A Dynamic Tree-Based Item Response Model for Visual World Eye-tracking Data

Sarah Brown-Schmidt[1], Matthew Naveiras[1], Paul De Boeck[2], Sun-Joo Cho[1] (1-Vanderbilt University; 2-The Ohio State University; KU Leuven, Leuven, Belgium)

In complex scenes, eye gaze is probabilistically directed to different fixation locations, with the likelihood of a fixation to any particular location driven by several competing or complementary cognitive processes. In cases where gaze is in service of performing a task, one of the locations can be considered a task-relevant “target” location (e.g., an object that a person will select), a “competitor” may be similar to the target on some dimension, resulting in potential confusion, and other locations may be "unrelated" to the target and less likely to receive visual attention. We expect that multinomial processing will guide the likelihood of fixating different types of object categories, with one cognitive process increasing the likelihood of fixations to the target and competitor, and a separate process that selects the target and rules out the competitor.

Analysis of binary time-series data considers visual attention to a single interest area, whereas polytomous (e.g., target, competitor, other) time-series data considers visual attention given to several competing options that may be associated with different cognitive processes. The motivation for the present work is a research question for which multiple cognitive processes are assumed to differentially map onto one or more competing response options.

A dynamic generalized linear mixed effect model (GLMM) provides a flexible framework for modeling the heterogeneity and dependencies in observations and allowing the inclusion of trend and serial autocorrelations in intensive binary time series data. Here we present a dynamic tree-based item response (IRTree) model as a novel extension[1] of the dynamic GLMM[2]. Unlike a dynamic GLMM, a dynamic IRTree model is capable of modeling differentiated processes indicated by intensive polytomous time series eye-tracking data. We illustrate a dynamic IRTree model using visual world eye-tracking data. A simulation study resulted in satisfactory parameter recovery and showed that the omission of trend and autocorrelation effects can result in biased estimates and standard errors of experimental condition effects.

We apply the dynamic IRTree model to an empirical dataset[3]. The motivating example concerns listeners’ interpretation of instructions, e.g., “Click on the small elephant” in scenes containing seven objects, including a small elephant (the Target, T), a small envelope (the Competitor, C), and five other Unrelated objects (Fig1-2). It is assumed that the likelihood of fixating the three object categories is guided by multinomial processing (Fig3): lexico-semantic processing narrows the set of candidate referents to T and C (e.g., small elephant and small envelope). Then, ambiguity resolution processes narrow down the search space, picking out the T (small elephant) over C (small envelope) in one of the experimental conditions. Lexico-semantic information concerns the meaning of words, and in this data set this information differentiates T&C vs. U. Ambiguity between T&C can be resolved using different sources of information, including the speaker's perspective. To model these multinomial processes, we use a nested design with nested contrasts. The first node in the tree distinguishes objects that match the lexico-semantic information in the unfolding expression vs. those that do not (e.g., small elephant & small envelope vs. everything else). Among the items that match the unfolding expression, the second node in the tree distinguishes the target object from the competitor object (e.g., small elephant vs. small envelope). The dynamic IRTree approach allows us to disentangle complex relationships among different cognitive processes and different factors of interest. For example, it is possible that a given factor has an effect only on the first node of the tree (lexico-semantic processing), but not on the second node (ambiguity resolution), or vice versa. Separate consideration of the distinct cognitive processes involved is possible by a response tree approach, leading to new, more differentiated findings vs. other approaches.

This new method supports differentiation of hypothesized cognitive processes that guide eye-gaze, and testing of distinct predictions regarding the mechanisms driving each process.

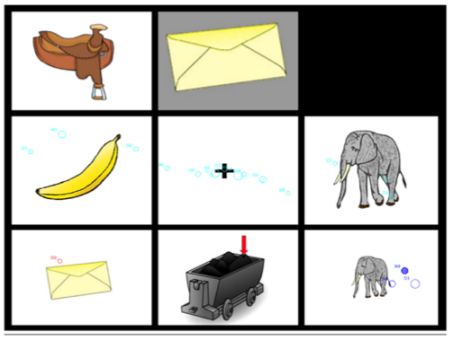

Figure 1. Example display from the empirical dataset[3], featuring images of a saddle, envelopes, elephants, banana, and coal (indicated by red arrow). Display shown from the perspective of one participant (P); their partner viewed a similar scene. Images in white visible to both Ps; images in gray visible to only one P (the other P saw a black box in this spot). Ps received instructions about which images they could both see (shared), and which images only they could see (non-shared); this afforded the critical manipulation of visual perspective. Superimposed on the example display are circles corresponding to individual fixations on one trial (dark blue = target; red = competitor; light blue = unrelated objects).

