Axiom Selection for Large Theory Reasoning

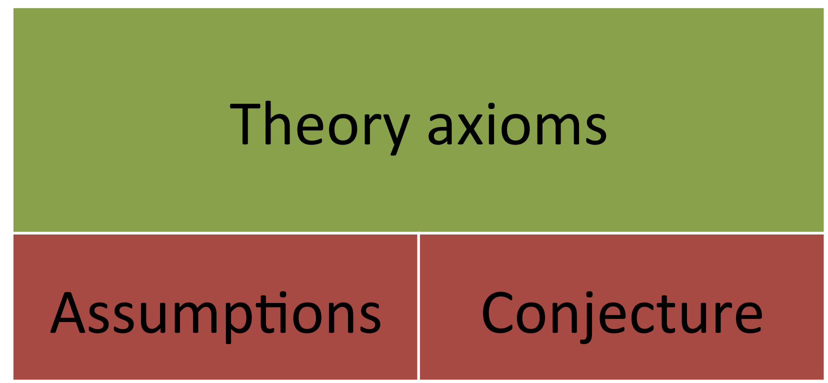
Kryštof Hoder

Andrei Voronkov

Large Theory Reasoning in First-Order Logic

- Traditional FO problems
  - Not too many axioms
  - Axiomatizations of algebras, set theory
- Large theory problems
  - Many axioms, most of them are irrelevant to the conjecture
- Axiom selection
  - attempts to remove the irrelevant and keep the important

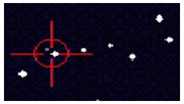
Structure of a First-Order problem



Sources of Large Theory Problems

- Ontology reasoning
  - SUMO, YAGO, CyC

  - Up to 10m axioms
  - Proofs involve few axioms, almost no equalities
- Mathematical libraries
  - Mizar Mathematical Library



  - Tens of thousands axioms
  - More complex proofs, equalities

Problems with Large Theories

- Preprocessing

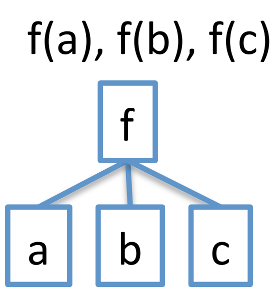
  Quadratic algorithm becomes a problem with 10m axioms

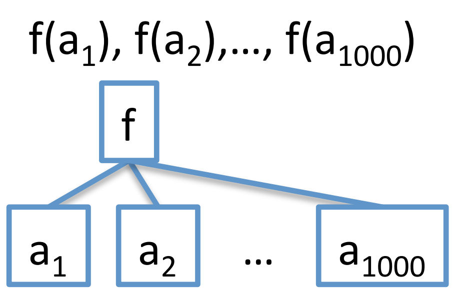
- Indexing

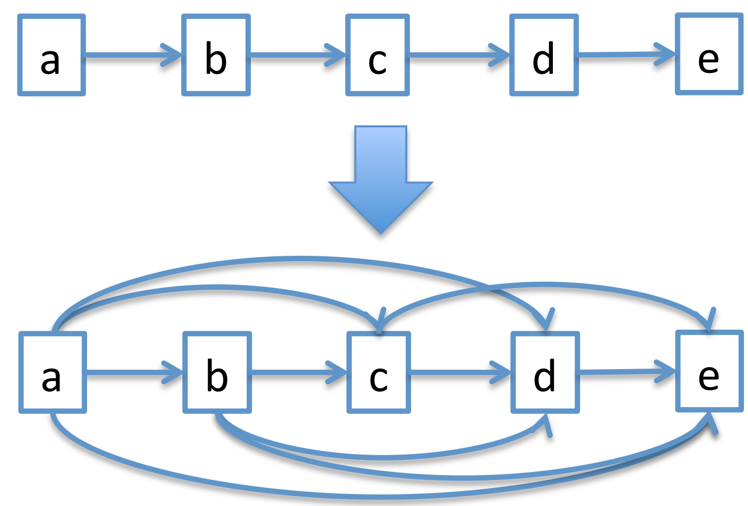
  Algorithms may assume small signature size

- Saturating irrelevant axioms

  E.g. transitive closure leads to quadratic amount of axioms (but it can be even much worse)







