

Providing Underlying Process Mining in Gamified Applications – An Intelligent Knowledge Tool for Analyzing Game Player’s Actions

Anna Tatsiopoulou*, Christos Tatsiopoulos, Basillis Boutsinas

Management Information Systems & Business Intelligence Laboratory, Dept. of Business Administration, University of Patras, 26504, Greece

ARTICLE INFO

Article history:

Received: 30 May, 2019

Accepted: 17 July, 2019

Online :30 July, 2019

Keywords:

Semantics

Knowledge mining

Extraction

Concepts

Gamification

Education

Sequential pattern mining

Process extraction

Adoptable knowledge

Management

Behaviour prediction

User-centered prediction

ABSTRACT

This work deals with the issue of understanding a user’s behaviour as this is expressed via a gamified application. The notion of ontologies and the association of concepts in relevance to decisions that have to be made is used. The current work introduces a new process-based approach, based on collected large log files and associations of underlying decisions based on them. Both of them deal with work in extracting information for intelligent use, the main difference being that the first discovers but stops on a concept relation basis, while the other based on processes, as knowledge transactions, further to the associations on a 1:1 level maybe applied on a multi associative model. The objective of the current work is to introduce the methodology into gamified environments (such as but not limited to) games, for semi-automated understanding of user behaviour and furthermore, prediction and in instances, guidance via optimal paths of decision making activities, that are useful in gamified applications in various areas like the education. Both the initial ontology based, and the extension work on it, are based on mining association rules, in one instance treated as knowledge nodes (concepts) and in the second as underlying knowledge processes, based on big log files. This may be applicable to online games, that generate big log files of user selections, that are available for study and examination for extracting user behavioural patterns. As a result, maximum length of sequential patterns and items in them, are discovered in an algorithmic based methodological approach, providing in this way a set of guidelines for designing gamified applications.

1. Introduction

The work in this paper is based on the work presented in the 9th International Conference on Information, Intelligence, Systems and Applications (IISA 2018) under the title “On Ontologies and Knowledge Associations in Gamified Environments” [1]. Its contribution is bifold (a) it extends on a more generalized manner the work so far in the above and (b) it is getting additional research by providing an additional mechanism for knowledge associations. More specifically, the case (a) above is mainly based on single concepts relations – associations and its application in mobile devices in the tourist domain, while in the part (b) contribution of additional research the notion of knowledge transaction is introduces as a set of serially associated concepts – nodes, with the presentation of its application in a gamified environment (on line game). For clarity purposes and the

distinction of concepts within the domain of gamification vs. the concepts in the game domain a brief introduction is provided below.

2. Gamification – Modeling User Behaviour

Gamification is the process of applying game techniques (e.g. dynamics, feedback, badges, leveling, etc.) to non-gaming contexts [2]. It is a new methodology that flourished just after 2008 [3] and refers to understanding a user’s behaviours on making decisions, made in a context of gaming and hence drive them to expected outcomes. As a process, gamification is under study within the scope of computer science, mathematics, artificial intelligence as well as psychology disciplines in the context of persuasive technology [4]. It is important to be noted here the difference between gamification and gaming itself [5]. The evolving processes in gamification are used today in various domains like education, training, marketing, health human

*Corresponding Author Anna Tatsiopoulou, & email: atatsio@upatras.gr

resources in enterprises, e-commerce and so on [6]. Understanding thus the above and applying them on involved processes on the above sectors, is of great importance. In the current work we will go through the gamification-based implications in the context of the designing educational content for computer-based learning and applications.

Designing and implementing gamification applications is an iterative process which consists of six main steps: 1) define learning objectives, 2) delineate target behaviours, 3) describe the players (either students and/or teachers), 4) device activity cycles, 5) implement fun and involve psychological factors that affect learning in it and finally, 6) deploy the appropriate tools. This process is performed so far by experience, by playtesting the designed gamification schemas, examine the corresponding KPIs, check what is working or not and in case that failures of the design and implementation are identified, return to the design phase and start the process all over [7].

Here, we attempt to formally introduce, as gamification is an emergent technology, the design of gamification applications, by analysing game player's actions stored in log files, to the domain of education as a process-based methodology. We try in other words, to study the user engagement into the knowledge acquisition cycle, in a domain agnostic way, based on a semi-formal approach, where, beneath it, underlies the concept of positive and negative reinforcement via the semantics, the mechanics and dynamics that the gamification-based applications are characterized of.

Education and gaming have many in common. Some of them are learning on achieving objectives, introduction and use as much as possible of engagement (interactivity in learning process) concept, motivation in achieving the previous, game elements, like positive and negative reinforcements, awarding and/or discouraging in specific cases and so on.

The above provided, our research was driven by the interest in studying what is a very important issue in the above context, i.e., what are the implications and the impact of the possible social and psychological factors that are being introduced by the gamification methodology to the achieving of the learning objectives. There are certain questions that have to be examined. For example, is the educational content being acquired and understood better if the subject is working in such an application alone, or in an antagonistic environment (one to one or one to computer), or finally in a cooperative team-based mode? Based on this, how social media and collaboration-based environment (electronic) assist this process? Should the subject "discover" the knowledge himself, in a role-based game manner, or should be driven/directed to it? If yes, in a "strict" or a more "flexible" way?

To study the above we introduce the concept of underlying processes in the gamification application that supports the learning process. We are going to examine the model of an educational application that underneath it exists a mechanism that discovers patterns in learning and at the same time tries to understand the individual's behavioural patterns and foresee future behaviour (knowledge extraction). This mechanism gets its input from a set of log files of collected raw data, that is generated from playing a respective game.

