Consumer Behaviour Timewise Dependencies Investigation by Means of Transition Graph

Anton Kovantsev(✉)

ITMO University, 16 Birzhevaya Lane, 199034 Saint Petersburg, Russia
ankovantcev@itmo.ru

Abstract. The investigation of consumption behaviour on the level of every single person or some certain groups put forward some new tasks different from the behavioural analysis of the whole population. One of them is the problem of temporal peculiarities of consumer behaviour, for instance, how to find those, who react on some critical events faster than the others. It could be useful for identifying a focus-group which would show the tendency and help to make more accurate predictions for the rest of the population. A graph-based method of consumer's behaviour analysis in the state space is developed in this research. The moments when the deviations from the usual behavioural trajectory occur are detected by incremental comparing the transition graph with its previous state. These moments collected for all customers help to separate the population by the delay time of their reaction to the critical situation. It's also noticed that the velocity of the reaction is a personal feature of a customer, hence, this separation stays actual for different external events which cause the behavioural anomalies.

Keywords: Transition graph · State space · Consumption behaviour · Reaction time

1 Introduction

In this research we propose a graph based method of timewise dependencies investigation in consumer behaviour in crisis situations caused by some external events. Usually these events ruin the previous plans and forecasts and put forward a problem of adjusting the prediction models in the face of enormous uncertainty. This is why it is important to know as soon as possible the main tendency in the certain economic process. Here we deal with the consumer's transactions in Russia in 2020 and in 2022 when those critical events happened. Analysing the customer's behaviour in general for the whole population we can notice these periods of stationarity loss in the cumulative expenses time series which is shown in Fig. 1. The cumulative daily expenses of 10,000 customers by major areas of interest are depicted.

We can also suggest that not all clients react on lockdown in 2020 and military affairs in 2022 simultaneously. Common sense and observations make us suggest

H. Cherifi et al. (Eds.): COMPLEX NETWORKS 2023, SCI 1144, pp. 80–90, 2024.
https://doi.org/10.1007/978-3-031-53503-1_7

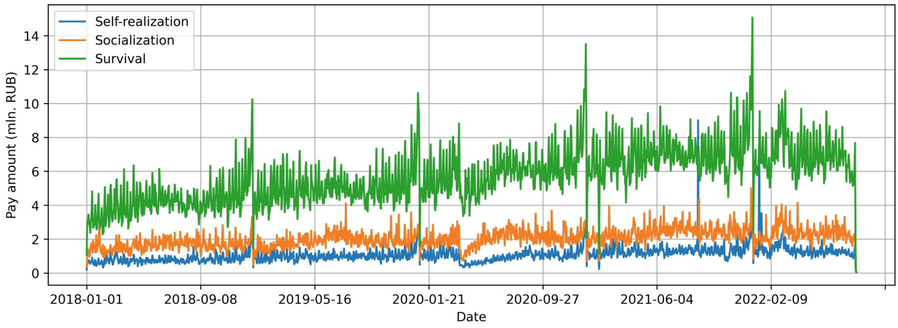

Fig. 1. Cumulative expenses of all customers.

that some of them changed their behaviour earlier than the others but in the same way.So, this group of the "quick reaction clients" can show how the rest of the population will react some time later and hence they become a leading group which shows the new tendency in the crisis behaviour.

Our solution is based on the dynamic graph analysis of the behavioural trajectory in the state space for every single customer. It helps to recognise abnormal behaviour and mark the time when it happens. In fact we deal with the number of different Hawkes processes with wide volatility of their rates along the time. Statistical methods to trace them are complicated and cannot perform the sufficient quality whereas the grapf methods applied to such processes could be fruitful, as it is shown by Paul Embrechts, Matthias Kirchner in [2]. So, the easiest way to recognise the deviations in customer's behaviour is to trace the days when the dynamic graph becomes not isomorphic to itself the day before. As we have found out, it works better and faster than a more complicated approach with the graph neural network. We can also calibrate this instrument on fully predictable anomalies like calendar holidays. Our intuition was that the same people who got started to prepare for the holiday earlier reacted to the critical events earlier. It was asserted within our experiments. Thus, we can separate the population according to the reaction time and distinguish a group of quick reaction clients which is about a quarter of the whole population. Besides, we can approximately calculate the time delay between these two groups by means of cross-correlation analysis. This time delay is helpful for forecast adjustment of the remaining three quarters.

