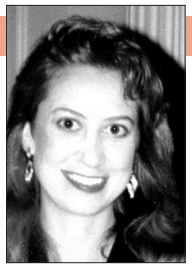

Marti A. Hearst
University of California, Berkeley
hearst@sims.berkeley.edu

Knowledge discovery from data?

Michael J. Pazzani
University of California, Irvine

This issue of *Intelligent Systems*, along with the Nov./Dec. 1999 issue, features articles on the topic of knowledge discovery in data (KDD). In this installment of Trends and Controveries, Michael Pazzani, chair of UC Irvine's Information and Computer Science Department, boldly challenges the KDD community to examine whether or not their work is really living up to its name: helping people discover new knowledge. As a prominent practitioner in the field, he is critiquing his own work as well as that of others. Not content to just point out the problem, Pazzani suggests a new direction: researchers should draw on cognitive psychology for insight about how to create tools to help design knowledge-discovery systems.

—Marti Hearst

The knowledge discovery and data mining (KDD) field draws on findings from statistics, databases, and artificial intelligence to construct tools that let users gain insight from massive data sets. People in business, science, medicine, academia, and government collect such data sets, and several commercial packages now offer general-purpose KDD tools.

An important KDD goal is to "turn data into knowledge." For example, knowledge acquired through such methods on a medical database could be published in a medical journal. Knowledge acquired from analyzing a financial or marketing database could revise business practice and influence a management school's curriculum. In addition, some US laws require reasons for rejecting a loan application, which knowledge from the KDD could provide. Occasionally, however, you must explain the learned decision criteria to a court, as in the recent lawsuit Blue Mountain filed against Microsoft for a mail filter that classified electronic greeting cards as spam mail.[1] In one early KDD success story, Robert Evans and Doug Fisher analyzed data from a printing press, found conditions under which the press failed, and identified rules to avoid these failures.[2]

Unfortunately, for every insightful nugget we find, there are many more obvious or trivial rules (such as "unemployed people don't earn income from work"[3]). Perhaps more troubling is that some rules are counterintuitive. For example, in screening for Alzheimer's disease, we found the following counterintuitive rule: "*If* the years of education of the patient is greater than 5 *and* the patient does not know the date *and* the patient does not know the name of a nearby street, *then* the patient is *normal*."[4]

KDD assumptions

The field of KDD contains far too many assumptions about what system users desire. It is time to question these assumptions and mount a research program that studies these issues. One common assumption is that users like some representations more than others. Unfortunately, there is a conflicting set of claims in the literature as to which format is easier to understand. Most assume that symbolic representations such as trees and rules are more comprehensible than alternatives such as neural networks, nearest-neighbor models, or logistic regression.[1] William Cohen argues for rule-based mail-filtering profiles.[2] Some have reported that simple Bayesian classifiers are more understandable,[3] although others argue for Bayesian networks.[4] Brian Ripley finds more insight in projection pursuit regression than decision trees.[5] Brian Gaines argues that "exception directed acyclic graphs" are more understandable than trees,[6] while Pat Langley argues for condensed determinations.[7] Richard Shiffman has proposed a decision table format for the representation of medical guidelines.[8]

An examination of any of the popular texts reveals that none have chapters devoted to making sure that knowledge is novel, useful, and understandable. While some KDD papers cover these topics, most contain unfounded assumptions about "comprehensibility" or "interestingness." For example, Roberto Bayardo and Rakesh Agrawal have a paper titled, "Mining the Most Interesting Rules."[9] However, on closer examination, the title "Mining Optimized Rules under Partial Orders" would be more appropriate, because the paper presents an impressive, efficient algorithm for searching the space of association rules with a variety of metrics that involve the confidence or support of association rules. The paper does not show that any of these metrics correlates with user judgments of what is interesting.

Aram Karalic's paper, "Producing More Comprehensible Models while Retaining Their Performance"[10] might just as well be titled, "Producing Smaller Models while Retaining Their Performance." It describes the use of the minimum description length principle to learn shorter rules. There has been no study that shows that people find smaller models more comprehensible or that the size of a model is the only factor that affects its comprehensibility.

