Learning MDL Logic Programs From Noisy Data

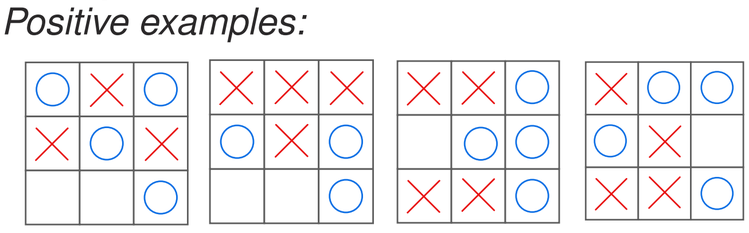
Céline Hocquette[1], Andreas Niskanen[2], Matti Järvisalo[2], Andrew Cropper[1]
[1]University of Oxford
[2]University of Helsinki
celine.hocquette@cs.ox.ac.uk

1 - Introduction

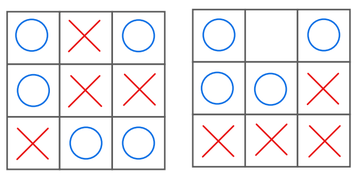
The goal of inductive logic programming (ILP) [1] is to induce a program (a set of logical rules) that generalises training examples.

Problem: learning optimal and complex programs from noisy data is difficult.

Example 1 (Game rules)
Positive examples:

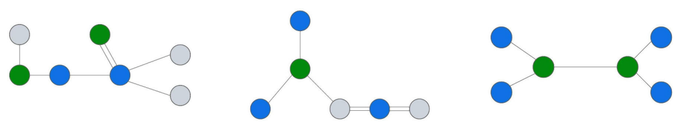


Negative examples:



```
win(Board,Mark) ← cell(Board,0,Mark),
                  cell(Board,1,Mark),
                  cell(Board,2,Mark)
win(Board,Mark) ← cell(Board,2,Mark),
                  cell(Board,5,Mark),
                  cell(Board,8,Mark)
win(Board,Mark) ← cell(Board,0,Mark),
                  cell(Board,4,Mark),
                  cell(Board,8,Mark)
```

Example 2 (Drug design)



```
pharma(Drug) ← zincsize(Drug,Zinc),
               hydrogen(Drug,Hydro),
               singlebond(Zinc,Hydro),
               hydrophobic(Hydro)
pharma(Drug) ← zincsite(Drug,Zinc),
               carbon(Drug,Carbon),
               doublebond(Zinc,Carbon),
               distance(Zinc,Carbon,Dist),
               leq(Dist, 2.6)
```

2 - Minimal Description Length

We search for a MDL program:

Definition 1 (**MDL cost function**). *The MDL cost of a hypothesis h is:*

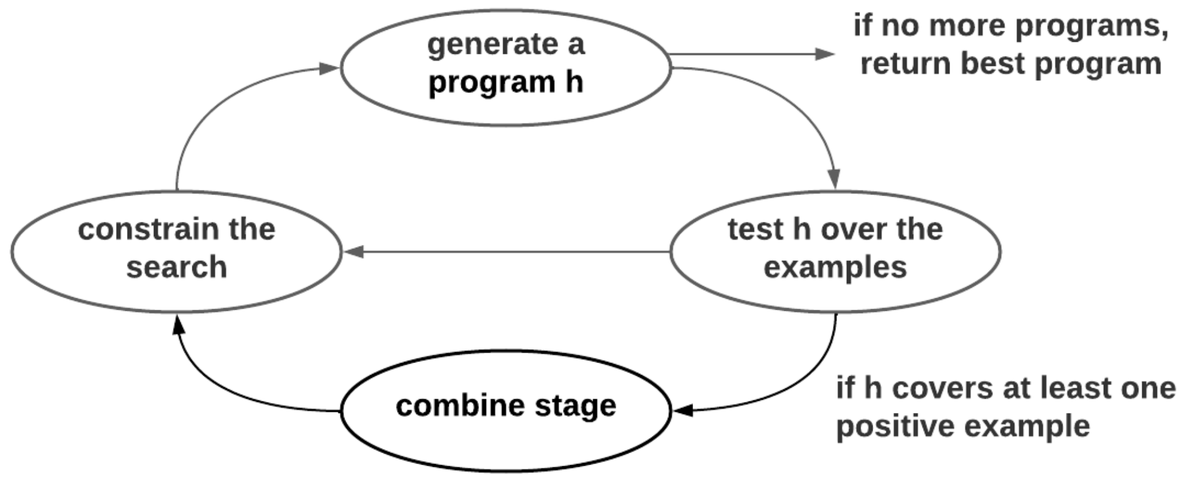
$$cost(h) = size(h) + falsenegative(h) + falsepositive(h).$$

3 - Theoretical Analysis

Theorem 1
MaxSynth *returns a MDL solution.*

4 - Our approach (MaxSynth)

Key idea: learn small programs independently and then try to find a *minimal description length* combination (a union) of these small programs.



