IBM Research

Unifying Learning and Reasoning

Achille Fokoue
Principal Research Staff Member
Manager, AI Foundations - Reasoning
IBM Research AI

IBM

Can a crocodile run a steeplechase?

How Can Machines Reason?

- Humans
  - Steeplechase requires jumping
  - Crocodile has weak, short legs
  - Therefore, can not jump

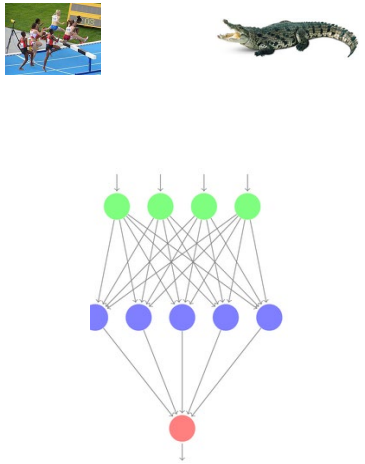
  => Can not run a steeplechase

- Information Retrieval
  - Pr(steeplechase, crocodile)

- Symbolic Reasoning

$$\ldots$$
$$\forall x.Crocodile(x) \supset WeakLegs(x)$$
$$\ldots$$
$$\forall x.WeakLegs(x) \supset \neg CanJump(x)$$
$$\ldots$$
$$\forall x.\neg CanJump(x) \supset \neg CanSteeplechase(x)$$
$$\ldots$$

- Neural Reasoners



- Answers need to be explicitly present in the text

- Provided knowledge needs to be complete and correct
- Complex knowledge bases need to be created

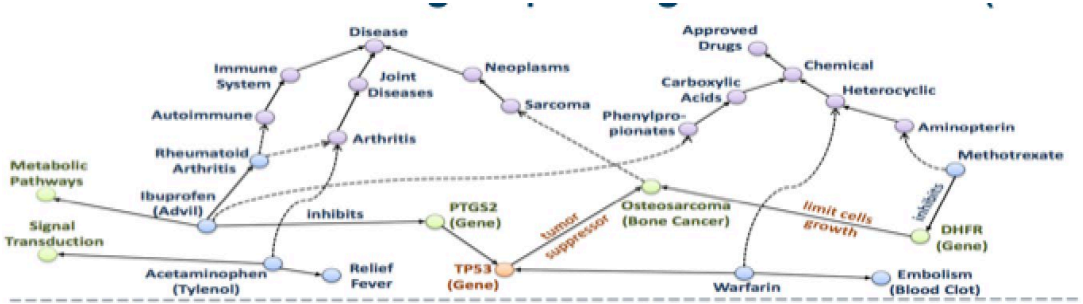
- Need large amounts of data to train
- Difficult for humans to interpret results

What do Machines Need for Reasoning?

- Knowledge
  - In form of rules, ontologies, knowledge bases, etc.
  - Knowledge Graph, Knowledge Induction efforts
- Ability to Use that Knowledge
  - Efforts in algorithmic developments for reading comprehension, deep learning, machine learning, etc.
  - Reasoning systems and tools to support these algorithmic advancements
  - Scalability

What Counts as Reasoning and Learning? What are Their Limitations?

Symbolic Inference over knowledge bases: Ontological, Rule and Probabilistic Inference



$$\frac{\text{NOT four\_sides}(X) \qquad \text{rectangle}(X) \Rightarrow \text{four\_sides}(X)}{\frac{\text{NOT rectangle}(X) \qquad \text{square}(X) \Rightarrow \text{rectangle}(X)}{\frac{\text{NOT square}(X) \qquad \text{square}(c)}{\square}}}$$

Symbolic Reasoning Limitations:

- Complexity & Scalability of Reasoning
- Complex knowledge bases needed

Natural Language Inference:

Two farmers are using two horses to help with farm work

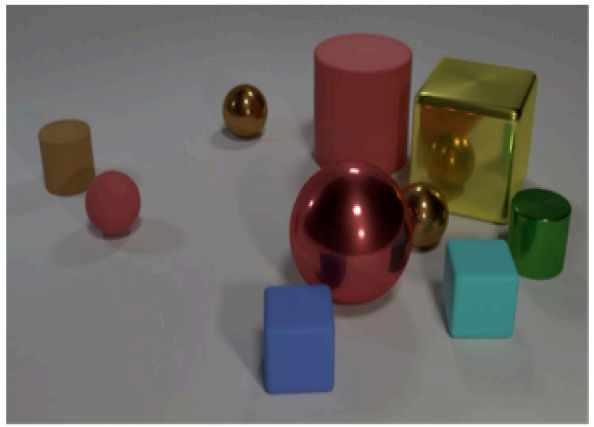
↓

The people are ploughing the fields

1. John picked up the apple.
2. John went to the office.
3. John went to the kitchen.
4. John dropped the apple.
5. Where was the apple before the kitchen?

Answer: office

Visual Q&A



Q: Are there an equal number of large things and metal spheres?
Q: How many objects are either small cylinders or red things?

