# Tracking the Evolution of Customer Purchase Behavior Segmentation via a Fragmentation-Coagulation Process

**Ling Luo[1,2,*], Bin Li[2], Irena Koprinska[1], Shlomo Berkovsky[2], Fang Chen[2]**
[1]School of IT, University of Sydney, Sydney, NSW 2006, Australia
[2]Data61, CSIRO, Eveleigh, NSW 2015, Australia
{ling.luo, irena.koprinska}@sydney.edu.au, {bin.li, shlomo.berkovsky, fang.chen}@data61.csiro.au

## Abstract

Customer behavior modeling is important for businesses in order to understand, attract and retain customers. It is critical that the models are able to track the dynamics of customer behavior over time. We propose FC-CSM, a Customer Segmentation Model based on a Fragmentation-Coagulation process, which can track the evolution of customer segmentation, including the splitting and merging of customer groups. We conduct a case study using transaction data from a major Australian supermarket chain, where we: 1) show that our model achieves high fitness of purchase rate, outperforming models using mixture of Poisson processes; 2) compare the impact of promotions on customers for different products; and 3) track how customer groups evolve over time and how individual customers shift across groups. Our model provides valuable information to stakeholders about the different types of customers, how they change purchase behavior, and which customers are more receptive to promotion campaigns.

## 1 Introduction

Customer behavior analysis is an essential component of business intelligence and marketing [Sheth *et al.*, 1999]. Constructing customer behavior models allows businesses to identify the customer segments that are most likely to buy their products and reach target customers effectively [Rossi *et al.*, 1996]. With a better understanding of their customers, it is easier for the stakeholders to develop cost- and time-efficient strategies, including promotions and tailored programs with social value.

While the traditional customer segmentation approaches are based on the demographic information, the recent approaches focus on identifying different types of customers in terms of purchase behavior. For example, mixture models [Bucklin *et al.*, 1998] can be used to discover latent groups in terms of customer behavior. In this way the purchase behavior of an individual customer can be interpreted as a mixture of different prototypes of behavior weighted by the membership in those groups.

However, the traditional mixture modeling is not suitable for behavior modeling in dynamic scenarios, e.g. for tracking a customer purchase behavior shifting across groups. For example, a customer Alice who bought potato chips regularly, may buy less of them after she embarked on a diet. Considering the group level, Alice may shift from Group 1 comprising customers who prefer snacks, to Group 2 comprising customers who prefer fruits. If all the other customers remain invariant, it is easy to track Alice's change by simply updating her group membership. However, other customers may also change their behavior at the same time, which leads to shifting the behavior of the entire group (prototype). Alice may still have a large or even unchanged membership of Group 1, if many other customers from Group 1 also shift to Group 2. This may provide misleading results to business management, who desire to know how the behavior of each customer changes across groups over time and which customers are more receptive to promotions.

A better way for dealing with the "group-level behavior shifting" problem is by adopting a split-merge approach. On one hand, a group can split into multiple smaller groups, when different types of behavior emerge in the group; on the other hand, the customers from two or more groups can merge into a large group, when they are affected by common factors, such as changes of seasons and price – which are of interest for the stakeholders. Due to the aforementioned limitations of using mixture models to address certain demands of the stakeholders, we propose a Bayesian nonparametric Customer Segmentation Model, named FC-CSM, which can track the varying customer groups via a Fragmentation-Coagulation Process (FCP).

The FCP is a random partition process, which can describe the split and merge of groups over time [Bertoin, 2006]. At each time interval, the model conducts fragmentation and coagulation operations on all the existing groups in the partition. Therefore, one group can split into smaller groups in the fragmentation step, when the members start to have divergent purchase behavior. On the contrary, several groups can merge in the coagulation step, when customers from different groups exhibit similar behavior. Another advantage of FCP is that, by configuring the split and merge rates, it can be employed to analyze individuals in a minor but interesting group of customers receptive to a promotion or with abnormal behavior, which is often desired by the stakeholders.

FC-CSM facilitates customer purchase behavior analysis and is capable of answering questions such as how many types of customers exist, when and how groups split and merge, how long a group lasts for, who are the customers in each group and how receptive they are to promotions.

We conduct a case study on a purchase dataset collected by an Australian national supermarket chain. We construct FC-CSM to explore the customer purchase behavior of various popular products and answer to the above practical questions. We first evaluate the fitness of FC-CSM and demonstrate that FC-CSM has achieved a higher accuracy than the mixture of Poisson processes. Then we examine the impact of promotions on purchase behavior and find that customers buying fresh products, soft drinks and confectionery are more receptive to the promotions. Most importantly, we illustrate how the customer segmentation for different products evolves over time, and track how individual customers shift across groups. The results of our work can be used by stakeholders to discover behavior change patterns that are worth noting, analyze the receptiveness of various customers to promotions, and identify target customers for the promotions.

