Knowledge-Based Systems 97 (2016) 11–23

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Data calibration for statistical-based assessment in constraint-based tutors

Jaime Gálvez[a], Eduardo Guzmán[a,*], Ricardo Conejo[a], Antonija Mitrovic[b], Moffat Mathews[b]

[a] E.T.S.I. Informática, Dpto. Lenguajes y Ciencias de la Computación, Universidad de Málaga, Bulevar Louis Pasteur, 35, Campus de Teatinos, 29071 Málaga, Spain

[b] Department of Computer Science and Software Engineering, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

ARTICLE INFO

Article history:
Received 28 September 2015
Revised 14 January 2016
Accepted 18 January 2016
Available online 25 January 2016

Keywords:
Learning analytics
Assessment
Constraint-Based Modeling
Intelligent Tutoring Systems
Item Response Theory

ABSTRACT

Intelligent Tutoring Systems (ITSs) are one of a wide range of learning environments, where the main activity is problem solving. One of the most successful approaches for implementing ITSs is Constraint-Based Modeling (CBM). Constraint-based tutors have been successfully used as drill-and-practice environments for learning. More recently CBM tutors have been complemented with a model derived from the field of Psychometrics. The goal of this synergy is to provide CBM tutors with a data-driven and sound mechanism of assessment, which mainly consists in applying the principles of Item Response Theory (IRT). The result of this synergy is, therefore, a formal approach that allows carrying out assessments of performance on problem solving tasks. Several previous studies were conducted proving the validity and utility of this combined approach with small groups of students, in short periods of time and using systems designed specifically for assessment purposes. In this paper, the approach has been extended and adapted to deal with a large set of students who used an ITS over a long period of time. The main research questions addressed in this paper are: (1) Which IRT models are more suitable to be used in a constrained-based tutor? (2) Can data collected from the ITS be used as a source for calibrating the constraints characteristic curves? (3) Which is the best strategy to assemble data for calibration? To answer these questions, we have analyzed three years of data from SQL-Tutor.

1. Introduction

Intelligent Tutoring Systems (ITSs) are probably the most well-known product of the Artificial Intelligence in Education (AIED) research community. ITSs are environments that help student learn a subject matter. To do that, they use a knowledge base that is comprised of a student model and a domain model, modeling what the student knows and what to teach, respectively. The teaching process of an ITS consists of consulting the knowledge base and adapting the content and tutorial actions according to the student model. This behavior tries to mimic an expert human teacher who adapts the process to every individual student. Perhaps the most extended interaction pattern an ITS provides is an environment where students can solve problems belonging to certain domain matter. According to Jonassen [18], "*most educators agree that problem solving is among the most meaningful and important kinds of learning and thinking*". A problem exists when a problem solver has a goal but does not know how to reach it. Problem solving is a mental activity aimed at finding a solution to a certain problem [3]. The challenge of solving a problem forces students to build models through a process of understanding, exploring and interacting with the world, developing several branches of science at all levels of education [30]. Thus, problem solving entails cognitive processing with the goal of transforming a given situation into a desired scenario when no obvious method of solution is available to the problem solver [21]. According to Mayer [22] problem solving expertise can be decomposed into four components:

1 *Problem translation*, where the student transforms the problem stem into an internal mental representation.
2 *Problem integration*, a mental model of the situation described in the problem stem is constructed.
3 *Solution planning*, where the strategy to solve the problem is determined, i.e. the steps to take in order to solve the problem. This component requires the student to apply his/her procedural knowledge.
4 *Solution execution*, that is, the previous plan is applied to solve the problem.

Constraint-Based Modeling (CBM) [39] is one of the most popular approaches for developing ITSs [8,43]. Its effectiveness as an

* Corresponding author. Tel.: +34 952137146.
E-mail address: guzman@lcc.uma.es (E. Guzmán).

http://dx.doi.org/10.1016/j.knosys.2016.01.024
0950-7051/

Nomenclature

ITS:	Intelligent Tutoring System
AIED:	Artificial Intelligence in Education
CBM:	Constraint-Based Modeling
IRT:	Item Response Theory
ECD:	Evidence Centered Design
ICC:	Item Characteristic Curve
BN:	Bayesian Network
CCC:	Constraint Characteristic Curve

instructional methodology has been proved in a range of tutors and studies performed over 15 years [33,35,37,38]. A characteristic that makes it a very attractive approach is its ability to be applied in a tutoring system easily since it does not require a complex architecture. Furthermore, it does not require identifying all possible steps a student could take to reach a solution to a problem. Instead, it only requires the identification of domain principles (represented as constraints) that no solution should violate.

