Educational Game Analysis Using Intention and Process Mining

Konstantin Nikitin[a]

[a] National Research University Higher School of Economics, 20 Myasnitskaya Ulitsa, Moscow, 101000, Russia

Abstract
Gamification is a common practice that however leads to more complex and difficult processes for modeling. In this work usage of Process Mining for educational game analysis is studied. For such purposes goal-oriented process mining seems to be efficient and intention mining in particular. While Map Miner Method seems to be relevant to indicate intentions within the game, it's not precise enough to analyze player behavior in the context of game design. Usage of classical process mining for modeling discovered strategies forms hierarchical model of gaming and education process with different levels of abstraction. Initialization of model properties was described and some assumptions about resulting two-perspective formalism features were made. Proposed method was tested on VR-Chemical Lab project and preliminary potential of this approach and directions for further studies were indicated.

Keywords 1
Intention Mining, Map Miner Method, Educational Process Mining, Spaghetti-like Processes.

1. Introduction

Gamification of education is a common practice nowadays [1]. However, development and analysis of such unstructured systems is a complex task that involves more than nine different disciplines [2]. In the fast-growing industry full of data it seems to be important to visualize and model processes of user-product interaction. So, methods for modeling player behavior in games are in the focus of present work.

To observe player behavior many data mining tools are used. Unfortunately such methods are usually "black boxes" and for analysis of behavior reasoning we need more detailed modeling. While educational gaming is definitely a process we suggest to use process mining techniques for such analysis. However, educational games have some features we should mention before modeling. First of all there are many different strategies of playing that, however, lead to the predefined results. Especially in so called "sand-box" projects, where player's actions do not have strict order. So, gaming process becomes "spaghetti-like" and difficult to interpret. Another problem is that we need to detect user intentions during the session, gaming or studying, to filter some cases. Educational games combine study and enjoyment, so we should analyze the reasoning of actions to balance these parts. Also, it's important to separate different reasons of deviations to exclude some. While inappropriate usage is out of scope of our analysis, player can be solving the task but being distracted for some time and this case we should consider.

In this paper exploration of process mining and particularly intention mining applicability for modeling an educational game process is done. We consider use goal-oriented process mining, but

discovered intention models seemed to be too abstract because of different perspective, so used Map Miner Method was modified and combined with activity-oriented modeling.

2. Related Work

Player Experience Modeling (PEM) is defined as the area responsible for collecting data about players and their behavior within the game [3]. It consists of three general groups of methods: subjective (based on player impressions), objective (based on physiology parameters) and gameplay-based (based on information from game objects).

While player-game interaction is definitely a process we suggest using process mining (PM) to model gaming behavior. It can be considered as gameplay-based player modeling. Process mining in games is not very common yet, however some researches are made. In [4] PM is used to differentiate players by their skill level in programming game. However, results show that process mining is not applicable for such task. In another research fuzzy and heuristic process mining were used to analyze player behavior in first-person shooter [5]. This research showed the promising results of using process mining. However mentioned study is based on process with little number of actions and discovers simple one-level models. As was mentioned earlier our goal is to model unstructured educational game with large number of possible actions and their sequence.

While we are focused on educational games, it is worth to review an area of educational process mining (EPM). Bogarin et al. [6] made the general survey of using EPM. Their paper provides complete information about frameworks, concepts, data, tools, techniques and application domains of process mining for education. Moreover, it states the great potential of intentional mining research that has not been discussed yet in articles. We consider this area as important for serious game studies and related to our topic.

 For better understanding of student's learning habits in [7] improvements for EPM were proposed. Researchers apply two-step clusterisation on the model and detect the most perfomant learning paths. PM is close to another area, data mining. In [8] objectives of educational data mining by various stakeholders are described. Moreover authors define reasons of detecting certain problems and data mining techniques to reveal them.

