

Online Large-margin Weight Learning for First-order Logic-based Abduction

Naoya Inoue, Kazeto Yamamoto, Yotaro Watanabe, Naoaki Okazaki, and Kentaro Inui
Tohoku University

Background

* Abduction is inference to the best explanation

Given:

- Observation: $\{get\text{-}gun(John), go\text{-}to\text{-}store(John)\}$
- Background knowledge: $(\forall x) hunt(x) \rightarrow get\text{-}gun(x)$
 $(\forall x) go\text{-}shopping(x) \rightarrow go\text{-}to\text{-}store(x)$
 $(\forall x) rob(x) \rightarrow get\text{-}gun(x)$
 $(\forall x) rob(x) \rightarrow go\text{-}to\text{-}store(x)$

Find:

- The best explanation (\equiv highest-score explanation)
- $H_1: \{hunt(John), go\text{-}shopping(John)\}$
- $H_2: \{rob(John)\}$**
- $H_3: \{rob(John), hunt(John)\}$

$score(H_1) = 4.3$
 $score(H_2) = 13.5$
 $score(H_3) = 10.8$

* Tuning of score function relies on:

- Manual tuning
- Probabilistic logic-based learning (e.g. Markov Logic Networks [Richardson & Domingos 06])

Problem: inference is not scalable; learning is even harder

* There are many applications: natural language processing, plan recognition etc.

Weight Learning for First-order Logic Abduction

* Desiderata for learning framework

- *Scalability*: computationally cheap, good results in a short time
- *Accurateness*: discriminative power
- *Usability*: learn from partially observed dataset

* Our solution

- \rightarrow *online*
- \rightarrow *large-margin* training
- \rightarrow with *latent variables!*

Designed by the user: e.g. 1 if "rob" and "gun" are included in H ; 0 otherwise.

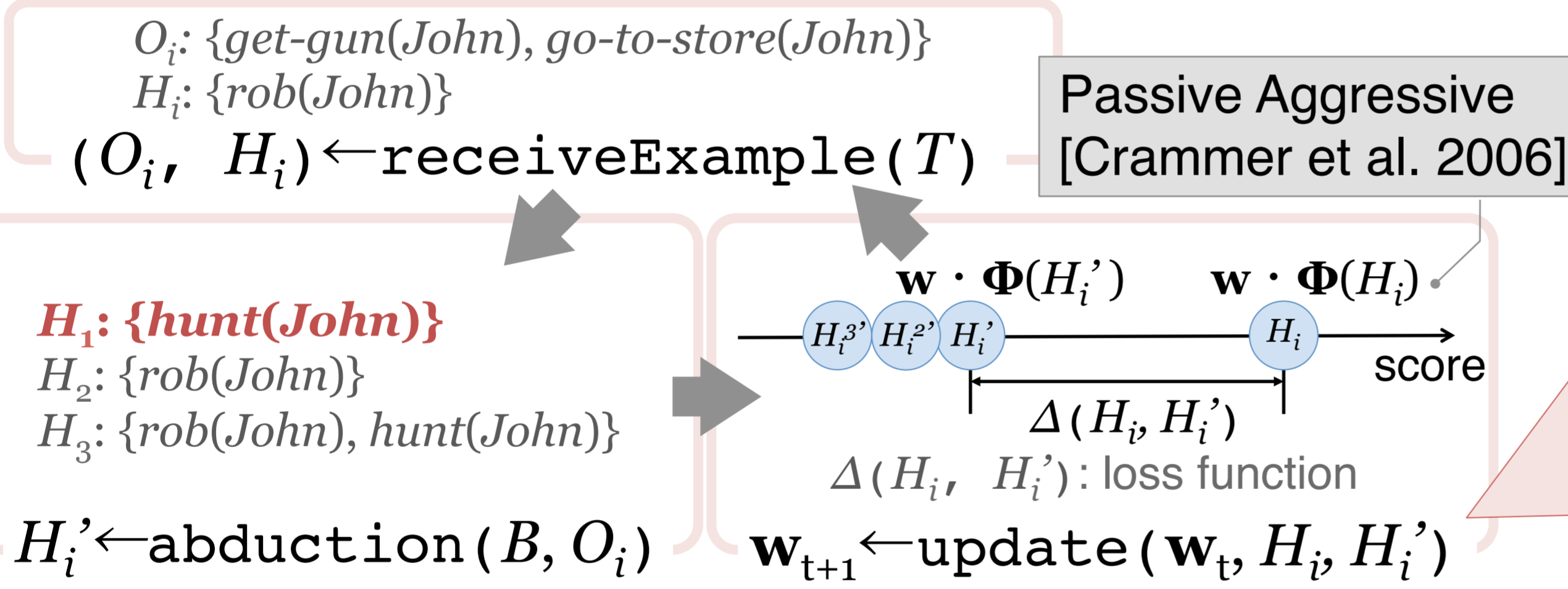
* The learning framework

(1) Assume weighted linear scoring model: $score(H; \mathbf{w}) = \mathbf{w} \cdot \Phi(H)$

