"I Didn't Pass the Exam Because ...": Testing the Viability of Conceptual Features for Actionable Analytics in the Context of Competency Exam Failure Reflection

Fanjie Li[1], Yeonji Jung[2], Alyssa Wise[1]
[1]Vanderbilt University, [2]The University of Memphis
fanjie.li@vanderbilt.edu, yeonji.jung@memphis.edu, alyssa.wise@vanderbilt.edu

ABSTRACT: Taking an actionability-oriented approach, this study explored strategies for constructing predictive analytics for instructors to support students following competency exam failure. Our core approach builds on efforts in explainable modeling by making features in predictive models not only inspectable but also *interpretable* and *actionable*. We provide an example of this in the context of detecting dental students' failure attribution through analytics of their written reflections for failed competency exams. Through human-in-the-loop linguistic modeling, we built conceptual features describing students' perceived causes of failure to inform instructors in providing targeted support. We trained and evaluated a random forest classifier using the conceptual features and compared its performance with the classifier built on baseline n-grams features. LIME-based explanations of both models were generated for human interpretation. Results support the viability of conceptual features both in improving model performance and in enhancing model interpretability.

Keywords: Actionability, human-in-the-loop, conceptual features, explainable modeling, failure attribution.

1 INTRODUCTION

Being able to learn from failure and recover from difficult experiences is important for students' long-term success. Students' failure attribution, in particular whether they see the cause of their failure as within or beyond their control to change, has profound impact on how they perceive and react to the challenge, providing indications of whether the student is likely to recover independently or if external support is needed (Zimmerman & Moylan, 2009). Although students' attributional thinking is not always articulated and available to instructors, analytics of student reflections on failed experiences offers an opportunity for instructors to gain insights into how students see causes of their failure, identify who needs timely support, and recognize the specific areas in which students struggle. Using a corpus of dental students' written reflections on failed competency exams, this study explored strategies for constructing analytics of failure attribution with actionability to support recovery in mind.

2 STRATEGY 1: PRAGMATIC FRAMING OF OUTCOME CLASSES

Although the attribution theory (Weiner, 1985) provides nuanced dimensions for characterizing students' causal attribution (i.e., locus (internal/external), stability, controllability), fine-grained categorizations of failure attribution are difficult for instructors to interpret and act on within their busy daily routines. Zimmerman and Moylan (2009) have underscored the importance of the *controllability* dimension of attribution in self-regulated learning, suggesting that attributing failures to uncontrollable factors (e.g., external barriers, lack of ability) often discourages students' efforts for further improvement, whereas learners who attribute failures to controllable factors (e.g., use of strategies) are more likely to sustain motivation during setbacks and engage in further self-regulation. The alignment between self-regulated learning theory and dental education's focus on developing future dentists' ability for self-assessment and self-improvement (Driessen et al., 2005) motivated us to adopt an actionable framing of the outcome class (attribution for failure) that can directly inform instructors' decision-making as to *which students to support*: (1) **External/no attribution**: Students did not make attributions or attribute their failure to external factors only *(need timely support)*; (2) **Internal attribution:** Students demonstrate self-evaluation of knowledge, effort, etc. but do not show clear intent or path to change *(need timely support)*; (3) **Internal, controllable attribution**: Besides reflecting on internal reasons, students also demonstrate reflections on strategy or plan/intention to change. Students of this type are more likely to self-regulate (*less need for support*).

3 STRATEGY 2: CONSTRUCTING CONCEPTUAL FEATURES

To further support instructor in identifying the *kind of support these students need*, we examined whether conceptual features describing students' perceived causes of failure can be constructed via human-in-the-loop linguistic modeling. Informed by our prior thematic analysis of specific reasons for failure, we extracted noun phrases and specific forms of verb phrases (e.g., negation + verb such as "not study") that often capture linguistic cues of causal attribution. These phrases were subsequently input into ChatGPT 3.5 to assist in finding commonly used phrases to describe each reason for failure. The prompts to ChatGPT took the general form of "From the list of phrases below, please find all phrases that can be used to describe [REASON]". Based on ChatGPT's responses, we summarized linguistic patterns for detecting whether students mentioned a reason in their reflections. These include: (1) **Exam difficulty or delivery** (e.g., presence of hard/difficult/tough); (2) **Course design** (e.g., "clinical experience" that describe clinical exposure in curriculum); (3) **School activities** (e.g., "scheduling conflict" that describe hectic school schedules); (4) **Luck** (the part-of-speech tag "*existential there"* was used to indicate descriptions of external conditions given that sentence subject has been found to indicate writer's locus of control (Rouhizadeh et al., 2018)); (5) **Lack of knowledge** (e.g., presence of understand, grasp, clear about), (6) **Ability** (e.g., presence of (un)able/(in)ability to); (7) **Mistakes** (e.g., presence of verb starting with "mis"); (8) **Efforts to prepare** (e.g., presence of study/prepare/review); (9) **Learning/exam strategies** (e.g., time management).