Neural Reasoning Limitations:

- Need large amounts of data to train
- Difficult for humans to interpret results

Addressing Limitations by Bridging the Divide

Reasoning to Address Learning Issues:
Data Inefficiency & Explainability

UReQA Project

Learning to Address Reasoning issues:
Scalability/Computational Complexity & KB creation

TRAIL Project

IBM Research

Complex Reasoning Over Contextualized Knowledge in Natural Language

Avinash Balakrishnan
Maria Chang
Kshitij Fadnis
Achille Fokoue-Nkoutche
Pavan Kapanipathi
Bassem Makni
Siva Patel
Kartik Talamadupula
Michael Witbrock
Mo Yu

Complex Reasoning Over Contextualized Knowledge in Natural Language

- It's important for full AI, hinges on other AI problems like reading comprehension, NLU, reasoning, KR

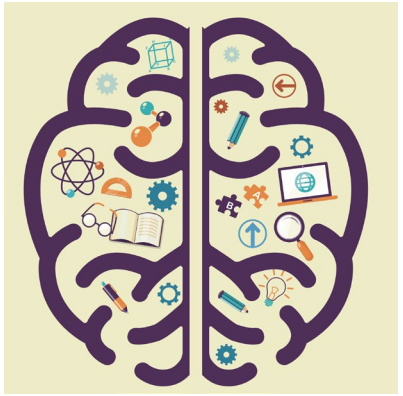
- NLP tasks that require reasoning
  - Complex Question Answering
    - Standardized Tests
  - Natural Language Inference/Textual entailment

- Solving these tasks requires:
  1. Ingest and/or acquire relevant background knowledge from many domains.
  2. Formalize this knowledge and update as necessary.
  3. Reason over the acquired knowledge

In your body, what two organs work together to make sure that oxygen gets to all the other organs of your body?

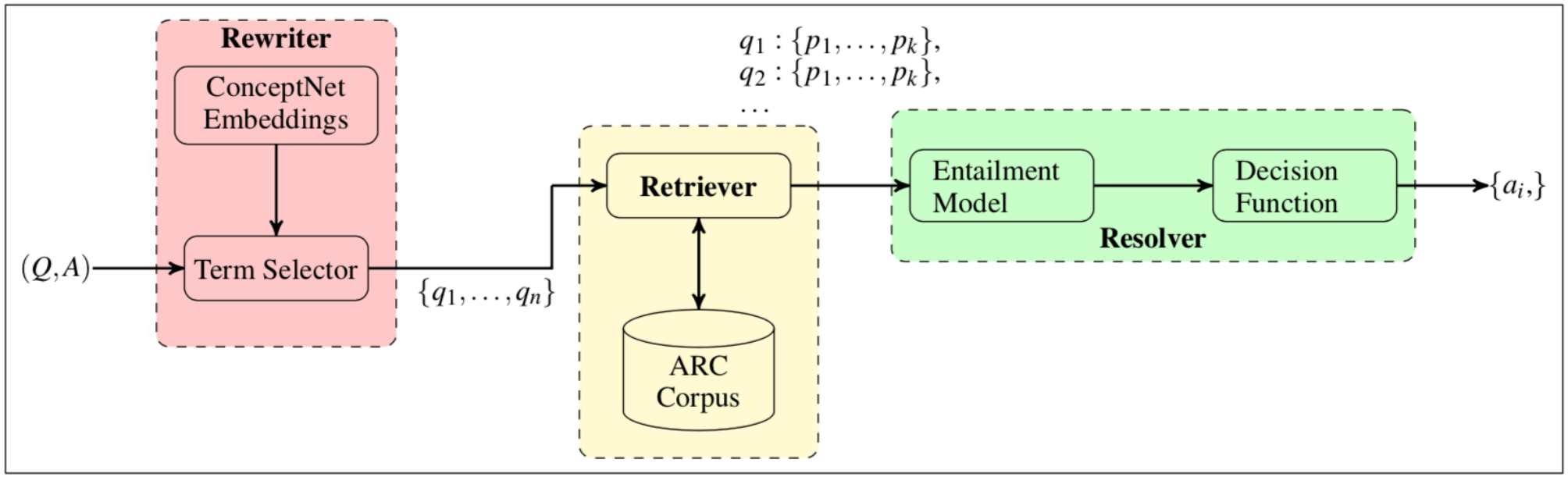
(A) Lungs and kidneys (B) Heart and lungs

(C) Brain and kidneys (D) Heart and liver



Query Rewriting with Background Knowledge (AKBC 2019)

- Use a Rewriter to select terms from the question and answers that make good queries.
- By combining query rewriting, background knowledge, and textual entailment our system is able to outperform several strong baselines on the ARC dataset.

