FCP Filter: A Dynamic Clustering-Prediction Framework for Customer Behavior

Yuanzhe Zhang[1]($\boxtimes$), Ling Luo[2], Yang Wang[3], and Zhiyong Wang[1]

[1] The University of Sydney, Sydney, NSW 2006, Australia
yzha9691@uni.sydney.edu.au, zhiyong.wang@sydney.edu.au
[2] The University of Melbourne, Melbourne, VIC 3010, Australia
ling.luo@unimelb.edu.au
[3] University of Technology Sydney, Sydney, NSW 2007, Australia
Yang.Wang@uts.edu.au

Abstract. Customer purchase behavior prediction plays an important role in modern retailing, but the performance of this task is often limited by the randomness of individual historic transaction data. In the meanwhile, Fragmentation and Coagulation Process (FCP), a stochastic partition model, has recently been proposed for identifying dynamic customer groups and modeling their purchase behavior. However, FCP is not able to forecast the purchase behavior because such a data-driven method requires transaction observations to conduct clustering. To tackle this challenge, we propose FCP filter, a clustering-prediction framework based on FCP, which can forecast purchase behavior and filter random noise of individual transaction data. In our model, FCP clusters customers into groups by their temporal interests to filter random noise of individual transaction data. Then a predictor is built on grouped data. The predicted results are also fed to FCP to adjust the parameter for prior knowledge at the next time step. Our model is superior in capturing temporal dynamics and having flexible number of groups. We conduct experiments on both synthetic and real-world datasets, demonstrating that our model is able to discover the latent group of individual customers and provides accurate predictions for dynamic purchase behavior.

Keywords: Dynamic clustering · Customer behavior · Prediction

1 Introduction

In modern retailing, accurate prediction on the quantity that customers are going to purchase over items helps retailers to design effective marketing and warehousing strategies. However, the purchase behavior of individual customer is often random, limiting the accuracy of prediction. For example, Tom normally purchases 1 bottle of milk in his weekly visit to supermarket, but may buy 2 bottles occasionally. The observation of purchase intensity on transaction data

is 1 or 2 but the real purchase intensity could be 1.2, so the observational noise is 0.2 and 0.8 for those two observations. In the above example, we can see that the purchase behavior of individual customer is not stable. To address this problem, we propose to cluster customers into groups by their historic transaction data because the purchase intensity of customer group is more stable and can represent the real purchase intensity of individuals. The random purchase noise of a customer can be filtered if his latent group could be accurately found.

For clustering customers into groups, a dynamic and flexible clustering model, Fragmentation and Coagulation Process (FCP) [7], has been recently proposed. FCP is a data-driven clustering model, with scalable number of customer groups, which does not need to be predefined, and can evolve with the data. This property enables FCP to capture the dynamic purchase behavior of customers accurately. However, FCP clustering can only be conducted when customer purchase data is given, which makes it hard to forecast the behavior of customer groups in the future. The significance of FCP in real-world applications is limited by the unavailability of future transaction data. For retailers, forecasting the behavior of customer groups is more important than just grouping customers in the past time.
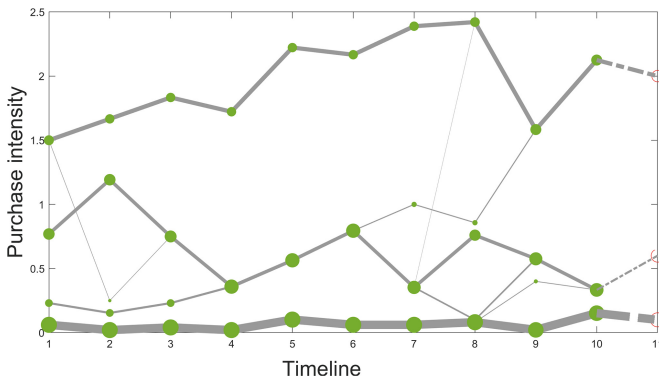
Fig. 1. The purchase intensities of customer groups evolve with time. Each bubble implies a customer group and the size of bubbles implies the size of groups. The line connecting two bubbles represents customers shifting between groups.

