The use of big data in adaptive gamification in collaborative location collecting systems: a case of traveling behavior detection

María Dalponte Ayastuy$^{1,2[0000-0002-1412-5694]}$ and Diego Torres$^{1,2[0000-0001-7533-0133]}$

^{1} Depto CyT, Universidad Nacional de Quilmes, Bernal, Argentina
`mdalponte@unq.edu.ar`
^{2} LIFIA, CICPBA-Facultad de Informatica, Universidad Nacional de La Plata, La Plata, Argentina.
`diego.torres@lifia.info.unlp.edu.ar`

Abstract. Collaborative location collecting systems (CLCS) is a particular case of collaborative systems where a community of users collaboratively collects data associated with a geo-referenced location. Gamification is a strategy to convene participants to CLCS. However, it cannot be generalized because of the different users' profiles, and so it must be tailored to the users and playing contexts. A strategy for adapting gamification in CLCS is to build game challenges tailored to the player's spatio-temporal behavior. This type of adaptation requires having a user traveling behavior profile. Particularly, this work is focused on the first steps to detect users' behavioral profiles related to spatial-temporal activities in the context of CLCS. Specifically, this article introduces: (1) a strategy to detect patterns of spatial-temporal activities, (2) a model to describe the spatial-temporal behavior of users based on (1), and a strategy to detect users' behavioral patterns based on unsupervised clustering. The approach is evaluated over a Foursquare dataset. The results showed four types of behavioral atoms and nine types of users' behavioral patterns.

Keywords: Adaptive gamification challenges · Spatial-temporal user profiling · Users behavioural patterns

1 Introduction

Collaborative location collecting systems (CLCS) is a particular case of collaborative systems where a community of users collaboratively collects data associated with a geo-referenced location. The community of users travels around the globe collecting data.There are a vast number of CLCS. For example, the so-called Location Based Social Network, are CLCS where people share with their friends visited places; Foursquare (https://foursquare.com/) is a well-known case. Another example is the citizen science collecting systems that allow users to collect location-based data with a scientific goal. For example, iNaturalist (https://www.inaturalist.org/) is a biodiversity mapping social system where users spot in a map, using a mobile application, the visualization of any living being. In many of these systems with users worldwide, a large amount of data is already stored[10, 8, 15] .

CLCS should develop strategies to "convene participants, keep them active and committed with the specific project's task, keep them engaged with the project, and make them feel part of it" [3]. The use of game elements in non-game contexts, known as **gamification**, is a widespread approach to increase user engagement [5]. Nevertheless, it is also well known that gamification cannot be generalized because of the users' different motivations, personalities, needs, or values. So it must be tailored to the users, and playing contexts [4]. This research field is known as **adaptive gamification** which is presented as a promising possibility to improve user engagement towards these systems [6].

One of the most used game elements in gamified collaborative systems is challenges[3]. A game challenge is a task or problem in which difficulty depends on the user's skills, abilities, motivation, and knowledge [7] and count toward progress and outcomes. However, most of the use of this game element is not tailored to the user.

There is a wide range of types of challenges detailed in the literature [14]. Particularly, those that require endurance faculties or those that require sustaining a temporality and rhythm can be mentioned, which are the ones considered to develop this proposal. To develop challenges of this type in a personalized way, it is necessary to categorize people based on how they interact with the CLCS in terms of distance traveled and time between data collection moments (check-in).

This work is focused on the first steps to detect users' behavioral profiles related to spatial-temporal activities in the context of CLCS. Specifically, this article introduces: (1) a strategy to detect patterns of spatial-temporal activities, (2) a model to describe the spatial-temporal behavior of users based on (1), and finally, a strategy to detect users' behavioral patterns. These patterns will be the input for *endurance* and *rhythm* challenge adaptation in CLCS.

Specifically, this work presents three contributions. (a) A way to characterize the playing activities in terms of the invested time, the traveled distance, and the number of performed actions given a short time frame, for example, a day or a couple of hours; then these activities are categorized into categories called behavior atoms. Then, (b) a description of each user traveling gaming behavior is presented by sequencing the behavior atoms in a long time interval, for example, a year. Finally, (c) categorization of the users detecting similarities in their temporal sequences (b).

