Scalable Assessment of Cross Category Promotion Effects Using Machine Learning

Soheil Sadeghi
Microsoft, sosadegh@microsoft.com

Jun Li
Point 72, junli.dyingpoet@gmail.com

Yannis Pavlidis
WalmartLabs, yannis@walmartlabs.com

November 15, 2015

Direct effects of promotions are studied extensively in marketing research although indirect effects (cross category research) are less considered. Indirect effects, cannibalization and halo, are typically studied in highly stylized setups (e.g. pancake and syrup) or in setups that involve a handful of categories. Parametric models, mostly regression models with heterogeneity, are usually used to study the indirect effects of promotions on pre-specified items. However, from retailers perspective that cares a great deal about indirect effects the retailing context involves a very large number of categories. Therefore, existing models are not scalable to study indirect effects of promotions for a retailer. In this paper, we propose a framework that is designed to uncover the magnitude and direction of indirect effects. The conceptual framework relies on behavior based Jaccardian index of similarity to assess indirect effects. Machine learning tools are used to uncover items that are independent, exhibit cannibalization or halo effects. The conceptual framework is simple to use, offers usable insights for a retailer and is scalable.

Key words: promotion, cannibalization, Jaccardian Index, machine learning

---

1. Problem

In this paper, we focus on the cannibalization effect of the promoted items on the non-promoted items' sales for a retailer. To clearly define the cannibalization effect of promotion, consider a customer facing with a dilemma: choice of a meal between two options; peanut butter-jelly sandwich and nutella sandwich. Suppose the customer has no specific preference (financial, personal, ...) for choosing either one meaning the chance of choosing peanut butter-jelly sandwich is the same as choosing nutella sandwich. To simplify, considers a quick grocery shopping from a small retailer that offers four items peanut butter, jelly, nutella, and bread without any specific promotions. We can then write the problem as a probability model in which the customer's chance of buying bread is one although its chance of buying each other item is a half, subject to three other conditions:

- The chance of buying peanut butter and jelly together is a half.
- The chance of buying peanut butter and nutella together is zero.
- The chance of buying jelly and nutella together is zero.

Table 1 includes this probability model in the no-promotion situation. Probability of other combinations of items and conditional cases are easily calculable using the numbers included in the same table. For instance, it is easy to show that $P(PeanutButter \cap Jelly \cap Nutella) = 0$ and $P(PeanutButter|Jelly) = 1$. Therefore, this probability model is unique and well-defined.

Table 1 **Probability Model of the Peanut Butter-Jelly and Nutella Sandwich Dilemma (Without Promotions)**

$P(PeanutButter) = 1/2$	$P(PeanutButter \cap Jelly) = 1/2$
$P(Jelly) = 1/2$	$P(PeanutButter \cap Nutella) = 0$
$P(Nutella) = 1/2$	$P(Jelly \cap Nutella) = 0$
$P(Bread) = 1$	

Next, assume the retailer launches a promotion on peanut butter. As a result, customer's chance of buying peanut butter increases. This is the direct effect of peanut butter's promotion on itself. The chance of buying jelly also increases since peanut butter and jelly come together. This indirect effect of peanut butter's promotion on jelly's sale trend is called the "halo" effect. On the other hand, the chance of buying nutella decreases. This indirect effect of peanut butter's promotion on nutella's sale trend is called the "cannibalization" effect. Finally the chance of buying bread does not change meaning peanut butter's promotion has "neutral" effect on bread. Promotion on peanut butter changes the probability model of the above dilemma to be the one in Table 2.

Table 2 Probability Model of the Peanut Butter-Jelly and Nutella Sandwich Dilemma (With Promotion on Peanut Butter)

$P(PeanutButter) > 1/2$	$P(PeanutButter \cap Jelly) > 1/2$
$P(Jelly) > 1/2$	$P(PeanutButter \cap Nutella) = 0$
$P(Nutella) < 1/2$	$P(Jelly \cap Nutella) = 0$
$P(Bread) = 1$	$P(PeanutButter \cap Jelly) = P(PeanutButter) = P(Jelly)$

In general, the relation between pairs of items can be considered as either a halo, a cannibalization, or a neutral relation. The focus of this work is to study the cannibalization effect of promoted items on the sales of other items for a retailer. In the next section, we propose two fundamental features in order to explain the potential cannibalization relation between pairs of items.

2. Features Foundation

A retailer mostly communicates with customers through purchase baskets. The set of all customers is then equivalent to the set of all purchase baskets from a retailer's perspective meaning to study customers, a retailer can investigate all the purchase baskets. Each purchase basket is a set of items. Therefore, one way to study the purchase baskets is to understand the way items interact with each other through exploring the placement of items in the purchase baskets. In this section, we introduce two fundamental concepts in order to explain the way items interact with each other. We will first define a feature that can measure the chance of items coming together in a purchase basket. To develop this concept, consider a retailer with six items A, B, C, D, E, and F. Suppose there are, in total, eight purchase baskets for this retailer. Table 3 represents all eight baskets B1, B2, ..., B8. For instance, B2 includes four items A, B, C, and F although B8 only includes two items D and E.

