Inferring Generative Model Structure with Static Analysis

Paroma Varma, Bryan He, Payal Bajaj, Nishith Khandwala, Imon Banerjee, Daniel Rubin, Christopher Ré

STANFORD INFOLAB

STANFORD DAWN

Summary

- **Generative Models to Label Training Data**
  Use generative models to combine sources of weak supervision to assign noisy labels
- **Complex Dependencies among Sources**
  Sources of labels are rarely independent
- **Inferring Model Structure**
  Use static analysis to infer dependencies and encode in generative model structure

Theoretical Results

- **Learning Dependencies**
  Learning k-degree dependencies among n heuristics requires $O(n^{k-1}\log n)$ samples
- **Inferring Dependencies**
  Analyzing the code can infer the dependencies among heuristics without data

  Given dependencies, learning heuristic accuracies requires $O(n \log n)$ samples

Experimental Results

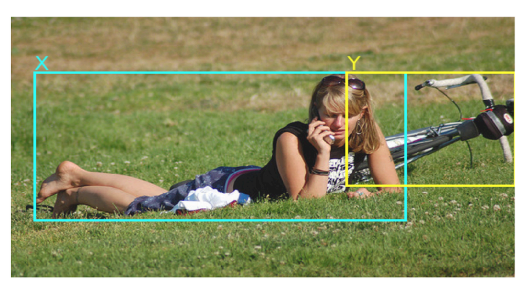
Application	Model	Improvement Over			
		MV	Indep	Learn Dep	FS
Visual Genome	GoogLeNet	7.49*	2.90*	2.90*	-0.74*
ActivityNet	VGGNet+LR	6.23*	3.81*	3.81*	-1.87*
Bone Tumor	LR	5.17	3.57	3.06	3.07
Mammogram	GoogLeNet	4.62	1.11	0	-0.64

** reports F1 scores, rest in accuracy (%)*

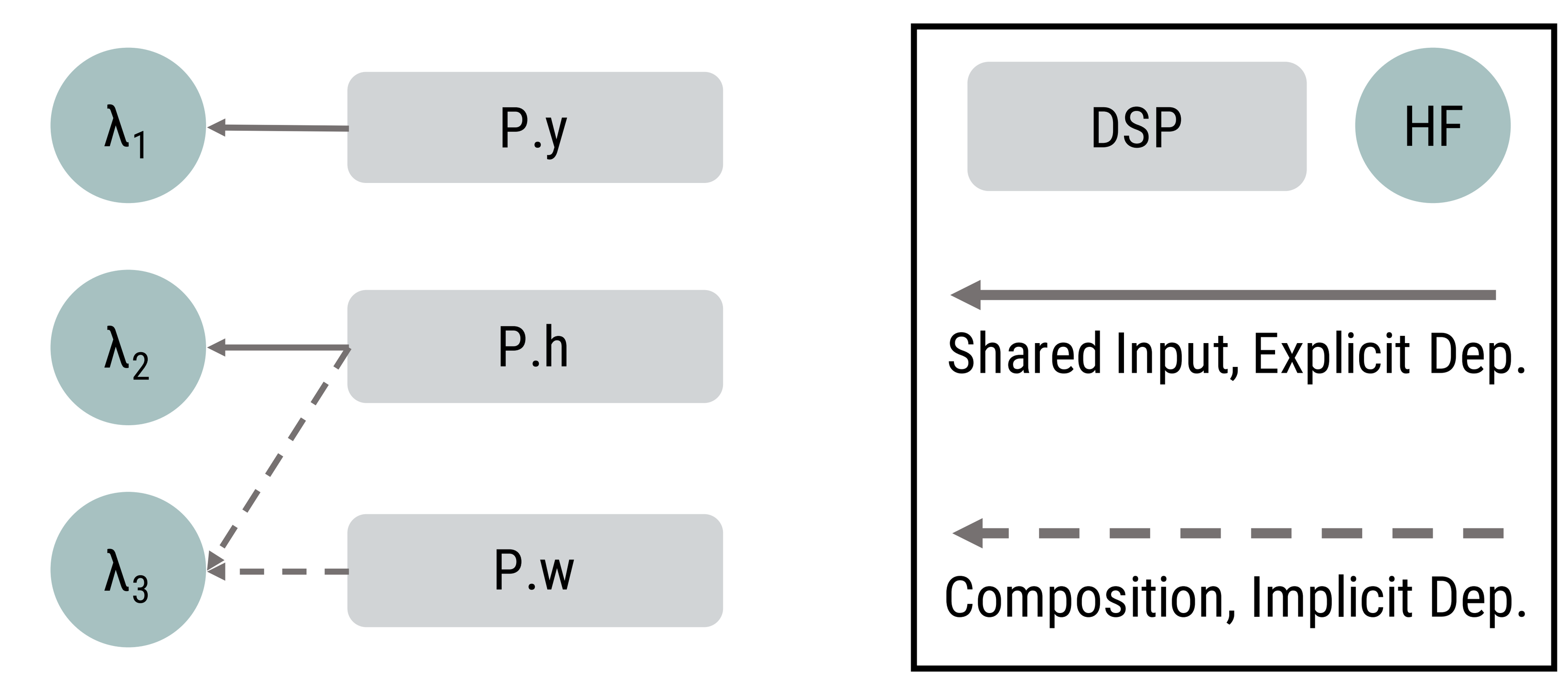
- Inferring dependencies outperforms learning dependencies
- Outperforms fully supervised model with additional noisy training labels