An unsupervised clustering strategy is proposed to infer user profiles from analyzing their traveling behavior with a time series strategy. The approach will be evaluated with a Foursquare dataset. The results show the detection of four behavioral atoms and nine types of users' behavioral patterns.

This paper is structured as follows: in Section 2 the related work is described, Section 3 gives the motivation of this work in terms of two specific problems.

Section 4 details the proposed approach to these problems, Section 5 describes the steps of the approach over a Foursquare dataset, Section 6 presents some discussions around possible improvements to this work, and finally, the conclusions and further work are given in Section 7.

2 Related Work

A user profile is a central component of information systems such as adaptive systems, and it has been widely studied. Ponciano et al. [12], and Aristeidou et al. [2] worked in profiling the users' motivations and contribution patterns in citizen science projects, looking at engagement metrics. Several works have been done specifically with Foursquare datasets to estimate the user's behavior. The work in [10] studies the geo-temporal dynamics of user activity to unfold place transitions and identify sequences of activities. Also, mobile users' spatial-temporal activity preference was inferred from the user-generated digital footprints in LBSNs [16]. Long et al. [9] focus on exploring the local geographic topics using the Latent Dirichlet Allocation (LDA) model to discover the local geographic topics from the check-ins datasets.

To estimate sequence similarity and feature representations for sequence classification and clustering is one of the main tasks of exploratory data mining and is used in many fields such as bioinformatics, pattern recognition, image analysis, or machine learning.

None of the mentioned contributions are related to personalizing the gaming experience based on the space-time behavior record, as is introduced in this article.

3 Problem Statement

The available literature records the work in the classification of challenges from different aspects. Vahlo et al. [14] perform an exhaustive classification of thirty-eight videogame challenge types into five challenge types: Physical, Analytical, Socioemotional, Insight, and Foresight. Challenges of *endurance* and *rhythm* are considered in this work as the types of challenges most related to the activities of the CLCS since they require a change of geographical position. In the context of the CLCS application, an *endurance challenge* can be related to the number of check-ins or the frequency. For example, to obtain five check-ins in a day or to travel ten kilometers per day for three days. A *rhythm challenge* would be a situation that involves check-in at certain times or repeating a sequence of activities. For example, to check in before noon and after dinner through four days.

This work aims to classify users according to their traveling behavior profile, focusing on distances traveled, invested time, and number of check-ins in a period. The input is a check-in set with information about their geographic position and a timestamp.

For this classification purpose, this work proposes a model of user behavior over time that involves the definition of two strategies. Firstly, a strategy to synthesize the user's interaction with the CLCS in a single value for each time frame based on the traveling aspects. Lastly, a strategy to classify users from the similarity of their time series.

3.1 Definitions

Before going into details about the approach, it is necessary to present some definitions that will give a conceptual context.

A *check-ins dataset* is a dataset that includes the log of users' check-ins. It is defined by user id, latitude, longitude, and timestamp.

In this work, the *time frame* is the interval applied to group the user's activities and aggregate them from the temporal and spatial point of view. Time frame size could be variable, for example, in terms of hours or days. Therefore, the check-ins dataset is divided into fixed-size time frames. The time frames are calculated from the interval covered by the dataset's samples and a frame size parameter. Time frames are used to define the size of the behavioral atom described below.

The activities aggregation are computed by the $checkInCounts$, $investedTime$, and $traveledDistance$ functions as are defined below.

Definition 1. *userDataFrame(user_id,timeFrame): List of checkins dataset entries by user_id,ordered by timestamp..*

Definition 2. *checkInCounts(user, timeFrame): the number of entries in userDataFrame(user,timeFrame)*

Definition 3. *investedTime(user, timeFrame): difference between the first and last check-in timestamp in userDataFrame(user,timeFrame)*

