BehaviorTracker: Visual Analytics of Customer Switching Behavior in O2O Market

Yaru Du
School of Computer Science and Software Engineering
East of China Normal University
51164500203@stu.ecnu.edu.cn

Changbo Wang*
School of Computer Science and Software Engineering
East of China Normal University
cbwang@sei.ecnu.edu.cn

Chenhui Li
School of Computer Science and Software Engineering
East of China Normal University
chli@sei.ecnu.edu.cn

Hong Yin
Faculty of Economics and Management
East China Normal University
hyin@jjx.ecnu.edu.cn

ABSTRACT

Visualization of customer behavior is urgently needed for an increasing number of customer orders on O2O (online to offline) platform. Although many works have been done on visualizing customer opinion or customer click events of one store, visualizing customer switching behavior among stores is still challenging. The challenge is to show customer order records over time and structure the inter-connection among different stores when customer switching behavior happens. In this work, we focus on Takeout O2O service to present a novel visual analysis system for retailers focusing on customer switching behavior patterns. Firstly we define five customer segments based on switching behavior. Then this system enables temporal-spatial driver exploration for different segments through several interactive views. Moreover, in order to visualize inter-connection sequences, augmented streamgraph with the bundled parallel coordinates is proposed as one alternative technique to visualize temporal event sequences. Case studies through collaboration with domain experts also demonstrate the usefulness and effectiveness of this system in helping customer relationship management.

CCS CONCEPTS

• **Human-centered computing** → **Visualization analytics**; • **Information systems** → *E-commerce infrastructure*;

KEYWORDS

Visualization of Customer Switching Behavior, Visual Analytic System, Temporal Event Sequence, E-commerce

ACM Reference Format:
Yaru Du, Changbo Wang, Chenhui Li, and Hong Yin. 2018. BehaviorTracker: Visual Analytics of Customer Switching Behavior in O2O Market. In *Proceedings of VINCI '18*. Vaxjo, Sweden, August 13-15, 2018, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

*Corresponding author

VINCI '18, August 13-15, 2018, Vaxjo, Sweden

ACM ISBN 978-1-4503-6501-7…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Takeout service [19] refers to prepared meals or other food items, purchased at a restaurant, that the customers intend to eat elsewhere. Driven by O2O e-commerce, Takeout O2O service develops rapidly as large numbers of online food orders rise. In such a big scale market, it is important to conduct customer relationship management (CRM) for locking in customers. Understanding and tracking customer behavior could help build profound, close and long-lasting customer relationship to retain and satisfy customers. Therefore, there is a growing need to extract customer behavior patterns from large collections of the O2O platform.

Extensive studies have been conducted on visual analysis of customer behavior. For example, Adobe Analytic has the basic visualization sales funnel chart (Figure 1a) that displays the proportion of customers from the beginning of online shopping up to the end of payment. Furthermore, visualization systems [9] [24] [5] [1] have been created to explore customer clickstream and customer feedback. However few of them present a visual analysis system of customer switching behavior among stores. Customer switching behavior [7] is a phenomenon that customers continue to use the service category (e.g., coffee service) but switch from one service provider (e.g. Starbucks) to another (e.g. Costa). In order to know where customers switch to and why they leave in the shadow area of Figure 1a, we focus on customer switching behavior between target store and other stores which have the same service category as target store (Figure 1b).

As one kind of temporal event sequence, order records are used to detect customer switching behavior patterns. However this process is complex and time-consuming caused by several characteristics of this temporal event sequence. First, order records in O2O service are disorganized with multi-situations. As switching costs from one online store to another become lower, customers may switch among stores casually which makes it difficult to identify critical switching behavior. Second, they have inner connections among different layers when switching behavior happens among different stores. Although a stacked graph is a good alternative to present temporal sequence, it emphasizes global features but not the connection between individual layers with detail information. Retailers prefer to see details (e.g. when they switch and where they switch to) of switching behavior to help them explore switching factors. Finally, they have multiple attributes which make color

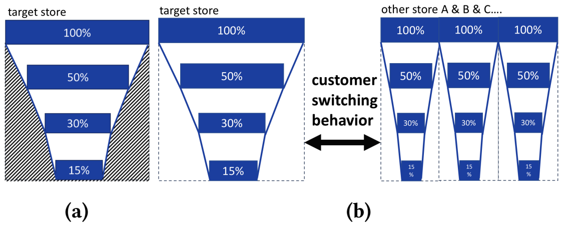


Figure 1: (a) Shadow area is lost customers. We don't know where they switch to. (b) In order to know the shadow area, we focus on customer switching behavior between target store and other stores with the same service.

encoding quite difficult to present visual cues. Therefore the visualization system has to be carefully designed for this temporal event sequence to combine customer-centered factors with store-centered factors, both include many facets. We design and develop BehaviorTracker to address challenges above, which is a visual analytics system to analyze and track customer switching behavior towards multiple stores using sales data. By pre-grouping different types of customers, this system supports deep exploration of tracing how and why customers change their buying behavior across stores over time. BehaviorTracker targets the tracking and analysis of customer switching behavior among stores in Takeout O2O service, but these techniques can be easily applied to the visual analysis of customer behavior on other O2O services, because the characteristics of switching behavior are shared across different fields.

