Evaluating Computational Discovery in the Behavioral and Brain Sciences

Sebastian Musslick[1,*], Joshua Hewson[1], Benjamin Andrew[1], Sida Li[2], George Dang[3], John Gerrard Holland[3]

[1] Carney Institute for Brain Science, Brown University, Providence, RI 02906, USA
[2] Department of Statistics, University of Chicago, Chicago, IL 60637, USA
[3] Center for Computation and Visualization, Brown University, Providence, RI 02906, USA
[*] Correspondence: sebastian@musslick.de

Introduction to Discovery Problem. The integration of behavioral phenomena into mechanistic models of brain function is a fundamental staple of the behavioral and brain sciences. Yet, researchers are accumulating increasing amounts of data without having the time or money to integrate these data into scientific theories and/or to test the resulting theories in follow-up experiments (Almaatouq et al., 2022; Oberauer & Lewandowsky, 2019). In other sciences—such as physics, biology, and chemistry—these issues have been partially addressed by integrating artificial intelligence and automation into scientific practice (e.g., King et al., 2009; Lindsay, Buchanan, Feigenbaum, & Lederberg, 1993; Udrescu et al., 2020). Our own efforts along these lines have focused on automating pieces of the empirical research process in psychology and neuroscience, including the discovery of interpretable mechanistic models with differentiable architecture search (Musslick, 2021) and the synthesis of counterbalanced experiment sequences (Musslick et al., 2022). This talk will present a novel framework that relies on computational discovery and closed-loop automation to identify interpretable models of human behavior and brain function. We will focus on how we used this framework to compare different model discovery techniques across different domains of human information processing. Specifically, we will highlight our adaptations of these techniques and discuss their comparative performance in recovering twelve established psychological and neuroscientific models of human information processing from synthetic data.

Formulation of Discovery Problem. Our framework, pieces of which are open-sourced as a pip package[1], consists of three integrated software components: (1) an autonomous theorist that constructs interpretable computational models linking experiment conditions to dependent measures; (2) an autonomous experimentalist that designs novel experiments; and (3) an environment that automates behavioral data collection via online experiments. This talk focuses on the evaluation of three computational discovery algorithms—differentiable architecture search (Liu, Simonyan, & Yang, 2018; Musslick, 2021) and two instances of numeric equation discovery (symbolic regression; Guimerà et al., 2020; Jin, Fu, Kang, Guo, & Guo, 2019)—to recover twelve established models of human information processing from synthetic data. To do so, we formalized recoverable models as computation graphs that take experiment parameters as input (e.g., the brightness or duration of a visual stimulus) and transform this input through a composition of functions to produce observable dependent measures as output (e.g., the probability of detecting said stimulus).

Data and Knowledge Provided to the Discovery Algorithms. To implement our approach, we generated synthetic data from twelve established models of human information processing, which we treated as "ground truth". Our objective was to recover these models from synthetic data using a discovery algorithm. Recoverable models take as input one to six experimental factors (comprising the experimental condition) relevant to the ground truth, characterizing psychological paradigms

[1] AutoRA, Documentation: https://autoresearch.github.io/autora/; GitHub: https://github.com/AutoResearch/autora.

in psychophysics, value-based decision-making, and cognitive control. Model outputs corresponded to dependent measures (e.g., subjective ratings of stimulus intensities or choice probabilities). The discovery algorithms knew about the number of experimental factors and the range of the dependent measure (either a real value or a probability). The space of candidate models included computation graphs with a fixed set of operations, including all operations used in the ground truth models.

Outputs Produced by the Discovery Algorithms. The discovery algorithms yielded interpretable equations in the form of computation graphs that link the independent variables (input nodes of the graph) to a dependent variable (output node of the graph). Intermediate nodes of the computation graphs correspond to latent variables resulting from operations applied to the preceding nodes (e.g., addition or multiplication).

Evaluation of Discovery Algorithms. We evaluated the discovery algorithms based on their ability to recover the ground-truth models from synthetic data. The discovery algorithms used different objective functions to identify candidate models, including mean-squared error, log-likelihood, and minimum description length. Thus, we evaluated candidate models based on all three metrics applied to a hold-out test set generated by the ground truth model. In addition, we compared candidate models against two baseline models—(logistic) regression and a multi-layer perceptron.

Results and Interpretation. We find that Bayesian symbolic regression (Guimerà et al., 2020) outperformed alternative discovery algorithms across a wide range of ground truth models, suggesting that the method is well suited to recover interpretable models of human information processing. Yet, for complex ground truths (e.g., prospect theory), the equations recovered by all algorithms rarely resemble the ground truth model, highlighting the potential for constraining search algorithms with common theoretical constructs (e.g., expected utility).

References

Almaatouq, A., Griffiths, T. L., Suchow, J. W., Whiting, M. E., Evans, J., & Watts, D. J. (2022). Beyond playing 20 questions with nature: Integrative experiment design in the social and behavioral sciences. *Behavioral and Brain Sciences*, 1–55.

Guimerà, R., Reichardt, I., Aguilar-Mogas, A., Massucci, F. A., Miranda, M., Pallarès, J., & Sales-Pardo, M. (2020). A bayesian machine scientist to aid in the solution of challenging scientific problems. *Science Advances*, *6*(5), eaav6971.

Jin, Y., Fu, W., Kang, J., Guo, J., & Guo, J. (2019). Bayesian symbolic regression. *arXiv preprint arXiv:1910.08892*.

King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., Byrne, E., ... others (2009). The automation of science. *Science*, *324*(5923), 85–89.

Lindsay, R. K., Buchanan, B. G., Feigenbaum, E. A., & Lederberg, J. (1993). Dendral: a case study of the first expert system for scientific hypothesis formation. *Artificial intelligence*, *61*(2), 209–261.

Liu, H., Simonyan, K., & Yang, Y. (2018). Darts: Differentiable architecture search. *arXiv preprint arXiv:1806.09055*.

Musslick, S. (2021). Recovering quantitative models of human information processing with differentiable architecture search. In *Proceedings of the 43rd Annual Conference of the Cognitive Science Society* (pp. 348–354). Vienna, AT.

Musslick, S., Cherkaev, A., Draut, B., Butt, A. S., Darragh, P., Srikumar, V., ... Cohen, J. D. (2022). Sweetpea: A standard language for factorial experimental design. *Behavior Research Methods*, *54*(2), 805–829.

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic bulletin & review*, *26*(5), 1596–1618.

Udrescu, S.-M., Tan, A., Feng, J., Neto, O., Wu, T., & Tegmark, M. (2020). Ai feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity. *Advances in Neural Information Processing Systems*, *33*, 4860–4871.