經營科學
第34卷 第2號
2017年 6月, pp.15-33
https://doi.org/10.7737/KMSR.2017.34.2.015

빅데이터를 통한 대형할인매장 촉진활동 전략 분석 : 베이지언 네트워크기법 응용을 중심으로*

김 범 수†
서강대학교 경영학부

Developing an Efficient Promotion Strategy for a Multi-Product Retail Store : A Bayesian Network Application

Bumsoo Kim†
Sogang Business School, Sogang University

■ Abstract ■

This paper considers a Bayesian Network analysis for understanding the heterogeneous cross-category effects of different promotion activities and developing an efficient overall promotion strategy for a large retail store. More specifically we differentiate price reduction promotion and floor promotion and study their heterogeneous effect on consumer purchase behavior under a market basket setting. We then utilize Bayesian networks in identifying complex association structure in market basket dataset by analyzing the effects of different promotional activities and also include the effects of time, family income and size. We find from our Bayesian network analysis that the dominant cross-category promotion effect of price promotion is the indirect effect whereas the dominant cross-category promotion effect of floor promotion is the direct effect. Also, among the demographic variables we find that family size of the household is linked with more product categories compared to income and see that there are differences in the extent of the effects by product category. Finally, we also show the existence of products acting as a network hub and how they can be utilized by retailers faced with a limited marketing budget and suggest a more efficient promotion strategy.

Keywords : Bayesian Network, Market Basket Analysis, Big-Data, MCMC Method, Cross-Category Promotion, Consumer Behavior

논문접수일 : 2017년 04월 09일 논문게재확정일 : 2017년 04월 14일
* This paper was supported by the Sogang University Research Grant of 201410039.01.
† 교신저자, bk4498@sogang.ac.kr

1. Introduction

Utilizing consumer's market basket (scanner) data to extract and understand consumer behavior, has been a problem which intrigued researchers in both marketing and data mining for a long time. One line of research is on modelling consumer's purchase decisions in terms of product categories and brand choice and how they are affected by marketing mix variables. Another line of research in this area focuses on modelling store choice problem and finding ways to increase customer loyalty. Finally, there are works which study the timing aspect of consumer purchase behavior such as time between purchases. In this study, we develop Bayesian networks to understand different aspects of consumer's multi-category purchase decisions and the effects of marketing-mix variables using market basket data.

The main problem within the context of consumer's choice behavior under multi-category setting is the complicated nature of interdependent structure between product categories. One may find number of products appearing in a given basket of goods because they are related *complimentary* products (such as beer and chips), have similar purchase cycle (such as beer and diaper) or by pure coincidence. Identifying the existence and the nature of these relationships becomes important, as marketing activity on one product will not only affect consumer's purchase decision on that product but on other dependent products as well. Most widely used models in regards to the multi-category choice problem is the *multivariate probit* model (MVP), as seen in Manchanda et al. [21] and Chib et al. [4]. These studies incorporate the *coincidence* and *complementary* effect in defining individual's indirect utility function and discover that significant cross-category effect exist in both price and promotion terms. There is also the multivariate logit model (MNL) suggested by Russell and Petersen [25], which also looks at the utility of individual households based on the extension of the multinomial logit model. However, the proposed theoretical models are limited in the sense that they look at bivariate cross-category relationships which may lead to biased results. Chib et al. [4] for instance, runs the MVP model simultaneously in a consumer choice setting with twelve product categories and finds that bivariate model is biased in both estimating the cross-category correlations and the effects of cross-category promotion effects. Note, that even Chib's model does not take into account the full interdependent structure among twelve product categories simultaneously but only looks at the relationship between *two* categories at a time. Hence, a more accurate picture of the full cross-category dependence structure may be found from incorporates the data-mining approach. These data mining approaches include structure construction through association rules found in Agrawal et al. [1], general network from Choi and Kim [7] and Kim et al. [19], growing neural networks from Decker [9] and Jung and Lee [17], and Bayes networks models proposed by Giudici and Passerone [14]. These approaches look at the joint dependent structure over the entire product categories and show that changes in consumer's choice in a category may affect not one but many different categories.

Also previous literature shows limitations in regards to incorporating time aspect into the general model. Chintagunta et al. [5] proposes the *conditional hazard function* approach and Seetharaman and Chintagunta [27] proposes the *proportional hazard model* (PHM) to model the purchase timing problem. Specifically, these models estimate the inter purchase times of a product using the hazard functions. However, these do not consider multi-category settings or look at the possible temporal effects of the marketing mix variables.

In this study, our objective is to develop models based on Bayesian network methods to address some of the problems not yet addressed in the marketing literature in regards to the market basket analysis. Specifically, we explore the effects of marketing-mix variables (such as coupons, display etc), demographic heterogeneity of households and seasonality on the joint association structure. Our approach is based on *Bayesian network method* of Heckerman et al. [15] and Markov Chain Monte Carlo model composition algorithm (MC^3) of Madigan and York [20], which was proposed to graphically model joint structures in discrete data. We show how our approach can potentially increase the *efficiency* in marketing activities by describing a more *accurate* joint association structure. In so doing, we extend the work of Giudici and Passerone [14] by incorporating marketing-mix, seasonality and heterogeneity effects in the model. One of the attractive feature of our approach is that we can handle the joint probability structure of the full model whereas classical methods find it computationally burdensome when the number of categories become large

Rest of the paper is organized as follows. First we present the basic model studied in the literature regarding cross-category association and promotion effects in section 2. In section 3, we briefly define and illustrate the Bayes network model for our problem and the implementation algorithm. We then present our finding in section 4 and conclude the paper with limitations and possible future research in section 5.

2. Literature Review

Utility based consumer choice models are among the most widely used methods to explain consumer choice behavior in both single category and multi-category setting. Here, researchers argue that the consumer's choice is based on some type of latent utility function related to each category, and model accordingly. In the multi-category setting, Manchanda et al. [21] model these latent utilities as a linear function of base utility from the product itself, utility arising from the direct marketing activity on the same product and utility from *cross-effects*. Specifically, *cross-effect* is defined as the changes in overall purchase utility from marketing activities of other categories. Manchanda et al. [21] also adds a random error vector which follows a multivariate normal distribution to capture *coincidences* in the shopping basket, and utilizes the multivariate probit model of Chib and Greenberg [3] to solve the problem. Russell and Petersen [25] defines the conditional choice models in determining the probability of purchasing in the final category given the choice decisions on other categories and solve them by utilizing the multivariate logistic distribution of Cox [8].

