Association for Information Systems

AIS Electronic Library (AISeL)

WHICEB 2020 Proceedings

Wuhan International Conference on e-Business

Summer 7-5-2020

Exploring Customer Behavior Patterns: A Process-based Perspective

Wenxue Sun
School of Economics and Management, Beijing Jiaotong University, Beijing, 100044, China

Lei Huang
School of Economics and Management, Beijing Jiaotong University, Beijing, 100044, China, lhuang@bjtu.edu.cn

Ying Wang
School of Economics and Management, Beijing Jiaotong University, Beijing, 100044, China, ywang1@bjtu.edu.cn

Exploring Customer Behavior Patterns: A Process-based Perspective

Wenxue Sun, *Lei Huang*[1], *Ying Wang*[2*]
School of Economics and Management, Beijing Jiaotong University, Beijing, 100044, China

Abstract: With the increasingly fierce competition among enterprises, it is important for enterprises to understand customer behaviors accurately in a dynamic environment. While data mining methods have been applied to investigate customer behavior patterns with high-quality objective data, the process perspective has been largely neglected. Given that customer behaviors can be reflected in process event logs, it is possible to mine the real behavior patterns from a process-based perspective. To this aim, this paper presents a method for exploring customer behavior patterns using process mining techniques. The method consists of five steps: data collection and preprocessing, customer service process modeling, identifying deviant behaviors, clustering analysis and discovering customer behavior patterns. This method provides a viable way to understand the customer behavior patterns from a process-based perspective.

Keywords: process mining, customer behavior pattern, conformance checking, cluster analysis

1. INTRODUCTION

A deep understanding for customers is crucial in customer relationship management for achieving competitive advantage [1]. Since customer behaviors are highly dependent on the specific context, it is often difficult to explore behavior patterns for a huge number of customers [2]. Traditional customer behavior analysis with questionnaires or interviews may be time consuming and involves high costs, with only subjective data obtained [3]. Therefore, data mining techniques have been widely applied for the identification of customer behavior patterns with high-quality objective data (e.g. web visit logs)[4,5]. Currently, complex customer service processes are executed in large service companies for delivering quality customer services[6]. These business processes compose an important dimension for understanding customer relationship management[7]. However, few studies have examined customer behaviors from a process-based perspective.

Process mining has emerged as a set of new techniques, aiming at the automatic construction of models explaining the behaviors observed in the business process data[8]. Over the last decades many algorithms, techniques and tools for process mining have been developed[9-10]. In various areas including education[11], e-commerce[1,12], government[13], logistics[14] and manufacturing[15], process mining techniques have been applied for extracting process knowledge in support of process improvement and compliance. For understanding customer behaviors reflected in the service processes, nevertheless, there still lacks a comprehensive method that can be applied in real scenarios. Hence, the purpose of this paper is to answer the following research question: *how to learn customer behavior patterns from a process-based perspective*?

The paper is organized as follows. In Section 2 we will briefly outline related concepts and techniques concerning customer behavior analysis and process mining. Section 3 presents the method based on process mining techniques. Section 4 elaborates on the case study and Section 5 provides the results. Finally, Section 6 concludes the paper.

2. RELATED WORK

This section outlines related concepts and techniques concerning customer behavior analysis and process mining, which will be utilized in the remainder of the paper.

2.1 Customer behavior analysis

In many previous studies, customer behavior data was gathered with questionnaires and interviews[3].

* Corresponding author. Email: lhuang@bjtu.edu.cn, ywang1@bjtu.edu.cn

These studies obtaining subjective customer behavior data often involve high costs and consume long time. Therefore, data mining techniques based on objective data have been widely used in customer behavior analysis. For example, cluster analysis has been considered helpful for divide customers into similar variety of teams that take high intra similarity and low external similarity[16]. For another example, association rule mining has been widely used to find the patterns of customers' consumption behavior[17] and discern social relationships between users[18]. The information systems like CRM, HR as well as ERP generate process event logs, which provide the possibility to explore customer behavior patterns using process mining techniques. However, there is still a lack of effective methods from a process-based perspective to explore customer behavior patterns.

