Session 1526

Case-Based Reasoning for Engineering Statistics

George Runger, Sarah Brem, Norma Hubele, Toniann Rotante, Kathryn Kennedy
Arizona State University

Abstract
In this paper, we report on the formulation and early results of research supported by the National Science Foundation's Experimentation and Laboratory-Oriented Studies Division (DELOS). Using findings from cognitive science, we discuss the design of an intelligent tutoring system (ITS) that utilizes case-based reasoning (CBR) to scaffold undergraduate engineering students in their learning of introductory probability and statistics. Such a system will:

- Assist students in extracting the underlying common structure from engineering statistics problems that illustrate the full range of engineering disciplines.
- Allow the students to generate, customize, and change a virtually infinite collection of exercises that can be solved with the assistance of the ITS. The students can explore the effect of changes to solutions.
- Help students formulate and solve "practical" and "open-ended" problems, a skill stressed by the ABET Engineering Criteria.

Introduction
Several trends have converged to make this an important project at this time:

- Psychological and computational advances in CBR that allow us to use processes that model human thought, rather than those that are simply computationally efficient.
- Increased natural language processing (NLP) capabilities that allow more powerful ITS and provide psychologically valid models of language and knowledge representation.
- Advances that make technology readily accessible to students.
- A demonstrated need for teaching problem formulation skills in engineering curricula, as evidenced by the EC 2000 criteria [1].

Our goal is a design for an ITS that teaches key concepts of probability and statistics, encodes and retrieves problems, and assists students in solving problems while based on psychologically valid models of reasoning. We believe this will have the following benefits:

- Students will be able to explore, adapt and augment a large database of examples with a computer-based tutor as their guide. According to cognitive science research, this should help them recognize critical similarities, represent concepts of probability and statistics, and practice solution procedures.
- The ITS will help empirically test if psychological validity facilitates students' understanding

and internalization of concepts and solution procedures.

- Through the students' use of the ITS we will gain a better understanding of the mechanisms that students employ in representing and solving problems in engineering statistics.
- The ITS will support non-expert instructors, quite common in large service courses.

There is the potential to catalyze restructuring of engineering courses and inform cognitive and instructional models of CBR.

The Accreditation Board for Engineering and Technology wants to infuse engineering curricula with complex, practical problems that present challenges as students will face them in the real world. In the context of an undergraduate engineering statistics course, students have to: 1) master difficult concepts such as randomness and probability distributions, 2) learn to represent complex everyday situations as mathematical problems, and 3) acquire the mathematical skills needed to solve those problems. For example, a student might be faced with modeling reliability for a variety of products, from a fan belt to a networked computer to windshield glass. These scenarios have important similarities though they look quite different on the surface; students cannot rely upon recognizing content or keywords to decide which techniques to apply. Students have to make decisions such as whether these units are best categorized as repairable or nonrepairable. They have to decide which probability distributions within these categories are most appropriate, and decide between alternative strategies for graphical representation. Even a problem involving only one basic equation and a few values can be extremely complex.

Furthermore, to impart knowledge that will be used outside the training environment, where instead of simply solving a textbook problem they are given, engineers must first envision the problem. Psychological research suggests that traditional classroom experience may not be enough. Sternberg [2] argued for separate "intelligences"—practical and analytical—based on dissociation between classroom skills and "everyday" reasoning skills. This position is backed by research on mathematical problem solving (e.g., [3]). Villagarcia [4] argued for real-life problems in the teaching of statistics to engineers. Work on knowledge representation in postsecondary statistics is quite limited; Quilici & Mayer [5] provided a rare exception. One key to success is access to many examples that provide authentic engineering situations, as well as an instructor who can help students bridge between their training and applications. However, these resources are typically unavailable for even modest individual guidance.

Intelligent Tutoring System

An ITS can present an intriguing approach to the challenges enumerated above. An ITS can:

- Provide help on-demand with an essentially limitless supply of exercises that students and instructors can explore, modify, and even create from scratch.
- Accompany students into the field to work on-the-job problems.
- Provide cognitive "scaffolding," challenging students while providing them with enough support to prevent frustration and abandonment [6].
- Alleviate anxieties that prevent students from seeking help [7].

  This ITS will be constructed in accordance with psychologically tested models of cognition.

We believe this will help students internalize the techniques and develop independence. Student interactions with the ITS will, in turn, inform our system, helping us to restructure exercises to capitalize upon and strengthen students' cognitive capabilities.

