Computational Intelligence in Serious Games: A Case Study to Identify Patterns in a Game for Children with Learning Disabilities

Andreia M. Domingues
University of Mogi das Cruzes (UMC)
Mogi das Cruzes/SP, Brazil

Sabrinna Delgado
Informatics and Knowledge Management Postgraduate Program, Nove de Julho University (UNINOVE)
São Paulo/SP, Brazil

Marcia A.S. Bissaco
Technological Research Center – Postgraduate Program in Biomedical Engineering, University of Mogi das Cruzes (UMC)
Mogi das Cruzes/SP, Brazil

Sidnei A. de Araújo
Informatics and Knowledge Management Postgraduate Program, Nove de Julho University (UNINOVE)
São Paulo/SP, Brazil

ABSTRACT
This work explores the application of computational intelligence techniques in a serious game (SG) for children with learning disabilities. Specifically, Data Mining (DM) techniques such as Decision Tree and Apriori algorithms were applied to identify the existence of patterns that allow a better understanding of the children's profiles involved in the game. The data analyzed are related to the interaction of twenty children with the considered SG, which consists of a three-dimensional virtual zoo, developed with features that appeal to the preferences of children about nine years old to assist and motivate their learning. The results obtained in the conducted experiments revealed patterns in the profiles of the game's players under analysis, allowing to identify some characteristics that can help the psychopedagogical team. These findings can also enable the improvement of the game making it adaptable to different player profiles.

Keywords
Computational Intelligence, Data Mining, Pattern Recognition, Decision Tree, Apriori Algorithm

1. INTRODUCTION
The development of individuals occurs with greater speed in the early years of childhood. It is in this phase of growth and development that the individuals build skills that are combined to produce sophisticated accomplishments such as walking, thinking, and communicating. From zero to eight years of age, children develop motor, emotional, social, mathematical logic, and vocabulary activities, some of which are indicative of healthy psychological development. Significant variations in the achievement of expected goals for their age can make children vulnerable to poor educational results, low self-esteem and social interaction [1].

Learning disability can be defined as a psychological disorder that usually begins in early childhood and compromises the development of functions linked to the biological maturation of the central nervous system, as the language development. According to the Diagnostic and Statistical Manual of Mental Disorders [2], 30% of school-age children have some type of learning disability.

Such disabilities may decrease with age, but milder deficits in reading, written expression, and math skills may persist into adulthood. However, these children are not incapable of learning, as their intellectual integrity and learning ability are preserved [3]. Education specialists believe that pedagogical interventions should consider the use of an educational SG, because when playing, the child acts spontaneously and, motivated by the challenge, uses all the acquired knowledge in addition to creating new mental schemes as they continue playing [4,5].

During the learning process, SG allow children to explore their imagination in a stimulating and motivating environment in search of texts with real meaning for them [4]. An educational SG can have an impact on children's skills and be potentially useful for teachers given the set of available resources in this kind of tool [1].

The SG development makes it possible to adapt teaching to the different psychological profiles of children [4]. SG can help children to adopt an active posture, as they are motivated by the challenge and the playful characteristics that these games provide.

Palma et al. [6] mention the importance of analyzing the profile of school-age players by observing their habits and motivations during the interaction with this type of game, as well as considering the differences between genders and how long they play. Thus, it would be possible to develop SG adaptable to the profile of the players.

The analysis of game data, as a decision-making tool, allows the recognition of player behavior patterns, making it possible to improve the design, quality, as well as predicting their errors to previously redirect them [7].

In this context, this work investigates the discovery of patterns in a SG developed for children with learning disabilities which stratify characteristics of their profiles through the application of computational intelligence techniques in DM tasks.

2. THEORETICAL BACKGROUND
This section presents the theoretical basis used for the development of this work. First, in Subsection 2.1, the concepts of Serious Games were presented. Subsection 2.2 describes Data Mining and, finally, in the subsection 2.3 concepts of decision tree and Apriori algorithms are presented as well.

2.1 Serious Games

Traditional educational methods applied in the teaching-learning process in the early years of formal education are not always suitable for children with learning disabilities. This problem affects one third of children in the initial phase of literacy and can cause school failure and dropout. To help solve this problem, experts suggest a special teaching method that includes the use of computerized educational games, since they allow the exercise of thinking voluntarily, with rules that involve the player's feeling of fun. In addition, they can establish a relationship between what the child is intended to know and the knowledge he already has [5].

A particular category of educational games is the “Serious Games”, which mainly aim at simulating practical day-to-day situations that are designed primarily for educational purposes rather than purely for entertainment. In addition, they can be a powerful tool to develop and train acquisition of new knowledge [1,4,5,8,9].