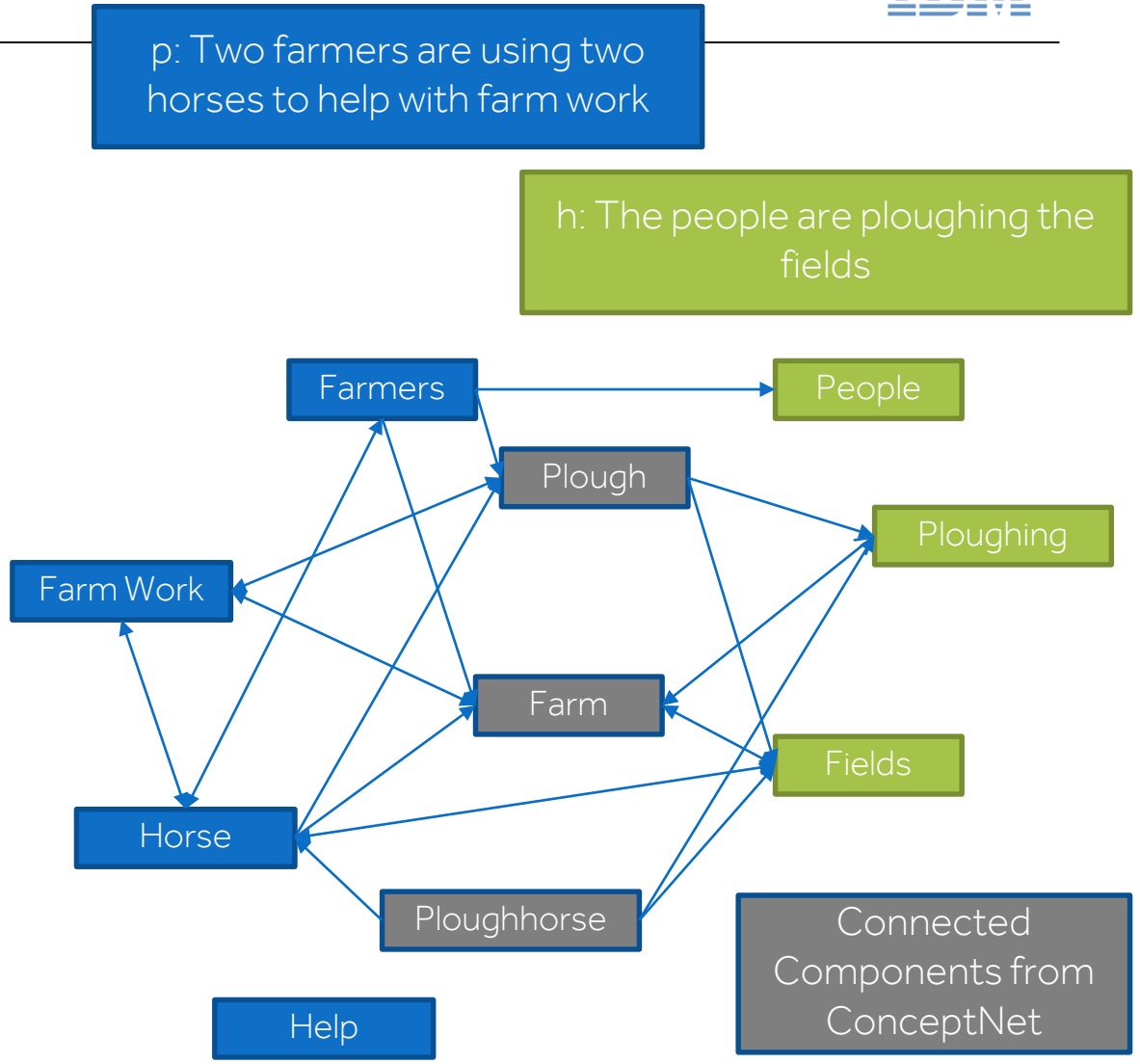


Paper Citations

Ryan Musa, Xiaoyan Wang, Achille Fokoue, Nicholas Mattei, Maria Chang, Pavan Kapanipathi, Bassem Makni, Kartik Talamadupula, Michael Witbrock. Answering Science Exam Questions Using Query Rewriting with Background Knowledge. AKBC 2019.

Natural Language Inference using External Knowledge (AAAI 2019)

- Understand the the relationship between terms using external knowledge.
  - Farmers, farm work and ploughing.
  - Working and ploughing corn.
  - Horses and ploughing.
- Evaluate multiple sources of this knowledge and their impact on entailment models.
  - DBPedia, WordNet, ConceptNet (worked best)
- We developed hybrid models that use both text and information from knowledge bases
  - Best performing system at the time of submission



