An ILP Formulation of Abductive Inference for Discourse Interpretation

Naoya Inoue, Kentaro Inui

Tohoku University

Abductive Discourse Interpretation

- **Abduction:** Inference of the best explanation to observations

hunting $\Rightarrow$ *get-gun*
shopping $\Rightarrow$ *go-to-store*
robbing $\Rightarrow$ *get-gun*
robbing $\Rightarrow$ *go-to-store*

get-gun *go-to-store*

- Formally,

Given

Background knowledge: B
Observations: O

Find

Hypothesis: H such that

$$H \cup B \models O$$

$$H \cup B \not\models \bot$$

B, O, H: sets of logical formulae

Abductive Discourse Interpretation

- **Abduction:** Inference of the best explanation to observations

hunting *shopping*

get-gun *go-to-store*

hunting* $\Rightarrow$ *get-gun
shopping* $\Rightarrow$ *go-to-store
robbing $\Rightarrow$ *get-gun*
robbing $\Rightarrow$ *go-to-store*

- Formally,

Given

Background knowledge: B
Observations: O

Find

Hypothesis: H such that

$$H \cup B \models O$$

$$H \cup B \not\models \perp$$

B, O, H: sets of logical formulae

Abductive Discourse Interpretation

- **Abduction:** Inference of the best explanation to observations

hunting $\Rightarrow$ *get-gun*
shopping $\Rightarrow$ *go-to-store*
robbing $\Rightarrow$ ***get-gun***
robbing $\Rightarrow$ ***go-to-store***

robbing

get-gun *go-to-store*

- Formally,

Given

Background knowledge: B
Observations: O

Find

Hypothesis: H such that

$$H \cup B \models O$$

$$H \cup B \not\models \perp$$

B, O, H: sets of logical formulae

Abductive Discourse Interpretation

- “Interpreting sentences is to find the best explanation to sentences” (Hobbs+ 93)
  – Various discourse phenomena can be handled in single unified framework

S_1: Sarah spent a long time for preparing a sales plan for a new product.
S_2: However, all her efforts were in vain.
S_3: Her company decided against putting the product on the market.

Background facts:

Understanding process:

want'(e3, x1, x4) ∧ money(x4)

Observations:

(∃e1,e2,x1,x2,x3,x4)
Sarah(x1) ∧ spend'(e1, x1, x2) ∧ long-time(x2) ∧ Rexists(e1) ∧ for(e1, e2)
∧ prepare'(e2, x1, x3) ∧ Rexists(e2) ∧ sales-plan(x3) ∧ product(x4) ∧ seg(e1, S1) ∧ seg(e2, S1)

S_1: Sarah spent a long time for preparing a sales plan for a new product.
S_2: However, all her efforts were in vain.
S_3: Her company decided against putting the product on the market.

Background facts:

Sarah(x1) ∧ spend’(e1, x1, x2) ∧ long-time(x2) ∧ Rexists(e1) ∧ for(e1, e2)
∧ prepare’(e2, x1, x3) ∧ Rexists(e2) ∧ sales-plan(x3) ∧ product(x4) ∧ seg(e1, S1) ∧ seg(e2, S1)
∧ want’(e3,x1,x4) ∧ money(x4)

y2 = x1

positive-result(e1)

Understanding process:

r = Contrast

female(y2)

negative-result(e4)

(∃e4,y1,y2,r)
however(e4) ∧ all(y1) ∧ her(y2) ∧ effort(y1) ∧ in-vain’(e4,y1) ∧ Rexists(e4) ∧ **coherent**(r, S1, S2) ∧ seg(e4, S2)
∧ rel(y1, y2)

S_1: Sarah spent a long time for preparing a sales plan for a new product.
S_2: However, all her efforts were in vain.
S_3: **Her company decided against putting the product on the market.**

Background facts:

Sarah(x1) ∧ spend'(e1, x1, x2) ∧ long-time(x2) ∧ Rexists(e1) ∧ for(e1, e2)
∧ prepare'(e2, x1, x3) ∧ Rexists(e2) ∧ sales-plan(x3) ∧ product(x4) ∧ seg(e1, S1) ∧ seg(e2, S1)
∧ want'(e3,x1,x4) ∧ money(x4) ∧ however(e4) ∧ all(y1) ∧ her(x1) ∧ effort(y1)
∧ in-vain'(e4,y1) ∧ Rexists(e4) ∧ coherent(Contrast, S1, S2) ∧ seg(e4, S2) ∧ rel(y1, x1)

z1 = x1

negative-result(e4)