The use of SGby children and adolescents with disabilities can contribute to improving their learning and meeting their needs in an inclusive educational environment [8]. According to Flogie et al. [5], SG environments integrating commercial game technologies combined with computational intelligence methods can offer personalized learning opportunities, providing more motivation and engagement for the learner, teach 21st-century skills, and provide an environment for authentic and relevant assessment.

2.2 Data Mining

The evolution of database technology (1980s onwards) made it possible for organizations to continuously collect and store huge amounts of data about customers, suppliers, products and services. However, in the 1990s it was realized that such data were being underutilized. Then, this scenario motivated the emergence of the concept of Data Mining (DM) [10].

DM can be defined as a process of exploring and analyzing data using statistical and computational intelligence algorithms, in order to discover patterns or rules that allow a better understanding of the information contained in the analyzed dataset [10,11].

The main tasks of DM are classification, discovery of association rules and clustering. Classification aims to examine the input data set and produce descriptions of the characteristics of these data for each class [10]. Association rules describe discovered patterns in the form of high-level IF...THEN rules[11]. Usually, the algorithms used in this task produce a large number of rules and it is up to the user to select those that he deems most useful for a given application. Finally, clustering aims to group data from a dataset into two or more groups, employing some measure to express the similarity between patterns based on their attributes.

The most common computational intelligence techniques and algorithms employed in DM are Decision Trees (DT), Artificial Neural Networks (ANN), Genetic Algorithms (GA), Fuzzy Logic, K-means, Support Vector Machines (SVM) and Bayesian Networks.

DTs and ANNs are very powerful techniques and widely used in classification tasks [12]. DTs are preferable in this kind of task, in many cases, since the acquired knowledge is represented through rules that facilitate the user's understanding. On the other hand, ANNs are important due to its ability for generalizing and to deal well with noise.

2.3 Decision Tree and Apriori Algorithms

Decision Tree (DT) is one of the simplest supervised learning algorithms and can be interpreted as a set of IF...THEN rules [12,13]. Because of this, DTs have been widely used in classification tasks, as an efficient way to build classifiers that predict classes based on the values of attributes that describe patterns. Thus, they can be used in various applications such as medical diagnostics, credit risk analysis, among other examples.

The key to the success of a DT algorithm is how to generate the tree, that is, how to choose the most important attributes to generate the rules and which rules can be discarded from the tree [12]. In this way, the attribute considered most important should be at the root of the tree. With this, a problem can be solved by applying the fewest rules. According to deAraujo et al. [14], since DTs do not always use all the attributes that represent a pattern for generating the set of rules, it has also the advantage of decreasing the computational time in classification tasks.

Among the DT algorithms are ID3 [13], C4.5 [15] and CART [16], being the first one of the most basic. The C4.5 algorithm, employed in this work, builds DTs from a training data set in the same way as the ID3 algorithm, using the entropy concept to define the importance of attributes. At each node of the tree, the C4.5 algorithm chooses the attribute that best partitions the set of samples into subsets tending to one category or the other. The attribute with the highest normalized information gain is chosen to make the decision [15].

The Apriori algorithm, proposed in[17], is one of the most well-known algorithms when it comes to mining association rules on large datasets. The name of the algorithm is because it uses prior knowledge of frequent itemset properties. It finds all sets of frequent items, called frequent itemsets. The first step of the Apriori algorithm is to count the occurrences of the items to determine the frequent itemsets of unit size. The later steps consist of two phases. First, the frequent itemsets found in the first step are used to generate candidate itemsets. Subsequently, a new search is performed in the dataset, counting the support of each candidate [17].

3. METHODOLOGY

To conduct the experiments, the WEKA Software [18] was employed. It consists of a tool that provides several algorithms from statistics and computational intelligence and, therefore, is widely used in DM tasks by the scientific community. Classification and generation of association rules tasks were performed using the DT and Apriori algorithms. Its valid to highlight that the J48 algorithm contained in the WEKA software and used in our experiments is an implementation of the known C4.5 algorithm.

The classification aimed to investigate the existence of patterns that could lead to the grouping of different player profiles. Nevertheless, the association rules made it possible to discover correlations between the data in order to derive new knowledge from the analyzed dataset.

The SG analyzed, whose opening screen is illustrated in Fig 1, has a three-dimensional virtual zoo as scenario and was developed with characteristics that appeal to the preferences of children around nine years old, in order to assist and motivate their learning.

