

Trends and Controversies
Reply
Read reponses

Relevance and insight in experimental studies

Pat Langley, Stanford University

As its name suggests, artificial intelligence is a science of the artificial.(1) As with other conscious creations, there is a great temptation to assume that we can understand the behavior of AI systems entirely through formal analysis. However, the complexity of most AI constructs makes this impractical, forcing us to rely on the same experimental approach that has been so useful in the natural sciences. Many of the same issues and methods apply directly to AI systems, including the need to identify clearly one's dependent and independent variables, the importance of careful experimental design, and the need to average across random variables outside one's control.

However, beyond these obvious features, a compelling experimental study of intelligent behavior must satisfy two additional criteria: it must have *relevance* and it must produce *insight*. I will illustrate these ideas with examples from machine learning, one of the most experimentally oriented subfields within artificial intelligence. Moreover, because AI researchers are often concerned with extending some existing method to improve its behavior, I will focus on this paradigm.

An experimental study of AI methods has relevance if it has implications for problems on which those methods will be used in practice. Researchers can best satisfy this criterion by running their experiments on *natural* domains from the real world. For example, in the machine-learning community, most papers report experimental results on data sets from the repository at the University of California, Irvine, a collection of files that contain data on a variety of natural classification tasks, such as medical diagnosis. Experiments with natural domains are essential because extensions to existing algorithms, although intuitively plausible, often make little difference in practice. Consider the naive Bayesian classifier, a simple learning method that uses training data to estimate the conditional probabilities of attribute values given the class. Because naive Bayes assumes that each attribute is conditionally independent, given the class, it would seem easy to improve it by using more sophisticated methods. However, experiments run by both Igor Kononenko and myself revealed little or no improvement with extensions to naive Bayes on a number of real-world data sets.(2,3) Those studies, although giving negative results, were relevant in that they tested our intuitions on natural domains.

However, experimental studies on natural domains alone do not satisfy our insight criterion. The machine-learning community, in particular, has come to rely almost exclusively on experiments that compare alternative methods on a variety of standard domains (20 or so data sets from the UCI repository), then conclude that one technique or another is superior

because it fares better on most of the domains. Such "bake-offs" tell us very little about the reasons for results, and thus do not provide the understanding about causes that we expect in science.

Researchers can best obtain insight by running experiments on synthetic domains that have been designed to test explicit hypotheses, typically motivated by the intuitions behind the original extension. For example, I have reported experiments on synthetic domains that involve target concepts with disjoint decision regions, which violate another assumption made by naive Bayes.(3) The importance of using synthetic domains is not because they let one generate some new task, but because they let one vary systematically some dimension of interest, and thus test hypotheses about the conditions under which one method will fare better than another. Of course, by themselves, studies with synthetic domains do not ensure relevance; my study found major differences between naive Bayes and its extension on the predicted synthetic domains, but these differences did not carry over to real-world induction tasks.

Thus, truly compelling studies, in machine learning and elsewhere, will include experiments on both natural and synthetic domains, the first to establish relevance and the second to achieve insight. Ideally, they will also *relate* the findings in the two types of study. For instance, if one finds the same shape of results (say when varying some other factor, such as number of training cases) in a synthetic and natural domain, this suggests that the natural domain has similar characteristics to the synthetic one. This strategy lets one move beyond causal accounts in artificial domains toward reasons for success or failure in natural ones, thus giving relevance and understanding at the same time.

Of course, insights about the sources of an algorithm's power are as important as insights about the effects of domain characteristics. Thus, a well-rounded experimental paper will also include lesion studies, which remove algorithm components to determine their contribution, and studies that examine sensitivity to specific parameter settings. Experiments that systematically vary external resources, such as the number of training cases available for learning, should also play a role in any complete empirical study. These recommendations are not new; Dennis Kibler and I proposed these research strategies in an early paper on the experimental study of learning,(4) and Paul Cohen makes a broader case for their use with any intelligent system.(5)

I have drawn my examples from machine learning, but the same basic arguments hold across AI. Research on planning, natural language, diagnosis, perception, and robotics would all benefit from a more balanced mixture of experiments. Although some work along these lines exists, papers in every AI subfield would be more compelling if they included systematic experiments designed with both relevance and insight in mind. I encourage AI researchers to take both of these criteria into account in their experimental evaluations, and thus to speed progress toward the day when our field becomes a true science of the artificial.

References

1. H.A. Simon, *Sciences of the Artificial*, MIT Press, Cambridge, Mass., 1969.
2. I. Kononenko, "Semi-Naive Bayesian Classifier," *Proc. Sixth European Working Session*

on Learning, Pittman, Glasgow, Scotland, 1991, pp. 206-219.

3. P. Langley, "Induction of Recursive Bayesian Classifiers," *Proc. 1993 European Conf. on Machine Learning*, Springer-Verlag, Vienna, 1993, pp. 153-164.

4. D. Kibler and P. Langley, "Machine Learning as an Experimental Science," *Proc. Third European Working Session on Learning*, Pittman, 1988, pp. 81-92. (Reprinted in J.W. Shavlik and T.G. Dietterich, eds., *Readings in Machine Learning*, Morgan Kaufmann, San Francisco, 1990.)

5. P.R. Cohen, *Empirical Methods for Artificial Intelligence*, MIT Press, 1995.



Pat Langley is director of the Institute for the Study of Learning and Expertise and a senior research associate at Stanford University. He received his PhD from Carnegie Mellon University, and he has published widely on the topics of machine learning and scientific discovery, including the recent text *Elements of Machine Learning*. He is an editor of *Machine Learning* and edits the Morgan Kaufmann series on that topic. Contact him at the Institute for the Study of Learning and Expertise, 2164 Staunton Court., Palo Alto, CA 94306; langley@cs.stanford.edu; <http://robotics.stanford.edu/users/langley/bio.html>.

IEEE Expert Online | Departments | Features | AI on the WWW | Editorial Board
Computer Society Publications | Computer Society home page

Send comments and questions about *IEEE Expert Online* to Anne Humphreys and Managing Editor Dick Price, ah34+@andrew.cmu.edu and dprice@computer.org. Send general comments and questions about the IEEE Computer Society's Web site to webmaster@computer.org

Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

Copyright (c) Institute of Electrical and Electronics Engineers, Inc. All rights reserved.