Contribution ID: **221**   Type: **Poster**

Using machine learning on multiple true-false item texts to predict the difficulty of best single-answer items: Identifying domain-specific text features beyond readability

Thursday 24 July 2025 17:10 (20 minutes)

Poster

Using machine learning on multiple true-false item texts to predict the difficulty of best single-answer items: Identifying domain-specific text features beyond readability

Author

Lubomír Štěpánek, Čestmír Štuka, Martin Vejražka, Patrícia Martinková

Keywords

item_difficulty_estimation, domain-specific_difficulty-related_textual_and_contextual_features, machine_learning_in_assessment, natural_language_processing, multiple_true-false_to_best_single-answer_items_transformation

Abstract

Accurately estimating item difficulty is crucial for designing fair and effective assessments, particularly in high-stakes exams such as medical faculty admissions. This study investigates subject-specific textual elements that significantly influence item difficulty beyond traditional readability features and explores the predictive potential of machine learning algorithms in estimating the difficulty of best single-answer items derived from multiple true-false items. Using historical admission test data from the First Faculty of Medicine, Charles University in Prague, we employ pre-calibrated difficulty estimates of multiple true-false items to predict the difficulty of their reformulated best single-answer counterparts.

Our approach goes beyond traditional textual features related to readability, such as word counts, vocabulary frequency, lexical similarity, and readability indices (Štěpánek, Dlouhá, & Martinková, 2023). Instead, we aim to leverage domain-specific contextual elements within item wording – particularly in subjects like physics, chemistry, and biology – that influence difficulty. These contextual and semantic elements include conceptual and knowledge representation features (such as domain-specific taxonomy or terminology abstractness), semantic embedding and contextual features (such as algorithm-estimated text complexity using large language models), syntactic and structural complexity (including text mode, sentiment density, and diction analyzed using language models), cognitive and conceptual load features (e.g., missing or aberrant information in the item wording), and domain-specific features (such as chemical or mathematical notation, formulas, or figures), among others. By adopting this approach, we seek to uncover key linguistic or conceptual patterns in item wording that strongly impact difficulty levels. Machine learning techniques are applied to identify these domain-specific difficulty-related textual and contextual features. The dataset includes multiple years of admission test responses, allowing us to match item wordings with test-takers'performance and apply the Rasch model for difficulty estimation. By comparing pre-calibrated multiple true-false item difficulties with the predicted and observed difficulties of their best single-answer versions, we evaluate the effectiveness of our approach in predicting difficulty shifts caused by item reformulation.

The findings of this study may contribute to the field of educational assessment by demonstrating how machine learning can enhance difficulty estimation, particularly when transitioning between item formats. The extracted textual features provide insights into the linguistic and cognitive factors influencing item difficulty, which can inform test construction and item design in high-stakes assessments.

References:

L. Štěpánek, J. Dlouhá, and P. Martinková, "Item difficulty prediction using item text features: Comparison of predictive performance across machine-learning algorithms", Mathematics, vol. 11, no. 19, p. 4104, Sep. 2023, issn: 2227-7390. doi: 10.3390/math11194104. [Online]. Available: http://dx.doi.org/10.3390/math11194104.

Primary author: Dr ŠTĚPÁNEK, Lubomír (Institute of Computer Science of the Czech Academy of Sciences; First Faculty of Medicine, Charles University)

Co-authors: Dr VEJRAŽKA, Martin (First Faculty of Medicine, Charles University);   MARTINKOVÁ, Patrícia (Institute of Computer Science of the Czech Academy of Sciences; Faculty of Education, Charles University);   Dr ŠTUKA, Čestmír (First Faculty of Medicine, Charles University)

Session Classification: Poster Session 4

Track Classification: Statistical analyses: Statistical analyses