# In the Zone: Towards Detecting Student Zoning Out Using Supervised Machine Learning

Joanna Drummond and Diane Litman

Department of Computer Science, Sennott Square,
University of Pittsburgh, Pittsburgh, PA 15260
{jmd73,litman}@cs.pitt.edu

**Abstract.** This paper explores automatically detecting student zoning out while performing a spoken learning task. Standard supervised machine learning techniques were used to create classification models, built on prosodic and lexical features. Our results suggest these features create models that can outperform a Bag of Words baseline.

**Keywords:** Zoning Out, Natural Language, Machine Learning.

## 1 Introduction

Recent investigations suggest detecting and adapting to student affect and other states could improve student learning and other performance measures for intelligent tutoring systems (e.g., [1,2,3,4,5,6]). Current detection methods include measuring response times [3] and using lexical, prosodic and other linguistic features [2,6].

Zoning out, a state defined as "thinking about other things while [performing a learning task]" [7], was shown to negatively impact student learning [7]. Thus, Moss calls for investigating intelligent tutoring systems that adapt to zoning out [7]. While little work has investigated detecting zoning out, detecting disengagement, a closely related phenomena, has been explored (e.g., [3,4]).

Given the promise of language-based affect-adaptive tutors [1,2] and the link between zoning out in a spoken learning task and normalized learning gains [7], we feel that spoken tutoring systems that adapt to students' zoning out have the potential for improving student learning. Therefore, we wish to show it is feasible to build models to automatically detect zoning out.

## 2 Dataset and Features

Our corpus, a subset of Moss's [7], contains novice undergraduate students reading aloud a biology paragraph, then performing a learning task (paraphrase, or self explain) aloud. Students' audio was recorded and human-transcribed, with transcriptions including common spoken disfluencies. At set intervals, the student took a short survey with the text "I found myself zoning out and thinking about other things when reading this text" with a Likert scale underneath, with 1 being "All the time," and 7 being "Not at all." So, we will attempt to classify

V. Aleven, J. Kay, and J. Mostow (Eds.): ITS 2010, Part II, LNCS 6095, pp. 306–308, 2010.

at this granularity. We combine everything the student read in that interval into one **Text**, and what the student produced via the learning task into one **Task**.

Since we wish to show detecting zoning out is possible, we group student self-reports into two categories: "*High*" if the student reports 1-3, and "*Low*" if the student reports 5-7, discarding borderline reports of 4. We have 52 instances of students self-reporting *High*, 63 reporting *Low*, and 20 discarded due to reporting 4. Therefore, we have a total of 115 data points, from 37 students.

We only present features chosen by the feature selection algorithm used in the machine learning experiments. ***Transcript-Based Features*** use our human transcriptions. **WC Text** calculated the number of words the student read, and **WC Diff** counted the difference between **WC Text** and the number of words the student said in their **Task**. We also created wordlists by investigating students' **Task** data. These wordlists generate word-count-based features. **Confusion** wordlist tried to capture when a student acknowledged that they were confused. **References** attempted to indicate when a student personalized the information. **Disfluencies** counted the number of human-annotated spoken disfluencies (filled pauses, unfilled pauses, and false starts) found in the student's **Task**. **Bag of Words** counts the number of times each word in the **Task** vocabulary is said in this data point's **Task**, making each word a unique feature. Our baseline model is built using only this feature.

***Audio-based Features*** are commonly used to classify user states in spoken systems. We used an implementation previously developed for detecting student affect. **Percent Text Silence** is the amount of internal silence divided by the time the student is actively speaking and their internal silence. **Text** and **Task Min Pitch** is the student's minimum pitch in that segment. **Text** and **Task Min Energy** describes loudness instead of pitch.

## 3 Machine Learning Experiments and Results

We present results from one machine learning algorithm, to show this task is feasible. The Bag of Words baseline and two experimental models were built using the J48 Decision Tree algorithm implemented by Weka, which includes a feature selection algorithm. Due to our small dataset, we used the leave-one-out cross-fold validation training/testing paradigm. We chose accuracy, precision and recall as our evaluation metrics. We then tested for differences between our models and the baseline using a two-tailed t-test.

The quantitative performance of our models can be found in Table 1. We evaluate our models using the three metrics, applied to both *High* and *Low*. We have highlighted the best performance for our experimental models in each metric in the table. This table also shows the results of the t-test. The Bag of Words row shows the performance of our baseline. The next row details our first experimental model, built with All designed features, a superset of those presented in Section 2, excluding **Bag of Words**. Qualitatively assessing this model, **Text Min Pitch**, **Disfluencies**, and **References** are the most important features. As **Text Min Pitch** was the root node of All's decision tree, we built the Text Min Pitch model, using only this feature. This model performed best in all metrics except *Low* Recall.

**Table 1.** Leave-One-Out Cross-Fold Validated Performance (N = 115); ∗ Significantly higher than baseline at $p = 0.05$, † at $p = 0.10$

| Model | Accuracy | *High* Precision | *High* Recall | *Low* Precision | *Low* Recall |
|---|---|---|---|---|---|
| Bag of Words | 0.522 | 0.474 | 0.519 | 0.569 | 0.524 |
| All Features | 0.583 | 0.548 | 0.442 | 0.603 | **0.698**∗ |
| Text Min Pitch | **0.643**† | **0.580** | **0.769**∗ | **0.739**∗ | 0.540 |

## 4 Conclusions and Future Work

Our long-term goal is to enhance spoken tutorial systems to detect and adapt to *High* zoning out students. We have shown that even with a small dataset, it is feasible to build a model to detect students' self-reported zoning out that outperforms a Bag of Words baseline. In addition, our automated audio-based features were very important in detecting zoning out, suggesting it's possible to automatically detect student zoning out in a real-time tutorial dialogue system.

To improve our results, we wish to explore different machine learning algorithms and different methods of feature selection. We also wish to explore fully automating all transcript-based features. In addition, we wish to apply our results to detecting disengagement in a spoken physics tutorial dialogue system [1].

**Acknowledgements.** We thank Dr. C. Schunn and Dr. J. Moss for our data, M. Lipschultz and ITSPOKE group for comments, and NSF grant #0631930.

## References

1. Forbes-Riley, K., Litman, D.: A user modeling-based performance analysis of a wizarded uncertainty-adaptive dialogue system corpus. In: Proc. Interspeech, Brighton, UK (September 2009)
2. Pon-Barry, H., Schultz, K., Bratt, E., Clark, B., Peters, S.: Responding to student uncertainty in spoken tutorial dialogue systems. Intl. Journal of AIED (2006)
3. Beck, J.: Using response times to model student disengagement. In: ITS (2004)
4. Cocea, M., Weibelzahl, S.: Log file analysis for disengagement detection in e-Learning environments. User Modeling and User-Adapted Interaction (2009)
5. Lehman, B., Matthews, M., D'Mello, S., Person, N.: What are you feeling? In: Investigating student affective states during expert human tutoring sessions. LNCS (2008)
6. D'Mello, S., Craig, S., Witherspoon, A., Mcdaniel, B., Graesser, A.: Automatic detection of learners affect from conversational cues. User Modeling and User-Adapted Interaction (2008)
7. Moss, J., Schunn, C.D., VanLehn, K., Schneider, W., McNamara, D.S., Jarbo, K.: They Were Trained, But They Did Not All Learn: Individual Differences in Uptake of Learning Strategy Training. In: Proc. of 30th Annual Meeting of the Cognitive Society (2009)