Table 3 Eight Purchase Baskets for a Retailer with Six Items A, B, C, D, E, and F

B1	B2	B3	B4	B5	B6	B7	B8
A	A	A	C	C	A	B	D
B	B	C	D	D	B	C	E
C	C	F	E		C	D	
	F					F	

Considering this as the population, we can define the purchase probability of each item to be the number of baskets including that item divided by total number of baskets. To find the purchase probability of item A, for instance, we define the purchase basket set $\mathbf{A}$ to be the set of all baskets that include item A i.e. $\mathbf{A} = \{B1, B2, B3, B6\}$. Note that, we use the boldface letter $\mathbf{A}$ to refer to the purchase basket set although A just means item A. Purchase probability of item A is hence defined to be the probability of occurring the purchase basket set $\mathbf{A}$ i.e. $P(\mathbf{A}) = 0.5$. Table 4 includes the purchase basket set and purchase probability for all eight items.

Table 4 Purchase Basket Set and Purchase Probability for Six Items A, B, C, D, E, and F

Item	Purchase Basket Set	Purchase Probability
A	$\mathbf{A} = \{B1, B2, B3, B6\}$	$P(\mathbf{A}) = 4/8$
B	$\mathbf{B} = \{B1, B2, B6, B7\}$	$P(\mathbf{B}) = 4/8$
C	$\mathbf{C} = \{B1, B2, B3, B4, B5, B6, B7\}$	$P(\mathbf{C}) = 7/8$
D	$\mathbf{D} = \{B4, B5, B7, B8\}$	$P(\mathbf{D}) = 4/8$
E	$\mathbf{E} = \{B4, B8\}$	$P(\mathbf{E}) = 2/8$
F	$\mathbf{F} = \{B2, B3, B7\}$	$P(\mathbf{F}) = 3/8$

So far, we have defined the purchase probability for each item meaning if a customer is trying to make a purchase basket we have an idea about the chance of each item being included in its basket. However, using the purchase probability of each item, we cannot claim anything about the way different items interact with each other. For instance, if a customer puts item A in its basket, what is the chance of item B being included in its basket as well? To address the purchase probability of item B given the fact that item A has been purchased, we use the conditional probability concept and find $P(\mathbf{B}|\mathbf{A})$ as follows:

$$P(\mathbf{B}|\mathbf{A}) = \frac{P(\mathbf{B} \cap \mathbf{A})}{P(\mathbf{A})} = \frac{|\mathbf{B} \cap \mathbf{A}|}{|\mathbf{A}|} \tag{1}$$

where $|.|$ denotes the size of a set.

The conditional probability $P(\mathbf{B}|\mathbf{A})$ measures the chance of purchasing item B resulting from purchasing item A. Similarly, $P(\mathbf{A}|\mathbf{B})$ measures the chance of purchasing item A resulting from purchasing item B. However, these two probabilities are not necessarily equal meaning the chance of purchasing item B resulting from purchasing item A is not necessarily the same as the chance of purchasing item A resulting from purchasing item B. Therefore, using the conditional probability as a similarity index between pair of items results in an unsymmetrical index meaning the similarity from A to B is not necessarily the same as similarity from B to A. To make it a symmetric index, in 1, we change the denominator to be the union of the two sets **A** and **B** instead of the marginal ones. The resulted similarity index is known as the Jaccard index $J(\mathbf{A}, \mathbf{B})$ calculated by

$$J(\mathbf{A}, \mathbf{B}) = \frac{|\mathbf{A} \cap \mathbf{B}|}{|\mathbf{A} \cup \mathbf{B}|}. \tag{2}$$

We can interpret the Jaccard index, defined on the purchase basket sets, as the chance of two items coming together in a basket. We use the name "normalized match" in our context to refer to the Jaccard index defined on the purchase basket sets. We use the word match as we consider number of baskets in which two items are included, and the word normalized as we divide by the size of their union. Therefore, we define the normalized match between A and B, $N.M(A, B)$, by

$$N.M(A, B) = J(\mathbf{A}, \mathbf{B}) \tag{3}$$

where sets **A** and **B** represents the purchase basket sets of item A and B, respectively. Table 5 includes the normalized match between all pairs of items A, B ..., F.

Normalized match is representing a customer's chance of purchasing two items together in a purchase basket. A high normalized match, for a pair of items, means that purchasing either one

Table 5 Normalized Match Between All Pairs of Items A, B, ..., F

Items	B	C	D	E	F
A	3/5	4/7	0	0	2/5
B		4/7	1/7	0	2/5
C			3/8	1/8	3/7
D				2/4	1/6
E					0

in a basket likely results in purchasing the other one in the same basket too. For instance, high $N.M(A,C)$ in Table 5 implies high $P(\mathbf{A}|\mathbf{C})$ and $P(\mathbf{C}|\mathbf{A})$.