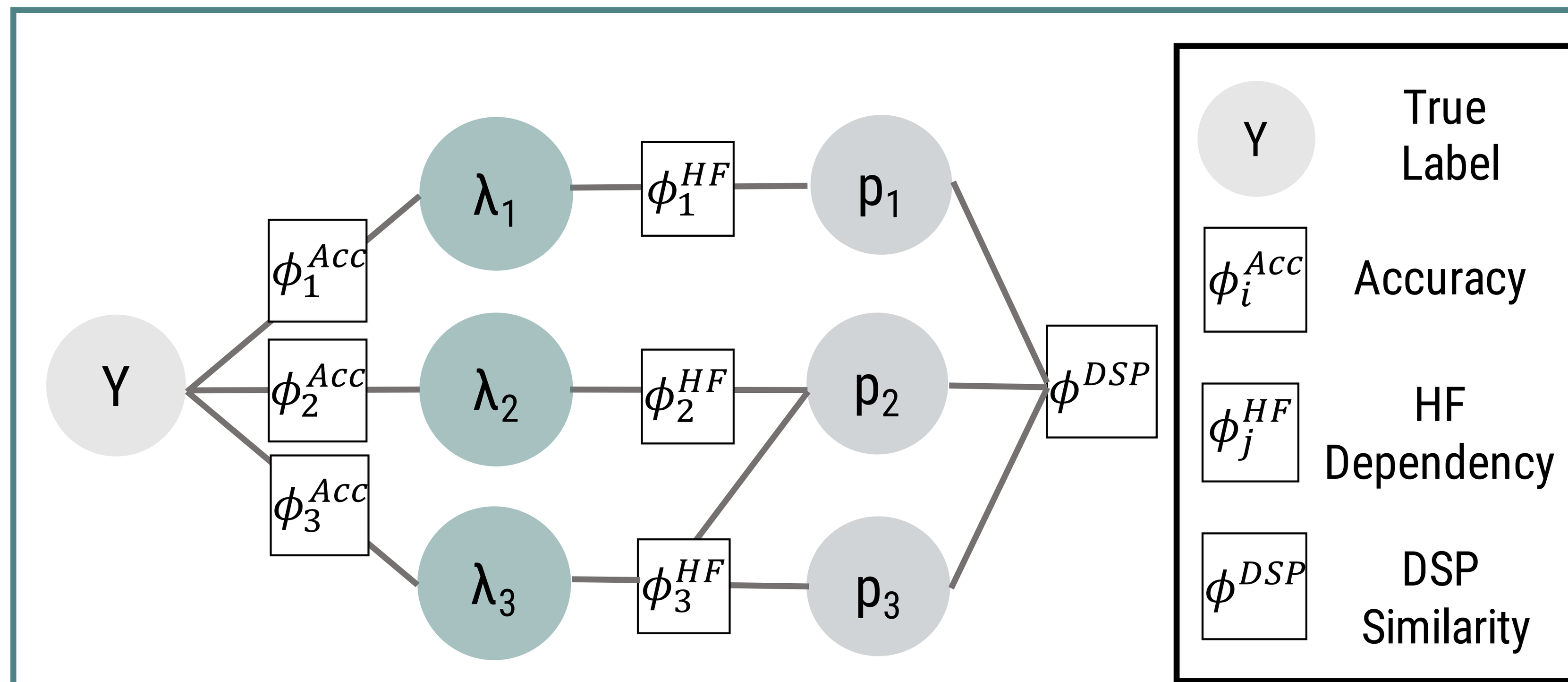
```
defλ_1(P.y):
    if P.y[‘person’]>= P.y[‘bike’]:
        return True
    else:
        return False
```

```
defλ_2(P.h):
    if P.h[‘person’] >= P.h[‘bike’]:
        return True
    else:
        return False
```



```
defλ_3(P.area):
    if P.area[‘person’]
            <=
        2*P.area[‘bike’]:
        return True
    else:
        return False
```





Heuristic Structure

- **Domain Specific Primitives (DSPs)**
  Interpretable characteristics of raw data
- **Heuristic Functions (HFs)**
  Programmatic rules that output noisy labels

Static Analysis

- **Shared Input** Sharing primitives as inputs leads to explicit dependencies
- **Compositions:** Primitives composed of others can lead to implicit dependencies

Statistical Modeling

- **HF Dependency** Represents the dependencies found using static analysis
- **DSP Similarity** Represents the learned correlations among the DSPs