Clarification

How did you arrive at the probabilities of the constituent grammar components of the discourse tree, which were then used to calculate the probability of the tree as a whole?

How is discourse segmentation related to frames (as in FrameNet)? They both seem to address meaning and association within a sentence (and perhaps over multiple sentences).

Why is there a difference between sentence level and document level parsing? - Is there any way to remove this difference?

Can you explain the "reranking" capability of the CODRA parsing algorithm? Specifically, what are the "global" features of the DT that can be used as evidence? Are these features found by the algorithm or specified by something else?

How can we identify if two words are semantically similar in order to improve the robustness of CODRA?

(2) Why are CRF's so great for NLP and why have they become so popular?

Algorithm details

Does it make sense to split the multi-sentential parser into one for paragraphs and one to connect paragraphs? It seems like paragraphs as a boundary makes a lot of sense since text is generally split into paragraphs by the author by design.

How is the CKY algorithm we looked at in class different from the one they used?

Is the forward-backward algorithm in a 'fat' chain structured CRF to calculate the posterior probability like the one we have seen in class as Smoothing?

(4) Is the sliding window of 2 sentences always enough? Are there instances where 3+ sentences may be related, and if so, is that common? Is it possible for a units in a sentence to merge with sentences further up and down stream?

(3) Why are the segmentation model and parsers detached? Would it be possible to jointly model these two things?

Would it be possible to utilize complex neural models to avoid having to rely on hand-picked features?

Does modeling parse trees as binary trees bring any edge cases that need to be handled?

How does the reranker compare discourse trees, is each tree given a score function and how is this score computed?

 (2) Why is it taken for granted that sentence boundaries are also EDU boundaries?

CODRA encodes nuclearity within its relation labels. Although this simplifies the algorithm, is it possible that there it is actually less efficient than assigning nuclearity separately (e.g. assigning nuclearity before determining a relation)?

Are there any key benefits to using the 1S-1S approach over the sliding window approach?

CODRA uses a pipeline approach. Would it be possible to use CODRA in parallel?

Linguistics, RST

It is written that the parsing models sometimes fail to distinguish relations that are semantically similar. Is this an important problem? If so, what solutions might fix this problem.

Does this approach consider theoretical analyses of discourse from the linguistics area of pragmatics?

(2) How do we distinguish between nucleus and satellite sentences?

What is Rhetorical Structure Theory - a brief overview of what is going on?

What is EDU boundary?

Data – domains, languages, quality, features

(16) Is it possible that CODRA performs exceptionally for these two genres of text, but is heavily outperformed for different genres? Why weren't more text sample genres (or generic text collections) included in the evaluation of the framework?

(10) What type of adaptions would CODRA have to make to work for other languages (some of which may be structured differently semantically or grammatically) and how easily can they be made.

Given that language is actually a dynamic social process, and varies from one community to another, is there any way to being independent from WordNet for the sense of words? Because for obvious reasons there will be several words out of WordNet, limiting the accuracy of the relations.

(14) How would the authors adapt CODRA to function on less grammatically correct text (or text with typos)? One assumption that precludes the direct application of the pipeline is the existing of clear sentence boundaries (used for the extraction of EDUs).

(10) How would CODRA handle idioms/metaphors/slang/sarcasm/sentiment/dialogue/etc.?

Getting experts to annotate entire documents with discourse relations is difficult. Instead, is it practical to just have experts check over documents parsed by CODRA and once fixed/confirmed use those documents as new training data?

Applications

<mark>(2) Has there been any extensions to this work currently? For example, the researches presented an idea of using CODRA on e-mail/blog type text. Has this been done since this paper's publishing?</mark>

Would this method work only with finite text or also with infinite streams of text (e.g. an ongoing conversation)?

How can we tell that the coherent structure found is actually coherent? Could the CODRA framework give a sense of how dense or sparse the overall meaning of a sentence is? IE. If the probability is very low (relative to some empirical standard) for the first-of-k-parse-trees then there isn't a very coherent sentence?

Could systems like CODRA have further improvements that involve recognizing the similarities between documents, such that they could be used predict who wrote them, based on input? Conversely, after recognizing the pattern for similar kinds of documents, could rhetorical analysis be used in NLP to generate certain styles of documents?

Could these algorithms somehow be applied to context free grammar parsing?

As mentioned, many NLP applications heavily depend on this analysis. Are there any recent successes in those applications using this framework?

It seems that although CODRA assesses tree structures of certain units (EDU's at the lowest level), its analysis concerns semantic, rather than syntactic, properties of sentences. Would analysis by CODRA reveal some insight into the relations between the semantics and syntax of language? Are there something parallel to structure and labeling in languages, at least insofar as NLP is concerned?

Performance

How does the simplified model used for multi-sentential parsing compare to performing exact inference on the chain structure?

<mark>Other than the accuracy compared to humans are there any other metrics that can be used to investigate the performance of a discourse parser?</mark>

Could we further increase the overall accuracy of labeling or selecting discourse trees etc. by applying ensemble methods such as bagging? (since there is always chance of error in this process)

Have there been any alternative more/less effective models proposed that implement both intrasentenial and multisentenial parsing? Have there been any models proposed that do not implement this discrimination shown to be equally or more effective in certain contexts?

<mark>(5+) Has CODRA been tested/applied to real world applications or use cases? How efficient is it compared to other existing similar tools here?</mark>

How does CODRA's accuracy compare to that of methods dependent on neural networks? SVMs…..

Could discourse parsing in general benefit from an any-time approximation algorithm with some sort of heuristics?

Personal/requests

What made you interested in writing about this topic and how long did this paper take to write?