We will next define another feature that can provide more information about the pairs of items with low normalized match. To develop this feature, consider two normalized matches $N.M(A,D) = 0$ and $N.M(A,E) = 0$ in Table 5. These numbers imply that, if a customer purchases item A in its basket, the chance of purchasing item D or E in the same basket as a result of A is zero, i.e., $P(\mathbf{D}|\mathbf{A}) = 0$ and $P(\mathbf{E}|\mathbf{A}) = 0$. However, we cannot claim that, because $N.M(A,D) = 0$ and $N.M(A,E) = 0$, the way item A interacts with item D is the same as it interacts with item E. To clarify this, compare the chance of purchasing item D as a result of not purchasing item A, $P(\mathbf{D}|\mathbf{A}^c) = 1$, with the chance of purchasing item E as a result of not purchasing item A, $P(\mathbf{E}|\mathbf{A}^c) = 0.5$, where $\mathbf{A}^c$ denotes the complement of the set $\mathbf{A}$. For a customer who is not going to purchase item A, item D is going to be in its basket ($P(\mathbf{D}|\mathbf{A}^c) = 1$) although item E may not going to be there ($P(\mathbf{E}|\mathbf{A}^c) = 0.5$). Therefore, there is a strong "connection" between A and D that does not exist between A and E. The existence of item D in a customer's purchase basket really depends on the existence of item A (if A is there, D is not going to be there although D will be there if A is not going to be there). However, the existence of item E in a customer's purchase basket while A is not there has only a chance of 0.5. We now develop a feature that can differentiate the strong connection between A and D from the week connection between A and E.

Figure 1 shows all the items in a graph. Two items are connected to each other if the normalized match between them is positive. In this graph, nodes and edges represent items and the normalized match between the pair of items, respectively. For instance, there is an edge between the pair A and C although there is no edge between the pair A and D.

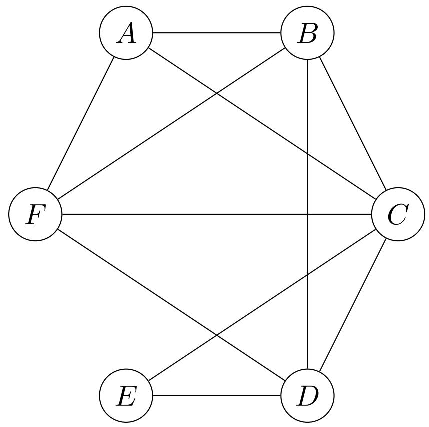

Figure 1 Graph Representing Items and the Normalized Match Between Pairs of Items.

There is no edge between the pair A and E as well as the pair A and D meaning they are not directly connected to each other. However, if we consider one-node connections between the pairs, there are three one-node connections between the pair A and D whereas there is only a single one-node connection between the pair A and E. In other words, the pair A and D is indirectly connected through three nodes B, C, and F although the pair A and E is indirectly connected only through the node C. Consider a basket with three items B, C, and F with free spots for any of three items A, D, and E. If item A is included in the basket then item D (or E) will not be included as $P(\mathbf{D}|\mathbf{A}) = 0$ (or $P(\mathbf{E}|\mathbf{A}) = 0$). If item A is not included in the basket then item D will be included as $P(\mathbf{D}|\mathbf{A}^c) = 1$ although item E only has a 0.5 chance to be included in the same basket ($P(\mathbf{D}|\mathbf{A}^c) = 1$). The main reason is that item D has a positive chance to be in the same basket with all three items B, C, and F although item E only has a positive chance with item C to be in the same basket. Therefore, number of one-node connections can differentiate the strong connection between A and D from the week connection between A and E. Indirect one-node connections can also be similarly defined for directly connected pairs.

We introduce a concept called “normalized common” in order to quantify the one-node connections between the pairs of items. To define the normalized common between the pair A and E, for instance, note that item A is directly connected to three items B, C, and F although item E is directly connected to two items C and D. For item A, we define the purchase co-item set $\mathbf{A}'$ to be the set of all items that have co-existed with item A at least in one basket including itself (if they have co-existed at least in one basket they have a positive normalized match and vise versa) i.e. $\mathbf{A}' = \{A, B, C, F\}$. Similarly, the purchase co-item set for item E is defined $\mathbf{E}' = \{C, D, E\}$. By this definition, obviously, the purchase co-item set of each item always includes that item itself. Table 6 shows the purchase co-item set of each item.

