Reranking the Berkeley and Brown Parsers

Mark Johnson and Engin Ural
Brown University
Macquarie University

June 2010

The Brown and the Berkeley parsers

- Both state-of-the-art, PCFG-based, generative parsers
- *Brown parser:*
  - conditions on a wide variety of manually-chosen information
  - simple training procedure, hand-designed smoothing
- *Berkeley parser:*
  - split-merge procedure learns refined non-terminals
  - complex but fully automatic training procedure

$\Rightarrow$ The parsers are *very different from each other*

See: Charniak and Johnson (2005), Petrov et al (2006)

MACQUARIE UNIVERSITY

Reranking the *n*-best parser output

- Reranking rescores the *n*-best trees produced by a parser
  - incorporates features difficult to use in generative models
  - discriminatively trained MaxEnt model with L2 regularisation
- Research questions:
  - will reranking work with the Berkeley parser?
  - if it does work, will the same features be most useful?
  - can we rerank the *combined* *n*-best trees of both the Brown and Berkeley parsers?
- Relevant previous work: Zhang et al (2009)
  - also combine *n*-best lists from Brown and Berkeley parsers
  - only use a small set of reranker features
  - their results are consistent with results reported here
  - also describe experiments using *self-trained* reranking parser

See: Collins and Koo (2005), Charniak and Johnson (2005), McCloskey et al (2006)

MACQUARIE UNIVERSITY

Experimental setup

- Brown parser run "out of the box"
- Berkeley trained with 6 splits, parsing in "accurate" mode
- Reranker training data consisted of PTB sections 2–21
  - 50-best parses produced using 20-fold cross-validation procedure
- Sections 22, 23 and 24 parsed using "out of the box" 50-best parser
- In order to avoid overtraining on section 23:
  - Folds 1–18 used as *main training data*
  - Folds 19 and 20 used as *development data*
  - PTB section 22 used as *test data*

See: Collins and Koo (2005)

MACQUARIE UNIVERSITY

Reranker features

- "Standard" features come "out of the box" with reranker
  - are probably tuned to Brown parser
- "Extended" features include more features that might help Berkeley parser
  - e.g., features that include heads, governors, head-to-head dependencies, etc.

	Reranker features	
	standard	extended
Number of feature super-classes	14	20
Number of feature classes	90	162
Number of features	1,333,950	4,256,553

Super-classes in extended feature set (1)

Parser: an indicator feature indicating which parsers generated this parse,

RelLogP: the log probability of this parse according to each parser,

InterpLogCondP: an indicator feature based on the binned log conditional probability according to each parser,

RightBranch: an indicator function of each node that lies on the right-most branch of the parse tree,

Heavy: an indicator function based on the size and location of each nonterminal (designed to identify the locations of "heavy" phrases),

LeftBranchLength: an indicator function of the binned length of each left-branching chain,

RightBranchLength: an indicator function of the binned length of each right-branching chain,

Super-classes in extended feature set (2)

Rule: an indicator function of parent and children categories, optionally with head POS annotations,

NNGram: and indicator function of parent and *n*-gram sequences of children categories, optionally head annotated, inspired by the *n*-gram rule features described by Collins and Koo

Heads: an indicator function of “head-to-head” dependencies,

SynSemHeads: an indicator function of the pair of syntactic (i.e., functional) and semantic (i.e., lexical) heads of each non-terminal,

RBContext: an indicator function of how much each subtree deviates from from right-branching,

SubjVerbAgr: an indicator function of whether subject-verb agreement is violated,

Super-classes in extended feature set (3)

CoPar: an indicator function that fires when conjoined phrases in a coordinate structure have approximately parallel syntactic structure,

CoLenPar: an indicator function that fires when conjoined phrases in a coordinate structure have approximately the same length,

Word: an indicator function that identifies words and their preterminals,

WProj: an indicator function that identifies words and their phrasal projections up to their maximal projection,

Super-classes in extended feature set (4)

WEdges: an indicator function that identifies the words and POS tags appearing at the edges of each nonterminal,

NGramTree: an indicator function of the subtree consisting of nodes connecting each pair of adjacent words in the parse tree, and

HeadTree: a tree fragment consisting of a head word and its projection up to its maximal projection, plus all of the siblings of each node in this sequence (this is like an auxiliary tree in a TAG).