Understanding process:

z3 = x4

female(z1)

negative-result(e5)

r = Resultative

(∃e5, e6, z1, z2, z3, r)
her(z1) ∧ company(z2) ∧ rel(z1, z2) ∧ decide(e5, z2, e6) ∧ **coherent**(r, S2, S3) ∧ seg(e5, S3) ∧ seg(e6, S3)
∧ against(e6) ∧ put(e6, z2, z3) ∧ product(z3) ∧ Rexists(e5) ∧ ¬Rexists(e6)

S_1: Sarah spent a long time for preparing a sales plan for a new product.
S_2: However, all her efforts were in vain.
S_3: Her company decided against putting the product on the market.

Background facts:

Sarah(x1) ∧ spend'(e1, x1, x2) ∧ long-time(x2) ∧ Rexists(e1) ∧ for(e1, e2)
∧ prepare'(e2, x1, x3) ∧ Rexists(e2) ∧ sales-plan(x3) ∧ product(x4) ∧ seg(e1, S1) ∧ seg(e2, S1)
∧ want'(e3,x1,x4) ∧ money(x4) ∧ however(e4) ∧ all(y1) ∧ her(x1) ∧ effort(y1)
∧ in-vain'(e4,y1) ∧ Rexists(e4) ∧ coherent(Contrast, S1, S2) ∧ seg(e4, S2) ∧ rel(y1, x1)
∧ her(x1) ∧ company(z2) ∧ rel(x1, z2) ∧ decide(e5, z2, e6) ∧ coherent(Resultative, S2, S3)
∧ seg(e5, S3) ∧ seg(e6, S4) ∧ against(e6) ∧ put(e6, z2, x4) ∧ product(x4) ∧ Rexists(e5) ∧ ¬Rexis

Understanding process:

- ✓ **S_1, S_2 are in contrast relation**
- ✓ **S_2, S_3 are in resultative relation**
- ✓ **Sarah wants money**
- ✓ **her = Sarah (x1)**
- ✓ **the product = new product (x4)**

