Corrected Co-training for Statistical Parsers

Rebecca Hwa
University of Pittsburgh
hwa@umiacs.umd.edu

Anoop Sarkar
Simon Fraser University
anoop@cs.sfu.ca

Miles Osborne
University of Edinburgh
miles@inf.ed.ac.uk

Mark Steedman
University of Edinburgh
steedman@inf.ed.ac.uk

ICML-2003 Workshop on the
Continuum from Labelled to Unlabelled Data

Overview

- Application: find the most likely parse for natural language sentences, using statistical methods trained on data annotated by experts
  labels are trees over strings of arbitrary length

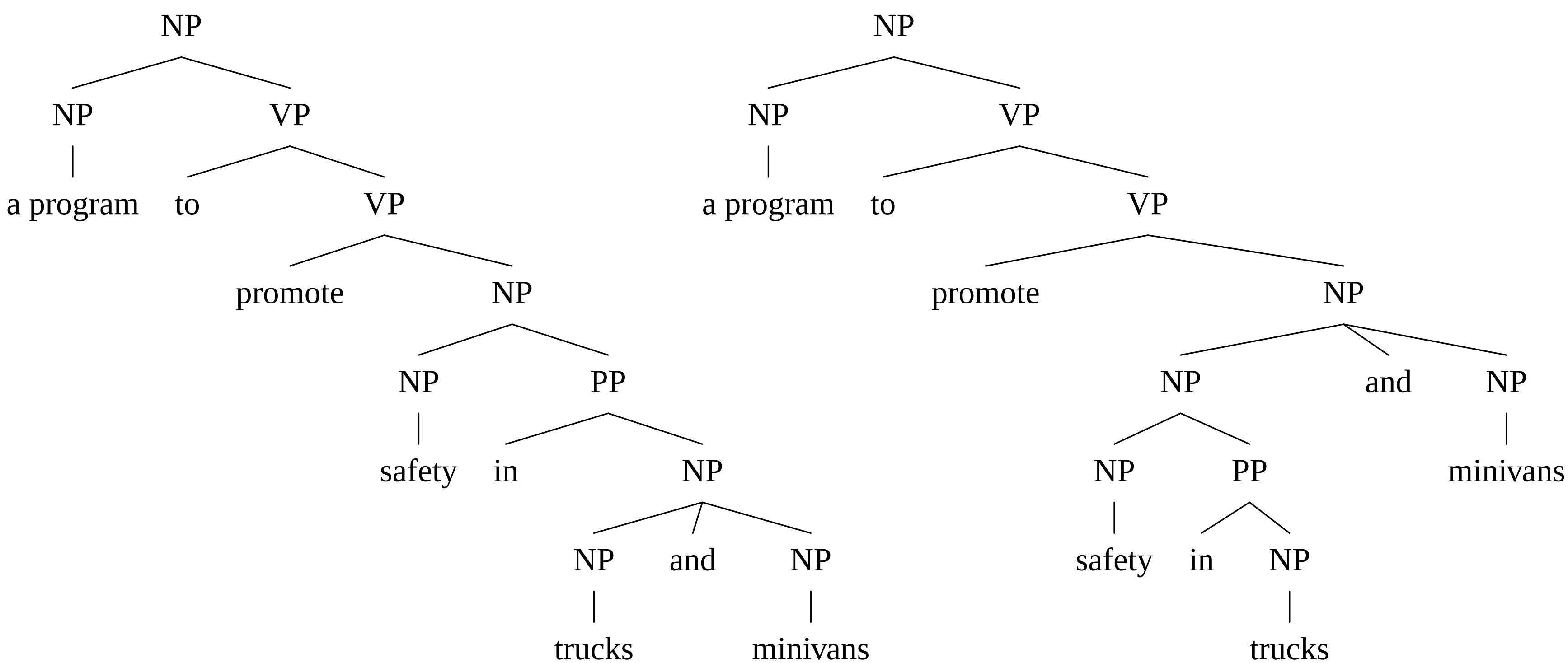
- Task: Reduce cost of annotating data by exploiting co-training in an active learning setting

- Highlights

  1. New method: one-sided corrected co-training

  2. Expts show it requires half as many manual annotation decisions

A Key Problem in Processing Language: Ambiguity: (Church and Patil 1982)