Another approach in determining meaningful relationships within a market basket dataset comes from the data mining literature. One of the first and most influential work in the data mining literature is the seminal work by Agrawal et al. [1] where they devise a comprehensive algorithm to mine association rules in a large dataset. Association rules, for example, may be statements as such : 75% of the baskets which contain chocolate and coffee also contains bananas, and as one can see there are multiple ways to utilize these rules in practice. More recently, there are also methods which uses artificial neural networks such as the works of Won [29]. Specifically, Decker [9] uses a dynamic algorithm is defined to identify and connect the nodes (market basket) similar to each other.

There are however, certain limitations associated with these approaches. Namely, in utility based consumer choice models, parameter estimation becomes extremely tedious and in many cases just infeasible when the number of product categories become large. One can see from previous works that at most these models look at bivariate relationships simultaneously. If the limitations in utility based models are computational issues, then the limitations we find in the data mining literature are more conceptual. Association rules and neural networks all work well computationally and are able to find those cluster of product categories which appear together in a basket, but fail in linking the observed patterns to consumer profiles or other exogenous variables (such as marketing mix variables) which may be the underlying reason behind such patterns.

With advances in computing and simulation algorithms, Bayesian belief networks pioneered by the works of Friedman et al. [11] and Heckerman et al. [12] first received great attention in general classification problems and is used in various fields such as environment systems modelling [6] and supply network risk analysis [13]. Bayesian network is a graphical representation of the probabilistic structure among a set of variables where variables themselves are represented as nodes of the graph and directed arrows between the nodes representing the probabilistic structure. Hence, it is only natural to utilize Bayes network in market basket analysis problem for the nodes can represent the product categories and the directed arrows can show the conditional probabilistic structure between the product categories present in the shopping basket. Another advantage in Bayesian networks is that the measure and direction of directed graph represents the conditional probabilities, and hence one can make probabilistic statements. Also with Bayesian network, one can also calculate the posterior probability of the network itself and hence the user has a clear measure to compare different candidate association structures.

Giudici and Passerone [14] previously utilized Bayesian network methods in regards to the market basket problem. Giudici and Passerone [14], mainly focuses on the strengths of the Bayesian network by comparing Bayes network methods to classical graphical methods. They find that In contrast to the classical methods, Bayesian network allows the user to directly analyze the full contingency table of the possible categories/nodes and can be used on very small and sparse datasets. They also introduces some marketing mix and demographic variables in the model, but other than the general form of the network structure, no quantification of the promotion effects were made. Therefore in this study, we extend previous literature and the use of Bayesian networks in identifying complex association structure in market basket dataset by analyzing the effects of different promotional activities within the Bayesian network framework and also include the effects of time by incorporating time-related variable.

3. Bayesian Network Model for Cross-Category Promotion

3.1 Bayes Network

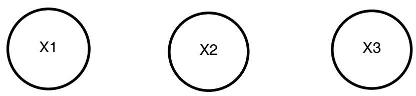
First we give a brief introduction of the Bayes network. Generally, a Bayes network, $N = (X, G, P)$ over a set of variables X, consists of an acyclic directed graph (DAG) $G = (V, E)$, where V is a set of nodes and E is a set of directed edges, and a set of conditional probabilities P. Here, a DAG is further defined to be a graph with directed edges between node υ without any directed cycles. Note there exist one-to-one to relationships between each nodes of the graph υ and mutually exclusive finite discrete random variables $X_v \in X$. Also given the network, the DAG specifies the nature of conditional and mutual independence structure between random variables with the directed edges E. Following is a more formal definition of the Bayes network, $N = (X, G, P)$ by Jensen [16].

Definition 1 : A (discrete) Bayes Network $N = (X, G, P)$ consists of

- a *DAG*; $G = (V, E)$ with nodes $V = (v_1, v_2, \cdots, v_n)$ and directed links E
- a set of discrete random variables, X, represented by the nodes V of the graph G
- a set of conditional probability distributions, $\boldsymbol{P}$, containing one distribution, $P(X_v | X_{pa(v)})$ for each random variable $X_v \in X$.

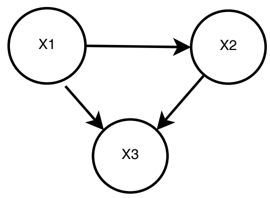
Given a clear description of the random variables, the most critical components of the Bayes network are the DAG, which shows the conditional independence structure between variables, and the corresponding set of conditional probability distributions. Furthermore, utilizing Bayesian principles, a Bayes network gives a joint probability distribution over a set of random variables X, meaning that the set of conditional probability distribution $\boldsymbol{P}$ specifies a multiplicative factorization of the joint distribution by the chain rule. The following simple example is a demonstration of the use of Bayes network in modelling different probabilistic structures.

Example 1 : Let $X = (X_1, X_2, X_3)$ be a set of random variables and assume also that the conditional probabilities are defined, then the following are some possible joint probability structures.



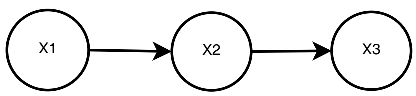
〈Figure 1〉 Mutual Independence

<Figure 1> shows no connections, meaning that no node has a parent or a child, and the joint probability is given by : $P(X_1, X_2, X_3) = P(X_1)P(X_2)P(X_3)$



〈Figure 2〉 Complete Dependence

In <Figure 2>, we see that X_3 has two parents in X_2, X_1 and that X_2 has one parent in X_1 and hence the joint probability is given by : $P(X_1, X_2, X_3) = P(X_3 | X_2, X_1)P(X_2 | X_1)P(X_1)$



〈Figure 3〉 Conditional Independence

<Figure 3> represents a distinctive conditional independence structure. Specifically, X_3 has one parent in X_2 and X_2 has one parent in X_1. This in turn implies that X_3 and X_1 are conditionally inde-

pendent given X_2. Therefore, the joint probability is given by : $P(X_1, X_2, X_3) = P(X_3|X_2)P(X_2|X_1)P(X_1)$

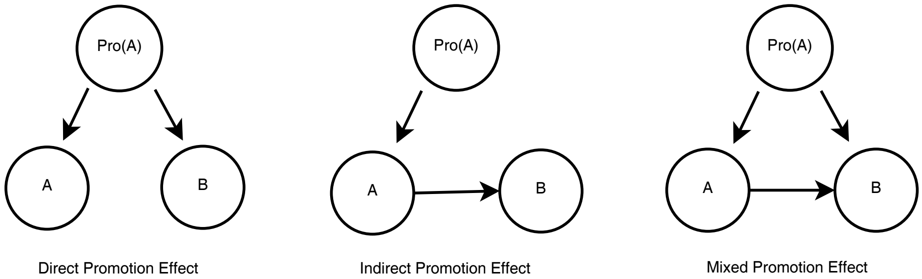
In sum, a probabilistic model with interactions between a set of random variables may be represented as a joint probability for inference and prediction purposes. In regards to the market basket problem it is only natural to utilize a Bayes network, which is a more compact and clear way of representing the joint probability structure of a large set of discrete random variables.