Table 6 **Purchase Co-item Set for Six Items A, B, C, D, E, and F**

Item	Purchase Co-item Set
A	$\mathbf{A}' = \{A, B, C, F\}$
B	$\mathbf{B}' = \{A, B, C, D, F\}$
C	$\mathbf{C}' = \{A, B, C, D, E, F\}$
D	$\mathbf{D}' = \{B, C, D, E, F\}$
E	$\mathbf{E}' = \{C, D, E\}$
F	$\mathbf{F}' = \{A, B, C, D, F\}$

Normalized common between two items A and E, $N.C(A, E)$, is then defined to be the Jaccard index defined on the purchase co-item sets:

$$N.C(A, E) = J(\mathbf{A}', \mathbf{E}') \tag{4}$$

where sets $\mathbf{A}'$ and $\mathbf{E}'$ represents the purchase co-item sets of item A and E, respectively. We use the word common as we consider how many co-items are in common between A and E, and the word normalized as we divide by the union. Table 7 includes the normalized common between all pairs of items A, B ..., F.

3. Modeling

Consider a retailer with large number of items that are classified into multiple categories. Items within each category are also grouped into several subcategories. In general, assume a retailer

Table 7 **Normalized Common Between All Pairs of Items A, B, ..., F**

Items	B	C	D	E	F
A	4/5	4/6	3/6	1/6	4/5
B		5/6	4/6	2/6	1
C			5/6	3/6	5/6
D				3/5	4/6
E					2/6

with c categories, s_r subcategories within the rth category ($r = 1, \cdots, c$), and n_{rk} items within the kth subcategory of the rth category ($k = 1, \cdots, s_r$). We introduce a response model to study the cannibalization at item level. We also generalize the model for the subcategory level as most retailers in industry are interested in studying the bigger pictures. We also discuss the challenges we face for generalizing the model to the subcategory level and how it impacts the interpretation at the subcategory level compared to the item level.

3.1. Item Level Model

Large retailers like Sam's Club usually hold several promotion campaigns in a year. During each promotion campaign, there are several promoted items in order to increase the customer engagement. As a result of promotion, the sale of promoted items usually increases, and many non-promoted items typically receive either halo or cannibalization effect although most receive no specific effect (neutral effect). To study and detect the cannibalization effect of promoted items on non-promoted items, we consider the historical data from a time window that includes T non-overlap promotion campaigns. These campaigns do not generally have the same time length and do not occur according to a specific pattern although the difference is not significant. Assume that for the tth promotion campaign, there are p_t promoted items ($t = 1, \cdots, T$). For each promoted item, all $\sum_{r=1}^{c} \sum_{k=1}^{s_r} n_{rk}$ items are considered in the model as the potential cannibalized candidates. We will add some features in the model in order to filter some of these items. For instance, if two items A and B are promoted, item B should not be considered as a potential cannibalized item for promoted item A as this indirect cannibalization effect is negligible compared to the direct effect of promoted item B on itself. The data for the tth promotion campaign then consists of $p_t \sum_{r=1}^{c} \sum_{k=1}^{s_r} n_{rk}$ rows in which each row represents a pair of items (promoted item, potential candidate). The data for all T promotion campaigns thus has $\sum_{t=1}^{T} p_t \sum_{r=1}^{c} \sum_{k=1}^{s_r} n_{rk}$ rows in total. For each row, we define a response variable that depends on the potential candidate, and some learning features that depend on both the promoted item and the potential candidate

To define the response variable for each row of the tth promotion campaign, we compare the actual sale and the expected sale of the potential candidate during that campaign. Expected sale of each potential candidate is the sale that is expected to occur during the promotion campaign if there were no promotion during that time. It is calculated based on a time series model that uses the sale data of earlier non-promotion campaigns in order to predict the expected sale during a promotion campaign. See De Gooijer and Hyndman (2006) for details on forecasting time-series data. Generally speaking, there are two major components contributing to the expected sale: trend and seasonality. Trend refers to the velocity of sales in the past few weeks, i.e. the rate at which sales have been increasing/decreasing Jha et al. (2015). Seasonality refers to periodic events that influence the sales, for e.g. Thanksgiving, Back-to-School sales Jha et al. (2015). Folks at @Walmartlabs are using clustering to forecast their sparse time-series data. For details on how they calculate the expected sale refer to Jha et al. (2015). Once the actual and the expected sale are calculated, the response variable for the potential candidate j in the promotion campaign t is then defined to be a three-level categorical variable:

$$Y_{jt} = \begin{cases} 1, & \text{if } \left(\frac{\text{Expected Sale}}{\text{Actual Sale}}\right)_{jt} < s_t^h \\ -1, & \text{if } \left(\frac{\text{Expected Sale}}{\text{Actual Sale}}\right)_{jt} > s_t^c \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where s_t^h and s_t^c are empirically determined by the distribution of $\left(\frac{\text{Expected Sale}}{\text{Actual Sale}}\right)_{jt}$'s for all j during the promotion campaign t. These are usually determined such that we have %10 of the data above s_t^h and below s_t^c. If, for a potential candidate, the expected sale is significantly smaller than its actual sale, then it is labeled 1 representing a haloed item as a result of the promotion campaign.