Limitations of Rule-Based ITS

There are many approaches to ITS construction. One successful method is using production rules to tell the computer how to solve problems that students submit; this method has been demonstrated in many areas, including geometry and algebra [8]. In statistics education, Grabowski and Harkness [9] used rules to build decision trees that help students choose the appropriate inferential tests. Although rule-based models can theoretically generate explanations for students, most have not focused on this, concentrating on finding solutions [10]. These systems are efficient, though they must be coded in advance. There is evidence for the psychological validity of rules, but there is debate as to when and how rules are invoked. Thus, there are pros and cons to their use.

Cased-Based Reasoning ITS

A second approach involves case-based, or analogical reasoning denoted as CBR [11]. When students submit a problem, the system searches for examples sharing critical structural similarities. Our proposed ITS uses these examples, or cases, to guide students to a solution. CBR helps students to extend existing knowledge to novel problems and domains, to understand why two problems that appear completely different are really "the same thing." It promotes student self-explanation, demonstrated to improve student performance (e.g., [12]). CBR is prominent in medical education and legal education, but does not play a prominent role in statistics education. Also, CBR is completely compatible with production rule systems, and an ITS could switch methods depending on what will best serve the student. For example, using rules to generate explanations and cases to generate illustrations. In this research we are in the process of developing a CBR system. The CBR experience will help determine the utility of integrating these approaches in the future and, if so, the most effective way to do this.

CBR is not merely a computational method, but a general approach to knowledge representation and problem solving. An important point is that it could be implemented by human instructors as well as via an ITS. Existing research suggests a number of CBR enhancements. For example, students face three phases of problem solution, moving back and forth as they reach dead-ends or see new ways to formulate the problem (as adapted in [10]):

1. situation conception: understanding the problem in everyday language
2. mathematized situation conception: creating a mathematical representation
3. solution method conception: carrying out the mathematical procedures

Successful problem solvers and experts have a better grasp of the underlying problem structure than poor problem-solvers and novices [13]. They look beyond surface similarity during situation conception, and more readily recognize relevant equations during mathematized situation conception [14], [15].

Expected Benefits to Users

1. Students and faculty have the potential to create an "infinite" set of examples and problems. They can submit novel problems, or alter existing ones. The ITS can moderate the process to ensure that the mathematical integrity of the problem is maintained.
2. Students can classify, identify, compare and solve many examples. Their investigations are limited only by electronic storage space and the willingness of themselves and others to submit problems. By extending lessons over many domains and situations, they can learn to look beyond superficial similarities and differences in order to recognize common elements that drive the representation and solution process.
3. Multiple levels of support can be provided. For example, an initial level of support might be to display a key example from a corpus of problems, asking the student to transfer from the solved example to their target problem. Additional examples can be added, up to a full explanation of how to solve the target problem. By scaffolding students, providing just enough support to maintain reasonable progress, students can learn at their own pace, challenged but not frustrated.
4. The ITS can coordinate multiple representations of the material. An ITS system can move between exposition, cases and rules to create the most effective hybrid approaches. It can also switch between numerical and graphical representations. Although this work is beyond the scope of this project.
5. Students can change the parameters of a problem in order to see how these affect the solution.
6. Students will view explanations of problems, and learn to generate their own explanations. The ITS can generate explanations for students by showing them the similarities between problems, and referencing rules or expository text from these illustrations. Furthermore, it can prompt students to create their own explanations, which can then be evaluated by an instructor/trainer, or compared with the ITS explanation.
7. The ITS can help crystallize knowledge for non-expert instructors. Introductory service courses are often taught by graduate students, who have little pedagogical experience and may themselves be unsure of certain concepts. The instructors can use the system to refresh their own knowledge, as well as learning ways to coordinate exposition, cases and rules.
8. The ITS can follow students to their first profession position. Depending upon what proves to be the most effective delivery system, students will be able to continue to access the system. The ITS may be held locally on a CD-ROM, or deployed across the Web, allowing constant and immediate updating.

As Reed [10] pointed out, the contributions of cognitive science to math education are largely unrealized, and applications to postsecondary statistics are quite rare. Textbooks are not written with models of analogy and transfer in mind, nor are instructors taught how to encourage abstraction and transfer. The same is true for workplace training. It is quite possible that improvements could be achieved simply by rewriting these materials to capitalize on existing theory. This will help us to create student materials and guides for instructors and trainers that maximize transfer.

