

A Conceptual Framework for Combination of Educational Data Mining and Software Cybernetics

Yi CAI*

College of Information Science and Technology
Beijing University of Chemical Technology
Beijing, China
E-mail: caiyi@mail.buct.edu.cn
+* Corresponding author

Qunxiong ZHU

College of Information Science and Technology
Beijing University of Chemical Technology
Beijing, China
E-mail: zhuqx@mail.buct.edu.cn

Abstract—Both educational data mining (EDM) and software cybernetics are emerging interdisciplinary research fields at present. This paper analyzes the combination mechanism between them in educational environment, and addresses a conceptual framework to improve the success of educational software. The framework regards educational software as the controlled object, EDM as the metric observer, and educational interpretation/evaluation as the feedback to satisfy the desired objectives relevant with the characteristics of integration, reliability and effectiveness.

Keywords—educational data mining; software cybernetics; combination; framework

I. INTRODUCTION

Software is widely used in the educational environment to facilitate educational institutions or those interest parties related through a variety of computer-based ways. For example, learning management system (LMS) can help educational institutions to manage educational records and deliver online courses [1]. And also, intelligent tutoring system (ITS) aims to provide immediate and customized instruction or feedback to learners in a meaningful and effective manner by using plenty of growing computing technologies [2]. Meanwhile, as an emerging interdisciplinary research field, educational data mining (EDM) is concerned with developing methods for exploring the unique types of data that come from both traditional educational environment and computer-based educational systems [3]. Furthermore, it seems that computer-based educational systems including educational software actually provide a huge amounts of data for the mining process. Although both software and EDM play different roles in educational environments, as the ultimate goal, they are inevitably to improve educational quality, and to promote the level of educational decision making.

As it was first proposed in 1994 [4], the idea of software cybernetics attempted to apply cybernetic approaches to solve software engineering problem and then in 2002 [5-6], it was defined to explore the interplay between software processes and control. There has been consistently expanding since then, mainly through the International Workshop on Software Cybernetics (IWSC) [7]. Since educational software is certainly one member of software family, it can

be considered from the viewpoint of software cybernetics. If software cybernetics is applied in educational environment that can also be called educational software cybernetics, it may collaborate with EDM to improve the role of educational software for reaching the educational goal.

The aim of the paper is to delineate a conceptual framework for combination of educational data mining and software cybernetics in order to improve software's success in the educational environment, and then to improve the success of education. Section 2 and 3 will present a brief introduction of EDM and the role of software within EDM. Section 4, 5 and 6 will address why educational software cybernetics and EDM can have an interaction for improving educational success. In section 7, the conceptual framework will be proposed. In the end, concluding remarks are contained in Section 8.

II. EDUCATIONAL DATA MINING

Although it is still a rather young research area, educational data mining is rapidly developing through the effort of world-wide researchers and institutions, such as International EDM Society aiming to support collaboration and scientific development in this new discipline. In fact, EDM covers the combination of various academic fields such as education, computer science and engineering, mathematics and statistics, and also is closely related to machine learning, and learning analytics (LA) [3]. Specifically, EDM is concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings, and using those methods to better understand students, and the settings which they learn in [8]. Its goal is not only to promote learner's success during learning or educational process, but also to improve the quality of managerial decisions in the educational environment.

EDM analyze data generated by a wide variety of educational systems, which can be classified into two groups depending on the involvement of information technology and computer science in educational practices. One group is called traditional education, and another one is computer-based education. Software exists in both groups, but plays different roles through respective ways. Software within traditional education mostly collect and store data into a static status to wait for someone discovering the knowledge

in the data. However, computer-based educational software may achieve the same function of traditional educational software, and also directly participate the educational process. Even to some extent, it has liberated some educational factors through intelligent methods and replaced some educational resource to improve the development of education. Actually, computer-based educational software already becomes one critical part of the current modern education. Therefore, data from educational software can be used to EDM as well as to check the effectiveness, usability and reliability of educational software for its relevant educational environment. It means that EDM can analyze data to improve software's success in educational environment. Some details and controllability of educational software will be presented in section 3 and 5.