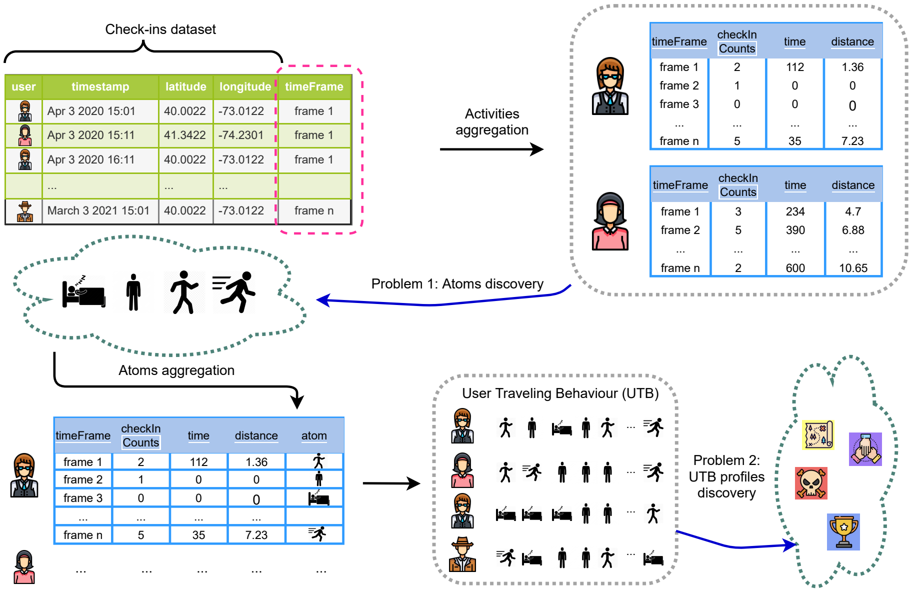
Definition 4. *traveledDistance (user, timeFrame): sum of all distances between two consecutive check-ins in userDataFrame(user,timeFrame)*

A **behavioral atom** is a categorical value that describes the user's interaction with the CLCS within a time frame from the mentioned activity aggregation.

A **user travel behavior (UTB)** is a sequence of behavioral atoms organized as a time series in chronological order to describe the user's behavior during the check-in dataset period.

Figure 1 shows an example of a check-in dataset and the transformation into a UTB set.

Lastly, given the users' time series, it is possible to define a similarity criterion between them based on the value patterns (or variation in the values) that make up the sequence. The second problem to be addressed was classifying users based on this idea of similarity, and with this objective, different machine learning techniques were applied. The following are the definition of the problems.



user	timestamp	latitude	longitude	timeFrame
👤	Apr 3 2020 15:01	40.0022	-73.0122	frame 1
👤	Apr 3 2020 15:11	41.3422	-74.2301	frame 1
👤	Apr 3 2020 16:11	40.0022	-73.0122	frame 1
	...	...	...	
👤	March 3 2021 15:01	40.0022	-73.0122	frame n

timeFrame	checkIn Counts	time	distance
frame 1	2	112	1.36
frame 2	1	0	0
frame 3	0	0	0
...	...	...	...
frame n	5	35	7.23

timeFrame	checkIn Counts	time	distance
frame 1	3	234	4.7
frame 2	5	390	6.88
...	...	...	...
frame n	2	600	10.65

timeFrame	checkIn Counts	time	distance	atom
frame 1	2	112	1.36	🚶
frame 2	1	0	0	🧍
frame 3	0	0	0	🛌
...	...	...	...	
frame n	5	35	7.23	🏃
...	...	...	...	...

Fig. 1. Data transformation into UTBs. The atoms alphabet in this example is made up with 4 elements: L (low), M (moderate), A (average), H (high)

Problem 1: Detect the atoms alphabet that better describes spatial-temporal activities patterns in the check-in dataset. The atoms alphabet will be the elements that shape the UTBs.

Problem 2: Detect users behavioural profiles by means of theirs UTBs.

4 Approach

For the analysis of the first problem, a KMeans-based clustering of the aggregated activities is performed. Clustering is an unsupervised machine learning technique to identify groups of samples based on their similarity. This technique allows the discovery of features from sample data, and in particular to this first problem, the similarity among aggregated activity records. Precisely, three aggregated values are calculated: the number of check-ins of the time frame, the time that elapses between the first and the last check-in of that time segment, and the distance traveled, based on the geographical distance between consecutive registered positions.

Regarding the second problem analysis, a study is carried out using time series k-means. To measure the similarity between user time series, and given that each user's activity may not occur on the same time frames, at least two ways of normalizing the observations can be considered: an absolute and a relative approach. In the former case, the temporary frames are fixed depending on the total period of the dataset (or period of analysis), and when a user has no activity in a time frame, it is filled in with zeros (or values that do not deviate the clustering). The drawbacks of the absolute approach are that users are compared in a synchronized way, apparently losing the possibility that two users with the same behavior pattern will result in the same cluster: a similar number of check-ins and distance traveled, but at different time frames. On the other hand, with the relative approach, the overall calendar is ignored.