**Related Work.** Extensive research has been carried out to build stochastic models of purchase behavior [Wagner and Taudes, 1987], for overall customers [Ehrenberg, 1959; Trinh *et al.*, 2014] or individual customers [Kim *et al.*, 2014]. However, modeling all customers together may overlook their heterogeneity, while building individual models can overfit the sparse and noisy purchase events of an individual customer. The customer segmentation based on behavior via mixture models can overcome the above disadvantages [Kotler and Armstrong, 2010]. The mixture models can discover various types of purchase behavior to support group-level analysis [Bucklin *et al.*, 1998]. The models can capture the behavior changes by a dynamic temporal component in the group-level model [Iwata *et al.*, 2013; Luo *et al.*, 2016a; 2016b; Li *et al.*, 2011; 2015]. Another option is Hidden Markov Models (HMM), in which the possible states of an HMM correspond to all groups of the partition [Xing and Sohn, 2007]. Since an HMM needs to label the states, it suffers from the label switching problem, where each permutation of labels generates a mode in the posterior distribution, which may lead to non-identifiability [Jasra *et al.*, 2005].

The random partition process FCP can describe the evolution of partitions using the fragmentation and coagulation operations [Bertoin, 2006]. Moreover, as the group of FCP is determined by its members rather than an explicit group label, the FCP can avoid the label switching problem. The idea of FCP has been applied in different fields such as financial markets [Eguiluz and Zimmermann, 2000], evolving networks [Tadić and Rodgers, 2010], and genetic variation [Teh *et al.*, 2011; Elliott and Teh, 2012]. To the best of our knowledge, there are no reported works that track the evolution of customer groups and how individual customers shift across the groups via a split-merge approach.

# 2 Methodology

Given a transaction dataset, we obtain customer purchase behavior matrix $\mathbf{X}^{U\times T}$ of $U$ customers during $(0,T]$, where $x_{it}$ is the behavior data for customer $i$ during $(t-1,t]$. $\mathbf{X}^{U\times T}$ can represent different types of purchase behavior. For example, $x_{it}$ can be a binary indicator of whether customer $i$ purchased a product during $(t-1,t]$, or it can represent the number of purchase events in week $t$ (time unit is one week), or it can reflect a preference level for a product. Our research tasks are to model customer behavior based on $\mathbf{X}^{U\times T}$, track the evolution of customer groups and capture how individual customers shift across groups.

In this section, we describe the construction of FC-CSM from a generative perspective. We also explain how to use the Gibbs sampler to infer the parameters of our model.

## 2.1 Construction of FC-CSM

At time $t$, the set of all customers $A$ can be divided into $n_t$ non-empty and non-overlapping groups, whose union is $A$. These groups form a partition $\pi_t$ of $A$, where each group contains customers with similar behavior at $t$. We aim to build a sequence of partitions $\{\pi_t\}$ $(t\in\{1,\dots,T\})$, so that we know the customers in a group at any time and how customer groups evolve over time, such as whether they split into subgroups or, on the contrary, merge.

FC-CSM is defined in a generative way as follows:

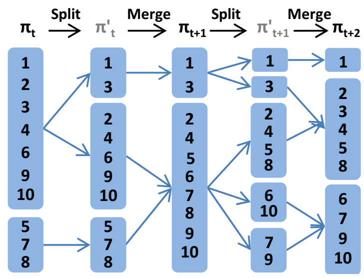
$$\begin{aligned}\pi_1 &\sim CRP(A,\rho,0)\\ \pi_{t+1}|\pi_t &\sim FCP(\pi_t,\rho,\delta)\\ x_{it}|\pi_t(i)=k &\sim Poisson(\lambda_{\pi_t(i)})\\ \lambda_k &\sim Gamma(\alpha,\beta)\end{aligned}\tag{1}$$

**Sample Sequence of Partitions:** We adopt the Chinese Restaurant Process (CRP) [Pitman, 2002], a nonparametric partition model, to describe the partitioning of all customers, that is, customer segmentation. CRP can avoid predefining the number of groups in the partition, which allows for a great flexibility of the model. From the perspective of the generative model, at $t=1$, we draw $\pi_1$ for customer set $A$ from $CRP(A,\rho,0)$, where $\rho$ $(\rho>0)$ is the parameter to control the number of groups in the partition.

We can obtain the following partition $\pi_{t+1}$ based on $FCP(\pi_t,\rho,\delta)$, where $\delta$ $(\delta\in[0,1))$ controls the temporal dependence between $\pi_t$ and $\pi_{t+1}$. When $\delta=0$, the partition fully depends on the previous one, which means the partition remains the same from $t$ to $t+1$; while $\delta\to 1$ means $\pi_t$ and $\pi_{t+1}$ are independent. Any group in $\pi_t$ can split into smaller groups and merge with the other groups to generate $\pi_{t+1}$.