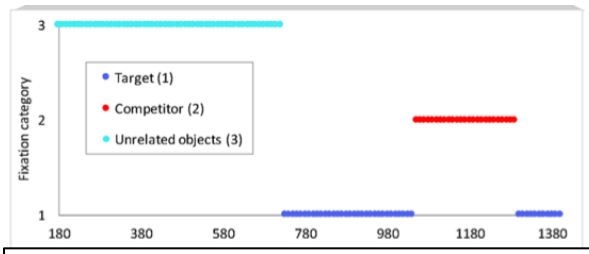

Figure 2. Example gaze data in the time region of interest (180ms after adjective onset in the small elephant) on one example trial, illustrating the polytomous nature of the data with the participant on this trial looking at an "other" unrelated object, then the target, the competitor and back to the target at the very end.

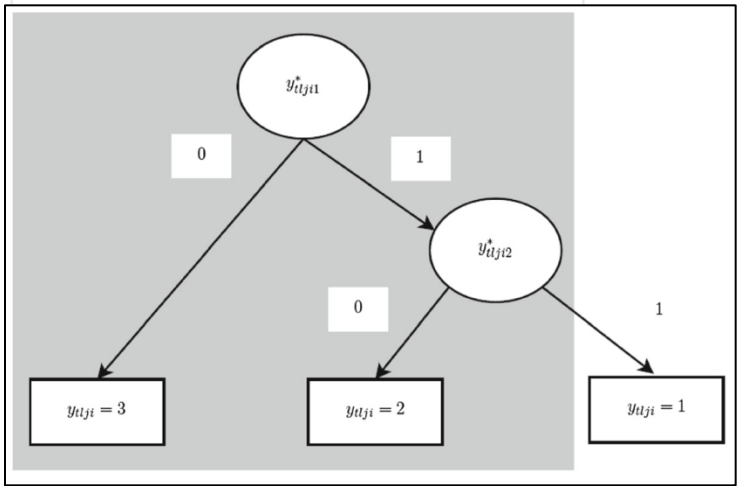

Figure 3. Tree diagram illustrating binary processes (two branches at each node in the tree) at each of two nodes within a three-category paradigm. In the empirical study, Node 1 captures lexico-semantic processing and Node 2 captures ambiguity resolution. Outcome 3 (y_{tlji} = 3) indicates a fixation to U for a particular timepoint (t), trial (l), participant (j), and item (i). Outcome 2 (y_{tlji} = 2) indicates a fixation to C at tlji. Outcome 3 indicates T fixation (y_{tlji} = 1) at tlji. At node 1 (y^*_{tlji1}), fixation to U is coded 0, and fixation to either T or C coded 1. At node 2 (y^*_{tlji2}), fixation to C coded 0, and fixation to T coded 1; at node 2, fixations to U are considered missing at random (MAR[4]).

References

1. Cho, S.-J., Brown-Schmidt, S., De Boeck, P., & Shen, J. (2020). Modeling Intensive Polytomous Time Series Eye Tracking Data: A Dynamic Tree- Based Item Response Model. Psychometrika, 85, 154–184. Supplemental Materials. Tutorial.
2. Cho, S.-J., Brown-Schmidt, S., & Lee, W.-Y. (2018). Autoregressive generalized linear mixed effect models with crossed random effects: an application to intensive binary time-series eye tracking data. Psychometrika, 83, 751-771. Raw data, R code, Model implementation details: https://osf.io/fz9j6/.
3. Ryskin, R. A., Benjamin, A. S., Tullis, J., & Brown-Schmidt, S. (2015). Perspective-taking in comprehension, production, and memory: An individual differences approach. Journal of Experimental Psychology: General, 144(5), 898.
4. Rubin, D. B. (1976). Inference and missing data. Biometrika, 63, 581–592. https://doi.org/10.2307/2335739.