2 Related Works

Many modern behaviour researchers pay attention to the timewise aspects of consumers' reaction to external circumstances. Mostly they deal with the time spent for purchasing, but the conclusions are valuable for our research because they support the hypothesis that time-related features in consumption behaviour remain persistent for every customer and represent some personal characteristic.

For instance, exploring the temporal aspect of consumption behaviour Jackob Hornik proposes a model that rests on the theory of choice under uncertainty [6]. Nevertheless, the conclusion of this article shows that some personal features of the customers persist over time. The approaches to the timewise aspects of consumption behaviour are also worth observing in this research. Riccardo Guidotti and his colleagues go further and design a model based on individual and collective profiles which contain temporal purchasing footprint sequences [4].

Chenhao Fu with coauthors investigate spatio-temporal characteristics and influencing factors of consumer behaviour [3]. They notice that by affecting consumer preferences, residents' social and economic attributes lead to the differences among consumers' spatial and temporal behaviour. They use structural equation modelling to further mine the relationships between influencing factors of spatio-temporal behaviour that are related to path relationships and decision-making processes.

Dr. Ling Luo proposes a systematic approach for tracking the customer behaviour changes induced by the health program [8]. According to this approach customer preferences are extracted from the transaction data and a temporal model for customer preferences is constructed so that the varying customer preference can be modelled via the series of preference matrices, taking the temporal domain into consideration. Victoria Wells, Marylyn Carrigan, and Navdeep Athwal in their research [11] pay attention to the behaviour changes during the critical process on the example of pandemy. Their analysis using foraging theory offers an explanation that locates these behaviours as both logical and understandable within the context of the pandemic, based on changing environment, constraints and currency assumptions.

These works show that there are definite persistent peculiarities in the behaviour of every single customer including the temporal features which we are going to investigate. Our confidence is supported by the prospect theory, which forms a basis for individual consumer neuroscience, and includes an overview of the most relevant concepts in consumer research and behavioural economics, such as the framing effect, the phenomenon of anchoring and the endowment effect, the phenomenon of temporal discounting, as well as decision-making under risk. All of them are related with the brain structures and processes, as Sven Braeutigam, Peter Kenning explain in their book [1].

Graph-based methods are also applied to behavioural analysis. Katsutoshi Yada and the other authors of [13] discuss how graph mining systems are applied to sales transaction data so as to understand consumer behaviour. They propose to represent the complicated customer purchase behaviour by a directed graph retaining temporal information in a purchase sequence and apply a graph mining technique to analyse the frequent occurring patterns. Constructing a recommender system Weijun Xu, Han Li and Meihong Wang enhanced the graph-based approach by using multi-behavior features [12]. They develop a new model named Multi-behavior Guided Temporal Graph Attention Network (MB-TGAT), tailor a phased message passing mechanism based on graph neural network and design an evolution sequence self-attention to extract the users' preferences from static and dynamic perspectives, respectively.

Zhiwei Li with colleagues in their research [7] use a temporal graph constructed on skeletal key points of a human shape to recognise abnormal behaviour of the people in video sequences. The spatio-temporal graph convolutional network (ST-GCN) is used to extract the temporal and spatial features of the skeleton, which are inputted into softmax to predict action classes.

So, graph-based methods are popular for the consumption behaviour analysis and particularly in its temporal aspect. It's surprising that we failed to find an attempt to separate the consumers by the reaction time on the critical events. Here we try to fill up this omission.