Given $\pi_1$, the transition from $\pi_t$ to $\pi_{t+1}$ at $t\in\{1,\dots,T-1\}$ can be completed in two sub-steps: the first step splits all groups in $\pi_t$ into smaller groups and produces an intermediate partition $\pi'_t$, while the second step merges groups in $\pi'_t$ to produce $\pi_{t+1}$. For example, Figure 1 illustrates the sequence of partitions for 10 customers from $t$ to $t+2$ with different scenarios. The FCP transition rules are: 1) each group in $\pi'_t$ either exists in $\pi_t$ or is a subgroup of another group in $\pi_t$; 2) each group in $\pi_{t+1}$ either exists in $\pi'_t$ or is a combination of other groups in $\pi'_t$. These two steps correspond to the fragmentation (split) and coagulation (merge) operations, respectively. FCP assumes that the partition sequence $\{\pi_t\}$ $(t\in\{1,\dots,T\})$ satisfies Markov property. It means

Figure 1: Sequence of partitions for 10 customers from $t$ to $t+2$. Each blue block represents one group.




that given $\pi_t$, $\pi_{t+1}$ is conditionally independent on the partitions at the other time intervals. The fragmentation and coagulation steps do not affect the marginal distribution, so $P(\pi_t)$ at any $t$ is $CRP(A, \rho, 0)$, and the marginal distribution of intermediate partition $P(\pi'_t)$ is $CRP(A, \rho, \delta)$.

**Sample Behavior Data:** The partition of customers is dictated by the behavior metrics determined by the requirements of analysis. For instance, if we desire to distinguish customers by the number of times they buy a product fortnightly, the purchase behavior data $x_{it}$ stores the number of purchase events in two weeks. The Poisson distribution can be used to model the number of events in each period of time. Specifically, the parameter $\lambda$ of the Poisson distribution can be interpreted as the expected number of events per unit of time. Therefore, we draw $x_{it}$ from a Poisson distribution with intensity $\lambda_{\pi_t(i)}$ for customer $i$ in group $\pi_t(i)$. In addition, we use the conjugate prior, $Gamma(\alpha, \beta)$ for the Poisson distribution, where $\alpha$ is the shape parameter and $\beta$ is the scale parameter of the Gamma distribution.

The main advantage of our model is that we can customize the grouping of customers according to the requirements of the analysis. As mentioned above, the customer behavior can be defined in different ways, by storing different data in $\mathbf{X}^{U\times T}$. The changes would only affect the sampling of behavior data, which means that we could replace Gamma-Poisson by other distributions, but the sampling of partition sequence remains unchanged. For example, if the behavior data $x_{it}$ is the binary purchase indicator, $x_{it}$ could be sampled from a Bernoulli distribution with a Beta prior; if $x_{it}$ is a preference rating, $x_{it}$ could be sampled from a Multinomial distribution with a Dirichlet prior.

For each customer $i$, we can trace how they shift across groups and their purchase rate at any time based on the sequence of their group membership $\{\pi_t(i)\}$, where $\pi_t(i)$ is the group of customer $i$ at $t$. The purchase rate $\lambda_i(t)$ is defined as a stepwise function $\lambda_{\pi_t(i)}(t)$. Therefore, we can obtain purchase rate curve $\lambda_i(t)$ of customer $i$, which captures the changes of their purchase behavior, besides the main output of FC-CSM – partition sequence.

## 2.2 Inference of FC-CSM

In FC-CSM (Equation 1), CRP has exchangeable and projective properties, which means that: 1) the order of allocating customers to groups does not affect the marginal distribution $P(\pi)$ and 2) the projection of $\pi$ on a subset $A'$ of the complete set $A$ follows $CRP(A', \rho, 0)$. Since FCP is a Markov chain with a series of split and merge operations, it is also exchangeable and projective [Teh *et al.*, 2011]. These two properties allow us to infer the group allocation for customer $i$ based on the current partition of the other customers, assuming customer $i$ is the last customer who needs to be mapped to a group. Then we can update the sequence of group allocations for customer $i$, complying with the split and merge rules. In one sampling round, due to the exchangeability, we go through all customers to update their group allocations from $t=1$ to $t=T$, as if this customer is the last one to join the existing customer groups.

The notation used in the description is summarized as follows: $\pi_t^{\neg i}$ represents the projection of $\pi_t$ on $A$ excluding customer $i$; $z_t$ is the latent group index for customer $i$ at $t$; set $S_k$ is the selected group for the customer, with group index $k$; $|S_k|$ is the number of customers in $S_k$; $|\pi_t|$ is the number of groups in $\pi_t$. The variables with the prime ($'$) are for the intermediate partition $\pi'_t$.

The inference of FC-CSM proceeds as follows. Firstly, we describe the conditional probabilities of allocating groups for a behavior sequence $\mathbf{x}_i$ according to FCP. At $t=1$, the conditional probability of group allocation is:

$$P(z_1 = k|\pi_1^{\neg i}) = \begin{cases} |S_k|/(|A|-1+\rho) & \text{if } S_k \in \pi_t^{\neg i} \\ \rho/(|A|-1+\rho) & \text{if } S_k = \phi \end{cases} \quad (2)$$

where the condition $S_k \in \pi_t^{\neg i}$ in the first case means that the selected group for customer $i$ is $S_k$, an existing group in $\pi_t^{\neg i}$. The second case $S_k = \phi$ means that customer $i$ will start a new group. Therefore, the possible space for $S_k$ is $\phi \cup \pi_t^{\neg i}$.