2.2 Process mining

Process mining is a new and emerging interdisciplinary field of data science and business process management. The basic idea of process mining is to discover, monitor and improve real processes by extracting knowledge from event logs[19].

In this paper, conformance checking, which is an important process mining technique, will be applied to identify deviant customer behaviors. Conformance checking aims at the detection and quantification of inconsistencies between a process model and its corresponding execution log [20]. Before performing the conformance checking, we need to pre-define the existing business process model (e.g. Petri Net model). The pre-defined process model will be compared with the real process executions reflected in the event log. As a result, the deviant customer behavior can be identified this way.

In addition, process cluster analysis[21] will be used in our study to categorize customer service process cases into similar groups. Subsequently, process discovery algorithms will be employed for each group of event log cases. In this paper, we will employ heuristic algorithm for it takes the frequency of the task order in the event logs into consideration and deals with the noise log data in a good way[22,23]. The discovered process models are able to provide illustrations for different customer behavior patterns.

3. THE METHOD

In this section, we propose a method for exploring customer behavior patterns using process mining techniques. The method consists of five steps: data collection and preprocessing, customer service process modeling, identifying deviant customer behaviors, cluster analysis for customer behavior patterns, and discovering customer behavior patterns in each cluster, as illustrated in Figure 1.

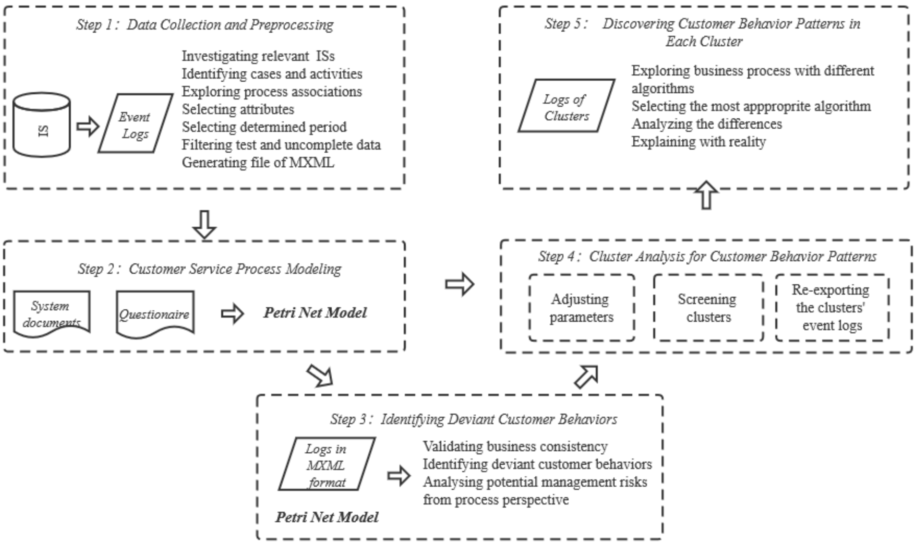

Figure 1. Comparison diagram of preprocessing

1) **Data collection and preprocessing:** Process data can be collected from multiple information systems which support the customer service. The obtained process data is then preprocessed to get the event log, containing information about customer service activities, their occurring time, and other activity context data.
2) **Customer service process modeling:** The obtained event log reflects the real process executions of customer service. This can be compared with the pre-defined process model. Thus, customer service process modeling is performed in this step to provide the illustration of pre-defined customer service process.
3) **Identifying deviant customer behaviors:** With the event log and the pre-defined process model for the customer service, conformance checking technique can be applied, comparing the event log with valid execution path specified by the process model. Deviant customer behaviors can be identified this way.
4) **Cluster analysis for customer behavior patterns:** Process cluster analysis[25], which categorizes customer service process cases into similar groups, will be performed in this step. Process cases within the same cluster demonstrate similar customer behavior patterns.
5) **Discovering customer behavior patterns in each cluster:** In this step, process discovery technique will be applied to the process cases in each cluster. Together with the comparative analysis of different clusters, the discovered process models illustrate different customer behavior patterns.

