INTELLIGENT TUTORS IN IMMERSIVE VIRTUAL ENVIRONMENTS

Peng Yan[1], Brian M. Slator[1], Bradley Vender[1], Wei Jin[1], Matti Kariluoma[1], Otto Borchert[1], Guy Hokanson[1], Vaibhav Aggarwal[1], Bob Cosmano[2], Kathleen T. Cox[2], André Pilch[2] and Andrew "Mazz" Marry[3]

[1]*Department of Computer Science and Operations Research, North Dakota State University*
1340 Administration Ave., Fargo, ND 58102, USA
[2]*WoWiWe Instruction Co LLC, 1854 NDSU Research Circle North, Suite 45, Fargo, ND 58102, USA*
[3]*Department of Biosciences, Minnesota State University, 1104 7th Ave South,Moorhead, MN 56563, USA*

ABSTRACT

Research into virtual role-based learning has progressed over the past decade. Modern issues include gauging the difficulty of designing a goal system capable of meeting the requirements of students with different knowledge levels, and the reasonability and possibility of taking advantage of the well-designed formula and techniques served in other research fields to improve role-based tutoring. In this paper, we attempt to develop a comprehensive and adaptable goal system and intelligent tutors in an educational bioscience game by proposing a hybrid approach. Our solution supports multi-user collaborations and competitive play, and integrates a data mining model to help discover student play patterns. The overarching aim is to make informed tutoring decisions, and improve student learning and efficiency, as they work through each module in the game.

KEYWORDS

Collaborative Learning, Exploratory Technologies, Student-Centered Learning

1. INTRODUCTION

Immersive virtual environments (IVEs) offer an excellent opportunity for science education. However, the tutoring designed in these games typically considers only the students' current actions, overlooking past activities where important play patterns may be hidden. In the real world, a good instructor teaches different students with different strategies, and thus, different advice might be given based on the severity of the mistakes made. For example, a lucky guess should be remediated when the correct answer is given by chance without reasoning and analysis. This work proposes to classify past mistakes by integrating data mining techniques to analyze student play history, in order to uncover important play patterns to create focused, individualized tutoring strategies for students.

1.1 Context

There have been a number of serious games developed using role-based learning, such as the Geology Explorer (Saini-Eidukat et al., 1998) for geosciences and earth science education, the Visual Program (Juell 1999) for AI education, the ProgrammingLand MOO (Hill and Slator, 1998) for computer science education, and the On-A-Slant village (Hokanson et al., 2008; Slator et al., 2001) for anthropology education.

1.2 Intelligent Tutoring

The intelligent tutors discussed in this work are deployed in the WoWiWe Instruction Co version of the Virtual Cell, a virtual, multi-user space where students fly around a 3D world and practice being cell biologists in a role-based, goal-oriented environment (Borchert et al., 2013; White, et al., 1999).

It focuses on providing an authentic problem-solving experience that engages students in the active learning of the structure and function of the biological cell (Slator et al., 2006). Three modules have been completed: the Organelle Identification module populated with sub-cellular components where students are required to identify the nucleus, endoplasmic reticulum, Golgi apparatus, and so forth using deductive scientific approaches (reasoning, analysis, assay etc.); the Electron Transport Chain (ETC) module which demonstrates the respiration process and requires the student to understand the movement of hydrogen and electrons when ADP is converted to ATP in the mitochondria; and the Photosynthesis module which teaches students the process of photosynthesis by asking them to repair damaged photosynthesis reactions in the chloroplast. In each module of the Virtual Cell game, a guide is on hand to direct them to their next task, and a tutor is available when students struggle in accomplishing their learning goals.

1.3 Project Overview

This paper can be summarized as follows. First, we design a comprehensive goal system that is adaptable to students with different knowledge levels in a role-based, goal-oriented immersive virtual environment (IVE). The learners are assigned specific tasks in accordance with their learning goals covering various components and organelles of a cell in the first module. The difficulty of the goals increases progressively as the student works through each task, and the level can be adapted to meet the requirements of educating different students such as high school students and college undergraduates.

Second, to provide more individualized tutoring to students who have difficulties in accomplishing their learning goals, we propose a data mining model to analyze student play history, aiming to discover non-obvious but important patterns to help make better tutoring decisions. For example, two kinds of tutoring are provided based on the specific type of mistake made: blind tutoring is initiated if the discovered patterns show the mistake was made by chance, while oriented tutoring is undertaken if the uncovered patterns implied the student may have fundamental confusion between two organelles.