When $t > 1$, the conditional probabilities of selecting a group for customer $i$ in the two sub-steps are:

$$P(z'_t = k'|z_t = k, \pi_t^{\neg i}, \pi_t'^{\neg i}) = \begin{cases} 1 & \text{if } S_k = S_{k'} = \phi \\ \delta|F_t(S_k)|/|S_k| & \text{if } S_k \in \pi_t^{\neg i}, S_{k'} = \phi \\ (|S_{k'}|-\delta)/|S_k| & \text{if } S_k \in \pi_t^{\neg i}, S_{k'} \in F_t(S_k) \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$P(z_{t+1} = l|z'_t = k', \pi_{t+1}^{\neg i}, \pi_t'^{\neg i}) = \begin{cases} \rho/(\rho+\delta|\pi_t'^{\neg i}|) & \text{if } S_l = S_{k'} = \phi \\ \delta|C_t(S_l)|/(\rho+\delta|\pi_t'^{\neg i}|) & \text{if } S_l \in \pi_{t+1}^{\neg i}, S_{k'} = \phi \\ 1 & \text{if } S_l \in \pi_{t+1}^{\neg i}, S_{k'} \in C_t(S_l) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

These two equations are for the split and merge steps, respectively. Specifically, $F_t(S_k)$ refers to the set of groups in $\pi_t'^{\neg i}$ which are split from $S_k$; it is formally defined as $F_t(S_k) = \{B|B \in \pi_t'^{\neg i}, B \subseteq S_k, B \neq \phi\}$. Similarly, $C_t(S_l)$ refers to the set of groups in $\pi_t'^{\neg i}$ which are merged into $S_l$; it is defined as $C_t(S_l) = \{B|B \in \pi_t'^{\neg i}, B \subseteq S_l, B \neq \phi\}$.

Given the conditional probabilities, we use the forward-backward (F-B) algorithm [Frühwirth-Schnatter, 1994] to infer the partition sequence for all customers. The F-B algorithm is a commonly used inference framework for dynamic latent variable models. The posterior distributions used to sample the group for customer $i$ at $t=1$ is represented as:

$$P(z_1 = k|\mathbf{x}_i, \{\pi_\tau^{\neg i}\}_1^T, \{\pi_\tau'^{\neg i}\}_1^{T-1}) \quad (5)$$

where $\{\pi_\tau^{\neg i}\}_1^T$ refers to the sequence of partitions projected on $A \setminus \{i\}$ from $\tau = 1$ to $\tau = T$.

When $t > 1$, we sample $k'$ for the split step as follows:

$$\begin{aligned} & P(z'_t = k' | z_t = k, \mathbf{x}_i, \{\pi_\tau^{\neg i}\}_1^T, \{\pi_\tau'^{\neg i}\}_1^{T-1}) \quad (6) \\ \propto \quad & P(z'_t = k' | z_t = k, \pi_t^{\neg i}, \pi_t'^{\neg i}) \\ & \times P(\{x_{i\tau}\}_{t+1}^T | z'_t = k', \{\pi_\tau^{\neg i}\}_t^T, \{\pi_\tau'^{\neg i}\}_t^{T-1}) \end{aligned}$$

After that, we sample $l$ for the merge step as follows:

$$\begin{aligned} & P(z_{t+1} = l | z'_t = k', \mathbf{x}_i, \{\pi_\tau^{\neg i}\}_1^T, \{\pi_\tau'^{\neg i}\}_1^{T-1}) \quad (7) \\ \propto \quad & P(z_{t+1} = l | z'_t = k', \pi_{t+1}^{\neg i}, \pi_t'^{\neg i}) P(x_{i(t+1)} | z_{t+1} = l) \\ & \times P(\{x_{i\tau}\}_{t+2}^T | z_{t+1} = l, \{\pi_\tau^{\neg i}\}_{t+1}^T, \{\pi_\tau'^{\neg i}\}_{t+1}^{T-1}) \end{aligned}$$

To compute the posterior distributions in Equations 5 – 7, we need the conditional probabilities of observations after the current $t$ given the partition sequences $\{\pi_\tau^{\neg i}\}_t^T$ and $\{\pi_\tau'^{\neg i}\}_t^{T-1}$, which are called "messages". The messages for the split and merge steps are denoted as $mf$ and $mc$:

$$mf_t(k') = P(\{x_{i\tau}\}_{t+1}^T | z'_t = k', \{\pi_\tau^{\neg i}\}_t^T, \{\pi_\tau'^{\neg i}\}_t^{T-1}) \quad (8)$$

$$mc_t(k) = P(\{x_{i\tau}\}_{t+1}^T | z_t = k, \{\pi_\tau^{\neg i}\}_t^T, \{\pi_\tau'^{\neg i}\}_t^{T-1}) \quad (9)$$

Given $mc_T(k) = 1$ at $t = T$, the messages at other time intervals can be computed in a backward manner recursively.