4. CASE STUDY

In this section, a case study is presented to illustrate our method. We describe the case scenario and event logs in our case study firstly, and then build the Petri Net model which will be used in conformance checking.

4.1 Case scenario

G port is an important comprehensive hub port in South China with more than 14,000 customers and 340,683 customer service process records. In this study, we collected process data from 1 Jan, 2018 to 31 Dec, 2019 for the case study. The complete customer service process includes four main sub-processes: documents handling, shipping, delivering and billing.

The process starts with the signing of contracts and ends with payment, including a total of 32 activities including ship plan, ship unloading, cargo lists checking, outbound, delivery and billing and so on. Table 1 shows that the logs contain the fields of case id (e.g. 'BBL02180117'), activity id, timestamp and resource, indicating the content, operation time and operator of activities respectively. Cargo name and customer record more detailed customer service information to support in-depth analysis of customer behavior patterns.

Table 1. An example of the extracted event log of the customer service process

Case ID	ACTIVITY ID	FORMDATE	FORMMAN	CARGO NAME	CUSTOMER
BBL0218017	Long-term contract	2018-01-08 00:00:00	Liu	Coil	Customer 1
BBL0218017	Ship plan	2018-04-11 09:24:51	Zhang	Coil	Customer 1
BBL0218017	Cargo list	2018-04-12 00:00:00	Qu	Coil	Customer 1
BBL0218017	Billing	2018-04-28 10:05:15	Lin	Coil	Customer 1

4.2 Modeling the customer service process

A process model was built describing the customer service process with Petri Net, as shown in Figure 3. In order to illustrate the relationship between processes and customers more clearly, we divided the main process into four sub-processes, ***documents handling, shipping, delivering and billing***. The relationship between processes and customers, as well as the relevant activities and documents of each sub-process are shown in Table 2. The ***documents handling*** sub-process reflects customer's habit of dealing with documents. And the ***shipping*** and ***delivering*** reflect the planned and actual quantity and frequency of cooperation, respectively. The last ***billing*** sub-process directly reflects the customer's consumption habits and economic capacity. To sum up, the event logs can reflect customer behaviors fully. And the more detailed illustration of each activity is shown in Table 3.

