different product categories, we use the market basket analysis with the multidimensional scaling technique to build an associated map which describes how likely multiple product categories are bought at the same time. Besides, we also build an inter-purchase time model for associated products to describe how likely a product will be bought after its associated product is bought. We classify inter-purchase time behaviors of multi-category products into nine types, and use a mixture regression model to integrate those behaviors under our assumptions of purchase sequences. Our sample data is from comScore which provides a panelist-label database that captures detailed browsing and buying behavior of internet users across the United States. Finding the inter-purchase time from books to movie is shorter than the inter-purchase time from movies to books. According to the model analysis and empirical results, this research finally proposes the applications and recommendations in the management.

***Keywords*—**Multiple-category purchase behavior, inter-purchase time, market basket analysis.

I. Introduction

DUE to the advance of technology and the fullness of customer database, it is possible for firms to execute database marketing programs for their customers. Data mining facilitates firms to well understand each customer's behavior, and provides useful information to cross-selling and product recommendation strategies, both of which are useful to increase customer value and loyalty to firms. Firms also can increase their profitability by successfully recommend their customers the right items at right time [1]. Therefore, nowadays, market basket analysis has become a trend.

Traditional market basket analysis only focuses the final the composition of shopping baskets, rather than the shopping sequences of items in a basket [2]. It also only consider the joint probability of two items bought at the same time, and ignores other factors resulting in the postpone buying behaviors. For example, after buying a printer, when to buy the compatible ink is determined by the customer's consumption rate of the complementary item of printers. To maximum the utility of a marketing strategy, firms may not only recommend a customer the right items, but also recommend them at the right time, derived from the accurate prediction of the customer's purchase cycle of complementary items. Therefore, the market basket analysis integrating with the inter-purchase time model generate more accurate prediction of customers' buying pattern.

Most of previous researches of inter-purchase time models focus on one single product rather than multi-category products. Their assumption of the behaviors of customers buying different products being independent of each other violets the market basket analysis's assumption of interdependence across multi-category product buying behaviors. Therefore, the inter-purchase time model considering only single product may generate biased prediction. Based on the above reason, this paper offers an inter-purchase time model for associated products. For academic purpose, our model can complete the previous models which only consider one single product or ignore the association among multi-category products. For practical purpose, our model provides more accurate information for marketing decision. There are two main research objectives listed as follows:

1) Use the traditional market basket analysis to detect the association between multi-category products. To find out how likely two items are bought at the same time, we use the interest measures in [3] which constitutes an associated matrix. Also, we use the multidimensional scaling technique to transform the associated matrix into distance of items in multidimensional space. The resulting associated maps which show the relative positioning of all items provide information which associated items are helpful for cross-selling strategy.
2) Derive an inter-purchase time model for multi-category products. We also determine the association between two items based on the expected length of inter-purchase time from one item to the other item. For this purpose, we offer a model of multi-category choice behavior integrated with inter-purchase time.

Our sample data is from comScore database which is a panelist-label database that captures detailed browsing and buying behavior of internet users across the United States. Data are analyzed according to the proposed model. The estimated mean inter-purchase time between different items are further processed to help predict purchase cycle and determine

C. I. Chen is with the Department of International Business Studies, National Chi Nan University, NantouHsien 545 Taiwan (phone: +886-49-2910960; e-mail: chingichen@ncnu.edu.tw).

C. I. Chen thanks the *National Science Council* of Taiwan (grants NSC102-2410-H-260 -049) for support.

association between items. The information is critical to make up-selling strategy and product recommendation.

II. Literature Review

A. Market Basket Analysis

Market basket analysis (MBA), also known as product affinity analysis, which explores association among multi-category choices, is an important issue in data mining literature [4]. Association rules which is mostly applied in market basket analysis includes three measures for evaluating of association between two products, such as indices for affinity measures, such as support index, confidence index, and confidence index [3]. Although the concept of market basket analysis is very popular, its underlying assumption of joint occurrence implying complementarity has been criticized [2]. Reference [5] indicates that the price cross-elasticity for most of pairs of products selected through association rules are in fact positive. In other words, most of the pairs of products which have higher measures of association rules are in fact substitutes determined by positive price cross-elasticity. The prediction of traditional market basket analysis seems not to be consistent with the observed purchase sequences [2].