In order to reduce the purchase random noise of individual customers as well as to enable FCP to forecast group purchase intensity, we propose our FCP filter based on FCP to predict purchase intensity at group level instead of individual customer. For example, as shown in Fig. 1, there are 3 groups at $t = 10$. We aim to predict the purchase intensity of these 3 groups at $t = 11$ and use the predicted value of each group as the purchase intensity of its members. The purchase intensity of individual customer changes rapidly and randomly, but the group-based purchase intensity is more stable and easier to discover regular patterns. We can take the group-level purchase intensity as the actual state

of its members, while taking the individual customer purchase intensity as the observation of group purchase intensity. By predicting over actual states instead of observations, the individual randomness can be filtered.

Also, in our model, not only the clustering results influence prediction but also the prediction results influence FCP clustering. In traditional FCP clustering, there is a hyperparameter to represent the prior knowledge of group purchase intensity. Since our FCP filter can get the prediction of group purchase intensity, we propose to update this parameter in a time-evolving manner instead of using a predefined value. This parameter can be calculated from prediction results and influence FCP clustering at the next time step. Theoretically, an accurate prediction leads to better clustering fitness than fixing the parameter.

In summary, we construct a dynamical clustering-prediction framework for modeling customer behavior. The main contributions of our model are (1) from **prediction perspective**, this framework helps to filter individual random purchase noise, (2) from **clustering perspective**, we enable FCP, a data-driven clustering model, to forecast group purchase intensity. The flexibility and dynamics of our FCP filter are appropriate for modeling customer behavior. It is *flexible* that the number of groups do not need to be predefined but estimated from customer transaction data. It is *dynamic* that the customer membership and group number can change with time. The hyperparameter controlling the priori knowledge of group purchase intensity is also updated dynamically so that the group purchase intensity can be estimated more accurately.

2   Related Works

Clustering on customers is also known as customer segmentation, which aims to identify the customers whose purchase behavior is in the same manner [10]. In order to identify customer groups, the data-driven approaches based on clustering analysis are formal and reliable solution [3]. Decision tree [5] was used to segment customers using their demographic information. Clustering models like K-means [4] for static clustering and mixture model based on Non-homogenous Poisson process [6] for tracking dynamic group interests were also proposed. However, the preferences and interests of customers may also change over time. In order to track the customers' temporal shifting across groups, a novel Bayesian non-parametric customer segmentation model FC-CSM [7] based on a random partition process, Fragmentation and Coagulation Process (FCP) [1], was proposed. It achieves high accuracy in fitting individual purchase frequency. Besides modeling the dynamics of segmentation, another advantage of FC-CSM is the flexibility. There is no need to set the number of customer groups manually, which can be learned automatically from data directly. However, the FC-CSM relies on the observed transaction data, so that the clustering can only be conducted for the past time. It is more meaningful to forecast the purchase behavior of groups instead of only analyzing past data. In this way, we propose to build prediction on FCP. Due to the efficiency of FCP to identify latent groups and model purchase behavior, the prediction could be more accurate than individual prediction.

3   Methodology

Our problem can be formally described as follows. Given the transaction data of a product, a matrix $\mathbf{X}^{U \times T}$ is generated to record the transaction quantity, for U customers during T time steps. Each entry x_{it} in $\mathbf{X}^{U \times T}$ refers to the purchase quantity of customer i at time step t. The task is to forecast the purchase quantities of customers at the next time step $T + 1$, i.e. $\hat{\mathbf{\Lambda}}_{T+1}^{U \times 1}$ in which $\hat{\lambda}_{i(T+1)}$ means the predicted purchase quantity of customer i at time $T + 1$.

Overall, our model has three main components: (1) **customer segmentation** based on FCP; (2) **tracking model** to track group purchase intensity trajectory and (3) **predictor** to forecast group purchase intensity at next time step.