Fig 1: Game Opening Screen

It works on the linguistic ability to identify words considering animal names (Rhino, Elephant, Giraffe, Zebra, Bear and Lion), as showedin Fig 2.

Fig 2: Game screen captured during a player's interaction with the game

Pre-test was applied to 74 children of 2nd and 3rd school year, identifying 12 girls and 8 boys (20 children) with reading difficulties in the language skills worked. They were divided into two groups, an experimental group (EG) and a control group (CG), to participate in the intervention with the virtual zoo. EG participated in all phases of the environment

Table 1. Data collected from the considered SG

Child	Gender	Age	School_year	Rhino	Elephant	Giraffe	Zebra	Lion	Bear	Points	Time
1	M	7	2	0.5	0.0	0.5	0.5	1.0	0.0	0-3	<10_min
2	F	7	2	0.0	0.0	0.5	0.0	0.5	0.0	0-3	10_20_min
3	M	7	2	0.0	0.0	0.0	0.0	0.5	0.5	0-3	>20min
4	F	7	2	0.0	0.0	0.0	0.5	1.0	0.0	0-3	10_20_min
5	M	8	2	0.5	0.0	1.0	1.0	1.0	0.5	>3<5	<10_min
6	F	8	2	0.0	0.0	0.5	0.0	0.5	0.5	0-3	>20min
7	M	8	2	0.0	1.0	0.5	0.5	1.0	1.0	>3<5	<10_min
8	F	8	2	1.0	0.5	1.0	1.0	1.0	1.0	>=5	<10_min
9	F	8	2	0.5	0.0	1.0	0.5	1.0	0.5	>3<5	<10_min
10	F	7	2	1.0	1.0	0.0	1.0	1.0	1.0	>=5	10_20_min
11	F	9	3	0.0	0.0	0.0	0.0	0.0	0.0	0-3	10_20_min
12	F	7	2	0.0	0.0	1.0	0.0	1.0	0.5	0-3	10_20_min
13	F	8	2	0.5	0.5	1.0	1.0	1.0	0.0	>3<5	<10_min
14	F	9	3	0.5	0.0	0.0	1.0	1.0	0.0	0-3	<10_min
15	M	6	2	0.0	0.5	0.0	0.0	0.5	0.0	0-3	>20min
16	F	7	2	0.5	0.5	1.0	1.0	0.5	0.0	>3<5	>20min
17	M	8	2	1.0	0.5	0.0	0.0	1.0	1.0	>3<5	<10_min
18	M	8	2	0.0	0.0	1.0	1.0	1.0	1.0	>3<5	<10_min
19	M	8	2	1.0	0.0	1.0	0.0	1.0	1.0	>3<5	<10_min
20	F	7	2	0.0	0.0	0.5	0.0	1.0	1.0	0-3	10_20_min

(recognition of words, syllables and letters of the alphabet) while CG participated only in the letter recognition phase applied at pre and post-tests, taking into account the time players spent on their answers. Then, these twenty children were invited to interact with the game without a predetermined time. The time spent to reach the objective of the game and the score obtained were collected to identify if the player was correct in the first (1 point), second (0.5 point) or third (0 point) attempt. The raw data extracted from the game to be analyzed by the considered DM techniques are shown in Table 1.

The data in Table I represent the players' performances during their interactions with the game, that is, these values were weighted according to the number of attempts they made to form the names of the six animals. It also shows the time spent by each child to form the names of the six animals, which varied from 5 to 28 minutes.

4. RESULTS

First, the BioEstat statistical software was used, which through the t-Student test showed bilateral p=0.0033 for the EG group, while the GC group showed bilateral p=0.0145. It shows that the time spent by children in the EG and CG groups decreased when comparing pre-test and post-test. Therefore, the number of attempts by the player to reach the tasks was recorded in both tests and a statistical analysis with t-Student test showed that virtual environment made possible a significant improvement (bilateral $p \leq 0.05$) in the post-test in relation with the pre-test.

Furthermore,it was applied the C4.5 algorithm in the data classification.The analysis performed considering the “Time” as class attribute demonstrates that it was possible to classify correctly 80% of the data, as can be seenin Fig 3, demonstrating the existence of patterns in the data. This is corroborated by the Kappa coefficient with a score of 0.67, which represents substantial agreement (values from 0.61 to 0.80) according toKundel and Polansky, 2003.

```
=== Summary ===

Correctly Classified Instances          16               80      %
Incorrectly Classified Instances         4               20      %
Kappa statistic                          0.6721
Mean absolute error                      0.1927
Root mean squared error                  0.3625
Relative absolute error                 45.2904 %
Root relative squared error             78.1897 %
Total Number of Instances               20
```

Fig 3: Statistical data obtained from the classification by DT considering “Time”as class attribute

Fig 4 illustrates the tree generated in this same analysis, that is, considering the “Time” as class attribute. It is possible to observe that the players who got the word Lion right on the first try are mostly children with 8 years old (the numbers of players are indicated in parentheses).