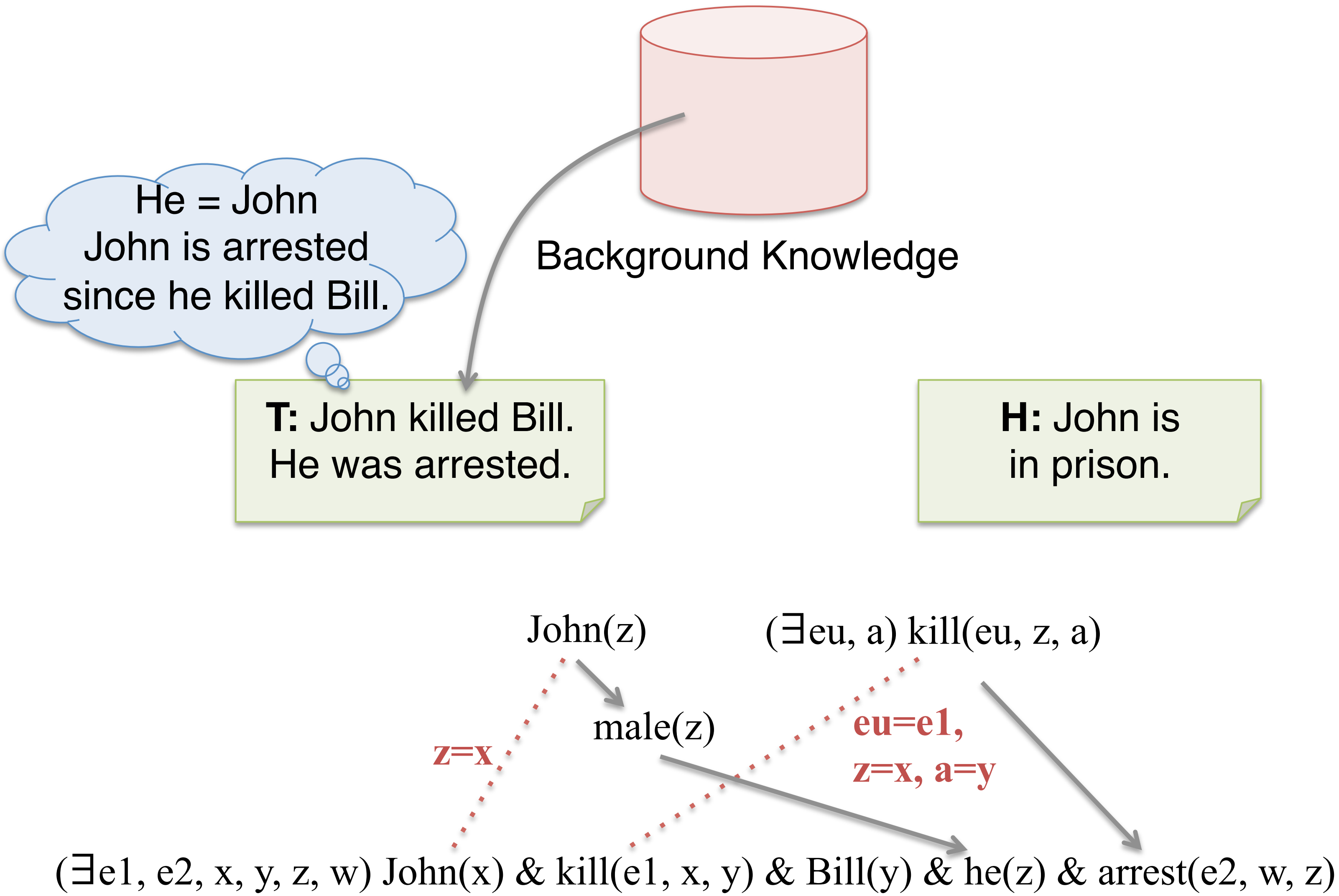
Discourse and RTE

T: John killed Bill. He was arrested.

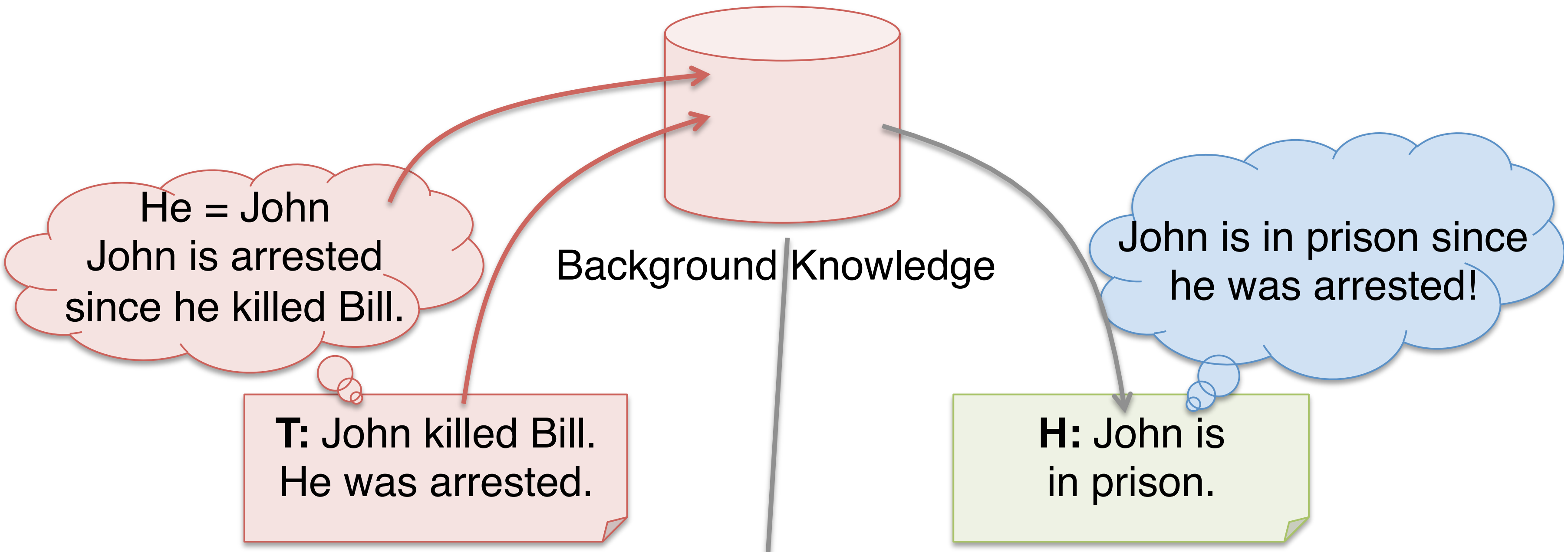
H: John is in prison.

Knowing T helps understanding H?

Discourse and RTE



Discourse and RTE



Useful for understanding H
→ T ⊃ H!!