The approach is detailed below through a case study using the Foursquare dataset for New York between April 2012 and February 2013.

5 Evaluation

The following steps are carried out to address both problems: First, a pre-processing of the dataset is carried out to eliminate null data, standardize the data, and analyze the correlation between the variables. After this, the elbow curve is analyzed to decide the number of clusters. Next, the clustering algorithm is run over the dataset with the number of cluster parameters defined before. Then, how the clusters are formed and distributed are plotted and analyzed. Finally, the categories are determined.

All the analysis was developed in Kaggle environment[3], using Python language with Pandas and sklearn as main libraries.

5.1 The Foursquare dataset

In Foursquare, users visit and comment about the places of interest, and share them with their friends.

The foursquare dataset has information about users' check-ins, particularly the venue id, venue category, geographic location, and timestamp. This work uses the Foursquare dataset from New York with 227,428 records between April 2012 and February 2013[16]. Each record includes check-in timestamp, GPS coordinates as a latitude/longitude pair, user id, venue id, venue category id, venue category name, and time zone offset. This work focus only on user id, GPS coordinates, and timestamp. Time zone offset was ignored because all of them had the same value.

The dataset describes the activity information of 1,083 users through 318 days. It is important to note that there are periods in which the dataset does not have check-ins. The largest occur between August 21, 2012, and September 12, 2012, and between September 19, 2012, and October 10, 2012. Although it would be nice to have these data, we consider that they will not alter the analysis. In time series figures, they can be seen as blank spaces.

A one-day time frame was used in this evaluation. All of the activities a user performed on the same day were grouped in the same time frame.

---

[3] https://www.kaggle.com

5.2 Activities aggregation

The aggregation step computed for each user and time frame the aggregated values defined in Section 3.1. The result was a pandas dataframe with user id, checkIncounts, investedTime, traveledDistance, and day on each row. The resulting dataframe counts 94,387 rows.

The mean of check-ins count was 2.4 (sd 2.5), between one check-in and 71, but the 75% of the samples had three check-ins. The mean invested time was 269 minutes (sd 404.4), equivalent to 4.5 hours, being zero the minimum value (when at most a check-in was made) and 1,438 minutes (almost 24 hours). The mean traveled distance in a day was 4.7 km (sd 9.87), between zero and 303 km.

The data reflects a high standard deviation in most of the variables. This revealed that users had different characteristics in their behaviors instead of monotony. Indeed, it is coherent and validates the purpose of this article.

As we can notice, this dataframe does not include rows representing the absence of activities in a day. For example, if a user did not perform a check-in on a given day, it does not appear.

5.3 Problem #1 analysis

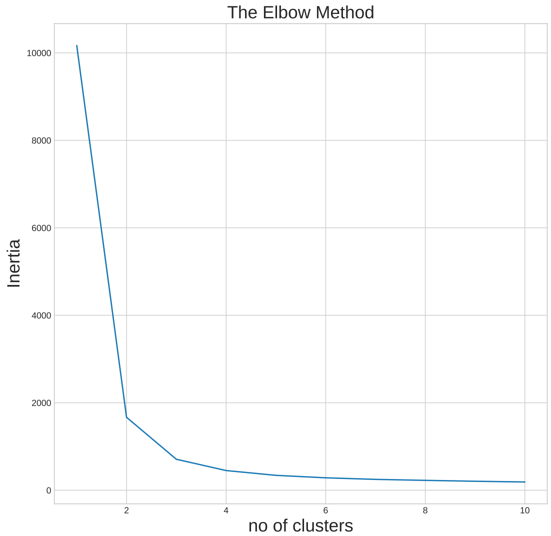


Fig. 2. Elbow method for behavioural atoms

As a pre-processing stage, two steps were developed. Firstly the normalization of the samples with the described absolute approach in such a way as to record the activity of all users in all time intervals, filling in with zero values when there was no user activity. As a result, the dataset's size is 344,394 rows, which means 250,007 zero-valued records were added. Lastly, a common practice is to standardize the aggregated values to the range between 0 and 1 to avoid a bias in the distance calculation carried out in clustering.