Log files in most cases, contain structured raw data that are coming and are registered as text files, coming directly from systems' interaction activities. This either comes from the user's or from other systems' components themselves activity or both them. In most cases, this raw data, comes in a structured and in a predefined format, while in other cases, due mainly to some external reasons and abnormalities, comes as incomplete data, introducing thus "noise" to the system. Out of these log files, the important aspect and the objective of the current work is to mine hidden knowledge, that can be characterized as knowledge produced during the game play (knowledge in action, while learning), in terms of underlying processes and patterns. Such concept of getting an advanced level of knowledge out of raw data, is also currently under both research and application, known as Big Data Analysis. To have such a repository of real time produced raw data in terms of log files that registers user behaviour in a game is in this step our objective and part of presenting our design methodology, in terms of underlying techniques.

More specifically, for experimentation purposes, we have selected and dealt with log files, that are being continuously generated out of the context of a game, played worldwide on the WEB 24/7/365, by thousands of players, in real time. Therefore, the generated log files are updated, using a generalized structured data model, that in many cases has the ability to expand or to reduce itself, in a dynamic way, as it will be presented in the following sections. Due to these idiomatic data type characteristics, initial and advanced parsing techniques had to be developed for two major reasons (a) to generate clear, coherent and consistent data sets and (b) to transform them, so to be persistent and hence to be stored and aggregated in a database for further exploration and processing. What we will concentrate on also here, is the discovery of interesting and meaningful patterns that signify the underlying existence of processes, towards a predictable and modifiable behaviour, a major research topic in the field of artificial intelligence [8].

3. Ontology Based Knowledge Discovery - The ONARM+ Methodology

In the previous presented work [1], we have introduced both the theoretical as well as the technical background of the ONARM+ methodology. Based on this, we have also presented in detail an application, in the domain of tourism. For this, we have considered the case of independent users that by potentially using our platform would exchange parts of their personal concepts, based on a centralized tourism [9] concept ontology, in a semi-automated and intelligent way. The scenario to be implemented based on our methodology, required 2 major components: (a) a client component named Concept-Net, that could be able to run either on mobile and/or desktop environments and (b) the Concept-Net Back End set of modules. Both of them and their subcomponents have been implemented as the following conceptual diagrams represent them (Figure 1):

Our work dealt on the concept relevance aspects with to the transformation of raw data obtained, via user selections and decision-making actions, in a specific domain (being domain agnostic), its comparisons with others on a concept basis, represented as an ontology of a specific domain.

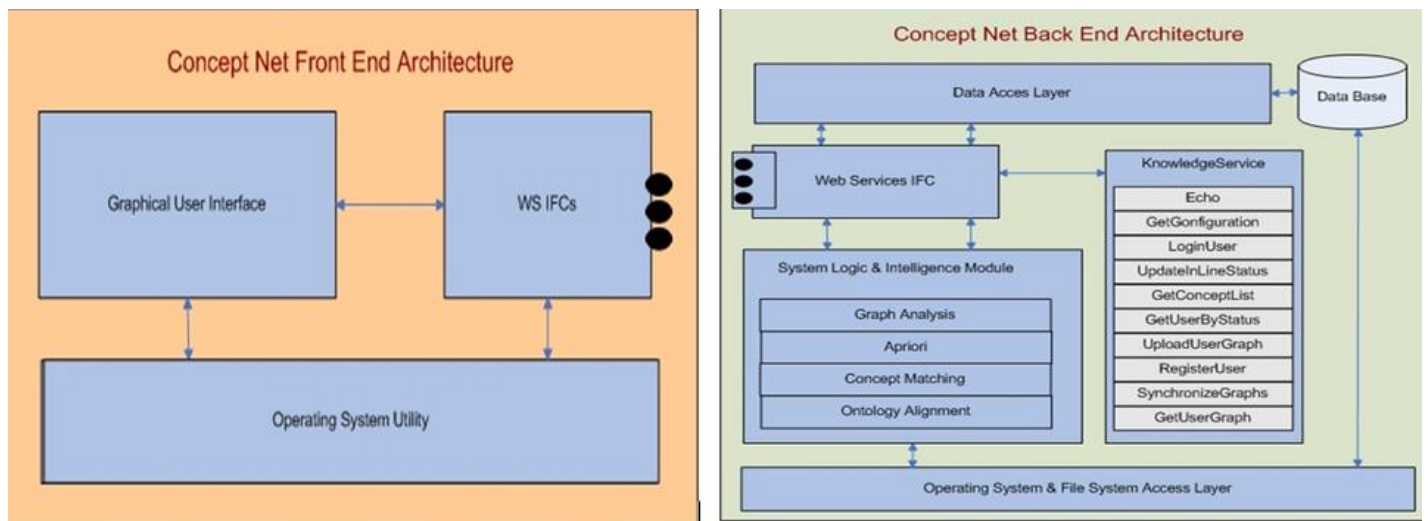


Figure 1: Left: (a) Client logical diagram, on the right (b) server-side components of the Concept-Net application

In short, during the process of knowledge extraction between the 2 users, firstly the discovery of concepts that are of common interest has to take place and then, the comparison via measures while on the final step, the merging of the parts of the specific ontology parts of each user to specific place of the ontology (aligning of them [10], [11]) takes place.

The methodology works under the assumption of existence of at least 2 user profiles, in terms of ontologies of interest. These represent the user's preferences and interests in a specific area and the categorization in a conceptual way that is depicted in a graphical way via the Protégé tool (The Protégé Ontology Editor, <http://protege.stanford.edu/>, last accessed on June 2, 2018). Also, there exists a central ontology that holds all concepts in the specific domain. As described in the work [1], the methodology is based on the Apriori algorithm [12], to identify. In general, the three major phases of the ONARM+ methodology are: (a) knowledge acquisition, (b) knowledge representation and (c) knowledge exchange.