Reference [2] presents the sequential market basket analysis which captures the observed purchase sequences. The empirical shopping trip data are collected by a RFID "tag" mounted under the shopping cart as in [6], producing a rich description of which items are put in the cart in sequence. For a pair of products, A and B, they use four kinds of frequency of all trips, including the number purchases of A→B sequence, B→A sequence, A alone, and B alone, to compute the conditional probability of the purchase sequence of P(A→B) and P(B→A), and marginal probability of P(A) and P(B). Then, using the definition of sequential probability, they define five types of pairs of A and B, including weak A→B sequence, strong A→B sequence, complements, substitutes, and independents. However, Kamakura's research only consider items bought in one trip rather than that customers may wait for a period of time to buy the items complement of the ones they bought in the last trip.

Firms can make a profitable sale by continuing up-selling the customers other add-ons or related items of the current bought item. It is critical for firms to accurately predict the purchase cycle of two related items, because only the up-selling at right time can effectively induce the customer to purchase the related items. In this paper, we build an inter-purchase model which captures the purchase cycle between a pair of associated items.

B. Inter-Purchase Time Model

Both of inter-purchase time and purchase frequency are measures of timing in a RFM model. The customer value evaluation which considers the relationship between purchase quantity and inter-purchase time can prevent from underestimation or overestimation [7]. If firms can accurately predict inter-purchase time of each customer, one-to-one sales promotion can be executed when the customer is most likely to buy items. Firms also can actively offer incentives to shorten inter-purchase time of the customers who have longer inter-purchase time.

Due to its positive value range, inter-purchase time is usually assumed to follow a Gamma distribution or exponential distribution. We model the inter-purchase time using a lognormal distribution, which is easily added in explanatory variables (e.g. marketing variables and customer variables) and formed a regression model.

Most previous research of multi-category choices uses purchase incidence model as in [8]-[10] rather than inter-purchase time model. Because the purchase incidence model only considers which items are bought on the same shopping trip, it ignores the sequence of multi-category choices. For example, how long after purchasing a printer until the customer gets its ink cartridge is the critical information for firms to execute an up-selling strategy. Therefore, an inter-purchase time model is more fit to modeling the sequence of multi-category choice than a purchase incidence model.

Only few researches discuss both of inter-purchase time and multi-category choice behaviors. Reference [11] classify the purchase choice of two items, A and B, on a shopping trip into four types, including a purchase made in both categories, made in Abut not in B, or vice versa, or neither of the categories purchased. They use inter-purchase time of two items, t_A and t_B, as two random variables, and four bivariate failure time models to construct the likelihood function. Reference [12] models the inter-purchase time to follow a generalized Gamma distribution, and its scale parameter is determined by the item purchased at the last shopping trip and the current ones. Both of the two studies only model the inter-purchase time between one item alone, rather between different items, such as $t_{A\to B}$ and $t_{B\to A}$, which more accurately describe the purchase sequences in different shopping trips.

III. Research Method

A. Association Rules

The most common practice in market basket analysis is the identification of association-rules [3]. We use the **interest** measure to evaluate whether each pair of products A and B tends to be complements or substitutes. The interest measure is defined as the ratio between the joint probability and the probability of joint occurrence under independence as follows:

$$Interest = \frac{Pr(A \cap B)}{Pr(A) \times Pr(B)} \tag{1}$$

However, the interest measure only focuses on the relationship between each pair of terms rather than among the whole items. For enhancing cross-selling efforts, it is helpful for firms to well understand the whole association structure and identify several subsets of complementary items. Therefore, we further use the multidimensional scaling technique to create an associated map displaying the relative positions of all items objects, given the associated matrix composed of interest measures of all pair of items.