3.2 Modelling Cross-Category Promotion Effect with Bayes Network

In this subsection, we aim to introduce the ideas in marketing which relate the effects of exogenous variables such as marketing-mix and customer demographic variables on the sales or the choice of products and describe how we plan to utilize the Bayes network methods to model them.

We will first consider the effects of marketing-mix variables and more specifically explore different models which explain the cross-category promotion effects. One of the most documented operational strategies which could be used by retailers to induce an increase in sales of a specific product, is that of product promotion. Blattberg et al. [2] give a thorough review of the promotion literature and there they define promotion activities as not only price promotions (*temporary* price discounts to the customer) but also include distribution of coupons, display and feature advertising activity direct to the consumer. Specifically, the study highlights that promotion activities can either change the magnitude of consumer's purchases and/or enhance store traffic. Another important study regarding promotion strategy is done by Mulhern and Leone [24], where they emphasize the presence of demand interrelationships among multiple products within a store. As we have discussed, the underlying problem is a multi-category problem and single product promotion strategy and implications of simple buyer-seller models have limited value in devising promotion strategies. Therefore, we will look at the promotion problem in a multi-product setting identifying and analyzing interdependent structure between products.

Throughout the marketing literature, researchers agree that promotion activities in one category have impacts on the sale of products within that category, but they differ in modelling cross-category promotion effects depending on the nature of the underlying relationship between categories. All in all, there are three modelling paradigms related to cross-category promotion effects. Direct promotion effect proposed by Manchanda et al. [21] can be described as promotion activities on category A directly affecting the purchasing behavior not only on category A but also category B. The second paradigm, "indirect promotion effect" described by Russel et al. [26], is where promotion activities in a category A only affect purchasing decisions in this category but the changes in purchasing patterns in category A affects the purchasing behavior in category B. Therefore promotion activities in category A indirectly affect purchasing decisions in category B. Finally there is the mixed promotion model where both direct and indirect promotion effects are present across product categories. Simple Bayes network representations of these different promotion effects are illustrated in <Figure 3> below.



<Figure 3> Bayes Network Representation of Cross-Category Promotion

The other important dependence structure between categories is that of a *coincidence* relation. For example, consumers may purchase different items in a shopping trip due to habit [18], mood [10], store layout [28] and similar product cycle or due to seasonal effects. Finally, consumer heterogeneity may also be an underlying reason behind certain joint purchase patterns. Although these factors are usually not controllable by the retailer, if data (such as consumer demographics) are available then one may investigate the true nature of the association between categories. Distinguishing between different dependence structures is fundamentally important, as different promotion strategy may be required to target different aspects of the relationships.

How one models cross-category promotion is entirely up to the researcher. If there exists some prior knowledge on product dependency structure then one can specify these structures a priori, where they could be direct, indirect or mixed promotion models. If there is no past data or expert opinion on the matter then one can also follow a pure data mining approach and try to find the best cross-category dependence structure from scratch or by starting with a uniform prior on possible network structures. In any case, we will use the Bayesian network as (i) it can handle all three promotion models described above, (ii) it is easy to represent and model the dependence structure between categories and (iii) it can be used to learn the structure when there is no prior structural knowledge available. In addition to the promotion variables, we propose to include seasonal indicator variables and consumer demographic variables to see how the association structure changes with respect to these variables.

3.3 Data and Model Algorithm

The market basket dataset we will use for our model include weekly sales data of multiple product categories from six grocery stores in the Chicago Metropolitan area, where 500 households were followed for a 2 year period. The dataset also includes store's promotion activity over the time period and customer demographic data. Now, recall that the Bayesian network is graphical model which consists of a DAG, a set of variables and a set of conditional probabilities corresponding to the edges between variables. Hence, in <Table 1> we define the necessary ingredients for the Bayesian network by discretizing the raw data and create the following variables for each k.

〈Table 1〉 Definition of Exogenous Variables Used in the Model

Variable	Definition
P_{ik}	1 if the product in category i is promoted in transaction k; 0 otherwise
I_k	1 if the household in transaction k has income greater than the median; 0 otherwise
F_k	1 if the household in transaction k has more than 1 member; 0 otherwise
S_{mk}	12 monthly binary variable which takes the value 1 if the transaction k occurred in that given month; 0 otherwise
X_{ik}	0 quantity of product in category i in transaction k is 0 1 quantity of product in category i in transaction k is below the median level and > 0 2 quantity of product in category i in transaction k is above the median level

Note that we do not need to estimate the probabilities related to P_{ik}, I_k, F_k and S_{mk} variables as they are exogenous variables known by the retailer. This information also allows us to know that nodes corresponding to P_{ik}, I_k, F_k and S_{mk} variables do not have parent nodes as promotion activity run by the stores were not targeted on any demographic groups or seasons. Also we assume that these exogenous variables are independent of each other and therefore eliminate the possibility of edges between them. Finally, from previous marketing literature it is reasonable to assume that the exogenous variables affect product purchase and not vice versa. Hence, when a link is identified between product purchase variable and exogenous variables we fix the direction of the link to be one-way (i.e. from exogenous variables to the purchase variable).

For estimation of the network, we assign *Dirichlet* priors for cell probabilities and *Uniform* priors for initial model likelihoods and utilize the MC^3 model selection algorithm of Madigan and York [20]. Our work is a direct and logical extension of Giudici and Passerone [14], where the focus was to identify the product clusters only. Once these clusters are identified, it can provide the retail managers with valuable information on planning some type of targeted marketing and operational strategies for different sections of consumers and at different times over the course of the year. The following algorithm outlines the steps we take in the study.

