Solving ML with ML: Effectiveness of A* for synthesizing ml pipelines

CSE3000 - Research Project

Rémi Lejeune - R.J.Lejeune@student.tudelft.nl

TU Delft

1. Introduction

Issues when creating machine learning pipelines :

- Complex to build
- Require advance knowledge
- Time-consuming

Possible solution: Automated machine learning using program synthesis.

2. Program synthesis

Definition:
Program synthesis is the task of automatically finding a program that satisfies certain constraints.
Implementation used:
Create a Context-Free-Grammar and search through it to find the required program

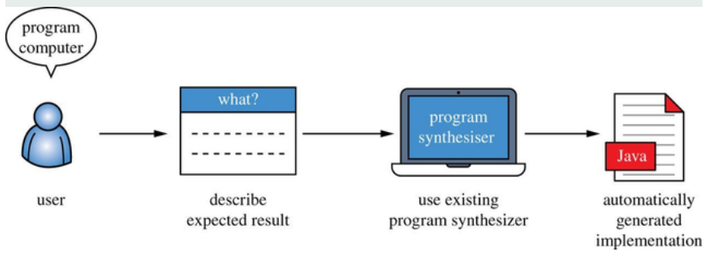

Figure 1: Program synthesis-based workflow for using a computer to solve a problem

3. Research question

Can A* improve the performance of autoML using program synthesis?

4. Method

a. Datasets

Simple datasets	Adv. datasets	Papers' datasets
iris	gas-drift	glass
seeds	musk	car-evaluation
blood transfusion	madelon	wdbc
monks-problem	gisette	wine-quality-red
ilpd	har	wine-quality-white
qsar-biodeg		spambase
tic-tac-toe		

Table 1: The three groups of datasets created

b. Grammar

```
START =
    Pipeline ( [ CLASSIF ] ) | Pipeline  ( [ PRE, CLASSIF ] )
PRE =
    PREPROC | FSELECT |
    ( ”seq”, Pipeline  ( [ PRE, PRE ] ) ) |
    ( ”par”, FeatureUnion ( [ BRANCH, BRANCH ] ) )
BRANCH =
    PRE | CLASSIF | ”seq”, Pipeline  ( [ PRE, CLASSIF ] ) )
```

Listing 1: The grammar used to generate ML pipelines

c. A*

Formula:

$$\underbrace{(1 - acc)}_{\text{Heuristic function}} + \underbrace{0.001 * (1 + log_{10}(T))}_{\text{Cost function}}$$

- acc is the accuracy
- T is the time in milliseconds

Goal:
Focus on accuracy and if two pipelines are almost equal split on time taken

5. Results

Goal:
Find if A* performs better than the other algorithms

Parameters used:

- Train on 300 samples
- Run the algorithm 10 times
- Evalue maximum 100 pipelines per run
- Datasets used: Seeds, HAR, and WDBC
- Max depth of the pipelines is 5

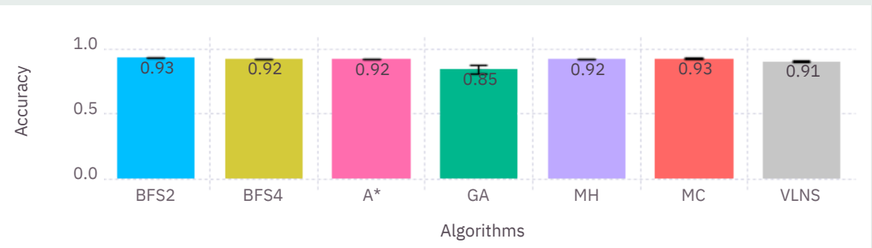

Figure 2: Results found for the Seeds dataset

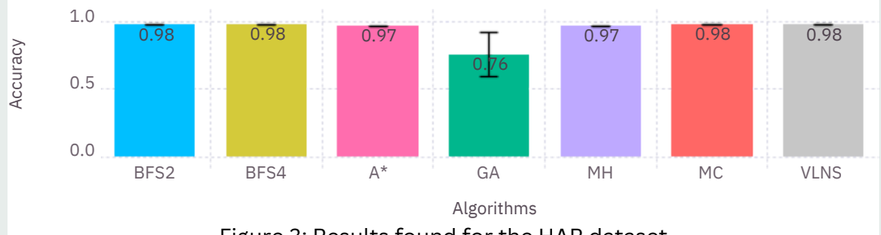

Figure 3: Results found for the HAR dataset

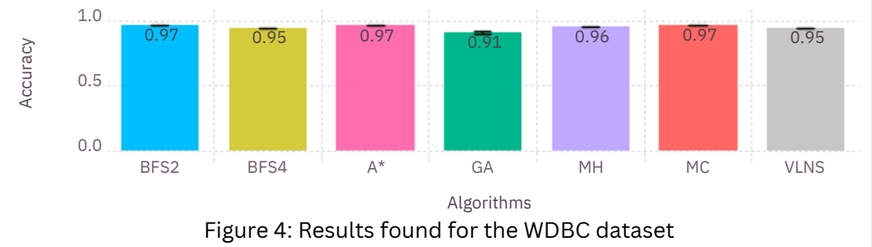

Figure 4: Results found for the WDBC dataset

6. Limitations

Due to the inability to access to DelftBlue:

- Only 100 pipelines were evaluated for each run
- Only trained on 300 samples for each run

7. Future work

- Use datasets that are more complex and require more preprocessing
- Create an adaptive version of A* that explore the breadth of the search space
- Remove limitations and use a supercomputer

8. Conclusion

From the results, we can conclude that A* did not improve performance since its accuracy was similar to the other algorithms. However, this cannot be generalized as only three datasets were evaluated