Scalable Rule Learning in Probabilistic Knowledge Bases

Arcchit Jain — KU Leuven
Tal Friedman — UCLA
Ondřej Kuželka — KU Leuven
Guy Van den Broeck — UCLA
Luc De Raedt — KU Leuven

SafeLearner

Knowledge Bases (KBs) are becoming increasingly:

- Larger
- Probabilistic
- Incomplete

What is new?
We use Lifted Inference to learn probabilistic rules for completion of such large and probabilistic KBs

Why Rule Learning?
- Handles probabilistic KBs (PDBs)
- Completes KBs in an explainable way

Why *SafeLearner*?
Significantly faster than ProbFOIL$^+$ and scales as good as AMIE+ (Runtime under 2.5 hours on KB with 14k+ tuples)

Example

researcher	paper	p
bob	plp	0.9
carl	plp	0.6
greg	plp	0.7
ian	db	0.9
harry	db	0.8

author/2

researcher	university	p
edwin	harvard	1.0
fred	harvard	0.9
alice	mit	0.6
dave	mit	0.7

location/2

researcher	researcher	p
alice	edwin	0.2
alice	fred	0.3
bob	carl	0.4
bob	greg	0.5
bob	harry	0.6
bob	ian	0.7
carl	greg	0.8
carl	harry	0.9
carl	ian	0.8
dave	edwin	0.7
dave	fred	0.6
edwin	fred	0.5
greg	harry	0.4
greg	ian	0.3
ian	ian	0.2

$target$: coauthor/2

AMIE+ Rules(H):
- coauthor(A, B) :− author(A, C), author(B, C).
- coauthor(A, B) :− location(A, C), location(B, C).

Query (Q): $\exists\ a, b \quad s.t. \quad p_{h_1} \vee (p_{h_2} \wedge location(a,c) \wedge location(b,c)) \vee (p_{h_3} \wedge author(a,d) \wedge author(b,d))$

Learned Rules (H^*):
- 0.105::coauthor(A, B) :− true.
- 0.687::coauthor(A, B) :− location(A, C), location(B, C).
- 0.333::coauthor(A, B) :− author(A, C), author(B, C).

Problem Specification

Given:
- A probabilistic KB (PDB) : $\mathcal{D} = \{\langle tuple, probability \rangle\}$
- A target relation : $target$

To Find: A set of rules H^* that it minimize cross-entropy

$$H^* = \underset{H}{\operatorname{argmax}}\ Cross\ Entropy\ (H, E, D)$$
$$= \underset{H}{\operatorname{argmin}} \sum_{\langle t_i, p_i \rangle \in E} (p_i \log q_i + (1 - p_i) \log(1 - q_i))$$

where E is the set of $target$ tuples in $\mathcal{D}$, and q_i is the predicted probability of i^{th} tuple t_i.

Algorithm

1. Get deterministic rules (H) by running AMIE+ on the deteriministic part of $\mathcal{D}$ (ignoring the probabilities associated with tuples)
2. To each rule in H, add the classical probability of $P(head = true \mid body = true)$ calculate on all $target$ examples.
3. Convert H to a single query Q
4. Within Stochastic Gradient Descent's loop:
   a) Randomly pick a *target* tuple $\langle t, p \rangle$ from E
   b) Get symbolic expression of $P(Q(t))$ over $\mathcal{D}$ using Lifted Inference Engine of *SlimShot*
   c) Calculate gradient of cross-entropy loss and update rule probabilities
5. Remove all the rules from H with insignificant probabilities

Key features

- Uses **Lifted Inference** in rule learning, thereby avoiding grounding for knowledge compilation
- Uses **memoization** to store the canonical structures of all the queries with their probability expressions
- Breaks larger queries into independent subqueries for better performance

Source code and full paper available at:

https://github.com/arcchitjain/SafeLearner/tree/AKBC19

DTAI Declarative Languages & Artificial Intelligence
European Research Council — Established by the European Commission
Research Foundation Flanders — Opening new horizons