3.1   Customer Segmentation

We adopt Fragmentation and Coagulation Process (FCP)[1,7], a dynamic random partition model, to segment customers and capture dynamic interests of customers. The schematic diagram of FCP from time step t to $t+1$ is illustrated in Fig. 2. FCP contains two procedures: fragmentation and coagulation. Given the initial customer partition π_t, at the fragmentation step, each customer group can remain the same or be split into several subgroups, forming the intermediate partition π'_t. Then, at the following coagulation step, a group can remain the same or be merged with other groups, forming the new partition π_{t+1}. In this way, FCP can capture the evolution of customer segmentation from t to $t + 1$. Theoretically, FCP is flexible to model any change of segmentation, which means that the new segmentation can be totally different from the previous one.
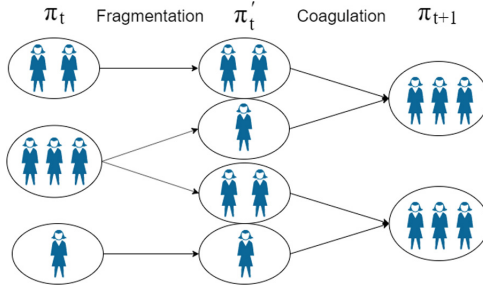

Fig. 2. Illustration of customer segmentation using FCP

Priori Probability of Customer Segmentation. FCP defines the priori transition probability from partition π_t to π_{t+1}. Formally, at $t = 1$, we adopt a random partition process Chinese Restaurant Process (CRP) [8] to model the initial partition of customers, and the probability of customer i in group g is:

$$P\left(\pi_1(i) = g | \pi_1^{-i}\right) = \begin{cases} |M_g| / (|U| - 1 + \rho) & \text{if } M_g \in \pi_t^{-i} \\ \rho / (|U| - 1 + \rho) & \text{if } M_g = \phi \end{cases} \tag{1}$$

where the hyperparameter ρ is to control the probability that the customer starts a new group, and M_g denotes the set of customer members in group g. π_t^{-i} is the partition of customers except for customer i at t, which assumes customer i is the last one who needs to allocate. In CRP model, the larger groups of a partition tends to attract more members and becomes larger.

Given partition and allocation at previous time step, for customer i, the transition probability from group g in the current partition to group g' in *fragmentation* step is defined as:

$$
P\left(\pi_t'(i) = g' | \pi_t(i) = g, \pi_t^{-i}, \pi_t'^{-i}\right)
$$
$$
= \begin{cases}
1, & \text{if } M_g = M_{g'} = \phi \\
\delta \left|F_t\left(M_g\right)\right| / \left|M_g\right|, & \text{if } M_g \in \pi_t^{-i}, M_{g'} = \phi \\
\left(\left|M_{g'}\right| - \delta\right) / \left|M_g\right|, & \text{if } M_g \in \pi_t^{-i}, M_{g'} \in F_t\left(M_g\right) \\
0, & \text{otherwise}
\end{cases} \tag{2}
$$

where $F_t\left(M_g\right)$ is formally defined as $\{B | B \in \pi_t'^{-i}, B \subseteq M_g, B \neq \phi\}$, which refers to the groups splitting from M_g. Equation (2) shows that a customer is more likely to join large groups splitting from M_g. The hyperparameter δ controls the probability that customer i starts a new group not existing in the previous group $\pi_t(i)$, which is also the temporal dependency of partitions between consecutive time steps.

Similarly, in the *coagulation* step, the transition probability of customer i joining group l from the intermediate group g' is:

$$
P\left(\pi_{t+1}(i) = l | \pi_t'(i) = g', \pi_{t+1}^{-i}, \pi_t'^{-i}\right)
$$
$$
= \begin{cases}
\rho / \left(\rho + \delta \left|\pi_t'^{-i}\right|\right), & \text{if } M_l = M_{g'} = \phi \\
\delta \left|C_t\left(M_l\right)\right| / \left(\rho + \delta \left|\pi_t'^{-i}\right|\right), & \text{if } M_l \in \pi_{t+1}^{-i}, M_{g'} = \phi \\
1, & \text{if } M_l \in \pi_{t+1}^{-i}, M_{g'} \in C_t\left(M_l\right) \\
0, & \text{otherwise}
\end{cases} \tag{3}
$$

where $C_t\left(M_l\right) = \{B | B \in \pi_{t+1}'^{-i}, B \subseteq M_l, B \neq \phi\}$ denotes the set of subgroups merged into M_l. The priori knowledge is that a customer is more likely to join the group that merged by more subgroups.