Network Search Algorithm

1. Initialization : Set initial network model M_1
2. For every iteration l = 1 to L+S-1
   - Sample u~$U_{[0,1]}$
   - Sample M'~Q(· | M_l)
   - if u < R then $M_{l+1} = M'$
   - else $M_{l+1} = M_l$
3. After burn in period of L, collect the sample of network models $\{M_L, M_{L+1}, \cdots, M_{L+S}\}$ checking the convergence of the MC chain
4. Set a desired level of marginal posterior probabilities for network edges to a given level α
5. Identify all edges E_{ij} that satisfies $P(E_{ij}|D) \geq \alpha$
6. Create the Bayes network with all identified edges in step 5
7. Updated the parameters given the final network in step 6

4. Results

After sorting the raw market basket dataset to conform to our setting, we find a total of 61567 transaction baskets each with 58 variables, and we break them down to a training dataset with 56567 baskets and a testing dataset with 5000 baskets. In more detail, these 58 variables consist of 22 product purchase variables, 22 promotion variables, 12 monthly indicator variables, an income variable and a family size variable. For this analysis we use SAS and R software on a personal computer with INTEL(R) i7-2600 CPU 3.40Ghz processor and 16GB RAM memory.

There are many validation techniques and metrics in the literature which evaluate model complexity, prediction performance, uncertainty and edge direction. Marcot [22] gives an excellent review on set of existing and new measures for evaluating Bayesian network model performance and uncertainty. In this study we use a validation statistic which is based on the similarities of edges and the direction of edges generated from the dataset. After randomly splitting the training dataset to two datasets D_1 and D_2, we computed the average network structure via the MC^3 algorithm, denoted N_1 and N_2. Now, by defining the set of edges present in N_1 and N_2 as E_1 and E_2, we compute the ratio of common edges in both networks $r = \frac{E_C}{E_U}$, where $E_C = (E_1 \cap E_2)$ and $E_U = (E_1 \cup E_2)$. After repeating the validation process 20 times for each α levels, we present the result in <Table 2>.

<Table 2> Model Validation Results

	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 0.9$
r	0.84	0.90	0.93

We can see that the ratio of common edges in independent networks created through the algorithm is relatively high, meaning that there are common patterns identified from our dataset. By reviewing these validation statistics, we can reasonably conclude that the average network structures generated from the MC^3 algorithm are consistent networks.

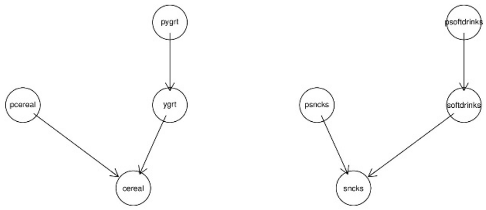
In determining the edge direction we follow the posterior probability approach of Margaritis [23] and Gamez et al. [12]. For example, when both direction is possible between nodes we choose the direction where the posterior probability, calculated from MCMC generated sample of networks, is higher.

4.1 Cross Category Promotion Effect of Coupons

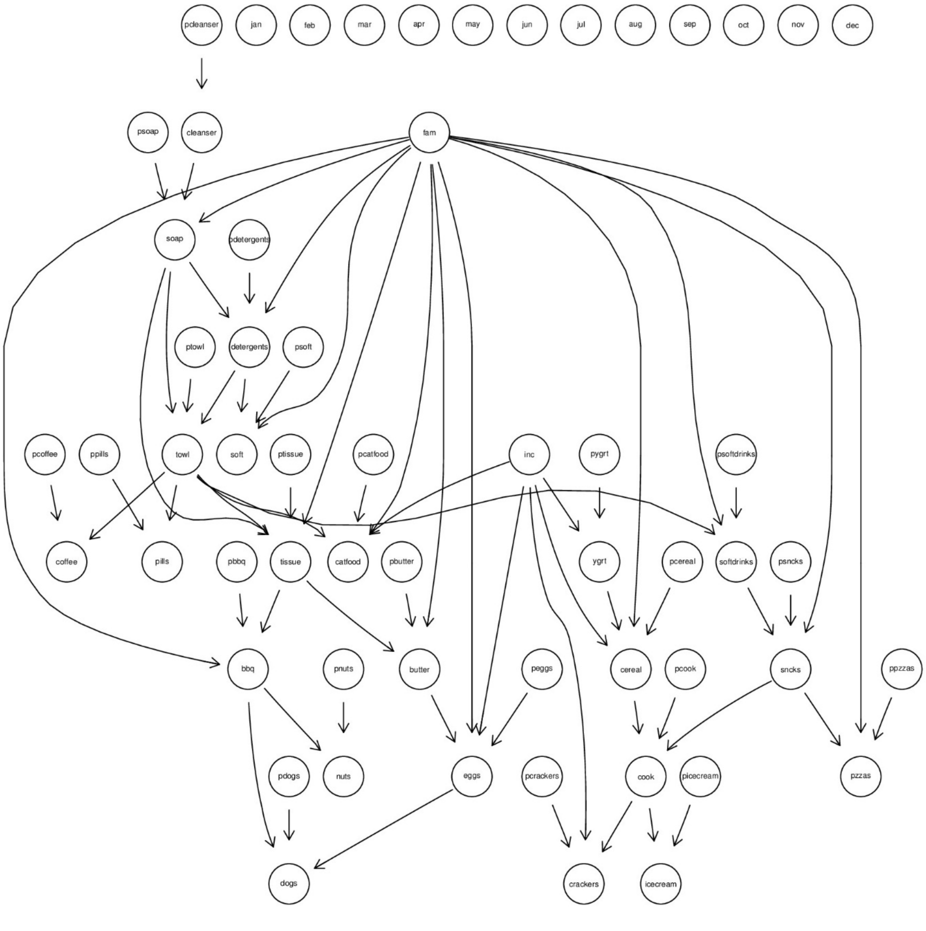
First, we only consider coupons as promotion variable and generated the Bayes network with posterior edge probability $\alpha = 0.75$. The resulting network N_1 in <Figure 4> shows many interesting characteristics. The most obvious pattern that we observe is that all of the monthly indicator variables are not connected with the rest of the network, meaning that the probability of a certain product being present in a basket is not affected by the month or time of the transaction. Considering the fact that the products considered here in the model are *staples*, or everyday goods, it is reasonable to conclude that time is not a significant exogenous factor in consumer's purchase decision.