Our work makes three contributions as follows:

- a new pipeline of identifying and visualizing switching behavior patterns based on O2O services.
- an interactive visualization system to enable analysts to better understand customer switching behavior in online stores.
- an augmented streamgraph with bundled parallel coordinates to show inter-connections of temporal event sequences.

2 RELATED WORK

2.1 Customer switching behavior

Customer switching behavior is one of the important factors that affect service companies' market share and profitability [7]. Keaveney et al. [8] proposed attitudinal, behavioral, and demographic factors to explore switching behavior of online service customers. [2] applied push, pull, and moorings (PPM) migration framework to explain customer switching behavior. Ghazali et al. [4] examined perceived switching costs and perceived attractiveness of alternatives as switching barriers on customer. With adopting aforementioned study, we took these factors into consideration for our visual analysis. Furthermore, a recent study made effort on making customer segments by extracting customer behavior [13]. Although the method of extracting behavior in [13]'s work is similar with the one in our work, we focus on proposing a new pipeline to identify different patterns through visualizing switching behavior on temporal-spatial dimensions.

2.2 Visual analysis of e-commerce data

As the rapid development of e-commerce, transforming e-commerce through data mining is very necessary. The data of customer feedback and customer clickstream can reveal behavior. OpinionSeer [24] is an effective system to compare customer opinions of different groups through visualizing online hotel customer feedback. Hao et al. [5] provided geo-temporal term associations to visually analyze the sentiment of customer feedback stream. Marques et al. [12] leveraged customer clickstream to identify customer profile and Lee et al. [9] provided a parallel coordinate visualization for online clickstream analysis. Brainerd et al. [1] first proposed an interactive, scalable visualization tool for analyzing customer behavior of buying products on the website. Aforementioned works focused on identifying customer-centered factors which are internal factors in one store. However, most of them ignore the external factors in different stores (e.g. competitors influence) [26], which can be evaluated through customer switching behavior among stores. Yaeli et al. [25] from IBM tried to understand customer switching behavior in different departmental stores by mapping customer inter-store movement into online customer browsing behavior. But there is still a big gap between online and offline shopping mode when analysing switching behavior. To our knowledge, interactive visualization system of tracking and analyzing customer switching behavior among online stores remains lacking.

2.3 Visualization of temporal event sequences

Due to increasingly large and multivariate temporal event sequences, a vast number of techniques have been developed. In order to visualize these high-dimensional sequences of multivariate events, Liu et al. [11] provided an analytic pipeline and effective method to extract sequential patterns and present them. Some studies focused on presenting relationship among different sequences. MatrixWave [21] explored visualization techniques for comparison of two sequences. PieceStack [23] were developed to understand intrinsic details of individual layers in stacked graph. Structuring the inter-connections among different time points is also essential to analyze different sequences. Wongsuphasawat and Gotz [22] proposed Outflow visualization technique to allow users to explore external factors that correlate with specific transitions in sports events. LoyalTracker [17] showed a visual summary to track user loyalty when users switch search engines over time. GameLifeVis [10] effectively illustrated the evolving behavior of massive game players as a sequence of time-oriented transitions. However, these methods are inappropriate to explore factors of customer switching behavior. Because this temporal event sequnece is disorganized with multi-situations but using systems above cannot effectively show detailed visualization of these situations. We explain our visualization work in detail in Sec. 6.2.

3 DATA AND TASK ABSTRACTION

3.1 Customer order data

Dianping.com [1] is a Chinese group buying website for locally found consumer products and retail services, which is the world's largest online and on-demand take-away platform. We collected customer

[1] http://www.dianping.com/

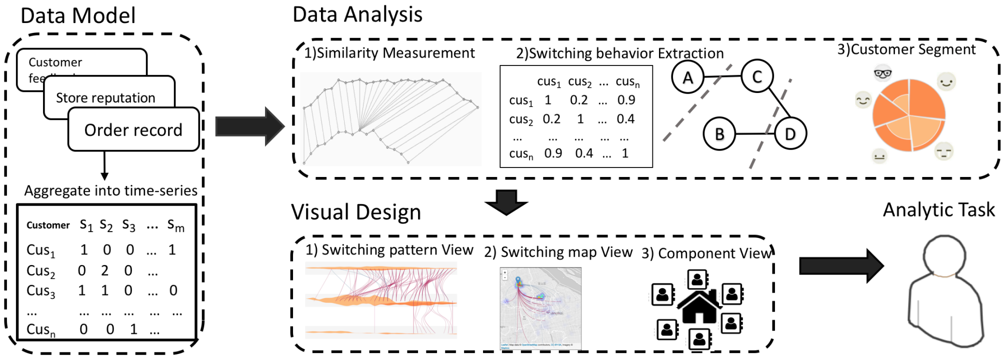


Figure 2: Overview of BehaviorTracker System

reviews from this webiste as our data samples. The data can be divided into three parts: order records, store data, and customer data. Order records contains basic order information including order time, order store, and customer feedback, while store and customer data include store reputation and customer sentiment on different features (e.g. delivery, service, and tasty). It is worth mentioning that click fraud [20] occurring for online advertising may be hidden in these order data. Click fraudsters take orders in different stores for dispersion instead of taking bulks of orders in one store to prevent their fraudulent clicks from detection.

