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 $P(T | S)$

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

- e.g. for PCFGs: $P(T, S) = \prod_{i=1...n} P(RHS_i | LHS_i)$

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

\[
L_A^0 \leftarrow L.
\]

\[
M_A^0 \leftarrow \text{Train}(A, L_A^0)
\]

**Loop:**

\[
U^i \leftarrow \text{Add unlabeled sentences from } U
\]

\[
M_A^i \text{ parses } U^i, \text{ assigns uncertainty scores } f
\]

Select the \( n \) parses \( \{P_A\} \) with highest \( f \) scores, and remove them from the unlabeled pool

Ask a person to correct \( \{P_A\} \)

\[
L_A^{i+1} \leftarrow L_A^i \cup \text{Corrected}(\{P_A\})
\]

\[
M_A^{i+1} \leftarrow \text{Train}(A, L_A^{i+1})
\]
Co-training for statistical parsing (unsupervised, no sample selection)

Initialize:

\[
L^0_A \leftarrow L^0_B \leftarrow L \\
M^0_A \leftarrow \text{Train}(A, L^0_A) \\
M^0_B \leftarrow \text{Train}(B, L^0_B)
\]

Loop:

\[
U^i \leftarrow \text{Add unlabeled sentences from } U \\
M^i_A \text{ and } M^i_B \text{ parse } U^i \text{ and assign scores } f_A \text{ and } f_B \\
\text{Select new parses } \{P_A\} \text{ and } \{P_B\} \text{ according to } S \\
L^{i+1}_A \leftarrow L^i_A \cup \{P_B\} \\
L^{i+1}_B \leftarrow L^i_B \cup \{P_A\} \\
M^{i+1}_A \leftarrow \text{Train}(A, L^{i+1}_A) \\
M^{i+1}_B \leftarrow \text{Train}(B, L^{i+1}_B)
\]
Corrected co-training (co-testing)

Initialize:

\[ L_0^A ← L_0^B ← L \]
\[ M_0^A ← Train(A, L_0^A) \]
\[ M_0^B ← Train(B, L_0^B) \]

Loop:

\[ U^i ← \text{Add unlabeled sentences from } U \]
\[ M_i^A \text{ and } M_i^B \text{ parse } U^i \text{ and assign scores } f_A \text{ and } f_B \]
Select new parses \( \{P_A\} \) and \( \{P_B\} \) according to \( S \)
\[ L_{i+1}^A ← L_i^A \cup \text{Corrected}(\{P_B\}) \]
\[ L_{i+1}^B ← L_i^B \cup \text{Corrected}(\{P_A\}) \]
\[ M_{i+1}^A ← Train(A, L_{i+1}^A) \]
\[ M_{i+1}^B ← Train(B, L_{i+1}^B) \]
One-sided corrected co-training

Initialize:
\[
\begin{align*}
L_0^A &\leftarrow L_0^B \leftarrow L \\
M_0^A &\leftarrow \text{Train}(A, L_0^A) \\
M_0^B &\leftarrow \text{Train}(B, L_0^B)
\end{align*}
\]

Loop:
\[
\begin{align*}
U^i &\leftarrow \text{Add unlabeled sentences from } U \\
M_i^A \text{ and } M_i^B &\text{ parse } U^i \text{ and assign scores } f_A \text{ and } f_B \\
\text{Select new parses } \{P_A\} \text{ and } \{P_B\} &\text{ according to } S \\
L_{i+1}^A &\leftarrow L_i^A \cup \text{Corrected}(\{P_B\}) \\
L_{i+1}^B &\leftarrow L_i^B \cup \{P_A\} \\
M_{i+1}^A &\leftarrow \text{Train}(A, L_{i+1}^A) \\
M_{i+1}^B &\leftarrow \text{Train}(B, L_{i+1}^B)
\end{align*}
\]
Treebank annotation: reducing number of constituents corrected by “humans”

![Graph showing parsing accuracy on test data vs. number of training constituents corrected by the annotator. The graph includes lines for single-learner sample selection, corrected co-training, and one-sided corrected co-training.](image)
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