Another key feature of the network we need to address is that there are no cross-category edges between a promotion node and product nodes. All 22 promotion nodes are directly connected to the product nodes of the same product category, meaning that there are no *direct* cross-category promotion occurring here. Note that in this Bayesian network model we only considered coupons as means for promotion. Coupons are usually used as a target promotion tool, as they are direct price discounts for a specific product and this may be the reason behind no cross-category connections. However, this does not mean that there is no cross-category promotion effect as *indirect* cross-category promotion effect may occur through the links between product nodes themselves.

For example, we see from <Figure 5> below that coupons on yogurt, cereal, snacks and softdrinks increase the conditional probability of these products being included in a consumer's basket. In addition, note that yogurt being present in the basket increases the conditional probability of cereal being present in the basket as well. Hence, one can conclude that coupons on yogurt increase the conditional probability of cereal being present in the consumer's basket by affecting the conditional probability of yogurt being included in the basket. Similar analysis can be applied to the product cluster of softdrinks and snacks.



〈Figure 5〉 Examples of Cross-Category Effect of Coupons Identified in Bayes Network N_1



〈Figure 4〉 Bayes Network (N_1) Detailing the Cross-Category Promotion Effect of Coupons

One can also identify certain products acting as a *hub* within the network. The node representing the paper towl is linked with 5 different product nodes, i.e. coffee, vitamin pills, tissue, catfood and softdrinks. We also see that soap is linked with 3 different product nodes. From these product *hubs*, retailers can identify key products where they can focus their promotion effort. Assuming that the cost of promotion through coupons is constant across products and that there is a limited marketing budget, it would be profitable for the retailers to focus their efforts on these key products as the presence of these key products affect the conditional probability of more *secondary* products being present simultaneously.

Finally, we can see the different effects of demographic variables, income and family size, on the conditional probabilities of the products being included in the consumer's shopping basket. What we see is that family size of the household affects more products in terms of purchase decisions compared to the household income. However, remember that we are mostly dealing with everyday product category and therefore, bigger families are more likely to purchase greater quantities of these products more frequently. Furthermore, in our analysis we are looking at product categories and not specific brands within a category. Given a certain product category, there are numerous brands of the product where some are perceived as luxury brands with higher prices. It would certainly be interesting to see if there are any income effects on the purchase decisions of these luxury brands compared to the generic brands in future studies. It is also worth noting that all products linked with the income variable are food products or consumable goods which cannot be stocked for future use.

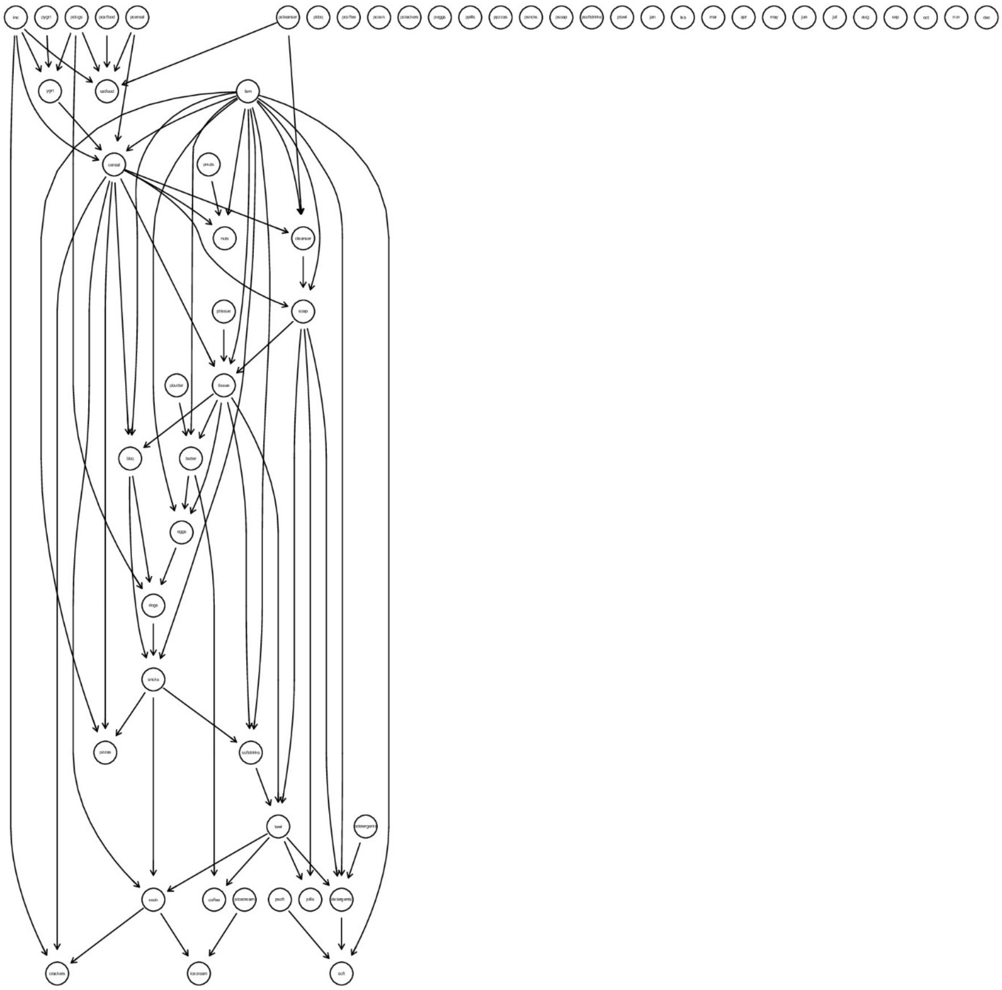
To sum up, we found from the full Bayesian network analysis of the cross-category promotion effect of coupons that indirect cross-category promotion is the dominant cross-category effect we find in the multi-category setting. Due to the nature of the product categories we consider, we found that time does not affect purchase decisions of consumers when considered as a exogenous variable. Also, among the demographic variables we consider family size of the household is linked with more product categories compared to income. Finally, we also showed the existence of products acting as a network hub and how they can be utilized by retailers faced with a limited marketing budget.