On the other hand, for a potential candidate, if the expected sale is significantly larger than its actual sale then it is labeled -1 meaning a cannibalized item as a result of the promotion campaign. The rest of the candidates are labeled 0. By defining the response variable as a categorical variable, we are approaching the problem from a probabilistic point of view. For each candidate, we are trying to estimate the probability of getting cannibalized, the probability of getting haloed, and the probability of getting no effect.

The features, for the row corresponding to the promoted item i and potential candidate j in the tth promotion campaign, are as follows:

- Normalized Match: $N.M(i,j)_t$ (based on the data up to the promotion campaign t)
- Normalized Common: $N.C(i,j)_t$ (based on the data up to the promotion campaign t)
- $C(i,j) = \begin{cases} 1, & \text{if promoted item } i \text{ and potential candidate } j \text{ belong to the same category} \\ 0 & \text{otherwise} \end{cases}$
- $S(i,j) = \begin{cases} 1, & \text{if promoted item } i \text{ and potential candidate } j \text{ belong to the same subcategory} \\ 0 & \text{otherwise} \end{cases}$
- $P(j)_t = \begin{cases} 1, & \text{if the potential candidate } j \text{ is a promoted item in the } t\text{th promotion campaign} \\ 0 & \text{otherwise} \end{cases}$
- $I_t = \begin{cases} 1, & \text{if the row belongs to the } t\text{th promotion campaign} \\ 0 & \text{otherwise} \end{cases}$

We can finally write the model as:

$$Y_{jt} = f\Big(N.M(i,j)_t < \pi, N.C(i,j)_t > \pi', C(i,j), S(i,j), P(j)_t, I_t\Big) + error_{ijt} \qquad (6)$$

where $f(.)$ is an arbitrary function ($t = 1, \cdots, T$; $i = 1, \cdots, p_t$; $j = 1, \cdots, \sum_{r=1}^{c} \sum_{k=1}^{s_r} n_{rk}$). Also, π and π' are empirically determined based on the distributions of the normalized match and normalized common. The region $(N.M(i,j)_t < \pi, N.C(i,j)_t > \pi')$ filters the data to be the potential region for studying the cannibalization, shown in Figure **??**.

The model in 6 is applied to a train data set in order to relate the learning features to the response variable. The learned model is then applied to a test data set with given features in order to estimate the probability of each candidate belonging to the cannibalized set, haloed set, and the neutral set. Note that each item can be considered as a potential candidate for more than one promoted item. Therefore, for each promoted item, the probability of a candidate getting cannibalized or haloed will be estimated. For each promoted item, all potential candidates will be thus ranked based on a function of estimated probabilities (halo probability − cannibalization probability) and the cannibalized items will be detected empirically using the distribution of that function. The output of this model is a list of cannibalized items for each promoted item. Therefore, one promoted item could have no cannibalized item although another one can have more than one cannibalized items. In addition, an item can be appeared in the cannibalized list of more than one promoted item meaning an item can receive cannibalization effect from multiple promoted items.

Based on the model in 6 all items in the region $(N.M(i,j)_t < \pi, N.C(i,j)_t > \pi')$ are considered as the potential cannibalized candidates for each promoted item. However, the features used in the model allow for filtering the data. For instance, to consider the non-promoted items as the only potential candidates, we will use the model:

$$Y_{jt} = f\Big(N.M(i,j)_t < \pi, N.C(i,j)_t > \pi', C(i,j), S(i,j), P(j)_t = 1, I_t\Big) + error_{ijt}. \tag{7}$$

Similarly, to restrict the model in 7 to within-category potential candidates, the model in 8 will be used.

$$Y_{jt} = f\Big(N.M(i,j)_t < \pi, N.C(i,j)_t > \pi', C(i,j) = 1, S(i,j), P(j)_t = 1, I_t\Big) + error_{ijt}. \tag{8}$$

3.2. Subcategory Level Model

To generalize the model in 6 to the subcategory level, we first define a “promoted subcategory” to be a subcategory with at least one promoted item. Based on this definition, we ignore the number of promoted items within each subcategory meaning there is no difference between a subcategory with only one promoted item and a subcategory with more than one promoted items (both are labeled as promoted subcategories). The goal is to detect and study the cannibalized subcategories caused by each promoted subcategory. Normalized match and normalized common are calculated based on the same formula with one difference that they are defined over the pairs of subcategories not the pairs of items. The feature $C(u,w)$ is defined to be one if two subcategories u and v belong to the same category. Feature $P(v)_t$ is defined to be one if subcategory v is a promoted subcategory in the promotion campaign t. Finally, the response variable is defined similar to the formula in 5 using the aggregated sales info for subcategories. Assuming that there are q_t promoted subcategories for the tth promotion campaign $(t=1,\cdots,T)$, we write the model as follows:

$$Y_{vt} = g\Big(N.M(u,v)_t < \pi, N.C(u,v)_t > \pi', C(u,v), P(v)_t, I_t\Big) + error_{uvt} \tag{9}$$

where $g(.)$ is an arbitrary function $(t=1,\cdots,T;\ u=1,\cdots,q_t;\ v=1,\cdots,\sum_{r=1}^{c} s_r)$. Similar to the item level case, data filtering is applied using the provided features. For instance to use the within-category non-promoted subcategories as the potential candidates the following model will be used:

$$Y_{vt} = g\Big(N.M(u,v)_t < \pi, N.C(u,v)_t > \pi', C(u,v)=1, P(v)_t=1, I_t\Big) + error_{uvt}. \tag{10}$$

Similar to the item level case, the model will provide a list of cannibalized subcategories for each promoted subcategory. However, the effect of aggregating the data from item level to the subcategory level changes the interpretation of the result.

3.2.1. Aggregation Effect

To use the subcategory-level model in 9, the two features normalized match and normalized common are calculated between the pairs of subcategories. The empirical data suggests that, as a result of aggregating the data from the item-level to the subcategory-level, the two features normalized match and normalized common tend to converge for subcategories within each category. Table 8 includes the two features between the subcategory “dry shave” and five other subcategories within the "health and beauty aids" category of our empirical study with Sam's club data. The normalized matches are all very close to each other as well as all the normalized commons. The model in 9 uses normalized match and normalized common as the primary features to study the pairs of subcategories. Therefore, as suggested by the empirical data, the model cannot distinguish the pair (dry shave, shaving) from the pair (dry shave, skin care). Consider the case where we have promotion on all five subcategories "shaving", "skin care", "bodywash", "face care", and "hair care". The sales data also suggests that the subcategory "dry shave" has received cannibalized effect during the promotion. Based on the model in 9 it is impossible to relate this cannibalization effect to only a subset of promoted subcategories. However, we can relate this cannibalization effect to the set of all five promoted subcategories.

Table 8 Aggregation Effect on Normalized Match and Normalized Common

Subcategory	Subcategory	Normalized Match	Normalized Common
dry shave	shaving	0.022	0.561
dry shave	skin care	0.021	0.565
dry shave	bodywash	0.021	0.556
dry shave	face care	0.022	0.578
dry shave	hair care	0.019	0.551

As suggested by the aggregation effect, the way we interpret the results at the subcategory-level is as follows. Suppose we have 100 subcategories within a category and there is a promotion on 10 of the subcategories. The model detects a set of subcategories, within the same category, that are cannibalized during the promotion campaign. We interpret the cannibalized set to be the result of the promoted set instead of relating each cannibalized subcategory to each promoted subcategory. In other words, all the cannibalized subcategories are considered to be the result of all promoted subcategories and no individual contribution will be reported.

4. Random Forests as The Learning Method

Random forests, proposed by Breiman (2001), is a combination of decision trees such that trees depend on randomly selected values of the features with the same distribution. To build a classification tree, literally speaking, we divide the feature space (possible values for features) into distinct and non-overlapping regions. For every observation that falls into a region, we let the training observations in that region vote for the most popular class. To learn about how to construct a decision tree, Breiman (1996), Dietterich (2000), and Breiman (1999) provide some decent examples and clarification. The significant disadvantage of decision trees is having a high variance. Generally speaking, by splitting the training data into several parts and applying a decision tree to each part could yield in quite different results. Bagging Breiman (1996) is a general learning method to reduce the variance of a learning method and is frequently applied to decision trees. To apply bagging to classification trees, we can construct multiple decision trees using multiple bootstrapped training sets and take the majority vote. Random forests is an improved version of bagged trees in order to decorrelate the trees. Similar to bagging, decision trees are applied to bootstrapped training sets. However, only a small fraction of features are randomly selected as split features for building the decision trees. It might seem irrational but there is a neat justification behind the idea. In bagged trees, for the case with one very strong feature along with some other moderately strong features, the strong feature will be used as the top split in most of decision trees. As a result, most trees are similar resulting in highly correlated set of trees. Bagging on highly correlated trees does not reduce variance a lot. Random forests overpower the correlated trees by forcing each tree to use only a small fraction of predictors. For deep discussion and technical details on random forests, refer to Breiman (2001) and Hastie et al. (2005).

To learn the model on the training data, in our empirical application, a lot of methods can be used such as random forests and gradient boostings. The results are not significantly different for different methods as far as the data is rich. The main reason for choosing random forests as the learning method here is that random forests is very fast and can be done in parallel. For retailers that are dealing with very large data sets and need to run algorithms every day, being able to process the task in parallel and fast is a big advantage. Gradient boosting computations, for instance, can not be done in parallel and is usually slower than random forests. Our empirical applications shows there is some significant information in the data we feed to the model. Therefore, no matter what learning algorithm is used, the results are not very different. Random forests is chosen because it is simple to understand and explain, it is fast, and the computation can be done in parallel.