The likelihood term $P(x_{it} | z_t = k)$ in the sampling and messages is determined by the selected distribution for $x_{it}$. Assuming $x_{it}$ has Poisson distribution, the likelihood of $x_{it}$ in group $k$ is defined as:

$$P(x_{it} | z_t = k) = \frac{\lambda_k^{x_{it}} e^{-\lambda_k}}{\lambda_k!} \quad (10)$$

Based on the partition sequence $\{\pi_t\}$ $(t \in \{1, \ldots, T\})$, we can estimate the intensity parameter $\lambda_k$ of the Gamma-Poisson distribution by maximum *a posteriori* (MAP):

$$\lambda_k = \begin{cases} \dfrac{\sum_{j \in S_k} x_{jt} + \alpha - 1}{|S_k| + (1/\beta)} & \text{if } S_k \in \pi_t^{\neg i} \\ \dfrac{\sum_{j \in A \setminus \{i\}} x_{jt} + \alpha - 1}{|A| - 1 + (1/\beta)} & \text{if } S_k = \phi \end{cases} \quad (11)$$

It means that when $S_k$ is an existing group in $\pi_t^{\neg i}$, we compute $\lambda_k$ based on the other customers in $S_k$. Otherwise, if $S_k$ is a new group without any customers, we compute the default $\lambda_k$ based on all the other customers in $A \setminus \{i\}$. When customizing our model to handle different types of behavior data, the only change in terms of inference is the computation of the likelihood term.

In a *sampling round*, for each customer, we first update the sequence of messages based on the current partition sequence using Equations 8 and 9. Then we sample new group allocations for the current customer based on Equations 5-7. After updating the group allocations of all the customers, we start the next sampling round and repeat this process until convergence. The output is the partition sequence of all customers.

To summarize, FC-CSM offers several notable advantages. First, it can track the *evolution of customer groups* and find the size, duration, ancestors and descendants of each customer group. Second, it can capture the *trajectories of customers* shifting across groups and the variations of individual purchase rates. Third, it is *customizable* and can analyze various types of purchase behavior.

# 3 Case Study

Our case study is based on a transaction dataset of an Australian national-wide supermarket chain, collected through the supermarket loyalty cards between January 1 and December 31, 2014. Each transaction contains a unique customer identifier, product metadata (id, category, brand and name), timestamp, purchased quantity and cost. There are 931 customers in this dataset. We select 38 most popular products based on the number of customers who bought these products at least 10 times during the observation period. There are 21 fresh products including fruits and vegetables and 17 other products such as soft drinks and biscuits.

The supermarket stakeholders desire to construct accurate models to understand the purchase behavior, including what types of customers they have and whether supermarket promotions are effective. They also want to track the dynamics of purchase behavior such as how the customer groups evolve over time and how individual customers shift across groups. Therefore, we conduct the case study aiming to: 1) evaluate the fitness of our model and compare it with two mixture models – Homogeneous Poisson Process (PP) and Non-Homogeneous Poisson Process (NHPP) [Luo *et al.*, 2016a; 2017]; 2) explore the impact of promotions on purchase behavior of customers for different products; 3) demonstrate the trajectories of customer partitions and how customer shifts across groups.

The parameters of our model are configured as: $\rho = 0.2$, $\delta = 0.1$, $\alpha = 2$, $\beta = 0.5$, and the unit of time is 14-day. We conducted 500 iterations of Gibbs sampling for each product to obtain the sequence of partitions.

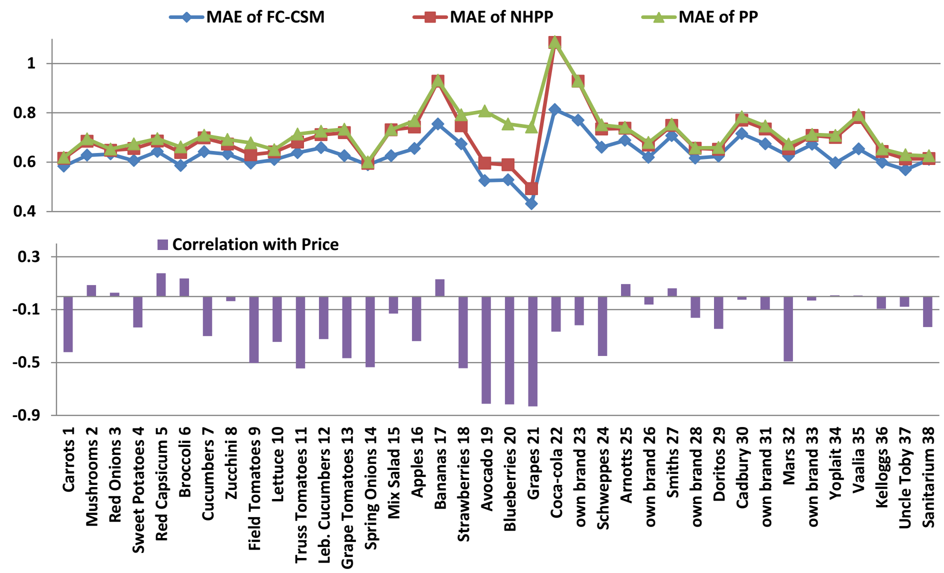
## 3.1 Fitness of Purchase Rate

We use the Mean Absolute Error (MAE) metric to evaluate the fitness of purchase rate. For each customer, we first compute the absolute difference between the estimated and actual number of purchase events for each interval, and average the values across $T$ intervals. Then, the MAE for a product is the mean absolute difference obtained for all customers, computed by $\sum_{i=1}^{|A|}(\sum_{t=1}^{T} |\lambda_i(t) - x_{it}|)/(|A|T)$.

