

자동 수정의 테스트 과적합 문제

이창공

2024년 1월 18일

Abstract

프로그램 자동 수정 기술의 테스트 과적합 문제는 오랫동안 대두되어온 고질적인 문제이다. 이러한 문제를 해결하기 위하여 수많은 방법론들이 제안되고 있다. 그러나, 이러한 방법론들은 모두 테스트 케이스를 가정하는 방법을 사용하고 있으며, 이는 결국 테스트 과적합 문제를 극복하였다고 할 수 없다. 이렇게 오랜 시간 동안 하나의 문제에 대한 뚜렷한 해결책이 없는 이유는 기술의 평가 지표에 있을지도 모른다.

프로그램 자동 수정 기술의 테스트 과적합 문제는 기술이 처음 제안된 2000년대 중후반 부터 지금까지 뚜렷한 해결책 없이 대두되어온 고질적인 문제이다. 이러한 문제는 대부분의 자동 수정 기술이 다수의 실패 테스트 케이스를 가정하기 때문에 발생한다. 자동 수정이 온전히 작동 하기 위해서는 고쳐야 할 결함의 발견이 선행되어야 한다. 실제 개발 환경에서 실용적으로 가장 많이 쓰이는 결함 발견 방법은 테스트 케이스를 이용하는 것이므로, 이러한 가정은 고질적인 문제에도 불구하고 상당히 현실적인 셈이다.

이러한 문제를 해결하기 위하여 수많은 방법론들이 제안되고 있다. 특히, 싱가포르 국립대학교의 아빅 샤티(Abhik Roychoudhury) 교수는 해당 문제에 대한 여러 방법론들을 적극적으로 제안하고 있다. 대표적인 테스트 케이스 과적합 문제의 해결 방안으로는 21년도에 제안된 논문과 같이 적은 테스트 케이스로 작동할 수 있는 방법과 24년도에 제안된 테스트 케이스 자동 합성기를 사용하여 테스트 케이스의 수를 극단적으로 늘리는 방법이 있다. 테스트 케이스 과적합 문제를 단적으로 생각했을 때 이러한 방식은 기존의 방법들보다 효과적이라고 볼 수 있다.

그러나, 이러한 방법론들은 모두 테스트 케이스를 가정하는 방법을 사용하고 있으며, 이는 결국 테스트 과적합 문제를 극복했다고 할 수 없다. 테스트 과적합 문제는 근본적으로 수정의 결과를 실패한 테스트 케이스로 검증하기 때문에 발생한다. 따라서, 테스트 과적합 문제를 해결하기 위해서는 테스트 케이스를 가정하지 않는 새로운 방법이 필요할 수 있다. 예를 들어, 페이스북에서 19년에 개발한 방식은 테스트 케이스를 사용하지 않고 다수의 수정을 적용한 뒤 개발자에게 선택을 맡기는 방식을 사용하고 있다. 해당 방식은 자동화를 포기함으로써 과적합 문제를 극복한 것으로 하나의 예시가 될 수 있다.

이렇게 오랜 시간 동안 하나의 문제에 대한 뚜렷한 해결책이 없는 이유는 기술의 평가 지표에 있을지도 모른다. 테스트 과적합 문제의 근본적인 해결책을 찾기 위해서는 수정의 안전성이 보장되어야 한다. 즉, 검사에 의존적이기 보다 보수적인 수정 생성 방법론이 필요하다. 이러한 지적은 과거부터 있었던 반면에 현재까지의 자동 수정 기술의 평가 지표는 수정의 개수에 맞추어져 있다. 따라서, 가시적인 성과에 초점을 둔다면 보수적인 수정 생성은 잘 드러나지 않는 성과일지도 모르겠다.