While the association of concepts in terms of independent knowledge nodes, is of major significance in the discovery and exchanging knowledge processes, when we strive for understanding decision making in terms of a series of choices, then the above provide only partial information. To add value to the previous and essentially to introduce similarities and therefore predict more accurate user choices, we have introduced as an extension to this, the methodology below, which identifies similarities among processes for decision-making, in terms of generalized knowledge transaction processes. This is particularly useful, when not only we need to extract such information so to find similarities among users' profiles, but also equally significantly, to predict such patterns and in a second step, to be able to guide them, so that to trigger dynamic choices for desired user behaviour driving into.

4. Process based Knowledge Transactions Pattern Discovery

4.1. General

In this part, we go through our methodology and we provide an example for supporting the process of designing and development www.astesj.com

of educational applications based on gamification by using it. For this, we use knowledge extraction methodologies for extracting users' behavioural sequential patterns from large on-line game log files and propose to learners an optimal path or set of actions in a specific order, towards knowledge acquisition. Large game log files stored on line, are collected, processed, stored in an event data base and then transformed to appropriately format, for applying Apriori algorithm sequential pattern mining technique. Extracted patterns then are translated to set of proposed actions. The future path of user preferences in terms of decisions, can be (a) reinforced and (b) predicted by our methodology, while step by step educational/game objectives are achieved. Therefore, designers of educational gamification applications are provided with a set of means to select appropriate type of game mechanics/dynamics and offer selected action types in a user-centered approach tailored to any individual's requirements. Following advancements in games based on electronic media and the advanced level of new generation familiarization with such environments, in conjunction with the fact that gamification has a wide application in education domain, as it is a useful mean for maintaining student participation and enhance engagement in the educational process, we focus on frequent sequential patterns extraction from game log files and transform them to powerful tools for implementing gamified educational applications.

4.2. Background

Knowledge is the ultimate resource in every enterprise's routine operations and signifies the holy grail to their success. Knowledge in human beings is better understood but not equally explained. Recent developments and technological advancement in all fields of IT technologies (hardware and software) generate on a continuous manner, on real time basis to all kind of unrelated at a first glance data, coming from simple to very complex applications and transactions. The rising need is therefore to discover underlying processes and provide meaning to them as upper level knowledge.

On the other hand, following this trend of advancements, games based on electronic media have evolved from simple applications in the very beginning, to very complex environments, many of them simulating real life experiences and scenaria. In the

domain of education, gamification has a wide application since it has been proved to be a useful mean for maintaining student participation and enhance their engagement in the educational process. Gamification is a mean that is widely used in the marketing domain targeting to foster both brand loyalty and awareness as well as to enhance customer engagement. As mentioned previously, these three key marketing concepts are relevant in the gamification context: engagement – “high relevance of brands to consumers and the development of an emotional connection between consumers and brands”, brand loyalty – “the relationship between relative attitude and repeat patronage” and brand awareness, “the rudimentary level of brand knowledge involving, at the least, recognition of the brand name” [13]. Transforming these to learning processes in education, we have the students that are the consumers and the emotional connection to branding in our case the sources of educational context (humans, books, eBooks, web, Internet in general, etc). Therefore, the students can be seen as “highly engaged, emotionally driven players”, in this gamified learning approach, where strategic decisions have to be made by them (intermediate learning objectives’ achievements, towards the final overall knowledge acquisition, within the provided domain).

These complex gamified applications, therefore, require the track of players’ actions and the estimation – calculation of potential responses to their decisions in an intelligent way. In recent years, the tracking and logging of this information and monitoring of the way that game engines and applications perform – has become widespread in the digital entertainment industry, providing sometimes very detailed and specific information on the performance of popular commercial titles with millions of players and installations in digital devices [14],[15]. It has been widely accepted in the academic and business fields that the so-called game telemetry can be a very powerful tool not only for game development but also to assist a variety of stakeholders via user modeling, behavior analysis, matchmaking and playtesting, [16] under the condition that it is processed in the right way. As a consequence, large amounts of game telemetry data are tracked, logged and stored, but not always is being correctly analyzed and used for drawing results and decisions useful to other applications. The challenge quickly became to deal with working with large-scale data, in a process based meaningful way, for extracting behavioral decisions [17]. Quickly, new methods have emerged [15] to assist analysts and decision makers to obtain the information they need to make better decisions. These included: automatic summarization of data, the extraction of the essence of the stored information, and the discovery of patterns in raw data. When datasets became very large (any dataset that does not fit into the memory of a high-end PC as large-scale, i.e. several GB and beyond) and complicated, the breakdown method to smaller sets of data was introduced. Applying on these raw data knowledge discovery processes will help identify and uncover patterns of behavior in it, whether user-derived or learning based-derived, and these, in turn, can be highly valuable.

In this work, we use sequential pattern mining to analyse game player’s actions and behaviours, stored in log files.

The objective of the sequential pattern matching process is to discover patterns that take place in sequence(s), so to be used in prediction of future events that generate similar data. It has a great

application domain, where human behaviour analysis plays a key role in making strategic role, that varies from students’ behaviour and its impact in achieving learning objectives, customer analysis behaviour (and therefore prediction) up to security and alarm analysis, manufacturing engineering, WEB pages’ prefetching and so on. In the sequential pattern matching a sequential database of records, expressed as patterns of events is provided to be analysed. The input must be sorted in absolute lexicographical order for the algorithm. The output is a subset of patterns that frequent occur within it, provided that a minimum support (at least), is entered as an expected indicator of the least desired frequency of occurrences of the patterns to be discovered within the total records of the database.