We compare the model with: 1) the baseline PP, and 2) the NHPP, which uses a mixture of Poisson processes with a polynomial and periodic component in the intensity functions. Figure 2 shows that the MAE of our model is lower than the MAE of both PP and NHPP for all 38 products. The MAE of our model is 0.63, which is significantly lower than the MAE of PP, 0.73 ($p < 0.001$) and of NHPP, 0.70 ($p = 0.002$). On average, our model has decreased MAE of NHPP by 10%. We also notice that our model achieves greater improvements for fruits (14.35%), soft drinks (21.72%) and chilled desserts (13.93%), which jointly have an average improvement of 16.67%. For the other products, the average increase of fitness is 6.99%. The possible reason is that the change of seasons have a stronger impact on fruits, soft drinks and chilled desserts than on the other products. Although NHPP captures the long-term patterns via the polynomial component, the dynamic partition of our model is more flexible and can generate a lower MAE for these products.

Figure 2: Top: MAE of purchase rates of 38 products for FC-CSM, NHPP and PP. Bottom: Correlation between purchase rates computed by FC-CSM and price. Table: categories of products.

| Category | Products |
|---|---|
| Vegetables | 1-15 |
| Fruits | 16-21 |
| Soft Drinks | 22-24 |
| Biscuits | 25-26 |
| Snacks | 27-29 |
| Confectionery | 30-32 |
| Chilled Desserts | 33-35 |
| Cereal | 36-38 |




### 3.2 Impact of Promotions

We examine the impact of promotion via the correlation between the average purchase rate and the price. Negative correlation means the purchase rate is higher when the price decreases, so smaller correlation values indicate a larger impact of promotions.

The results are presented by the bar chart in Figure 2. Among all products, 20 correlation values are lower than -0.2, which means the customers buying these products are more receptive to promotions. At the category level, the average correlation of all fruits except for bananas is strongly negative, -0.67. The vegetables also have low correlation values, -0.23 on average. The prices of fruits and vegetables generally have more significant changes than other products, due to the varying product availability. It makes sense that the price of these fresh products impacts the purchase decisions. As for the other categories, the promotions of soft drinks and confectionery are generally more effective than of biscuits, snacks and chilled desserts. Thus, regular promotions with high discount rates of the soft drinks may play an important role in attracting customers.

### 3.3 Trajectory of Purchase Behavior Changes

Given the partition sequence for all customers, we first summarize the group and customer statistics. The total number of groups for different products over the observation period varies between 21 and 48. The average lifetime of one group is 2.15, which corresponds to 30.1 days, as the unit of time is 14-day. About 14% of the groups lasted for more than 8 weeks, which indicates that the partition changes at a moderate pace. The number of groups at each time interval varies from 1 to 6, and the average value is 2.57, which shows the dynamics of the group allocation. As for PP and NHPP, their numbers of groups remain at 3 over time.

We analyze the **trajectory of customer partition** for each product and illustrate the duration, size and purchase rate of all customer groups using the bubble plots in Figure 3. We select 3 representative plots as examples, corresponding to products with: 1) stable behavior (Arnott's biscuits, top), 2) behavior strongly affected by seasonal availability (grapes, middle) and 3) varying behavior, with customers dynamically shifting across groups (avocado, bottom).

The top plot for Arnott's biscuits shows that there are two different values of $\lambda$. About 25% customers have $\lambda$ above 1.5, while the other 75% have $\lambda$ below 0.7. We notice that there are not many shifts between the two levels, which means the purchase behavior is relatively stable.

The middle plot for grapes shows the impact of the seasonal product availability on the purchase behavior. There is a major group steadily including about 90% of customers. Overall, the $\lambda$ curve of this group has a U shape, which is the result of the high price and low supply in mid-year, when it is winter in Australia. Our model has also identified minor groups of customers (in contrast to mixture models which would have included these customers in the major group); this information can be used by stakeholders to discover customers who differ from the majority but are worth noting.

The bottom plot for avocado illustrates different responses to promotions from different customers. There is only one major group at the first 7 time intervals. The $\lambda$ values increased from $t = 8$ for all customers and the major group split into two streams, due to different responses to promotions. About 30% of the customers are more receptive to promotions and form groups with $\lambda$ values from 1.5 to 2.2. The remaining customers are less receptive to promotions and have $\lambda$ values between 0.5 and 1. Some customers shift from one level to the other. For example, at $t = 18$, most of the customers in Group 19 increased their purchase rate and formed

Figure 3: Trajectory of customer partitions based on the number of purchases. Each bubble represents a customer group with the id in it, and its size shows the proportion of customers in the group. The subplots are for Arnott's biscuits (top), grapes (middle) and avocado (bottom). The dashed lines highlight the time intervals discussed in the text.