\mathbf{w} : weight vector
 Φ : feature vector

(2) Learn \mathbf{w} from training examples online, following the large-margin principle:

Training examples $T = \{(O_i, H_i)\}_{i=1}^N$ $\left\{ \begin{array}{l} O_i: \text{observation (input)} \\ H_i: \text{gold-standard explanation for } O_i \text{ (output)} \end{array} \right.$



partially-specified update:

update \mathbf{w} s.t. any explanation that includes H_i is the best

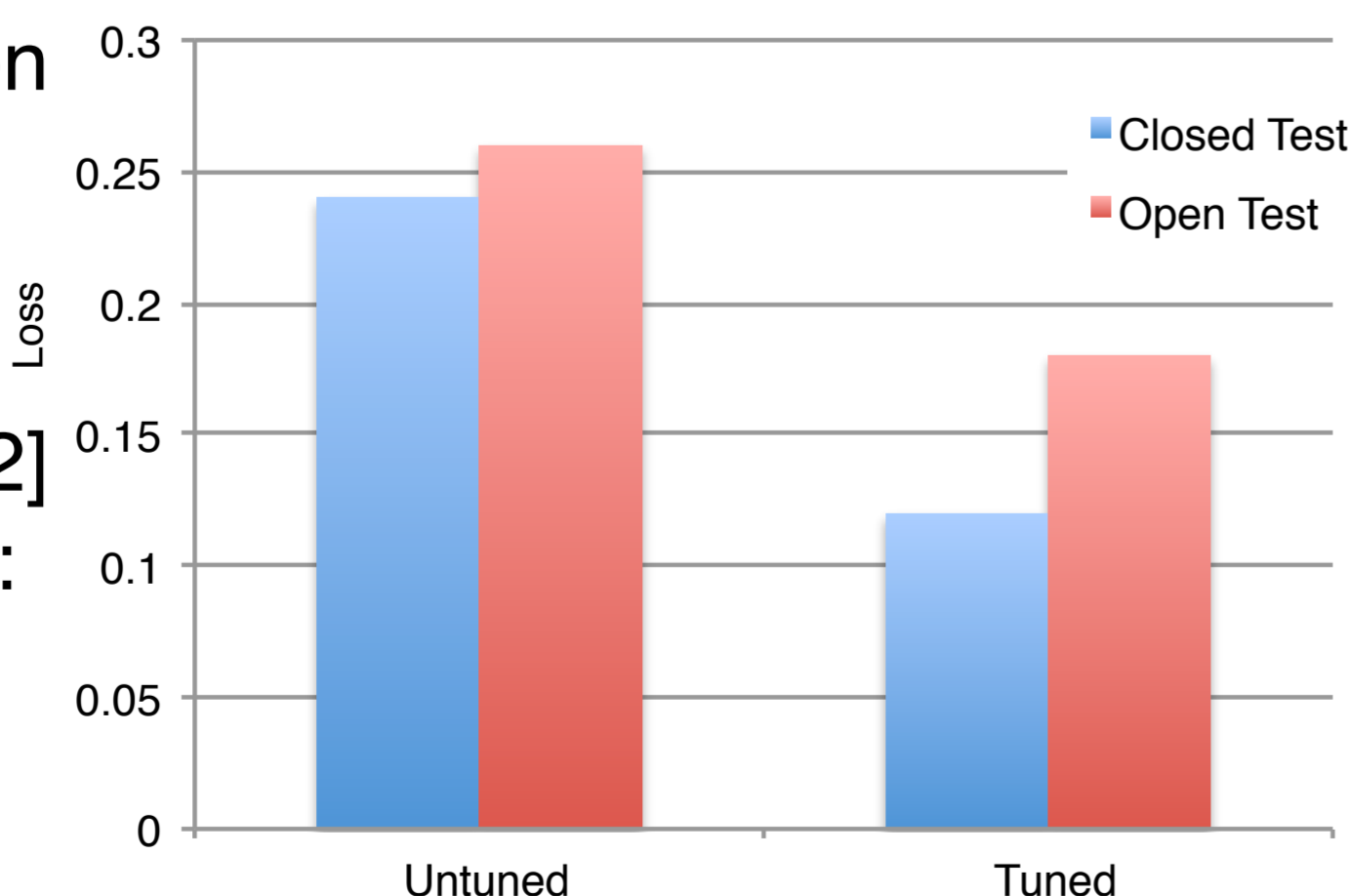
- (\times) $H_1': \{hunt(John), go\text{-}shopping(John)\}$
- (\circ) $H_2': \{rob(John)\}$
- (\circ) $H_3': \{rob(John), hunt(John)\}$

$H_i \leftarrow abduction(B, O_i)$ s.t. H_i is included [Yamamoto et al. 12]