Historically, AprioriAll is the first sequential pattern mining algorithm. The authors of AprioriAll then proposed an improved version called GSP. The AprioriAll and GSP algorithms are inspired by the Apriori algorithm for frequent itemset mining [16].

4.3. Methodology Definitions and Assumptions

There are prerequisites that log files have to support to be useful for the pattern mining extraction. A non-exhaustive list of them is the “case-id”, the “supporting activity”, “event timestamping”, “connectivity with other events, triggering or triggered by”, information on resources, such the “user name”, “system cost on event” and so on. These, so called event attributes, will allow the extraction of underlying processes, by associating such events and using a generalized model of event logs, consisting of cases, that in turn consist of events, consisting in turn of attributes. That introduces [18]:

1. A set of case(s) based process(es)
2. A set of events associated with processes via respective cases
3. An ordered set (tuple) of such events on a per case basis
4. A set of attributes per event (empty set allowed)

As a final point to be noted, is the difficulty and the rareness of availability of user data in terms of raw data (and not in aggregated form), from the game providers, that generate and own these big data files.

At this point, the current work proposes a new methodology for discovering these meaningful patterns, using minimal human intervention (in terms of pre-processing, association and conclusions’ analysis). In this section, the main steps of it are going to be presented below:

1. Data Source (Game) Selection. That was a very important step, so the proper game to be selected that conforms with the above requirements and restrictions. At the same time, the game had to offer in an open and continuous way the data being produced, in terms of log files.

2. Data Collection. Based on the fact that the previous criterion would have been met, to develop a method, that in an automatic way would be in a position to collect on a timely basis the generated data.

3. Data Maintenance. Provided the following were met, to further elaborate and develop an automatic way of synchronizing the data collection, so to avoid repeated and unnecessary steps, that

would lead in duplication of data collection. This in turn, most probably, would lead the game data provider to ban the data collection process and stop further experimentation. Unnecessary hits therefore for data collection purposes, had not only to be completely avoided, but at the same time to be carefully scheduled when to occur, avoiding Denial of Service as previously mentioned.

4. Data Transformation. This step was necessary so to introduce unique event game ids, that could be easily backwards traversed, in a meaningful way, reflecting all their attributes, that would lead to processes.

5. Data Persistence. Log files as previously explained are continuously updated in real time, as a stream of data. Out of this, chunks are not useful and have to be thrown away from mining, yet though to be identifiable in case of suspect of usefulness of it. Space and storage limitations had to be considered in the next step, being the storage in a local database.

6. Data Association. In sequence, the asynchronously generated data, has been through data cleansing and in this step, is provided in an automatic way as input to algorithms for association of it. In this step, any type of associative – rule-based algorithm could be used. It is important to be noted that since data can be easily transformed during input process to any given and expected requirement of the algorithm to be used can be applied to any type of these. For the purposes of this work the Apriori algorithm has been used [18].

7. Process Extraction via association rules. For this, the Apriori based algorithm GSP is used, with the objective to speed up the generation of interesting association rules, that will in turn lead to the mining of interesting process as proposed by [8]. The input of GSP is a sequence database and a user-specified threshold named minimum support (minsup, a value in [0,1] representing a percentage).

4.4. Technical Framework and Architecture

Based on the previous criteria, the selected game was the Alter Aeon. According to Wikipedia (https://en.wikipedia.org/wiki/Alter_Aeon), Alter Aeon was built on 1995, continuously and free provided since to thousands of people worldwide. It generates everyday thousands of lines of players’ transactions, that are publicly available as streamed log files from the game developers. More details on the game and rules of playing can be found on its web site <http://www.alteraeon.com>.

For the specific log files, in short, the process was to download the HTML files and write them in local log files. The HTML page was created by the provider to provide the information and had to be parsed until at every entry in formation of value to be found regarding transactions. Then, the files were broken in different ways, so to match the mysqlite data base schema that was adapted. After that, the sequential files were generated and fed into the Apriori algorithm for sequential process associations.

The overall procedure from scheduled data collection from the WEB Data Source, storage of unstructured log files, transformation to structured local log files, database storage, Sequential data sets generation, Apriori as described in (Figure 2)

have been implemented using Python 2.7 programming language and were tested in a distributed manner of 2 machines, connected via cloud. The first was used (a) for connection and data collection from game web site (server) of relevant data, (b) the storage of it as raw unstructured data to log files and (c) the transformation of it to local structured log files. These were implemented as a daemon set of processes, on an Ubuntu Linux server-based machine. The second part of the task was performed on a Mac Book Pro OSX with 16 GB RAM and 1 TB solid state HD.

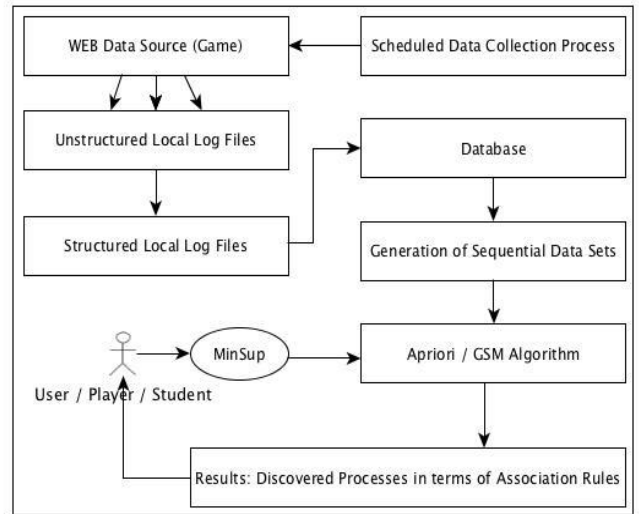


Figure 2: Overall Methodology Architecture

The separation of the overall task into two different sub tasks, in independent hardware platforms (machines), was a critical issue, since the generation of the sequential databases, as the log files were increased in size, was an intense and resources demanded issue. Additionally, the first machine has been configured to update its log files on an hourly basis (dynamically reconfigured). After this repetitive process (as many times as proper data required, here the system was collecting data for 3 consecutive months), the Apriori algorithm was applied, implemented in a Java application environment.

