ACCURATE PARSING USING RERANKING

Dominick Ng, Matthew Honnibal, and James R. Curran

a-lab, School of Information Technologies
dominick.ng@sydney.edu.au

SUMMARY

- We implement a discriminative parser reranker for the C&C statistical natural language parser, which improves parsing accuracy significantly.
- Reranking returns the best of an \( n \)-best set of parses; the process is more flexible than parsing and can consider a wide variety of features that model topology, context, and linguistic fidelity.

BACKGROUND

- Combinatory Categorial Grammar (CCG) is a lexicalised grammar formalism based on combinatory logic [Steedman, 2000].
- Each word is assigned a category that describes how it behaves in the sentence.
- Transitive verbs such as \( \text{likes} \) require an object \( \text{NP} \) to the right and a subject \( \text{NP} \) to the left to produce a valid sentence.

FEATURES

- Tree Topology - describes the overall shape of the parse tree and attempts to capture general conventions of English, e.g.
  - preference for right-branching trees
  - parallelism in conjunctions

- Local Context - adds context that is difficult to capture in the parser model, e.g.
  - ancestor paths of nodes in the tree
  - words at the edges of larger constituents

- CCG - features that may indicate an overly complicated or undesirable derivation based on the grammar, e.g.
  - combinations of unary rule applications
  - heads of conjunctive structures

- C&C - the features used in the parser model, to give the reranker the same evidence used by the parser itself
  - word-category pairs
  - rule applications and dependencies

RESULTS

- All experiments improved over choosing a parse at random from the \( n \)-best list.
- Regression over 10-best parses without pruning achieved a significant improvement in parsing accuracy.

- More features generally improved performance, and the novel features that we developed performed best in isolation.
  - Charniak and Johnson’s features alone did not improve performance for the more expressive CCG formalism.

- We experiment with various combinations of unary rule applications and heads of conjunctive structures.

FUTURE WORK

- Fully integrating the reranker into the parser for improved speed.
- Drawing features from wider corpora, such as the North American Newspapers and Wikipedia.

CONCLUSION

We have shown that reranking with expressive formalisms improves parsing accuracy. Our experiments also demonstrate that a wide variety of features is necessary for reranking to positively impact results.

Acknowledgements

This work was supported by Australian Research Council Discovery grants DP06655973 and DP1097291, the Capital Markets Cooperative Research Centre, and a University of Sydney Merit Scholarship.

References