A K-means clustering was executed in order to detect behavioral atoms. A critical step for any unsupervised clustering algorithm is to determine the optimal number of clusters into which the data can be clustered. The *elbow method* [13] is one of the most popular methods for determining this optimal value of *k*. The values of k are iterated from 1 to 11 and calculate the inertia or cost function (the sum of squared distances of samples to their closest cluster center) for each value of k in the given dataset. The algorithm is described in the following code chunk:

```
inertia=[]
for i in range(1,11):
    kmeans = KMeans(n_clusters= i, init='k-means++', random_state=0)
    kmeans.fit(data)
    inertia.append(kmeans.inertia_)
```

To determine the optimal number of clusters, the selected value of k is the one at the "elbow". Using the elbow of a curve as a cut-off point is a common heuristic in mathematical optimization to choose a point at which the cost function (average of the distances or inertia) decreases with respect to the number of clusters, do not justify adding another cluster. In other words, adding another cluster does not provide much better modeling of the data because the total within-cluster variation is minimized. As can be seen in Figure 2, four clusters were a reasonable number.

The KMeans clustering was executed with $k = 4$. In a complementary way, an analysis of the clusters' arrangement in 2d and 3d plots and the clusters' distribution was done. Figure 3 shows the cluster arrangement in 2-D and 3-D plots, where a clear relationship with the time variable can be seen. The Figure 4 explains the distribution of the atom clusters sizes. The x-axis details each cluster, and the y axis the cluster's element count. The first cluster (red color) has the vast majority of the records (89.5 percent) and the reason, as was expected, is that it includes all the zero entries introduced by the normalization process and the activity records of a single check-in (58,399 items, 16,95 percent). This cluster was called the *low* cluster because it includes the records with little or no activity. It is followed by the *moderate* cluster (green color) with 11,208 records representing 3.25 percent of the total, including records with moderate activity. The *average* cluster (blue color) with 11,195 records representing almost the same percent; and finally, the *high* cluster (cyan color) with 13,539 records representing the 3,93 percent of the total.