4.5. Results

4.5.1. Initial Results Presentation

In the diagram below (Figure 3), is shortly depicted the process for running the cleared and synthesized log files through the GSP Algorithm.

On the left-hand side of the above image, is a snapshot of the unstructured game log files, collected for a day transactional activity, from the game Web page and on the right is presented a sample of GSP process discovery outputs.

It is worth mentioned that such files were also aggregated to monthly basis and can be further aggregated to any time interval, depending on what time would like to be examined for process extraction. Initially the date / time has been registered and then event type has been recorded. Finally, on the last column, the detailed event description can be found, in a syntactically dynamic format, according the circumstances that the specific event took place.

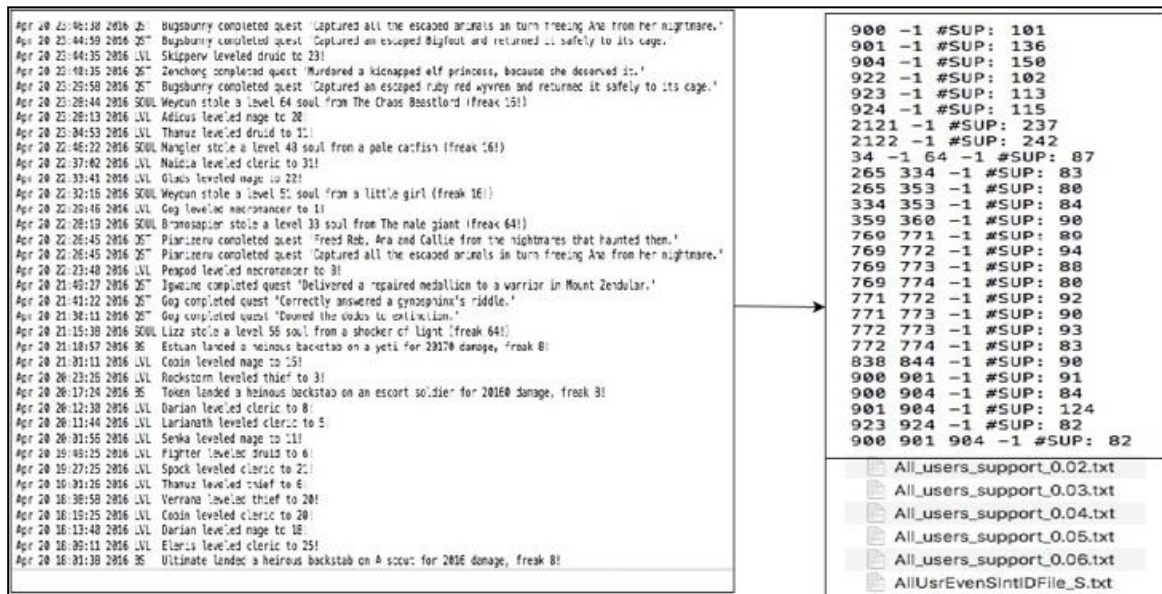


Figure 3: Initial Results Discovered by GSP algorithm.

On the right-hand side, there is (a) a sample of the process discovery after the GSP application; (b) on the lower right-hand side, the format of the filenames of the produced log files are presented, that contain the interesting associations, for processes found. It can be noted at the end of each filename, the respective minsup measure, as previously introduced, under which the process associations took place. (c) finally, on the right top, a sample of the output is presented. The (-1) stands as a separator of the frequently associated process id, that in previous steps has been uniquely composed for a player and player's activity in general to be able to be reversed tracked. As it can be seen in the sample output, there are interesting associations of frequency of 1-member, 2-member and 3-member activities. At the end, the number represents the measure of interestingness, as a measure of frequency of associated processes.

The application of the GSP Apriori implementation algorithm, provides all possible associations that could be found [12], by applying the specific minsup measure. Any sequential pattern mining algorithm could be used instead, as the VMSP [19], that works in a similar way taking the same data format as input and producing a similar file format as output, being different on the fact that provides the maximal model for the associations discovered by the sequential data files.

4.5.2. Processing of Initial Results – Extracting Behavioural Process Patterns

Sequential pattern mining is a research field studied for more than two decades, focusing now in two main areas a) Applications in the aspect of utilizing sequential pattern mining either in new applications or in new ways for existing applications and b) Algorithms either for designing more efficient algorithms or designing algorithm for handling more complex data or for finding more complex and meaningful patterns [20].

The present work contributes to the first area providing results that can be used in the design and development of new gamified

applications and in the improvement of existing ones in any domain. The work performed during the experimental phase lead to the following main results:

1. The maximum length of sequential patterns varies inversely with the support value used. The finding was in line with the results stated by R. Agrawal and R. Srikant [8]
2. Sequential patterns discovered only for the events of the type “QUEST”. This can be seen as the “repeated” choice of users at different “levels” to perform tasks in the game play in order to achieve certain objectives from simple to more complicated ones.
3. Sequential patterns mined were of maximum length three events for 0.02 support, of two events for support 0.03 when for support 0.04, 0.05 and 0.06 only single items were discovered, which lead to the conclusion that users for various reasons either internal (interest, resources etc.) or external (time availability etc.) are more familiar with this type of pattern.
4. Item members of two or three component sequential patterns were not present as single items in the higher support level sequential patterns mined. This fact may be interpreted and used to the design phase of the potential application(s), as a valuable hint, that helps designers combine “QUEST” elements, being at present not very popular, but of high interest, especially when associated with interaction with other facts / achievements, thus providing added value to them.
5. Single items discovered for support 0.04, 0.05 and 0.06, where not part of the maximum sequential patterns discovered with either two or three event sequence lengths, which can be seen as a cross checking point for implementing gamification strategy, where the importance is the relevance of the “QUEST” with the desired task to be performed.