3.2 Analytical questions

To better understand the problem domain and identify the potential uses of switching behavior patterns, we worked closely with one business management scientist *M* and several retailers *Ri* who have online stores. We interviewed retailers first to collect actual problems they concern about when launching a store. Then we exploded these collected problems to expert *M* who can give advice from the business management perspective. Through a series of interviews with our experts, we refined design requirements, presented prototypes, and collected feedback to improve the system iteratively. The analytical tasks are described below.

Q1 Is there any customer group having same switching behavior? What kind of customers are they?
Q2 How do customers behave over time who have the same switching behavior?
Q3 How do customers behave in spatial dimension who have the same switching behavior?
Q4 Is there any potential driver for the appearance of such a customer group ?
Q5 Is a customer who took more than one order in same day abnormal? Is he a click fraudster? Or service-loyal customer?
Q6 Are other stores which customers switch to competitive to target store?

4 SYSTEM OVERVIEW

Figure 2 shows the system overview of BehaviorTracker. The input of the system is a set of online customer takeout order records with related customer feedback and store reputations from Dianping.com. Data model module preprocesses order records of each customer into time series with store identification (equation2). Data analysis module leverages Dynamic Time Warping (DTW) algorithm [18] to measure time series similarity and spectral clustering to recognize customer segments. Based on calculated similarities, spectral clustering is used to recognize customer segments. Visual design module shows customer switching behavior patterns through visualizing order details of the extracted segments in temporal and spatial dimensions. Related components including customer feedback and store reputation are also visualized for deeply exploring reasons of customer switching behavior.

5 DATA ANALYSIS

In this section, we introduce our analytical approaches used in the data analysis module (Figure 2) as a new pipeline of segmenting customers based on switching behavior. We first clear all customers who only have one order record because there is no switching feature. Then we process our data into temporal sequences with switching features. Finally, we provide an effective method for clustering customers based on switching behavior.

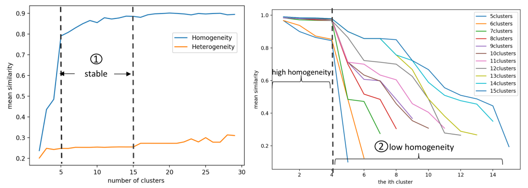
5.1 Data model

It is necessary to process the data for reflecting customer behavior. We focused on customers who took orders in target store and tracked their buying behavior in all other stores whose service category is same as target store's. Let

$$S_m = \{s_1^{(m)}, s_2^{(m)}, \ldots, s_i^{(m)}, \ldots s_k^{(m)}\} \tag{1}$$

denote ordering sequence of a customer with time steps, where $s_i^{(m)}$ represents the semantics that customer m takes the ith order. The equation 2 is defined to evaluate $s_i^{(m)}$. If a customer takes more than one order in one day, $s_i^{(m)}$ is evaluated as 2. If a customer takes one order in target store, $s_i^{(m)}$ is evaluated as 1. If a customer takes one order in one of other stores, $s_i^{(m)}$ is evaluated as 0. We seperately defined the condition customers take more than one order in one day because expert *M* commented that those customers seem abnormal for frequent daily orders of one service and maybe click fraudster [20] as we mentioned in Sec.3.1.

$$s_i^{(m)} = f(dt, pt)^{(m)} = \begin{cases} 2 & \text{if dt(i) = dt(i+1)} \\ 1 & \text{if dt(i)} \neq \text{dt(i+1)} \wedge \text{pt(i) = tid} \\ 0 & \text{if dt(i)} \neq \text{dt(i+1)} \wedge \text{pt(i)} \neq \text{tid} \end{cases} \tag{2}$$



(a) **(b)**

Figure 3: Cluster evaluation. (a) no.1 is a stable area with maximum homogeneity and minimum heterogeneity; (b) no.2 shows low homogeneity when there are more than 4 clusters.

where *dt(i)* and *pt(i)* denote the date and the store id of the ith record; and *tid* denotes the target store id.

As a result, each vector S_m represents a sequence with temporal shift. Aggregated dataset is shown in Data Model of Figure 2.

5.2 Switching behavior similarity and cluster

Similarity distance calculation is a critical component to the segmentation process. We adopted DTW to measure the distance between two order sequences S_i and S_j because it can effectively align time series [18]. The calculation is as follows:

$$\begin{aligned} D(S_i, S_j) = \min\{D(S_{i-1}, S_{j-1}), D(S_{i-1}, S_j), \\ D(S_i, S_{j-1})\} + C(S_i, S_j); \\ p \in [1, N], q \in [1, M] \end{aligned} \tag{3}$$

C is the local cost matrix representing all the pairwise distances between S_i and S_j.

Then we leveraged spectral cluster [14] method to segment customers based on switching behavior. We chose this graph-clustering technique instead of data-clustering like K-means because K-means easily fails when clusters do not correspond to convex regions. Spectral cluster algorithm takes steps as follows: Firstly, it is required to form a similarity matrix (equation 4) where the (i,j)-th entry is similarity distance you define in equation 3; Secondly, it is to construct Graph Laplacian from the similarity matrix (equation 4) and select k eigenvectors to define a k-dimensional subspace; Thirdly, clusters are formed in this subspace using K-means or any other algorithm.