Paper Citation

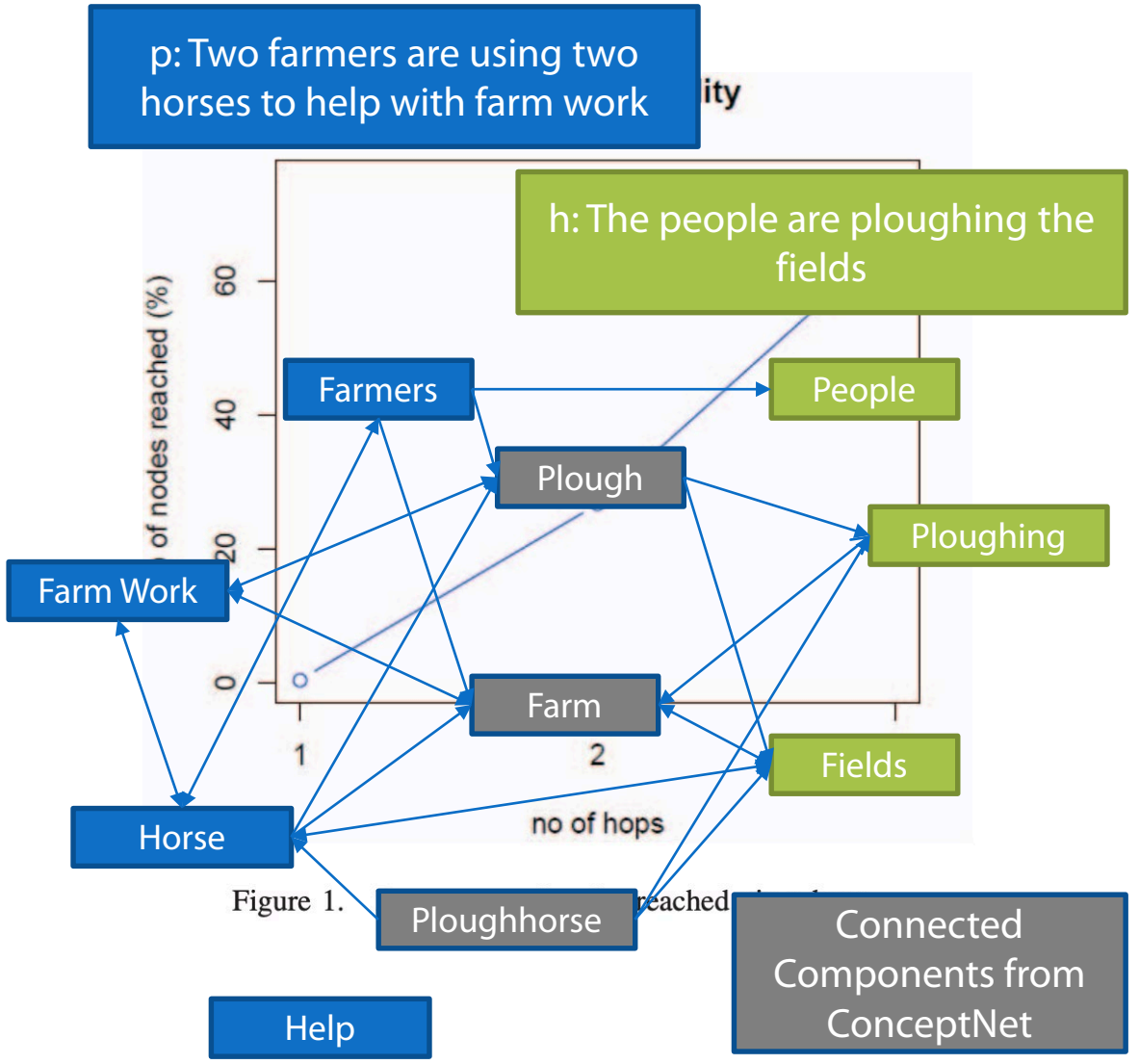
Xiaoyan Wang, Pavan Kapanipathi, Ryan Musa, Mo Yu, Kartik Talamadupula, Ibrahim Abdelaziz, Maria Chang, Achille Fokoue, Bassem Makni, Nicholas Mattei, Michael Witbrock. Improving Natural Language Inference Using External Knowledge in the Science Questions Domain. AAAI 2019.

Contextualizing Knowledge for Natural Language Processing Tasks

- Knowledge Bases/Knowledge Graphs are used to enrich unstructured text with semantics

- We have shown that harnessing external knowledge can improve performance on downstream NLP tasks
  - Complex Question Answering
  - Natural Language Inference

- Contextualizing Knowledge Graphs
  - Existing approaches: **N-hops**
  - Challenge: Optimal subgraphs/knowledge that is useful for end task
    - Relevancy, Context, and Noise



Results

- **Experimental setup:**
  - Dataset: SciTail -- NLI Dataset for Science domain question answering. Motivation: Ability of the dataset to be used for downstream QA
  - Overall Goal is to improve performance of QA systems using External Knowledge and Reasoning.
- **Results:**
  - Our system that utilizes the graph based model was the state of the art at the time of publication on Scitail dataset
  - BERT based models perform the best as of now

Model	Dev	Test
Decomp-Attn (Parikh et al. 2016)	75.4	72.3
DGEM* (Khot, Sabharwal, and Clark 2018)	79.6	77.3
DeIsTe (Yin, Roth, and Schütze 2018)	82.4	82.1
BiLSTM-Maxout (Mihaylov et al. 2018)	-	84.0
mLSTM (Wang and Jiang 2015)	88.2	84.1
Our implementation		
mLSTM (GRU)	88.5	84.2
mLSTM+Wordnet* (Chen et al. 2018)	88.8	84.3
mLSTM+Gmatch-LSTM* (**ConSeqNet**)	**89.6**	**85.2**

Table 1: Performance of entailment models on SciTail in comparison to our best model that uses match-LSTM as the text and the graph model with *Concept-Only graph* and CN-PPMI embeddings. * indicates the use of external knowledge in the approach.

IBM Research

Trail Reasoner: DRL based Theorem Prover

Team: Ibrahim Abdelaziz, Maria Chang, Cristina Cornelio, Maxwell Crouse, Achille Fokoue, Pavan Kapanipathi, Bassem Makni, Ryan Musa, Aldo Pareja, Edwin Pell, Kavitha Srinivas, Veronika Thost, Spencer Whitehead, Michael Witbrock

IBM

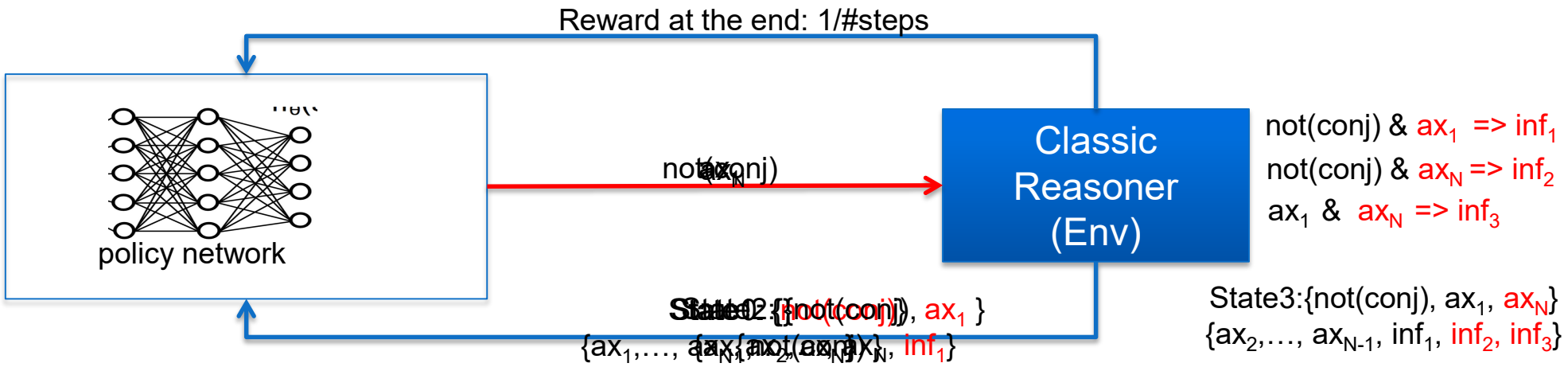
- Background: Reasoning as a search/decision problem:
  - Premise Selection: Axioms or facts at each step in a proof?
  - Rule Selection: Inference rules to apply?