3 Data Description

In our research we deal with the transaction data on debit cards of clients of a commercial bank in Russia for January 2018 - July 2022. It contains information about 10,000 active consumers' expenses, which is sampled daily. Payment categories, standardly marked in transactional data by MCC (Merchant Category Code), referring to the place of payment of the code category are brought into line with three basic values: "survival", "socialisation", "self-realisation". The mapping of MCC to the values is described in the article by Valentina Guleva et al. [5]. The cumulative expenses of all consumers look like the time series which are shown in Fig. 1

Despite the data itself cannot be provided for public access, this data-set is chosen because it includes the periods of significant critical events in March-April 2020 and in February-March 2022, which are mostly interesting for this research. Thus, we obtain the 10,000 time sequences of three-component vectors of every-day behaviour for each customer. The example of such a sequence is plotted in Fig. 2.

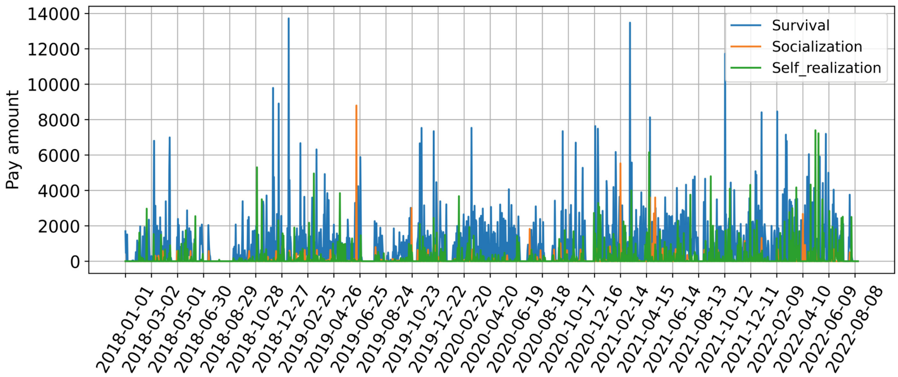

Fig. 2. Transaction sequences for a single customer.

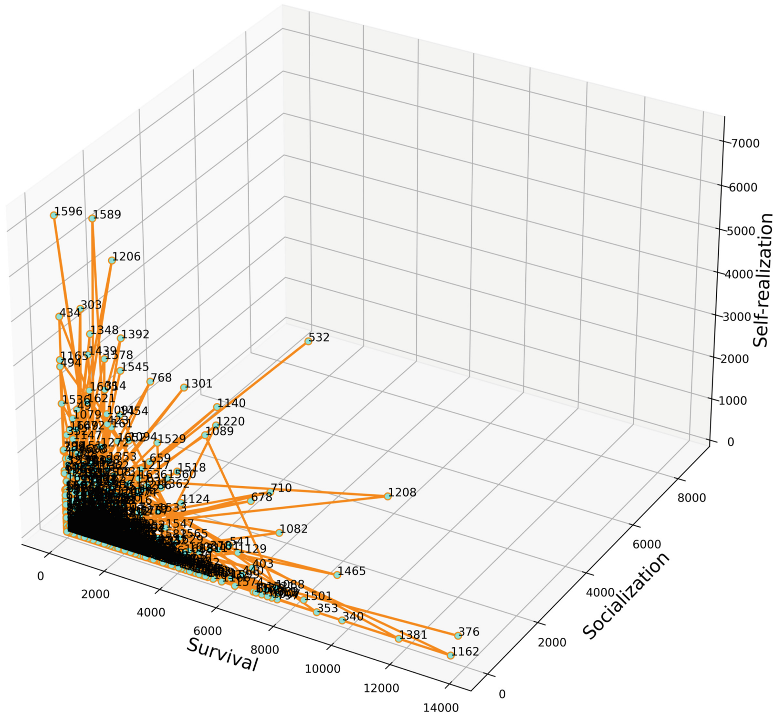

Fig. 3. Trajectory in the state-space for a single customer. *Numbers next to the points are the ordinal numbers of days*

4 Transition Graph Construction and Applying

The method which we use to construct a dynamic transition graph for the transactional sequence for every single customer is similar to that proposed for the continuous time series by Dmitriy Prusskiy et al. in [10]. The difference is that we do not need to reconstruct the state space because those three components of the transactional sequence for "survival", "socialisation", "self-realisation" already define a state space (abbreviated SSS-state space) of the consumer behaviour dynamics. The every-day expenses by three basic values trace the consumer's trajectory in this space, as it is shown in Fig. 3.