The most related study to ours research propose the User Behavior Pattern Detection methodology that is based on detecting patterns in user behavior in gamified course with some level of freedom in system usage [1]. In this paper there was shown that goal-oriented process mining is more suitable for complex and non-structured gamified systems. This is still a modest area, but it can improve interpretability of unstructured models for stakeholders [9]. Two literature reviews we made in this area. One observes goal-oriented PM in general and divides models into three categories: goal modeling and requirements elicitation, intention mining and performance indicators [9]. The second concentrates directly on intention mining [10].

Intention mining is an emerging field focused on detecting reasoning behind the actions [11]. The most significant work on this field is doctoral thesis by Khodabandelou [12] where Map Miner Method is described for mining intentional process model. According to mentioned literature review this technique has not been developed further yet.

3. Overview of the Proposed Method

As was mentioned earlier, classic activity-oriented process mining techniques tends to discover very complex "spaghetti-like" models (Figure 1). So for modeling player behavior in educational game goal-oriented approach was chosen, more particularly intention mining.
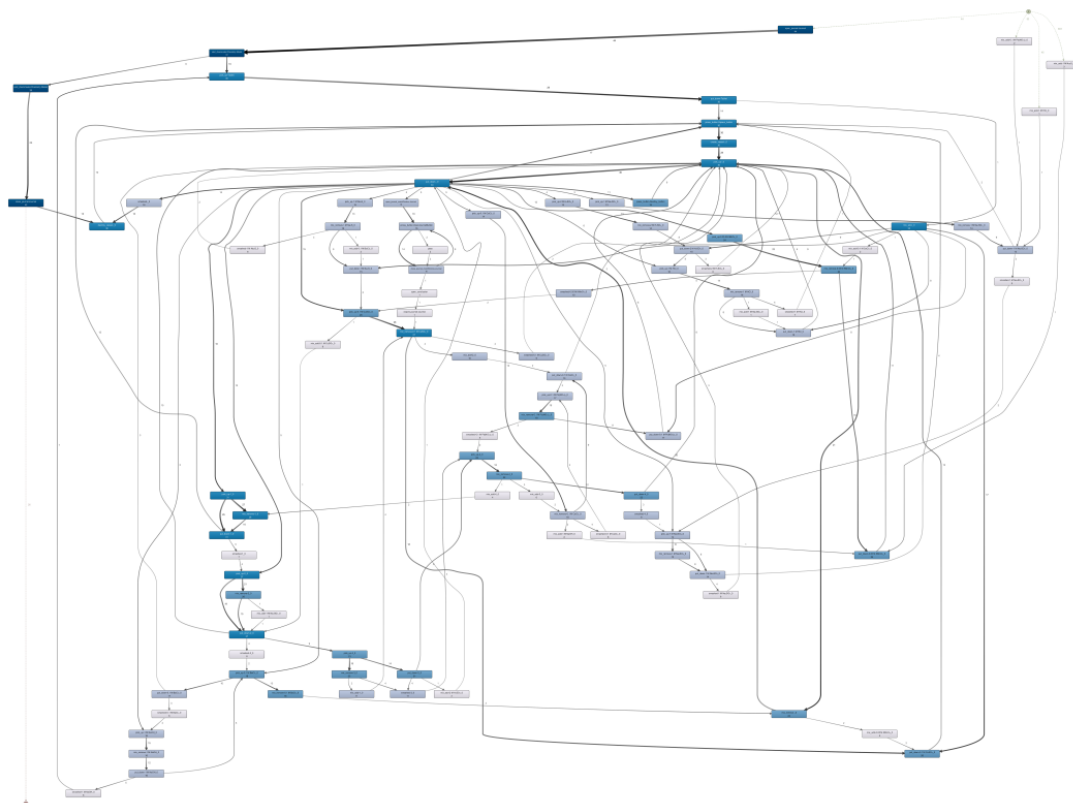
Figure 1: Process map based on VR-Chemistry Lab logs made with Disco tool with 100% of activities and 0% (only most important) of paths.