$(\exists eu, a)$ arrest(eu, a, x)

(Ovchinnikova+ 11)

$(\exists e1, x)$ John(x) & in-prison$(e1, x)$

Motivation

- **Goal:** Discourse understanding
  - Anaphora, implicit discourse relation, etc.
- **Solution:** Abductive interpretation
  - Weighted abduction (Hobbs+ 93)

- **Problem:** Scalability of abduction
  - Most of existing work (Santos 95, Ishizuka & Matsuo 98, etc.): propositional logic
  - No tools available for large-scale inference

☑Introduction

☐Weighted Abduction

☐ILP Formulation

☐Evaluation

Weighted Abduction (Hobbs+ 93)

- "Interpreting sentences is to prove the logical forms of the sentences."
  - Merging redundancies where possible
  - Making assumptions where necessary
- Important features
  - Best proof (or *explanation, hypothesis*) is selected by assumability costs

Scheme of Weighted Abduction

AXIOM:
assumability weight is assigned

$$\text{robbing} \xRightarrow{1.4} \text{go-to-store}$$

$$\text{robbing} \xRightarrow{1.6} \text{get-gun}$$

$$\text{shopping} \xRightarrow{1.2} \text{go-to-store}$$

BACKCHAINED:
cost is propagated

UNIFICATION:
smaller cost is taken

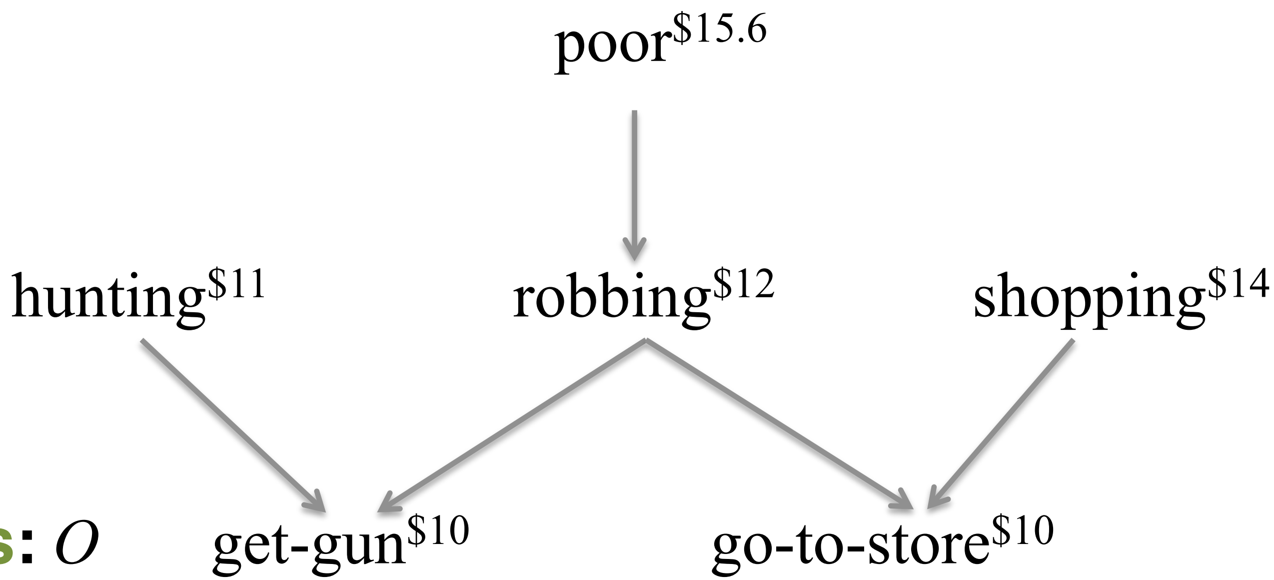
OBSERVATION:
assumability cost is assigned

Background knowledge: B

- $\textbf{robbing}^{1.2} \rightarrow \textbf{get-gun}$
- $\textbf{robbing}^{1.5} \rightarrow \textbf{go-to-store}$
- $\text{hunting}^{1.1} \rightarrow \text{get-gun}$
- $\text{shopping}^{1.4} \rightarrow \text{go-to-store}$
- $\textbf{poor}^{1.3} \rightarrow \textbf{robbing}$

Hypothesis: H

- $\{\text{hunting}^{\$11}, \text{shopping}^{\$14}\}$ ... \$25
- $\{\text{robbing}^{\$12}\}$ ... \$12
- $\{\text{poor}^{\$15.6}\}$ ... \$15.6



