Modeling Engagement Dynamics in Spelling Learning

Gian-Marco Baschera[1], Alberto Giovanni Busetto[1,2], Severin Klingler[1], Joachim M. Buhmann[1,2], and Markus Gross[1]

[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 [1, 2] 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.

Feature Extraction We identified a set of recorded features which are consistent with previous work [3, 4, 6]. Table 1 lists the features, which are evaluated for each word entered by the learner. The set contains measures of input and error behavior, timing, and variations of the learning setting induced by the system controller.

Engagement states are inferred from the repetition behavior of committed errors and without external direct assessments. We subscribe to the validated hypothesis of interplay between human learning and affective dynamics [7]. Committed errors and the knowledge state at subsequent spelling requests of the same word are jointly analyzed. Error repetition acts as a noisy indicator for learning and forgetting. We restrict the analysis on phoneme-grapheme matching (PGM) errors [12], which is an error category representing missing knowledge in spelling, in contrast to, e.g., typos. We extracted 14 892 observations of PGM errors with recorded word repetitions from the log files.

Feature Processing The processing of continuous features is based upon the following central assumptions: emotional and motivational states come in spurts [4], and they affect the observed features on a short-to-medium time scale. Time scale separation enables a distinction between sustainable progress in the observed input behavior ($f(i)$) and other local effects ($p(x_i)$), such as the

Table 1. Extracted features and abbreviations (bold) used in the following.

Feature	**Description**
Timing	
Input **R**ate	Number of keystrokes per second.
Input **R**ate **V**ariance	Variance of seconds per keystroke.
Think **T**ime	Time from dictation of word to first input letter of student.
Time **f**or **E**rror	Time from last correct input letter to erroneous input letter.
Time **t**o **N**otice **E**rror	Time from error input letter to first corrective action.
Off **T**ime	Longest time period between two subsequent letter inputs.
Input & Error Behavior	
Help **C**alls	Number of help calls (repeating the dictation).
Finished **C**orrectly	True if all errors are corrected when enter key is pressed.
Same **P**osition **E**rror	True if multiple errors occur at one letter position of a word.
Repetition **E**rror	State of previous input of the same word (three states: *Correct / Erroneous / Not Observed).*
Error **F**requency	Relative entropy [10] from observed to expected error distribution (given by the student model [12]) over last five inputs. Positive values are obtained from larger errors numbers, negative values from smaller ones.
Controller Induced	
Time **t**o **R**epetition	Time from erroneous input to respective word repetition.
Letters **t**o **R**epetition	Number of entered letters from erroneous input to respective word repetition.

influence of engagement states. The terms are separated as

$$t(x_i) = f(i) + p(x_i), \tag{1}$$

with independent additive normal $p(x_i) \sim \mathcal{N}(0, \sigma^2)$. The transformation $t(\cdot)$ of the original feature x_i consists of scaling and outlier detection. The separation of long-term variation $f(i)$ depends on the temporal input position i in the student input history. The finally obtained additive terms $p(x_i)$ are referred to as processed feature. Table 2 lists the employed processor modules. Whereas scaling and outlier detection operate point-wise on the individual words, regression subtraction is time- and user-dependent. The selection of processing steps and corresponding coefficients for each feature are the result of a downhill simplex optimization of the differential entropy (with fixed variance) [13, 11], resulting in a distribution of $p(x_i)$ with maximal normality. Figure 1 illustrates the processing of the Time for Error (TfE) feature. The low-pass and variance filters, listed in Table 2, allow for a separation of low frequency components from rapid fluctuations of the processed features and are tested in the feature selection step.

Feature Selection and Model Building The relation between processed features $p(x_i)$ and error repetition γ_r is estimated via LASSO logistic regression [11] with 10-fold cross-validation for different filter and filter parameters. The regression parameters are denoted by b_i. Figure 2 illustrates the comparison between Error Repetition Probability (ERP) predictions obtained from unprocessed and processed features. The model based on processed features exhibits a better BIC score ($-6\,369$) compared to unprocessed regression ($-6\,742$). In the selected features (see Table 3), we identified three main effects influencing the knowledge state at the next repetition:

Table 2. Employed feature processing modules and abbreviations (bold).