Performance of our Algorithm

2008

LTB	SInE-VD 0.3	SInE 0.3	Vampire-LTB 10.0	iProver-LTB 0.5c	MaLARea 0.3	E-LTB 1.0pre	EP-LTB 1.0pre	randoCoP 1.1
Attempted	150	150	150	150	150	150	150	150
Solved	88	86	76	62	52	34	32	23
Av. Time	389.45	402.85	446.71	520.06	568.66	1058.82	1125.00	1388.28
Solutions	88	85	76	0	52	0	32	23

2009

LTB	Vampire-LTB 11.0	iProver-SInE 0.7	SInE-LTB 0.4	leanCoP-SInE 2.1	E-LTB 1.1pre	EP-LTB 1.1pre
Attempted	100	100	100	100	100	100
Solved	69	67	64	35	18	18
Av. Time	24.53	76.46	75.33	110.81	63.39	77.79
Solutions	69	0	64	35	0	18

2010

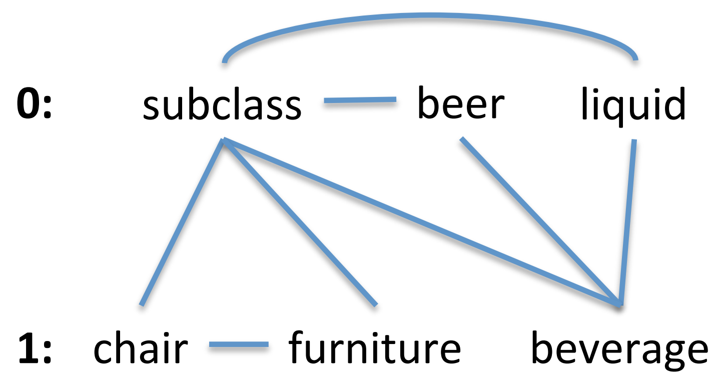
LTB	Vampire- 0.6	Currahee 0.1	iProver-SI 0.8	*Vampire-* *11.0*	iProver-E 0.6	E-LTB 1.2pre	leanCoP-S 2.2
Solved	104/150	87/150	87/150	58/150	52/150	40/15	3/150
Av. CPU Time	13.50	84.42	94.91	22.67	127.39	22.38	1.80
Av. WC Time	6.38	23.00	30.29	22.94	38.83	6.83	2.00
Solutions	94/150	0/150	0/150	58/150	0/150	0/15	3/150

Idea: Simple Relevance

- Based on mutual occurrences of symbols in axioms
- Symbol *s* is 0-relevant if it occurs in the goal
- If s is *d*-relevant and appears in axiom A, A and all symbols in A become (*d*+1)-relevant
  - *d*-relevance implies also (*d*+N)-relevance
- Select *d*-relevant axioms $d \in \{1,\ldots,\infty\}$

subclass(beverage, liquid)
subclass(beer, beverage)
subclass(chair, furniture)

? subclass(beer, liquid)

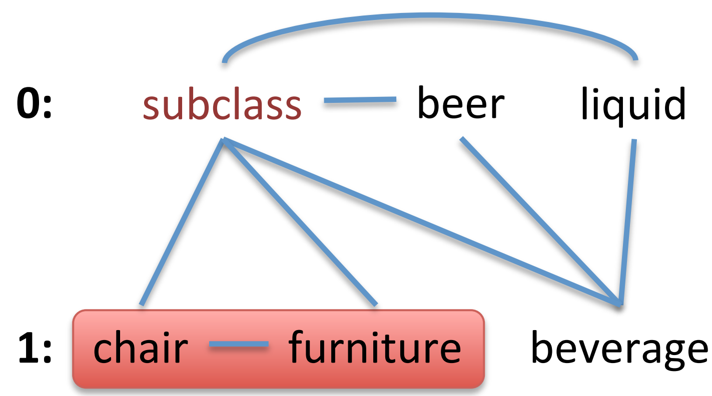


Problem: Common Symbols

- With this notion of relevance almost all axioms are usually selected
- Common symbols (such as ‘subclass’ or ‘subsumes’) make relevant otherwise unrelated symbols

subclass(beverage, liquid)
subclass(beer, beverage)
subclass(chair, furniture)

? subclass(beer, liquid)



Solution: Trigger-Based Selection

- We had

  If s is *d*-relevant and appears in A, A and all symbols in A become (*d*+1)-relevant

