Artificial Intelligence Research and Development
I. Sanz et al. (Eds.)

doi:10.3233/FAIA230687

Sequence Pattern Mining for Citizens Behaviour Learning in Fair-Purpose Social Games. Some Preliminary Results

Beatriz LÓPEZ [a,1], Roger CAMPDEPADROS [b] and Albert SABATER [c]
[a] *eXiT research group, University of Girona*
[b] *Grup de Recerca Interdisciplinar en Gènere i Desigualtats Socials (IGESI), University of Girona*
[c] *Observatori d'Ètica i Intel·ligència Artificial (OEIAC), University of Girona*
ORCiD ID: Beatriz López https://orcid.org/ 0000-0001-9210-0073, Roger Campdepadros https://orcid.org/ 0000-0003-3744-162X, Albert Sabater https://orcid.org/ 0000-0003-3532-6572

Abstract. Artificial Intelligence can be used to understand how citizens interact in a society with the aim to work together towards a better world. The application of AI in this domain has shown immense potential, particularly in leveraging games as powerful tools to simulate individual roles within a society. This research endeavors to explore how sequence pattern mining are applied to data coming from a game, in which humans are posed with the dilemma of progressing through their jobs, while enjoying life, trying to catch an equilibrium between happiness and healthiness. Preliminary results on a game played with 13 citizens shows the feasibility of sequence learning for this purpose. Our findings shed light on the intricate interplay of individual aspirations, communal synergies, and the pursuit of equilibrium, laying the foundation for future advancements in AI-driven methodologies for sociological investigations.

Keywords. pattern mining, sequence learning, social games

1. Introduction

Studying how Artificial Intelligence (AI) will change public service provision and its underlying societal systems has become key in different contexts worldwide, as governments and organizations increasingly recognize its potential to revolutionize the way services are delivered, streamline administrative processes, and enhance overall efficiency. For instance, one of the most significant impacts of AI on public service provision is the potential to improve decision-making through data analysis and predictive algorithms. This data-driven approach enables policymakers and public service providers to make more informed decisions, allocate resources more effectively, and address societal needs proactively.

[1]Corresponding Author: beatriz.lopez@udg.edu

However, which policies, behavioural changes and institutional developments are necessary and appropriate to prevent or support certain scenarios has focused so far on how AI-based social assessment technologies for public service provision categorise present and future human behaviour on scales such as legal recipients/fraudulent recipients, deserving/non-deserving or acceptable/not-acceptable ones. Since delegating decisions on such value judgements to AI raises ethical and social considerations related to responsibility, accountability, transparency and the quality of social decision making about the distribution of resources, further research needs to be conducted from a multidisciplinary perspective of social sciences and AI. This is the case of the AI FORA project [2] in which we are involved. In this project, heterogeneous societal organizations interact in their specific socio-cultural context to negotiate and co-produce AI social assessment technology. This approach is particularly useful in social service provision as often public opinion and discourse in these areas can be emotional and empathetic, not least because societal core values are affected and at stake, and decisions on providing or refusing public social services have far-reaching consequences for the concerned individuals.

Because games have a unique ability to engage and empower participants and they can play a significant role in facilitating the co-production of AI social assessment technology and achieving more responsible and empathetic social service provision, we employ gamification elements, such as challenges, rewards, and competition in order to motivate stakeholders to contribute their knowledge, insights, and concerns more willingly.

Games have been previously used as a key tool to understand self-identity and values and by designing collaborative games that require input from different stakeholders, societal organizations can work together in an interactive and inclusive environment that recognise the citizens major implications in society [1]. Games allow participants to explore different scenarios and consequences and data gathered from games can be further used by agent-based social simulations to understand the impact of behaviours at the long run. In the context of AI social assessment technology, stakeholders can use simulation games to understand the potential impacts of AI-driven decisions on social service provision. This promotes thoughtful reflection on the technology's implications and helps identify and address potential biases and unintended outcomes [2,3,4]. Despite their potential complexity, games not only offer the opportunity to simulate what-if scenarios about possible automated decisions regarding in social provision, but also provide the potential to engage the wider public in discussions surrounding AI and social service provision [5,6]. Increased public understanding and input contribute to a more democratic and transparent approach to AI development and the use of concepts as ethnicity [7] and transparency [8] in decision-making models in virtual environments before implementing them in real-world scenarios allows for identifying potential flaws and biases, minimizing the risk of harmful consequences in actual social service provision.