The similarity matrix W_{n*n} was aggregated by

$$w_{(i,j)} = 1/(1 + DTW(S_i, S_j)),\ 0 \le i, j \le n,\ w_{(i,j)} \in W \tag{4}$$

Moreover, we evaluated clustering results through comparison of inter-cluster homogeneity and intra-cluster heterogeneity of similarity [13]. When measuring similarity, all items in the same cluster are desirable to minimize heterogeneity distance and maximum homogeneity distance. As Figure 3a shown, when the quantity of clusters is more than five and less than 15 (no.1), it is more stable with maximum homogeneity and minimum heterogeneity. Then, we measured mean similarity of each cluster when the quantity of clusters is more than 5 and less than 15 (no.1)(Figure 3b). Finally, we found clusters after the 4th cluster (no.2) are all with low homogeneity. We showed this experiment result to expert M and he told us switching behavior below 4 clusters has no regular pattern and it is meaningless to segment them separately. Therefore we chose 5 as the number of clusters including a no pattern cluster as the 5th cluster. This level of segmentation suits the managerial needs of local retailers. Every cluster is a specific type of customer switching behavior and we define them with emotion labels for better visualization in Figure 4.

type	description	label	homogeneity
stable	never switch		0.96
inflow	switch into target store from other stores		0.89
outflow	switch into other stores from target store		0.86
abnormal	more than one order in a day		0.84
no pattern	switch frequently among stores		0.11

Figure 4: Customer segment based on switching behavior

6 VISUAL DESIGN

Analyzing customer switching behavior interactively through visualization system is urgently needed. This is due to no existing visualization techniques working for tracking dynamic switching behavior patterns and customer switching drivers in one coherent view. Moreover, as end users like retailers do not have much background in information technology, a system should be designed more simply and intuitively. By iteratively presenting system prototype to retailers, we designed a set of visualization techniques as well as user-friendly interactions.

6.1 Data overview

Data overview (Figure 5A) with pie chart and matrix diagram is designed to illustrate customer segments based on switching behavior and customer characteristics (**Q1**). The pie chart shows customer proportion of different switching types. Emotion pictures (defined in Figure 4) around the pie chart reveal the switching behavior is/is not beneficial to the target store. For example, a crying emotion in Figure 5A intuitively notices end users should be highly vigilant because outflow happens in the target store. Matrix diagram shows customer categories based on customer characteristics. As Figure 5A shown, the horizontal direction represents the attribute of order time including three types(only weekend, weekend and workday, only workday). The vertical direction represents the attribute of distance from customer location to store location including two types (far, close). As a result, six categories are aggregated into six grids and the number of customers in each category is encoded with color shades in each grid. This segmentation can give end users an overview of customer distribution (e.g., customers in area 6 are most likely office workers because they only take target store orders at work days).

6.2 Switching pattern view

Switching pattern view (Figure 5B) aims at showing customer switching behavior in target store and other stores along with time (**Q2**).

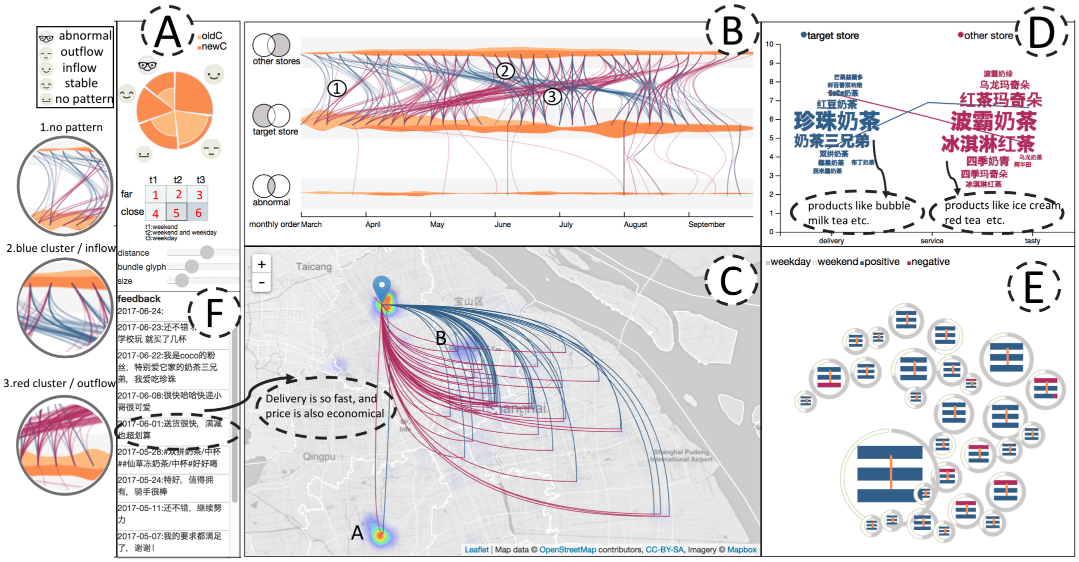
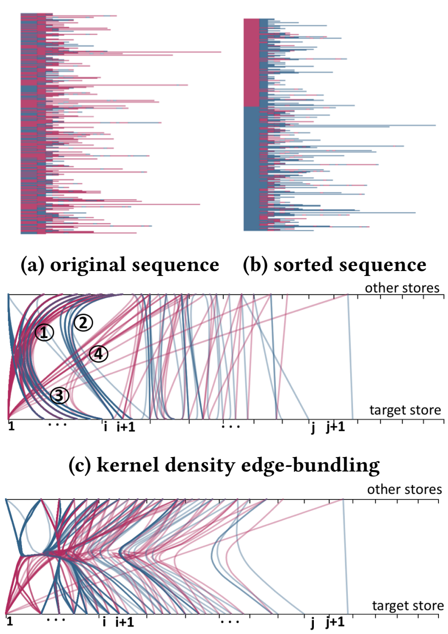