The above findings are of great significance for designing gamification-based applications for behavioural prediction of users. This is because “QUEST”s can be seen as user behavioural

transactions that include events of high importance, for achieving the maximum objectives set in such applications, in any type of environments supporting well defined processes like, e.g. market(s), sales, e-learning regardless the type of gamification type applied (intrinsic or extrinsic) and so on. The current methodology overcomes the obstacle of the manual observation, association, extraction and in the end generalization, in a (semi-) manual way, based on intuitive approaches. On the contrary, the current work provides an automated algorithmic based methodological approach, for provision of a set of guidelines, to be used in the design process during the application development phase, which so far is based only on empirical findings. In the specific application selected in the current game, this is focused on using “QUEST”s, in the content of gamified applications, with the objective to drive the game user to maximal results, [6] in terms of “QUEST” usage, in a way that is familiar and transparent to him/her. Therefore, based on the fact that the maximum sequence length is 3, provides designers of gamification strategies, for the specific domain, means to select each time the appropriate “QUESTS” for achieving specific goals and drive via behavioural changes in terms of decision making in presented challenges, to the desired performance for selecting either desired actions, or products or services or education-based objectives available in a respective gamified application. On the other hand, the findings that single items gain added value when associated among other, is of great importance and can be used as guide for decision making vis-a-vis to the way that “QUESTS” are presented to the users of these specific gamified application(s).

Getting back to our case in the game selected for experimentation, as mentioned, a sample of 36.000 approximately user interactions have been collected into a respective SQLite database, categorized under 10 relevant tables of specific events types supported by the game. This data was collected on a periodic per hour and per day basis and then aggregated to a monthly based per user model. Running then the specific Apriori sequence algorithm, our methodology provided as an outcome, the conclusion that there was an interesting path of predefined processes expressed in terms of game terminology as “QUEST”s that was followed by the user as a sequence towards the game objectives. These (as all game events), have been coded as follows (processID, questID, questName):

900: q678 -> “Searched the fetid Caves of Pestilence and destroyed the weather machine”

901: q679 -> “Ruined the plans of Malkon the Plague Vampire and saved Archais from his evil machinations”

904: q682 -> “Breached the Fetid Caves of Pestilence to rescue a dwarven child”

It is important to be mentioned here, that the associations of the type (900 -> 901 -> 904) were discovered under the support measure of 0,02 and with frequency of 82 times among all possible processes interactions discovered. The number of the three (3) associated processes, is directly relevant to the sample data collected from the large log files produced by the gamified applications and they are also language/term agnostic. Finally, in the tuples discovered, the order is of significance, providing the sequence of these processes.

What this means, is that there is a significance in users’ interestingness, when they use the gamified application, in getting, in a sequential way, from process 900 to 901 and to 904 by achieving their respective objectives. In other words, from the point of view of the behavioural analysis, the future path of potential user preferences getting from a point A to a point B, in terms of decisions, that can be achieved through multiple ways, can be (a) reinforced and (b) predicted by our methodology, while step by step game objectives are achieved.

This is a very important analysis and conclusion for applications that deal and are composed by decision making, based on a process-oriented approach.

5. Comparisons and Advancements – Generalizations & Potentials

The above findings, if applied in the education domain, will be of great significance and will provide initial guidance for developing gamified applications for the educational industry (eProducts, eServices and so on). Such applications need to be compliant and respect the overall model of “Input, Process and Output” (Heis, 2008) in the education domain, where input are the educational objectives that by the appropriate learning process are transformed to learning outputs that in turn are evaluated against the initial objectives (Figure 4).

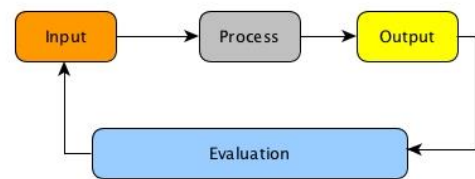


Figure 4: Input Process Output Model (Heis 2008, [220])

As a consequence, for drawing and implementing interesting, effective and challenging input for an educational application, being aware of the patterns that users are familiar with, through the “Evaluation” of their behaviour in such applications, would allow to provide to them more tailored to their needs “Input” for the educational “Process” that will allow the maximum “Output”.

In terms of gamification resources, our methodology will enrich the decision process for creating and applying more personalised experience to end users and suitable quests, challenges, rewards etc. for being also compliant with the fact that designing user experience for educational applications that incorporate gamification strategies needs to facilitate the “ABCD” model educators use for writing and using learning objectives, whereas “A” means audience, “B” behaviour, “C” conditions under which the objectives have to be completed and “D” is the degree that learners have to achieve [21].

It is known that the type of proper motivation varies depending on the audience characteristics therefore for succeeding in adopting the above-mentioned model, given that learners are willing to participate in educational processes if they are subjects to appropriate education methods and resources, informed on the foreseen behavioural change and triggered enough for completing the set of achievements they are obliged to.