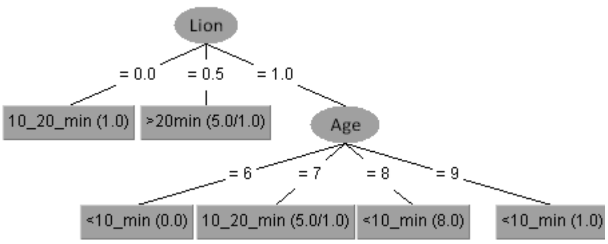

Fig 4:The DTproduced considering “Time” as class attribute

When analyzing players interaction data considering specific words, it was observed, for example, that those who got the word “Lion” right on the first attempt finished the game in up to 10 minutes. In addition, it is observed that most boys finished the game in up to 10 minutes while most girls finished it spendingfrom 10 to 20 minutes (see Fig 5).

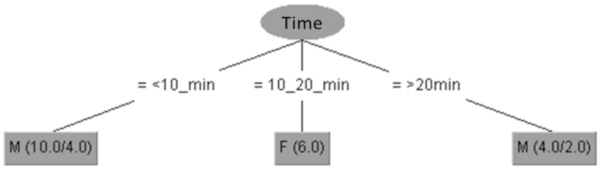

Fig 5: The DT produced considering “Lion” as class attribute

Still considering the players' performance for the “Lion” attribute (Fig 6), it was possible to classify correctly 90% of the data observing a Kappa coefficient of 0.74, also considered as substantial agreement.

```
=== Summary ===

Correctly Classified Instances          18               90      %
Incorrectly Classified Instances         2               10      %
Kappa statistic                          0.7436
Mean absolute error                      0.12
Root mean squared error                  0.2683
Relative absolute error                 36.699  %
Root relative squared error             67.19   %
Total Number of Instances               20
```

Fig 6: Statistical data obtained from the classification by DT considering “Lion” as class attribute

Finally, using the Apriori algorithm, association rules were generated as demonstrated in Fig 7. The rule 6, for example, shows that those players who got the word “Giraffe” right on the first attempt were inthe second year of school, while the rule 1 shows a knowledge that has already been discovered by using the DT illustrated in Fig5.

```
 1. Time=<10_min 10 ==> Lion=1.0 10    <conf:(1)>
 2. Age=8 9 ==> School_year=2 9    <conf:(1)>
 3. School_year=2 Time=<10_min 9 ==> Lion=1.0 9
 4. Gender=M 8 ==> School_year=2 8    <conf:(1)>
 5. Age=7 8 ==> School_year=2 8    <conf:(1)>
 6. Giraffe=1.0 8 ==> School_year=2 8    <conf:(1)>
 7. Points=>3<5 8 ==> School_year=2 8    <conf:(1)>
 8. Age=8 Lion=1.0 8 ==> School_year=2 8    <conf:(1)>
 9. Age=8 Time=<10_min 8 ==> School_year=2 8    <conf:(1)>
10. Age=8 Time=<10_min 8 ==> Lion=1.0 8    <conf:(1)>
```

Fig 7: Association rules generated by Apriori algorithm

The SG applied in the present study was evaluated by ten specialists (being five pedagogues, four psychologists, and one psychopedagogue) who individually opined about the existence of game characteristics referring to their specialties. According to them, the SG used has playful characteristics, educational content, and situations that promote motivation and, consequently, help children with learning difficulties. They attested that the environment puts players in front of challenges, rules, and limits, as they are delimited in a space, having to follow rules to continue playing to reach a new stage.

Additionally, according to these specialists the game also works positively on the frustration of losing and on self-esteem, as there are encouraging phrases and when an incorrect choice occurs, the player is led to try again, thus having other chances to get it right. The implemented sound stimuli, in turn, can promote the learning of the language skills necessary for reading that were addressed in the mentioned game. In addition, it was possible to observe that the children participated with interest in the activities proposed by the game.

The results obtained suggest that Computational Intelligence can identify patterns during the interaction of players with SG. In this way, it is possible to systematically approach the strategy applied in the present study, seeking to help specialists to use this same methodology in SG developed for children with similar learning difficulties.