Evaluation

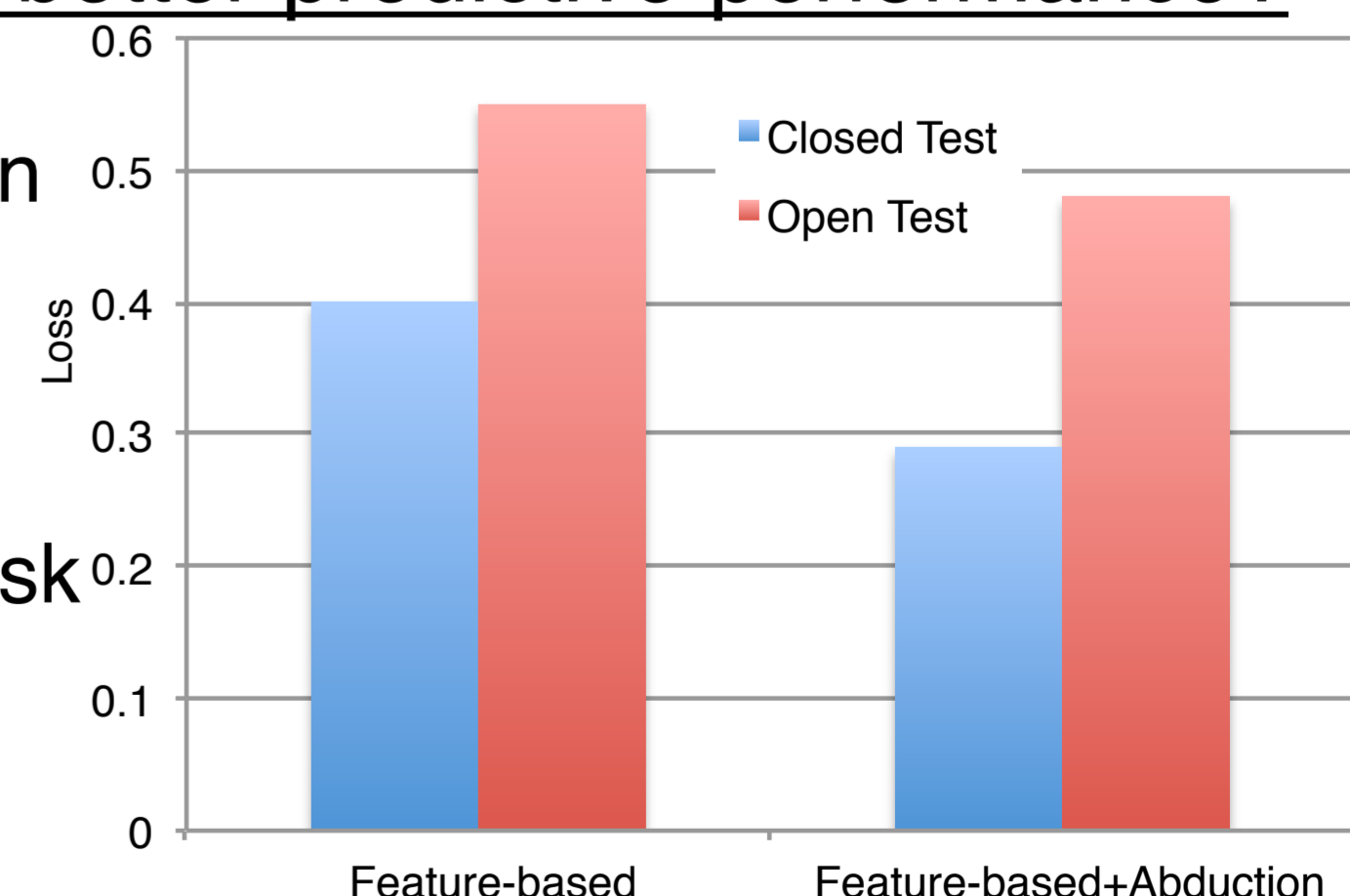
* Q1: Does learning have positive impact?

- Task: Plan recognition
- Gold-standard: plan literals
- Dataset: Ng & Mooney [92]
- Training/Testing: 25 examples
- BK: 107 axioms



* Q2: Does combining logic-based reasoning with existing classifier give better predictive performance?

- Task: Coreference resolution
- Gold-standard: equalities
- Dataset: CoNLL-2011 Shared Task
- Training/Testing: 100 documents
- BK: 300,000 axioms



Findings

- * Weight learning reduces predictive loss
- * Combining abductive reasoning with feature-based classifier reduces predictive loss
- * Generalization ability on unseen dataset

Future work

- * Use k-best explanations for update
- * Comparison with feature-based classifier exploiting world knowledge as features

This work was partially supported by Grant-in-Aid for JSPS Fellows (22-9719), Grant-in-Aid for Scientific Research (23700157, 23240018), and JST, PRESTO.