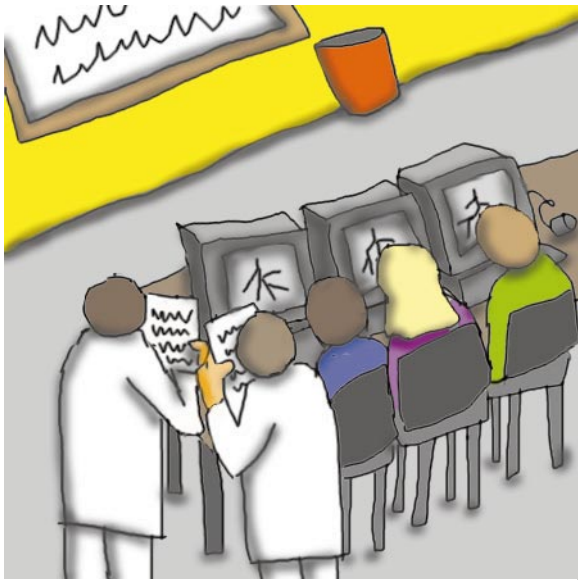

Illustration by Sally Lee

We would usually associate the behaviors in the precondition with an impairment of memory, yet the conclusion was that of normal memory. My experience has been that finding counterintuitive results is not unusual in practice, yet it is rarely mentioned in literature.

Is this knowledge, and is it useful?

KDD papers usually focus on how the authors acquired the knowledge. The few authors who systematically discuss what knowledge they found make comments such as "This rule is puzzling as it would be expected that a positive test for chromosome aberration would be a test for carcinogenesis, not a negative test."[5] John Major and John Mangano examined rules mined from a hurricane data set, and of the 161 rules, they found 10 "genuinely interesting" rules.[6]

Most published papers also concern the development of new algorithms or the challenges of scaling existing algorithms to larger data sets. That is, although KDD is defined as "the process of identifying valid, novel, useful, and understandable patterns in data,"[7] most literature is about validity and process and very little is about novelty, utility, and understandability. Although there are a few papers on these topics in KDD or related conferences, most contain assumptions about comprehensibility or interestingness that have never been tested (see the "KDD assumptions" sidebar). Our only evidence comes from the authors' and collaborators' intuitions. As Brian Gaines reports, "Psychological studies of the nature of comprehensibility of knowledge structures are necessary to give substance to the intuitions."[8]

Meeting user demands

Familiarity with a representation system is at least one determinant of the representation's ease of use. Why else would the US stick with the old imperial measurement system instead of the metric system? Rather than ask which is the better representation, we should ask whether there are visualization tools or variations within a particular representation system that might be more acceptable to users. Several KDD packages offer means to visualize representations,[9] but none have shown that users prefer such visualization tools over textual representations. Furthermore, it's not clear whether 3D visualizations of learned models provide benefits over 2D visualizations.[10]

Another possibility is alternative models within a representation system, offering a particular model that is more acceptable to users. After all, it's rarely the case that there is one decision tree, rule set, linear model, or Bayesian network that jumps out as much better than all alternatives in fitting the data. Perhaps we can use secondary criteria such as comprehensibility and interest to select among alternative models that are statistically indistinguishable.

KDD texts and tutorials stress that it is a process in which someone familiar with data mining interacts with a domain expert. Together, they help determine which problems are interesting and important. We use feedback on whether the domain expert finds the solution acceptable to adjust the data's format (such as by adding or deleting variables) or the learning algorithm's parameters (such as the significance level of the overfitting avoidance method) until we find an acceptable solution. Because our algorithms do not have parameters for

References

1. U.M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery: An Overview," *Advances in Knowledge Discovery and Data Mining*, U.M. Fayyad et al., eds., AAAI/MIT Press, Menlo Park, Calif., 1996, pp. 1–34.
2. W. Cohen, "Learning Rules that Classify E-Mail," *1996 AAAI Spring Symp. Machine Learning in Information Access*, AAAI Press, Menlo Park, Calif., 1996.
3. I. Kononenko, "Comparison of Inductive and Naïve Bayesian Learning Approaches to Automatic Knowledge Acquisition," *Current Trends in Knowledge Acquisition*, B. Wielinga, eds., IOS Press, Amsterdam, 1990.
4. D. Heckerman, "Bayesian Networks for Knowledge Discovery," *Advances in Knowledge Discovery and Data Mining*, U.M. Fayyad et al., eds., AAAI/MIT Press, 1996.
5. R. Ripley, "Statistical Aspects of Neural Networks," *Networks and Chaos Statistical and Probabilistic Aspects*, O.E. Barndorff-Nielsen et al., eds., Chapman & Hall, London, 1996, pp. 40–123.
6. B. Gaines, "Transforming Rules and Trees into Comprehensible Knowledge Structures," *Advances in Knowledge Discovery and Data Mining*, U.M. Fayyad et al., eds., MIT Press, Cambridge, Mass., 1996, pp.205–226.
7. P. Langley, "Induction of Condensed Determinations," *Proc. Second Int'l Conf. Knowledge Discovery & Data Mining*, E. Simoudis, J. Han, and U.M. Fayyad, eds., AAAI Press, Menlo Park, Calif., 1996, pp. 327–330.
8. R. Shiffman, "Representation of Clinical Practice Guidelines in Conventional and Augmented Decision Tables," *J. American Medical Informatics Association*, Vol. 4, No. 5, 1997, pp. 382–391.
9. R. Bayardo and R. Agrawal, "Mining the Most Interesting Rules," *Proc. Fifth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, ACM Press, New York, 1999.
10. A. Karalic, "Producing More Comprehensible Models While Retaining Their Performance," *Information, Statistics and Induction in Science*, Melbourne, Australia, pp. 54–65.

