Decalog 2007: Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue, pages 189–190.
Trento, Italy, 30 May – 1 June 2007. Edited by Ron Artstein and Laure Vieu.

Automatic Discourse Segmentation using Neural Networks

Rajen Subba
Department of Computer Science
University of Illinois at Chicago
Chicago, Illinois 60607
rsubba@cs.uic.edu

Barbara Di Eugenio
Department of Computer Science
University of Illinois at Chicago
Chicago, Illinois 60607
bdieugen@cs.uic.edu

Discourse segmentation is the task of determining minimal non-overlapping units of discourse called elementary discourse units (EDUs). It can be further subdivided into sentence segmentation and sentence-level discourse segmentation. This paper addresses the latter, more challenging subtask, which takes a sentence and outputs the EDUs for that particular sentence.

(1) Saturday, he amended his remarks to say that he would continue to abide by the cease-fire if the U.S. ends its financial support for the Contras.

(1a) Saturday, he amended his remarks
(1b) to say
(1c) that he would continue to abide by the cease-fire
(1d) if the U.S. ends its financial support for the Contras.

In example (1), a sentence from a Wall Street Journal article taken from the Penn TreeBank corpus is further segmented into four EDUs, (1a), (1b), (1c) and (1d) (RST, 2002). Discourse segmentation, clearly, is not as easy as sentence boundary detection. The lack of consensus with regards to what constitutes an elementary discourse unit adds to the difficulty. Building a rule based discourse segmenter can be a tedious task since these rules would have to be based on the underlying grammar of the particular parser that is to be used. Therefore, we adopted a neural network model for automatically building a discourse segmenter from an underlying corpus of segmented text. We chose to use part-of-speech tags, syntactic information, discourse cues and punctuation. Our ultimate goal is to build a discourse parser that uses this discourse segmenter.

The data that we used to train and test our discourse segmenter is the RST-DT (RST, 2002) corpus. The corpus contains 385 Wall Street Journal articles from the Penn Treebank. The training set consists of 347 articles for a total of 6132 sentences, whilst the test set contains 38 articles for a total of 991 sentences. The RST-DT corpus provides us with pairs of sentences and EDUs. For the syntactic structure of the sentences, we have used both the gold standard Penn Treebank data and syntactic parse trees generated by (Charniak, 2000). As regards the discourse cues, we used a list of 168 possible discourse markers.

Problem formulation Like (Soricut and Marcu, 2003), we formulate the discourse segmentation task as a binary classification problem of deciding whether to insert a segment boundary after each word in the sentence. Our examples are vectors that provide information on POS tags, discourse cues and the syntactic structure of the surrounding context for each word in the sentence. The categories that we decided to use in our vector representation for each example are given in table 1. We used binary encoding of the values for each category in order to convert them into numeric values and compress our data. For all the 12 categories, we needed a total of 84 bits. After processing our data we obtained about 140,000 examples (vectors) to train the model. Each vector also indicated whether a segment boundary followed that particular word or not. We used a Multi-Layer Perceptron. The weights of the network were initialized using a random uniform distribution. Back-Propagation was used to update the weights. Each training run was limited to 50 iterations. We trained both a single model and a bagged model.

No.	category type
1	Prev. word POS
2	Prev. word Next Label
3	Prev. word Parent
4	Cur. word POS
5	Cur. word Parent
6	Next word POS
7	Next word Next Label
8	Next word Parent
9	Common ancestor CFG Rule for Cur. word and Next word
10	Cur. word CFG Non-Terminal
11	Next word CFG Non-Terminal
12	Is Next word a Discourse Cue ?

Table 1: Categories used for training the model.

Experiments and Results We evaluate our discourse segmenter against the test set of 38 articles with 991 sentences from the RST-DT corpus. We compare our results on the RST-DT test set with that of (Marcu, 2000) and (Soricut and Marcu, 2003). (Soricut and Marcu, 2003) used a probabilistic model (SynDS) and (Marcu, 2000) implemented a decision tree based model (DT). (Soricut and Marcu, 2003) measures the performance of the segmenter based on the it's ability to insert inside-sentence segment boundaries. Table 2 reports the results for the RST-DT test set for four systems using their metric. NNDS (Neural Network Discourse Segmenter) is our system. NNDS-B is the bagged model. SynDS is the best reported system that we are aware of. The results show that NNDS, a neural network based discourse segmenter can perform as well as SynDS. Bagging the model increases the performance of the segmenter. More importantly recall is higher since a bagged model is less sensitive to overfitting. The human segmentation performance as reported by (Soricut and Marcu, 2003) is 98.3% F-Score.

We also compare our system to (Huong et. al, 2004). (Huong et. al, 2004) is a symbolic implementation. Unlike (Soricut and Marcu, 2003), they used a flat-bracketing measure to compute performance. This measure accounts for both the start and end boundaries of a segment for precision and recall. They report an F-Score of 80.3% using the Penn TreeBank parsed trees. Our segmenter using bagging obtains a performace of 84.19% F-Score according to this measure. While our evaluation is based on the full test set of 38 articles, (Huong et. al, 2004) used only 8 articles for testing their symbolic segmenter.

System	Parse Tree	Precision	Recall	F-Score
DT	-	83.3	77.1	80.1
SynDS	C	83.5	82.7	83.1
SynDS	T	84.1	85.4	84.7
NNDS	C	83.66	80.17	82.03
NNDS	T	85.35	83.8	84.56
NNDS - B	C	83.94	84.89	84.41
NNDS - B	T	85.56	86.6	86.07

Table 2: Performance on the RST-DT corpus.
(Parse Tree: C - Charniak, T - Penn TreeBank)

Conclusion We have presented a connectionist approach to automatic discourse segmentation. Bagging the model yields even better performance. The performance of our discourse segmenter is comparable to the best discourse segmenter that has been reported. In the future, we intend to exploit additional features, namely lexical head features from the syntactic parse trees. We also plan to test our discourse segmenter on other discourse corpora, where segmentation decisions are based on a different coding scheme to test how well our model can generalize.

Acknowledgments

This work is supported by award IIS-0133123 from the National Science Foundation (NSF), and additionally by awards ALT-0536968 from NSF and N000140010640 from ONR.

References

Charniak, E. 2000. A maximum-entropy-inspired. In Proceedings of the NAACL 2000, pages 132139, Seattle, Washington, April 29 May 3.

Daniel Marcu. 2000. The Theory and Practice of Discourse Parsing and Summarization. The MIT Press, November 2000.

Huong Le Thanh, Geetha Abeysinghe and Christian Huyck. 2004. Automated Discourse Segmentation by Syntactic Information and Cue Phrases, In Proceedings of the IASTED, Innsbruck, Austria.

Radu Soricut and Daniel Marcu. 2003. Sentence Level Discourse Parsing using Syntactic and Lexical Information. In *Proceedings of the HLT/NAACL-2003*, Edmonton, Canada, May-June.

RST-DT. 2002. RST Discourse Treebank. Linguistic Data Consortium.