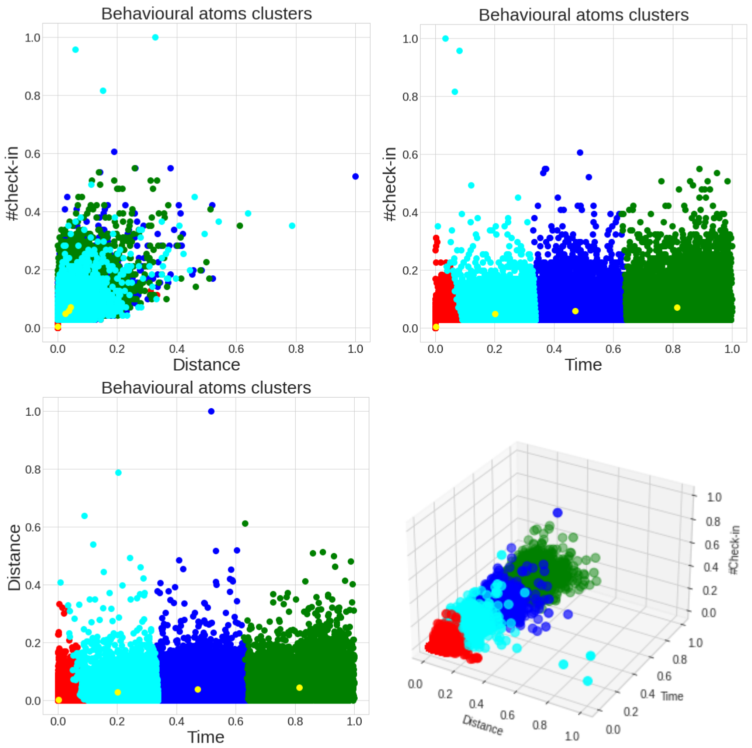


Fig. 3. Clusters

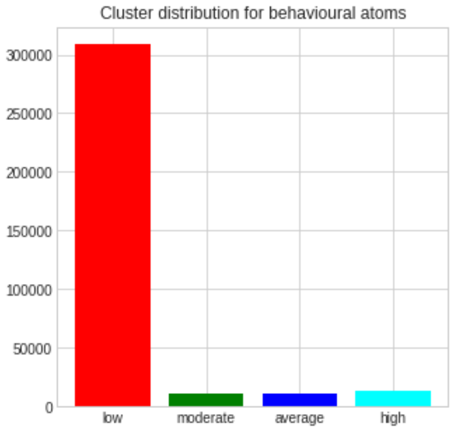


Fig. 4. Behavioural atoms distribution in clusters

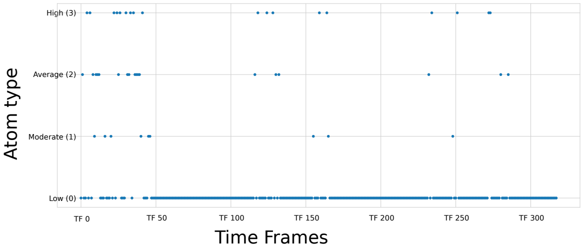


Fig. 5. User 14: 318 days series (UTB)

5.4 Problem #2 analysis

A time series is a sequence of observations of a continuous variable, and a sequence is analogous to a discrete or categorical variable. For this reason, they are considered a particular type of time series. One of the most popular approaches in time series classification is the use of the nearest neighbor (NN) classifier in conjunction with a distance function[1].

Having computed the behavioral atoms for each aggregated time frame using four categories (L: low, M: medium, A: average, H: high), the UTB sequences were constructed for each user. Figure 5 shows an example of an UTB series: The *y axis* details the atoms type, and the *x axis* the time frames. Each point represents a type of atom the user performed in the time frame. As the UTBs are time series, they can be classified using time series k-means clustering. On the other hand, to address the problem of synchronization of the absolute approach described in section 4, dynamic time warping (DTW) was chosen as a function of distance, instead of euclidean distance, which was applied in the approach of Problem 1 (Section 5.3). In time series analysis, DTW is one of the algorithms for measuring similarity between two temporal sequences, which may vary in speed. It has been shown that dynamic time warping (DTW) distance with an NN classifier worked effectively [11].

As was done to analyze Problem 1, the elbow method was developed to determine the optimal number of clusters for the classification of UTBs. The result is shown in Figure 6, and it can be seen that an optimal cluster number is 9. The time series are plotted in Figure 7. Each cell of the graph plots a UTB cluster using gray-colored lines. Above these, the cluster average is plotted with a red line. It can be seen that cluster 6 includes the users who had less activity; that is, they spent more time in a low or moderate state (see red line). Furthermore, it is seen that cluster 7 represents the most active users, or in other words, with more periods in the high state. This cluster is followed by cluster 0 and cluster 4.

As was done to analyze Problem 1, the elbow method was developed to determine the optimal number of clusters for the classification of UTBs. The result is shown in Figure 6, and it can be seen that an optimal cluster number