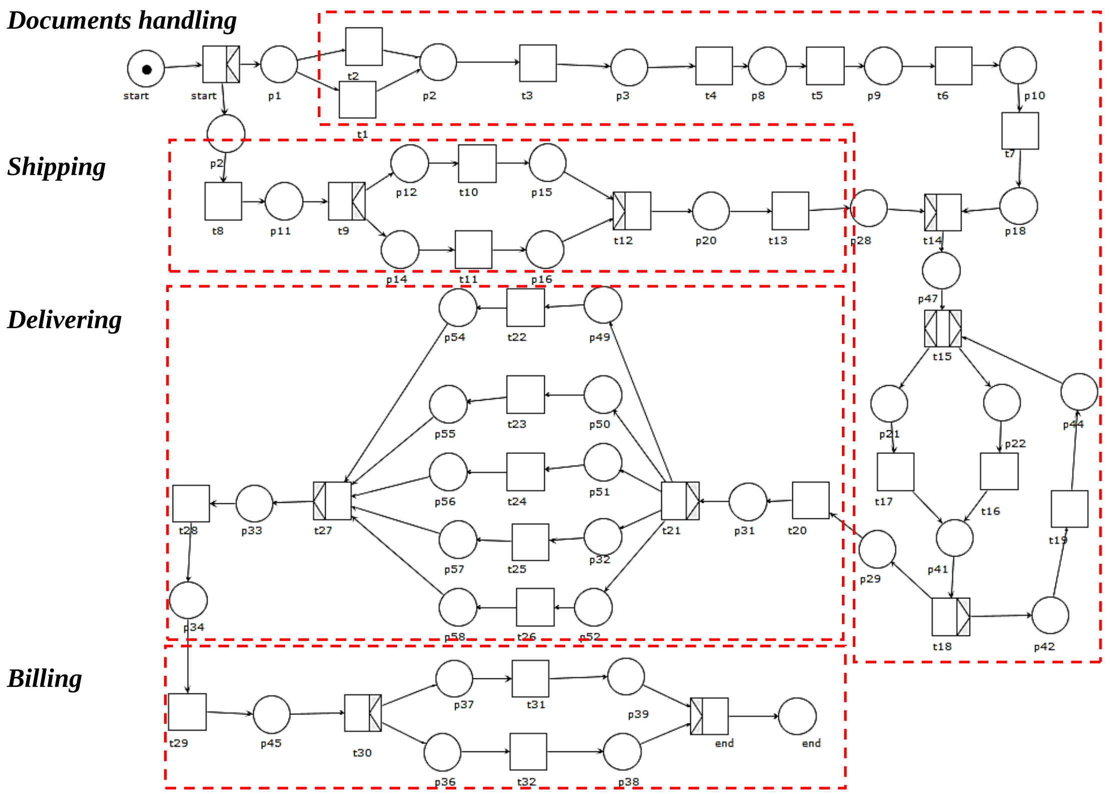

Figure 3. Petri Net model of pre-defined business process

Table 2. The relationship between sub-processes and customer behavior patterns

Sub-process	Relevant activities or documents	Operator	The reflected information of customers
Documents handling	Contract, Cargo-list, outbound order, prebill	Customer	Contract type, the order and type of handling documents
Shipping	Forecast, shift planning, day and night planning, ship unloading	Company	Planned operation quantity, frequency of cooperation
Delivering	Delivery card, loadometer records	Customer	Delivery habits, actual operation quantity
Billing	Billing, payment, invoice	Customer	Consumption habits, economic capacity

Table 3. The activities of the pre-defined business process model

Activity ID	Activity Name	Sub-process	Activity ID	Activity Name	Sub-process
T1	Long-term contract	Documents handling	T18	Outbound check	Documents handling
T2	Single ship contract	Documents handling	T19	Outbound sub-order	Documents handling
T3	Cargo list	Documents handling	T20	Loadometer	Delivering
T4	Cargo list check	Documents handling	T21	Delivery card	Delivering
T5	Cargo list sign	Documents handling	T22	Truck-weigh record	Delivering
T6	Cargo list claim	Documents handling	T23	Barge-weigh record	Delivering
T7	Outbound	Documents handling	T24	Barge-unweigh record	Delivering
T8	Ship forecast	Shipping	T25	Train-weigh record	Delivering
T9	Day and night plan	Shipping	T26	Train-unweigh record	Delivering
T10	Shift plan	Shipping	T27	Outbound complete	Delivering
T11	Ship berth	Shipping	T28	Outbound settlement	Delivering
T12	Ship unload	Shipping	T29	Billing	Billing
T13	Unload tally	Shipping	T30	Settlement handover	Billing

Activity ID	Activity Name	Sub-process	Activity ID	Activity Name	Sub-process
T16	Prebill	Documents handling	T31	Transfer income	Billing
T17	Bind prebill	Documents handling	T32	Invoice	Billing

5. RESULTS

Results are presented in this section in terms of the deviant customer behaviors and different customer behavior patterns.

5.1 Deviant customer behaviors analysis

Deviant customer behaviors were identified through conformance checking. After data collection and preprocessing, the event logs were exported into MXML forma. The pre-defined Petri Net model is another input source for conformance checking. The event logs were mapped with Petri Net model one by one to compare the consistency between the actual business process and the pre-defined process. Figure 4 shows the conformance checking results.