For our research MAP formalism was used. Map metaphor was firstly introduced in [13] and then developed for Disco tool. Let's consider map as a directed graph, usually labeled. Map metaphor in Disco is used to represent actions as nodes and transitions between them with transition frequencies as edge labels. In intention mining nodes represent some users' intention while edges are different strategies to achieve new intention after another. In both cases graphs have start and end nodes.

In the current research we use intention mining for process modeling. The most significant intention mining method is Map Miner Method. There are two main components of the model. Intention is defined as an objective to achieve a goal with clear criteria, which users have in mind that can be fulfilled [12]. A Strategy is a trace of actions needed to achieve an intention [9]. One intention can be achieved by many strategies [12].

The Map Miner Method is based on Hidden Markov Models (HMM) and analyzes traces of groups of users. To discover map two algorithms are being used. After estimating HMM parameters using supervised or unsupervised learning, the first algorithm, Deep Miner, discovers sub-intentions from logs, using metrics of fitness and precision. After that Map Miner algorithm groups sub-intentions into intentions using K-means [12].

Our study is focused on sub-intention level of Map Miner Method. Sub-intentions are low-level intentions mined directly as results of fulfilling strategies defined by HMM. To construct map in this method to every strategy corresponding sub-intention is assigned. Sub-intention in this case is an abstraction needed only to represent nodes of the graph. While every sub-intention has only one corresponding strategy that leads to it from other nodes, we can shift the map to use more formal representation closer to HMM used to mine strategies while keeping the same graph topology. We consider intention map as a directed graph with Strategies as nodes and transitions between them as edges. Every transition takes place with certain probability after completing corresponding strategy Example of such map is presented in Figure 2 (b). Formal semantics of map in our work is described as following:

Let's assume $P \in R, P \in [0; 1]$. A Strategy map (or Intention map) is $(A, S, \delta, \rho, \pi)$, where

$A = \{a_1, a_2, \dots\}$ – Set of all possible actions (observations of HMM).

$S = \{s_1, s_2, \dots\}$ – Set of all possible strategies (hidden states of HMM).

$\delta: S \times S \to P$ – Transition probabilities.

$\rho: S \times A \to P$ – Emission probabilities.

$\pi = (\pi_1, \pi_2, \dots), \pi_i \in p$ – Probabilities of initial states.

Strategies are connected with each other according to the transition matrix obtained during HMM creation. For each strategy there is a corresponding set of actions according to emission probabilities. In the current paper strategies and states are used as synonyms.

We used unsupervised learning and Baum-Welch algorithm to train HMM because no predefined strategies are available.

While Map Miner Method discovers very readable map of player intentions it is seems to be too abstract for detailed game-design analysis. First of all, we can not analyze strategies themselves. While strategies represent set of actions and actions thus are unordered, it can differ from real system. Moreover we don't mine any supporting information about time spent, for example. Having probabilities of transitions between states we still can not easily indicate reasons of moving from one intention to another. There is no formal definition of completing the strategy. That's why we suggest use intentional mining for modeling player behavior with conjunction with classical process mining techniques to model player enactment within particular strategy (**Figure 2**).