Although the dataset considered in the experiments is small, the results suggest relevance in the analysis and motivate new experiments on larger datasets. Therefore, the continuity of this research is promising since the proposed methodology could be easily applied in other SG to provide similar inferences.

5. CONCLUSIONS

In this work, DM techniques were applied in a SG(a virtual zoo) designed for children with learning disabilities. The results obtained led to discovering patterns in the players' profiles of this SG, allowing the identification of some characteristics that help the psychopedagogical team in classifying the children (players) into groups based on different points of view. The findings make it possible to improve and enrich this virtual zoo making it dynamic and adaptable to different player profiles. In future research will be considering the development of a knowledge-based framework applicable to a large variety of SGs.

ACKNOWLEDGMENTS

The authors would like to thank UMC and UNINOVE for the financial support. Additionally, the author S. A. de Araújo would like to thank CNPq – Brazilian National Research Council for his research scholarship (Process 313765/2019-7).

REFERENCES

[1] Cornejo, R., Martínez, F., Álvarez, V. C., Barraza, C., Cibrian, F. L., Martínez-García, A. I., &Tentori, M. “Serious games for basic learning mechanisms: reinforcing Mexican children's gross motor skills and attention”,*Personal and Ubiquitous Computing*, 25(2), 375-390, 2021.

[2] American Psychiatric Association. “Diagnostic and statistical manual of mental disorders: DSM-5 (Vol. 5)”. Washington, DC: American psychiatric association. 2013.

[3] dos Santos, I. J., Frighetto, A. M., & dos Santos, J. C. “Dyslexia: A Learning Disability”. *Nativa–Revista de Ciências Sociais do Norte de Mato Grosso*, 2(1). 2013.

[4] Avila-Pesantez, D., Delgadillo, R., & Rivera, L. A. “Proposal of a conceptual model for serious games design: A case study in children with learning disabilities”. *IEEE Access*, 7, 161017-161033.2019.

[5] Flogie, A., Aberšek, B., Aberšek, M. K., Lanyi, C. S., &Pesek, I. “Development and evaluation of intelligent serious games for children with learning difficulties: observational study”. *JMIR Serious Games*, 8(2), 2020.

[6] Palma, N. M. & Ramos, J. L. “Physical activity, obesity and active video games at school: study of gaming habits and practices in elementary and secondary school youth”. *Doctoral dissertation*, Evora University (Portugal). 2013.

[7] da Silva Lourenço, W., de Araujo Lima, S. J., & Alves de Araújo, S. “TASNOP: A tool for teaching algorithms to solve network optimization problems”. *Computer Applications in Eng. Education*, 26(1), 101-110.2018.

[8] Keller, C., Döring, A. K., & Makarova, E. “The potential of Serious Games to foster learning among children and adolescents with disabilities: A systematic review”. *Digital Culture and Education*, 13(2), 6-36. 2021.

[9] Rocha, T., & Barroso, J. “PLAY for LEARNING: Serious Games to Assist Learning of Basic Didactic Concepts: A Pilot Study”. In:*International Conference on Human-Computer Interaction* (pp. 62-71). Springer, Cham. 2021.

[10] Fayyad, U, Piatetsky-Shapiro, G., Smyth, P., &Uthurusamy, R. “Advances in knowledge discovery and data mining”. *American Association for Artificial Intelligence*. Menlo Park: MIT Press. 1996.

[11] Bhat, P., Malaganve, P., & Hegde, P. “A new framework for social media content mining and knowledge discovery”. *International Journal of Computer Applications*, 182(36), 17-20. 2019.

[12] Mitchell, T. M. “Machine Learning”. New York: McGraw-Hill. 1997.

[13] Quinlan, J. R. “Induction of decision trees”. *Machine Learning*, 1(1):81-106. 1986.

[14] de Araújo, S. A., de Barros, D. F., da Silva, E. M., & Cardoso, M. V. “Applying computational intelligence techniques to improve the decision making of business game players”. *Soft Computing*, 23(18), 8753-8763.2019.

[15] Quinlan, J. R. “C4.5: programs for machine learning”. Morgan Kaufmann Publ. Inc., San Francisco, USA. 1993.

[16] Bramer,M. “Principles of data mining”. Springer, London. 2007.

[17] Agrawal, R., & Srikant, R. “Fast algorithms for mining association rules”. In:*Proc. 20th int. conf. very large data bases*, VLDB (vol. 1215, pp. 487-499). 1994.

[18] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Logan S, Medford E, &Hughes, N.“The importance of intrinsic motivation for high and low ability readers' reading comprehension performance”. *Learning and Individual Differences* 21(1), 124-128. 2011.