- **Challenges:**
  - Reasoning in FOL or HOL is very hard (undecidable)
  - Infinite number of actions

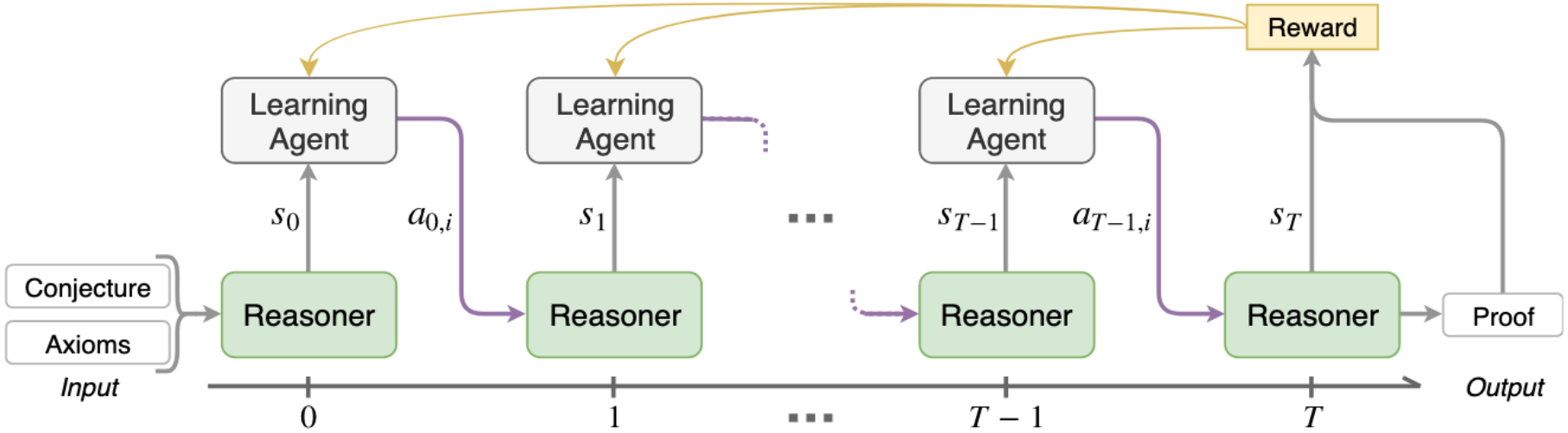
- State-of-the-art:
  - Decades of research for building reasoners based on heuristics:
    - Vampire – world champion in ATP
    - Beagle
    - E-prover, Princess …
- Goal:
  - Learn sophisticated optimization techniques applicable to a variety of logical formalisms: e.g., FOL, high order logics

RL based Proof Guidance

- Environment: Unoptimized Classical Reasoner (w/o decision making)
- Initial State:
  - Already selected premises: {}
  - Applicable Actions: (Axioms U {negated conjecture}) x Inference Rules



RL based Proof Guidance



- Reward function:
  - Reward = 1/#steps
  - Normalized reward to improve stability of training
    - The inverse of the number of steps performed using standalone reasoners
    - The best reward obtained in repeated attempts to solve the same problem