This method will provide a hierarchical model with different levels of precision and different perspectives. On the top level process is considered as number of high-level intentions, that are consist from different sub-intentions fulfilled with strategies that in their turn analyzed from an activity perspective using Fuzzy miner.
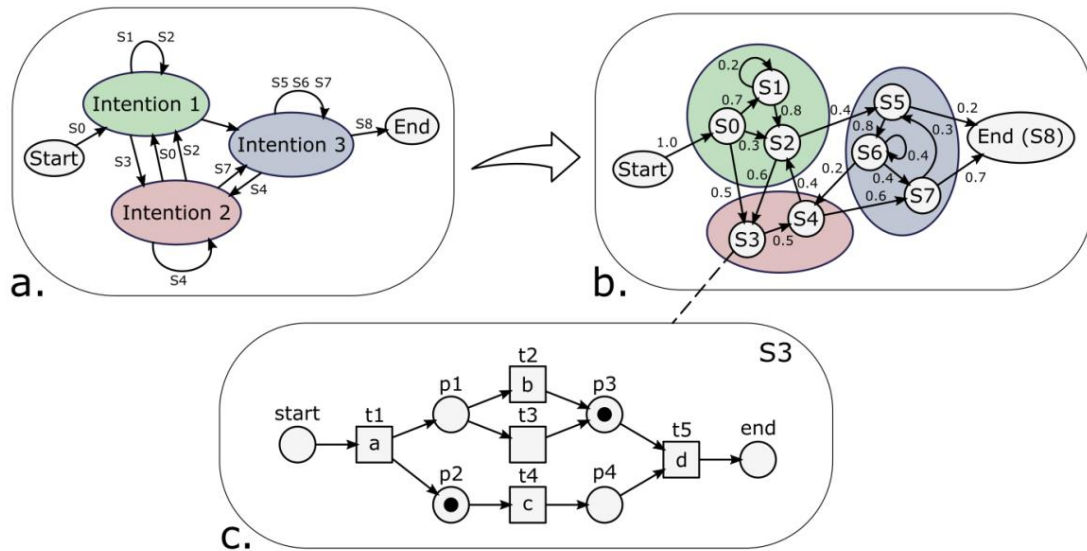

Figure 2: Abstract example of proposed hierarchical model. Intention map with high-level abstract intentions and strategies needed to move from one intention to another (a); Strategy level with probabilities of transitions between nodes (b); Action level for one particular strategy as Petri Net (c).

4. Intention Mining Model Discovery

To start intention mining we should initialize some parameters for Hidden Markov Model. Such attributes are: number of hidden states, initial probabilities for transition and emission matrices, vector π of initial state probabilities. Transition matrix defines probabilities of one state (or in our case strategy) to follow another. Emission matrix states what actions with what probability are included in some strategy. Vector π needed to identify initial state of the model. Usually such parameters are set randomly or with some assumption based on the data nature.

4.1. Estimating number of strategies

According to thesis to determine number of states it's possible to create number of models and find the threshold effect [12]. This effect states that after some point when strategy number increase, actually mined different strategies number remains the same. So some strategies are duplicated. In our case no such threshold effect was observed. On the figure dependence between number of strategies to mine and different strategies mined as long with completely different, when no strategy is a part of some another, is shown (Figure 3 (a)). Strategy is considered as completely different if it has no duplicates and is not a part of another strategy in terms of included actions.

To compare different models with different number of strategies information criteria were used. In this study we tried three criteria: AIC, BIC and ICL. Unfortunately in HMM criteria tend to prioritize models with larger number of states [14] (Figure 3 (b)). Moreover creating large number of models with high number of strategies is a time-consuming operation. Nevertheless information criteria could indicate the best model in some range. So this method was combined with heuristic assumption. During modeling we visualized event occurrences for each strategy. While it does not help to analyze what model is better between similar ones, it provides visual information for analyst when strategies become too sparse (many strategies with little number of events in them). The compromise is restriction of maximum number of hidden states according to heuristic method and usage of information criteria to find local minimum in a limited range.