5. Empirical Application

The data for our empirical study belongs to Sam's Club with 100 categories and 100 subcategories within each category. Sam's Club usually holds promotion campaigns for almost 20 days every other month. Therefore, there are typically six promotion campaigns and six non-promotion campaigns in a year. The data for our study was for a six month window starting January 2015 to June 2015 and the goal was to build a learning algorithm that uses this historical data in order to predict for promotion campaign in July 2015. Table 9 includes duration and number of promoted subcategories in each promotion campaign. The study is at the subcategory-level that uses the model in 10 to learn from the historical data. The learned model is then applied to the data from July 2015 (without sales information) in order to detect the cannibalized subcategories using the learning features in the model. The result of the model is then compared with sales data during July 2015 to check how accurate the learning model is.

Table 9 Descriptive Statistics of Sam's Club Empirical Data

Promotion Campaign	January 2015	March 2015	May 2015
Duration (in days)	31	28	25
# Promoted Subcategories	133	129	132

Table 10 includes the description of the data that is prepared for our study. Sale ratio is the ratio of the expected sale and the actual sale which is calculated using the Sam's Club's transaction data and the times series model mentioned in Subsection **??**. The response variable of the model in then defined based on the definition in 5 using the sale ratio data. Normalized match, normalized common, category binary, and promotion binary are all calculated based on the definitions mentioned in Section 3. There are also six binary variables to identify what rows belong to each promotion campaign.

Table 10 Data Description of Sam's Club Empirical Data

Variable	Type	Description
Prom-ID	smallint	Unique identifier for the promoted subcategory
Poten-ID	smallint	Unique identifier for the potential subcategory candidate
Sale Ratio	double	Potential candidate's expected sale divided by its actual sale
Norm-Match	double	Normalized match between the pair of subcategories
Norm-Common	double	Normalized common between the pair of subcategories
Cat-binary	tinyint	1 if the pair of subcategories belongs to the same category, otherwise 0
Prom-binary	tinyint	1 if the potential candidate is a promoted subcategory, otherwise 0
May2015	tinyint	1 if the row belongs to promotion campaign May 2015, otherwise 0
March2015	tinyint	1 if the row belongs to promotion campaign March 2015, otherwise 0
Jan2015	tinyint	1 if the row belongs to promotion campaign January 2015, otherwise 0

To estimate the probability of each candidate belonging to each categories of the response variable, two binary classifications are used instead of a three-level classification. Two binary variables "cannib" and "halo" are defined as follows:

$$\text{cannib}_{jt} = \begin{cases} 1, & \text{if } \; Y_{jt} = -1 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

$$\text{halo}_{jt} = \begin{cases} 1, & \text{if } \; Y_{jt} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

One classification is used to model *cannib* on the learning features to estimate the probability of the candidates belonging to the cannibalized set. Another classification is used to model *halo* on the learning features to estimate the probability of candidates belonging to the haloed set. By subtracting the sum of the two cannibalization and halo probabilities from one, for each candidate, neutral probability can be calculated. Random Forest is used to learn the classifications and to estimate the probabilities. Tables 11 and 12 includes features importance, in percentages, for the two classification learners.

The two classifier are then applied to the test data set from July 2015 in order to estimate the probability of each candidate belonging to the cannibalized set, haloed set, and the neutral

Table 11 Features Importance for Cannib Classification

Features	Importance
Norm-Match	%36
Norm-Common	%24
March2015	%14
Jan2015	%14
May2015	%12

Table 12 Features Importance for Halo Classification

Features	Importance
Norm-Match	%45
Norm-Common	%33
March2015	%8
Jan2015	%7
May2015	%7

set. For the set of promoted subcategory within each category, all potential candidates will be thus ranked based on a function of estimated probabilities (halo probability − cannibalization probability) and the cannibalized subcategories will be detected empirically using the distribution of that function. The output of this model is a list of cannibalized subcategories for all promoted subcategories within each category. For the promotion campaign in July 2015, subcategories within 20 categories are received promotions. Tables 13-16 includes the cannibalized subcategories within each category. In the same tables, third column is the sale ratio for the cannibalized subcategories detected by the model. As shown in the tables, most of the sale ratios are above 1 which implies the expected sale is much larger than what is actually sold during the promotion campaign. The model has detected these cannibalized subcategories without using the sale ratio for July 2015. This comparison verifies that the model has done a solid prediction for July 2015.

References

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Breiman, L. (1999). Using adaptive bagging to debias regressions. Technical report, Technical Report 547, Statistics Dept. UCB.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

De Gooijer, J. G. and Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting*, 22(3):443–473.

Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning*, 40(2):139–157.

Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85.

Jha, A., Ray, S., Seaman, B., and Dhillon, I. S. (2015). Clustering to forecast sparse time-series data. In *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*, pages 1388–1399. IEEE.