B. Inter-Purchase Time between Associated Items

The inter-purchase time is defined as the time difference form the last shopping trip to the current shopping, illustrated in Fig. 1. The random variable t_{ij} represent the time difference from thej-1th purchase to the jth purchase for customer i, using 1 day as the time unit. Let A and B are two item categories. We assume that each customer can choose to buy A, B, or both of A and B in each shopping trip. Then, we define four inter-purchase time random variables, including t_{ij}(A→A), t_{ij}(B→B), t_{ij}(A→B), and t_{ij}(B→A), to capture the different combinations of the purchase sequences of two items.

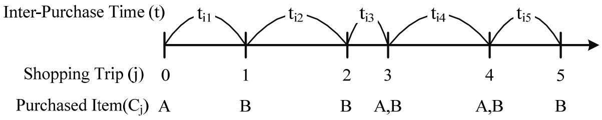

Fig. 1 Inter-purchase time and purchased items

There are nine types of purchase sequences of two items illustrated in Table I. Type 1 to type 4 indicate that only one item is bought at the (j-1)th and at the jth shopping trip; type 5 to type 8 indicate that both two items are bought either at the (j-1)th or at the jth shopping trip; type 9 means that both two items are bought at the both shopping trips. We develop a mixture inter-purchase time model which can integrate the nine situations.

TABLE I
NINE TYPES OF PURCHASE SEQUENCES OF TWO ITEMS

jth Trip (j-1)th Trip	A	B	A and B
A	(1) A→A	(3) A→B	(7) A→AB
B	(4) B→A	(2) B→B	(8) B→AB
A and B	(5) AB→A	(6) AB→A	(9) AB→AB

C. Assumptions of the Inter-Purchase Time Model

We assume that the inter-purchase time of each customer, t_{ij}, follows a lognormal distribution with explanatory variables. In other words, the logarithm of t_{ij} follows a normal distribution. First, we construct nine lognormal distributions of the inter-purchase time to model the nine types of purchase sequences illustrated as table. Second, we integrate the nine models into a mixture regression model as the final inter-purchase time model.

1) **Type1~Type4**: Only one item bought at each of two trips

Assume a customer only purchases one item both at the (j-1)th and the jth shopping trip. For example, in type 1 situation, the customer only purchases item Aat both of trips. We assume that the inter-purchase time between two consecutive shopping trips, tij (A→A), follows a lognormal distribution as follows:

$$lnt_{ij}(A\rightarrow A)\sim Normal(\mu_{AA},\ \sigma^2) \tag{2}$$

As the same logics, we can build the other three distributions of t_{ij}(B→B), t_{ij}(A→B), and t_{ij}(B→A).

2) **Type5~Type8**: Both two items bought either at the (j-1)th shopping trip or at the jth trip

We assume that choice of A and B bought at the same day can be classified into two situations, including buying B after having bought A or buying A after having bought B within one shopping trip. For example, the purchase sequence of type 5 is AB→A, can be further classified into two situations as follows:

a. $C_{j-1}=A\rightarrow B,\ C_j=A$

In the first situation, the customer is assumed to bought B after having bought A at the (j-1)th shopping trip within one day. Therefore, the inter-purchase time of this purchase sequence is t_{ij}(A→B)<1, i.e., lnt_{ij}(A→B)<0. Then the customer bought A at the jth shopping trip, and the inter-purchcase time is t_{ij}(B→A), the logarithmic value of which is assumed to follow a normal distribution, i.e., $lnt_{ij}(B\rightarrow A)\sim Normal(\mu_{BA},\ \sigma^2)$ 。 Therefore, lnt_{ij}(AB→A) is derived to follow a weighted normal distribution as follows:

$$\begin{aligned} lnt_{ij}(AB\rightarrow A) &\sim Pr(lnt_{ij}(A\rightarrow B)<0)\times Normal(\mu_{BA},\ \sigma^2) \\ &= \Phi\left(-\frac{\mu_0^{AB}}{\sigma}\right)\times Normal(\mu_{BA},\ \sigma^2) \end{aligned} \tag{3}$$

where Φ(.) is the standard normal cumulative distribution function.