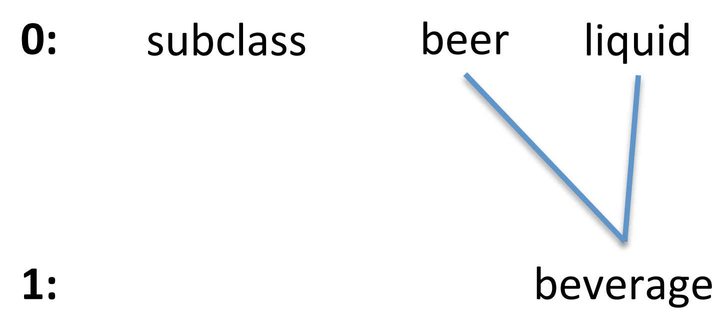
- Assuming a 'triggers' relation between symbols and axioms:

  If s is *d*-relevant and triggers A, A and all symbols in A become (*d*+1)-relevant

subclass(beverage, liquid)
subclass(beer, beverage)
subclass(chair, furniture)

? subclass(beer, liquid)

We want:



What Is a Common Symbol?

- There is no a priori information on symbol commonness
- We approximate it by number of occurrences
  - more common symbols appear in more axioms

subclass(beverage, liquid)
subclass(beer, beverage)
subclass(chair, furniture)

Occ.	Symbols
3	subclass
2	beverage
1	liquid, beer, chair, furniture

‘Triggers’ relation

Occ.	Symbols
3	subclass
2	beverage
1	liquid, beer, chair, furniture

- Should penalize common symbols
- But not ignore them completely

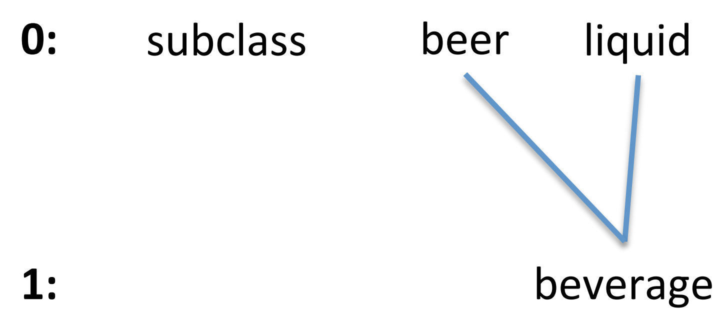
$$\text{subclass}(x, y) \wedge \text{subclass}(y, z) \rightarrow \text{subclass}(x, z)$$

- Our solution:

Only the least common symbols trigger an axiom.

1: subclass(beverage, liquid)
1: subclass(beer, beverage)
subclass(chair, furniture)

? subclass(beer, liquid)



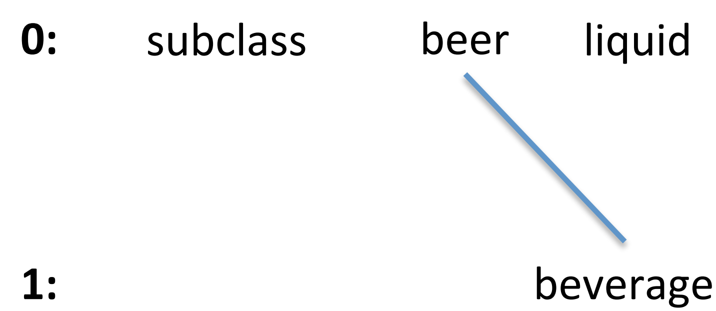
Incompleteness, Unstability

- Small difference in number of occurrences can lead to loss of important axioms

1: subclass(X,Y) ∧ subclass(Y,Z) → subclass(X,Z)
subclass(petrol,liquid)
¬subclass(stone,liquid)
2: subclass(beverage,liquid)
1: subclass(beer,beverage)
subclass(guinness,beer)

Occ.	Symbols
7	subclass
3	liquid
2	beer, beverage
1	petrol, stone, guinness

? subclass(beer,liquid)



Incompleteness, Unstability

- Small difference in number of occurrences can lead to loss of important axioms
- Or simply too little axioms may be selected
- We need a possibility to extend the 'triggers' relation