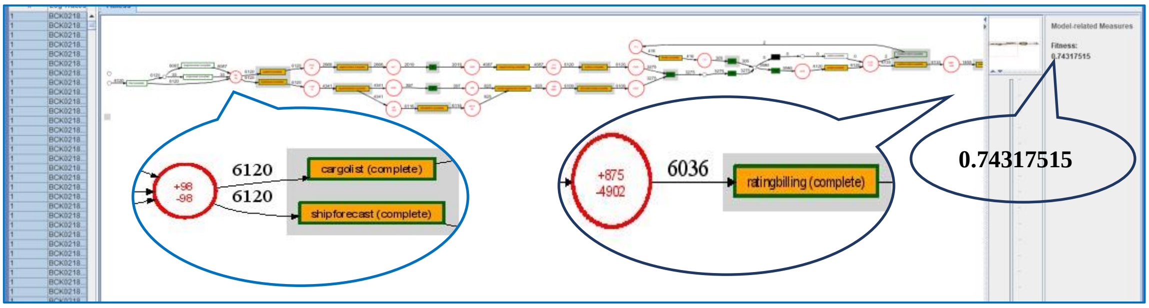

Figure 4. Screenshot of Conformance checking results

As can be seen in Figure 4, the fitness value is **0.74317515**, indicating a good fit and the actual customer behavior patterns contained in actual event logs is meaningful. However, there are still many missing tokens (identified by "-") and redundant tokens (identified by "+") when we-"replay" the pre-defined model using the real process execution event log[14] (see Figure 4). By continuing to analyze the number of missing and redundant tokens between each event, the deviant customer behaviors can be identified and should receive more attention from managers.

Table 3. Deviant Customer Behavior Identification

Activity	Ideal process model	Token (+/-)	Deviant Customer Behavior Identification
T29: Billing	The activity should be operated after the of the activity *outbound settlement* is completed	+875 -4902	80% of the delivery and billing phases were without settlement operations. The deviant is that may lead to inaccurate billing content, or the delivery operation still exists during billing phase. Thus, the deviant behavior associated with a group of customers should be corrected promptly to reduce the goods loss.
T3: Cargo list	The activity should be operated after signing the contract	-98	1.6% of cargo list documents were completed before the contract was signed. The deviant is that the goods occupy the resources of berth and warehouse after unloading. The delivering and billing phase can only start after the contract is signed. Thus, this deviation of customer behavior should be eliminated for increasing productivity.

5.2 Analysis of customer behavior patterns

Before further analysis, we first regrouped the pre-processed event logs according to customers, cargos to improve the performance of the clustering analysis. Event logs were clustered using a clustering algorithm integrated in ProM. In our study, we set the value of the parameter ‘max-diameter’ to 0.7, as repeated experiments showed that the boundaries of the clustering results are very clear with this parameter value. As a result, we got 10 clusters with at least 3 traces as shown in Figure 5. The clusters with more than 100 traces were then selected for further analysis.

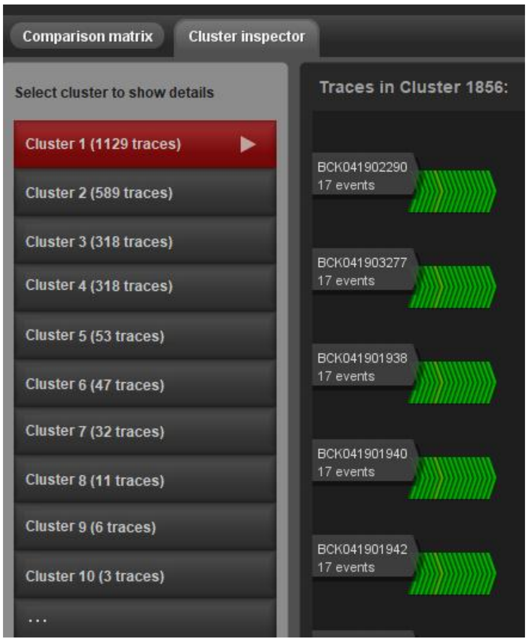

Figure 5. Screenshot of Cluster analysis results

Process discovery was then performed to explore the traces contained in each cluster. As shown in Figure 6, the process models discovered indicate significantly different customer behavior patterns.