Figure 5: BehaviorTracker System View: (a) Data overview navigating different customer segments. (b) Switching Pattern View showing temporal patterns of different segments. (c) Switching Map View displaying with a radiance map to visualize spatial distribution of different segments. (d) Store View showing relative store reputation. Customer View containing (e) customer profile and (f) original customer feedback.



(a) original sequence **(b) sorted sequence**

(c) kernel density edge-bundling

(d) force-directed edge-bundling

Figure 6: Discussion of bundled parallel coordinatates

It combines streamgraph with bundled parallel coordinates to show inter-connections of temporal event sequences. Through this view, end users can explore temporal drivers of switching behavior they focus on (**Q4, Q5**).

Streamgraph for temporal event sequences. Three streamgraphs in Figure 5B represent three types of orders (other stores', target store and *abnormal* orders), which is coherent with equation 2. Each stream is stacked by one light yellow layer (old customers) and one dark yellow layer (new customers). Old or new customers are identified by their most recent order time. For example, if a specific time is defined as July, a customer who takes his most recent orders after July is a new customer, otherwise an old customer. Therefore, the new customer also may have order records before July. Besides, end users can define this specific time through slider bar.

Bundled parallel coordinates for inter-connection. Figure 6a shows order sequences of target store customers , much like rotated icicle plot. Each line starting from left to right represents an order sequence of one customer and it is color-coded by store categories (blue is target store and red is other stores). Visualizing raw sequences like this is not useful for recognizing switching behavior patterns from temporal data, which is also mentioned in [11]. Although we tried to sort sequences with order types (Figure 6b), it only can be seen most customers belong to inflow type because sequences with blue color are more than others. Therefore, to visualize order sequences with switching features between the target store and other stores, bundled parallel coordinates in Figure 6c is applied. It is motivated by edge-bundling layout for interactive parallel coordinates [15]. Two axises represent timeline with weeks interval and Bezier curves are leveraged for easily following with the eyes. These smooth curves represent switching behavior: 1) blue line represents flowing into target store, 2) red line represents flowing out from target store. Furthermore, some lines may be overlapped when two customers switch at the same period. As a result, lines are intensive with darker color in a specific area if switching behavior is quite frequent in some periods (e.g., no.3 & no.4 in Figure 8a on July and September). Edge bundling based on kernel density [15] is applied to intuitively focus on clusters with such frequent switching behavior. As Figure 6c shown, four clusters of frequent switching behavior can be presented in different time periods: $cluster1$ with red lines from 1 to i; $cluster2$ with blue lines from i to $i+1$; $cluster3$ with blue lines from 1 to i; $cluster4$ with red lines from 1 to $i+1$. At first, we also considered force-directed edge bundling [6] as one state-of-the-art alternative. However, it doesn't perform well in our experiment (Figure 6d) because lines are all evenly extended to the left-middle direction with ambiguous clusters. Force-directed edge bundling is more suitable for the data which are considered as one cluster.

However, expert M doubted the distorted encoding way of inter lines in Figure 6c. It is puzzling to measure switching frequency with numbers of lines while line is encoded as customer order in one specific store. Finally, we combined Streamgraph with bundled parallel coordinates to compensate for this drawback. As Figure 5B shown, this design was approved by expert M.

Personal switching behavior. To deeply explore one specific period end users are interested in(**Q4, Q5**), switching behavior with daily order level (Figure 9a) is provided by brushing the related month in the axis. Besides, the grey path with color-coded rectangular nodes is one order sequence with its related order satisfaction. To be consistent with the encoding of blue inflow and red outflow lines, blue rectangle represents satisfied sentiment and red one represents unsatisfied sentiment possibly causing outflow behavior.

6.3 Switching map view

Switching map view with Heatmap and Streetmap shows where customers switch to (**Q3**). Heatmap is to overview stores in the district level while Streetmap is to explore competitive stores in the street level (**Q6**). End users can toggle between these two freely and explore spatial factors (**Q4, Q5**) of switching behavior.

Heatmap. In Figure 5C, as it is the one-to-others relationship, we use a blue mark icon to represent target store location and heatmap schemed with rainbow colormaps to represent other stores where customers switch to. Rainbow colormap as one prevalent visualization tool can increase the amount of detail perceivable in the image. The area where customers take more orders in other stores is rendered by darker color (e.g., district A rendered in red). The name of a district is also shown when end users hovered the pointer over its related area on the map. If end users click the blue mark icon, *inflow/outflow* geo-condition will be shown with Bezier curves. The red curve is for outflow and the blue one is for inflow. The thickness of curve represents the number of customers who flow in/out.