In order to be successful in the application of statistical knowledge to problems in the workplace, students must understand a complex set of interrelated concepts and terminology. Though this is not perfect agreement, there is some consensus about the elements of this set (e.g., [16], [17]). Complete understanding of these concepts is fundamental to the ability to appropriately apply them to new situations.

Making the link between the description of the problem and the underlying structure, in order to apply the appropriate solution technique, is particularly difficult for students with limited language skills compounded with mathematics anxiety. A disproportionate number of minority students fall into this category, and hence, are at a disadvantage in reaching a level of true understanding of statistical concepts. Strategies for improving their success rate in this domain may be related to the findings reported on by Harris and Schau [18] for women and girls. The knowledge gained in this proposed research will support this strategy of focusing on the underlying structure, and hence the level of transfer, required to improve the success rate of minorities and women.

Initial Assessment

In parallel with our development of the ITS we incorporated an early assessment. The objective of this assessment was to help us better understand transfer and difficulty and to provide an initial demonstration of efficacy of the system. Also, the design of the ITS needs to address the issues of exercise difficulty and exercise transfer; an initial assessment was used to provide this information. The system will eventually be based on a corpus of a large number of engineering statistics problems. However, this would require that we construct a method of grouping similar problems. Although this will eventually be addressed, after numerous discussions, the team decided to create rules to distinguish exercises and then write new exercises based on these rules. This would enable us to create a more targeted test of the case-based methodology. We focused on one particular type of exercise: hypothesis tests. This is typically a challenging topic for undergraduate engineering students. Exercises were generated and related via two criteria: difficulty and transfer.

Difficulty measures the degree of challenge of a problem. Transfer measures the differences in surface features from problem to problem. These features include the subject matter of the problem, the statistical words, and symbols used to present the problem. Transfer also includes replacing key words of the exercise with synonyms. For example, if the alternative hypothesis in one exercise uses “greater than”, then a transfer is to use “more than” in another exercise. The following set of rules were used to measure difficulty and transfer:

- Difficulty is derived from the sequence of steps for hypothesis testing provided by Montgomery, Runger, and Hubele [19]. See pages 155-156. Only the presence of steps 1, 2, 3, and 5 are changed in the related problems. Steps 2 and 3 are considered one step. Either both the null hypothesis and the alternative hypothesis are expressly stated in the problem, or neither is stated.

  Difficulty is measured using the following scale from 0 to 3: 0 if all three steps are explicitly stated in the problem; 1 if only one step is not explicitly stated in the problem; 2 if any combination of 2 steps are not explicitly stated in the problem; and 3 if none of the steps are explicitly stated in the problem.

- Transfer is derived by departures from the original problem in *setting*, *presentation*, and *computation*.

  Transfer is measured using the following scale from A to C: A is the root problem; B is if there are synonym replacement and changes in the data values; and C if there are changes in all three categories.

Much additional work is needed to study these measures, but the initial scoring is designed to match the ordinal increase in transfer (or difficulty) as more elements of an exercise are changed.

The following definitions of terms were used.

Setting: The engineering discipline or practical application from which the problem is derived. Example: The original problem was about cement. The new problem deals with shampoo.

Presentation: The syntax, word choice and order, and symbols used to state the problem. Example: The original problem used the words 'sample standard deviation' to express the alternative hypothesis. The new problem replaces the words with the symbol 's'.

Computation: The type of alternative hypothesis presented. Example: The original problem used the two-sided alternative hypothesis. The new problem uses a one sided alternative hypothesis. Note: a change in the values of the data used in the problem is considered a transfer of 0.

The task then became one of generating these problems to ensure that everyone interpreted and applied the rules in the same way. Each of five members of the team wrote problems and distributed them for review. It became evident that the group understood the rules and generated appropriate changes in transfer and difficulty for each problem. A family of exercises consisted of 8 exercise instances: 4 levels of difficulty times 1 level of transfer.

The following two exercises are an example of a transfer from synonym changes only:

SOFT DRINK MACHINE

A soft drink mixing device is modified to discharge a particular quantity of syrup into a container where it is combined with carbonated water. A random sample of 25 servings was determined to have a mean syrup composition of 0.068 fluid ounces and a standard deviation of 0.008 fluid ounces. Is there evidence to claim the mean quantity of syrup dispensed is different from 0.050 fluid ounces? Assuming normality of the data, use $\alpha = 0.01$.

BEVERAGE MACHINE

A post-mix beverage machine is adjusted to release a certain amount of syrup into a chamber where it is mixed with carbonated water. A random sample of 25 beverages was found to have a mean syrup content of 1.098 fluid ounces and a standard deviation of 0.016 fluid ounces. Do the data presented support the claim that the mean amount of syrup dispensed is not 1.0 fluid ounce? Assuming normality of the data, use $\alpha = 0.05$.