We also developed a library of problem-oriented knowledge to help locate the confusions that students may have as they explore in the Virtual Cell. Supported by this library, tutors act as sub-topic experts, keep an eye on student progress, and match the current student’s actions to previous students’ actions, allowing the tutor to ask relevant questions that may be preventing the student from moving forward.

Last, we employ ontology mapping techniques to improve the data quality of the student play history and model student learning activities. For example, tutoring decisions depend to a high degree on the type of diagnostic assay the student is performing. Therefore, the agents developed in this work are capable of offering more individualized and problem-oriented tutoring.

2. THE GOAL SYSTEM

There are three modules in the Virtual Cell: the Organelle Identification (ID) module, the Electron Transport Chain (ETC) module and the Photosynthesis module. They are designed to help students improve scientific reasoning and their understanding of the scientific method (Borchert et al., 2010).

2.1 ID Module Goal System

The ID module was developed to provide an introduction to game play, with the student tasked with making hypotheses, gathering data in the form of required experiments, and finally identifying seven different organelles contained in the cell. These tasks represent seven parallel goals (identifying nucleus, endoplasmic reticulum, Golgi apparatus, mitochondria, chloroplast, ribosome and vacuole) which together form the structure of the goal system in the ID module. Figure 1 illustrates part of the goal system. Each goal is represented by a series of tasks to be completed. Through performing the tasks, students learn the scientific/deductive process to follow in order to confirm the identification of an unknown substance, in this case an organelle. For example, a student might be asked to identify the Golgi apparatus in the Virtual Cell. To complete this goal, three tasks must be performed, and the goal of identifying this organelle cannot be completed if any of the tasks are skipped or performed in the incorrect order.

- ***Task 1:*** hypothesize that an unkown organelle is a Golgi apparatus and scan it.
- ***Task 2:*** perform the marker assay for glycosyl transferase to confirm the hypothesis.
- ***Task 3:*** make a report identifying it as a Golgi apparatus.