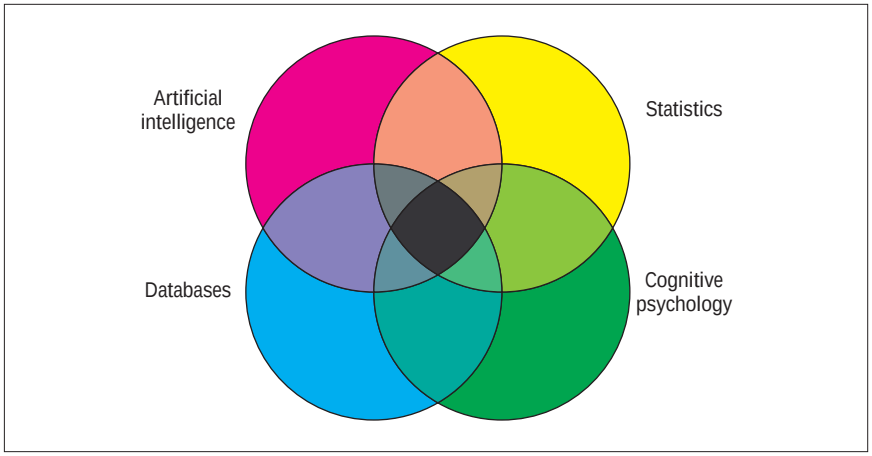

Figure 1. KDD should combine cognitive psychology with artificial intelligence, databases, and statistics to create models that people find insightful.

novelty, utility, and understandability, we must adjust the available parameters with indirect control over these criteria until we satisfy the domain expert. If we knew more about the factors that affect user acceptance of learned models, instead of generating alternatives and testing them for these criteria, we could bias the search toward models that meet these criteria.

The need for cognitive psychology

It's time for KDD to draw on cognitive psychology in addition to databases, statistics, and artificial intelligence. By taking the human cognitive processes into account, we might be able to increase the usefulness of KDD systems. After all, people's perceptions of novelty, utility, and understandability ultimately determine the acceptance of data mining. Figure 1 illustrates how these four fields combine to form KDD. People have been learning representations of the environment for millennia and have been using these learned category models to guide decision-making. Psychological investigation has revealed factors that simplify the learning, understanding, and communication of category information. I am not necessarily suggesting that our KDD systems emulate the way people learn from data. After all, people have difficulty finding subtle patterns in terabytes of shopping data. However, the knowledge KDD systems acquire should be integrated into the knowledge of the cashier, store manager, and marketing director.

Therefore, the biases of human learners are relevant to the acceptance of knowledge acquired through data mining, and KDD systems could benefit from incorporating some of the human learning biases we discuss in the following subsections. There is rich literature on human learning and category representation, and other findings are undoubtedly relevant to KDD. Furthermore, KDD can benefit from using cognitive psychology methodologies by systematically varying aspects of the KDD system and measuring the effects of these manipulations on the acceptability of learned models to users.[11]

Consistency with prior knowledge. A series of findings has shown that people prefer a model consistent with prior knowledge when more than one model is consistent with the data.[12] For example, prior knowledge might tell us that people buy more furniture, draperies, and other household goods after changing residences. However, the data might tell us that expenditures of such goods in a household increases by 160% in the first six weeks after moving. This more specific fact is much more valuable than the general knowledge for marketing purposes. Without such a general piece of knowledge, we might have instead noticed that expenditures in general on these items increase 8% in August and September (because more people move in the summer than other times of year). A marketing director could use the more specific knowledge to recommend a promotion targeted at new homebuyers—the latter knowledge would only support a seasonal sale.

Occasionally you hear anecdotes about finding "nuggets," such as an increase in credit purchases of expensive items, such as Rolex watches, preceded by a credit card gasoline purchase at an automated pump, might be a sign that someone is fraudulently using the card. The explanation is that the gasoline purchase provides an easy and anonymous way of determining if the credit card is valid, and the expensive items can easily be sold for cash. However, a person, not the KDD system, created this post hoc explanation. Most KDD systems learn a single concept or set of associations and do not attempt to update a knowledge base that contains a network of interrelated concepts. The experts who are expected to gain insight from such systems by definition have a considerable amount of knowledge about the field. Studying how people assimilate new knowledge could help us design better KDD systems.