b. $C_{j-1}=B\rightarrow A,\ C_j=A$

Following the logic,inthe second situation the customer is assumed to bought A after having bought B at the (j-1)th shopping trip within one day. Thenthe customer bought A at the jth shopping trip. In this situation, lnt_{ij}(AB→A) is derived to follow a weighted normal distribution as follows:

$$\begin{aligned} lnt_{ij}(AB\rightarrow A) &\sim Pr(lnt_{ij}(B\rightarrow A)<0)\times Normal(\mu_{AA},\sigma^2) \\ &= \Phi\left(-\frac{\mu_0^{BA}}{\sigma}\right)\times Normal(\mu_{AA}\sigma^2) \end{aligned} \tag{4}$$

Combining the two situations, a and b, the logarithmic inter-purchase time, lnt_{ij}(AB→A), follows a mixture normal distribution as follows:

$$\begin{aligned} lnt_{ij}(AB\rightarrow A) \sim\ & \Phi\left(-\frac{\mu_0^{AB}}{\sigma}\right)\times Normal(\mu_{BA},\sigma^2) \\ & + \Phi\left(-\frac{\mu_0^{BA}}{\sigma}\right)\times Normal(\mu_{AA,}\sigma^2) \end{aligned} \tag{5}$$

3) **Type 9**: Both two items bought at both the (j-1)th shopping trip and the jthtrip

This is the most complex purchase sequences. According to the above assumptions, the logarithmic inter-purchase time, lnt_{ij}(AB→AB), can be derived to follow a mixture normal distribution of type 1 to type 4. The purchase sequence "AB→AB" can be further classified into four situations, including:

a. $C_{j-1}=A\rightarrow B,\ C_j=A\rightarrow B$
b. $C_{j-1}=A\rightarrow B,\ C_j=B\rightarrow A$
c. $C_{j-1}=B\rightarrow A,\ C_j=A\rightarrow B$
d. $C_{j-1}=B\rightarrow A,\ C_j=B\rightarrow A$

The distributions of the above four situations is integrated as a mixure normal distribution as follows:

$$
\begin{aligned}
Int_{ij}(AB \rightarrow AB) \sim & \left[\Phi\left(-\frac{\mu_0^{AB}}{\sigma}\right)\right]^2 \times Normal(\mu_{BA},\sigma^2) \\
& + \Phi\left(-\frac{\mu_0^{AB}}{\sigma}\right)\Phi\left(-\frac{\mu_0^{BA}}{\sigma}\right) \times Normal(\mu_{BB},\sigma^2) \\
& + \Phi\left(-\frac{\mu_0^{AB}}{\sigma}\right)\Phi\left(-\frac{\mu_0^{BA}}{\sigma}\right) \times Normal(\mu_{AA,}\sigma^2) \\
& + \left[\Phi\left(-\frac{\mu_0^{BA}}{\sigma}\right)\right]^2 \times Normal(\mu_{AB},\sigma^2)
\end{aligned} \qquad (6)
$$

IV. Empirical Results

The results of the study are based on data of the transaction with Amazon.com from 2010/1/1 to 2011/12/31, obtained from comScore's worldwide database of 2 million people who have provided comScore with explicit permission to monitor their online behavior. Customers who have more than five transaction records are kept as the sample with 35,036 records. The top five best selling product categories are listed as the Table II.