In addition, games generate empirical data, and machine learning methods could be used to analyze them, as we demonstrate in this work. In particular, we use the data from the unemployment game[3] that has been set up to understand the concept of socially fair regarding provision of jobs. To that end, the game has been designed with the aim that players have the possibility to agree in the algorithm that assigns jobs to citizens. Players interact in several rounds, selecting some actions to choose in each round. Therefore, a

[2] https://www.ai-fora.de/

[3] https://github.com/micrology/AIFORA

sequence of actions per player are gathered at the end of the game, and sequence pattern mining is applied as a suitable method to discover common patterns.

By applying machine learning techniques, such as sequence pattern mining, to the data generated by the unemployment game, we can gain valuable insights into how players approach the concept of socially fair job provision. The gathered sequences of actions represent the decision-making processes and preferences of the players throughout the game, offering a rich dataset to analyze. Thus, sequence pattern mining allows us to identify recurring patterns in the players' actions, which can shed light on their strategies and priorities through patterns. For example, these patterns can offer valuable clues about potential biases and areas for improvement in the job assignment algorithm. Moreover, machine learning methods can help quantify the impact of different player actions on the overall fairness and effectiveness of the job provision process. The combination of serious games and sequence pattern mining provides a powerful approach for understanding complex social dynamics and decision-making processes, unveiling underlying patterns and trends that may not be apparent through traditional analytical methods.

This paper is organized as follows. In Section 2 we provide a literature review of the work related to us. Section 3 provides a basic explanation of the game we use. Section 4 details the methodology designed, and Section 5 describes and discuss the results achieved. Finally, the paper ends with some conclusions and future work.

2. Related work

Sequence pattern mining has been applied to games according to the purpose the game has been set up. For example, in [9], the authors use the Prefixspan algorithm [10] to learn patterns in a game with tutoring purposes. Tutoring systems offer the possibility to conduct users towards a useful achievement; on the other hand, in our work we aim to analyse self-identify values in a game.

Learning environments are also explored in [11] but with the goal to understand players' emotions. To that end, Taub et al. use the SPAM algorithm[12]. Taub et al. work is similar to us since they use sequence pattern mining to analyse different students skills regarding the games progresses. In educational games, sequence pattern mining helps track learners' progress and identify areas where they may be struggling. This information allows educators to tailor their instructional approach and provide personalized feedback, enhancing the learning experience. Another similarity of the Taub and colleagues approach is that we also use information on different person's skills.

[13] focuses on finding the given context of a frequent sequence. Context is a mattern of future research of our work. Although our game allow changes according to different external events (i.e. decrease of available jobs), our preliminary approach does not consider such dynamics of the game.

In this work, we use VEPRECO[14], a more powerful algorithm than its ancestors, Prefixsan and SPAM, and we apply sequence learning with the aim to understand self-identity and values, according to [1].

3. The game in a nutshell

In this research, the social game serves as an interactive platform where a group of users can participate and experience various aspects of social service provision and decision-making. Each player is assigned an identity card that outlines their citizen profile, which includes information such as their marital status (single or parents), level of education, current state (happiness, wealth, training level, communication skills), and job status (unemployed, poor job, good job). The game aims to simulate real-life scenarios, and players are expected to act according to their assigned identity and state.

During each round of the game, players have the opportunity to visit six different places or services, each offering unique interactions and choices:

- The Labour Office: Here, players can apply for a new job based on a predefined algorithm that determines job assignment. The outcome of this interaction can impact the player's job status and overall well-being.
- The Working Place: If the player has a job, they can attend their workplace, reflecting the daily work routine and the experience of having a job.
- Staying at Home: Players can choose to stay at home, which may signify taking a break or focusing on personal well-being.
- The Travel Agency: This allows players to simulate going on holidays, offering an opportunity for leisure and relaxation.
- The School: Players can visit the school to acquire new knowledge and skills, reflecting the importance of education in their lives.
- The City Hall: The city hall acts as a central meeting place where players can interact with each other, propose changes to the algorithm governing job assignment, participate in voting processes, and make donations.

Throughout the game, players make choices at each place, and these decisions impact their happiness, wealth, skills, training level, and job position, if applicable. For example, getting a new job might increase a player's happiness and wealth, while attending the school can enhance their skills and training level. It is also worth noting that the game offers certain exceptions to the rule of visiting a single place in a round. For instance, players can go to work after obtaining a new job from the labor office, showcasing how work and job transitions are interconnected. Additionally, players can always visit the city hall to make donations, emphasizing the importance of social engagement and contributing to the community.