Likelihood of Observations. The individual purchase quantity is modeled by Poisson distribution. Given the purchase quantity of customer i at time step t, x_{it}, the likelihood of customer i belonging to group g at t is represented as follows:

$$
P\left(x_{it} | \pi_t(i) = g\right) = \frac{\lambda_g^{x_{it}} e^{-\lambda_g}}{x_{it}!} \tag{4}
$$

where the purchase intensity for customer group g is λ_g.

The purchase intensity has Gamma distribution as its prior, due to the conjugacy of Poisson and Gamma distributions. Therefore, we have the Maximum A Posteriori (MAP) of λ_g as follows:

$$
\lambda_g = \begin{cases} \dfrac{\sum\limits_{j \in M_g} x_{jt} + \alpha_t - 1}{|M_g| + \beta^{-1}} & \text{if } M_g \in \pi_t^{-i} \\[4mm] \dfrac{\sum\limits_{j \in U \setminus \{i\}} x_{jt} + \alpha_t - 1}{|U| - 1 + \beta^{-1}} & \text{if } M_g = \phi \end{cases} \tag{5}
$$

where the purchase intensity of a group can be interpreted as the average purchase quantities of its members, and impacted by the hyperparameters of α_t (i.e. shape parameter) and β (i.e. scale parameter) of the Gamma prior.

3.2   Tracking Purchase Intensity

For each customer i, we need to determine the purchase intensity trajectory $\{\lambda_{it}\}_1^T$ in order to predict for the future. An intuitive idea is to use the purchase intensity of the group that customer i belongs to along the time as the trajectory of purchase intensity, i.e. $\lambda_{it} = \lambda_{\pi_t(i)}$ for any t. However, the customer interests are evolving with time that the groups from the past may not fit the customers' current interests, and those $\lambda_{\pi_{T-n}(i)}$ may demonstrate misleading trends for prediction.

Therefore, we propose to predict their purchase intensity $\hat{\lambda}_{i(T+1)}$ only considering the current group membership, $\pi_T(i)$ and *backtrack the intensities of this group* in the past time steps, instead of using the actual groups the customers belonged to. The difficulty for tracking the purchase intensity of group $\pi_t(i)$ is that the group members could be totally different in consecutive time steps i.e. $M_{\pi_t(i)} \neq M_{\pi_{t-1}(i)}$. To address this problem, we build a *backward tracking model* to get the series of purchase intensities for the current group $M_{\pi_T(i)}$ in partition π_T.

Assume the group we are going to track is denoted as $g_{tracking}$ and the members of $g_{tracking}$ as $M_{g_{tracking}}$. The group $g_{tracking}$ is initialized as $\pi_T(i)$ for current time step $t = T$. If there exists $g \in \pi_{t-1}$ satisfying tracking rules (Eq. (6)), we update the group g as the new group to be tracked.

$$
g_{tracking} \Leftarrow g : \frac{|M_g \cap M_{g_{tracking}}|}{|M_{g_{tracking}}|} > \eta_1 \text{ and } \frac{|M_g \cap M_{g_{tracking}}|}{|M_g|} > \eta_2 \tag{6}
$$

In the tracking rules (Eq. (6)), we require that the majority of group g has shifted to group $g_{tracking}$ and the majority of group $g_{tracking}$ come from group g. The hyperparameter η_1 and η_2 are generally set as >0.5, so there could only be at most one or no tracked group. If there is no group $g \in \pi_{t-1}$ satisfying the tracking rules, $g_{tracking}$ remains the same:

$$
M_{g_{tracking}} = \begin{cases} M_g, & \text{if } \exists\, g \in \pi_{t-1}, s.t.\ g_{tracking} \Leftarrow g \\ M_{g_{tracking}}, & \text{otherwise} \end{cases} \tag{7}
$$

As to the individual purchase intensity, it is defined as follows based on whether there is a group g found:

$$\lambda_{i(t-1)} = \begin{cases} \lambda_g, & \text{if } \exists\, g \in \pi_{t-1}, s.t.\ g_{tracking} \Leftarrow g \\ \dfrac{\sum_{j \in M_{g_{tracking}}} \lambda_{\pi_{t-1}(j)}}{|M_{g_{tracking}}|}, & \text{otherwise} \end{cases} \quad (8)$$

If there is no tracked group g found, we use the average purchase intensity at $t-1$ of all members of group $g_{tracking}$ to represent tracked group intensity.

