

Statistical Software Debugging

Alice Zheng
Carnegie Mellon University

Traditional software debugging is an arduous task that requires time, effort, and a good understanding of the source code. Given the scale and complexity of the task, the development of methods for automatically debugging software seems both essential and very difficult. However, several trends make such an endeavor increasingly realistic: (1) the wide-scale deployment of software, (2) the establishment of distributed crash report feedback systems, and (3) the development of statistical machine learning algorithms that can take advantage of aggregate data over multiple users.

In this talk, I present a statistical software debugging framework that applies machine learning techniques to run-time reports of instrumented programs. The problem has a relatively simple solution under the single-bug assumption. However, in the more realistic case of multiple bugs, the problem can no longer be dealt with using feature selection and classification techniques. I describe the challenges and present a solution inspired by bi-clustering algorithms.

This is joint work with Ben Liblit (U. Wisconsin, Madison), Michael Jordan (U.C. Berkeley), Alex Aiken and Mayur Naik (Stanford).

Speaker's Bio:

Alice Zheng received her PhD from UC Berkeley in 2005. She is interested in applications of machine learning in software and computer systems. Her recent work includes statistical software debugging, failure diagnosis in computer networks, and models of social networks. She is currently a postdoctoral fellow at Carnegie Mellon University.