In addition to perceived reasons for failure, we also constructed linguistic patterns for detecting **plan/intention to change**. Specifically, we used the presence of "*need to/more/further*", "*will*", and *comparatives* (e.g., more) to capture students' use of future-focus language, and used "*should/could have('ve)* [e.g., paid more attention]" to capture students' use of intention-intensive language.

4 MODEL PERFORMANCE AND LIME-BASED MODEL EXPLANATIONS

We trained a random forest (RF) classifier using the conceptual features and a second baseline RF classifier for comparison using n-grams features. For both classifiers, we performed a five-fold cross-validation on the training set for hyperparameter tuning and evaluated its performance on a 10% hold-out test set. The conceptual feature model achieved better classification performance (AUC: 0.83, Kappa: 0.49, precision: 0.68, recall: 0.67) than the baseline n-grams model (AUC: 0.79, Kappa: 0.47, precision: 0.66, recall: 0.64). We generated and compared LIME-based model explanations of three test-set reflections randomly selected from each class. Figure 1 presents model explanations for baseline and conceptual feature models for one of these instances (external/no attribution class) as an illustration due to space limitations.

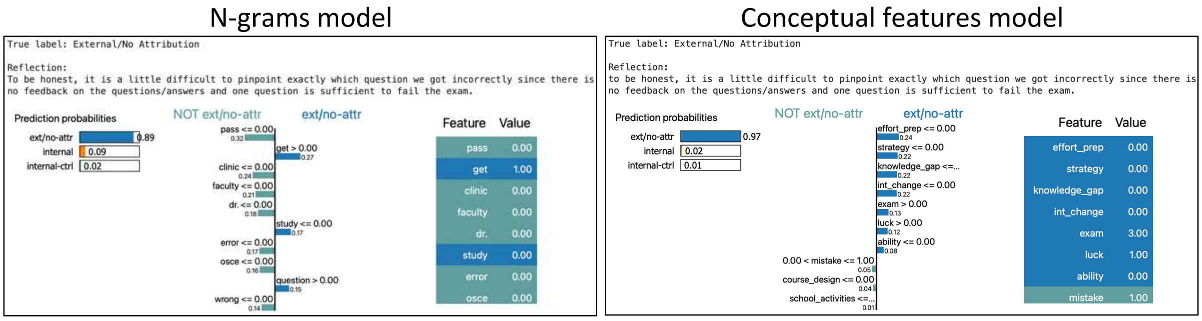

Figure 1: LIME explanations for the instance randomly selected from external/no attribution class

5 DISCUSSION

The results support the viability of conceptual features both in improving model performance and in enhancing model interpretability. Inspecting the model predictions for an instance from the *external/no attribution* class (Figure 1), both conceptual feature model and n-grams model correctly classified this instance. As aligned with our coding protocol, the reason that the conceptual feature model made the resulting prediction is that the extracted features do not indicate students' *intention to change* or attention to internal reasons (e.g., *lack of knowledge*, *efforts to prepare*, *strategy*, *ability*) (features <= 0) but does point to their reference to *exam difficulty* and *luck* (features > 0). Differential importance of such features across students (e.g. lack of knowledge versus efforts to prepare) is promising as a tool to point instructors towards different modes of support. It is noteworthy that the presence of the part-of-speech tag *"existential there"* (the matched pattern underlying the *luck* feature) rightly contributed to the prediction of the external/no attribution class. However, referring to the actual text of this reflection, *existential there* was actually used to describe the unavailability of feedback. This suggests that the *"luck"* feature may need to be redefined or split to avoid confusing instructors. For future work, we will further refine the conceptual features (e.g., the naming issue identified from the above observation) and evaluate the interpretability and actionability of the conceptual features with dental instructors and advisors.

REFERENCES

Driessen, E. W., Van Tartwijk, J., Overeem, K., Vermunt, J. D., & Van Der Vleuten, C. P. (2005). Conditions for successful reflective use of portfolios in undergraduate medical education. *Medical Education*, 39(12), 1230–1235.

Rouhizadeh, M., Jaidka, K., Smith, L., Schwartz, H. A., Buffone, A., & Ungar, L. (2018). Identifying locus of control in social media language. In *Proceedings of EMNLP 2018* (pp. 1146–1152).

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In *Handbook of metacognition in education* (pp. 299–315). Routledge.