1: subclass(X,Y) ∧ subclass(Y,Z) → subclass(X,Z)
subclass(petrol,liquid)
¬subclass(stone,liquid)
subclass(beverage,liquid)
1: subclass(beer,beverage)
subclass(guinness,beer)
subclass(pilsner,beer)

Occ.	Symbols
7	subclass
3	liquid, beer
2	beverage
1	petrol, stone, guinness, pilsner

? subclass(beer,liquid)

0: subclass beer liquid

Tolerance

t=1.5:

- We had

  Only the least common symbols trigger an axiom

- Having tolerance parameter t:

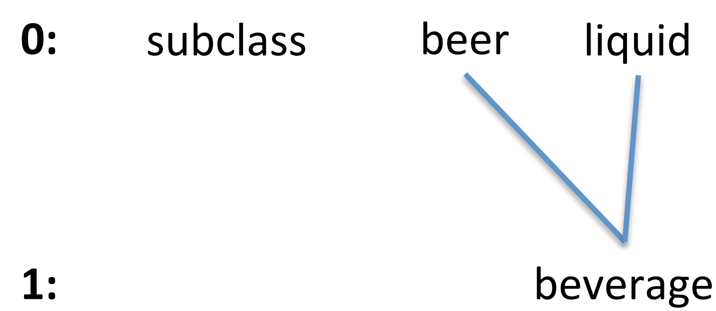
  Only symbols with *t* times more occurrences than the least common symbol trigger an axiom

- For t=∞ the selection degrades to the simple relevance

1: subclass(X,Y) ∧ subclass(Y,Z) → subclass(X,Z)
subclass(petrol,liquid)
¬subclass(stone,liquid)
2: subclass(beverage,liquid)
1: subclass(beer,beverage)
subclass(guinness,beer)
subclass(pilsner,beer)

Occ.	Symbols
7	subclass
3	liquid, beer
2	beverage
1	petrol, stone, guinness, pilsner

? subclass(beer,liquid)



Implementation

Preprocessing

- Linear in the size of theory axiomatization
- Two passes through the theory axioms:
  - Count symbol occurrences
  - Record axioms triggered by each symbol

Selection

- Linear in the size of the resulting set of axioms
  - goal + selected axioms
- Iteratively selecting *d*-relevant axioms based on (*d*-1)-relevant symbols discovered in previous iteration

---

- Preprocessing can be modified to support selection with different tolerance values

Experiments

Two parameters:

t Tolerance

d Depth limit (selects *d*-relevant axioms)

Problem sizes

problems	average size (axioms)	average size (atoms)
SUMO	298,420	323,170
CYC	3,341,990	5,328,216
Mizar	44,925	332,143

Numbers of selected axioms

SUMO:

$d\backslash t$	1.0	1.2	1.5	2.0	3.0	5.0
1	12	13	14	16	21	28
2	70	82	115	158	272	654
3	188	230	372	762	1950	5980
4	316	470	942	3021	8720	23440
5	540	979	2417	8179	22644	52241
7	1027	2708	8517	24445	54958	97481
∞	1116	8361	26959	57322	82379	107926

CYC:

$d\backslash t$	1.0	1.2	1.5	2.0	3.0	5.0
1	29	35	41	47	60	72
2	142	287	442	607	1027	1476
3	505	937	1451	2484	5311	10482
4	1784	3232	5716	11603	29963	69015
5	4432	8870	16806	37599	110186	249192
7	10698	25607	56337	150277	431875	832935
∞	36356	495360	1310965	1562064	1822427	2057597

Mizar:

$d\backslash t$	1.0	1.2	1.5	2.0	3.0	5.0
1	4903	4911	4921	4936	4973	5038
2	5296	5395	5553	5823	6427	7743
3	6118	6451	7068	8280	10841	16337
4	6893	7556	9001	12176	18300	28878
5	7432	8517	11165	16945	26842	37284
7	7897	9991	15788	26203	36507	41443
∞	8047	15987	28353	35345	39389	41762

Experiments

Solved problems

atoms	only with Sine	only without Sine	together
10,000	243	64	721
20,000	217	10	542
40,000	208	7	464
80,000	187	3	373
160,000	138	1	243
320,000	80	1	168
640,000	50	0	100
1,280,000	50	0	50
rating 1	232	25	402

Implemented in Vampire (http://vprover.org)

```
vampire --mode axiom_selection --sine_selection axioms
        --sine_tolerance t --sine_depth d
```