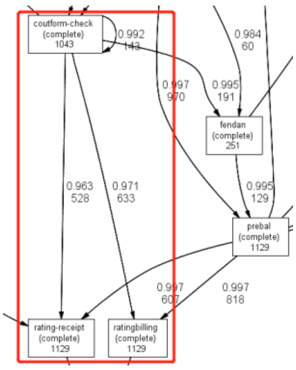

Figure 6 (a). Cluster 1

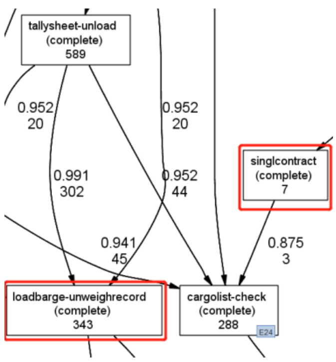

Figure 6 (b). Cluster 2

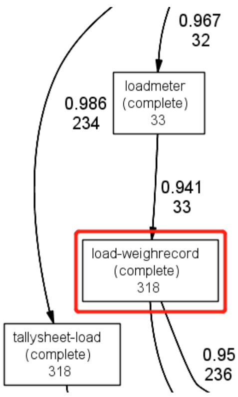

Figure 6 (c). Cluster 3

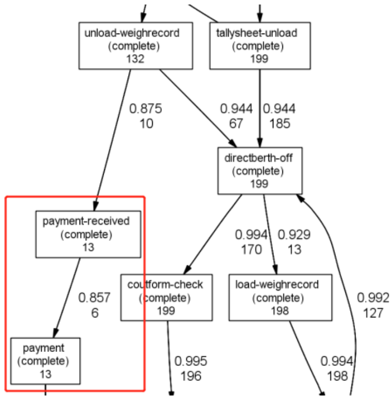

Figure 6 (d). Cluster 4

1) First, Figure 6 (a) reflects that customers in ***Cluster 1*** has no delivery process and goes directly into the billing phase after completing the documents. The cargos of the cluster are directly sent to the customers through pipelines or belts, without counting or weighing. The efficiency of the whole process can be improved as long as the efficiency of **documents handling** sub-process is improved.
2) Figure 6 (b) shows that customers in ***Cluster 2*** have signed more single ship contracts rather than long-term contracts, and most of the delivery way is *barge-unweigh record*. Through interviews with system personnel, we acknowledged that most contracts are long-term ones. Because the single-ship contracts need to be signed at the port site when the customers deal with documents in the selected systems.
3) Different from the previous cluster, almost all the customers of C***luster 3*** have chosen the delivery way of *truck-weigh record*. When the way of delivery is obtained in advance, the warehouse and weighbridge can be ready as soon as the ship plan is completed for high efficiency.
4) Surprisingly, unlike other customers who take the payment activity as the end of the main process, most customers of ***Cluster 4*** have completed the payment activity at the documents phase which is popular for enterprises. Managers can pay more attention to these customers and cooperate with them more often.

Thus, our study provides a viable method for exploring event logs generated by customers. Different customer behavior patterns can be effectively identified. For different customer clusters, further customer relationship management measures can be carried out accordingly.

6. CONCLUSIONS

With the increasingly fierce competition among enterprises, it is really important for enterprises to understand the behavior of customers. Considering that event logs generated by information systems contains the complete behaviors of customers, it is possible to mine customer behavior patterns from a process-based perspective. This paper proposes a complete method. The required data is easy to obtain, and the method is easy to operate. The method has some important managerial implications for enterprises. The customer clustering allows for more effective customer management with less cost. In addition, discovery of the behavior patterns of different customer clusters makes it possible to perform the service operations with better preparations, leading to higher service efficiency and less resource consumptions. What's more, timely detection and correction of customer deviation behaviors would help reduce potential risks in the customer services. Our case study of an important Chinese port shows that this method is applicable. Future work includes the optimization of algorithms and techniques.

ACKNOWLEDGEMENT

This work was supported by National Natural Science Foundation of China [grant number 71502010] and the China Railway Corporation [grant number K2019S011].

REFERENCES

[1] Zhang X , Gong W , Kawamura Y . Customer Behavior Pattern Discovering with Web Mining[C]// Advanced Web Technologies and Applications, 6th Asia-Pacific Web Conference, APWeb 2004, Hangzhou, China, April 14-17, 2004, Proceedings. DBLP, 2004.

[2] JU Chunhua, SHUAI Zhaoqian. Customer behavior pattern mining and change detection with context intensity constraints [J]. Journal of Management Sciences in China,2014,17(08):60-73.