Overall, depending on the chosen action, the agents change their happiness, wealth, skills and training level, as well as the job position, if it is the case. In summary, the game used in this research provides a dynamic and immersive environment for players to experience social service provision, decision-making, and societal interactions. By assigning unique citizen profiles and offering multiple places to visit with different outcomes, the game allows players to explore various aspects of life and gain insights into the complexities of socially fair job provision and the factors that influence individual well-being and opportunities. Further details of the game can be found elsewhere[4].

[4] https://github.com/micrology/AIFORA

4. Methodology

Figure 1 shows an overview of the methodology followed. The starting point is the data collected from the social game. Data has been collected in the basis of each entity providing services in the game; therefore, we need to track users behavior from all the data to build individual sequences per each user. Once a database of sequences is obtained, two possible branches are explored to obtain frequent patterns. First frequent patterns from the full dataset are generated. And second, the sequence database is split according to a given dimension of the profile of users (see Table 1). This second step is repeated for each dimension. Next, sequence learning mining is applied in each split and the corresponding sequence patterns are found. The patterns that characterize a given dimension are finally obtained by selecting the ones that are present for a given value of the dimension, and do not in the opposite one.

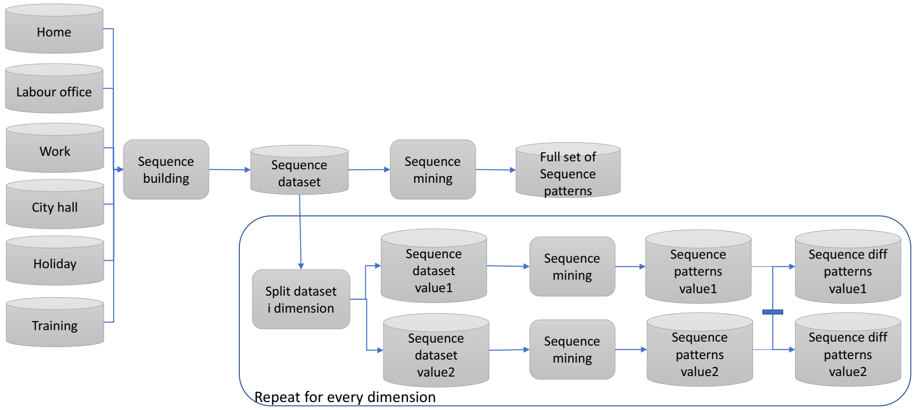

Figure 1. Flow chart of the methodology

Dimension	value	
Gender	men	women
Family	single	parent
Ethnicity	white	minority
Skills	high	low
Vulnerability	high	low
Residence area	rich	poor

Table 1. Dimensions that configure the players profiles.

4.1. Dataset

We get the data from a game played by 13 players ($N = 13$). Each action available to choose by the agents is labelled according to the map provided in Table 2. There are a total of $M = 6$ places to visit. A total of $R = 10$ rounds have been played.

From each round quantitative data about the players is gathered, as well as qualitative data (i.e. what changes in the algorithm has been proposed; which externalities could be arisen, as firing a person from a job, ...). We focus in this work on quantitative data, as a preliminary exploratory analysis, and we left for future work the use of qualitative data.

Place, activity (if many)	label
Home	H
Labour office	L
Work	W
City hall, proposal to change the algorithm	C1
City hall, voting	C2
City hall, donation	D
Travel agency	Hol
School	T

Table 2. Mapping of the places a player could attend, and the activities she could perform (if more than one), and the labelled used for the method presented in this work.

We have a total of $R = 10$ sets of data, one per round, sorted according to the visits of players ($P = \{p_1, \ldots, p_N\}$) to the different places. We note by $\langle D_1, \ldots, D_R \rangle$ the sequence of data gathered in each round, where each $D_j = \{S_1, \ldots, S_M\}$. Each S_i contains the particular activities available in the service i that has been chosen by the players, $S_i = \{a^i_1, \ldots, a^i_{|S_i|}\}$, each $a^i_j = \{p^{i,j}_1, \ldots, p^{i,j}_{|a^i_j|}\}$ with $p^{i,j}_k \in P$. Note that a^i_j could be empty if no player has chosen this action in a given round.