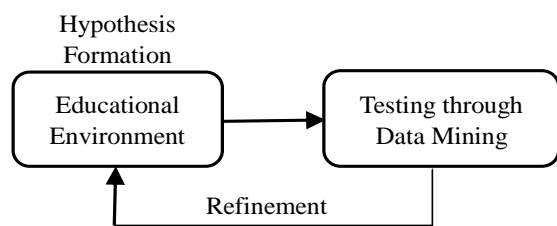


Figure 1. Application process of EDM

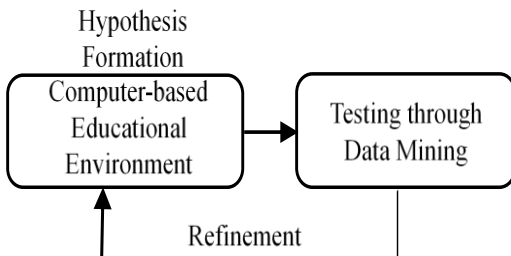


Figure 2. Application process of EDM in computer-based educational environment

From an educational and an experimental viewpoint, the application process of EDM can be seen as an iterative cycle of hypothesis formation, testing through data mining, and refinement (see Fig. 1) [3]. Furthermore, depending on the computer-based type of educational environment, the process can be depicted as the cycle in Fig. 2. In the cycle, hypothesis formation defines the educational object of data mining differing from other domains, and refinement is a feedback to propose educational decisions to improve or optimize the educational environment.

A variety of methods are widely used within EDM. Some of them are popular in the application of general data mining, such as prediction, clustering, relationship mining, social network analysis, process mining, and text mining. Meanwhile, some methods are particular with EDM, such as

the distillation of data for human judgment, discovery with models, knowledge tracing (KT) and nonnegative matrix factorization. Definitely, it is not clear that which method is adaptive to the new conceptual framework in this paper. It needs to clarify the problem over the future application of this framework.

III. EDUCATIONAL SOFTWARE

Human has always tried to explore appropriate technological processes and resources in order to facilitate and improve success of education. The emergence of computer with software has developed educational technology into a new stage. Consequently, software is widely used throughout nearly all of the educational area. It is not doubtful that software is pretty important for modern education.

Educational software is computer software, the primary purpose of which is teaching or self-learning [9]. Due to the complexity, diversity and hierarchy of education, educational software presents a variety of performances in terms of content, object, human, time and technology. For example, educational software can be classified by academic subjects or contents, such as various instructional software for chemistry, physics, and mathematics. It also may be sorted by institutions such as primary school, middle school, high school, and college or university. In particular, some of educational software have used intelligence techniques for educational modeling, adaptation, personalization, and so forth. A number of integrated software systems, such as learning management system (LMS) and intelligent tutoring system (ITS), have emerged to facilitate educational success. Meanwhile, the wide use of Internet overwhelmingly promotes the transmission of knowledge and improves the communication within educational software. Over all, the development of educational software is changing the educational world.

However, it is a big challenge that the application of software in educational environments and progresses has brought active influence as well as uncertain effect. There are at least three problems that those educational software related should be carefully considered. First, the conflict between conventional and aggressive opinions facing the practical application of educational software may lead to game or deviation in education. For example, senior teachers do not enjoy using ITS because they have developed their life-long teaching styles and habits. Second, not all software are initially reliable to its application environment. Maybe, a terrible software even plays a bad role during a teaching process. Such as, frequent running faults in a physics software may lead to a severe failure in the class teaching. Third, educational effectiveness is also a problem that whether student can learn much more efficient through educational software or system. And also, teacher and student may not satisfy the learning outcome with educational software or system. In fact, all of these three problems can be concluded into the interaction characteristics including Integration, Reliability, and Effectiveness between traditional education and developing software.

And then, these three problems are daily working in the current educational environment with software. Some of the running information or activities are unfolding through the data relevant with the three problems, which can be collected and analyzed through data mining technique to improve the success of educational software. Furthermore, it can help to reach the goal of improving educational quality, and promoting educational decision making. That's the critical point for the desired objective of education, and also the potential link between EDM and educational software.

IV. CONTROLLABILITY OF EDUCATIONAL SOFTWARE

Cybernetics is concerned with the scientific study of control and communication in animals and machines [10]. Philosophically, both software and educational system can be controlled from the view point of primary cybernetics. For all the controlled objects, controllability is one of the fundamental ideas in cybernetics. Roughly speaking, a dynamic system is controllable if an arbitrary state of it can be moved into another arbitrary state in a finite length of time [11]. And controllability is the ability of a system to have control over its responses.