By computing backwards from $t = T$ to $t = 1$, we can finally get the trajectory of group purchase intensity $\{\lambda_{it}\}_1^T$ for customer i.

3.3   Predicting Purchase Intensity

Finally, the prediction model can be applied on the tracked purchase intensity trajectory $\{\lambda_{it}\}_1^T$ of customer i.

$$\hat{\lambda}_{i(T+1)} = Pred(\lambda_{i(T)}, \lambda_{i(T-1)}, \ldots, \lambda_{i1}) \quad (9)$$

In traditional FCP, the priori distribution of group purchase intensity is modeled by Gamma distribution with static predefined hyperparameter α_t and β in Eq. (5). Since our FCP filter can get the prediction of group purchase intensity, we propose to update this prior hyperparameter with the prediction results so that the priori knowledge of group purchase intensity could be more accurate. For Gamma distribution, we estimate the parameter α_{T+1} by Maximum Likelihood Estimation (MLE), taking the predicted group purchase intensity $\hat{\lambda}_g$ at time step $T+1$ as observations, and we have:
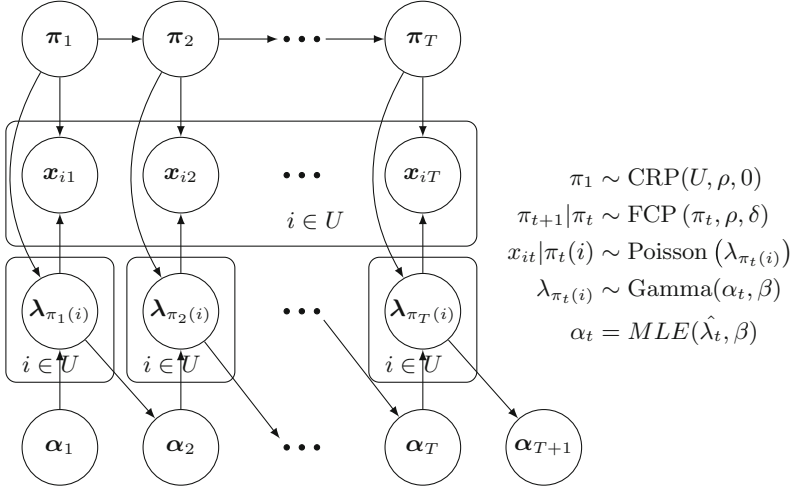
$$\alpha_{T+1} = \frac{\sum_{g \in \pi_T} \hat{\lambda}_g}{|\pi_T|\beta} \quad (10)$$

where $|\pi_T|$ is the total number of groups in the partition.

Our framework does not restrict the prediction models to use, and we have tested the performance of using the framework with various models including linear regressions and Long Short Term Memory (LSTM) in our experiments.

3.4   Graphical Model of the Framework

The generative graphical model of our FCP filter is shown in Fig. 3. The initial partition π_1 is sampled based on CRP rules and the partitions in following time steps obey FCP rules as described in Sect. 3.1. Given customer i belonging to group $\pi_t(i)$ at time t, the individual purchase intensity x_{it} is drawn from Poisson distribution with parameter $\lambda_{\pi_t(i)}$, which is the purchase intensity of the group he belongs to. The group purchase intensity $\lambda_{\pi_t(i)}$ at time t is drawn from Gamma distribution with hyperparameters α_t (i.e. shape parameter) and β (i.e. scale parameter). It is worth noting that α_t is dynamic, which means that different α_t at different time step t, that is different from original FCP. The parameter α_t is computed by using MLE of Gamma distribution with the predicted group purchase intensity $\hat{\lambda}_t$ and the scale parameter β as shown in Sect. 3.3.

$$\pi_1 \sim \mathrm{CRP}(U, \rho, 0)$$
$$\pi_{t+1}|\pi_t \sim \mathrm{FCP}\left(\pi_t, \rho, \delta\right)$$
$$x_{it}|\pi_t(i) \sim \mathrm{Poisson}\left(\lambda_{\pi_t(i)}\right)$$
$$\lambda_{\pi_t(i)} \sim \mathrm{Gamma}(\alpha_t, \beta)$$
$$\alpha_t = MLE(\hat{\lambda}_t, \beta)$$

Fig. 3. The generative graphical model of our FCP filter

3.5 Inference

The customer partition and allocation are inferred by sampling using the posterior transition probabilities, computed by Eqs. (11) and (12) based on the priori transition probabilities and the observation likelihood terms.