TABLE II
Top Ten Product Categories on Amazon

Rank	Category	Records	Rank	Category	Records
1	Books & Magazines	11,403	6	Toys & Games	2,211
2	Movies & Videos	6,529	7	Apparel	2,075
3	Online Content Sales	3,011	8	Electronics & Supplies	2,075
4	Health & Beauty	2,632	9	Food & Beverage	1,409
5	Music	2,444	10	Art & Collectibles	1,247

A. Associated Map

For the ten product categories, there are 45 pairs of items. Therefore, we can compute 45 interest measures which compose an associated matrix illustrated as Table III. The matrix shows that "Health & Beauty" has the strongest complementary relationship with "Food & Beverage" are the strongest compliments. "Art & Collectibles" also have interest measures which are larger than one with "Toys & Games", "Apparel", and "Food & Beverage".

TABLE III
Association matrix of Ten Product Categories

	P1	P2	P3	P4	P5	P6	P7	P8	P9
P2	0.691								
P3	0.118	0.146							
P4	0.318	0.455	0.079						
P5	0.664	1.356	0.494	0.417					
P6	0.502	0.813	0.082	0.55	0.719				
P7	0.481	0.721	0.136	1.002	0.733	1.037			
P8	0.214	0.309	0.109	0.636	0.356	0.423	0.61		
P9	0.509	0.732	0.132	1.795	0.78	0.875	1.248	0.604	
P10	0.609	1.041	0.129	0.922	1.017	1.466	1.497	0.692	1.475

We further use multidimensional scaling technique to visualize the associated matrix. The created associated map is shown as Fig. 2. It shows that "Online Content Sales" and "Electronics & Supplies" are independent or substitute items with other eight products because they are located far away. The remaining eight product categories can be classified into three groups according to their nearness with each other. For example, if one customer purchases "Book & Magazines", the best recommending items are "Movies & Videos" and "Music".

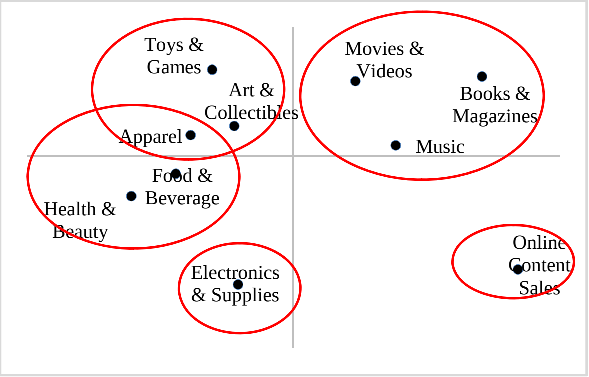

Fig. 2 Associated map of ten product categories

B. Mixture Inter-Purchase Time Model

Our mixture inter-purchase time model can fit the inter-purchase time between two shopping trip at which customers may choose to buy one or two items. The inter-purchase time models provided by the previous studies only can fit the inter-purchase time between same items such as type 1 and type 2, or between different items such as type 3 and type 4. We further discuss how inter-purchase time changes with different multi-category choices at two concessive shopping trips such as type 5 to type 9.

TABLE IV
Estimation of Mixture Inter-Purchase Time Model

Purchase Sequences	Model 1		Model 2	
	coefficient	p-value	coefficient	p-value
A→A	2.368	0.000^{***}	2.357	0.000^{***}
B→B	2.493	0.000^{***}	2.493	0.000^{***}
A→B	2.678	0.000^{***}	2.606	0.000^{***}
B→A	2.685	0.000^{***}	2.652	0.000^{***}
F-value	2000.68	0.000^{***}	2654.1	0.000^{***}

Taking "Books & Magazines" and "Movies & Videos"for example, the estimated regression coefficients by our mixture regression model as illustrated in Table IV. Model 1 only consider the inter-purchase time data which match type 1 to type 4 purchase sequences. The empirical results indicate that the average logarithmic inter-purchase between two same items bought such as $\hat{\beta}_{A\rightarrow A}$=2.368 and $\hat{\beta}_{B\rightarrow B}$=2.493, are both smaller than the ones between different items bought such as $\hat{\beta}_{A\rightarrow B}$=2.678 and $\hat{\beta}_{B\rightarrow A}$=2.685. The estimated values imply that there is little difference between the inter-purchase time from buying A to buying B and ones from buying B to buying A.