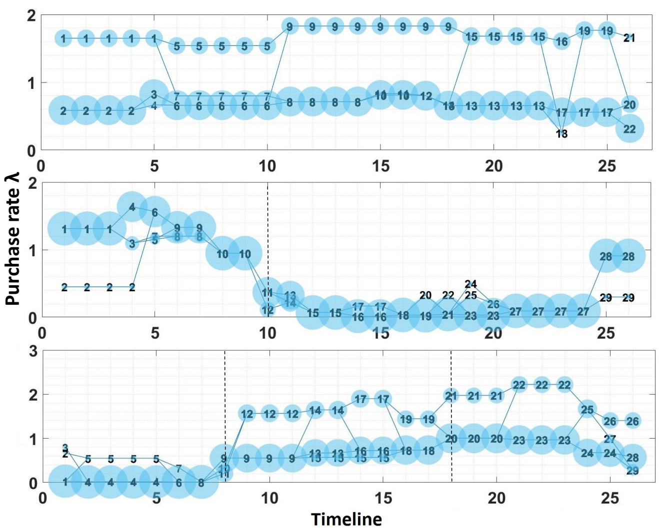
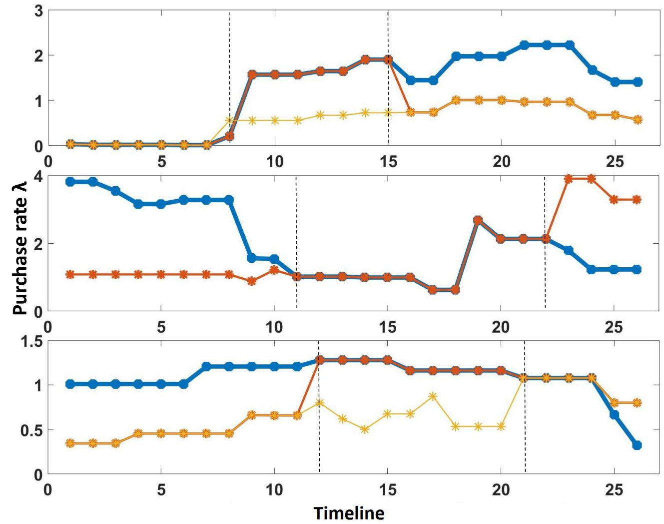



Figure 4: Individual purchase rate curves of different customers for avocado (top), Coca-Cola (middle) and broccoli (bottom). Each plot demonstrates contrasting behavior of multiple customers for one product, and each line corresponds to one customer. The dashed lines highlight the time intervals discussed in the text.




Group 21, while a small number of customers decreased $\lambda$ and joined Group 20.

For an **individual customer**, the purchase rate curve $\lambda_i(t)$ can track how they change behavior and shift across groups, which supports the fine-grained personal analysis. It also facilitates the comparison of customers in order to identify their similarities and differences. We present 3 sets of customers with contrasting behavior in Figure 4 (the group index has been omitted for clarity). When the lines with different colors join together, it means that the corresponding customers are in the same group during that period of time.

The top plot includes three customers buying avocado. The turning points of their purchase curves are at $t = 8$ and $t = 15$, when the price changed substantially. For example, at $t = 15$, customer 2 (red) decreased $\lambda$ and joined customer 3 (yellow), whereas customer 1 (blue) maintained her high purchase rate until the end. The middle plot displays two customers buying Coca-Cola who are in the same group for 22 weeks. However, customer 1 (blue) used to have significantly larger $\lambda$ than customer 2 (red) before $t = 11$. They split again after $t = 22$, when customer 1 decreased $\lambda$ gradually, whereas customer 2 increased $\lambda$ abruptly. It shows that customers with similar total number of purchases can have contrasting behavior changes. Similarly, the bottom plot contains three customers buying broccoli. The model successfully captures that customer 2 joined customer 1 at $t = 12$ and all of them joined the same group at $t = 21$.

## 4 Conclusions

In this paper, we propose the Bayesian customer segmentation model FC-CSM, which can track the evolution of customer groups using a fragmentation-coagulation process. The FC-CSM can split groups when customer behavior diverges and merge multiple groups when their behavior becomes similar. The model outputs the sequence of customer partitions over time, which can capture the dynamics of customer behavior and address the group-level behavior shifts. Moreover, FC-CSM is also flexible and can be applied to various types of behavior data.

Our case study uses a real-world supermarket transaction dataset. We analyze the evolution of customer groups and how it is affected by product promotions and seasonal changes. The case study shows that FC-CSM is more accurate than a mixture of Poisson processes. We find that customers were more receptive to promotions of fresh fruits, soft drinks and confectionery. The sequence of customer partitions shows that for products with high impact of promotions, customer groups tend to split during promotions and merge when the price stabilizes. FC-CSM also shows the trajectories of individual customers, which track how they shift across groups and how customers with similar number of purchases may have contrasting behavior patterns.

In summary, FC-CSM offers a strong and sensitive tool for analyzing customer purchase behavior and tracking the effects of promotions and seasonal campaigns. In future work, we will explore how to process the model outputs with a systematic and comprehensive procedure, so that we can provide analytical results to the stakeholders based on their needs, such as identifying interesting groups based on different criteria, and generating a list of target customers for marketing campaigns. We believe that the information provided by FC-CSM can be used by stakeholders in order to understand customer behavior changes, identify customer groups, and also optimize the timing and focus of promotion campaigns and business strategies.