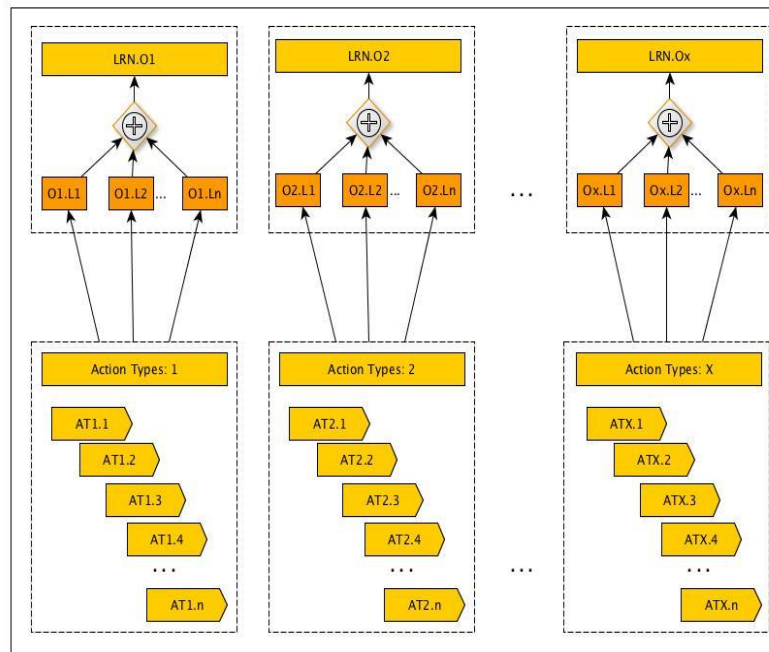


Figure 5: Design of an Educational Gamified Application

Even in Fogg's Behavior Model [22] is clearly stated that providing sufficient motivation, adequate ability and efficient activation to learners, results to the desired from the educator point of view "behaviour" achievement. As gamification contributes towards the provision of both motivation and activation of learners, it is important to employ suitable type and "portion" of them for resulting to successful applications and learning outcomes.

We select to present the potential application of our methodology to the education domain. For this purpose, we consider the education related processes, such that to be related in an input-process-output schema of relations, composing thus a system-oriented view point to it (education).

Our methodology will allow gamification applications in the above-mentioned education model, to use the "right" pattern each time, for supporting the desired behaviour for the learners, overcoming the theoretical obstacle that "the use of challenges in Information Systems though, may be limited because of a high number of monotone and standard tasks" [23]. Such strategy, hence, based on optimal path reinforcement of educational processes, will lead both teachers and students, to be informed and getting up to their educational objectives with a high degree of performance, understanding and productivity in the area.

Based on the above findings of our methodology, we attempt to design a gamified application. We consider again the educational domain for the target group of the application, that is independent of the subject that has to be introduced – taught.

The main course objectives are for the paradigm below (Figure 5), introduced as LRN.O1, LRN.O2, ..., LRNOX. It shows how action types will be connected to Learning Objectives for the design of an Educational Gamified Application. These are the necessary objectives to be achieved so that the course to be successful for the learner (or the learner group).

Every such Learning Objective Ox, in turn, can be achieved in the gamified application to be designed, by the achievement of respective objective game levels (Ox.Ln, meaning for the objective LRN.Ox, the Level n and so on). We assume that the number of levels per learning objective may vary, (but at least may have a value of 2).

On the other side, in our game design methodology we offer, the ability to the developer but at the same time to the player – learner, to design a variety of action types for the achievement of such levels, as described above and play freely by navigating and making actions – decisions through the game scenario. Such action types that are introduced by the game designer and implemented by the game developer, may be represented as a virtual library of actions, containing – being categorized further in sub-action types. For example, somebody may select to achieve the respective level, to read material, then to play a local educational game and then take a quiz to collect points towards next level. On the other hand, somebody else, may select to study much more material in sequence, so that immediately as soon as will take the quiz, to be sure that will also be awarded the next level. This model introduces the independent process in sequence selection towards one, the same objective achievement, allowing edutainment weight to be balanced by the user, on real time basis, i.e. how much game, how much, in app study, out app (text – notes) study, cooperative action types, info from social networks and so on will on demand use.

Collecting this information in large log files, as our application will run, will in turn produce frequent patterns of selected actions, where players achieved the objective much faster, as a function of selected action types. Our methodology then, when enough knowledge and processes will be collected, will propose to next player i, a proposed optimal path towards knowledge acquisition of the respective level, i.e., a set of Action Types, in a specific order, so the player – learner will be more productive. Of course,

this will be only a suggestion and the player will be always free to reject it and hence select his own path, towards the level achievement.

6. Conclusions and Future Work

As an extension to the ONARM+ previously presented work, we developed a formal methodology to support the design of gamified applications. It is based on analyzing game player's actions stored in log files to extract frequent patterns of actions. The extracted patterns are processed and classified according to the expected outcomes which can be supported in a gamification application.

This knowledge base for gamification of frequent sequential patterns of actions, along with the outcomes, they are considered as knowledge process-based transactions. Thus, given a specific domain and the expected outcomes for a gamification application, the designer could build it by selecting the proper patterns of the knowledge base above, so to drive via optimal decision making paths user behaviour changing, which is the basis for a variety of diverse domains, like marketing, learning, consumer/branding loyalty and so on. Apart from patterns, the knowledge base above. Based on this, we plan to extend our work so that the knowledge base above, to contain also meta-rules for controlling game techniques (e.g. dynamics, feedback, badges, levelling, etc.) and their relative association on a per user profile basis.

Finally, we have presented an educational game design concept, based on our proposed methodology, where the proposed application, introduces to new players – learners, accumulated optimal knowledge experience, towards game level achievements and their knowledge – learning objectives. In this game we tried to examine, using our methodology associations in user decisions and choices, as a set of concept-based processes, so based on an algorithmic approach to predict and drive choices of similar user profiles, on the same domain. While availability by developers of such real time on line games have been proven a task difficult to be achieved, future work will be focused in collecting more than an additional year's supplementary data to verify our findings vis-à-vis user behaviour.