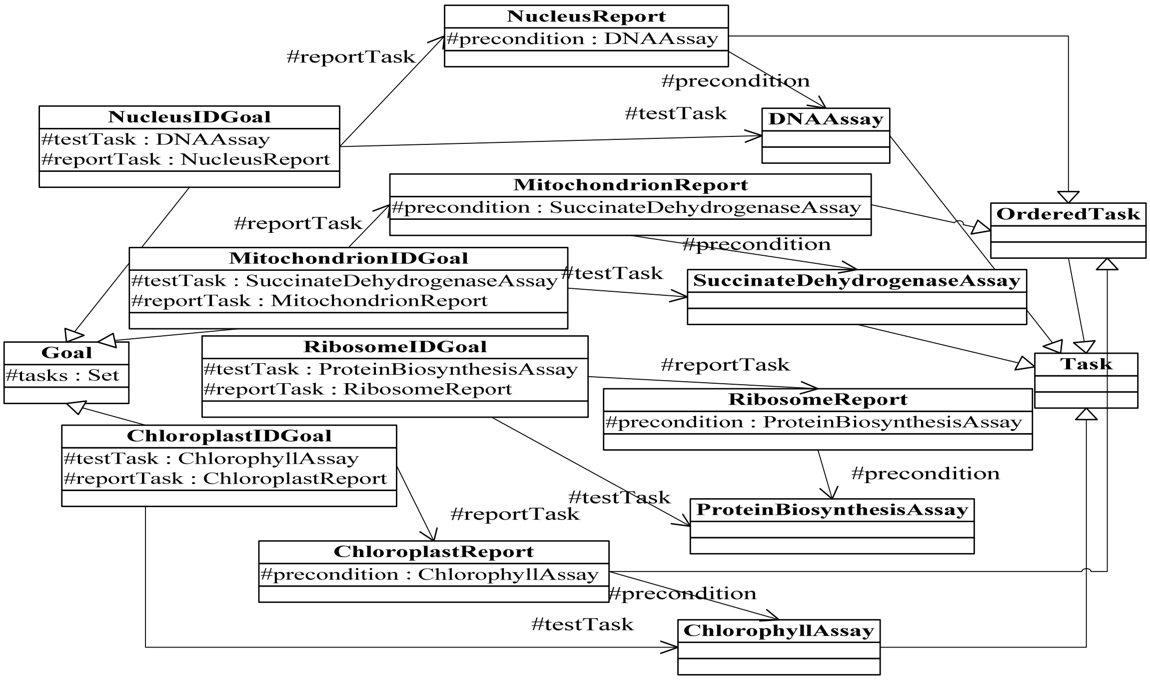

Figure 1. Goal structure for the ID module:A goal is completed if its corresponding tasks are all completed. For example, to complete the "NucleusIDGoal", two tasks need to be completed: the "DNAAssay" task which requires the player to perform the marker assay for DNA synthesis, and the "NucleusReport" task which requires the player to file a report to confirm the hypothesis. Note that the "DNAAssay" task must be completed before the "NucleusReport" task.

2.2 ETC and Photosynthesis Goal Systems

The ETC module introduces the player to the electron transport chain and cell respiration by first presenting a healthy mitochondrion, and then guiding the player to a damaged mitochondrion. The player is then required to repair the damaged ETC by accomplishing 3 tasks.

- ***Task 1:*** investigate possible reasons for the cell's low ATP production;
- ***Task 2:*** repair the damaged electron transport chain.
  - ***Task 2.1:*** hypothesize the broken ETC complex.
  - ***Task 2.2:*** leave the mitochondrion and find a ribosome.
  - ***Task 2.3:*** purchase a new version of the broken complex.
  - ***Task 2.4:*** re-enter the broken mitochondrion.
  - ***Task 2.5:*** replace the broken complex with a new one.
- ***Task 3:*** file an incident report

There are six complexes involved in a functional ETC, as well as a healthy supply of the "raw material" substrates: succinate dehydrogenase, hydrogen, and oxygen. Any one of the complexes could be broken which would bring the ETC to a halt. As remediation, a substrate pointed at the proper complex will jump-start the ETC from that point up until the broken complex is encountered.

The photosynthesis module is designed to teach one of the most important biochemical processes in plants. Its goal structure is similar to the ETC module except for the final task; the player is required to repair an inactive section of the chloroplast to produce ATP. The detailed goal structure of the photosynthesis module is shown in Figure 2.

To make a game suitable for students with different backgrounds e.g. high school students and college undergraduates, the tasks or goals in the ETC and photosynthesis can be easily adjusted. For example, in the photosynthesis module, one task is to let students gather the needed substrates to improve the health of the cell. It is obvious that the more missing substrates, the harder the task. Many students know that CO_2 is necessary in the photosynthesis process. But that photons serve as tools to shake electrons free from chlorophyll, and move energy through the rest of the electron transport chain, is more esoteric knowledge. Therefore, the conceptual difficulty of the task of collecting necessary substrates could be decreased by adding protons to the student's inventory at the beginning.

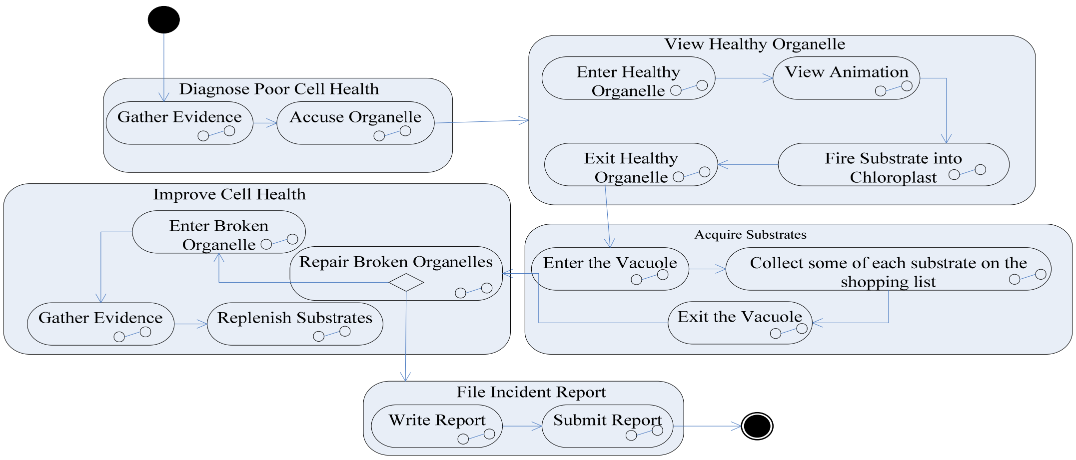

Figure 2. Goal structure for the Photosynthesis module:A goal is completed if its corresponding tasks are all completed. For example, to complete the "ViewHealthyOrganelle" goal, the player must enter a healthy organelle first, and then view an educational animation. After that they learn how to use substrates to test organelles by firing an available substrate into the chloroplast, and finally exit the healthy organelle.