Parsing accuracy (f-score) on section 22

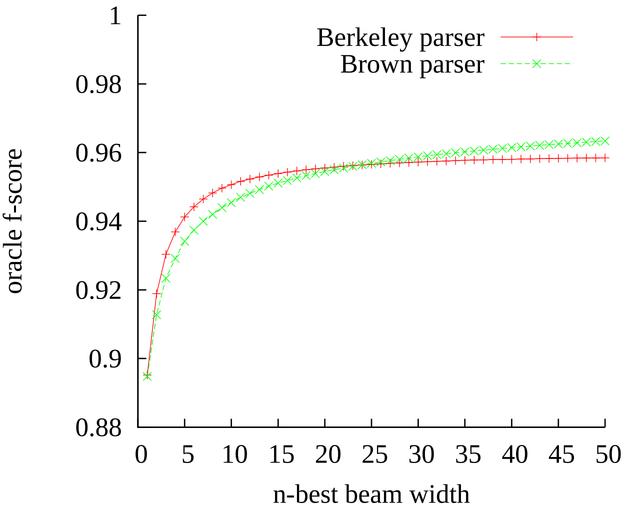
	No reranker	*Reranker features*	
		standard	extended
Berkeley trees	89.5	91.6	91.7
Brown trees	89.5	**91.8**	91.6
Combined trees		**91.8**	**91.9**

- Feature weights estimated by minimising EM-based log-loss with L2 regularisation using L-BFGS

See: Riezler et al (2000)

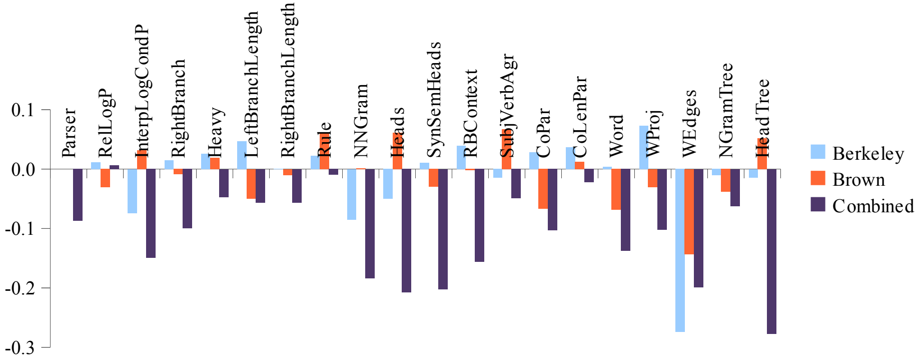
MACQUARIE UNIVERSITY

Oracle f-score on section 22



MACQUARIE UNIVERSITY

Feature super-class ablation experiment



- *Average f-score change* on folds 19–20 and section 22
- Rerankers used *extended feature set* trained with *averaged perceptron algorithm*
  - **91.2%** f-scores on both *Berkeley and Brown trees*, and
  - **91.6%** f-scores on *combined trees*.

MACQUARIE UNIVERSITY

Conclusions from feature super-class ablation experiment

- Linguistically-informed features (e.g., Heads, SynSemHeads, HeadTree) are more important when reranking combined trees than single parser output
  - perhaps log prob scores from individual parsers are effective when used on their own trees, but need recalibration on combined trees?
- Log prob scores from parsers also supply important information
- *Edge features* are particularly useful for Berkeley parser

See: Collins (2002), Collins and Roark (2004)

Conclusions

- Reranker on *section 23 combined trees* achieves **91.49% f-score**
  - only 0.1% higher than standard reranker on Brown trees
- Reranking the output of the Berkeley parser or a combination of Berkeley and Brown trees is *not significantly more accurate than reranking the Brown trees alone*, even with the extended feature set
  - perhaps the reranker features are still too oriented around Brown trees?
- *There is still room for improvement in parsing!*

See: Huang (2008)

MACQUARIE UNIVERSITY

Interested in parsing?

Macquarie University (Sydney, Australia)
is recruiting *PhD students* and *post-docs*.

Contact *Mark.Johnson@mq.edu.au* for more information.

MACQUARIE UNIVERSITY