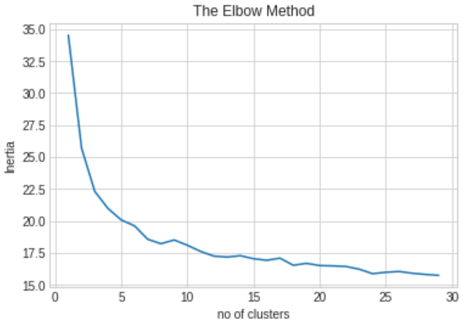


Fig. 6. Elbow method computation for users' UTB up to 30 clusters

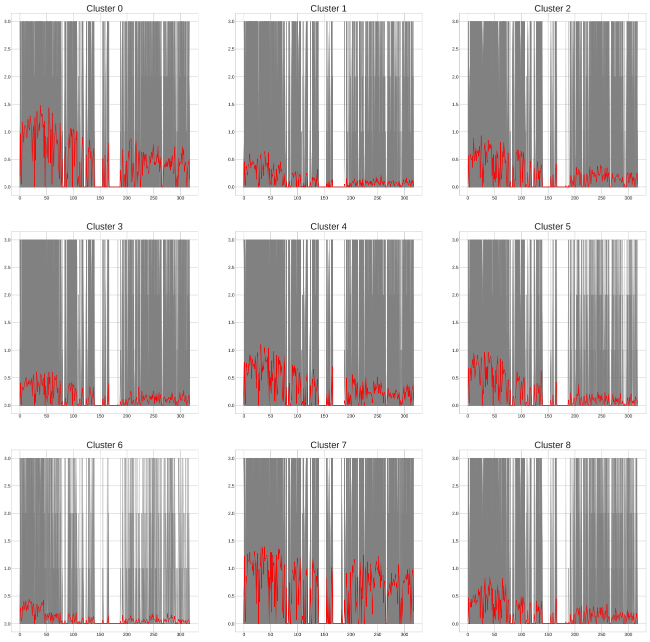


Fig. 7. Clusters of UTBs

is 9. The time series are plotted in Figure 7. Each cell of the graph plots a UTB cluster using gray-colored lines. Above these, the cluster average is plotted with a red line. It can be seen that cluster 6 includes the users who had less activity; that is, they spent more time in a low or moderate state (see red line).

Furthermore, it is seen that cluster 7 represents the most active users, or in other words, with more periods in the high state. This cluster is followed by cluster 0 and cluster 4. Also, it can be noticed the empty areas where the dataset does not include check-ins as was mentioned in Section 5.1.

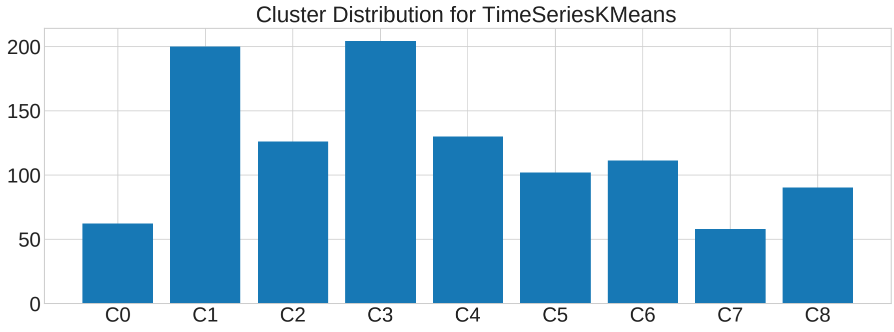


Fig. 8. Time series clusters distribution

Therefore the time series k-means clustering with $k = 9$ lead to a cluster distribution depicted in Figure 8. It can be noted that the clusters that were identified as the most active are those with the fewest number of users, denoting a relatively small group of people who are committed to the game.

6 Discussion

The way the UTBs clusters' average is computed assigns a numerical representation to each behavioral atom. Consequently, when the red line has high values, it identifies clusters with a considerable amount of atoms of type H and A. Nevertheless, this line does not plot categorical values because it represents the average among the numerical values given to the categories, i.e., L=0, M=1, A=2, H=3. Consequently, it is necessary to carry out an analysis where the clusters' central lines can be plotted in such a way that they truly represent the complete series. For example, taking the value that occurs the most in each timeframe, considering the large number of zero values, does not bias this centroid and make it fall. Building a more real centroid would allow a more refined grain analysis on each cluster's users' behavioral atoms.

On the other hand, based on what is observed in 8, it can be thought that some clusters are more cohesive and therefore more focused on the recommendation of challenges. As we have described in Section 3, traveled distance and invested time can be an input for endurance game challenges. In this work, the behavioral atoms describe activity levels. The height of the red line in Figure 7 shows part of this information.

The user level can be a reference to provide challenges that can motivate them to improve. On the one hand, these challenges should exceed the user's current level and allow users to overcome the challenges. Otherwise, the challenge will be unattainable and will therefore generate disappointment or boredom in the users.

A high red line can be related to a high level of user endurance in a period. The analysis could help tailor endurance challenges with a high-intensity level to users that could reach those challenges.

Regarding the rhythm game challenges, it is possible to consider the similarity of sequences and the repetition of sub-sequences among users. In the analysis carried out, grouping the clusters using DTW takes into account grouping people with similar sequences of atoms. However, the analysis of recurrent sub-sequences is a pending task to be addressed in the future.

7 Conclusions and further work

In this work, a user classification mechanism was presented in terms of their movement behavior. A possible future work is to reconsider the temporal aspect taking the frequency of check-ins instead of the time elapsed between the first and the last check-in. The results show that several users only perform two check-ins, one at the beginning of the day and another at the end of the day. So the intermediate activity is measured but may not be significant.

It may also be of interest for user profiling to calculate the frequency of check-ins for each user in a venue category (such as hotel, cinema, mall) and then measure the similarity between users who have the same frequencies.

