STROHMAIER, Anselm R.; MORA-RUANO, Julio G.; SCHONS, Christian & OBERSTEINER, Andreas
München, Bamberg

Can a Machine Learning Algorithm Tell Right from Wrong in Eye Movements of Mathematical Word Problem Solving?

Theoretical Background

In mathematical word problem solving, processes of reading and mathematical thinking interact. Because information processing during reading is closely associated with eye movements, the method of eye tracking has been used in the past to investigate word problem solving processes (Strohmaier et al., 2020). Previous studies showed that many aspects of the solution process, for example, solution strategies, the effect of inconsistent operational terms, differences in the mental representation or item difficulty are reflected in students' eye movements (e.g., Strohmaier et al., 2019).

The analysis of eye movement data is challenging, in particular for two reasons: First, the same parameters often have different, nonlinear, and possibly opposite interpretations. For example, longer fixation times on relevant areas can both mean higher processing effort or more targeted attention (Strohmaier et al., 2020). Second, hundreds of different parameters can be extracted from eye tracking data, and it is challenging for researchers to determine which parameters are relevant in their particular research context. This is problematic both for theoretical and statistical reasons, as it leads to the risk of focusing on isolated or incidental findings, or missing complex underlying interactions.

Both these issues can be addressed by the use of Machine Learning (ML) algorithms. ML is particularly useful in handling datasets including large numbers of interacting predictors with nonlinear effects. ML algorithms can be distinguished in supervised and unsupervised learning: In unsupervised learning, an ML algorithm exploratively identifies patterns in a dataset. This approach has been used to analyze eye movements in mathematics education by Schindler et al. (2020; 2022) to identify counting strategies in primary school children. In supervised learning, a ML algorithm is trained with examples for a given classification, tries to identify patterns in these data and then applies these "learned" patterns to classify an unknown dataset. Supervised learning has been used to analyze eye movements in other areas, for example, in physics education (Küchemann et al., 2021; Rebello et al., 2018). However, the approach has rarely been used in mathematics education (Schindler et al., 2019). There are several methods for supervised ML, for example, Support Vector Machines (Küchemann et al., 2021), Neural

In: P. Ebers, F. Rösken, B. Barzel, A. Büchter, F. Schacht & P. Scherer (Hrsg.),
Beiträge zum Mathematikunterricht 2024.
57. Jahrestagung der Gesellschaft für Didaktik der Mathematik. WTM.
https://doi.org/10.37626/GA9783959872782.0

Networks, Decision Trees, or Random Forests. The choice of an appropriate algorithm depends on both the data and the classification problem.

In sum, prior studies showed that eye movements are associated with the process of word problem solving, and that ML algorithms are effective in handling eye tracking data. In the present study, we exploratively tested how accurately an ML algorithm can predict whether a word problem is solved correctly or incorrectly based solely on eye movement parameters.

Method

We used a dataset of 649 word problem solutions by 42 undergraduate students. The data were collected on a Tobii Spectrum eye tracker with a sampling rate of 600Hz. 9-point calibration was done with an accuracy below 0.5° and a precision below 0.1°. Stimuli were three-line word problems on percentages, combinatorics, exponential growth, and fractions. We extracted total of 115 meaningful eye tracking parameters, including average and total fixation times, saccade data, and pupil data, distinguishing between different areas of interest (e.g., numbers, text, relevant information) and times of interest (e.g., initial reading, subsequent reading). Prior to model fitting, data were preprocessed using k-nearest-neighbors imputation, Yeo-Johnson transformation, a 3 *SD* outlier trim and the exclusion of highly correlated parameters. For the remaining parameters, recursive feature elimination was run to identify which parameters contributed to the model accuracy.

Model fitting was done using a Random Forest Classification. Random Forest was chosen due to its robustness, computational efficiency and its ability to handle large sets of nonlinear predictors (Breiman, 2001). The final model consisting of 500 trees was trained with 80 parameters in a cross-validation procedure, where 90% of the dataset were used for training and 10% for validation. This procedure was done for 10 different partitions and repeated 20 times. All analyses were conducted in R using the packages *caret* (Kuhn, 2008) and *randomForest* (Liaw & Wiener, 2002).

Results

The trained model performed with an accuracy of .68 in correctly identifying the true, observed solutions ($\kappa = .21$, precision = .90, recall = .70, F1 = .79; Figure 1). The cross-table shows that the algorithm performed rather well on accurately classifying observed correct answers (90%), but only classified 28% of observed wrong answers accurately. Figure 2 illustrates how the estimated probability of a correct answer was distributed across the observed correct and wrong solutions. Figure 3 shows a partial dependence plot of the six most relevant predictors and their association with the probability for a classification as correct solution. It shows a nonlinear relation with the classification for some predictors, such that the predicted solution probability quickly increased or decreased after a certain cutoff value.

