FROM THE EDITORS

OVER-THE-HORIZON: NOT JUST FOR RADAR ANYMORE

By Norman Chonacky

A FEW MONTHS AGO, I HEARD AN INTERESTING COLLOQUIUM BY JOHN HOPCROFT, ONE OF THE PREEMINENT THINKERS IN THE AREA OF THEORETICAL COMPUTER SCIENCE. HIS TALK MIRRORED HIS CURRENT INVESTIGATIVE PASSION—FUTURE DIRECTIONS IN THEORETICAL COMPUTER SCIENCE—BUT HIS

approach can be instructive to anyone concerned with what I call adaptive anticipation (discerning trends and adapting behavior to changes before events require it).

He begins with a simple inquiry: when does a question in theory become a significant question? His answer: when it is associated with a matter that has great practical importance. He asserts that such a question is arising in computer science because the nature of computing is changing. These changes are marked by the sophistication of user interfaces, huge gains in speed and memory capacity at the desktop, the merging of computing and communications, and the overwhelming availability of data in digital form.

But what are some indicators of the changing nature of computing? Hopcroft argues that one such category is Internet users’ expectations. Current queries—“What information is there?”—yield pages of hits, but we expect that we should be able to ask, “How can I apply the information that’s there?” One of several examples he gave is easy to relate to, given the cost of higher education and the current cultural passion for picking the “right” school. Now a student can ask, “Where can I find information on colleges?” But what an anxious 18-year-old really wants to ask is something like, “Where should I go to college?” Rather than a pile of college catalogs, the student wants to know how that information relates to whether this or that college is likely to serve his or her goals and needs.

Such an expectation could well be fueled by our experiences with e-merchants that provide us with suggestions, such as “Others who ordered this item also ordered A, B, and C.” However vast the stores of general expository data accessible on the Web might be, the way they’re organized and indexed isn’t conducive to extracting the features that would help in responding to the college query. Hopcroft enumerated many examples of such queries to conclude that we could derive great practical value from theory in our efforts to construct methods that permit efficient access to “relevant” information from large bodies of data in real time.

Hopcroft then went on to articulate the challenge for theory, posed by this significant practical need, in the form of two questions related to the methods for identifying pattern changes in large data sets:

- How soon can we detect a change in patterns in a large volume of data?
- How large must a change be in order to distinguish it from random fluctuations?

He proceeded to develop cogent examples of how the use of large graphs, spectral analysis, high-dimensional data reduction, clustering, collaborative filtering, and extraction of signal from noise were related to this methodological challenge for theory. It was at this point that my thoughts began to wander toward how his line of inquiry might relate to my own question of discerning trends in computational applications to science and engineering. Anticipating and responding to such trends is, after all, one of the editor in chief’s principal responsibilities.

Hopcroft’s concluding tour de force—a discussion of the prototypical application of these methods to a study of Citeseer—refocused my attention. Citeseer, a database of computer and information science papers, contains some 300,000 technical documents linked via references. The study, conducted at two time marks several years apart, identified several natural communities and categorized them as stable, ascending, or declining. I fancied that we could also use such an application to identify natural communities within the content domain from which CiSE derives its target reader population. When I approached Hopcroft after his talk, he smiled and confessed that the Citeseer study re-
quired a vast level of effort. The caveat, I surmised, is that although the directions for theoretical development are emerging, the theory itself isn’t ready for methodological developers. The consequent methodologies are even less ready for potential users like me.

Although my vision of pressing a button and receiving reports about computational use trends within the science and engineering community turned out to be a pipe dream, I was nonetheless inspired by Hopcroft’s talk in two respects. First, it validated the notion that the state of natural communities is a valid and potentially fruitful concept on which to construct anticipatory analyses. Second, it suggests that I can reasonably apply his approach of inquiry to peek over my own knowledge horizon, evaluating the state of our natural communities and drawing at least qualitative conclusions that will help me direct CiSE’s editorial policy.

Have you been thinking about future directions for theoretical computer science? CiSE welcomes manuscripts for feature articles from anyone thinking about computer science theory as it relates to the needs of computing in science and engineering, especially in the light of recent advances in computer hardware and software. Please visit www.computer.org/cise/author.htm for details.