4.2. *Preprocessing*

From the sequence $\langle D_1, \ldots, D_R \rangle$, we design Algorithm 1 to obtain a sequence of activities per user in each round, $\{Seq_1, Seq_2, \ldots, Seq_N\}$, each $Seq_j = \langle A_1, \ldots A_R \rangle$, and $A_i \subset \mathscr{A}$, being $\mathscr{A}$ the full set of available activities (see Table 2). Observe that A_i is a set of actions (itemset), since a user can eventually perform more than one action in a given round.

The algorithm bassically iterates over the different rounds (step 2), and on the services (nested iteration of step 3), to collect the different activities performed by each players (iteration of step 3). An example of the pre-processing output is the following:

$$\langle [T], [L,W], [W], [H], [C1], [W,D], [W], [W], [Hol], [W] \rangle$$

4.3. *Sequence pattern mining*

Given a data base of sequences of itemsets (set of events or items that occur at once), sequential pattern mining algorithms learn sequential patterns in the form of $\langle [a,b,...], [a',b',...], ..., [a'',b''] \rangle$ that are frequent in a the database, meaning that the itemsets in $[a,b,...]$ happens befere itemsets in $[a',b',...], ..., [a'',b'']$, etc. This kind of algorithms requires a support threshold $\gamma \in [0,1]$ that determines the frequency required by the pattern in the database to be considered a pattern. A low *gamma* value means that a lot of patterns could be gathered, which could be eventually difficult to interpret; a high *gamma* value could led to find no patterns at all. After some preliminary trials, we finally set up *gamma* $= 0.8$, mining that a pattern will have the 80% of sequences in the

Algorithm 1 Sequence building

Require: $R, P = \{p_1, \dots, p_N\}, \langle D_1, \dots, D_R \rangle$
each $D_j = \{S_1, \dots, S_M\}$
each $S_i = \{a^i_1, \dots, a^i_{|S_i|}\}$
each $a^i_j = \{p^{i,j}_1, \dots, p^{i,j}_{|a^i_j|}\}$ with $p^{i,j}_k \in P$

Ensure: $Seq = \{Seq_1, Seq_2, \dots, Seq_N\}$

$sequence_j \leftarrow null$ for all $p_j \in P$
for $r \in [1, R]$ **do**
    $itemset_j \leftarrow null$ for all $p_j \in P$
    for $i \in [1, M]$ **do**
        for every $a^i_j \in S_i$ **do**
            for $k \in [1, N]$ **do**
                if $p_k \in a^i_j$ **then**
                    append(a^i_j, $itemset_j$)
                end if
            end for
        end for
    end for
    append($sequence_j$, $itemset_j$) for all $p_j \in P$
end for

database that match the pattern. This *gamma* value allows to get a manageable amount of patterns; with a lower *gamma* value, the number of patterns would be huge, and difficult to interpret; while with a higher *gamma* value, very few patterns would be found.

An example of the patterns found in our setting is $[[T], [L], [W, D], [W]]$, which can be read as follows: players first go to the school to get some training (T); next (not necessarily the immediate next action) they go to the labour office (L); afterwards they go to the work office, and make some donation (W,D); and finally they go again to work (W).

Sequential pattern mining have a high computational cost, and there has been a lot of research to make the mining process efficient. In this wok we use the VEPRECO algorithm [14] that has been shown to be a competitive algorithm.

5. Results

In this section, we provide the results in terms of the number of the patterns found, for the full data set (first branch of the methodology, see Figure 1), and for each dimension (second branch on the Figure). Regarding the full set of patterns, we found 46 patterns, with the length shown in Table 3 . The maximum length of a pattern was 5. The identification of distinct patterns for each dimension allows us to understand how different attributes impacted decision-making and interactions within the game environment.

In Table 4 the results of the analysis for each individual dimension of the players' profiles are presented. In doing so, we provide a detailed breakdown of the patterns found within each dimension, along with additional statistics that offer valuable insights into player behavior and decision-making. Table 4 begins by showing the distribution of the different profile dimensions among the players. Here it is worth noting that there is a bias regarding gender representation in the dataset. There are more players from one gender

Description	#
1-length patterns	6
2-length patterns	14
3-length patterns	14
4-length patterns	9
5-length patterns	3

Table 3. Length of patterns found for the full set of sequences.

compared to the other and this imbalance should be addressed in future gamification experiments to ensure a more equitable and representative dataset. However, the remaining dimensions (other than gender) show a balanced representation, with an equal number of players (6 versus 7) falling under each given condition. For each dimension, the table presents the total number of patterns found (# patterns) specific to that dimension. These patterns represent recurring sequences of actions and interactions associated with a particular attribute of the players' profiles. The table also provides the number of patterns that make a given dimension-value distinct from the other dimensions (# diff patterns), along with the percentage of distinct patterns from the total patterns. This highlights the uniqueness and specificity of certain patterns associated with each profile dimension. For example, only 45.71% of the patterns found for men are truly distinct from the patterns found for women, suggesting some overlapping behavior between the genders.