Overall, initial interesting results of such associations were found, for specific players. Further work will concentrate also in five main directions (a) application of the maximal VMSP mining algorithm and comparison of results generate; (b) generalization of results found to a broader set of players; (c) application of the proposed methodology to other gamified domains, for mining underlying processes; (d) analysis of more other computer games (types) to extract more patterns and (e) the generalization of our work for groups of users instead of focusing to specific ones.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] C. Tatsiopoulos, D.M. Katsanta, "On ontologies and knowledge associations in gamified Environments", 2018, 9th International Conference on Information, Intelligence, Systems and Applications (IISA), July 23-25, 2018, IEEE, Zakynthos, Greece, DOI: <https://doi.org/10.1109/IISA.2018.8633589>, e-ISBN: 978-1-5386-8161-9

[2] S. Deterding, Dixon, D. R. Khaled, L. Nacke, (2011). From game design elements to gamefulness. In Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11. <http://doi.org/10.1145/2181037.2181040>

[3] Z. Epstein, (2012). Enterprise Gamification for Employee Engagement. Marioherger.At.

[4] H. Oinas-Kukkonen, M. Harjumaa, (2009). Persuasive systems design: Key issues, process model, and system features. Communications of the Association for Information Systems.

[5] T. Xu, F. Feng, D. Buhalis, J. Weber, H. Zhang, (2015). Tourists as Mobile Gamers: Gamification for Tourism Marketing. Journal of Travel and Tourism Marketing, 33(8), 1124–1142. <http://doi.org/10.1080/10548408.2015.1093999>

[6] E. B. Passos, D. B. Medeiros, P. A. S. Neto, E. W. G. Clua, (2011). Turning real-world software development into a game. In Brazilian Symposium on Games and Digital Entertainment, SBGAMES. <http://doi.org/10.1109/SBGAMES.2011.32>

[7] K. Werbach, D. Hunter, (2012). The Gamification Toolkit: Game Elements. For the Win: How Game Thinking Can Revolutionize Your Business. <http://doi.org/10.1017/CBO9781107415324.004>

[8] R., Srikant, R. Agrawal, (1996a). Mining sequential patterns: Generalizations and performance improvements. Advances in Database Technology — EDBT '96. <http://doi.org/10.1109/ICDE.1995.380415>

[9] C. Tatsiopoulos, B. Boutsinas, "Automatic knowledge exchanging between tourists via mobile devices," in Journal of Hospitality and Tourism Technology, vol. 1, No2, pp.163-173, 2010.

[10] C. Tatsiopoulos and B. Boutsinas, "Ontology mapping based on association rule mining", in Proceedings of 11th International Conference on Enterprise Information Systems, vol.3, pp.33, Milan, Italy, 2009.

[11] C. Tatsiopoulos, B. Boutsinas, K. Sidiropoulos, "On aligning interesting parts of ontologies", in Proceedings of International Joint Conference on Knowledge Engineering and Ontology Development (KEOD2009), pp. 363-366, 2009, Madeira, Portugal.

[12] R. Srikant, R. Agrawal, (1996b). Mining Sequential Patterns: Generalizations and Performance Improvements. In Proc. 5th Int. Conf. Extending Database Technology, EDBT (Vol. 1057, pp. 3–17). <http://doi.org/10.1109/ICDE.1995.380415>

[13] A. K. B. Bharathi Gopinath, A. Singh, C. S. Tucker, H. B. Nembhard, (2016). Knowledge discovery of game design features by mining user-generated feedback. Computers in Human Behavior. <http://doi.org/10.1016/j.chb.2016.02.076>

[14] L. Mellon, (2009). Applying Metrics Driven Development To Mmo Costs and Risks. Versant Corporation.

[15] A. Drachen, A. Canossa, (2011). Evaluating motion: spatial user behaviour in virtual environments. International Journal of Arts and Technology. <http://doi.org/10.1504/IJART.2011.041483>

[16] D. Kennerly, (2003). Better Game Design Through Data Mining. In Gamasutra.

[17] A. Drachen, M. Seif El-Nasr, A. Canossa, (2013). Game Analytics – The Basics. In Game Analytics. http://doi.org/10.1007/978-1-4471-4769-5_2

[18] W.M.P. Van der Aalst, (2011). Process Mining. Process Mining. <http://doi.org/10.1007/978-3-642-19345-3>

[19] P. Fournier-Viger, C. W. Wu, A. Gomariz, V. S. Tseng, (2014). VMSP: Efficient vertical mining of maximal sequential patterns. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). http://doi.org/10.1007/978-3-319-06483-3_8

[20] P. Fournier-Viger, J. Chun, W. Lin, R. U. Kiran, Y. S. Koh, R. Thomas, (2017). A Survey of Sequential Pattern Mining. Ubiquitous International.

[21] A. Salam, (2015). Input, process and output: System approach in education to assure the quality and excellence in performance. Bangladesh Journal of Medical Science, 14(1), 1–2. <http://doi.org/10.3329/bjms.v14i1.21553>

[22] B. Fogg, (2009). A behavior model for persuasive design. In Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09. <http://doi.org/10.1145/1541948.1541999>

[23] S. Thiebes, S. Lins, D. Basten, (2014). Gamifying information systems A synthesis of gamification mechanics and dynamics. Twenty Second European Conference on Information Systems, (August 2016), 1–17. Retrieved from <http://aisel.aisnet.org/ecis2014/proceedings/track01/4/>

[24] U. K. Heis, (2008). Input Process Output, 1–3. http://doi.org/10.1007/1-4020-0611-X_462