[3] M. Alex Syaekhoni,Chanseung Lee,Young S. Kwon. Analyzing customer behavior from shopping path data using operation edit distance[J]. Applied Intelligence,2018,48(8).

[4] Goel A , Mallick B . Customer Purchasing Behavior using Sequential Pattern Mining Technique[J]. International Journal of Computer Applications, 2015, 119(1):24-31.

[5] Cheng Mingzhi, Xin Yang, Tian Yangge, Wang Cong, Yang Yixian. Customer Behavior Pattern Discovering Based on Mixed Data Clustering[P]. Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on,2009.

[6] Ray G, Muhanna W A, Barney J B, et al. Information technology and the performance of the customer service process: a resource-based analysis[J]. Management Information Systems Quarterly, 2005, 29(4): 625-652.

[7] Chen I J, Popovich K. Understanding customer relationship management (CRM): People, process and technology[J]. Business Process Management Journal, 2003, 9(5): 672-688.

[8] W.M.P. van der Aalst,H.A. Reijers,A.J.M.M. Weijters,B.F. van Dongen,A.K. Alves de Medeiros,M. Song,H.M.W. Verbeek. Business process mining: An industrial application[J]. Information Systems,2006,32(5).

[9] Mannhardt F , Leoni M D , Reijers H A , et al. Data-Driven Process Discovery - Revealing Conditional Infrequent Behavior from Event Logs[C]// International Conference on Advanced Information Systems Engineering. Springer, Cham, 2017.

[10] Van der Aalst W , Weijters T , Maruster L . Workflow mining: discovering process models from event logs[J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(9):1128-1142.

[11] Sedrakyan G , Weerdt J D , Snoeck M . Process-mining enabled feedback: “Tell me what I did wrong” vs. “tell me how to do it right”[J]. Computers in Human Behavior, 2016, 57:352-376.

[12] Nicolas Poggi, Vinod Muthusamy, David Carrera1, Rania Khalaf. Business Process Mining from E-Commerce Web Logs[J]. BPM, 2013, 65–80

[13] W. M P , van, der, et al. Business process mining: An industrial application[J]. Information Systems, 2007.

[14] Wang, Y, Caron F Vanthienen J, Huang L, Guo Y. Acquiring logistics process intelligence: Methodology and an application for a chinese bulk port. Expert Systems with Applications, pp. 195–209 (2014)

[15] Rozinat A , De Jong I S M , Gunther C W , et al. Process Mining Applied to the Test Process of Wafer Scanners in ASML[J]. IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS PART C-APPLICATIONS AND REVIEWS, 2009, 39(4):474-479.

[16] Anton, Jon. The past, present and future of customer access centers[J]. International Journal of Service Industry Management, 11(2):120-130.

[17] So Young Sohn, Yoonseong Kim. Searching customer patterns of mobile service using clustering and quantitative association rule[J]. Expert Systems with Applications, 34(2):1070-1077.

[18] Erlandsson F , Bródka, Piotr, Borg A , et al. Finding Influential Users in Social Media Using Association Rule Learning[J]. Entropy, 2016, 18(5).

[19] Erdem S, Demirörs O. An exploratory study on usage of process mining in agile software development[C]// 2017.

[20] Rozinat, A. and Van der Aalst, W.M.P. , Conformance checking of processes based on monitoring real behavior, Information Systems, 2008, 33:64-95

[21] Pedro M. Esposito, Marco A. A. Vaz, etc. MANA: Identifying and Mining Unstructured Business Processes[M]. BPM 2012 Workshops, LNBIP. 2013.

[22] Gang Xue, Xiaohu Ye, Jinwu Yang. Pseudo-parallel genetic algorithm in process mining[P]. Information Science and Technology (ICIST), 2012 International Conference on,2012.

[23] Dongen B F V , Aalst W M P V D . Multi-phase Process Mining: Building Instance Graphs[C]// International Conference on Conceptual Modeling. 2005.

[24] Caruso G , Gattone S A , Fortuna F , et al. Cluster Analysis as a Decision-Making Tool: A Methodological Review[M]// Decision Economics: In the Tradition of Herbert A. Simon's Heritage. 2018.