

Combining learning and reasoning for Bio-AI

Nil Geisweiller

SingularityNET & OpenCog Foundations



SingularityNET



Why combining machine learning and reasoning?



Ultimate answer to overfitting

Why?

What about simulation?

Impractical without abstractions

Reasoning
⇓
Abstractions

Why?

Help learning (and reasoning)

- Reasoning for meta-learning
 - Filter relevant features
 - Guide optimization
- Learning for meta-reasoning
 - Discover inference control patterns
 - Create contextual Hebbian links

Learning & reasoning over the Bio-AtomSpace

- Learning:
 - MOSES (program evolution)
⇒ Predictive models
 - Pattern Miner (frequent pattern mining)
⇒ Discover abstractions
- Reasoning:
 - Pattern Miner
 - PLN (Probabilistic Logic Networks)
⇒ Use existing and discovered background knowledge

Example: Knowledge Base

- From GO

```
(InheritanceLink  
  (ConceptNode "GO:0000001")  
  (ConceptNode "GO:0048308"))
```

- From BioGRID

```
(EvaluationLink  
  (PredicateNode "interacts_with")  
  (SetLink  
    (GeneNode "MYPN")  
    (GeneNode "ACTN2")))
```

- From Reactome

```
(EvaluationLink  
  (PredicateNode "has_location")  
  (ListLink  
    (GeneNode "SCO2")  
    (ConceptNode "mitochondrial matrix")))
```

- From SMP

```
(MemberLink  
  (GeneNode "GPD1")  
  (ConceptNode "SMP0038938"))
```

Example: Preprocess KB

- `(ConceptNode "GO:0005575" (stv 0.960931 0.960846))`
- `(ConceptNode "GO:0036343" (stv 0.000203749 0.960846))`
- `(SubsetLink (stv 1 1)
 (ConceptNode "GO:1900454" (stv 0.000305623 0.960846))
 (ConceptNode "GO:1900452" (stv 0.000713121 0.960846)))`
- `(SubsetLink (stv 0.428571 0.017199)
 (ConceptNode "GO:1900452" (stv 0.000713121 0.960846))
 (ConceptNode "GO:1900454" (stv 0.000305623 0.960846)))`
- `(AttractionLink (stv 0.165899 0.213373)
 (ConceptNode "GO:0003013" (stv 0.0110534 0.960846))
 (ConceptNode "GO:0008015" (stv 0.00183374 0.960846)))`

Example: Pattern Miner

Discover abstractions (possibly across databases)

- GO and SMP

```
(EvaluationLink (stv 0.9998396 1)
  (PredicateNode "nisurp")
  (ListLink
    (LambdaLink
      (VariableNode "$Gene")
      (PresentLink
        (MemberLink
          (VariableNode "$Gene")
          (ConceptNode "SMP0000113"))
        (MemberLink
          (VariableNode "$Gene")
          (ConceptNode "GO:0019371"))))
      (ConceptNode "db-77891201549-p4S0oCVx7m21Nm1l"))))
```

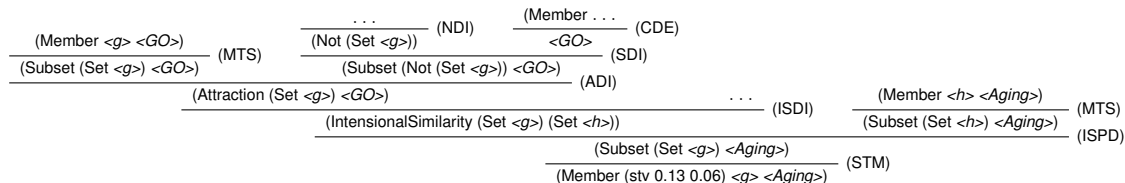
- Convert to probabilistic relationship

```
(Subset (stv 0.73 0.01)
  (ConceptNode "GO:0019371")
  (ConceptNode "SMP0000113"))
```


Example: Inference

Acronym	Rule
CDE	Concept Direct Evaluation
NDI	Negation Direct Introduction
SDI	Subset Direct Introduction
ISDI	Intensional Similarity Direct Introduction
ADI	Attraction Direct Introduction
MTS	Member To Subset
STM	Subset To Member
ISPD	Intensional Similarity Pattern Deduction

Shorthand	Term
<g>	(Gene "FCGR2B")
<h>	(Gene "ITPR3")
<GO>	(Concept "GO:0030889")
<Aging>	(Concept "Gene Expression Increase with Aging")



Status

- Moderately complex models with MOSES
- Simple patterns with Pattern Miner
 - Pattern size: 2 conjuncts
 - GO + SMP dataset: 1M atoms
 - Time: couple hours
- Short inference trees with PLN
 - Inference tree size: dozen steps
 - GO dataset: 650K atoms
 - Time: couple hours
- Focused on longevity

Difficulties

- Porting data into the atomspace (consistency checking)
- Very resource hungry (millions of atoms)
 - CPU: 1 single step can take 20+ minutes
 - RAM: 1 single step can take 64GB+
 - Need ECAN!
- Advanced forms of reasoning
 - Push PLN to its limits and beyond

To do

- Bio-AtomSpace
 - Keep importing more data
 - Experiment with more domains, COVID-19, Cancer, Regenerative Therapy
- OpenCog
 - Rule-engine
 - Multi-threaded Rule Engine
 - Back-propagate strength (only confidence so far)
 - PLN
 - Fix, add rules
 - Improve uncertainty calculations
 - Integrate ECAN
 - Integrate spatio-temporal reasoning
 - Experiment with inference control meta-learning