Table 4 also includes the mean and standard deviation of the pattern lengths for each dimension. From the analysis, it is observed that, on average, patterns had a length of around 4 actions in the game. The standard deviation indicates slight variations from this average length, suggesting some players followed more extended sequences of actions compared to others. Finally, at the bottom of table 4, three actions are highlighted as possible key distinguishable features between dimensions. For instance, going to the work office (W) is found to be much more common in women and single individuals compared to other profiles. Additionally, players living in poor areas were less likely to visit the work office frequently. This highlights how certain attributes influence players' choices and job-seeking behavior in the game. Donations (D) were more often performed by men, white individuals, and high-skilled and high-vulnerable players, suggesting differences in charitable behavior across various demographics. Holidays (Hol) were found to be more often selected by white and high-skilled individuals, indicating distinct preferences for leisure activities among different groups.

5.1. Discussion

The results shown in this work are a preliminary analysis of sequence learning from information coming from a social game. Patterns are important to understand the behaviour of people in the game according to their skills and the selection of the different actions they have available. In the case of social games, patterns provide insights about self-identify regarding, for example, fairness in job assignment. Some patterns are consistent with expectations, such as women and single individuals going more often to work, or individuals of certain ethnic backgrounds frequently selecting holidays. However, the length of the patterns indicates a diverse range of behaviors exhibited by players. While the findings are insightful, the interpretation of these patterns requires further in-depth

	Genre		Family		Ethnics		Skills		Vulnerability		Residence area	
	men	women	single	parent	white	minority	highS	lowS	highV	lowV	rich	poor
Distribution	9	4	6	7	6	7	6	7	6	7	6	7
# patterns	35	37	55	213	363	49	190	47	49	235	151	66
# diff patterns	16	18	20	11	43	7	42	8	7	49	11	20
% diff patterns	45.71%	48.65%	36.36%	5.16%	11.85%	14.29%	22.11%	17.02%	14.29%	20.85%	7.28%	30.30%
Pattern length (mean)	2.3	3.1	3.0	3.5	4.1	2.9	3.4	2.9	2.9	3.7	3.4	3.0
Pattern length (std)	0.8	1.2	1.2	0.9	1.2	1.1	1.0	1.1	1.1	1.2	1.0	1.2
Diff patterns length (mean)	2.5	4.1	4.0	3.7	4.2	3.2	3.6	3.0	3.0	4.0	3.6	3.7
Diff patterns length (std)	1.2	0.7	1.4	0.8	1.3	0.9	0.9	0.9	0.8	1.1	0.8	1.1
Work (W)												
City Hall- Donation (D)												
Holidays (Hol)												

Table 4. Description of the data and the found patterns. Black boxes at the bottom represent a high correlation between players profiles (columns) and the actions at the three bottom rows. Grey boxes means a low correlation; and the cell in white means that no correlation has been found.

analysis. In addition, several considerations have been identified as potential areas of improvement and expansion for future research.

First, it is important to consider a dynamic perspective and context awareness. Currently, the analysis focuses on static patterns without considering changes in the game environment, such as alterations in the job assignment algorithm. Incorporating sequence learning from a dynamic perspective, along with context-aware information, could enhance the understanding of how players' decisions are influenced by changing conditions. Second, the game could be further improved by using an integration of player condition information. We acknowledge the omission of critical information at the onset of the game, such as players' happiness or wealth, which can significantly impact their decision-making. Indeed, integrating this additional player condition data poses a challenging problem for sequence learning. Third, while quantitative analysis provides valuable insights, complementary approaches are needed to gain a deeper understanding of the patterns observed. Within this context, qualitative data, such as player annotations and observations, can contribute to a more nuanced view and rich interpretation of the findings, including further information on the context dimension to the quantitative patterns. Thus, to provide truthful and meaningful insights, it is essential for the machine learning community to consider and incorporate qualitative data and ethical perspectives in their analyses.