4.2 Cross Category Promotion Effect of Display and Feature Promotion

We now study the effect of display and feature promotion by generating the Bayes network with posterior edge probability $\alpha = 0.75$. The resulting network N_2 in <Figure 6> shows some similarities in its structure compared to the previous network N_1, however there also exist some important differences.

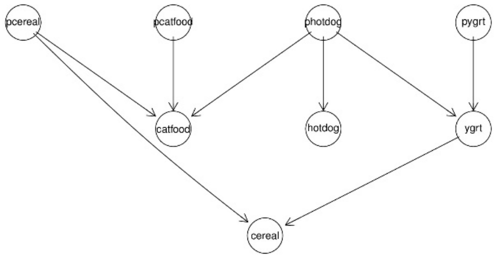
Regardless of different promotion activities, we find that there exist certain product categories which work as a network hub as before. Products such as paper towl and tissue reappear as those which has a link with many product categories. Also, as before, family size of the household is linked with more product categories compared to household income. Note that the product categories we consider in the model have not changed and we are still mainly dealing with everyday products and there is no reason to believe that the marginal effect of demographic variables on purchase behavior changes with different promotion strategy. Finally, we see that the monthly indicator nodes are still disconnected from the main network and hence, we can re-confirm that there are no exogenous time effects on consumer purchase behavior. Again, the underlying reason for this is the composition of the products considered in the model, however the interpretation of the result may give some managerial insight to the retailers. To be more specific, remember that we are looking at promotion here as promoting a product as a special feature item or putting them on a promotional display.

The result here shows that in case of everyday goods that we consider, devising a promotion strategy based on time is ineffective. The results regarding time variable may change if we look at seasonal products, such as certain fruits or holiday items, however in regards to everyday products the retailers may benefit by devising a promotion plan based on different criteria. The biggest difference, in terms of network structure, we find here is that there are cross-category links between promotion nodes and product nodes.



〈Figure 6〉 Bayes Network (N_2) Detailing the Cross-Category Promotion Effect of Display and Feature Promotion

As identified in <Figure 7>, the promotion variable on hotdog and cereal (i.e. *photdog* and *pcereal*), does not only increase the conditional probabilities of hotdog and cereal being included in the basket but also increases the conditional probabilities of catfood and yogurt in the basket. These cross-category links are evidence of direct and mixed cross-category promotion effects that we discussed earlier. Recall that promotion activities are known to have positive effects in directing and enhancing store traffic and by strategically positioning products that are on promotion, retailers can expect to direct customers in such a way where they are exposed to other products seemingly not related to the original promoted product. For example, we see from our learned network that sales of hotdog, catfood and yogurt are not related to each other, meaning that they are not *complementary* products. However, they may *coincide* in a consumer's basket through the presence of direct promotion effect of hotdog. Hence, retailers may gain additional benefits by devising a promotion plan based on the strategic placement of promotional products. In other words, they may profit by displaying promotional products in an area with seemingly unrelated products as one will be able to stimulate the purchase of true *complementary* products via underlying links between products themselves.



〈Figure 7〉 Examples of Cross-Category Effect of Display and Feature Promotion Identified in Bayes Network N_2

The other notable difference from the coupon promotion model is that there are many product promotion variables which are not linked with corresponding product categories. For example, we find that floor/display promotion of coffee, cookies, eggs, vitamin pills, frozen pizzas and snacks do not increase the probability of those products being present in the customer's basket. One may interpret this result in two different ways. The first is that floor/display of said products are ineffective in stimulating consumer purchase. However, this interpretation goes against conventional intuition and contradicts established marketing literature where previous researchers argue that promotion of a product, whether it is a price reduction or display promotion, positively impacts the sales of promoted product. The second interpretation, which is based on the composition of variables, is that in floor/display promotion is not as effective in product categories where there exist brand loyalty. Recall that our model looks at a product category as a whole. However, in certain product categories where brand loyalty is prevalent, consumers may be reluctant to change to another brand just because it is under floor promotion. In this model, we focused our attention on floor feature/display promotion of products. It seems that without actual price discount as an incentive, feature and display promotion on its own may not be enough to lure customers away from certain brands within the product category. By contrast, one may argue that product categories such as catfood, tissue and hotdogs, where promotion variables are linked to the corresponding product purchase nodes are either less sensitive products in terms of brand loyalty or that there is small variety of available brands.

4.3 Heterogeneous Effects of Promotion

As mentioned before, the causal link between promotion and increase in the volume of sales is well established in the literature. In order to validate this positive effect of promotion from our model, we now incorporate both coupons and floor/display promotion variables in the model and generate the Bayesian network structure, and then use the network to update the multinomial parameters accordingly and compute the necessary conditional probabilities.

We then calculate two *conditional odds ratios* to see the effect of promotion on the likelihood of product purchase probabilities. Define OR_p as the conditional odds ratio of product purchase, i.e. the ratio between P (no purchase) and P (purchase) when product is promoted or not. Further, we define OR_h as the conditional odds ratio of high/abnormal product purchase when product is promoted or not. For example, in case of the icecream product category, we find OR_p (Icecream) to be 1.45 and OR_h (Icecream) to be 1.78. This result shows that when ice cream is on promotion consumers are 1.45 times more likely to purchase ice cream and 1.78 times more likely to buy high/abnormal quantity of ice cream compared to when it is not on promotion. In all product cases we find this pattern in terms of conditional odds ratios. OR_p and OR_h are all above 1, which confirms that product promotion increases sales probability. Furthermore, in most cases OR_h were greater than OR_p meaning that promotion more often than not leads to the purchase of high/abnormal quantity.

Let us now turn our attention to the heterogeneous effects of promotion. From our learned Bayesian network, we found that demographic variables such as income and family size are linked with product purchase probability of certain categories. In order to analyze the heterogeneous effect of promotion, we compare the conditional odds ratio OR_p and OR_h of these product categories for different levels of income and family size.

By analyzing the OR_p for product categories linked with income levels in <Table 3>, we see that other than crackers category, the OR_p are mostly similar. This result shows that promotion increases the likelihood of purchase for the promoted products and the increase in the likelihoods are similar in consumers in both income groups. Hence, one can argue that there are no heterogeneous effect of promotion, in terms of purchase decision alone, on customer demographic with different income levels. However, OR_h tell a different story. Recall, that OR_h represent the likelihood of abnormal quantity purchase for the product on promotion. From the table we see that the OR_h are consistently higher for low income households compared to the high income households, meaning that low income households are more likely to purchase high quantity of promoted products. Therefore, one can conclude that the heterogeneous promotion effect can be seen from the difference in high quantity purchase likelihoods for different levels of household income.