The following is an example of a transfer (from the previous exercises) of setting and computation:

EXERCISE: CABLE HARNESS

The diameter of holes for cable harness is known to have a standard deviation of 0.02 in. A random sample of size 10 yields the following data: 1.76, 1.69, 1.74, 1.73, 1.76, 1.77, 1.75, 1.78, 1.75, and 1.76, with sample mean 1.749. Is the true mean diameter less than 1.75? Use $\alpha = 0.01$.

The following two exercises are an example of a change of difficulty:

EXERCISE DIFFICULTY 0

A photoconductor film is manufactured at a nominal thickness of 25 mils. The product engineer wishes to decrease the energy absorption of the film, and he believes this can be achieved by reducing the thickness of the film to 20 mils. Eight samples of each film are manufactured in a pilot production process, and the film absorption (in $\mu J/in^2$) is measured. For the 25-mil film the sample mean is 1.179 and the sample standard deviation is 0.088, and for the 20-mil film the sample mean is 1.036 and the sample standard deviation is 0.093. Test H_1: $\mu_1 = \mu_2$ versus H_2: $\mu_1 > \mu_2$. Use the two sample t-statistic, $\alpha = 0.10$, and assume that the two population variances are equal and the underlying population is normally distributed.

EXERCISE DIFFICULTY 3

A photoconductor film is manufactured at a nominal thickness of 25 mils. The product engineer wishes to decrease the energy absorption of the film, and he believes this can be achieved by reducing the thickness of the film to 20 mils. Eight samples of each film are manufactured in a pilot production process, and the film absorption (in $\mu J/in^2$) is measured. For the 25-mil film the sample mean and standard deviation are 1.179 and 0.088, and for the 20-mil film the sample mean and standard deviation are 1.036 and 0.093 respectively. Do the data support the claim that reducing the film thickness decreases the mean energy absorption of the film? Use $\alpha = 0.10$ and assume that the two population variances are equal and the underlying population is normally distributed.

Experiment

As part of our research agenda, an early assessment was conducted of this CBR ITS with undergraduate engineers from a broad scope of academic disciplines who were enrolled in an

introductory engineering statistics course at Arizona State University in Fall, 2002. A computer approach was not used in this experiment. Hardcopy materials were used. Because our software design was not complete there was a concern that a weak implementation would be a substantial distraction from the core element of our research.
All exercises were centered around one-sided hypothesis tests on the mean using either a z or t statistic from the introductory engineering statistics course. The students had previous lectures, but not tests on this material.

Approximately 80 students volunteered for the experiment. At random, the participants were divided into two groups. One group was asked to work 6 exercises without examples. The second group worked the same exercises but with low-transfer examples available to be used as example cases. All exercises were of equal difficulty as measured by our scoring system. Both groups then worked post-test exercises. Detailed scoring rubrics were developed for the post-test exercises and the results for each group were compiled. Standard protocols for pre- and post-test information for the participants were followed. Written instructions were used in all cases. The student work was completed in one (very long) day.

Results from the full data set:

H_0: $\mu_{\text{examples}} = \mu_{\text{no examples}}$
H_1: $\mu_{\text{examples}} > \mu_{\text{no examples}}$
Sample mean test score for the group with examples: 21.42, sample size n = 38
Sample variance for the group with examples: 108.95
Mean test score for the group without examples: 17.31, sample size n = 39
Sample variance for the group without examples: 105.53
P-value for test: 0.043
We conclude that the mean test score for the group with examples is greater at significance level of 0.05.

Results after eliminating scores of post tests with no work or determined to not have made an effort:

H_0: $\mu_{\text{examples}} = \mu_{\text{no examples}}$
H_1: $\mu_{\text{examples}} > \mu_{\text{no examples}}$
Sample mean test score for the group with examples: 24.67, sample size n=33
Sample variance for the group with examples: 43.42
Mean test score for the group without examples: 20.91, sample size n=32
Sample variance for the group without examples: 53.89
P-value for test: 0.017
We conclude that the mean test score for the group with examples is greater at significance level of 0.05.

Although this is an initial, limited assessment and much work remains, the case-based method (with examples) resulted in significantly better student performance. This initial experiment was dramatically limited in the number of examples that could be used because it was not computer based. We expect future work to be able to use computers and consequently include many more related exercises into a study. From these results, with only a single example, we are excited about the possibilities with many more cases available.