In conclusion, the preliminary analysis of sequence learning from the social game data has revealed interesting patterns in player behavior. However, further research is needed to integrate additional information, consider dynamic perspectives, and incorporate qualitative data to gain a more comprehensive understanding of the intricacies of player interactions within the game. By addressing these challenges and limitations, we believe that it is possible to provide deeper insights into the complexities of social dynamics, leading to more informed and responsible AI-driven decision-making in social service provision.

6. Conclusions

Social games are a powerful mechanism to self-identify on fairness, and machine learning provides a tool to understand the human behaviors in the games. This works presents a sequence learning approach over data gathered in a social game. Results show that some interpretations can be gathered from the patterns found, though some results could be biased due to gender issues in the players distribution. Therefore, future work should repeat the analysis in a non-biased game. Moreover much more future work is required to provide truthful patterns that include the complexity of the dynamics of the game, in order to infer behaviors that could be used by social science. This dynamics include the consideration of qualitative information (i.e. externalities).

Acknowledgements

This work received support from the Generalitat de Catalunya via the Consolidated Research groups 2021 SGR 01125 and 2021 SGR 00838 as well as the AI FORA research project funded by the VolkswagenStiftung in the programme: Artificial Intelligence and the Society of the Future.

References

[1] Oliver TH, Doherty B, Dornelles A, Gilbert N, Greenwell MP, Harrison LJ, et al. A safe and just operating space for human identity: a systems perspective. The Lancet Planetary Health. 2022 11;6:e919-27.

[2] Ahrweiler P. Agent-based simulation for science, technology, and innovation policy. Scientometrics. 2017 1;110:391-415.

[3] Epstein JM. Generative social science : studies in agent-based computational modeling. Princeton University Press; 2006.

[4] Epstein JM. Agent-based computational models and generative social science. Complexity. 1999;4:41-60.

[5] Herget F, Kleppmann B, Ahrweiler P, Gruca J, Neumann M. How Perceived Complexity Impacts on Comfort Zones in Social Decision Contexts—Combining Gamification and Simulation for Assessment. Springer Proceedings in Complexity. 2022:203-15.

[6] Szczepanska T, Antosz P, et al . GAM on! Six ways to explore social complexity by combining games and agent-based models. International Journal of Social Research Methodology. 2022;25:541-55.

[7] Anzola D, Barbrook-Johnson P, Gilbert N. The Ethics of Agent-Based Social Simulation. Journal of Artificial Societies and Social Simulation. 2022 10;25.

[8] Özge Dilaver, Gilbert N. Unpacking a Black Box: A Conceptual Anatomy Framework for Agent-Based Social Simulation Models. Journal of Artificial Societies and Social Simulation. 2023 1;26.

[9] Bermudez RS, Sison AM, Medina RP. Extended PrefixSpan for Efficient Sequential Pattern Mining in a Game-based Learning Environment. APIT '20: Proceedings of the 2nd Asia Pacific Information Technology ConferenceJanu. ACM; 2020. p. 118-22.

[10] Pei J, Han J, Mortazavi-Asl B, Pinto H, Chen Q, Dayal U, et al. PrefixSpan: Mining sequential patterns efficiently by prefix-projected pattern growth. Proceedings - International Conference on Data Engineering. 2001:215-24.

[11] Taub M, Azevedo R. Using Sequence Mining to Analyze Metacognitive Monitoring and Scientific Inquiry based on Levels of Efficiency and Emotions during Game-Based Learning. Journal of Educational Data Mining. 2018;10(3):1-26.

[12] Ayres J, Flannick J, Gehrke J, Yiu T. Sequential pattern mining using a bitmap representation. Proceedings of the eighth ACM SIGKDD International Conference on Knowledge discovery and Data Mining - KDD. ACM Press; 2002. p. 429-35.

[13] Dumartinet J, Foppolo G, Forthoffer L, Marais P, Croitoru M, Rabatel J. Contextual Sequential Pattern Mining in Games: Rock, Paper, Scissors, Lizard, Spock. Research and Development in Intelligent Systems XXXII. 2015:375-80.

[14] Mordvanyuk N, Bifet A, López B. VEPRECO: Vertical databases with pre-pruning strategies and common candidate selection policies to fasten sequential pattern mining. Expert Systems with Applications. 2022 10;204:117517.