Observations: O

Background knowledge: B

robbing$^{1.2}$ $\longrightarrow$ get-gun

robbing$^{1.5}$ $\longrightarrow$ go-to-store

hunting$^{1.1}$ $\longrightarrow$ get-gun

shopping$^{1.4}$ $\longrightarrow$ go-to-store

poor$^{1.3}$ $\longrightarrow$ robbing

Hypothesis: H

{hunting$^{\$11}$, shopping$^{\$14}$} ... \$25

{robbing$^{\$12}$} ... \$12

{poor$^{\$15.6}$} ... \$15.6

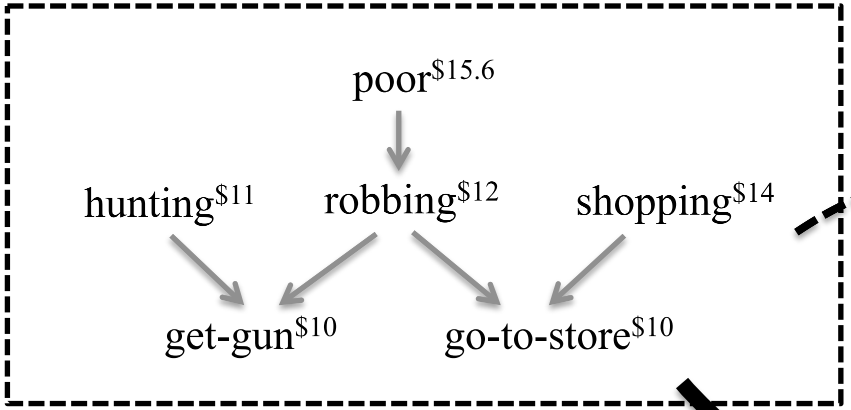
:

lanation is least-cost explanation

ate specificity is selected

Implementation Issue:
The combinatorial explosion of candidate hypotheses.

Observations: O get-gun$^{\$10}$ go-to-store$^{\$10}$



Naive approach

$H_1 = \{\text{robbing}^{\$12}\}$
$H_2 = \{\text{poor}^{\$15.6}\}$
$H_3 = \{\text{hunting}^{\$11}, \text{shopping}^{\$14}\}$
$\vdots$

$$\arg\min_i C(H_i)$$

$$C(H_i) = \sum_{h \in H_i} c(h)$$

Our solution: ILP-based Reasoning

Key ideas:

- Candidate hypotheses space is represented as **0-1 variables** and **constraints**

$$h_{\textit{get-gun}},\ h_{\textit{hunting}},\ h_{\textit{robbing}},\ \dots$$
$$r_{\textit{get-gun}},\ r_{\textit{hunting}},\ r_{\textit{robbing}},\ \dots$$
$$u_{\textit{robbing1,robbing2}}\ \dots$$
$$u_{\textit{robbing1,robbing2}} \leq h_{\textit{robbing}} + \dots$$

- Cost of hypothesis is represented as **sum of ILP variables**

$$C(H) = 10 \cdot h_{\textit{get-gun}} + 11 \cdot h_{\textit{hunting}},\ \dots$$

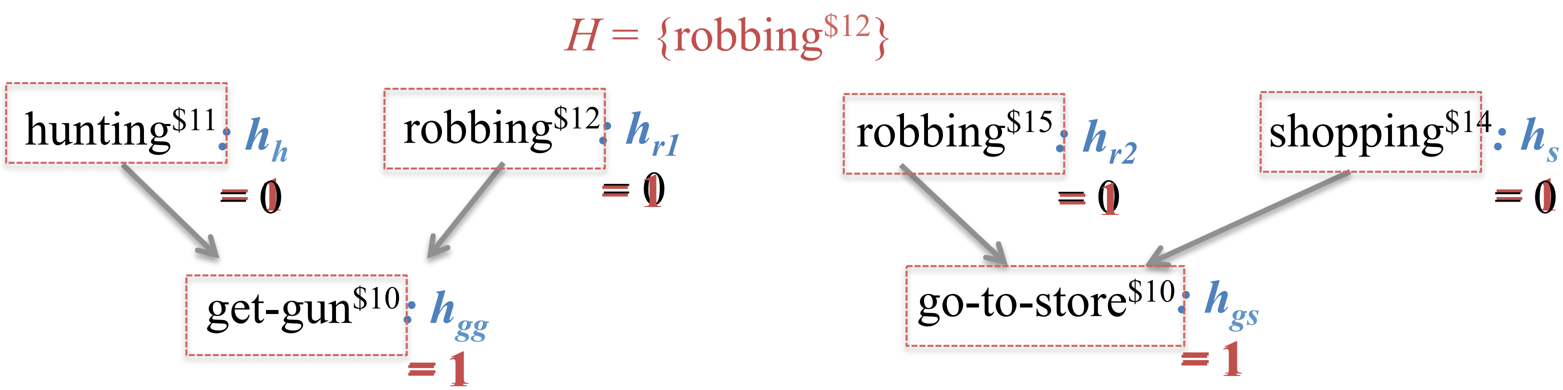
- There is efficient algorithms to find optimal assignment of 0-1 variables