3. INTELLIGENT TUTORS

From the perspective of intelligent tutoring systems, the agents of interest must fundamentally support models of the knowledge of a domain expert and an instructor (Slator, 1999). We maintain software tutoring should not only consider current activities, but also should be aware of past performance, in order to give contextualized, individualized tutoring.

3.1 Knowledge Base

We maintain various information resources in the library of the game. First, based on a concern for information canonicity and coverage, we integrate the relevant biological knowledge derived from our content experts into the backend knowledge base. Students can easily use the toolbox provided in the game to navigate to the game's online encyclopedia, using it to find desired information. Second, multimedia educational materials are incorporated to complement the system's existing knowledge, e.g. the animations introducing the process of electron transport and photosynthesis. Students can actively study these materials at any point in the game. Or in another scenario where the player is stuck, problem-oriented and individualized knowledge will be offered as often as needed. This kind of knowledge, like non-obvious student play patterns, goes beyond domain-specific knowledge, and constitutes another significant aspect that helps in customizing the tutoring for individuals. In addition, mini-games that go along with each lesson module are developed for situations where students finish early or where they have trouble understanding lesson concepts. For example, the mini-game designed for the ETC module teaches ATP production by requiring the student to re-orient complexes to correct positions so that ATP can be steadily produced.

3.2 ID Module Tutors

There are currently three modules in the Virtual Cell. The first is the Organelle Identification (ID) module, where students need to identify seven organelles correctly to accomplish their learning goals. During this activity, play history is recorded by the system which serves as important data for determining the type of tutoring to be provided at later stages in the game.

3.2.1 Modeling Method for Capturing Learning Activities

Table 1. Semantic Type - Concept Mapping

Semantic Type	Instances (Examples)
Organelle	nucleus, endoplasmic_reticulum, Golgi_apparatus, mitochondrion, chloroplast, ribosome, vacuole
Correct Report	correct_report_no_assay, correct_report_correct_assay, correct_report_incorrect_assay
Incorrect Report	nucleus_as_ mitochondria, endoplasmic_reticulum_as_ chloroplast vacuole_as_Golgi_apparatus
Assay	DNA_synthesis, succinate_dehydrogenase, phospholipid_biosynthesis, glycosyl_transferase, protein_biosynthesis, chlorophyll

To model student play history, we define a student profile which is essentially a set of related concepts that together represent the student learning activities. This model is inspired by the closed text mining algorithm (Srinivasan, 2004). To further differentiate between different concepts, semantic type (ontological information) is employed in profile generation. Table 1 illustrates part of the semantic type - concept mappings. Here, each student profile is defined as a vector composed of a number of semantic types.

$$profile(Stud_i) = \{ST_1, ST_2, ..., ST_n\} \tag{1}$$

Where ST_i represents a semantic type to which the related concepts representing the student's learning activities belong. Each semantic type can be further expanded by an additional level of vector composed of concepts that belong to this semantic type and relevant to the student's play activity.

$$ST_i = \{w_{i,1}m_1, w_{i,2}m_2, ..., w_{i,n}m_n\} \tag{2}$$

Where $m_j \in ST_i$ represents a concept under the semantic type ST_i and $w_{i,j}$ denotes its weight. When generating the profile we replace each semantic type in (1) with (2). To compute the weight for each concept in (2), we employ a variation of the *TF*IDF*weighting scheme (Jin and Srihari, 2006) and then normalize the weight under each semantic type:

$$w_{i,j} = s_{i,j} / highest(s_{i,l}) \tag{3}$$

Where $l = 1, 2, ..., r$ and there are totally r concepts for ST_i, $s_{i,j}$ = the number of occurrences of m_j, where $m_j \in ST_i$. By using the above three formulae we can build the corresponding profile for any given student. To summarize, the procedure for building a student profile is composed of the following four major steps:

- ***Step 1:*** Concept Retrieval: retrieve all relevant concepts from the student play history.
- ***Step 2:*** Semantic Type Employment: each concept is associated with and grouped under one semantic type (e.g., Assay, Incorrect Report) in which it belongs.
- ***Step 3:*** Weight Calculation: for each concept, a variation of the *TF*IDF* scheme is used to calculate the weight (i.e., $s_{i,j}$ as shown in Formula 3).