Being agglomerated into clusters by the distance between them the trajectory points in the SSS-state space space become the nodes of a graph network while the transition along a trajectory in time determines its directed edges. As the amount usually spent by different clients may significantly differ, for the sake of the reasonable number of clusters and nodes in the graph, the distance between points for hierarchy clustering has been chosen by an automated elbow method based on a geometric approach proposed by Eduardo Morgado et al. in [9]. The

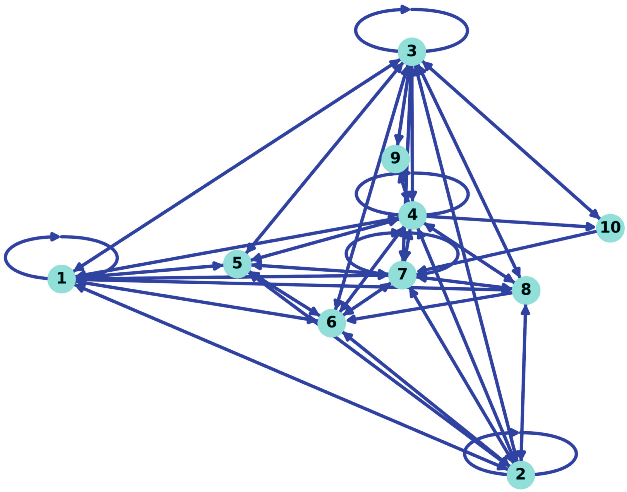

Fig. 4. Transition graph along the whole period for a single customer. *The numbers of nodes correspond the order in which the behaviour trajectory hits the certain points in the state space.*

example of the initial graph for one of the customers is presented in Fig. 4. This initial graph is constructed on the first part of the sequence.

Further to make it dynamic every new point in the behavioural trajectory is checked for proximity to one of the clusters, nodes of the graph. If the distance is less than that defined when the initial graph has been created the new point is added to the node, otherwise a new node appears and a new edge connects it with the node where the previous point is. This procedure continues incrementally. We represent the history of transactions as the time sequence of graphs. Every time when a new node or an edge appears we may suspect an anomaly in customer behaviour. In other words, if once on the new time step the dynamic graph occurs not isomorphic to itself on the previous step, we conclude that the customer behaviour has changed. At first following the method of [10] or [7] we applied the graph neural network (GNN) to detect the differences in the graphs. This consists of three convolutional layers for node embeddings, a readout layer of global mean pooling and a final classifier with $\frac{1}{2}$ dropout and linear transform for the result. It's precisely similar to that used in [10] where the architecture and parameters were chosen in the experimental studies. But in out work we made sure that the isomorphism check works much faster and provides better results. An isomorphism of temporal graphs on the time step $t : G(t)$ and $t-1 : G(t-1)$ is a bijection between the vertex sets of these graphs $f : V(G(t-1)) \rightarrow V(G(t)$ such that any two vertices u and v of $G(t-1)$ are adjacent in $G(t-1)$ if and only if $f(u)$ and $f(v)$ are adjacent in $G(t)$. The graphs with which we deal in our