Additionally it is worth to mention that no advantage of one information criterion over another was noticed.
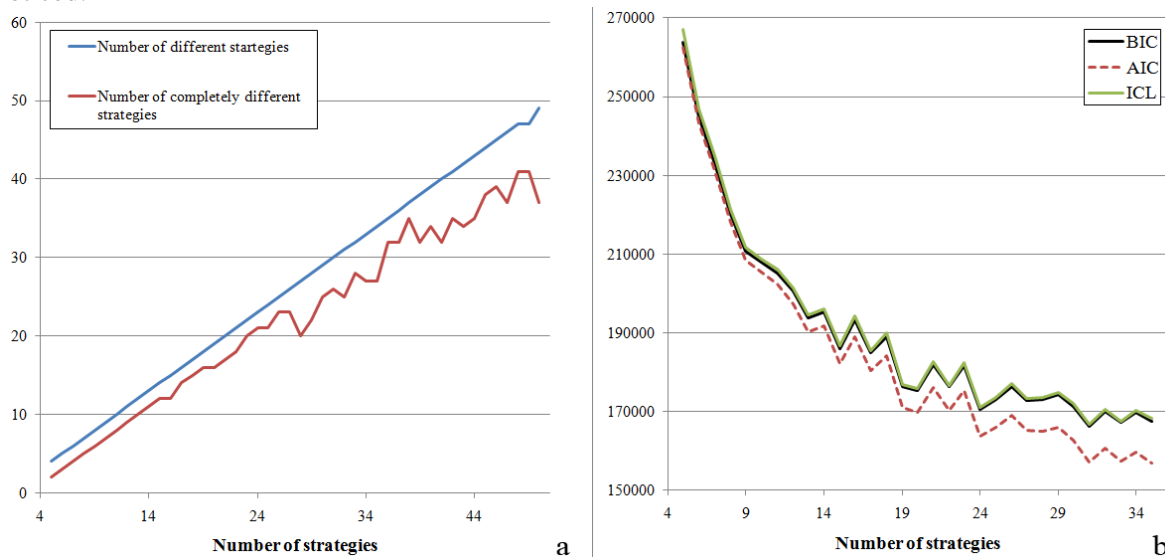

Figure 3: Number of different and completely different states mined depending on defined number of states (a); information criteria values for models with different number of hidden states (b).

4.2. Initialization of Hidden Markov Model

According to thesis end strategy is strategy with no outgoing transmissions [12]. Tests on our data showed that we can both have many such strategies (especially after transition elimination due to accuracy setting) and no such strategies. But on the map there should be only one start and end. To solve this problem we insert into event logs an artificial event "End Event" as last for each case. After HMM training all states that are can be treated as end strategies should be connected with strategy containing this inserted event (and only it).

Because of this insertion we modify initial transition end emission matrices. Last state is assumed as end state. Therefore its transition probabilities to other strategies are considered as 0, while transition to itself is 1. Corresponding row of emission matrix states that this strategy has one and only one event − "End Event". Finally, in π there is 0 on probability that this state will be an initial in our model.

To indicate the starting strategy no such insertion is required. Start is defined by values of vector π. If several strategies have nonzero probability to be the initial state, Start state is injects onto the map with outgoing transitions to all such strategies with corresponding to π transition probabilities to each.

5. Implementing Activity-Oriented Process Mining

As was mentioned earlier we suggest using hierarchical model with increasing level of precision from high-level intentions to strategy workflows. To do so we extend intention mining by using activity-oriented process mining technique. After creating the intention map we analyze strategies individually. According to emission matrix there is a set of actions each of them is in the strategy. Fuzzy miner and map formalism are used to model process based on the same logs including only chosen actions. Such model shows the actual process of fulfilling specific sub-intention with current strategy.

To show some properties of the discovered model we need to introduce some definitions:

We say that action a_k is in strategy s_m when $\rho(s_m, a_k) > 0$.

We say that two strategies s_k and s_m are connected when $\delta(s_k, s_m) > 0$.

We say that two actions a_k and a_m are connected when there is transition between corresponding events in the discovered by Fuzzy miner map.

Action a_k follows a_m when they are connected and direction of the corresponding graph edge goes from a_k to a_m. Opposite direction states that action a_k is followed by a_m.

Because of using HMM for mining strategies we state that discovered map has following property:

Two actions a_k and a_m are connected only if one of the conditions below is true:

1) a_k is in some strategy s_t and a_m is in the same strategy s_t;

2) a_k is in s_t, a_m is in s_r, s_t and s_r are connected;

Using this property we can expand model of strategy. Start and end of the activity map could be replaced by corresponding labels of previous end next strategy. Let's define all actions start event is connected to as start actions of current one. All actions connected to the end are considered as end actions of the strategy. If any starting activity a_k follows some other activity a_m on the map, then we can say that a_k follows s_m if a_m is in s_m. This statement will be reflected on the map by additional node for s_m. In the same way end actions of the map are connected with strategies containing actions that follows end ones.