- ***Step 4:*** Weight Normalization: the weight of each concept is further normalized by the highest concept weight observed for its semantic type as given in Formula 3. Within each semantic type, all concepts are ranked according to the normalized weights.

3.2.2 Tutoring Strategies

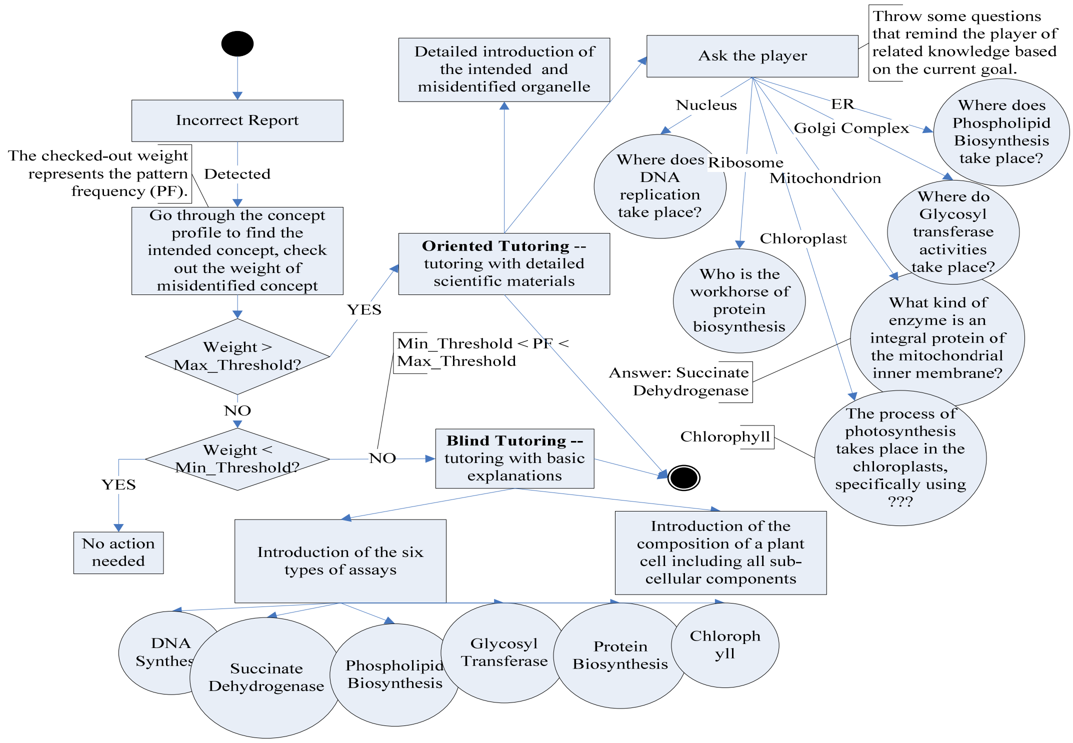

Figure 3. The tutor's decision making process for incorrect reports: Once an incorrect report is filed, the system will examine the player profile and check out the weight of the misidentified concept. If its weight exceeds the threshold defined in the system, either blind or oriented tutoring will be activated.

The generated profile represents the student's play history and potentially includes valuable patterns to be mined and utilized. Each time the student conducts an activity, e.g., performing a DNA synthesis assay on a nucleus or mistaking a mitochondrion for a chloroplast, the profile is updated (i.e., the weight of each concept in the profile is recalculated) to reflect the up-to-the-minute learning status of the student. The tutoring system makes tutoring decisions based on the discovered pattern frequency (i.e. weight computed using formula 3). If the pattern frequency does not reach the threshold predefined in the system, blind tutoring will be offered, otherwise oriented tutoring will be launched. Here blind tutoring means the tutor provides general conceptual information like the structure and composition of a plant cell, and does not further investigate the student's learning problem (e.g., "did the student perform an incorrect assay?"). On the contrary, oriented tutoring indicates the tutor starts to explore the student's past activities, such as "the student is mistaking what for what?" and then attempts to offer the best problem-oriented advice.

For example, a student may have confusion between a nucleus and mitochondrion. If this mistake has been captured frequently enough to reach the threshold set in the tutoring system, a student-oriented tutoring session (with a specific focus on explaining the nucleus and mitochondria) will be activated. Figure 3 demonstrates the detailed remediation strategies for incorrect reports.