Streetmap. As Figure 9b shown, each store location with a pie chart is shown on the street-level map. Each circle size of the pie chart represents the number of customers in one store. Blue/red part in the pie chart represents inflow/outflow proportion of this store. For example, in the store D, the red circle represents all customers who flow out from store D and come into target store. In the store E, the red pie represents a small part of customers who flow out from store E and come into target store while the blue pie represents another part of customers who flow into store E from target store.

6.4 Store view

As Figure5D shown, store reputation and popular dishes are clearly presented through line chart and word cloud. End users can understand store reputation directly through each line. The line is generated by three dots and each dot represents store performance of each aspect (delivery, service and tasty). When end users click one specific line, relative popular dishes will be presented with word cloud. The bigger the word is, the more popular its related dish is. An intuitive comparison between target store and the other store is also provided. Blue is encoded for the target store and red is for the other store. This is effective for end users to confirm competitiveness of other stores (**Q6**) as well as making business strategy in the next step.

6.5 Customer view

We designed Figure 5E to present numbers of adjustable customer glyph with force-directed layout.

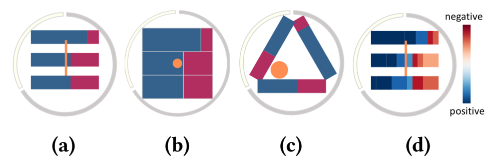
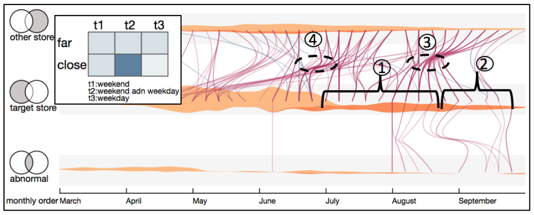


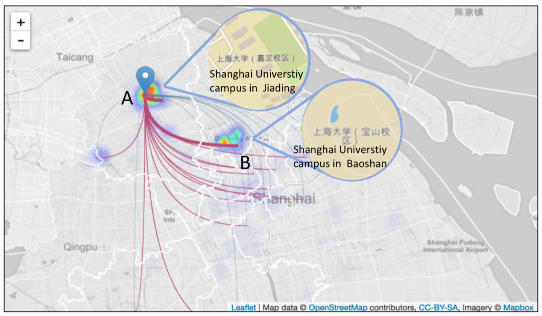
Figure 7: Glyph alternatives. a is the accepted one. a, b, c are all encoded with two-color sentiment but with three different layouts: rectangle, treemap, triangle; d is encoded with rectangle layout and multi-color sentiment.

Customer glyph. As Figure 7a shown, we designed a glyph which has three components to represent customer profile. (1) The outer part of the glyph represents the proportion of orders in the weekend with grey color and in the workdays with white color. The circle size is measured by the number of target store orders for a certian customer . (2) Three horizontal rectangles inside respectively represent feature sentiment in delivery, store service or tasty aspect. If the red band in a rectangle appears, it represents negative feedback in the related aspect. If the blue band appears, it represents positive feedback. This encoding is consistent with the view of Personal switching behavior in Figure 9a. (3) A vertical rectangle in the middle is designed to show whether a customer is an old or new customer, and color encoding is consistent with the pie chart in Figure 5A.

Glyph alternatives. We considered several design alternatives during the establishment of glyph prototypes. Initially, we applied a tree map to visualize feature sentiment (Figure 7b) and customer type is marked by a circle in the middle. However, end users always ignore the little shape in the middle which also may cover the tree map if it is enlarged. Then a triangle layout (Figure 7c)was considered inspired by OpinionSeer [24]. The circle is positioned by the weight of negative sentiment to attract end users to focus on the aspects with negative feedback. In Figure 7c, the inside circle is positioned in left-bottom because more negative feedback occurred



(a)



(b)

Figure 8: Geographic migration at outflow behavior (a) outflow cluster of no.4 causes the decrease of orders on July and August (no.1); outflow cluster of no.3 causes the decrease of orders on September (no.2) (b) a large outflow between district A and B which are both Shanghai University campus.

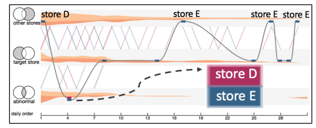
in left and bottom rectangles. However, this creative design wasn't accepted by expert M. He argued that the scattered layout with triangle make him hardly focus on the overview of sentiment in all aspects, although it is aesthetically pleasing. Finally, we designed multi-colored sentiment (Figure 7d) which has different levels of positive and negative sentiment (e.g., The sentiment in *"It took a very long time to arrive"* is more negative than in *"The delivery time is a little longer than before"*). But expert M commented this colorful band distracts end users from analyzing switching behavior and it is better with only two colors instead.