Aforementioned property leads to the assumption that complete activity-oriented process model can be discovered as union of strategy models. So intention mining is a potential tool not only to construct goal-oriented maps but also for activity grouping. However, groups have some intersecting activities, but emission matrix shows the frequency of each such action appearance for each group.

These properties are not proved formally yet, however all experimental tests made so far confirms current assumption. On Figure 4 very simple example is shown. For sandbox Disco project, used in this program as default example, intention map was made. Than for each strategy corresponding activity map was discovered. Complete map was constructed independently in Disco just as usual activity-oriented model. Comparison of two resulting maps shows both validity of the union of specific strategies and accordance of connections between activities and strategies. So, mined strategies could be considered as logical groups of activities of overall process.
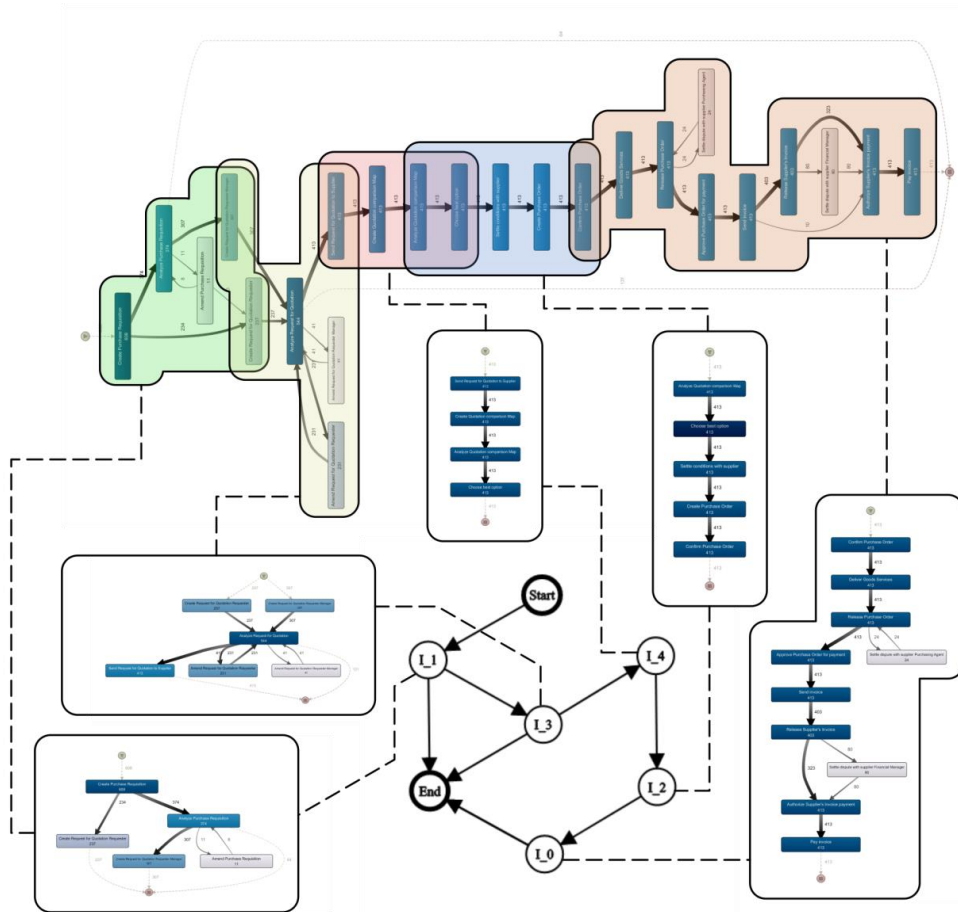
Figure 4: Example of simple process union. Mined intention map with corresponding activity-oriented maps for each strategy and their equivalence in separately mined full process map at the top.