Module	**Operation on feature** x	**Parameters**
Scaling		
Logarithmic	$\log(s + x)$	s
Exponential	$\exp(-\frac{a+x}{b})$	a, b
Splitting	$\mathrm{I}_{x>s}$	s
Outlier detection		
Deviation **C**ut	$\min(\mu + \sigma, \max(\mu - \sigma, x))$ $\quad \mu = \mathrm{mean}(x)$	σ
Regression subtraction		
Learning **C**urve	$x_i - f(i)$ $\quad f(i) = a\exp(-bi) + c$	a, b, c
Filtering		
Low-Pass	$x_i = \sum_{j=0}^{n} x_{i-j} G(j, n)$ [1]	n
Variance	$x_i = \mathrm{var}([x_{i-n}, ..., x_i])$	n

[1] $G(j, n)$ corresponds to the sampled Gaussian kernel $G(j, n) = \frac{1}{\sqrt{2\pi n}} e^{-\frac{j^2}{2n}}$.

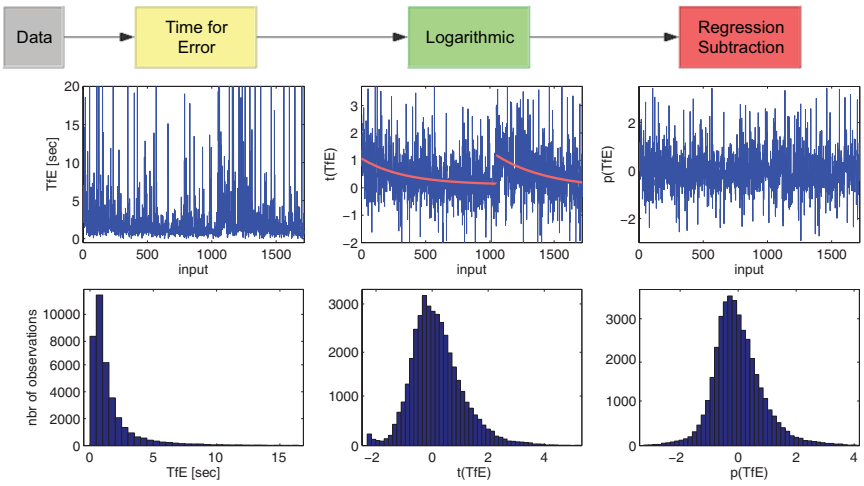

Fig. 1. Top line exemplifies the processing pipeline for the TfE feature. On the second row, the signal of the processing steps is plotted for the data recorded from two learners: extracted feature (left), transformation (center), and separation (right). The third row shows the respective histogram plots for the data of all 54 students.

Focused state indicates focused or distracted state of the student. In non-focused state more non-serious errors due to lapse of concentration occur, which are less likely to be committed again at the next repetition (lower ERP).

Receptive state indicates the receptiveness of the student (receptive state or beyond attention span). Non-receptive state inhibits learning and causes a higher ERP.

Forgetting: the time (decay) and number of inputs (interference) between error and repetition induce forgetting of learned spelling and increase the ERP.

The parameters of the logistic regression indicate how features are related to the ERP. We inferred the affiliation of features to engagement states based on

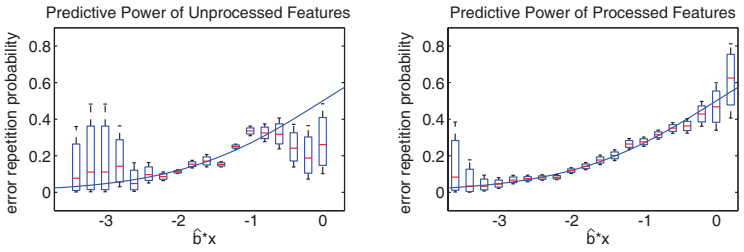

Fig. 2. ERP prediction (10-fold cross-validation) from unprocessed (left) and processed features (right). Predictions are plotted as blue curve and accompanied by mean (red stroke), 68% (box), and 95% confidence intervals (whisker) of the observed repetitions for bins containing at least 10 observations.

the relations extracted from the regression analysis and expert knowledge about desired input behavior. For example, the parameter $b = 0.06$ of EF demonstrates that a higher than expected error frequency is related to a lower ERP. This indicates that a student is non-focused and commits more but rather non-serious errors. On contrary, if a student does not finish an input correctly ($FC = 0$), the ERP increases ($b = -0.49$). This indicates that students, which are not correcting their spelling errors, are less likely to pick up the correct spelling.