Parsing as a machine learning problem

- S = a sentence
  T = a parse tree
  A statistical parsing model defines $P(T \mid S)$

- From $P(T, S)$ find best parse: $\arg\max_T P(T \mid S)$
  models are referred to as A, B, . . . in this talk

- e.g. for PCFGs: $P(T, S) = \prod_{i=1\ldots n} P(\mathrm{RHS}_i \mid \mathrm{LHS}_i)$

- Accuracy is measured by *F-score* = $\frac{2 \cdot LP \cdot LR}{LP + LR}$
  LP = label+span precision, LR = label+span recall

Sample selection for Statistical Parsing: Our previous work

- Unsupervised selection of eligible parse trees using co-training **(EACL-2003)**
  uses two parsers A and B with independent views of the parse tree

- $P_A(T \mid S)$ or $P_B(T \mid S)$ are our **uncertainty scores** f

- Select the n parses using the ***difference*** method **(NAACL-2003)**
  select parse for A if parse score $f_B >$ score f_A by some threshold n
  select parse for B if parse score $f_A >$ score f_B by some threshold n

Single-learner sample selection for statistical parsing

Initialize:

$L_A^0 \leftarrow L.$

$M_A^0 \leftarrow Train(A, L_A^0)$

Loop:

$U^i \leftarrow$ Add unlabeled sentences from U

M_A^i parses U^i, assigns uncertainty scores f

Select the n parses $\{P_A\}$ with highest f scores,

and remove them from the unlabeled pool

Ask a person to correct $\{P_A\}$

$L_A^{i+1} \leftarrow L_A^i \cup Corrected(\{P_A\})$

$M_A^{i+1} \leftarrow Train(A, L_A^{i+1})$

Co-training for statistical parsing (unsupervised, no sample selection)

Initialize:

$L_A^0 \leftarrow L_B^0 \leftarrow L$
$M_A^0 \leftarrow Train(A, L_A^0)$
$M_B^0 \leftarrow Train(B, L_B^0)$

Loop:

$U^i \leftarrow$ Add unlabeled sentences from U
M_A^i and M_B^i parse U^i and assign scores f_A and f_B
Select new parses $\{P_A\}$ and $\{P_B\}$ according to S
$L_A^{i+1} \leftarrow L_A^i \cup \{P_B\}$
$L_B^{i+1} \leftarrow L_B^i \cup \{P_A\}$
$M_A^{i+1} \leftarrow Train(A, L_A^{i+1})$
$M_B^{i+1} \leftarrow Train(B, L_B^{i+1})$

Corrected co-training (co-testing)

Initialize:

$L_A^0 \leftarrow L_B^0 \leftarrow L$

$M_A^0 \leftarrow Train(A, L_A^0)$

$M_B^0 \leftarrow Train(B, L_B^0)$

Loop:

$U^i \leftarrow$ Add unlabeled sentences from U

M_A^i and M_B^i parse U^i and assign scores f_A and f_B

Select new parses $\{P_A\}$ and $\{P_B\}$ according to S

$L_A^{i+1} \leftarrow L_A^i \cup$ Corrected($\{P_B\}$)

$L_B^{i+1} \leftarrow L_B^i \cup$ Corrected($\{P_A\}$)

$M_A^{i+1} \leftarrow Train(A, L_A^{i+1})$

$M_B^{i+1} \leftarrow Train(B, L_B^{i+1})$

One-sided corrected co-training

Initialize:

$L_A^0 \leftarrow L_B^0 \leftarrow L$

$M_A^0 \leftarrow Train(A, L_A^0)$

$M_B^0 \leftarrow Train(B, L_B^0)$

Loop:

$U^i \leftarrow$ Add unlabeled sentences from U

M_A^i and M_B^i parse U^i and assign scores f_A and f_B

Select new parses $\{P_A\}$ and $\{P_B\}$ according to S

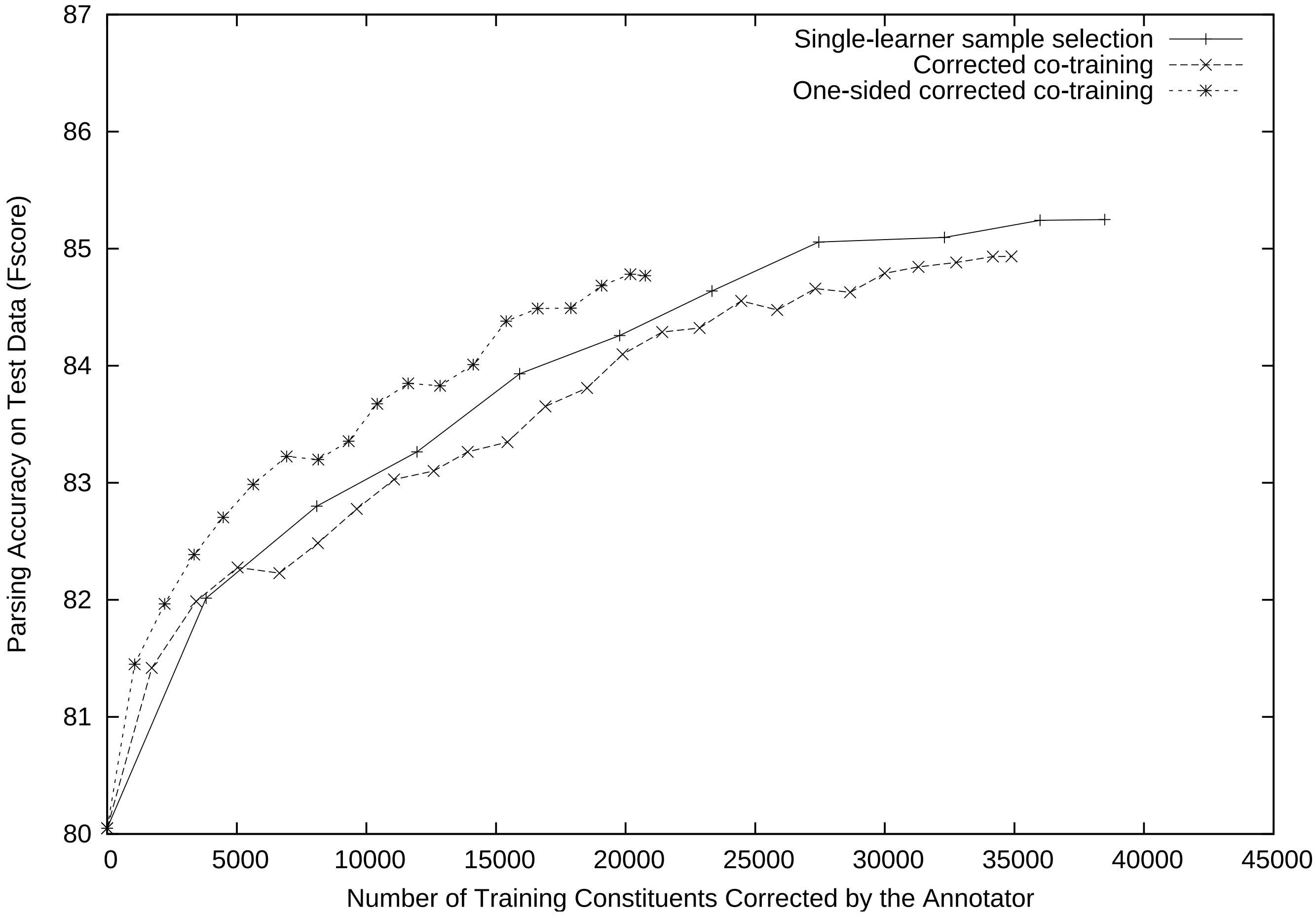
$L_A^{i+1} \leftarrow L_A^i \cup$ Corrected($\{P_B\}$)

$L_B^{i+1} \leftarrow L_B^i \cup \{P_A\}$

$M_A^{i+1} \leftarrow Train(A, L_A^{i+1})$

$M_B^{i+1} \leftarrow Train(B, L_B^{i+1})$

Treebank annotation: reducing number of constituents corrected by “humans”



Summary

- Application: find the most likely parse for natural language sentences, using statistical methods trained on data annotated by experts

- Task: Reduce cost of annotating data by exploiting co-training

- Highlights

  1. New method: one-sided corrected co-training

  2. Expts show it requires half as many manual annotation decisions