For the customer segmentation component, we use Gibbs sampler to infer the group membership of each customer over time $\pi_t(i)$. In more detail, since the FCP is exchangeable and projective [9], we assume that customer i is the last customer to be sampled, which means that we can allocate customer i given the allocation of all the other customers. According to Bayesian theorem, the sampling posterior transition probabilities for split and merge steps are defined respectively as:

$$
\begin{aligned}
&P\left(\pi'_t(i) = g'|\pi_t(i) = g, \mathbf{x}_i, \left\{\pi_\tau^{-i}\right\}_1^T, \left\{\pi_\tau'^{-i}\right\}_1^{T-1}\right) \\
&\propto P\left(\pi'_t(i) = g'|\pi_t(i) = g, \pi_t^{-i}, \pi_t'^{-i}\right) \\
&\times P\left(\left\{x_{i\tau}\right\}_{t+1}^T |\pi'_t(i) = g', \left\{\pi_\tau^{-i}\right\}_t^T, \left\{\pi_\tau'^{-i}\right\}_t^{T-1}\right)
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
&P\left(\pi_{t+1}(i) = l|\pi'_t(i) = g', \mathbf{x}_i, \left\{\pi_\tau^{-i}\right\}_1^T, \left\{\pi_\tau'^{-i}\right\}_1^{T-1}\right) \\
&\propto P\left(\pi_{t+1}(i) = l|\pi'_t(i) = g', \pi_{t+1}^{-i}, \pi_t'^{-i}\right) P\left(x_{i(t+1)}|\pi_{t+1}(i) = l\right) \\
&\times P\left(\left\{x_{i\tau}\right\}_{t+2}^T |\pi_{t+1}(i) = l, \left\{\pi_\tau^{-i}\right\}_{t+1}^T, \left\{\pi_\tau'^{-i}\right\}_{t+1}^{T-1}\right)
\end{aligned}
\tag{12}
$$

where the priori terms in the equations above can be calculated based on Eqs. (2) and (3) by forward and backward algorithm as used in Hidden Markov Model [2], with the likelihood terms given in Eq. (4).

In summary, the dynamic customer segmentation is firstly modeled by FCP. Then we build tracking model to get intensity trajectory of each latent group. After that, predictor can be used to predict the purchase intensity of tracked groups. Finally, the predicted results also influence FCP clustering at the next time step by updating α_t.

4   Experimental Results

We conducted experiments on synthetic and real-world datasets to illustrate that our model can (1) identify dynamic customer groups based on purchase behavior, (2) achieve more accurate prediction results by filtering individual random noise. The hyperparameters are empirically set using validation dataset as follows: $\rho = 0.8$, $\delta = 0.4$, $\eta_1 = \eta_2 = 0.65$, $\alpha_1 = 2$ and $\beta = 0.5$. The evaluation metrics in our study is the Mean Absolute Error (MAE). The MAE measures the average error between predicted purchase intensity and the ground truth.

4.1   Synthetic Dataset

We generate a synthetic dataset to demonstrate our model's capability to identify the latent group and customer shifting over groups. There are 40 products in the synthetic dataset. For each product, we generate purchase quantity $X^{100 \times 10}$ of 100 customers from 3 latent groups with 10 time steps.

Specifically, we firstly generate the group purchase intensity of those 3 groups at the first 5 time steps randomly $\Lambda^{3 \times 5}$. At each time step, we sort the group-level purchase intensities from large to small values, so that those 3 groups can show relevant purchase patterns continuously. To fill in the intensity matrix $\Lambda^{3 \times 10}$ of 10 time steps, λ_{gt} from $t = 6$ to $t = 10$ is generated by linear regression of 3 orders: $\lambda_{gt} = \sum_{n=1}^{3} a_n * \lambda_{g(t-n)} + b$. Then all customers are allocated into those 3 groups randomly at $t = 1$. We assume that a customer changes group membership over time with probability of 0.1, which means that the customers have 10% of chance shifting into another group. Finally, we generate customer purchase quantities using Poisson distribution with parameter $\lambda = \lambda_{\pi_t(i)}$ based on their allocation.

