Final project

20211195 하태관

Result of GP project

1. Data 1

Fitness function and prediction points

|  |  |
| --- | --- |
|  |  |

* 1. number of nodes of best\_individual:49
  2. RMSE value of best\_individual:1.15
  3. deth of best\_individual:12
  4. average number of nodes:46.8
  5. average fitness value=6 (다만, depth가 12를 초과하면 페널티로 10이 추가되기 때문에, fitness가 rmse와 일치하지는 않습니다)
  6. average depth=11.654

1. Data 2

Fitness function and prediction points

|  |  |
| --- | --- |
|  |  |

* 1. number of nodes of best\_individual:58
  2. RMSE value of best\_individual:0.550
  3. deth of best\_individual:10
  4. average number of nodes:51.8
  5. average fitness value=2.14
  6. average depth=9.492

Performance of MLP vs GP

1. Data 1

|  |  |  |
| --- | --- | --- |
|  | GP | MLP |
| Prediction |  |  |
| performance |  |  |

Number of parameter in MLP: 61202

MLP rmse loss: 2.6~4.2 (진동)

Epochs = 1000

Number of nodes in GP: 49

GP rmse loss: 1.15

1. Data 2

|  |  |  |
| --- | --- | --- |
|  | GP | MLP |
| Prediction |  |  |
| performance |  |  |

Number of parameter in MLP: 10401

Epochs = 1000

MLP rmse loss: 0.4~0.5 (진동)

Number of nodes in GP: 58

GP rmse loss: 0.55

Conclusion

* + 1. Optimal solution vs Computing speed:

Genetic programming은 search space 내부 수많은 포인트에서 generation을 진행하기 때문에 optimal solution을 수월하게 찾을 수 있지만, MLP는 적은 포인트에서 gradient descent를 진행하기 때문에 local solution에 빠질 수 있다. 다만, Genetic programming은 전체 population을 순차적으로 evaluation하기 때문에 적은 점으로 학습을 진행하는 mlp 에 비해 계산시간이 느리다. (실제 실험에서 GP가 훨씬 오래 걸렸기 때문에, GP에 multiprocessing 적용 필요)

* + 1. Black box,

실험에서 1만개, 혹은 6만개의 MLP 내부 파라미터가 각각 무엇을 의미하는지, 어떠한 계산 과정을 거쳤는지 알지 못하는 Black box 문제가 존재한다. 이에 비해, GP는 내부 노드 symbol (add, sub, x, 1, etc)을 직접 확인할 수 있고, 노드의 일부분을 다른 task에 활용할 수 있다.

* + 1. Node number

학습 이후에, GP에서 사용하는 노드의 규모(50)와 MLP에서 사용한 파라미터(10000~60000)의 규모가 확연히 차이 났다. 다른 task에 trained data를 활용할 때, GP는 더 적은 메모리로 task 진행이 가능하다.

* + 1. Datawise

MLP는 Data-Wise이기 때문에 더 많은 데이터가 주어졌을 때, Genetic programming에 비해 더 나은 성능을 보일 수 있을 것이다.