〈Table 3〉 OR_p and OR_h for Different Levels of Income

Category	Low income OR_p	High income OR_p	Low income OR_h	High income OR_h
Cereal	2.23	2.38	4.53	3.38
Catfood	1.71	2.49	3.18	2.71
Crackers	4.75	2.13	5.31	3.21
Eggs	3.75	3.43	5.13	3.96
Yogurt	2.77	2.82	4.21	2.81

〈Table 4〉 OR_p and OR_h for Different Family Size

Category	Single member family OR_p	Multi-member family OR_p	Single member family OR_h	Multi-member family OR_h
BBQ	2.53	2.77	3.28	4.38
Butter	2.26	2.54	3.16	3.75
Catfood	2.04	1.97	3.82	3.81
Cereal	1.85	1.68	2.43	4.15
Detergent	1.41	1.39	3.49	2.53
Egg	3.10	2.93	4.88	4.69
Soap	4.22	4.45	5.34	4.17
Softner	1.89	2.02	3.60	2.81
Snacks	3.55	2.89	3.93	4.75
Softdrink	2.37	2.05	3.24	5.73
Tissue	1.67	1.95	3.71	2.11
Frozen Pizza	1.93	1.85	4.78	5.84

We performed the same analysis on conditional odds ratios OR_p and OR_h for households with different family size, and summarized the findings in <Table 4>. Similar to the results for income variable, we find that there are no discernible difference in OR_p for both levels of family size. Once again, we can conclude that there are no heterogeneous promotion effect on product purchase likelihood across different levels of family size. However, there are some interesting results regarding OR_h for single and multiple member households. Other than catfood and eggs, OR_h varies across the two levels of household size, indicating the presence of heterogeneous promotion effect. In more detail, we find that in product categories, barbeque sauce, butter, cereal, snacks and softdrinks, households with multiple family members are more likely to purchase high quantity of the promoted products. Whereas, in product categories detergents, soap, softner, tissue and frozen pizza, single member households are more likely to purchase high quantity of the products on promotion. This result may give an important insight to consumer behavior and their reaction to promotion in that, for single member households promotion yields *stock up* effect on storable products. In the case of consumable food products, it is seen that households with multiple members have greater likelihood of purchasing higher quantity of the product, and this may impact the consumption behavior, or in more detail the rate of consumption, for the promoted products.

5. Concluding Remarks and Future Research Ideas

In this essay we proposed a Bayes network approach to identify and evaluate the cross-category promotion effect of coupons and other promotion activities, accounting for various exogenous effects. Our approach is based on the Bayesian market basket analysis method of Giuidici and Passerone [14] and extends their methodology to include marketing-mix, demographic and seasonal variables. We used MCMC methods such as the MC^3 of Madiagan and York [20] and found that it can handle 56 node variables and over 60000 sparse data points. By incorporating more variables into the model, we were able to see more robust results and paint a clearer picture of association structures than previous works in the literature. Furthermore, we have incorporated an added layer of complexity in modeling the quantity of sales in the study. Previous work on utilizing Bayes networks use a binary variable to define purchase and non-purchase decisions only, whereas in our case we further distinguish purchase decision to high quantity purchase and normal purchase.

We also explored various ways in defining and incorporating the exogenous variables into the model, such that the exogenous effects on the association structure can be effectively measured. Further, we investigated the heterogeneous effect of promotion for households with different levels of income and family. The results showed that promotion does indeed have different impact on different demographics and we explained how this result could be integrated in management's promotion policy.

There are many managerial implications to take away from this study. First and foremost, we showed that cross-category promotion works in various ways depending on promotion strategy. In the case of coupon promotion, there was no evidence of direct or mixed cross-category promotion effects and cross-category promotion effect only occurs through to other product categories via indirectly. Hence, retail managers considering coupon promotion may want to focus on product categories which are network hubs such that the indirect promotion effect to secondary product categories can be maximized. For example, in the case of yogurt and cereal categories which were identified as closely related product categories, it was found that 37% of the times both categories were being promoted simultaneously. By analyzing the cost of running promotion for individual product categories and incorporating the changes in purchase likelihood, retail managers may want to reduce the proportion of simultaneous promotion of closely related product categories. In the case of display/feature promotion, we found evidence of direct promotion effects across product categories which are seemingly not related in function. Retail managers may want to analyze how the products were placed and displayed for these cases and devise a more consistent display plans. Another interesting finding was that display/feature promotion was not effective on categories with established brand loyalty. Again, when creating promotion strategies, managers may want to focus on those products which are affected by display/floor promotion, to maximize the utility of a limited retail space and efficiently spend their marketing budget. Another key finding of this paper which may be useful to management is the evidence of heterogeneous effect of promotion on differing customer demographic. We found that there are certain products where income level of the customer affect the product purchase probabilities. Furthermore, when those product categories are promoted it was shown that the likelihood of high quantity purchase is greater for low income households compared to high income households. Hence to stimulate sales of said products, managers may want to consider targeting low income households with coupons to promote purchase. Investigation of family size and how its association with promotion revealed that, for storable goods promotion increases the likelihood of purchase greater for single member households and for consumable food products promotion is more *effective* on larger households. Like before, when devising a system for target promotion, management can incorporate our findings and distribute coupons accordingly to minimize the cost of promotion and at the same time maximize its effectiveness.