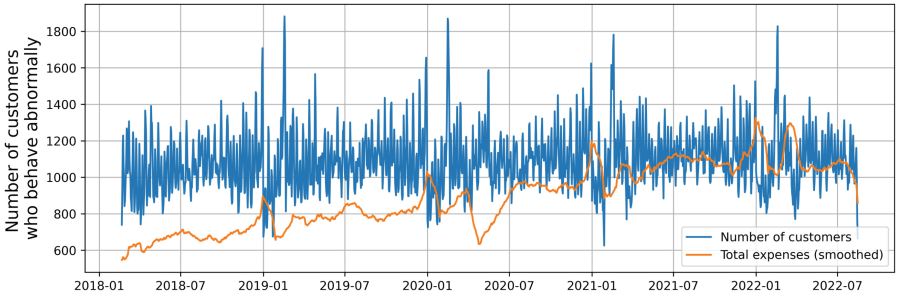

Fig. 5. Number of customers whose graph significantly changed at a certain date.

research are weighted and directed, so the notion of isomorphic may be applied by adding the requirements to preserve the corresponding of edge directions, node weights. But the tests show that it is not necessary. The point is that we should only detect a deviation from the usual trajectory of behavior, regardless of the way it occurs.

Of course, the circumstances of the personal life of any client make him deviate from the usual trajectory rather often and these anomalies may not be caused by some critical events which affect the whole population. But if such an anomaly appears simultaneously in the behaviour of a large group of consumers, it could be a sign of some change in the environment, which we tend to trace.

5 Timewise Dependencies Investigation

The first step of the experiment was the initialising the transition graphs for all customers in the data set. We used the first fifty-day period so that the monthly and weekly seasonalities get included into the normal state of the consumers' behaviour. Then incrementally day by day the new transactional data was added and on each step the self-isomorphism of the graph was checked. The date when the dynamic graph occurs not isomorphic to its previous state was saved in memory for the certain client. It took about 80 minutes for a humble personal desktop computer to recognise the special dates for 10,000 clients over 1638 days. Thus we obtain a set of dates with abnormal behaviour for every customer.

The next step is to count the clients with abnormal behaviour for each day. The result of this calculation is shown in Fig. 5. The plot of total population expenses is also shown there to compare with calculated numbers. We may notice that during the periods of our interest in 2020 and in 2022 the increasing number of customers with abnormal behaviour precedes the changes in total consumption tendency.

Now the collected sets of dates may be used for calibrating the customers. We do not have labels for normal and deviant consumer behavior, but we can assume that certain holidays tend to trigger these deviances. As we noticed before, the calendar holidays are fully predictable crises which give us an opportunity to see which of the clients start preparing in advance. We choose the International Women's Day celebrated annually on March 8 which is very popular in Russia and besides it is a day off. In Fig. 5 we can see that the most of behavioural anomalies happens exactly around this date. For definiteness we took March 8, 2021, when no crisis happened. We assumed that the early customers started preparing beforehand and sought those who behave abnormally from March 3 till March 5. Of course, the date range influences the set of the customers which we want to obtain as a result. The wider range provides the more clients, the earlier dates may have less connection with the holiday. These dates were chosen by the series of experiments. So, 2519 customers were separated as the early ones. The group of customers who were selected by abnormal behaviour during this period indeed was the leading group, what we found out when comparing their cumulative expenses to that of the residue customers for the critical period of February 2022 as it is shown in Fig. 6. The time series data was rescaled for better timewise feature comparison.

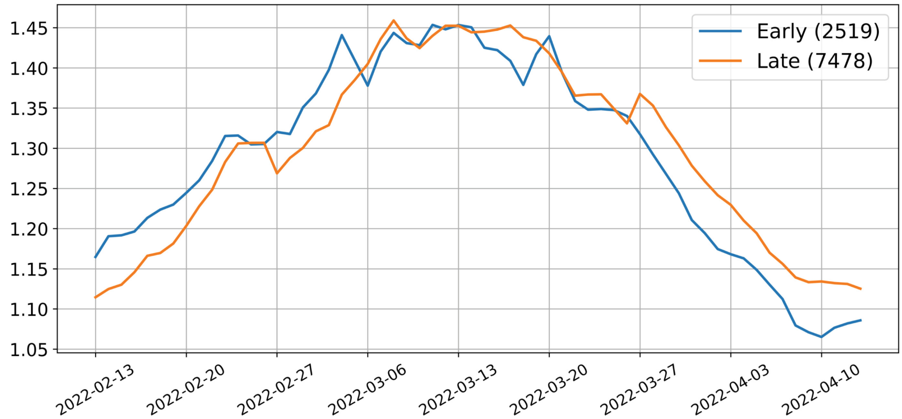

Fig. 6. Cumulative expenses of "early" and "late" customers in February-March 2022. *Rescaled for better timewise feature comparison.*