$$\arg\min_{h_{get-gun}, h_{hunting}, \dots} C(H)$$

ILP formulation (h → r → u)

$$\arg\min_{h} \sum_{p \in \{p \mid p \in P, h_p = 1\}} c(p)$$

- P: set of literals potentially included in hypothesis
- h_p: 1 if literal p is included in hypothesis

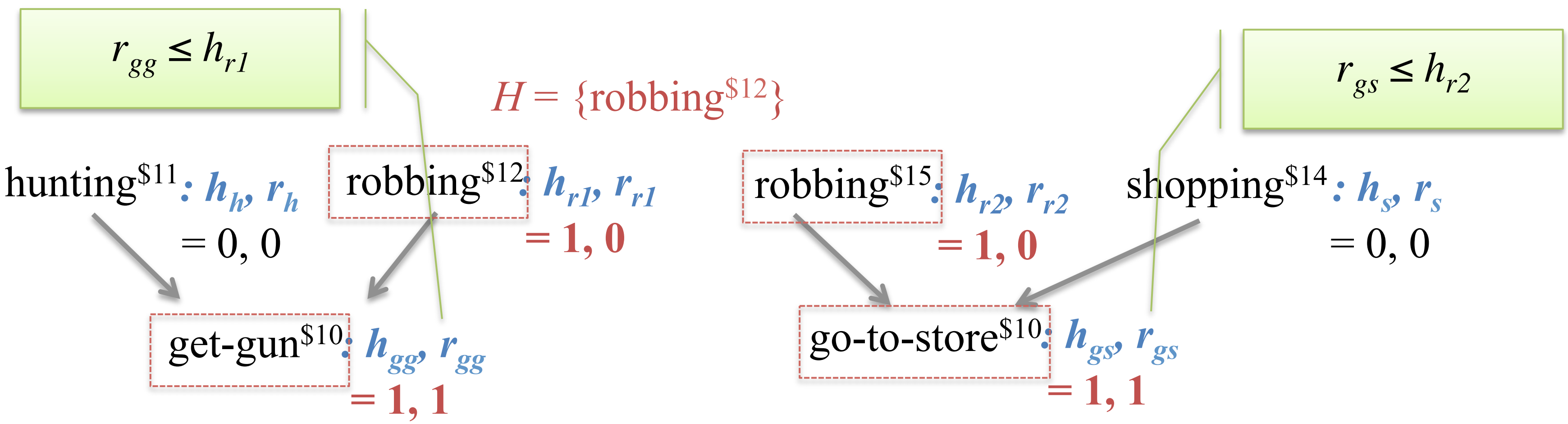


P = {get-gun, go-to-store, hunting, robbing1, ...}

ILP formulation (h → r → u)

$$\underset{h,r}{\arg\min} \sum_{p \in \{p \mid p \in P, h_p = 1, r_p = 0\}} c(p)$$

- P: set of literals potentially included in hypothesis
- h_p: 1 if literal p is included in hypothesis
- r_p: 1 if literal p doesn't pay its cost