7 CASE STUDY

To show effectiveness of the system, we conducted tasks listed in Sec. 3.2 with the collaboration of retailers. We chose five top sale stores with special locations (campus or office building surrounded) as target stores to explore their customers' switching behavior. During this conduction, we found several patterns which are also evaluated by expert M from the business perspective. More can be found in our video as a supplementary material. [2]

7.1 Geographic migration at outflow behavior

When retailers were asked to explore the drivers of outflow behavior, retailer $R1$ who owns his store near an university is particularly concerned about the decrease of orders since July (Figure 8a). Retailer $R1$ said it is reasonable for the decrease in July and August



(a)

(b)

Figure 9: Normal switching behavior. As no.1 shown, when a customer was unsatisfied with the first order, it would result in taking another order in the same day.

(no.1 in Figure 8a) because students went home for the summer holiday and took orders in other places. However, he was confused why the number of orders didn't return back to the normal level as before when the new semester started in September (no.2 in Figure 8a). Furthermore, there are no negative feedback when he checked Customer View. Then he focused on the red area of outflow in August. By brushing August in the time axis, the related outflow/inflow condition of target store was showed in Switching Map View (Figure 8b). An obvious red line is crossed from the target store located in Jiading district to stores located in Baoshan district. It attracts retailer $R1's$ interest and he zoomed in to see the detail of Baoshan district. Finally he found there is a same campus called Shanghai University. Therefore, he deduced the outflow is caused by migration of Shanghai university students from one campus to another. To validate this conclusion, we interviewed some students at Shanghai University and confirmed the fact that senior students in Jiading campus are required to move to Baoshan Campus when a new semester starts. Traditional techniques such as line chart or funnel chart can not intuitively identify this potential driver.

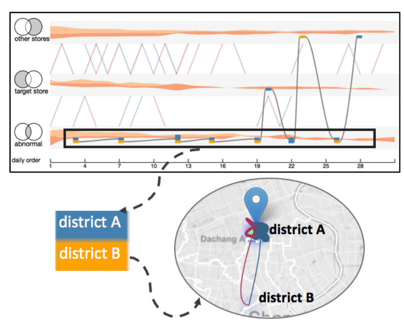


Figure 10: Anomaly in ordering behavior. The customer takes more than one orders in one day frequently and these orders are from two districts (A&B).

7.2 Anomaly in ordering behavior

Retailers also care about *abnormal* orders which are more than one order in the same day. There are not so many customers having such behavior so one of the retailers $R2$ decided to explore each customer separately for providing customized service. However, it

[2]The case studies video of the BehaviorTracker system proposed in this paper can be seen in https://vimeo.com/265327062

is not as abnormal as we considered because it is reasonable if a customer wasn't satisfied with the first order, it would result in taking another order. For example, as Figure 9a shown: The customer only had one *abnormal* order at the third streamgraph, which is one negative order at Store D (red rectangle) and one positive order at Store E (blue rectangle); Afterwards, this customer positively take turns to order at Store E and target store (blue rectangles at the first and second streamgraphs). Compared with this normal situation, Figure 10 shows one *abnormal* situation that the customer took two orders in one day frequently. Furthermore, these orders are located in district A and district B respectively with equal outflow and inflow (map in Figure 10). Expert *M* took two possible deductions for this phenomenon: (1) this customer is a service-loyal customer who prefers to enjoy service both in workplace and living place. And district A and B may be work and living places of this customer. (2) this customer is a click fraudster as we mentioned in Sec. 3.1.

8 CONCLUSION AND FUTURE WORK

In this paper, we have presented BehaviorTracker for analyzing customer switching behavior among stores under O2O e-commerce model. Through several interactive visualization views with coherent red-blue encoding, switching behavior patterns in each customer segment can be identified by inter clusters between temporal sequences. Multiple interactions are applied to flexibly explore potential drivers of switching behavior. We have also conducted case studies by using order records data of milk tea in Takeout O2O service. The results validate that BehaviorTracker is effective in customer switching behavior analysis and can be useful for visual analysis on other products or services under O2O model.

However, it still exists some drawbacks with increased scalability for BehaviorTracker system usage. For algorithm perspective, similarity computation in switching behavior takes several minutes. With growing order records, the algorithm should be optimized through fast DTW-based recognition [16]. From the visual design perspective, as we have only collected seven months order records in this work, time axis performs well to present this scale data. But when records are more than two years, a scalable timeline is essential for exploration in a long term rather than in a short term. Furthermore, for the number of other stores increasing, circles in Streetmap of Switching Map View may be overlapped if store locations are very close, which is unfriendly for visual analysis. This problem can be solved partially by applying force-directed techniques [3]. Therefore, we will optimize these works in the next step and aim at deploying our system on O2O platform to make it available for retailers.

ACKNOWLEDGMENTS

This paper is supported by Shanghai Science and Technology Development Fund Sofe Science Research Project under grant 1769210440.

REFERENCES

[1] J. Brainerd and B. Becker. 2001. Case study: e-commerce clickstream visualization. In *Information Visualization, 2001. INFOVIS 2001. IEEE Symposium on.* 153–156.

[2] Hsin Hsin Chang, Kit Hong Wang, and Shi Yu Li. 2017. Applying Push-Pull-Mooring to Investigate Channel Switching Behaviors: M-shopping Self-Efficacy and Switching Costs as Moderators. *Electronic Commerce Research and Applications* 24 (2017).