3.3 ETC and Photosynthesis Tutors

We adopt similar tutoring strategies in the ETC and Photosynthesis modules since they have similar goal structures. The ETC occurs in mitochondria as the third and last stage of cellular respiration. Tutoring strategies for the ETC are illustrated in Figure 4.

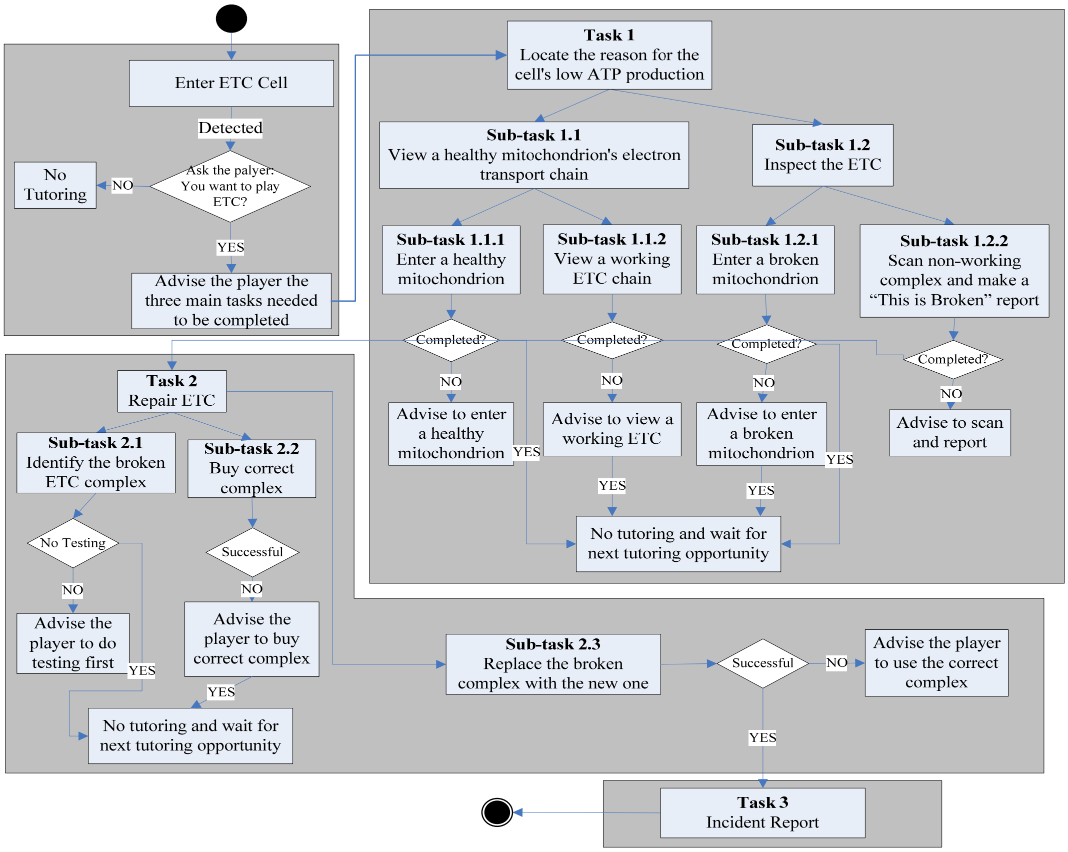

Figure 4. Tasks and tutoring opportunities in the ETC module:Tutoring is offered if a failed task completion is detected. For example, "sub-task 2.2" requires the player to buy a complex. If the player fails to buy the correct one (e.g., an incorrect one is bought), a tutoring message will be sent.

However, appropriate tutoring opportunities might be difficult to recognize in some complex cases. For example, in the ETC module, the final task is to enable a broken mitochondrion to start transporting electrons. In order to activate the electron transportation, broken carrier substrates that hold electrons like NADH dehydrogenase must be replaced. Before that replacement, a sequence of activities need to be conducted: 1) check the current inventory to find a healthy complex; 2) buy a healthy complex if the current one is running out; 3) replace the broken complex with the healthy one. In this case, it is not certain all of the activities must be involved, which means one or more of them might not be needed based on practical considerations such as the current inventory status.

