

SKILL - a Stochastic Inductive Logic Learner

J. Côte-Real¹, T. Mantadelis¹, I. Dutra¹, R. Rocha¹, and E. Burnside²

¹CRACS & INESC TEC, University of Porto
{jcr, theo.mantadelis, ines, ricroc}@dcc.fc.up.pt

²Department of Radiology, University of Wisconsin, WI, USA
eburnside@uwhealth.org

Contributions

SKILL is a PILP tool which can extract non-trivial knowledge from probabilistic data. It takes as input relational data annotated with probabilities and it produces a logical model that can output probabilities as well.

The contributions of this work are:

- ✓ the SKILL system, which is a general purpose PILP tool that supports logical constructs, arbitrary variable logical terms and annotated disjunctions
- ✓ an efficient search strategy to traverse the PILP search space
- ✓ an analysis of a real-world medical dataset pertaining breast cancer

Results (Biopsies)

Biopsy is very important in determining malignancy of a lesion and usually yields definitive results; however, in 5% to 15% of cases, the results are non-definitive.

Probabilistic ILP allows for incorporating in the PBK the confidence of physicians in observations and known values from the literature, namely the confidence in malignancy for each case (before excision), assigned by a group of physicians analysing that case. Learning from non-discrete classes is a unique characteristic of SKILL that combines interpretable rules with a non-boolean classification model.

```
is_malignant(Finding) ←  
  is_mass(Finding, Mass) ∧  
  mass_shape(Mass, irregular) ∧  
  mass_density(Mass, high) ∧  
  mass_margin(Mass, microlob) ∧  
  0.70::is_malignant(mass_margin(microlob))  
is_malignant(Finding) ←  
  is_mass(Finding, Mass) ∧  
  mass_shape(Mass, irregular) ∧  
  0.50::is_malignant(mass_shape(irregular)) ∧  
  mass_margin(Mass, indistinct) ∧  
  0.20::is_malignant(mass_margin(indistinct))
```

SKILL's hypotheses can be used for probabilistic prediction of malignancy of a tumor. We found that the probabilities produced by SKILL are much closer to the expected values than the probability values. In other words, SKILL can produce predictors that are better calibrated than other traditional probabilistic models.

SKILL's Algorithm

Initially, the algorithm uses the TopLog engine to generate all possible hypotheses of length one. A hypothesis of length one is constructed from literals that are contained in the PBK and are selected by the user. This approach results in hypotheses of length one that mirror patterns contained in the observations w.r.t. the PBK. The algorithm proceeds by generating hypotheses with length greater than one until reaching MaxHypLength.

Algorithm 1: SKILL Algorithm

```
1 Input = PBK, PE, MaxHypLength, Psize, Ssize, PRankMetric,  
  SRankMetric, EvalMetric  
2 Output = Best hypothesis according to EvalMetric  
3 Hyps1 = HypsN = AllHyps = generate_hyps_length_one(PBK, PE)  
4 for Length = 2; Length ≤ MaxHypLength; Length++ do  
5   Primary = select_members(HypsN, Psize, PRankMetric)  
6   Secondary = select_members(Hyps1, Ssize, SRankMetric)  
7   HypsN = generate_combinations(Primary, Secondary)  
8   AllHyps = AllHyps ∪ HypsN  
9 end  
10 return best_hypothesis(AllHyps, EvalMetric)
```

Combining hypotheses to generate new hypotheses with larger length is not a trivial task; possible combinations N choose K , N being the number of length one hypotheses and K the maximum hypothesis length.

Results (Comparison)

SKILL's execution time and probabilistic accuracy were compared against the probabilistic rule learner ProbFOIL+ using a probabilistic dataset of 44 probabilistic facts about family relations. The dataset is composed of literals mother, father and parent (Prolog rule), as well as of 10 examples of the target predicate grandmother, all with probability 1.00.

Setting	Execution time (s)	
	SKILL	ProbFOIL+
mother	0.46	0.95
mother + father	0.73	2.91
mother + father + parent	0.96	45.31

ProbFOIL+'s independent error minimization technique allows for higher accuracy in every experiment, but as the dataset grows up to 44 facts, SKILL's accuracy is only marginally lower to that of ProbFOIL+ (97% and 99.7%, respectively). SKILL is clearly more scalable than ProbFOIL+, since its runtime only doubles for 3 literals when compared to 1, whereas ProbFOIL+ gets 50 times slower in the same case.

SKILL's search strategy selects candidate hypotheses for two different sets, named Primary and Secondary and new hypotheses are then generated by combining members from each set. In each iteration of the algorithm, the primary set is filled with the Psize best hypotheses, according to a given ranking metric (argument PRankMetric), from the set of hypotheses generated in the previous iteration (1 clause hypotheses when searching for 2 clauses hypotheses; 2 clauses hypotheses when searching for 3 clauses hypotheses; etc).

The secondary set is populated with Ssize hypotheses from the set of hypotheses of length one, according to SRankMetric. Depending on the ranking metrics chosen, the system can generate hypotheses in a fully stochastic way, use best hypotheses or create a heterogeneous mix. The stochastic component of the selection is distinct for each iteration. Finally, all hypotheses are evaluated according to the given evaluation metric (argument EvalMetric), and the best generated hypothesis for all different lengths is returned.

Conclusion

This work presents the PILP learner SKILL, which generates FOL rules (hypotheses) that can be used for classification and prediction and that produce probabilistic values.

SKILL addresses efficiency issues in hypotheses generation by limiting the number of candidate hypotheses in its search space.

SKILL was compared against the PILP learner ProbFOIL+ using the family dataset and both systems were found to generate the same final hypotheses. SKILL's algorithm performed up to 2 orders of magnitude faster and presented the same accuracy as ProbFOIL+ for larger datasets.

Finally, SKILL was used to extract non-trivial knowledge from a dataset of non-definitive biopsies which was annotated with probabilistic literature values and rules. Results showed that these annotations to the BK were in fact used in the physician's mental models and therefore useful for prediction of malignancy.

Acknowledgements

J. Côte-Real is funded by the FCT grant SFRH/BD/52235/2013. This work is financed by the FCT (Portuguese Foundation for Science and Technology) within project UID/EEA/50014/2013, and by NIH NLM grant R01LM010921.