

2023 Algorithm HW OPT : GP + DP

due: 6월 9일(금) 오후 23시 59분

1. Problem Statement

샘표로 구별된 dataset.csv 데이터 파일이 주어진다. 6가지 성질(feature)를 계측하였고, 각각의 성질은 A, B, C, D, E, F 열에 저장하였다. 이 데이터가 조합되어 나타나는 현상(result)을 계측하였고, 이는 G열에 저장하였다. 계측한 데이터의 수는 총 200개(200행)다. 6가지 성질의 조합으로 현상이 나타나며, 이 현상을 수식으로 표현하려고 한다. 또한 A, B, C, D, E, F 중 일부 데이터는 필요가 없는 데이터임을 알고 있다. Genetic programming과 dynamic programming을 이용해 가장 적합한 수식을 찾아라.

[입력]

입력파일은 dataset.csv다.

[출력]

출력 파일의 이름은 GP.out이다. 수식을 중위표현식으로 출력한다. 출력 파일의 연산자, 피연산자, 함수의 구별은 공백으로 한다. A, B, C, D, E, F 열 데이터는 각각 해당 열 대문자로 표기한다. 단, 연산 표기 형식의 다양성을 고려하여, 수식 해석에 문제가 없다면 모두 정답으로 인정한다.

출력 예 : $(3 + A \times e^{(2 + \log E)})$

3 + A * e (2 + log E)

또는

3.0 + A * e (2 + log (E))

또는

3 + A * e ^ (2 + log E)

등 모두 정답으로 인정

2. Restriction

조합론적 폭발이 발생하지 않도록 사용 가능한 operation은 사칙연산(+, -, x, /)과 일부 함수(sin, cos, tan, sinh, cosh, tanh, log, e)로 제한한다. 수식의 총 길이 역시 제한을 하며, 리프 노드의 개수를 기준으로 64개 이하다. 피연산자는 -10 이상, 10 미만의 소수점 2자리 미만의 임의의 상수가 5개 이하로 포함될 수 있다.

3. Language

프로그램 언어는 C++을 사용한다. Genetic algorithm과 genetic programming과 관련된 어떠한 라이브러리도 사용할 수 없으며, HW7, HW8와의 연속성을 위해 C++로 프로그램 언어를 제한한다.

4. Compilation and Execution

컴파일은 제출한 소스 파일을 이용해 채점 컴퓨터가 LINUX에서 컴파일한다. 채점 컴퓨터는 Intel Xeon Scalable 4210 이며, OS는 Ubuntu Server 22.04.2 LTS다.

5. Team

팀을 구성하여 문제를 해결할 수 있다. 팀의 구성 인원은 최소 1인, 최대 5인이며, 지난 숙제와 다른 친구들과 팀을 구성해도 무방하다.

6. 제출

제출 파일은 아래 **3파일** 2파일을 학번.zip 으로 압축하여 제출한다. 폴더 형태로 압축하지 않는다.

- 학번.cpp
- ~~학번.exe~~
- 학번.pdf

학번은 팀원 중 가장 빠른(가장 작은) 숫자이며, LMS 제출은 해당 학번 학생만 제출하고, 다른 팀원은 제출하지 않는다. 학번 숫자에 반드시 하이픈을 포함한다.

예시.

수하루

2021-11111	2022-11110	2021-11113	2021-11114	2021-11115
김서아	김이서	김지안	김하윤	김지아

제출 파일: 2021-1111.zip

압축 해제 시 다음 3파일이 .zip 파일과 같은 경로에 위치한다.

- 2021-1111.cpp
- ~~2021-1111.exe~~
- 2021-1111.pdf

7. Report

프로젝트에 대해 보고서를 작성하여 제출한다. 분량은 A4 4페이지 이하이며, 팀 이름과 멤버의 학번, 이름을 적고 주어진 형식으로 작성한다. 보고서 내에 소스 코드를 그대로 포함해서는 안된다. 포함 시 큰 패널티를 받는다. 보고서 본문에는 최소한 다음 내용을 포함시킨다.

- 1) 사용한 GA 전체 구조
- 2) Dynamic programming을 활용한 방안
- 3) 사용한 GA 연산자에 대한 설명
- 4) 테스트 컴퓨터 사양과 정답 도출에 걸린 시간
- 5) Discussion(느낀 점, 잘 안 되는 점, 의외의 현상, 예상대로 된 점 등)

8. Grading

- 평가는 보고서와 코드를 1:1 비율로 평가한다. 제출한 코드 중 일부는 공개되며, 제출 후 팀의 코드를 본인의 포트폴리오에 올려도 무방하다. 둘 모두 상대평가에 의해 점수가 부여된다.
- 보고서에 정답이 기재되어 있는 경우, 컴파일한 파일을 8개로 병렬 처리하여 각각 3,600초 실행한 뒤 종료하여 정답을 도출하는지 확인한다. 하나라도 정답이 도출되는 것을 확인하면 코드에 점수를 부여한다.
- 보고서는 위의 최소 내용을 모두 포함할 경우 기본 점수가 부여되며, 나머지 성적은 보고서 내용이 충실한 정도에 따라 평가한다. 보고서에 불필요한 내용(표지, 소스코드, 별도의 분석 없는 raw data 등)을 포함하는 경우 감점 요인이 된다.
- Copy detection program에 의해 카피 판별 후, 원본과 복사본 공히 0점 처리하며 최종 학점에서 두 grade 강등함. 일부든 전부든 모두 카피로 처리함.
- 보고서와 프로그램을 따로 제출할 수 없다.