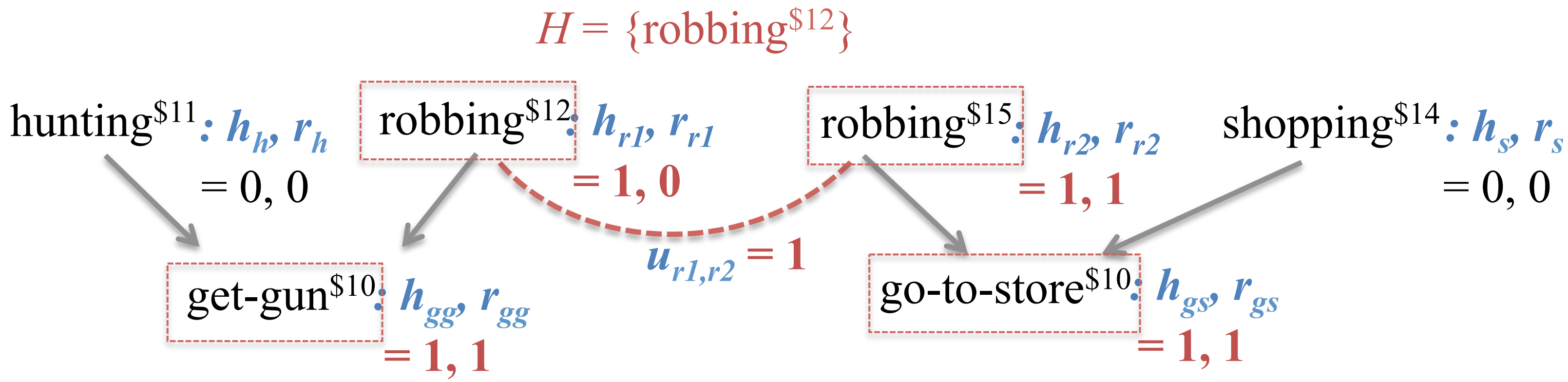


P = {get-gun, go-to-store, hunting, robbing1, ...}

ILP formulation (h → r → u)

$$\arg\min_{h,r} \sum_{p \in \{p \mid p \in P, h_p = 1, r_p = 0\}} c(p)$$

- P: set of literals potentially included in hypothesis
- h_p: 1 if literal p is included in hypothesis
- r_p: 1 if literal p doesn't pay its cost
- $u_{p,q}$: 1 if literal p is unified with literal q

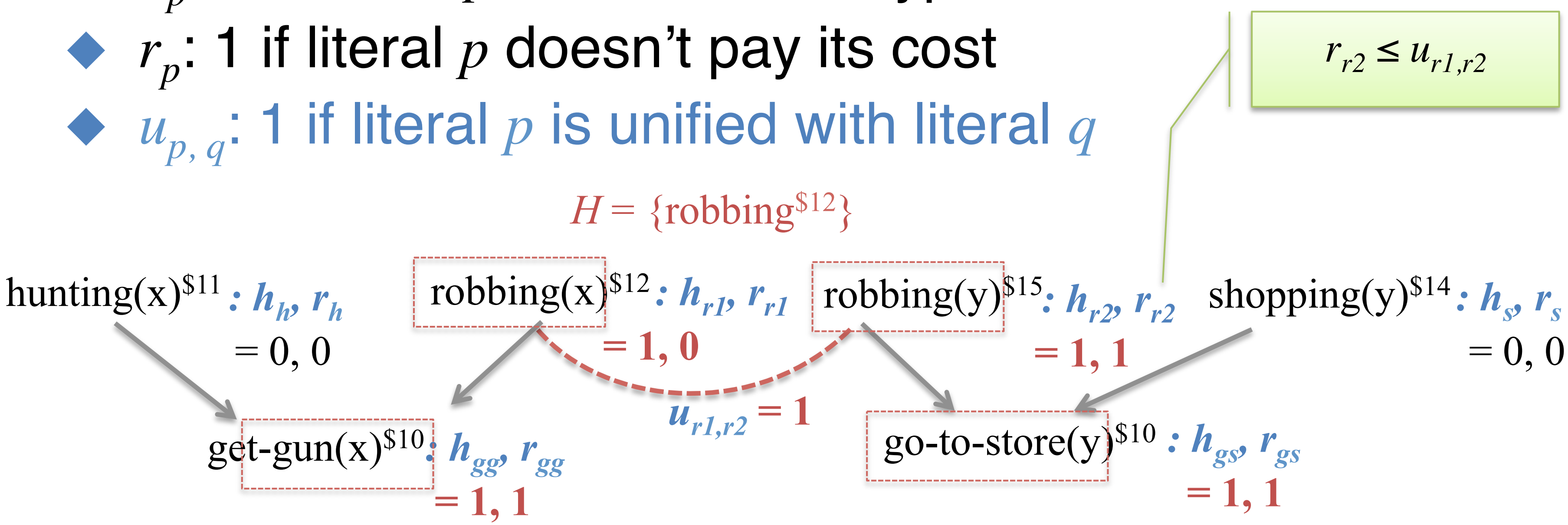


$P = \{\text{get-gun, go-to-store, hunting, robbing1, ...}\}$

ILP formulation (h → r → u)

$$\underset{h,r}{\arg\min} \sum_{p \in \{p \mid p \in P, h_p = 1, r_p = 0\}} c(p)$$