Method 1 : Herbrand Vectorizer

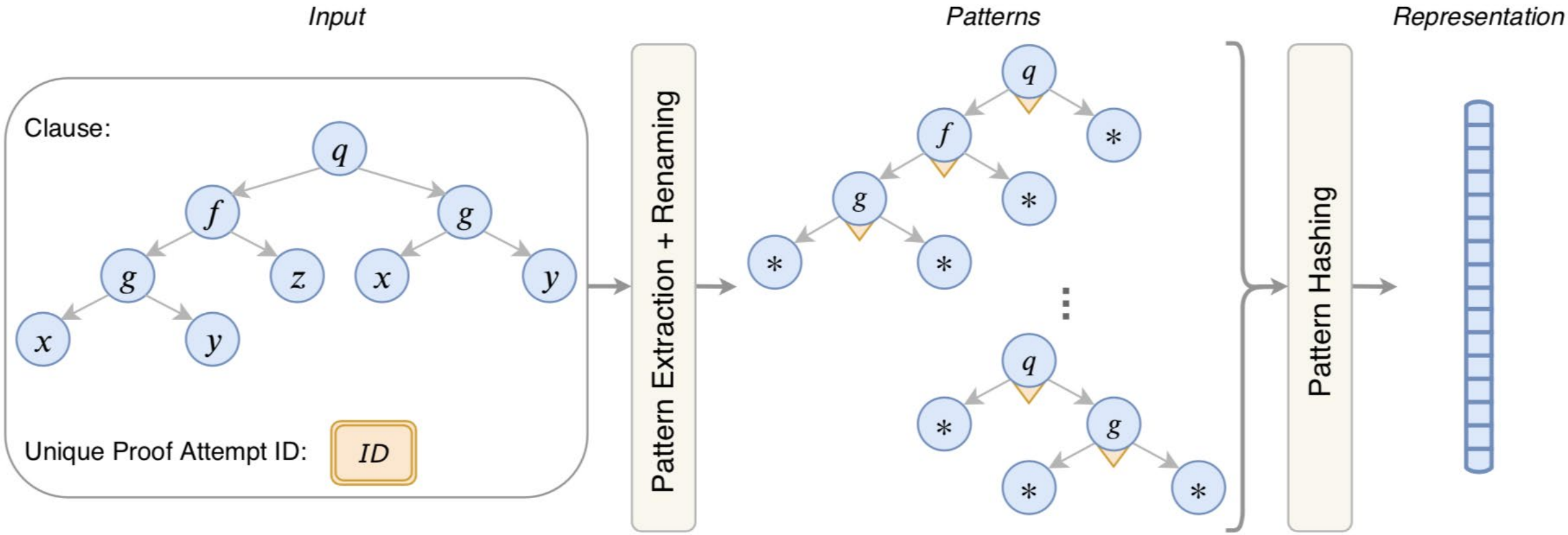


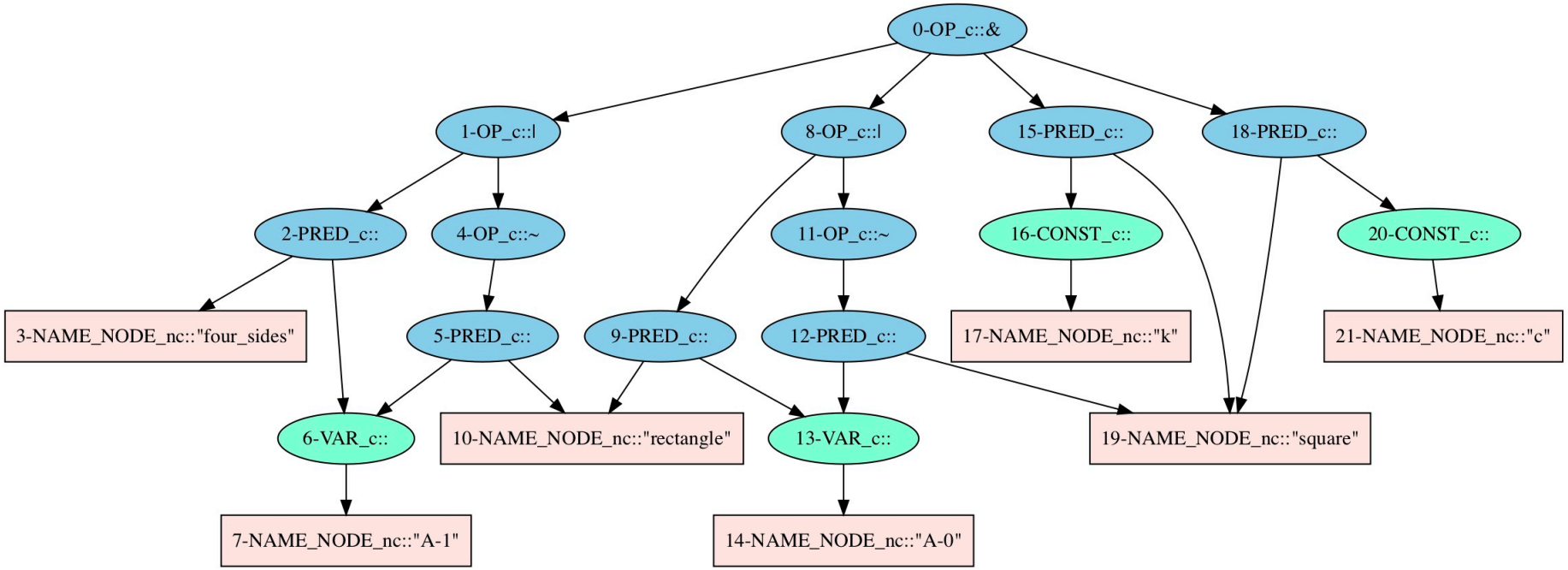
Figure 2: Overview of the vectorizer operating on the clause $q(f(g(x, y), z), g(x, y))$.