6. Experimental results

Presented method was tested on real data of VR-Chemistry Lab project. For this study 44 participants were solving giving task in a chemical laboratory in virtual reality. Users were trying to indicate substances in four vessels by mixing them with different reagents and observing reactions. Data was collected in .csv format. Discovered by fuzzy miner map is shown on Figure 1.

Data includes recorded user actions within the game with information about action subjects (mixing specific reagents, breaking vessels, interacting with laboratory journal etc.). At preprocess it was filtered to remove duplicates caused by nature of logging. Then data was uploaded to the developed prototype and strategy map was discovered with modified Map Miner Method. Combination "action + subject" was considered as input action. Order of actions was defined by timestamps. Number of mined strategies was chosen using combination of heuristics and information criteria calculating to avoid both strategy duplicates and high level of diversity. Local minimum in range from 15 to 25 strategies is 20 (Figure 3 (b)). Event logs were filtered of system information not relevant to player behavior. As activities combination of activity and subject was treated to differentiate operations with different substances. Discovered map of sub-intentions is shown on Figure 5 (a).

We used subsequences of event related to particular strategy to model player enactment. For example on Figure 5 (b) modeled Strategy 10 is shown that, according to included actions, is related to mixing in created by user vessel reagent with sulfur-containing salts. This strategy models valid

intention in the given task. Both strategies connected to start and end activities of the modes are Strategy 0.
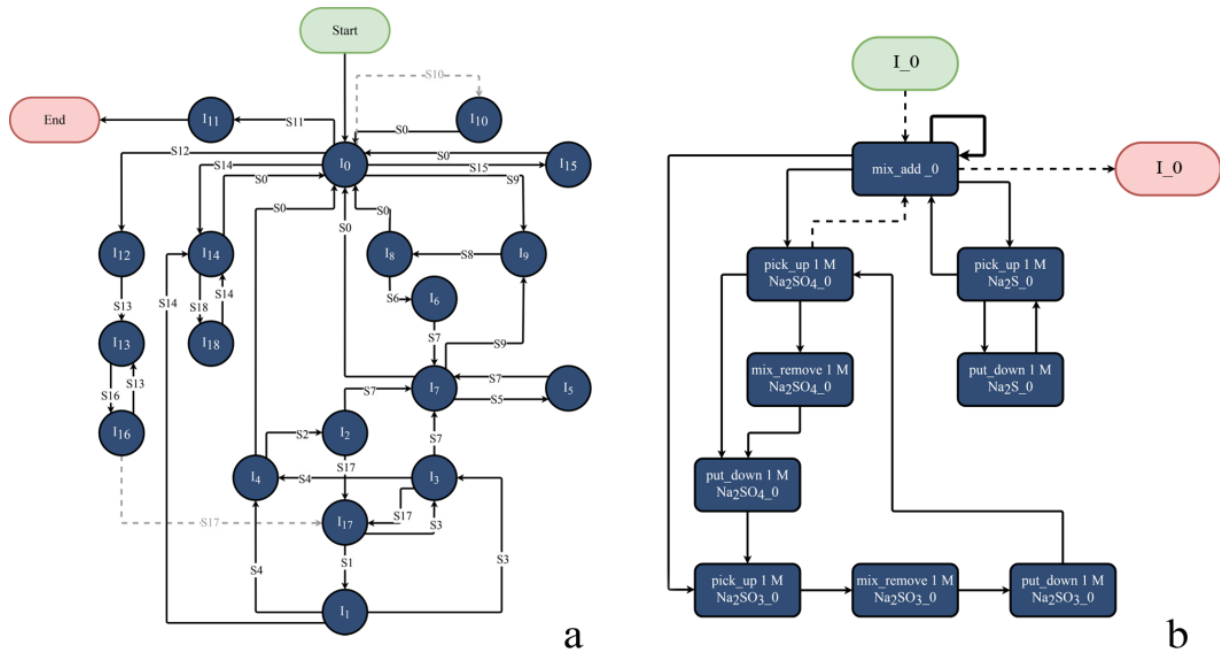

Figure 5: Intentional map of analyzed educational game for 20 strategies (a); Modeled Strategy 10 "Sulfur-containing salts application" (b).