We test the predicting performance using FCP filter model and using individual records. As the purchase intensity of each group in our synthetic data evolves according to the rule of linear regression of 3 orders, the same regression predictor is used for both cases. Accurate prediction results could demonstrate the capability of our model to identify latent groups for customers. The results are shown in Fig. 4 comparing these two models. We can see that our FCP filter achieves lower MAE on almost all products. The average MAE over 40 products of individual prediction and our FCP filter are *5.58* and *3.34*, respectively. This means that FCP filter successfully tracked the evolving purchase intensity of latent groups in this dynamic dataset and predicted accurately.

To illustrate the flexibility of FCP filter, we also compared the average MAE using prediction models built on static K-means clustering with different number
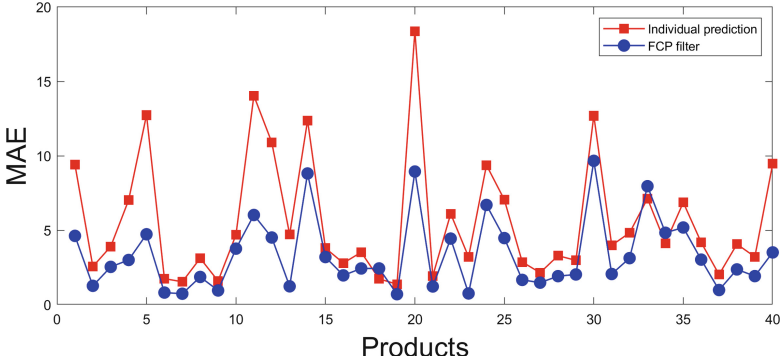
Fig. 4. Comparison of prediction results (MAE) using FCP filter and individual records on 40 synthetic products

of clusters (k). The average MAE results of K-means are *3.93* and *4.76* for $K = 3$ and $K = 5$, respectively, which are higher than FCP filter result (*3.34*). Our model outperformed static K-means clustering with $K = 3$ or 5, even when the ground truth for the number of clusters is 3. It shows that the importance of dynamics and flexibility of FCP filter in capturing the evolution of the purchase intensities. Moreover, there is no need to pre-define the number of clusters in our model.

4.2   Real-World Dataset

In this section, we use a real-world supermarket dataset[1] to illustrate our model's capability of filtering random purchase noise of individuals to get accurate prediction and usage of various predictors. The dataset contains 2,595,732 transaction records of 2,500 frequent customers on 2,383 products in 711 days (about two years). The transaction data is sparse in the first several months, so that we use the transaction data from 141 days to 420 days (40 weeks) for experiments. We divide 40 weeks into 10 time steps with 4 weeks in each time step. We select 24 popular products, which had the largest number of records and common in our daily life for experiments such as milk, cereal, eggs and so on. For each product, we discard customers who never bought that product and who ranked at the top 5 % based on purchase quantities as outliers. We randomly sample 100 customers for computational convenience and the purchase frequency is defined as the quantity purchased by a customer at one time step (4 weeks).

Several predictors are applied in our experiments to show FCP filter can generally achieve better prediction accuracy. They are LSTM network, 1-order and 3-order linear regressions, and a last-step predictor which takes the value at last time step as predicted value $\lambda_{i(T+1)} = \lambda_{\pi_T(i)}$. The average MAE is shown in Table 1. Similar to the results on the synthetic data, our FCP filter achieves the

---

[1] https://www.dunnhumby.com/careers/engineering/sourcefiles.