Table 13 **Result**

Category: snacks		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
seeds		
choose any two		
tortilla chips	regional	1.38
variety pack chips	bulk chips	1.15
snack mixes		
Category: audio		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
	portable audio	4.07
	streaming players	1.41
outdoor speakers	headphones	4.68
sound bars	home theater	2.89
	apple tv	2.07
Category: soda		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
	pr - can kola champagne	1.08
	pepsi csd ms pet	1.08
	coke flavors / seasonal	1.12
energy drinks	energy shots	1.12
ready to drink coffee / milk	coke csd ss pet	1.14
pepsi flavors seasonal	pr - small p.e.t pinapple	1.20
pepsi csd ss pet	pr - non alcoholic malta	1.20
pepsi csd cans	other flavors / seasonal	1.19
	dpsg csd ss pet	1.13
	other premium / glass csd	1.74
	coke premium / glass csd	1.45
Category: cooler		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
kids yogurt	breakfast snacks	1.07
greek yogurt	cookie dough	2.96
sliced cheese	protein salads	1.47
lunch kits	seasonal cheese	12.55
snacking	chunk cheese - wholesale	1.34
	breakfast sausage - precooked	1.11
Category: printers and ink		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio

Table 14 Result - Continued

Category: gourmet deli and home meals		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
	sauces	7.17
	prime rib	10.59
	premade party platters	1.06
	hms party trays	2.69
soups/stew	brisket	2.23
snack cheese	goat cheese	1.86
hummus	cheese cubes	2.05
dips/spreads	guacamole dip	1.29
snacking	cheese variety packs	2.48
	asian foods	0.91
	sushi	1.02
	regional dsd cheese	1.35
Category: mattresses		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
king	full	1.24
queen	twin	1.20
	regular foam	1.20
Category: diamonds		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
	hoop earrings	6.02
	wedding bands	3.84
	solitaire stud earrings	5.03
silver and diamond necklaces	composite bridal	4.84
	heart pendants	4.17
	fashion ring- composites	4.54
	special orders	1.09
Category: seasonal - accessories		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
storage	video game accessories	2.89
tablet access	treasure hunt	2.26
phones	computer accessories	0.91
av accessories	video game hardware	6.13
hearing aid batteries	video game software	4.58
variety packs	rechargeables	1.23
combo packs	chargers	1.63

Table 15 Result - Continued

Category: otc		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
diet drinks and powders	diet pills	0.93
diet bars	gnc	0.83
smoking	business dispensers	1.4
coq10 supplements	diabetic	1.04
child multivitamins	energy	1.11
minerals	herbals/supplements	0.96
laxatives	letter vitamins	1.01
adult multivitamins	footcare	1.15
antacids		
Category: canned protein condiments pasta soup		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
salsa / picante sauce	ketchup - business	1.05
	bbq sauce - business	1.24
Category: oil - rice - fruits - vegetables		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
fruit cups	business gravy	1.27
fruit pouches	canned fruit - retail	1.23
	tomatoes 10	1.22
Category: candy - snacks - business		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
fruit snacks	big bag choc	2.37
cookies	premium chocolate	3.67
Category: juice - water - sport drinks		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
kids drinks	flavored water	1.04
Category: janitorial		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
air care	aerosols	2.01
Category: frozen foods		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
potatoes		
multiserve pizza		
family snacks	shellfish	1.18
appetizers	bbq/chili/pork	1.29
value added shrimp	business fries	1.09
[illegible]	[illegible]	[illegible]

Table 16 Result - Continued

Category: health and beauty aids		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
power accessories		
shaving		
face care		
bodywash	liquid soap	1.02
cotton swabs	dry shave	3.34
deodorant	intimacy	0.76
skin care	hair regrowth	1.05
bar soap	hair clippers and accessories	3.51
mouthwash	cosmetic and gift sets	3.05
toothpaste	dentures	1.18
hairstyling		
hair care		
Category: televisions		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
	uhd led 55 - 59 inch	1.14
	uhd led 60 inch	0.8
	dvd combo tv	2.79
lcd 60 inch	lcd 46 - 54 inch	1.76
lcd 55 - 59 inch	lcd 19 - 31 inch	2.23
uhd led 46 - 54 inch	lcd 32 - 36 inch	1.96
	lcd 37 - 45 inch	1.5
	uhd led 61 - 70 inch	3.37
	75 inch and larger	2.01
Category: dairy		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
	milk ss multi pack	1.10
milk plant based	cream cheese - wholesale	1.48
	portion control-butter margari	1.42
	milk organic	1.04
Category: laundry and home care		
Promoted Subcategories	Cannibalized Subcategories	Sale Ratio
air care		
quick clean	disinfectant spray	1.07
bathroom cleaners	sponges	1.34
furniture polish	moisture absorber	0.8
light duty dish det	septic tank	0.98