Testing and debugging remain the most practical means of assuring the quality of software. In particular, partition testing and random testing are widely adopted in test case selection. According to classic studies, partition testing is better than random testing by only a small margin in many cases, but requires considerably more overhead. Random testing is, therefore, popular as a simple alternative.

In previous work [1, 2], Chen et al. introduced the innovative idea of adaptive random testing (ART) to further enhance random testing in test case selection. The fundamental concept is to improve the diversity or even spreading of test cases in the input domain. ART improves over random testing by more than 40 percent in terms of the $F$-measure, which is the number of test cases expected to reveal the first failure.

In real-life projects, however, the software under test may exhibit numerous failures. It is more customary to continue testing with a view to identifying more failures before debugging commences. This paper proposes a new metric to gauge the number of test cases expected to reveal the first $n$ failures. Using this new $F^n$-measure, the authors study the empirical results of two standard scenarios, in which testing and debugging are performed either simultaneously or in turn. They find ART to be statistically more effective than random testing in relation to the $F^n$-measure when testing and debugging are conducted in turn.

For readers interested in comparing test case selection approaches, this paper provides innovative insights and practical solutions. The $F^n$-measure will undoubtedly become a new yardstick for test case selection. Further studies of its application to test case prioritization and fault localization should provide additional insights.

References