Method 1: Herbrand Template

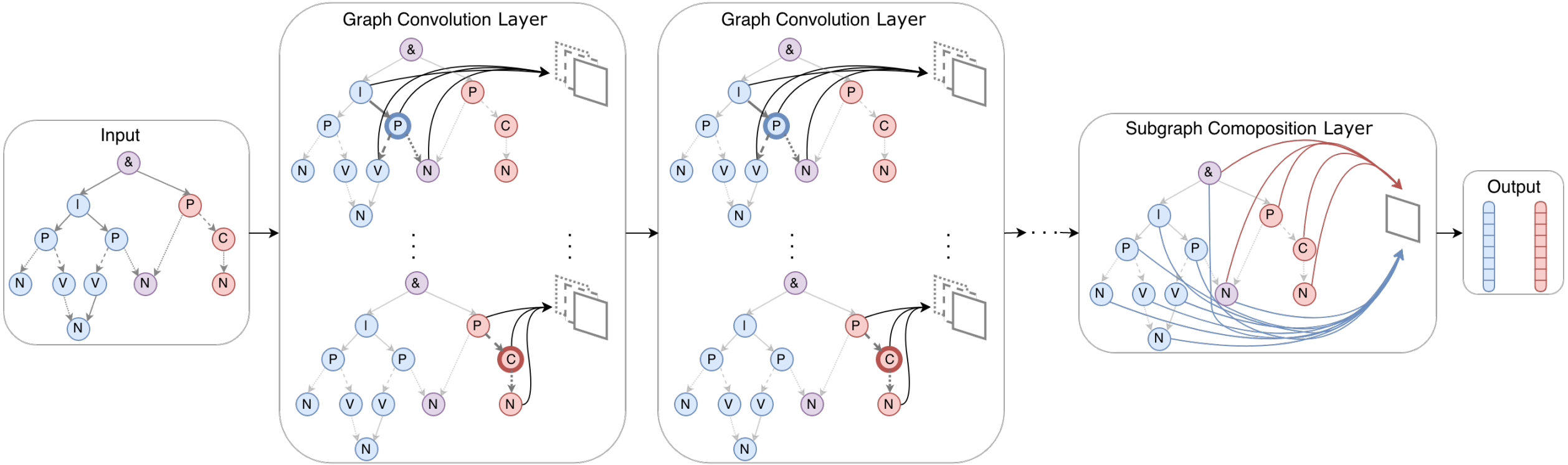
- To represent clauses, hash each of its parts into a location in a feature vector
- To represent actions, concatenate clause feature vector and one-hot vector of action type
- Advantages:
  - Easy to generate and interpret
  - Does offer some similarity comparison
- Disadvantages:
  - Not completely name invariant
  - Does not capture structure well

Method 2: Graph Convolutional Network with Subgraph Composition Layer

- To capture clause structure, represent entire set of clauses in CNF as a graph
- Use type information to enforce name invariance



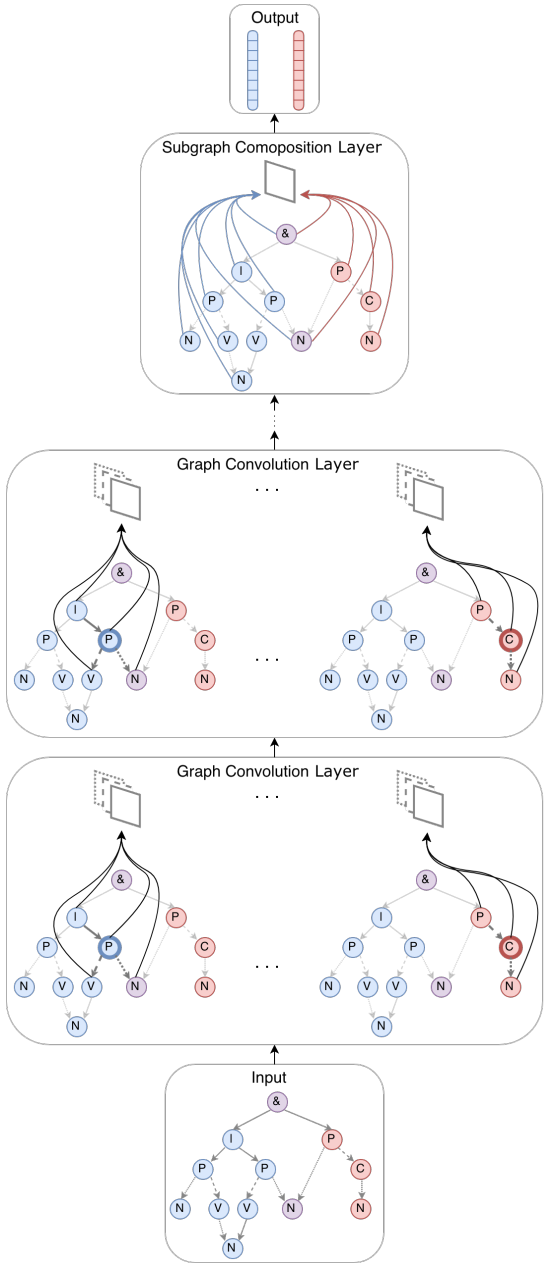
Method 2: Graph Convolutional Network with Subgraph Composition Layer



& = And, | = Or, P = Predicate, V = Variable, C = Constant, N = Name

Method 2: Graph Convolutional Network with Subgraph Composition Layer

- Advantages:
  - Name invariant and captures structure of clauses
  - Propagates information about other related clauses
- Disadvantages:
  - More expensive
  - Can be difficult to learn