Discussion

Our results show that the ML algorithm had a limited accuracy in classifying whether a mathematical word problem is solved correctly or incorrectly based solely on eye movements. Although the algorithm did classify a majority of cases correctly, the uncertainty remained very high. This underscores previous research which argues that while a relation between eye movements and solution processes exists, their use for predictions or classifications is more challenging (Strohmaier et al., 2019). In this particular case, the question whether a word problem is solved correctly or not might have various reasons, ranging from the process of decoding information to calculation. Possibly, these processes are too complex and diverse to be captured solely through eye movements, even by a flexible and powerful ML algorithm. Still, our analyses provided meaningful new information, including

	Observation	
Prediction	Wrong answer	Correct answer
Wrong answer	64	42
Correct answer	163	380

Figure 1. Cross table with predicted and observed word problem solutions.

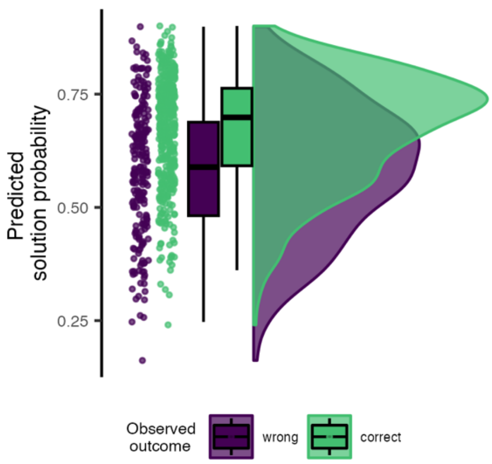

Figure 2. Predicted probability of correct solutions by observed answer

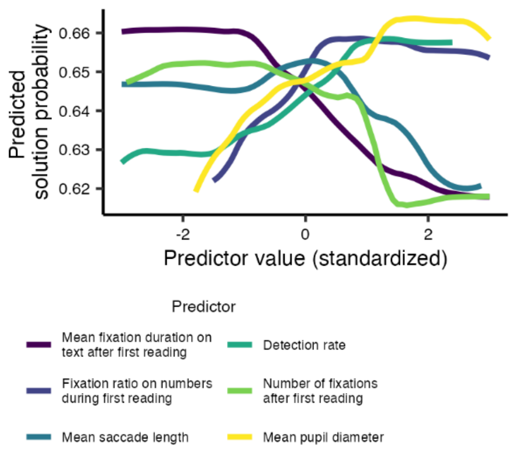

Figure 3. Partial dependency plot for the six most important predictors

which parameters were related to successful solutions despite possible non-linear and interacting effects, which are difficult to detect with correlational analyses. Overall, ML seems a promising tool to handle the complex relational structure and dimensionality of the data. In future research, we aim at addressing more well-defined cognitive processes like specific solution strategies, which might hold a higher potential for successful classification than word problem solving success.

References

Breiman, L. (2001). Random Forests. *Machine Learning 45*, 5–32.

Küchemann, S., Becker, S., Klein, P., & Kuhn, J. (2021). Gaze-Based Prediction of Students' Understanding of Physics Line-Graphs: An Eye-Tracking-Data Based Machine-Learning Approach. *Computer Supported Education, 12th International Conference, CSEDU 2020*, Virtual Event.

Kuhn, Max (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software, 28*(5), 1–26.

Liaw A, Wiener M (2002). Classification and Regression by randomForest. *R News, 2*(3), 18-22.

Rebello, N., Nguyen, M., Wang, Y., Zu, T., Hutson, J., & Loschky, L. (2018). Machine learning predicts responses to conceptual tasks using eye movements. *Physics Education Research Conference 2018, Washington, DC.*

Schindler, M., Doderer, J., Simon, A., Schaffernicht, E., Lilienthal, A., & Schäfer, K. (2022). Small number enumeration processes of deaf or hard-of-hearing students: A study using eye tracking and artificial intelligence. *Frontiers in Psychology, 13*, 909775.

Schindler, M., Schaffernicht, E., & Lilienthal, A. (2020). Identifying student strategies through eye tracking and unsupervised learning: The case of quantity recognition. In Inprasitha, M, Changsri, N., & Boonsena, N. (Eds.). *Proceedings of the 44th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 9-16). Khon Kaen, Thailand: PME.

Schindler, M., Schaffernicht, E., & Lilienthal, A. J. (2019). Differences in Quantity Recognition Between Students with and without Mathematical Difficulties Analyzed Through Eye: Analysis Through Eye-Tracking and AI. In M. Graven, H. Venkat, A. A. Essien, & P. Vale (Eds.), *Proceedings of 43rd Annual Meeting of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 337-344). Pretoria: PME.

Strohmaier, A. R., Lehner, M. C., Beitlich, J. T., & Reiss, K. M. (2019). Eye movements during mathematical word problem solving - Global measures and individual differences. *Journal für Mathematik-Didaktik, 40*, 255-287.

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye Tracking Methodology in Mathematics Education Research: A Systematic Literature Review. *Educational Studies in Mathematics, 104*, 147-200.