In order to evaluate the time delay between slow and fast customers we applied cross-correlation analysis and calculated the Pearson rank correlation coefficient for these two time-series with the time lag from two weeks earlier to two weeks later. The coefficient values by the time lag are presented in Fig. 7, which shows that the maximal correlation is for the three-day lag. This is how the early ahead of the late.

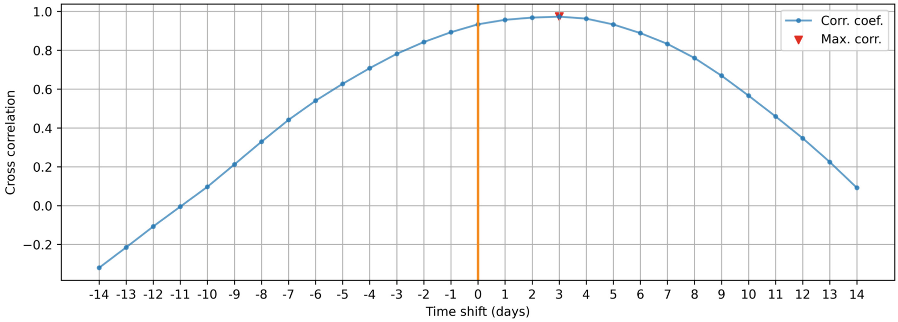

Fig. 7. Cross-correlation between "early" and "late" customers shows three-day outrunning for "early" ones.

These three days may be somehow connected with our choice of the calibration dates which were from three to five days before the holiday. But nevertheless, our approach showed a quite sufficient result. The same result was when we turned to the critical days of 2020. As we can see in Fig. 8 early customers group shows the change of tendency even with the greater advance of 6–7 days in the beginning of the growth around March 9, 2020.

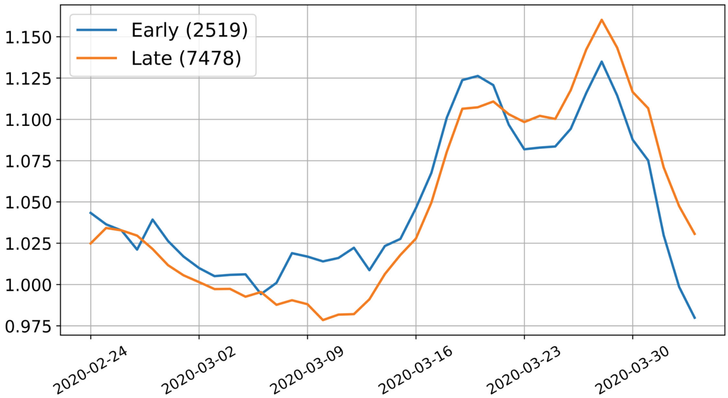

Fig. 8. Cumulative expenses of "early" and "late" customers in February-March 2020. *Rescaled for better timewise feature comparison.*

As it is mentioned in the Related works section, we could not find any practice or research for the task of separating the customers by reaction velocity. So, the only way to evaluate the proposed method is to make sure that the results are relevant to the problem and useful for the forecasting practice, and the requirements for computing resources are not so great.

