Tutorial on Model Comparison Methods

Jay I. Myung (myung.1@osu.edu)
Department of Psychology, Ohio State University
1835 Neil Avenue, Columbus, OH 43210 USA

Mark A. Pitt (pitt.2@osu.edu)
Department of Psychology, Ohio State University
1835 Neil Avenue, Columbus, OH 43210 USA

Keywords: Cognitive modeling; model evaluation; model selection; model complexity; Occam's razor; Akaike Information Criterion; cross-validation, minimum description length; Bayes factor.

Background

The question of how one should choose among competing explanations of data is at the heart of the scientific enterprise. Computational models of cognition are increasingly being advanced as descriptions and explanations of behavior in the cognitive sciences. The success of this line inquiry depends on the availability of robust quantitative methods to guide the evaluation and selection of these models.

The basic tenet in model comparison may be summarized in terms of the following three quantitative criteria of model evaluation:

> *Goodness of fit:* A good fit (or simulation of observed data) is a necessary, but not a sufficient, condition for judging the adequacy of a model;
>
> *Complexity:* When comparing models, one should avoid choosing an unnecessarily complex model that overfits, and instead, should try to identify a model that is sufficiently complex, but not too complex, to capture the regularity in the data;
>
> *Generalizability:* Model comparison should be based not upon how well a model fits a particular pattern of observed data, but upon generalizability, which refers to how well a model fits not only the observed data at hand but also new, as yet unseen, data samples from the same underlying process that generated the observed data.

The relationship among the three criteria is illustrated in Figure 1. In the figure, goodness of fit and generalizability are represented as curves whose performance can be compared as a function of complexity. The three smaller graphs in the lower panel contain the same data set (dots) and the fits to these data by increasingly more complex models (lines). The left-most model in the figure underfits the data. The data are curvilinear whereas the model is linear. In this case, goodness of fit and generalizability produce similar outcomes because the model is not complex enough to capture the bowed shape of the data. The model in the middle graph is a bit more complex and does a good job of fitting only the regularity in the data. Because of this, the goodness of fit and generalizability measures are higher and also similar. Where the two functions diverge is when the model is more complex than is necessary to capture the main trend. The model in the right-most graph captures the experiment-specific noise, fitting every data point perfectly. Goodness of fit rewards this behavior by yielding an even higher fit score, whereas generalizability does just the opposite, penalizing the model for its excess complexity.

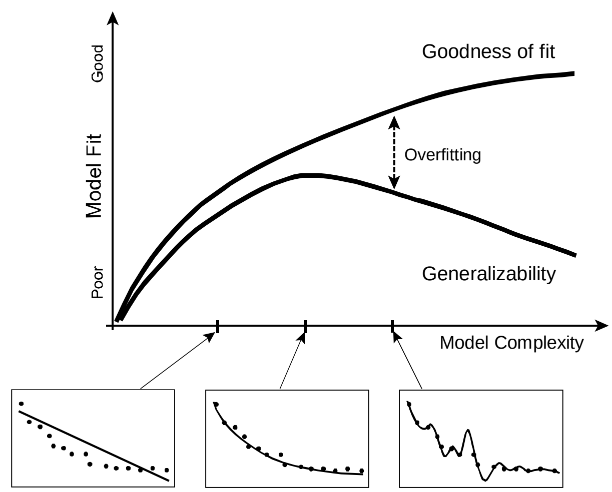

Figure 1. The relationship among goodness of fit, complexity, and generalizability. Reprint from Pitt and Myung (2002).

The foremost goal of model comparison is achieving good generalizability by trading off between two opposing forces, goodness of fit and complexity. Scientists in statistics and computing sciences have proposed a number of model comparison methods that formally implement this trade-off principle.

The purpose of this tutorial is to introduce these methods of model comparison to cognitive scientists who are engaged in computational modeling of cognitive behavior. Our goal in the tutorial is to provide a good conceptual overview of the methods with illustrative examples using selected models in cognitive science. By the end of the tutorial, the attendee should have a grasp of the basic issues in model comparison and an awareness of the tools available to address them.

Target Audience

Graduate students in cognitive sciences who have completed graduate-level statistics series.

Outline of the Tutorial

The topics to be covered in the tutorial lecture include

1. Model evaluation criteria (goodness of fit, complexity, and generalizability) and their relationship; problem of over-fitting; principle of Occam's razor.

2. Model selection methods and their properties: Akaike Information Criterion, Bayesian Information Criterion, cross-validation, accumulative prediction error, Bayes factor, minimum description length.

3. Other tools for model evaluation: Global model analysis by landscaping and parameter space partitioning.

4. Model comparison at work: Illustrative example applications of model evaluation and comparison for quantitative models of memory retention and connectionist models of word reading.

.

Suggested Readings

Gluck, K., Bello, P., & Busemeyer, J. (2008). Special issue on model comparison. *Cognitive Science, 32,* 1245-1424.

Myung, I. J., Forster, M., & Browne, M. W. (2000). Special issue on model selection. *Journal of Mathematical Psychology, 44,* 1-231.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review, 4*, 79-95.

Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review, 113,* 57-83.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6,* 421-425.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109,* 472-491.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science, 32,* 1248-1284.

Wagenmakers, E-J., & Waldorp, L. (2006). Special issue on model selection. *Journal of Mathematical Psychology, 50,* 99-213.

Zucchni, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology, 44,* 41-61.

Acknowledgment

This tutorial is supported by U.S. National Institute of Health grant R01-MH57472 to both authors.

Lecturers

Jay Myung is a Professor of Psychology at the Ohio State University. He received his PhD in 1990 in psychology at Purdue University and spent a year as a postdoc at the University of Virginia before moving to Ohio State as an Assistant Professor. His researcher interests in the fields of cognitive and mathematical psychology include model comparison, design optimization, Bayesian inference, and neural networks. He is currently serving as Editor of *Journal of Mathematical Psychology*. Homepage: http://faculty.psy.ohio-state.edu/myung /personal/.

Mark Pitt is a Professor of Psychology at the Ohio State University. . He received his PhD in 1989 in psychology at Yale University, twiddled his thumbs for a while, and then joined the faculty at OSU. His research interests are in model evaluation, design optimization, and in the field of language, spoken word recognition. He is currently an Associate Editor of *Cognitive Science*. Homepage: http://lpl.psy.ohio-state.edu/.