The issue of controllability of educational software actually refers to two questions: controllability of software and controllability of education.

Controllability of software is the fundamental question of whether or not software behavior can be controlled. Although software is dramatically different from conventional controlled objects such as automobile, aircraft, and chemical processes, it can be controlled while software serves as the controlled object and control is applied to the software. Because a controller can transform a software or modify its internal parameters during its process of requirement, design, development, maintenance, testing, rejuvenation and execution to achieve desired objectives [7] [11], it is noted that the controller in software cybernetics is diversified by the circumstance that software serves as the controlled object.

It is known that education is an extremely complex system, but it can respond to or make adjustments based upon input from the educational environment. Previous studies have always explored the laws or patterns that dominated the inherent activities or actions within education. For example, learning theory focuses on the integration of two very different processes, namely an external interaction process between the learner and his or her social, cultural or material environment, and an internal psychological process of elaboration and acquisition [12]. Another example, Student Development Theory has developed a number of models to improve student's success, such as 'Input-Environment-Output' model [13], and student involvement theory [14].

As what the examples exposed, based upon various educational views, a variety of factors obviously retain different influences on the success of education that leads the controllability of education into totally difficult circumstances. Nevertheless, among these factors, some of them can be controlled obeying the educational laws or patterns, and others cannot be controlled due to the terms of

time and space. It seems that the controllability of education really exists, but it is partially developed by specifying for some factors. Someone of current educational models may be regarded as a controlled object, and those educations related including faculty/staff, and student may become a controller.

Since educational software is applied to achieve some definite functions, it is often developed based on educational theory or model. While educational software serves as a controlled object in educational environment, the controllability existing in both software and education will play roles to contribute to the controllability of educational software. In context, those the controllability of educational software related become the controller. Consequently, the controllability of educational software can be extracted through considerate research work as an academic viewpoint.

V. SOFTWARE CYBERNETICS IN EDUCATIONAL ENVIRONMENT

Cai defined software cybernetics as the interplay between software processes and control [11]. In general, software cybernetics addresses issues and questions on: (1) the formalization and quantification of feedback mechanisms in software processes and systems; (2) the adaptation of control theory principles to software processes and systems; (3) the application of the principles of software theories and engineering to control systems and processes; and (4) the integration of the theories of software engineering and control engineering.

Since software cybernetics is concerning the relation between software engineering and control engineering, what issues it addresses become more complicated due to the involvement of educational settings. First, the interaction characteristics of Integration, Reliability, and Effectiveness between traditional education and developing software should be considered within feedback mechanisms in educational software processes and systems. Second, according to the characteristics of education, such as complexity, diversity, hierarchy and being partially controlled, the adaptation process of control theory to educational software needs to be dealt with carefully through educational theory. Third, if educational software serves as a controlled object, which parts of software processes or internal parameters can reveal the educational characteristics? Is there a theoretical model to delineate the controlled object? Fourth, undoubtedly, the integration of the theories of software engineering, control engineering and education also becomes more complex than that between software engineering and control engineering.

In particular, data relevant with the controlled object-educational software should be analyzed before it starts a feedback process, because the data cannot be used directly to the feedback to compare with the desired objectives. It means that the data needs to be transformed into appropriate educational information by some kinds of methods. Here EDM is a good choice as a metric observer that can transform quantitative data into valuable information through mining technology. If data about the interaction characteristics of Integration, Reliability, and Effectiveness

is collected, corresponding methods or models of EDM will be selected to analyze the data.

In addition, the notion of feedback is a fundamental principle in educational software cybernetics. On the one hand, the feedback makes educational software cybernetics to closed-loop or feedback control. On the other hand, the role of feedback in educational environment is to provide appropriate information that can compare with the desired objective of education. Here interpretation/evaluation is proposed to analysis the outcome of EDM as educational feedback.

VI. EDUCATIONAL DATA MINING AS METRIC OBSERVER

Why does educational software cybernetics need metric observer? First, while educational software serves as a controlled object, quantitative data from it cannot be directly used as a feedback, because data is not information, and needs to be analyzed through metric tools. Second, since educational software is developed based on educational theory, operates in educational environment, and tries to achieve the desired educational objectives, an appropriate metric observer should be set to transform the data into the information matching the objectives, environment and theory. Therefore, through explanation in previous sections, it is necessary to select EDM as the metric observer.