Table 1. Average MAE for FCP filter, individual and static K-means with different predictors

Average MAE	LSTM	Last-step predictor	Regression (1-order)	Regression (3-order)
Individual	1.05	1.06	1.16	2.27
K-means (K = 3)	1.18	1.04	1.10	1.77
K-means (K = 5)	1.16	1.02	1.10	4.29
FCP filter	**0.98**	**1.00**	**1.07**	**1.53**

best prediction accuracy with all the predictors. It is mainly because our dynamic model is suitable for modeling customers' dynamic interests and identifies the latent groups covered by random individual purchase behavior. We notice that the 3-order regression is not accurate and stable, and the possible reason could be that it is sensitive to the input purchase intensity series data.
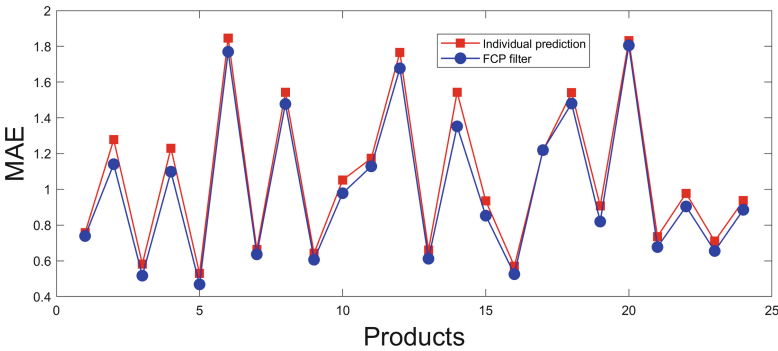
Fig. 5. Comparison of MAE for FCP filter and individual model with the last-step predictor on all products

Specially, Fig. 5 shows the prediction results of our model and individual model with the simple last-step predictor on all products. Our FCP filter gets higher prediction accuracy than individual prediction for every product. Since the predictor is quite simple, this result implies that the evolving customer purchase intensity is closer to group purchase intensity than individual one, and our FCP filter is able to find the latent group of customers and filter the noise in the individual records to produce accurate prediction results.

5   Conclusion

We build a dynamic and flexible clustering-prediction framework FCP filter to predict customer purchase intensity regardless of the random noise of individual

customer behavior. Our model segments customers by FCP and then predict customer purchase intensity on the tracked group purchase intensity. After that, the prediction result adjusts priori knowledge of clustering at next time step. We conduct experiments on both synthetic and real-world datasets, and show that FCP filter model is able to (1) identify the latent group and track purchase intensity evolving trends of groups; (2) improve the accuracy of customer purchase intensity prediction. Our framework is scalable with the datasets, without the needs of defining the number of clusters and is flexible to work with different predictors. Generally, our proposed model is not restricted to the domain of customer behavior modeling. It is also useful for other sequential data containing subjects that shifting among latent groups. In our future work, our model will be built on other domains with sequential data to improve prediction accuracy.

References

1. Bertoin, J.: Random Fragmentation and Coagulation Processes. vol. 102. Cambridge University Press, Cambridge (2006)
2. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Heidelberg (2006)
3. Carnein, M., Trautmann, H.: Customer segmentation based on transactional data using stream clustering. In: Yang, Q., Zhou, Z.-H., Gong, Z., Zhang, M.-L., Huang, S.-J. (eds.) PAKDD 2019. LNCS (LNAI), vol. 11439, pp. 280–292. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-16148-4_22
4. Dennis, C., Marsland, D., Cockett, T.: Data mining for shopping centres-customer knowledge-management framework. J. Knowl. Manage. **5**(4), 368–374 (2001)
5. Kim, S.Y., Jung, T.S., Suh, E.H., Hwang, H.S.: Customer segmentation and strategy development based on customer lifetime value: a case study. Exp. Syst. Appl. **31**(1), 101–107 (2006)
6. Luo, L., Li, B., Koprinska, I., Berkovsky, S., Chen, F.: Discovering temporal purchase patterns with different responses to promotions. In: International on Conference on Information and Knowledge Management (2016)
7. Luo, L., Li, B., Koprinska, I., Berkovsky, S., Chen, F.: Tracking the evolution of customer purchase behavior segmentation via a fragmentation-coagulation process. In: International Joint Conference on Artificial Intelligence (2017)
8. Pitman, J.: Combinatorial stochastic processes. Technical Report (2002)
9. Teh, Y.W., Blundell, C., Elliott, L., Gatsby, T.: Modelling genetic variations with fragmentation-coagulation processes. In: Advances in Neural Information Processing Systems (2011)
10. Tuma, M.N., Decker, R., Scholz, S.W.: A survey of the challenges and pifalls of cluster analysis application in market segmentation. Int. J. Mark. Res. **53**(3), 391–414 (2011)