6 Conclusion and Future Work

In this research a dynamic graph based method was developed to recognise the response time of the consumer behaviour to the critical events for the groups of quick and slow reacting clients. The idea of the method is that incremental graph isomorphism analysis can detect the abnormality in every customer's behaviour for each day of the period. The dates of calendar holidays, which are the days when the majority of the population change their usual behaviour, help to distinguish the groups of fast and slow responding clients. This temporal characteristic of a single customer as well as of a certain group persists that approves the hypothesis that it is a personal customer's feature which is either changing much slower than the observation time or even constant. The time delay between the groups of fast and slow customers is about three days for the data set which is explored and the ratio of quick and slow in population is about 1 to 3, respectively. As the method is incremental it can be used in the real-time mode adding new transactional data day by day for early detection of trend changes in consumption behaviour.

For the future work we plan to improve the calibration procedure to choose the customers in the leading group more precisely and to reduce the number of clients in it. It would be also good to choose the customers with the greater outrunning. Besides, it would be interesting to try some other graph features to see whether they could help to recognise the abnormal behaviour more accurately.

Acknowledgements. This research is financially supported by the Russian Science Foundation, Agreement 17-71-30029, with co-financing of Bank Saint Petersburg, Russia.

References

1. Braeutigam, S., Kenning, P.: Individual consumer neuroscience (2022). https://doi.org/10.1093/oso/9780198789932.003.0007
2. Embrechts, P., Kirchner, M.: Hawkes graphs. Theor. Probab. Appl. **62**, 132–156 (2018). https://doi.org/10.1137/S0040585X97T988538
3. Fu, C., Zhou, S., Yan, X., Liu, L., Chen, W.: Spatio-temporal characteristics and influencing factors of consumer behavior in retailing centers: a case study of guangzhou in guangdong province. Dili Xuebao/Acta Geographica Sinica **72**, 603–617 (2017). https://doi.org/10.11821/dlxb201704004
4. Guidotti, R., Gabrielli, L., Monreale, A., Pedreschi, D., Giannotti, F.: Discovering temporal regularities in retail customers' shopping behavior. EPJ Data Sci. **7**, 1–26 (2018). https://doi.org/10.1140/epjds/s13688-018-0133-0
5. Guleva, V., Kovatsev, A., Surikov, A., Chunaev, P., Gornova, G.: Value-based modeling of economic decision making in conditions of unsteady environment. Sci. Tech. J. Inf. Tech. Mech. Opt. **23**, 121–135 (2023). https://doi.org/10.17586/2226-1494-2023-23-1-121-135
6. Hornik, J.: The temporal dimension of shopping behavior. J. Serv. Sci. Manag. **14**, 58–71 (2021). https://doi.org/10.4236/jssm.2021.141005

7. Li, Z., Zhang, A., Han, F., Zhu, J., Wang, Y.: Worker abnormal behavior recognition based on spatio-temporal graph convolution and attention model. Electronics (Switzerland) **12**, 2915 (2023). https://doi.org/10.3390/electronics12132915
8. Luo, L.: Tracking purchase behaviour changes (2020). https://doi.org/10.1007/978-3-030-18289-2_4
9. Morgado, E., Martino, L., Millán-Castillo, R.S.: Universal and automatic elbow detection for learning the effective number of components in model selection problems. Digital Sig. Process. Rev. J. **140**, 104103 (2023). https://doi.org/10.1016/j.dsp.2023.104103
10. Prusskiy, D., Kovantsev, A., Chunaev, P.: Dynamic transition graph for estimating the predictability of financial and economical processes (2023). https://doi.org/10.1007/978-3-031-21127-0_39
11. Wells, V., Carrigan, M., Athwal, N.: Pandemic-driven consumer behaviour: a foraging exploration. Mark. Theory (2023). https://doi.org/10.1177/14705931231175695
12. Xu, W., Li, H., Wang, M.: Multi-behavior guided temporal graph attention network for recommendation (2023). https://doi.org/10.1007/978-3-031-33380-4_23
13. Yada, K., Motoda, H., Washio, T., Miyawaki, A.: Consumer behavior analysis by graph mining technique. In: Negoita, M.G., Howlett, R.J., Jain, L.C. (eds.) KES 2004. LNCS (LNAI), vol. 3214, pp. 800–806. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30133-2_105