Attention Based Policy Network

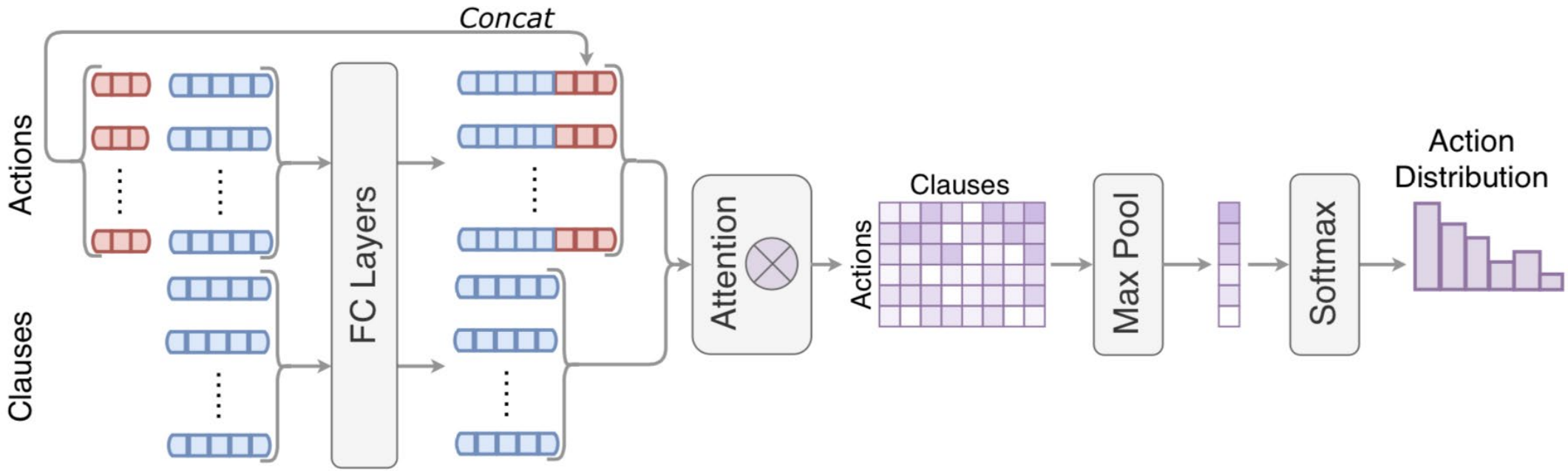


Figure 3: Policy network architecture.

Evaluation

- Dataset:
  - Mizar:
    - Real dataset of mathematical problems
    - 32K problems – 10%
  - TPTP
    - 2000 problems
- Time limit – 100 seconds per problem

- Baselines:
  - Beagle:
    - Well-established reasoner with competitive performance on ATP datasets
  - Manually optimized reasoner
    - Implements some of the common state-of-the-art optimization techniques such as subsumption checking, demodulation, and term indexing.

Effectiveness of Trail

	MIZAR	TPTP
Beagle (Fully Optimized)	61.0	32.4
Trail	57.0	24.3

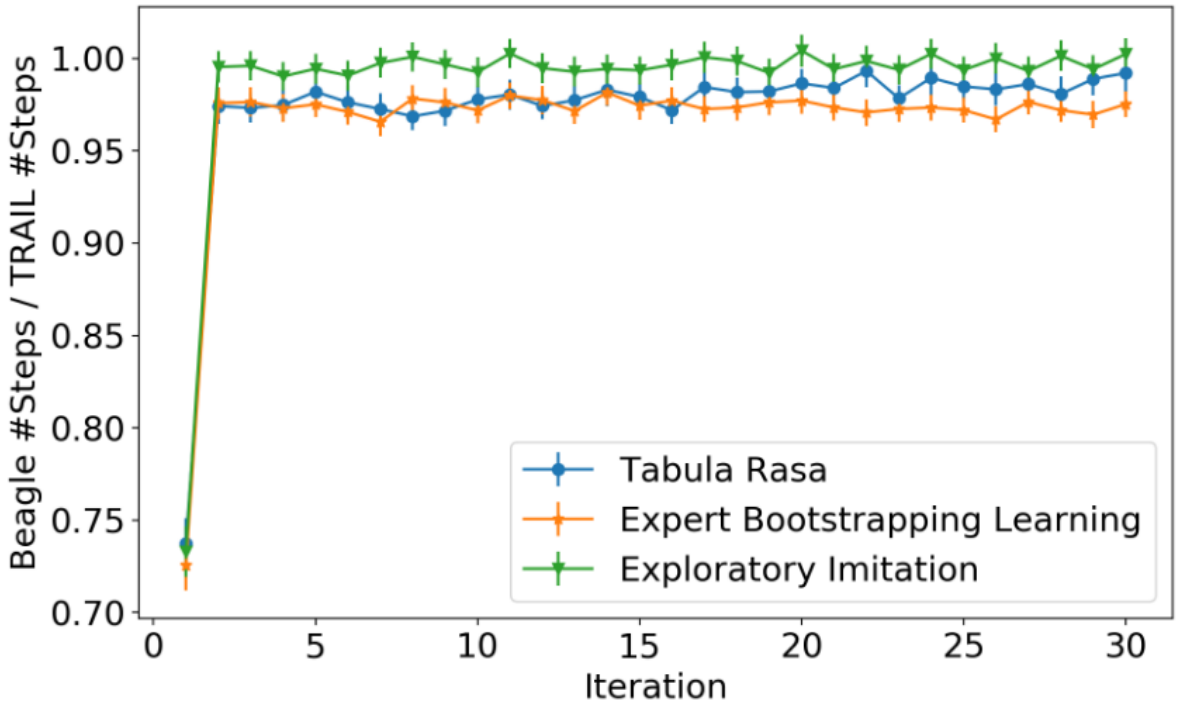
Percentage of problems solved on Mizar and TPTP datasets at testing

(Kaliszyk et al. 2018):

- RL based ATP system
- Solved only 50% in the same setting for which Vampire (heuristics-based like Beagle) solved 90% of the problems TRAIL is good.

Kaliszyk et al. "Reinforcement Learning of Theorem Proving“, NIPS (2018)

Effectiveness of Trail



Validation: Average number of steps relative to Beagle on Mizar
(with standard error bars)