ITS Design Plans

Based on the promising results of our initial assessment, our design of the ITS continues. As mentioned previously, an exercise family currently consists of 16 related exercises. We are in the process of incorporating these exercises into a prototype software shell that will allow for navigation among and between these exercise families, thus forming a large corpus of problems. Although the initial experiment was not computer based, we expect future work to be able to use computers and consequently include many more related exercises into a study. Our immediate design goal to integrate our difficulty and transfer families in a systematic manner to best assist student transfer.

References

1. EC (2000). *Accreditation Policy and Procedure Manual*. Baltimore, MD: Accreditation Board for Engineering and Technology, Inc.
2. Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.
3. Saxe, G. B. (1988). The mathematics of child street vendors. *Child Psychology* 59, 1415-1425.
4. Villagarcía, T. (1998). The use of consulting work to teach statistics to engineering students. *Journal of Statistics Education*, online at http://www.amstat.org/publications/jse/v6n2/villagarcia.html
5. Quilici, J.L. and Mayer, R.E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology* 88(1), 144-161.
6. Hogan, K. & Pressley, M. (Ed.) (1997). *Scaffolding student learning: Instructional approaches and issues*. Cambridge, MA: Brookline Books.
7. Ryan, A. M. & Pintrich, P. R. (1997). Should I ask for help?" The role of motivation and attitudes in adolescents' help seeking in math class. *Journal of Educational Psychology* 89, 329-341.
8. Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of Learning Sciences, 4,* 167-207.
9. Grabowski, B. L. & Harkness, W. L. (1996). Enhancing statistics education with expert systems: More than an advisory system. *Journal of Statistics Education*, online at http://www.amstat.org/publications/jse/v4n3/grabowski.html
10. Reed, S. K. (1999) *Word problems: Research and curriculum reform*. Hillsdale, NJ: Erlbaum.
11. Kolodner, J. L. (1993). *Case-Based Reasoning.* San Francisco, California: Morgan Kaufmann.
12. Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science* 13, 145-182.
13. Chi, M.T.H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R.J. Sternberg (Ed.), Advances in psychology of human intelligence (pp. 7-75). Hillsdale, NJ: Erlbaum.
14. Ross, B. H. (1987). This is like that: the use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13, 629-639.
15. Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 15, 456-468.

16. Garfield, J. (2000). The role of statistical reasoning in learning statistics. Paper presented at the annual meeting of the American Educational Research Association, April, New Orleans.
17. Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Hueristics and Biases*. Cambridge: Cambridge University Press.
18. Harris, M. B., & Schau, C. (1999). Successful strategies for teaching statistics. In S. N. Davis, M. Crawford, & J. Sebrechts (Eds.), *Coming into her Own: Educational Success in Girls and Women* (pp. 193-210). San Francisco: Jossey-Bass.
19. Montgomery, D. C., Runger G. C., & Hubele, N. F. (2001*). Engineering statistics, 2nd edition*. New York: Wiley.

GEORGE RUNGER
Professor Runger is Professor, Industrial Engineering, Arizona State University. He has researched in statistical methods for large engineering and scientific databases. Professor Runger's interest in this project stems from his experience as a class instructor, but also from extensive industrial consulting experience where statistical methods are applied. He holds a Ph.D. in statistics.

SARAH BREM
Professor Brem is Assistant Professor, Division of Psychology in Education, Arizona State University. She has researched analogical reasoning, student argument and explanation, and human knowledge representation for analogical reasoning. She holds a Ph.D. in cognitive psychology and completed an NSF Postdoctoral Fellowship in Science, Mathematics, Engineering and Technology Education at the University of California, Berkeley.

NORMA HUBELE
Professor Hubele is Professor, Industrial Engineering, Arizona State University. She has researched in statistical methods and data analysis. She has also published research in engineering education. She is the course director and lead instructor for the interdisciplinary undergraduate engineering statistics course at Arizona State University and coauthor of a leading engineering statistics textbook. She has a Ph.D. in engineering systems.

TONIANN ROTANTE
She is a graduate student in Industrial Engineering at Arizona State University. She has a B.S. in industrial engineering and industrial experience. Her research interests include the use of statistical methods in engineering analysis.

KATHRYN KENNEDY
Ms. Kennedy is a recent B.S. graduate in Industrial Engineering at Arizona State University and she will begin her graduate studies in the same department in 2003. She recently completed and also was a teaching assistant for an introductory engineering statistics course.