[3] Peter Eades and Mao Lin Huang. 2004. Navigating clustered graphs using force-directed methods. In *Graph Algorithms And Applications 2.* World Scientific, 191–215.

[4] Ezlika Ghazali, Nguyen Bang, Dilip S. Mutum, and Amrul Asraf Mohd-Any. 2016. Constructing online switching barriers: examining the effects of switching costs and alternative attractiveness on e-store loyalty in online pure-play retailers. *Electronic Markets* 26, 2 (2016), 1–15.

[5] M. C Hao, C Rohrdantz, H Janetzko, D. A Keim, U Dayal, L. E Haug, M Hsu, and F Stoffel. 2013. Visual sentiment analysis of customer feedback streams using geo-temporal term associations. *Information Visualization* 12, 3-4 (2013), 273–290.

[6] Danny Holten and Jarke J Van Wijk. 2009. Force-Directed Edge Bundling for Graph Visualization. In *Computer graphics forum*, Vol. 28. Wiley Online Library, 983–990.

[7] Susan M. Keaveney. 1995. Customer Switching Behavior in Service Industries: An Exploratory Study. *Journal of Marketing* 59, 2 (1995), 71–82.

[8] Susan M. Keaveney and Madhavan Parthasarathy. 2001. Customer switching behavior in online services: An exploratory study of the role of selected attitudinal, behavioral, and demographic factors. *Journal of the Academy of Marketing Science* 29, 4 (2001), 374.

[9] Juhnyoung Lee, Mark Podlaseck, Edith Schonberg, and Robert Hoch. 2001. Visualization and Analysis of Clickstream Data of Online Stores for Understanding Web Merchandising. *Data Mining and Knowledge Discovery* 5, 1-2 (2001), 59–84.

[10] Zhiqi Liu, Yandi Shen, Junhua Lu, Dingke Kong, Yinyin Chen, Jingxuan He, Shu Liu, Ye Qi, and Wei Chen. 2016. *Visual Exploration of Virtual Lives in Multiplayer Online Games.* Springer International Publishing.

[11] Z. Liu, Y. Wang, M Dontcheva, M Hoffman, S Walker, and A Wilson. 2017. Patterns and Sequences: Interactive Exploration of Clickstreams to Understand Common Visitor Paths. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 321–330.

[12] Orlando Marques, Alice|Belo. 2011. Discovering Student Web Usage Profiles Using Markov Chains. *Electronic Journal of e-Learning* 9 (2011), 63–74.

[13] Paul W. Murray, Bruno Agard, and Marco A. Barajas. 2017. Market segmentation through data mining: A method to extract behaviors from a noisy data set. *Computers and Industrial Engineering* 109 (2017), 233–252.

[14] Andrew Y Ng, Michael I Jordan, and Yair Weiss. 2002. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems.* 849–856.

[15] Gregorio Palmas, Myroslav Bachynskyi, Antti Oulasvirta, Hans Peter Seidel, and Tino Weinkauf. 2014. An edge-bundling layout for interactive parallel coordinates. In *Visualization Symposium (PacificVis), 2014 IEEE Pacific.* IEEE, 57–64.

[16] Yasushi Sakurai, Masatoshi Yoshikawa, and Christos Faloutsos. 2005. FTW: fast similarity search under the time warping distance. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems.* ACM, 326–337.

[17] C. Shi, Y. Wu, S. Liu, H. Zhou, and H. Qu. 2014. LoyalTracker: Visualizing Loyalty Dynamics in Search Engines. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1733–42.

[18] Xiaoyue Wang, Abdullah Mueen, Hui Ding, Goce Trajcevski, Peter Scheuermann, and Eamonn Keogh. 2013. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery* (2013), 1–35.

[19] wikipedia. 2004. https://en.wikipedia.org/wiki/Take-out. (2004).

[20] wikipedia. 2005. https://en.wikipedia.org/wiki/Clickfraud. (2005).

[21] Alan Wilson, Alan Wilson, Alan Wilson, Alan Wilson, and Alan Wilson. 2015. MatrixWave: Visual Comparison of Event Sequence Data. In *ACM Conference on Human Factors in Computing Systems.* 259–268.

[22] K. Wongsuphasawat and D. Gotz. 2012. Exploring Flow, Factors, and Outcomes of Temporal Event Sequences with the Outflow Visualization. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2659–2668.

[23] Tongshuang Wu, Yingcai Wu, Conglei Shi, Huamin Qu, and Weiwei Cui. 2016. PieceStack: Toward Better Understanding of Stacked Graphs. *IEEE Transactions on Visualization and Computer Graphics* 22, 6 (2016), 1640–1651.

[24] Yingcai Wu, Furu Wei, Shixia Liu, Norman Au, Weiwei Cui, Hong Zhou, and Huamin Qu. 2010. OpinionSeer: Interactive Visualization of Hotel Customer Feedback. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1109–1118.

[25] A Yaeli, P Bak, and G Feigenblat. 2014. Understanding customer behavior using indoor location analysis and visualization. *Ibm Journal of Research and Development* 58, 5/6 (2014), 3:1–3:12.

[26] Zhilin Yang and Robin T Peterson. 2010. Customer perceived value, satisfaction, and loyalty: The role of switching costs. *Psychology and Marketing* 21, 10 (2010), 799–822.