In the following we investigate the mutual dependence of the two engagement states, which are considered as dynamic nodes. We compared three models: (1) based on a mutual independence assumption ($F \nleftrightarrow R$); (2) with dependence of focused state on receptivity ($F \leftarrow R$); (3) with dependence of receptivity on focused state ($F \rightarrow R$). The parameters of the DBN are estimated based on the expectation maximization (EM) algorithm implemented in Murphy's Bayes net toolbox [9]. The mutual dependence of the engagement states is inferred based on the estimated model evidence (BIC).

3 Results

Figure 3 presents the graphical model ($F \rightarrow R$) best representing the data with a BIC of $-718\,577$, compared to $-724\,111$ ($F \nleftrightarrow R$) and $-718\,654$ ($F \leftarrow R$). The relation between the Focused and Receptive state is illustrated by their joint probability distribution in Figure 4 (left). In a fully focused state, students are

Table 3. Optimal processing pipeline, estimated parameter b and significance for features selected by the LASSO logistic regression. Note that the exponential scaling inverts the orientation of a feature. The last two columns show the influence of the engagement states on the features modeled in the DBN: for binary nodes the probability p_1 of being *true*; for Gaussian nodes the estimated mean m of the distribution.

Feature	**Processing Pipeline**	**b**	**sig.**	$\mathbf{p_1}$[%]/**m**	
Focused State				focused	non-f.
EF	Exp	0.06	2e-4	0.16	-0.34
IR	Log - DevC - LearnC - Var	-0.12	4e-6	-0.41	0.87
IRV	Log - DevC - LearnC	-0.22	2e-11	-0.36	0.78
REc		-0.28	8e-8	45%	32%
TfE	Log - DevC - LearnC - LowP	-0.50	1e-9	-0.13	0.28
Receptive State				receptive	non-r.
FC		-0.49	1e-7	95%	88%
HC	Split(zero/non-zero)	0.29	2e-4	4%	28%
OT	Log - DevC - LearnC - LowP	0.27	1e-9	-0.35	1.20
REe	LowP	0.20	1e-9	0.07	-0.24
TtNE	Exp - DevC - LearnC	-0.18	1e-5	0.11	-0.36
Forgetting					
TtR	Exp	-0.29	2e-8		
LtR	Log	0.34	1e-9		

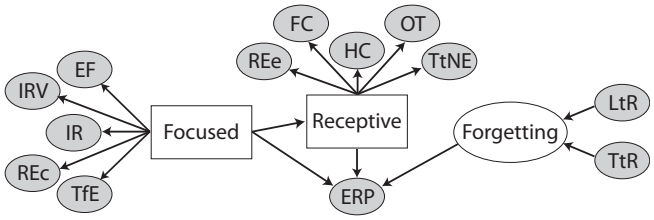

Fig. 3. The selected dynamic Bayesian net representation. Rectangle nodes denote dynamic states. Shaded nodes are observed.

never found completely non-receptive. In contrast, students can be distracted (non-focused) despite being in a receptive state.

The ERP conditioned on the two states is presented in Figure 4 (right). One can observe that the offset between top plane (forgetting) and bottom plane (no forgetting) is greater in the focused compared to the non-focused state. This underpins the assumption that in the non-focused state more non-serious errors are committed, of which the correct spelling is actually already known by the student. Therefore, the forgetting has a lower impact on their ERP. As expected, the non-receptive state generally causes a higher ERP. Again, this effect on learning is reduced for non-serious errors in the non-focused state. The estimated parameters of the conditional probability distributions for all the other observed nodes are presented in Table 3 (right).

The investigation of the age-dependence of engagement states shows that students below the median of 10.34 years exhibit a significantly ($p < 0.001$) higher probability of being classified as non-receptive (24.2%) and non-focused (32.5%) compared to those above the median (20.0% and 27.0%, respectively). This indicates that younger students tend to fall significantly more frequently into non-focused and non-receptive states.