- P: set of literals potentially included in hypothesis
- h_p: 1 if literal p is included in hypothesis
- r_p: 1 if literal p doesn't pay its cost
- $u_{p,q}$: 1 if literal p is unified with literal q



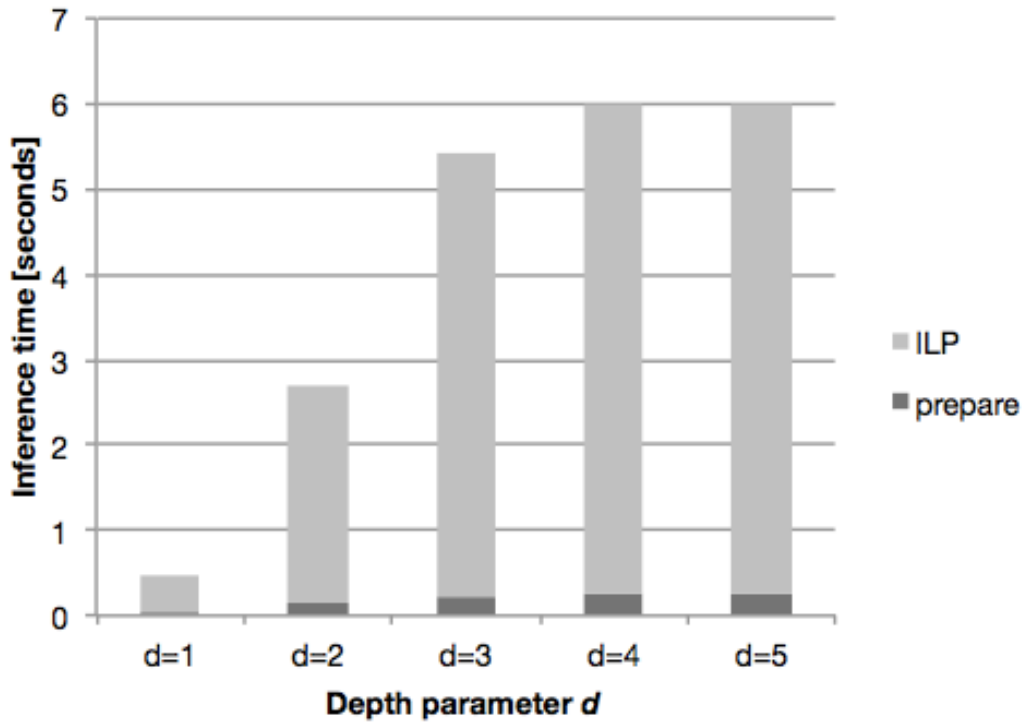
P = {get-gun, go-to-store, hunting, robbing1, ...}

Evaluation

- How scalable is our approach?
  - Depth limit v.s. Inference time
    - Depth: the depth limit of back-chaining
    - Inference time: averaged for all problems
- Dataset

Dataset	Problems	Avg. of Observation Literals	Axioms
Plan Recognition 1 (Ng & Mooney 93)	50	11.6	107
Plan Recognition 2 (Blaylock & Allen 04)	500	18.5	202
RTE (Ovchinnikova+ 11)	200	26.6	611, 903

Results on plan recognition dataset



# of candidate hypotheses $2^{|P|}$: 2^{57} 2^{90} 2^{100} 2^{103} 2^{103}

- The increase of inference time is not exponential to the number of candidate hypotheses

➢ Indicates the efficiency of our approach!

Additional Large-scale Evaluation

- **Knowledge base:** 611,903 axioms
  – WordNet 2.0 + 3.0: 556,735 axioms
  – FrameNet 1.5: 55,168 axioms
- **Observations:** 200 texts in RTE2 dataset
  – 26.6 literals per text

Setting	# of optimal solution found	# of candidate hypotheses
d = 1	97.5% (195/200)	2^{82}
d = 2	85.0% (170/200)	2^{396}
d = 3	22.0% (44/200)	2^{2061}

Timeout: 2 minutes

Summary

- Addressed the issue of scalability for abductive reasoning
- Proposed ILP-based approach to Hobbs et al. (93)'s weighted abduction
- Results of our experiments showed that:
  - our approach efficiently finds the best explanation
- Future work
  - Exploring the semantics of weights, costs
  - Applying weighted abduction to RTE, SRL, discourse relation recognition

THANK YOU!