Two alternative tutoring strategies are proposed to handle this issue. One is to wait until an incorrect substrate is chosen to replace the broken one, no matter whether the student inventory has the correct substrate or not. In this case, the student might get stuck if their inventory is running out of available substrates. The other is to check the student inventory before the replacement task is initiated.

4. CONCLUSIONS

This paper proposes a new solution to intelligent software agent tutoring that integrates data mining techniques with intelligent agents. This instructional system aims to individualize learning experiences through the incorporation of a data mining model based on student learning history. Blind tutoring is provided to meet the requirements of a majority of the students, while oriented tutoring is customized for struggling students. The integration of data mining techniques would also benefit other related tasks such as educational psychology and student-centered learning.

To support the development of such tutors, a comprehensive goal system that covers various pedagogical scenarios, targeted to focus the learning process on academic goals is presented. In this virtual environment, students demonstrate mastery of required knowledge and skills through the completion of these learning goals.

The solution introduced in this work is implemented in an educational 3D game for biology students, Virtual Cell (Borchert et al., 2013). Besides cellular biology, there is potential for adapting this solution to other applications that involve scientific reasoning and scientific methods. For example, the proposed data mining model can be re-used in other educational game arenas: psychology, math, physics and social science and humanities. The goal system supports cellular biology education and can also be extended to meet the needs of sub-disciplines like bacteriology, and many other specialized cells in multicellular organisms.

For future work, the ontologies developed in this task can be further improved to fit more learning activities. Furthermore, more tutoring opportunities might be discovered by collecting and analyzing the logs generated from mini-games. And common patterns can be generalized to construct typical learning cases for library enrichment. We will be exploring these issues and evaluating their performance in our future work.

ACKNOWLEDGEMENT

This work was supported by NIH Award #R44RR024779. The content is the responsibility of the authors, and not the official views of the National Center For Research Resources or the National Institutes of Health .

REFERENCES

Borchert, O. et al. 2010. Principles and Signatures in Serious Games for Science Education. *In Gaming and Cognition: Theories and Practice from the Learning Sciences*. IGI Global, Hershey, USA, pp. 312-338.

Borchert, O. et al. 2013. A 3D Immersive Virtual Environment for Secondary Biology Education. *Proceedings of Society for Information Technology & Teacher Education International Conference 2013*. New Orleans, USA.

Hill, C. and Slator, B. M., 1998. Virtual Lecture, Virtual Laboratory, or Virtual Lesson. *Proceedings of the Small College Computing Symposium (SCCS98).* Fargo-Moorhead, USA, pp. 159-173.

Hokanson, G., et al, 2008. Studying Native American Culture in an Immersive Virtual Environment. *Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT-2008).* IEEE Computer Society Press. Santander, Spain. pp. 788-792.

Jin, W. and Srihari, R. K., 2006. Knowledge Discovery across Documents through Concept Chain Queries. *Proceedings of The Sixth IEEE International Conference on Data Mining – Workshops(ICDMW '06).* Hong Kong, China, pp. 448-452.

Juell, P., 1999. Educational Opportunities using VRML in the AI Classroom. *Proceedings of the International Conference on Mathematics/Science Education and Technology (M/SET-99).* San Antonio, USA, pp. 119–123.

Saini-Eidukat, B. et al, 1998. Text-Based Implementation of the Geology Explorer, a Multi-User Role-Playing Virtual World to Enhance Learning of Geological Problem-Solving. *In GSA Abstracts with Programs*, Vol. 30, No. 7.

Slator, B. M. et al, 2001. Teaching with Immersive Virtual Archaeology. *Proceedings of the 7th International Conference on Virtual Systems and Multimedia (VSMM-2001).* Berkeley, CA, Oct. 25-27.

Slator, B. M. et al, 2006. *Electric Worlds in the Classroom: teaching and learning with role-based computer games.* Teachers College Press, New York, USA.

Slator, B. M., 1999. Intelligent Tutors in Virtual Worlds. *Proceedings of the 8th International Conference on Intelligent Systems (ICIS-99).* Denver, USA, pp. 124-127.

Srinivasan, P., 2004. Text Mining: Generating hypothesis from Medline. *In Journal of the American Society for Information Science and Technology*, Vol. 55, No.5, pp. 396-413.

White, A. et al, 1999. The Virtual Cell: An Interactive, Virtual Environment for Cell Biology. *Proceedings of World Conference on Ed Media and Hypermedia (ED-MEDIA 99).*Seattle, USA, pp. 1444-1445.