7. Discussion and Conclusion

In this work usage of process mining for educational game analysis was studied. According to literature review goal-oriented process mining is more suitable for modeling gamified processes. Intention mining was indicated as potentially relevant technique for such task and Map Miner as the most elaborated method. However for our case it lacks of precision so we suggest combine it with other process mining algorithms to analyze mined strategies and create hierarchical model.

Described method was modified for more specific initialization. Problems with HMM initial properties setting were solved. However choice of mined strategies number still requires manual assumption. Nevertheless it's now reinforced by usage of information criteria and implies only setting the maximum limit.

Sub-intention level of Map Miner Method was formally redefined as strategy level in our approach. This helped to combine goal-oriented map with activity-oriented model in hierarchical way. Using information gained from Map Miner we construct process model for each strategy. Collection of such models forms the third level of the proposed formalism. Relation between strategy and activity levels is shown using heuristic assumptions about model properties.

Preliminary results demonstrate that PM can be potentially used for educational game analysis and discovering user behaviors. Mined strategies tend to model real user intentions or, at least, group actions in logical clusters. Resulting model is obviously superior to the initial fuzzy mined process model in terms of interpretability. The most relevant for future research directions are:

- Discover the formal proof of the stated model properties;
- Use HMM nature of mined strategies and relations between goal-oriented and activity-oriented models to formalize the replay process for intention mining;
- Develop two-level conformance checking method for presented technique. First check discovered intention map and then particular strategies.

8. Acknowledgements

9. References

[1] Codish, David, Eyal Rabin, and Gilad Ravid. "User behavior pattern detection in unstructured processes–a learning management system case study." Interactive Learning Environments 27.5-6 (2019): 699-725.

[2] Aarseth, Espen. "Playing Research: Methodological approaches to game analysis." Proceedings of the digital arts and culture conference. 2003.

[3] Yannakakis, Geogios N. "Game AI revisited." Proceedings of the 9th conference on Computing Frontiers. 2012.

[4] Richter, N. G. Process mining in programming game logs to differentiate between skill levels. BS thesis. University of Twente, 2019.

[5] Ramadan, Souad, et al. "Process Mining of Logged Gaming Behavior." 2019 International Conference on Process Mining (ICPM). IEEE, 2019.

[6] Bogarín, Alejandro, Rebeca Cerezo, and Cristóbal Romero. "A survey on educational process mining." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8.1 (2018): e1230.

[7] Ariouat, Hanane, et al. "A two-step clustering approach for improving educational process model discovery." 2016 IEEE 25th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). IEEE, 2016.

[8] Romero, Cristóbal, and Sebastián Ventura. "Educational data mining: a review of the state of the art." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 40.6 (2010): 601-618.

[9] Ghasemi, Mahdi, and Daniel Amyot. "From event logs to goals: a systematic literature review of goal-oriented process mining." Requirements Engineering 25.1 (2020): 67-93.

[10] Diaz, Oswaldo E., María Gabriela Perez, and Jorge Edison Lascano. "Literature Review about Intention Mining in Information Systems." Journal of Computer Information Systems (2019): 1-10.

[11] Khodabandelou, Ghazaleh, et al. "Process mining versus intention mining." Enterprise, Business-Process and Information Systems Modeling. Springer, Berlin, Heidelberg, 2013. 466-480.

[12] Khodabandelou, Ghazaleh. Mining intentional process models. Diss. 2014.

[13] Günther, Christian W. "Process mining in flexible environments." (2009).

[14] Pohle, Jennifer, et al. "Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement." Journal of Agricultural, Biological and Environmental Statistics 22.3 (2017): 270-293.