

서울대학교 컴퓨터공학부
최적화 및 금융공학 연구실
김진현 (jh@soar.snu.ac.kr)

Designing costs in cost-based syntax error repair schemes

개요

- ▶ Repairing syntax errors in LR parsers
 - ▶ Joint work with 김익순(ROPAS alumnus)
 - ▶ Corchuelo, et al. (ACM TOPLAS 2002)
 - ▶ 문법 오류 발생시 이를 복구해가며 파싱 진행
 - ▶ 파서가 도중에 중단되지 않음
 - ▶ 해당 지점 이후의 문법 오류도 보고 가능
 - ▶ 국소적인(local) 에러 복구
 - ▶ 에러 발생 위치에서 토큰의 삽입/삭제 시도
 - ▶ 복구 후 남은 입력 스트링 분석 시도(shift)
 - ▶ 가능한 경우가 여러이면 최소 비용 연산 선택
 - ▶ 각각의 삽입과 삭제 연산의 비용은 동일
 - ▶ 즉, 최소 횟수의 연산으로 복구 시도
 - ▶ “연산 비용을 결정하는 것이 오류 복구의 품질을 결정하는데, 굳이 동일한 비용을 사용할 필요가 있을까?”

접근 방법

- ▶ GA를 이용한 연산 비용 결정
 - ▶ Genetic Algorithm
 - ▶ 최적화 문제를 풀기 위한 기법
 - ▶ 진화의 원리를 이용(교배, 변이)
 - ▶ 비용을 결정할 하나의 최적화 문제로 생각
 - ▶ 하나의 해 = 하나의 비용 할당 내역(cost table)
 - ▶ 이를 토대로 파서를 수행하여 에러 복구 시도
 - ▶ 그 결과에 따라 해의 적합도(fitness)를 평가
 - ▶ 해의 범위
 - ▶ 각각의 토큰의 삽입/삭제마다의 비용 결정(99×2)
 - ▶ 비용의 제곱합이 1이 되도록 normalize (198차원 구면)
 - ▶ 해의 적합도
 - ▶ Testset - 학생들의 과제물로 작성된 자바 코드
 - ▶ 복구 비용, 수정된 토큰 개수, 복구 내역 활용 가능

실험 및 결과

▶ 실험 개요

▶ 각 비용의 범위

- ▶ $[0, 1]$, $[0.06, 0.08]$

- ▶ 최소 비용과 최대 비용이 크게 차이나지는 않도록

▶ 해의 적합도 평가

- ▶ Testset을 복구하는데 든 총 비용의 합

- ▶ 미리 최적해를 알 수 있음

 - ▶ Testset에서 복구에 사용하는 토큰의 개수가 크지 않음

 - ▶ 이 토큰의 비용을 최소로 하는 것이 최적

- ▶ 사전 예비 실험 성격

▶ 실험 결과

- ▶ 각각의 경우에서 성공적으로 최적해를 찾아냄

- ▶ 5~6시간 정도 소요(Testset 1회 수행에 약 5초)

문제점 및 향후 진행 방향

- ▶ 해의 적합도 평가 방법 확립
 - ▶ 토큰 수정 횟수, 복구 후 shift 횟수 활용
 - ▶ 몇 가지 아이디어를 적용하여 실험
 - ▶ Testset 확대
 - ▶ 무작위적으로 에러를 추가한 자바 코드 사용
 - ▶ 예상 외로 좋은 결과가 얻어지지 않음
- ▶ 원인 분석
 - ▶ Cost table이 변해도 파서의 에러 복구 방식이 유사
 - ▶ 이후에 파서가 진행되는 shift 패턴도 유사
 - ▶ $A = B + ;$
 - ▶ Literal, this, null, boolean literal 등의 삽입 가능
 - ▶ 그러나 이후 shift가 진행되는 것은 동일
 - ▶ 예상 외로 cost의 효과가 미미한 것으로 보임

문제점 및 향후 진행 방향(cont.)

- ▶ Cost 범위 확대
 - ▶ Cost가 에러 복구에 미치는 영향 확인
 - ▶ 토큰별 삽입/삭제 외의 요소 고려
 - ▶ 파서의 상태, 토큰, 남아있는 입력등의 요소
- ▶ 해의 적합도 평가 방법 확립
 - ▶ Shift 횟수 정보 활용
 - ▶ 에러가 작성자의 의도에 가깝게 복구되는지 여부
- ▶ 에러 복구 기법 변경
 - ▶ 기법 특징상 cost 영향이 약할 수 있음
- ▶ 대상 언어 확대
 - ▶ Java, C/C++

A decorative header at the top of the slide consists of several overlapping horizontal bars. From left to right, there are three vertical bars in shades of red, orange, and light orange. Below these is a long, thin light purple bar that spans most of the width. On the right side, there are two overlapping rectangular blocks, one in a darker purple and one in a light pinkish-purple.

▶ Thank you!

▶ Q&A