Computers in Human Behavior 41 (2014) 486–503

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Process mining analysis of conceptual modeling behavior of novices – empirical study using JMermaid modeling and experimental logging environment

Gayane Sedrakyan [a,*], Monique Snoeck [b,1], Jochen De Weerdt [c,2]

[a] K.U. Leuven, Faculty of Business and Economics, Department of Decision Sciences and Information Management, Research Center for Management Informatics (LIRIS), Naamsestraat 69, B-3000 Leuven, Office number: HOG 03.114, Belgium

[b] K.U. Leuven, Faculty of Business and Economics, Department of Decision Sciences and Information Management, Research Center for Management Informatics (LIRIS), Naamsestraat 69, B-3000 Leuven, Office number: HOG 03.118, Belgium

[c] K.U. Leuven, Faculty of Business and Economics, Department of Decision Sciences and Information Management, Research Center for Management Informatics (LIRIS), Naamsestraat 69, B-3000 Leuven, Office number: HOG 03.122, Belgium

ARTICLE INFO

Article history:
Available online 20 November 2014

Keywords:
Teaching/learning conceptual modeling
Process-oriented feedback
Conceptual modeling pattern
Information systems education
Process mining
Learning data analytics

ABSTRACT

Previous studies on learning challenges in the field of modeling focus on cognitive perspectives, such as model understanding, modeling language knowledge and perceptual properties of graphical notation by novice business analysts as major sources affecting ***model quality***. In the educational context ***outcome feedback*** is usually applied to improve learning achievements. However, not many research publications have been written observing the characteristics of a ***modeling process*** itself that can be associated with better/worse learning outcomes, nor have any empirically validated results been reported on the observations of modeling activities in the educational context. This paper attempts to cover this gap for conceptual modeling. We analyze modeling behavior (conceptual modeling event data of 20 cases, 10.000 events in total) using experimental logging functionality of the JMermaid modeling tool and process mining techniques. The outcomes of the work include ***modeling patterns*** that are indicative for worse/better learning performance. The results contribute to (1) improving teaching guidance for conceptual modeling targeted at ***process-oriented feedback***, (2) providing recommendations on the type of data that can be useful in observing a modeling behavior from the perspective of learning outcomes. In addition, the study provides first insights for learning analytics research in the domain of conceptual modeling.

1. Introduction

Empirical studies show that more than half the errors that occur during systems development are requirements errors (Endres & Dieter Rombach, 2003; Lauesen & Vinter, 2001). Requirements errors are also the most common cause of failure of development projects (Moody, 2005; Schenk, Vitalari, & Davis, 1998). The success of the analysis of requirements depends heavily on models. Formalization of requirements through ***models*** enables quality control at a level that is impossible to reach with requirements articulated in natural language (Sikora, Bastian, & Pohl, 2011). With the growing importance of compliance between business strategy and ICT realizations as well as the emergence and evolution of Model Driven Engineering (MDE), ***conceptual modeling*** gains more relevance. Teaching conceptual modeling skills is however a challenging task. In their early careers novice modelers produce incomplete, inaccurate, ambiguous, and/or incorrect models (Schenk et al., 1998). Errors occurring early in the systems analysis process are much more expensive and time-consuming to resolve when only detected later in the engineering process than those that may occur at any other time in systems engineering (Schenk et al., 1998). Studies on *learning quality improvements* indicate a self-regulative approach as major source of impact on learning outcomes which in turn is closely intertwined with ***feedback*** research (Nicol & Macfarlane-Dick, 2006; Zimmerman, 2008), i.e. for all self-regulative activities, external feedback is considered as an inherent catalyst (Barber, Bagsby, Grawitch, & Buerck, 2011; Butler & Winne, 1995). As proposed by the constructivist approach (Hadjerrouit, 2005) the method of ***dialogue*** is the most optimal way to address learning difficulties by delivering personalized

* Corresponding author. Tel.: +32 (0)16 32 68 82.
E-mail addresses: Gayane.Sedrakyan@kuleuven.be (G. Sedrakyan), Monique.Snoeck@kuleuven.be (M. Snoeck), Jochen.DeWeerdt@kuleuven.be (J. De Weerdt).
[1] Tel.: +32 (0)16 32 68 79.
[2] Tel.: +32 (0)16 32 62 68.

http://dx.doi.org/10.1016/j.chb.2014.09.054
0747-5632/

feedback. Usually feedback is not available during modeling activities but is given after a task has been completed. In feedback literature this is referred to as ***outcome feedback***, the simplest form of feedback, indicating whether or not results are correct, thus providing minimal external guidance (Butler & Winne, 1995). Several researchers highlighted the effectiveness of more informative types of feedback paired with content-related information that guide the process of cognitive activities (Butler & Winne, 1995; Narciss, 2008; Shute, 2008). Studying the process of conceptual modeling of novice analysts might provide insights on what type of feedback would be effective in ***guiding a modeling process*** by determining the characteristics of a process of conceptual modeling that have a positive impact on conceptual model quality. Within this study we will therefore focus on revealing the aspects of the modeling process that might affect the quality of a model, and subsequently will formulate our research questions as: "(1) *Is it possible to identify patterns of a modeling process that can be associated with better/worse learning outcomes ?* (2) *What type of data is relevant to support the identification of such patterns ?*".