Consistent contrast. People prefer category representations that define contrasting categories with different values on the same attributes.[13,14] For example, carnivores have sharp teeth and a short distance from eye to eye; herbivores have flat teeth and their eyes are further apart. However, machine-learning algorithms such as decision trees do not have this bias. For example, decision trees typically use different tests on different subtrees.

Global biases. There is some evidence that people prefer a category description in which each attribute value in the description is individually predictive of that category.[4] Tree and rule learning systems partition the data as learning progresses. A variable value correlated with one category in a partition might be inversely correlated with the same category on the entire database. This is the source of counterintuitive rules in which being forgetful is used as evidence against dementia. This expressive power of trees and rules is rarely needed and reduces the acceptability of learning models.

There is no question that today's KDD tools provide value to organizations that collect and analyze their data. We expect more from knowledge discovery tools than simply creating accurate models as in machine learning, statistics, and pattern recognition. We can fully realize the benefits of data mining by paying attention to the cognitive factors that make the resulting models coherent, credible, easy to use, and easy to communicate to others. ■

Acknowledgments

This work has benefited from discussions with Dorrit Billman, Marti Hearst, Pat Langley, Heikki Mannila, Geoff Webb, and the KDD group at UCI. The National Science Foundation grant IRI-9713990 partly funded this research.

References

1. B. Caulfield, "Hallmark Without the Revenue?" *Internet World*, 22 Feb. 1999.
2. R. Evans, and D. Fisher, "Overcoming Process Delays with Decision Tree Induction," *IEEE Expert*, Vol. 9, No. 1, Jan/Feb 1994, pp. 60–66.
3. S. Brin et al., "Dynamic Itemset Counting and Implication Rules for Market Basket Data," *Proc. ACM SIGMOD Int'l Conf. Management of Data*, ACM Press, New York, 1997, pp. 255–264.
4. M. Pazzani, S. Mani, and W.R. Shankle, "Comprehensible Knowledge-Discovery in Databases," *Proc. 19th Annual Conf. Cognitive Science Soc.*, M.G. Shafto and P. Langley, eds., Lawrence Erlbaum, Mahwah, N.J., 1997, pp. 596–601.
5. A. Srinivasan et al., "Carcinogenesis Predictions Using ILP," *Proc. Seventh Inductive Logic Programming Workshop, Lecture Notes in Artificial Intelligence 1297*, Springer-Verlag, Berlin, 1997, pp. 273–287.
6. J. Major and J. Mangano, "Selecting Among Rules Induced from a Hurricane Database," *J. Intelligent Information Systems*, Vol. 4, 1995, pp. 39–52.
7. U.M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery: An Overview," *Advances in Knowledge Discovery and Data Mining*, U.M. Fayyad et al., eds., AAAI/MIT Press, Menlo Park, Calif., 1996, pp. 1–34.
8. B. Gaines, "Transforming Rules and Trees into Comprehensible Knowledge Structures," *Advances in Knowledge Discovery and Data Mining*, U.M. Fayyad et al., eds., MIT Press, Cambridge, Mass., 1996, pp. 205–226.
9. C. Brunk, J. Kelly, and R. Kohavi, "MineSet: An Integrated System for Data Mining," *Third Int'l Conf. Knowledge Discovery and Data Mining*, AAAI Press, Menlo Park, Calif., 1997.
10. M. Sebrechts et al., "Visualization of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces," *22nd Annual Int'l ACM-SIGIR Conf. Research and Development in Information Retrieval*, ACM Press, New York, 1999.
11. M. Pazzani and D. Billsus, "Representation of Electronic Mail Filtering Profiles: A User Study," *Proc. ACM Conf. Intelligent User Interfaces*, ACM Press, New York, 2000.
12. M. Pazzani, "The Influence of Prior Knowledge on Concept Acquisition: Experimental and Computational Results," *J. Experimental Psychology: Learning, Memory & Cognition*, Vol. 17, No. 3, 1991, pp. 416–432.
13. D. Billman, "Structural Biases in Concept Learning: Influences from Multiple Functions," *The Psychology of Learning and Motivation*, D. Medin, ed., Academic Press, San Diego, 1996.
14. D. Billman and D. Davila, "Consistency Is the Hobgoblin of Human Minds: People Care but Concept Learning Models Do Not," *Program 17th Annual Conf. Cognitive Science Society*, Lawrence Erlbaum and Associates, Hillsdale, N.J., 1995.

Mike Pazzani is a professor and the chair of the Information and Computer Science Department at the University of California, Irvine. His research interests include data mining and intelligent agents. He received his BS and MS in computer engineering from the University of Connecticut and his PhD in computer science from UCLA. He is a member of the AAAI and the Cognitive Science Society. Contact him at the Dept. of Information and Computer Science, Univ. of California, Irvine, CA 92697; pazzani@ics.uci.edu.