Finally, there are possible limitations to the study, which were identified in the course of our analysis. Note, that our dataset follows a fixed set of households over a 2 year time period. In an effort to include the effect of time in consumer purchase behavior, we took the data as a whole and used monthly indicator variables to model the temporal effect as an exogenous factor. However, in the case of grocery items, where most of the products are everyday goods, it was seen that there were no exogenous time effect. From these findings, our initial belief that time should be modeled from a system perspective, i.e. dynamically, is strengthened and it is only natural to extend our analysis to a dynamic Bayesian network model. Market basket analysis utilizing dynamic Bayesian model will provide us a deeper understanding of promotion effects across time, such as the existence of long-term effect of promotion and whether the network structure itself changes over time. With more data, it would also be insightful to repeat the analysis on monthly transaction data and enhance the validity of the analysis as well as investigating the possible differences in promotion effects across time periods. Another, limitation stemming from the nature of our dataset is that it is a repeated measure from a fixed set of households from Chicago area. We divided the dataset as randomly as possible for training and testing purposes, but we need to exercise caution in generalizing the results of the purchase pattern/behavior learned from our model to other general population such as the Korean population. It would be extremely useful to run our analysis on a different market basket dataset from Korea and compare the results as this may lead to interesting findings regarding the difference in consumer behavior and reaction to promotion activities of consumers in different cultures.

References

[1] Agrawal, R., T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," *Proceedings to the ACM SIGMOD*, ACM(1993), pp.207-216.

[2] Blattberg, R.C., R. Briesch, and E.J. Fox, "How promotions work," *Marketing Science*, Vol.14, No.3(1995), pp.122-132.

[3] Chib, S. and E. Greenberg, "Analysis of multivariate probit models," *Biometrika*, Vol.85, No.2 (1998), pp.47-361.

[4] Chib, S., P. Seetharaman, and A. Strijinev, "Analysis of multi-category purchase incidence decisions using IRI market basket data," *Advances in Econometrics*, Vol.16(2002), pp.57-92.

[5] Chintagunta, P.K., D.C. Jain, and N.J. Vilcassim, "Investigating heterogeneity in brand preferences in logit models for panel data," *Journal of Marketing Research*, Vol.28, No.4(1991), pp. 417-428.

[6] Chen, S.H. and C.A. Pollino, "Good practice in Bayesian network modelling," *Environmental Modelling and Software*, Vol.37(2012), pp.134-145.

[7] Choi, I.Y. and J.K. Kim, "Product Network Analysis to Analyze the Purchase Behavior of Customers," *Journal of the Korean Operations Research and Management Science Society*, Vol.34, No.4(2009), pp.57-72.

[8] Cox, D.R., "Regression models and life-tables," *Journal of the Royal Statistical Soceity Series B*, Vol.34(1972), pp.187-220.

[9] Decker, R., "Market basket analysis by means of growing neural network," *The International Review of Retail, Distribution and Consumer Research*, Vol.15, No.2(2005), pp.151-169.

[10] Donovan, R.J., J.R. Rossiter, G. Marcoolyn, and A. Nesdale, "Store atmosphere and purchasing behavior," *Journal of Retailing*, Vol.70, No.3 (1994), pp.283-294.

[11] Friedman, N., D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, Vol.29, No.2(1997), pp.131-163.

[12] Gamez, J.A., J.L. Mateo, and J.M. Puerta, "Learning Bayesian networks by hill climbing : efficient methods based on progressive restriction of the neighborhood," *Data Mining and Knowledge Discovery*, Vol.22, No.1(2011), pp.106-148.

[13] Garvey, M.D., S. Carnovale, and S. Yeniyurt, "An analytical framework for supply network risk propagation : A Bayesian network approach," *European Journal of Operational Research*, Vol.243, No.2(2015), pp.618-627.

[14] Giudici, P. and G. Passerone, "Data mining of association structures to model consumer behavior," *Computational Statistics and Data Analysis*, Vol.38, No.4(2002), pp.533-541.

[15] Heckerman, D., D. Geiger, and D.M. Chickering, "Learning Bayesian networks : The combination of knowledge and statistical data," *Machine Learning*, Vol.20, No.3(1995), pp.197-543.

[16] Jensen, F.V., "Bayesian Networks and Decision Graphs," Springer, 2001.

[17] Jung, S.-M. and G.-H. Lee, "Artificial Neural Network, Induction Rules, and IRANN to Forecast Purchasers for a Specific Product," *Journal of the Korean Operations Research and Management Science Society*, Vol.30, No.4 (2005), pp.117-130.

[18] Khan, R., M. Lewis, and V. Singh, "Dynamic customer management and the value of one-to-one marketing," *Marketing Science*, Vol.28, No.6(2009), pp.1063-1079.

[19] Kim, J.K., I.Y. Choi, H.K. Kim, and N.H. Kim, "Social Network Analysis to Analyze the Purchase Behavior of Churning Customers and Loyal Customers," *Korean Management Science Review*, Vol.26, No.1(2009), pp.183-196.

[20] Madigan, D. and J. York, "Bayesian graphical models for discrete data," *International Statistical Review*, Vol.63, No.2(1995), pp.215-232.

[21] Manchanda, P., A. Ansari, and S. Gupta, "The shopping basket : A model for multicategory purchase incidence decisions," *Marketing Science*, Vol.18, No.2(1999), pp.95-114.

[22] Marcot, B.G., "Metrics for evaluating performance and uncertainty of Bayesian network models," *Ecological Modelling*, Vol.230(2012), pp.50-62.

[23] Margaritis, D., "Learning Bayesian model structure from data," Ph.D Thesis, Carnegie Mellon University, 2003.

[24] Mulhern, F.J. and R.P. Leone, "Implicit price bundling of retail products : A Multiproduct approach to maximizing store profitability," *Journal of Marketing*, Vol.55, No.4(1991), pp. 63-76.

[25] Russell, G.J. and A. Petersen, "Analysis of cross category dependence in market basket selection," *Journal of Retailing*, Vol.76, No.3 (2000), pp.367-392.

[26] Russell, G.J., S.R. Shocker, A.D. Bell, D. Bodapati, A. Degeratu, L. Hildebrandt, L. Kim, N. Ramaswami, and V.H. Shankar, "Multiple-category decision making : Review and synthesis," *Marketing Letters*, Vol.10, No.3(2003), pp.319-332.

[27] Seetharaman, P. and P.K. Chintagunta, "The

proportional hazard model for purchase timing," *Journal of Business Economic Statistics*, Vol. 21, No.3(2003), pp.368-382.

[28] Swinyard, W.R., "The effects of mood, involvement and quality of store experience on shopping intentions," *Journal of Consumer Research*, Vol.20, No.2(1993), pp.271-280.

[29] Won, Y., "Effective Artificial Neural Network Approach for Non-Binary Incidence Matrix-Based Part-Machine Grouping," *Journal of the Korean Operations Research and Management Science Society*, Vol.31, No.4(2006), pp.69-87.