In order to answer the research questions, we opted for an empirical/experimental approach. Data on modeling activities of novice modelers (86 students in total) have been collected by means of experimental logging functionality of the JMermaid[3] modeling environment. Students' group works over one semester period of time were observed. For data analysis we opted for process mining techniques motivated by the fact that process mining techniques process mining has built a reputation of being capable of analyzing rich data trails and activity streams in various contexts (De Weerdt, Schupp, Vanderloock, & Baesens, 2013). In addition, process mining diagrams make it easier to visually extract useful information and quantify relevant properties on process-oriented modeling approaches. We further elaborate the findings with quantitative and qualitative analysis.

While findings showed that certain behavioral patterns can indeed be associated with better/worse outcomes in terms of reaching a satisfactory model quality, further examinations are needed to identify more generic patterns. The results of the study can be used to provide recommendations on ***process-oriented feedback***. This study presents first insights to support research on learning analytics (e.g. type of data needed) as well as artificial intelligence (e.g. automation of feedback) in the domain of conceptual modeling. In addition, this study can be inspirational for the application of process-oriented learning analytics outside of the topic of conceptual modeling, as learning event data is becoming more readily available through digital learning systems and other educational information systems.

The remainder of the paper is structured as follows: Section 2 describes the educational context and assumptions used within this paper. Section 3 gives an overview of related work and the research contribution. Section 4 describes the research method followed by Section 5 that describes the data analysis and subsequently reports on the results. Section 6 discusses the contributions and limitations of the work. Finally, Section 7 concludes the work proposing some future research directions.

2. Educational context and assumptions

To facilitate further reading of this paper, information on the educational context as well as some basic concepts used throughout the paper will be briefly discussed.

A *conceptual model* (also known as domain model) is a complete and holistic view of a system based on conceptual but precise qualitative assumptions about its concepts "entities" and their interrelationships (Embley & Thalheim, 2012). A conceptual model of an information system is defined as an "abstract model" of an enterprise and conceptual modeling in information systems development as the creation of an enterprise model for the purpose of designing the information system (Wand, Monarchi, Parsons, & Woo, 1995). A model is often represented visually as a diagram, by the use of a modeling language. In this paper the modeling language used is UML (Unified Modeling Language) motivated by the fact that UML is the widely accepted standard used for modeling systems throughout software engineering processes. A UML class diagram is the main structural diagramming approach widely used to visually represent an information system's components and relationships (Szlenk, 2006) that are used both in high level conceptual modeling as well as in more detail for lower level design and documentation of programming code (Berardi, Calvanese, & De Giacomo, 2005; Marshall, 2000; Szlenk, 2006). There are several UML diagramming approaches to capture the dynamic view of a system. Within this study we make use of UML statecharts.

The JMermaid tool used in this work is a conceptual modeling environment that has been developed by the Management Informatics research group at the faculty of Business and Economics, University of Leuven. It uses the UML as modeling language, but underneath it relies on the concepts of MERODE[4], an Enterprise Information Systems engineering methodology developed at KU Leuven (Snoeck, 2014). MERODE uses a limited subset of UML relevant for conceptual modeling that allows removing or hiding details irrelevant for a conceptual modeling view. The framework is based on three kinds of model views: restricted class diagrams called existence dependency graph (EDG), finite state machine (FMS) and an interaction model to combine the structural and behavioral view in a single model, called object event table (OET). To ensure inter and intra model consistency, the tool makes use of built-in intelligence such as automatic checks, as well as a "consistency by construction" (Snoeck, Michiels, & Dedene, 2003) approach that completes missing model elements automatically. This makes the approach easy to use in an educational context. The tool has been subsequently expanded with an experimental logging functionality to collect data on modeling activities.

In order to measure the effects of the modeling process on learning outcomes we need to distinguish between worse/better models. In this work we will refer to the quality dimensions of the Conceptual Modeling Quality Framework (CMQF) (Nelson, Poels, Genero, & Piattini, 2012) which is rooted in the seminal framework of Lindland and Sindre (Lindland, Sindre, & Solvberg, 1994) and presents a unified view of conceptual model quality. Within this framework, teaching conceptual modeling involves different types of modeling quality. The final objective is to achieve the capability of producing physical models with high external quality. External model validity – also called *semantic quality* – refers to the level to which the statements in a model reflect the real world in a valid and complete way (feasible completeness, feasible validity) (Lindland et al., 1994). Within this work we will thus focus on the modeling activities that can potentially affect the semantic quality of a conceptual model.

Since it would not be possible to actually track a human mind in our experiment, the concept of ***modeling effort*** was used throughout the paper to refer to the prevailing number of specific modeling activities as approximation of modeler's mental effort.

3. Related work

Prior studies on improving model quality have been focusing on the cognitive perspective of the modeling process, model

[3] http://merode.econ.kuleuven.ac.be/mermaid.aspx.

[4] MERODE is an Object Oriented Enterprise Modeling method. Its name is the abbreviation of Model driven, Existence dependency Relation, Object oriented DEvelopment. Cfr. http://merode.econ.kuleuven.be.