We use a MaxSAT solver to search for a MDL combination (a union) of programs.

Why does it work?
1. We can harness recent progress in MaxSAT solving [2].
2. We can build effective constraints based on the MDL principle.

5 - Experiment

Q1 Can MaxSynth learn programs from noisy data?
Q2 How well does MaxSynth handle progressively more noise?

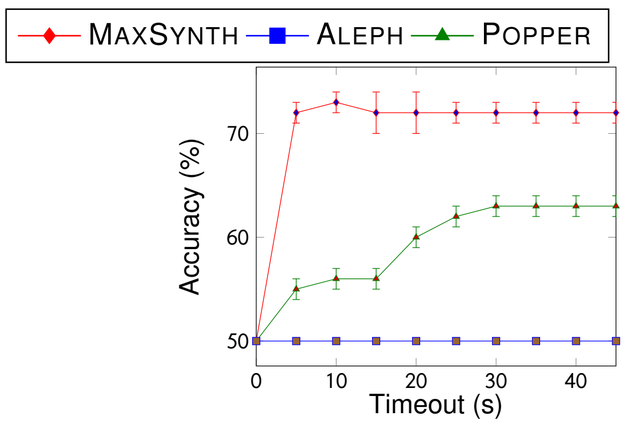


Fig. 1: Predictive accuracy with larger timeouts on *alzheimer-toxic*.

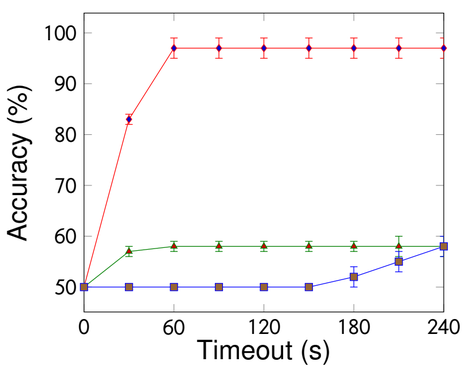


Fig. 2: Predictive accuracy with larger timeouts on *zendo2 (20)*.

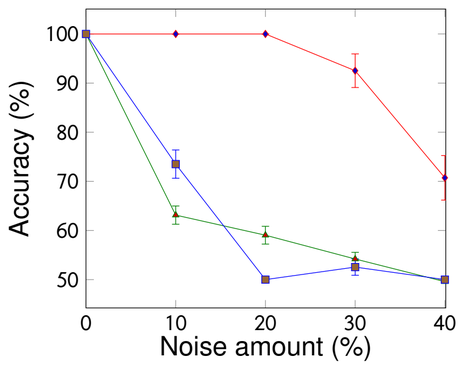


Fig. 3: Accuracy versus the noise amount on *iggp-rps*

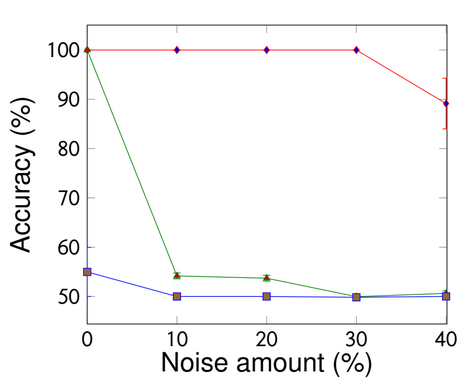


Fig. 4: Accuracy versus the noise amount on *dropk*.

- **MaxSynth can (i) learn programs, including recursive programs, with high accuracy from noisy data, and (ii) outperform existing systems in terms of predictive accuracies.**
- **MaxSynth can scale to moderate amount of noise.**

6 - Conclusion and Limitation

- **An approach that efficiently learns MDL programs from noisy examples.**

Future work: better cost functions

Article

References

[1] A. Cropper and S. Dumančić. Inductive logic programming at 30: A new introduction. *J. Artif. Intell. Res.*, 2022.

[2] Marek Piotrów. Uwrmaxsat: Efficient solver for maxsat and pseudo-boolean problems. In *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 132–136, 2020.