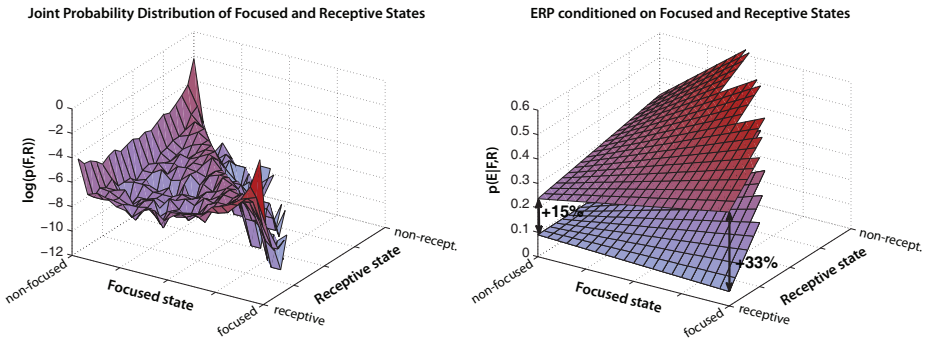

Fig. 4. Left: joint probability distribution of Focused and Receptive states. Right: ERP conditioned on engagement states for forgetting (top) and no forgetting (bottom plane). The ERP is plotted for all observed combinations of engagement states only.

4 Conclusion

We presented a model of engagement dynamics in spelling learning. We showed that domain knowledge can be systematically incorporated into data preprocessing to increase predictive power. In particular, the regression analysis demonstrates the advantages of feature processing for engagement modeling. Our approach enables the identification of the dynamic Bayesian network model directly from spelling software logs. The model jointly represents the influences of focused and receptive states on learning, as well as the decay of spelling knowledge due to forgetting. This core model can be extended with assessments of engagement of a different nature, such as sensor, camera or questionnaire data. This would allow to relate the identified states to the underlying fundamental affective dimensions (e.g., boredom, flow, confusion and frustration) of a student.

Acknowledgments We would like to thank M. Kast and Kay H. Brodersen for helpful suggestions. This work was funded by the CTI-grant 8970.1.

References

1. Cooper, D.G., Muldner, K., Arroyo, I., Woolf, B.P., Burleson, W.: Ranking Feature Sets for Emotion Models used in Classroom Based Intelligent Tutoring Systems, *UMAP 2010*, pp. 135–146 (2010)
2. Heray, A., Frasson, C.: Predicting Learner Answers Correctness through Brainwaves Assessment and Emotional Dimensions, *AIED 2009*, pp. 49–56 (2009)
3. Baker, R.S., Corbett, A.T., Koedinger, K.R.: Detecting Student Misuse of Intelligent Tutoring Systems, *ITS 2004*, pp. 531–540 (2005)
4. Johns, J., Woolf, B.: A Dynamic Mixture Model to Detect Student Motivation and Proficiency, *UMAP 2006*, pp. 163–168 (2006)
5. Busetto, A.G., Ong, C.S., Buhmann, J.M.: Optimized Expected Information Gain for Nonlinear Dynamical Systems, *ICML 2009*, pp. 97-104 (2009)
6. Arroyo, I., Woolf, B.: Inferring Learning and Attitudes from a Bayesian Network of Log File Data, *AIED 2005*, pp. 33–40 (2005)
7. Kort, B., Reilly, R., Picard, R.W.: An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy - Building a Learning Companion, *Advanced Learning Technologies*, pp. 43–46 (2001)
8. Gross, M., Vgeli, C.: A Multimedia Framework for Effective Language Training. *Computer & Graphics*, 31, pp. 761–777 (2007)
9. Kast, M., Meyer, M., Vögeli, C., Gross, M., Jäncke, L.: Computer-based Multisensory Learning in Children with Developmental Dyslexia, *Restorative Neurology and Neuroscience*, 25(3-4), pp. 355–369 (2007)
10. Kullback, S., Leibler, R.A.: On Information and Sufficiency, Annals of Mathematical Statistics, 22(1), pp. 79–86 (1951)
11. Bishop, C.: Pattern Recognition and Machine Learning, Springer (2006)
12. Baschera, G.M., Gross, M.: Poisson-Based Inference for Perturbation Models in Adaptive Spelling Training, *International Journal of AIED*, 20(1), in press (2010)
13. Nelder, J.A., Mead, R.: A Simplex Method for Function Minimization, *Computer Journal*, 7, pp. 308–313 (1965)
14. Murphy, K.: The Bayes Net Toolbox for Matlab. *Computing Science and Statistics*, 33, (2001).