Educational assessment characterizes aspects of student knowledge, skill, abilities, or other attributes. For this characterization it makes inferences from the observation of what they say, do, or make in certain kinds of situations [5]. Furthermore, educational assessment provides at least three different uses in instructional improvement [3]: first, results obtained through assessment motivate students and educational staff to achieve the academic goals set by policy makers. In addition, it represents a way of helping teachers to plan or revise their pedagogical strategies. Finally, assessment can be used to help stimulate deep understanding. The use of computers in testing is extensive nowadays. In the area of problem solving, however, there is still an enormous range of opportunities to explore [3,52]. Problem solving activities require students to apply their knowledge in constructing a solution to a certain situation [23]. One of the most recognized assessment techniques is Item Response Theory (IRT), which gave rise to a set of different models with different assumptions (see next section).

In our previous work [14,15] we made a first proposal of a model of assessment combining CBM with IRT. This proposal can also be seen as an implementation of the Evidence Centered Design (ECD) framework [1,29,41], which focuses on providing a generic methodology to perform assessments of problem solving. This synergy between the AIED and psychometric mechanisms opens the door to enhancing ITSs with new methods to perform automatic assessment of tasks that, if carried out by a human expert, would be highly difficult and prone to subjectivity. As will be explained later, the utilization of IRT makes it possible to apply new formal psychometric methods in CBM that were not possible before. In the same way, some of the fundamentals of CBM extend the typical use of IRT in testing environments, where theoretical concepts are assessed, to ITS, which requires applying practical knowledge to solve a problem.

Initially, in order to explore the validity of the approach for assessment purposes, two educational systems were developed and tested with undergraduate students of the University of Malaga in Spain [13–15]. Although the knowledge base of these ITSs was developed in well-defined domains, according to the classification made by Mitrovic and Weerasinghe [36], the tasks involved were completely different. In the first system, focused on the Simplex algorithm for mathematical optimization, the number of constraints was small and the tasks were well-defined (i.e. those tasks for which the process of solving them is known). On the other hand, the second system, focused on teaching fundamentals of Object Oriented Programming, had a relatively large number of constraints and the tasks were ill defined with a complex solution procedure (having more than one solution or many ways to achieve it).

Initial results obtained using CBM and IRT showed that the methodology was feasible and promising in these types of domains. Nevertheless, the experiments were carried out in systems constructed for assessment purposes, with a small group of students, using a particular IRT model and strictly following the restrictions imposed by the IRT to guarantee valid assessment results under this theory. To the contrary, the most successful CBM-based systems have been used mainly for learning purposes in drill-and-practice environments. That means that a student is allowed to solve the same problems several times which leads to the violation of the IRT models assumed hypotheses (i.e. student knowledge is constant during a session). This difference makes it necessary to explore the scalability and validity of the existing models based on the combination of IRT with CBM in tutoring systems used for learning purposes and with a large number of students.

The research carried out in this paper tries to cover the aforementioned problems by extending the existing methodology (explained in detail in the following sections) and performing a study with a larger dataset obtained over three years of use of the SQL-Tutor [34]. The aims of the study are: (1) to define an appropriate methodology to accommodate IRT models to constraint-based tutors; (2) to determine the most appropriate IRT models in this case; and (3) to explore different strategies for grouping and filtering existing ITS data to be used for the IRT calibration process. The advantages of using this approach are that it provides a data-driven technique that does not require heuristic knowledge. The resulting ITS would be adjusted by standard statistical calibration procedures that are not biased with the designer subjectivity.

The paper is structured as follows: Section 2 presents the theoretical background needed to understand both the model and the calibration strategies presented in this paper. Section 3 describes the related work in the field of AIED. Section 4 is devoted to a formalization of our assessment model and a generalization of that model to be used for ITS under the Evidence-Centered Design framework; it also outlines the drawbacks of the early approach. Section 5 proposes a new methodology to overcome the limitations of our proposal with several strategies that can be performed in the process of calibration. Section 6 describes the experiments and the methodological issues and Section 7 presents and discusses the results. Finally, conclusions are summarized in Section 8.

2. Theoretical background

The approach for assessment in ITSs is based on two main pillars, corresponding to the two methodologies already mentioned: CBM for modeling the ITS domain, and the IRT for assessing the student's knowledge in terms of the evidence provided by him/her while solving problems. Both techniques are summarized here. Moreover, the system used in this paper, i.e. SQL-Tutor, is also described briefly.

2.1. Constraint-Based Modeling

The first element of the methodology is the CBM paradigm for building ITSs, which will be the instrument through which students' evidence is gathered. CBM is based on Ohlsson's theory of learning from performance errors [39,40], according to which incomplete or incorrect student's knowledge can be used within an ITS to provide guidance. This faulty knowledge is detected using constraints, which are the key element of CBM. Constraints are principles that must be followed by all correct solutions in the given instructional domain. If the student's solution violates any constraints, it is incorrect and the system provides the student with the appropriate feedback for remediation. Each constraint