Modeling Engagement Dynamics in Spelling Learning

Gian-Marco Baschera¹, Alberto Giovanni Busetto¹,², Severin Klingler¹, Joachim M. Buhmann¹,², and Markus Gross¹

¹ Department of Computer Science, ETH Zürich
² Competence Center for Systems Physiology and Metabolic Diseases, Zürich, Switzerland
{gianba,busettoa,kseverin,jbuhmann,grossm}@inf.ethz.ch

Abstract. In this paper, we introduce a model of engagement dynamics in spelling learning. The model relates input behavior to learning, and explains the dynamics of engagement states. By systematically incorporating domain knowledge in the preprocessing of the extracted input behavior, the predictive power of the features is significantly increased. The model structure is the dynamic Bayesian network inferred from student input data: an extensive dataset with more than 150,000 complete inputs recorded through a training software for spelling. By quantitatively relating input behavior and learning, our model enables a prediction of focused and receptive states, as well as of forgetting.

Keywords: engagement modeling, feature processing, domain knowledge, dynamic Bayesian network, learning, spelling.

1 Introduction

Due to its recognized relevance in learning, affective modeling is receiving increasing attention. There are two reasons why modeling affective dynamics is considered a particularly challenging task. First, ground truth is invariably approximated. Second, experimental readouts and state emissions often exhibit partial observability and significant noise levels. This paper entertains the idea that intelligent tutoring systems can adapt the training to individual students based on data-driven identification of engagement states from student inputs.

Problem Definition. The goal of this study consists of modeling engagement dynamics in spelling learning with software tutoring. In our scenario, student input data and controller-induced interventions are recorded by the training software. Input behavior is assumed to be time- and subject-dependent.

Related Work. Affective models can be inferred from several sources: sensor data [12] and input data [3,4,6]. These sources differ in quality and quantity. On the one hand, sensor measurements tend to be more direct and comprehensive. They have the potential to directly measure larger numbers of affective features.
On the other hand, input measurements are not limited to laboratory experimentation. The measurement of student interaction with a software tutoring system offers a unique opportunity: large and well-organized sample sets can be obtained from a variety of experimental conditions. Recorded inputs have the potential to characterize the affective state of the student in a learning scenario. It has been shown that highly informative features, such as seconds per problem, hints per problem, and time between attempts, can be extracted from log files [6]. The identification of informative features and the incorporation of domain knowledge, either as implicit or as explicit assumptions, can substantially increment the predictive power of the inferred models [5]. Median splitting [6], thresholding [4], and input averaging [3] are conventional preprocessing techniques in affective modeling.

**Contributions.** We introduce a model which relates input behavior to learning, and explains the dynamics of engagement states in spelling training. We show how domain knowledge about dynamics of engagement can be incorporated systematically in the preprocessing of extracted input behavior to significantly increase their predictive power. The dynamic Bayesian network (DBN) is inferred from user input data recorded through a training software for spelling. Focused and receptive states are identified on the basis of input and error behaviors alone.

## 2 Methods

Our approach is articulated in four steps: (1) description of training process; (2) specification of extracted features; (3) feature processing based on domain knowledge; (4) feature selection and model building.

**Learning Environment.** The tutoring system consists of Dybuster, a multimodal spelling software for children with dyslexia [8]. During training, words are prompted orally and have to be typed in via keyboard by the student. As soon as incorrect letters are typed, an acoustic signal notifies the error. The system allows prompt corrections, which prevent the user from memorizing the erroneous input. Every user interaction is time-stamped and stored in log files.

Our analysis is based on the input data of a large-scale study in 2006 [9]. The log files span a time interval of several months, which permits the analysis of multiple time scales: from seconds to months. The German-speaking participants, aged 9-to-11, trained for a period of three months and with a frequency of four times a week, during sessions of 15-to-20 minutes. On average, each user performed approximately 950 minutes of interactive training. The training predominantly took place at home, except once per week, when the children attended a supervised session at our laboratory to ensure the correct use of the system. Due to technical challenges, a subset of 54 log files were completely and correctly recorded (28 dyslexic and 26 control). This dataset records 159,699 entered words, together with inputs, errors, and respective timestamps.