References

1. Abu Alfeilat, H.A., Hassanat, A.B., Lasassmeh, O., Tarawneh, A.S., Alhasanat, M.B., Eyal Salman, H.S., Prasath, V.S.: Effects of distance measure choice on k-nearest neighbor classifier performance: A review. Big Data **7**(4), 221–248 (2019). https://doi.org/10.1089/big.2018.0175, https://doi.org/10.1089/big.2018.0175, pMID: 31411491
2. Aristeidou, M., Scanlon, E., Sharples, M.: Profiles of engagement in online communities of citizen science participation. Computers in Human Behavior **74**, 246–256 (2017). https://doi.org/https://doi.org/10.1016/j.chb.2017.04.044, https://www.sciencedirect.com/science/article/pii/S0747563217302868
3. Ayastuy, M.D., Torres, D., Fernández, A.: Adaptive gamification in Collaborative systems, a systematic mapping study. Computer Science Review **39**, 100333 (2021). https://doi.org/https://doi.org/10.1016/j.cosrev.2020.100333, https://www.sciencedirect.com/science/article/pii/S1574013720304330

4. Böckle, M., Novak, J., Bick, M.: Towards adaptive gamification: a synthesis of current developments. In: Proceedings of the 25th European Conference on Information Systems (ECIS). Guimarães, Portugal (2017), https://aisel.aisnet.org/ecis2017_rp/11
5. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From Game Design Elements to Gamefulness: Defining "Gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. pp. 9–15. MindTrek '11, ACM, New York, NY, USA (2011). https://doi.org/10.1145/2181037.2181040, http://doi.acm.org/10.1145/2181037.2181040, event-place: Tampere, Finland
6. Göbel, S., Wendel, V.: Personalization and adaptation. In: Serious Games, pp. 161–210. Springer (2016)
7. Iversen, S.: In the double grip of the game: Challenge and Fallout 3. Game Studies **12** (2012), http://www.gamestudies.org/1202/articles/in_the_double_grip_of_the_game
8. Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., Zimmerman, J.: I'm the mayor of my house: Examining why people use foursquare - a social-driven location sharing application. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. p. 2409–2418. CHI '11, Association for Computing Machinery, New York, NY, USA (2011). https://doi.org/10.1145/1978942.1979295, https://doi.org/10.1145/1978942.1979295
9. Long, X., Jin, L., Joshi, J.: Exploring trajectory-driven local geographic topics in foursquare. In: Proceedings of the 2012 ACM conference on ubiquitous computing. pp. 927–934 (09 2012). https://doi.org/10.1145/2370216.2370423
10. Noulas, A., Scellato, S., Mascolo, C., Pontil, M.: An empirical study of geographic user activity patterns in foursquare. Proceedings of the International AAAI Conference on Web and Social Media **5**(1) (Jul 2011), https://ojs.aaai.org/index.php/ICWSM/article/view/14175
11. Petitjean, F., Ketterlin, A., Gançarski, P.: A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognition **44**(3), 678–693 (2011). https://doi.org/https://doi.org/10.1016/j.patcog.2010.09.013, https://www.sciencedirect.com/science/article/pii/S003132031000453X
12. Ponciano, L., Pereira, T.E.: Characterising volunteers' task execution patterns across projects on multi-project citizen science platforms. In: Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems. IHC '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3357155.3358441, https://doi.org/10.1145/3357155.3358441
13. Tibshirani, R., Walther, G., Hastie, T.: Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **63**(2), 411–423 (2001). https://doi.org/https://doi.org/10.1111/1467-9868.00293, https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00293
14. Vahlo, J., Karhulahti, V.M.: Challenge types in gaming validation of video game challenge inventory (CHA). International Journal of Human-Computer Studies **143**, 102473 (2020). https://doi.org/https://doi.org/10.1016/j.ijhcs.2020.102473, https://www.sciencedirect.com/science/article/pii/S1071581920300756
15. Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., Belongie, S.: The inaturalist species classification and detection dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)

16. Yang, D., Zhang, D., Zheng, V.W., Yu, Z.: Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns. IEEE Transactions on Systems, Man, and Cybernetics: Systems **45**(1), 129–142 (2015). https://doi.org/10.1109/TSMC.2014.2327053