## References


[Bertoin, 2006] Jean Bertoin. *Random Fragmentation and Coagulation Processes*, volume 102. Cambridge University Press, 2006.

[Bucklin *et al.*, 1998] Randolph E. Bucklin, Sunil Gupta, and Saurabh Siddarth. Determining segmentation in sales response across consumer purchase behaviors. *Journal of Marketing Research*, pages 189–197, 1998.

[Eguiluz and Zimmermann, 2000] Victor M. Eguiluz and Martin G. Zimmermann. Transmission of information and herd behavior: An application to financial markets. *Physical Review Letters*, 85(26):5659, 2000.

[Ehrenberg, 1959] Andrew S.C. Ehrenberg. The pattern of consumer purchases. *Applied Statistics*, pages 26–41, 1959.

[Elliott and Teh, 2012] Lloyd Elliott and Yee W. Teh. Scalable imputation of genetic data with a discrete fragmentation-coagulation process. In *Advances in Neural Information Processing Systems*, pages 2852–2860, 2012.

[Frühwirth-Schnatter, 1994] Sylvia Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2):183–202, 1994.

[Iwata *et al.*, 2013] Tomoharu Iwata, Amar Shah, and Zoubin Ghahramani. Discovering latent influence in online social activities via shared cascade Poisson processes. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 266–274. ACM, 2013.

[Jasra *et al.*, 2005] Ajay Jasra, Chris C. Holmes, and David A. Stephens. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, pages 50–67, 2005.

[Kim *et al.*, 2014] Hideaki Kim, Noriko Takaya, and Hiroshi Sawada. Tracking temporal dynamics of purchase decisions via hierarchical time-rescaling model. In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management*, pages 1389–1398. ACM, 2014.

[Kotler and Armstrong, 2010] Philip Kotler and Gary Armstrong. *Principles of Marketing*. Pearson Education, 2010.

[Li *et al.*, 2011] Bin Li, Xingquan Zhu, Ruijiang Li, Chengqi Zhang, Xiangyang Xue, and Xindong Wu. Cross-domain collaborative filtering over time. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pages 2293–2298. AAAI Press, 2011.

[Li *et al.*, 2015] Bin Li, Xingquan Zhu, Ruijiang Li, and Chengqi Zhang. Rating knowledge sharing in cross-domain collaborative filtering. *IEEE Transactions on Cybernetics*, 45(5):1054–1068, 2015.

[Luo *et al.*, 2016a] Ling Luo, Bin Li, Irena Koprinska, Shlomo Berkovsky, and Fang Chen. Discovering temporal purchase patterns with different responses to promotions. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*. ACM, 2016.

[Luo *et al.*, 2016b] Ling Luo, Bin Li, Irena Koprinska, Shlomo Berkovsky, and Fang Chen. Who will be affected by supermarket health programs? Tracking customer behavior changes via preference modeling. In *Pacific Asia Conference on Knowledge Discovery and Data Mining*, pages 527–539. Springer, 2016.

[Luo *et al.*, 2017] Ling Luo, Bin Li, Shlomo Berkovsky, Irena Koprinska, and Fang Chen. Online engagement for a healthier you: A case study of web-based supermarket health program. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 1053–1061. International World Wide Web Conferences Steering Committee, 2017.

[Pitman, 2002] Jim Pitman. Combinatorial stochastic processes. *Technical Report 621*, 2002.

[Rossi *et al.*, 1996] Peter E. Rossi, Robert E. McCulloch, and Greg M. Allenby. The value of purchase history data in target marketing. *Marketing Science*, 15(4):321–340, 1996.

[Sheth *et al.*, 1999] Jagdish N. Sheth, Banwari Mittal, and Bruce I. Newman. *Customer behavior: Consumer behavior and beyond*. Dryden Press Fort Worth, TX., 1999.

[Tadić and Rodgers, 2010] Bosiljka Tadić and G.J. Rodgers. Modelling conflicts with cluster dynamics in networks. *Physica A: Statistical Mechanics and its Applications*, 389(23):5495–5502, 2010.

[Teh *et al.*, 2011] Yee W. Teh, Charles Blundell, and Lloyd Elliott. Modelling genetic variations using fragmentation-coagulation processes. In *Advances in Neural Information Processing Systems*, pages 819–827, 2011.

[Trinh *et al.*, 2014] Giang Trinh, Cam Rungie, Malcolm Wright, Carl Driesener, and John Dawes. Predicting future purchases with the Poisson log-normal model. *Marketing Letters*, 25(2):219–234, 2014.

[Wagner and Taudes, 1987] Udo Wagner and Alfred Taudes. Stochastic models of consumer behaviour. *European Journal of Operational Research*, 29(1):1–23, 1987.

[Xing and Sohn, 2007] Eric P. Xing and Kyung-Ah Sohn. Hidden Markov Dirichlet process: Modeling genetic inference in open ancestral space. *Bayesian Analysis*, 2(3):501–527, 09 2007.