Then, Model 2 considers all the inter-purchase time data. The estimated regression coefficients $\hat{\beta}_{A\rightarrow A}$ =2.357 and $\hat{\beta}_{B\rightarrow B}$=2.493 are almost the same as model 1. However, the estimated inter-purchase time values between buying different items such as $\hat{\beta}_{A\to B}$=2.606 and $\hat{\beta}_{B\to A}$=2.652 are smaller than ones of model 1, showing more complementary to each other. The results of $\hat{\beta}_{A\to B}<\hat{\beta}_{B\to A}$ indicates that buying "Movies & Videos" after "Books & Magazines" shows more complementary than vice versa.

V. Conclusions

This study develops an inter-purchase time model for associated products which have more strategic implication than that for one single product. The empirical results show that when the transaction records of both items bought are not considered, the inter-purchase time between two concessive shopping trips of buying same items (A→A or B→B) is shorter than buying different items (A→B or B→A). When we use our mixture regression model to fit the whole data, the estimated inter-purchase time from buying "Books & Magazines" to buying "Movies & Videos" is shorter than vice versa. It implies that firms may advance the recommendation time of movies & videos to the customers who have bought books & magazines.

The advantage of our model is that it can fit the multi-category choices in one shopping trip. We also use demographic data as explanatory variables to estimate its effects on inter-purchase time. The empirical results can help firms to predict the purchase behaviors of new customers who have no transaction records with the firms.

References

[1] H. J. Lee, J. W. Kim, and S. J. Park, 2007. "Understanding collaborative filtering parameters for personalized recommendations in e-commerce," *Elect. Com. Res.*, vol. 7, no. 3, pp. 293–314, Oct. 2007.

[2] W. A. Kamakura, "Sequential market basket analysis," *Market. Lett.*, vol. 23, no. 3, pp. 505–516, Sep. 2012.

[3] T. Brijs, G. Swinnen, K. Vanhoof, and G.Wets, "Building an association rules framework to improve product assortment decisions," *Data Min. Knowl. Disc.*, vol. 8, no. 1, pp. 7–23, Jan. 2004

[4] S. S. Anand, A. R. Patrick, J. G. Hughes, and D. A. Bell, "A data mining methodology for cross-sales,"*Knowl-Based Syst.*, vol. 10, no. 7, pp. 449–461, May 1998.

[5] B. Vindevogel, D. Van den Poel, and G. Wets, "Why promotion strategies on market basket analysis do not work," *Expert Syst. Appl.*, vol. 28, no. 3, pp. 583–590. Apr. 2005.

[6] H. Sorensen, "The science of shopping," *Market. Res.*, vol. 15, no. 3, pp. 30–35, Sep. 2003

[7] L. Jen, C. H. Chou, and G. M. Allenby, "A bayesian approach to modeling purchase frequency," *Market. Lett.*,vol. 14, no. 1, pp. 5–20, Jan. 2003.

[8] P. Manchanda, A. Ansari, and S. Gupta, "The "shopping basket": a model for multicategory purchase incidence decisions," *Market. Sci.*, vol. 18, no. 2, pp. 95–114. May 1999.

[9] S. Chib, P. B. Seetharaman, and A. Strijnev, "Analysis of multi-category purchase incidence decisions using IRI market basket data," *Adv. Econom.*, vol. 16, no. 1, pp. 55–90, Jan. 2002.

[10] R. Niraj, V. Padmanabhan, and P. B. Seetharaman, "A cross-category model of households' incidence and quantity decisions," *Market. Sci.*, vol. 27, no. 2, pp. 225–235,310, Mar. 2008.

[11] P. K. Chintagunta and S. Haldar, "Investigating purchase timing behavior in two related product categories," *J Market. Res.*, vol. 35, no. 1, pp. 43–53, Feb. 1998.

[12] R. S. Guo, "A multi-category inter-purchase time model based on hierarchical Bayesian theory," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6301–6308, Apr. 2009.