In order to evaluate the success of educational software, there are at least three questions to be answered. And these three questions, which have been discussed in previous section, can be depicted by the interaction characteristics of Integration, Reliability, and Effectiveness. They are all involved in the cybernetics process of objective, controlled object, environment, observer and feedback. It is a critical issue that EDM, as metric observer, collects and analyzes the data relevant with these three characteristics to produce the information matching the cybernetics process. The following types of data may be transported to EDM as the metric criteria.

A. Data of Integration

Here integration refers to whether or not educational software can replace some conventional parts of education, and also refers to the ability of educational software to reach harmony with other educational settings. Basically, we can use a Yes/No binary data to evaluate the integration between software and other settings within a same educational system. Furthermore, statistics data can be used to evaluate the integration ability of educational software in a wide view. For example, the probability of choosing to use a physics

software among all faculty who are teaching physics in an educational district can present the integration between the software and conventional classes in physics.

B. Data of Reliability

In this context, reliability refers to the probability that educational software will work properly for a specified period of time in current educational environment. Measuring software reliability is difficult, thus quantifying software reliability in educational environment becomes more complex due to the involvement of diversified educational factors. Theoretically, one way is to count the running failures of educational software in periodic intervals. For instance, while a new software is introduced into an educational environment, data of operating failures can be recorded and calculated in the weekly interval to display the reliability in context. However, it must be noted that some of running failures are derived from internal faults in software, it is difficult for educational users to solve these failures, but the result can be fed back to the developer of software.

C. Data of Effectiveness

In this paper, effectiveness refers to the capability of educational software in function realization when it is used in educational environment. The aim of developing educational software is to provide specific abilities or functions that can benefit for education. Although it is impossible that an educational software can realize all the functions of the corresponding educational environment, it is good that software can achieve enough functions according to the designed requirement. For a simple example, when a chemical software with 30 functions is applied in an experiment class, we can testify how many functions have been realized during its application.

The feasibility of EDM as metric observer, and data types are discussed in this section, but how to mine the data needs to be further explored. It means that selecting appropriate mining methods is not presented in the paper.

VII. A CONCEPTUAL FRAMEWORK

Based upon the combination mechanism between educational data mining and software cybernetics, a conceptual framework is delineated in this section (See figure 3). The framework includes the following parts: desired input, compare, educational software as controlled object, educational data mining (EDM) as metric observer, output and interpretation/evaluation as feedback.

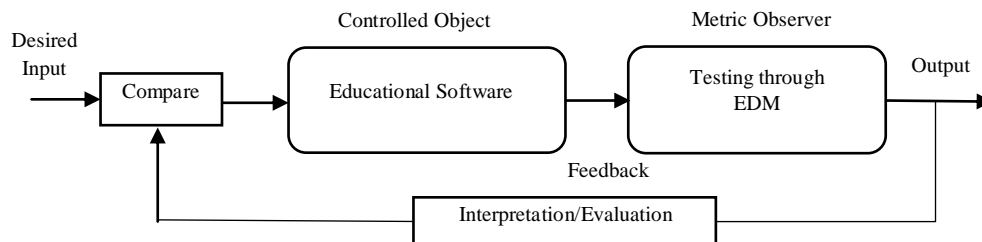


Figure 3. A conceptual framework

A. *Desired Input*

Here desired input, or desired objective refers to what the controller expects to achieve through the coming operation of the framework. Since it is concerned with the problems associated with the characteristics of Integration, Reliability, and Effectiveness, the criteria of desired input can be defined according to these the characteristics. Usually, if the criteria of desired input is decided, other parts of the framework will work for reaching the criteria through specific ways.

B. *Compare*

After several rounds of operations, whether or not the desired input is achieved needs the comparison between the desired input and the feedback. The part of compare maybe adopt both qualitative and quantitative ways to realize the comparison. It depends on the fact type of the desired input and the feedback. Compare in this context is probably more complex than that in a conventional control system.

C. *Educational Software as Controlled Object*

Since educational software serves as the controlled object and the corresponding control is applied, it will operate or respond through its internal mechanism, and then deliver raw data to next part of the framework. It must be noted that some attributes associated with the characteristics of Integration, Reliability, and Effectiveness have been embedded into educational software during its development process. Otherwise, raw data from this part may be unavailable for EDM.

D. *Educational Data Mining (EDM) as Metric Observer*

In this part, various methods of EDM can be used to analysis the raw data from the previous step. Although we have not clarified which methods are adaptive to the framework, we still believe that EDM can transform the raw data into valuable information, because EDM has successfully been applied in a wide area for finding much new knowledge hiding in education. In addition, developing new specific mining techniques for this framework is also necessary.

E. *Output*

Here output refers to the outcome of EDM. Besides of being interpreted or evaluated through educational or engineering opinions as the feedback, output can be recorded into a history list, which can display the growing process of the current framework, and also will be regarded as the experience for the future application.

F. *Interpretation/Evaluation as Educational Feedback*

This part is very important to apply the mining result acquired to the feedback. At least two tasks need to be completed by the part. First, it has to satisfy the demand of comparing with the criteria of the desired input. Second, it also can provide some meaningful suggestions about how to improve the framework to the controller.

In order to achieve the desired objective, the conceptual framework may keep on operating while the controlled object-educational software becomes much more robust.

From this view, the framework seems like an iterative cycle for improving the success of educational software.

VIII. CONCLUDING REMARKS

This paper analyzes the associate mechanisms between software cybernetics and educational data mining, and also presents a framework to help to achieve the success of educational software. However, the framework is only a conceptual one. Someone may doubt that it seems to a simple combination for some objectives. In fact, there really are a number of questions to be answered for this issue. For example, whether there is a theoretical model depicting the controlled object while educational software serves as a controlled object. How to mine the educational data has not been fixed. How do we implement the framework in educational practice? What are the details of metric criteria for the interaction characteristics of integration, reliability, and effectiveness? What methods of EDM will be selected to analyze the data? These questions will lead us to focus on the issue and to discover more details about it in the future.

REFERENCES

- [1] Learning management system[DB/OL]. https://en.wikipedia.org/wiki/Learning_management_system, 2016-12-11.
- [2] Joseph Psotka, Sharon A. Mutter. "Intelligent Tutoring Systems: Lessons Learned," Lawrence Erlbaum Associates, 1988.
- [3] Romero, Cristobal, Sebastian Ventura. "Data mining in education," Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery 3.1 (2013): 12-27.
- [4] K.Y.Cai, "On the Concepts of Total Systems, Total Dependability and Software Cybernetics," (unpublished manuscript), Centre for Software Reliability, City University, London, Draft version, October 1994; revised version, July 1995.
- [5] K.Y.Cai, "Optimal software testing and adaptive software testing in the context of software cybernetics," Information & Software Technology 44.14 (2002): 841-855.
- [6] K.Y.Cai, T.Y.Chen, T.H.Tse, "Towards Research on Software Cybernetics," Proc. 7th IEEE International Symposium on High Assurance Systems Engineering (2002): 240-241.
- [7] H. Yang, F. Chen, S. Aliyu, "Modern Software Cybernetics: Trends with New Cybernetics," The Journal of Systems & Software (2016), doi:10.1016/j.jss.2016.05.044.
- [8] Definition of educational data mining [DB/OL]. <http://www.educationaldatamining.org/>, 2016-12-11.
- [9] Educational software [DB/OL]. https://en.wikipedia.org/wiki/Educational_software, 2016-12-11.
- [10] N. Wiener, "Cybernetics: or Control and Communication in the Animal and the Machine," Boston, MA: Technology Press, 1948.
- [11] K. Y. Cai, J. W. Cangussu, R. A. DeCarlo, and A. P. Mathur, "An overview of software cybernetics," Proc. 11th International Workshop on Software Technology and Engineering Practice, IEEE Computer Society Press, 2003, pp. 77-86.
- [12] Illeris, Knud, ed, "Contemporary theories of learning: learning theorists in their own words," Routledge, 2009.
- [13] Alexander W. Astin, Anthony Lising Antonio, "Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education," Published by Rowman & Littlefield Publishers, 2012.
- [14] Alexander W. Astin, "Student Involvement: A Development Theory for Higher Education," Journal of college student development, 1999, vol. 40(5), pp. 518-529.