Multi-view and multi-task training of RST discourse parsers

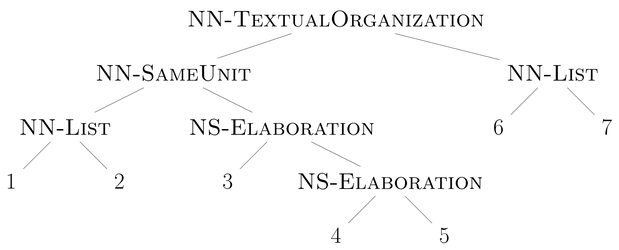
Chloé Braud* & Barbara Plank◇ & Anders Søgaard*

*CoAStaL, Dep. of Computer Science, Univ. of Copenhagen ◇ Computational Linguistics, CLCG, Univ. of Groningen

Introduction

Discourse parsing: identifying the structure describing the organization of a document

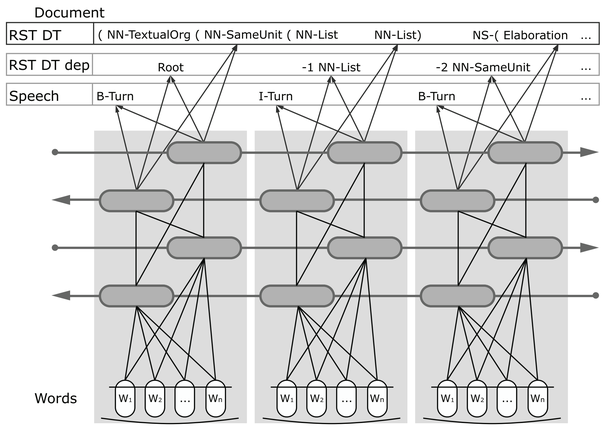
→ Issue: **sparsity** = limited amount of training data



Method

Multi-task and **multi-view** training of bi-LSTM networks

→ regularization using task supervision from related tasks and alternative views of the data



Setting

- Data: **RST Discourse Treebank**, 385 doc from the WSJ
- **Sequence prediction** task: trees are encoded as sequences preserving all the information + heuristics at evaluation time
- Systems: **bi-LSTM** with or without auxiliary tasks

Alternate views

	Constituent (main)	Dependency	Nuclearity	Relation	Fine-grained
EDU 1	(NN-TextOrg(NN-SaUnit(NN-List	Root	(NN(NN(NN	(TextOrg(SaUnit(List	(TextOrg(SaUnit(List
EDU 2	NN-List)	-1 NN-List	NN)	List)	List)
EDU 3	(NS-Elab	-2 NN-SaUnit	(NS	(Elab	(Elab-set-member

Auxiliary tasks

	Speech	Factuality	Aspect	Modality	Polarity	Tense	Coreference	PDTB
Corpus	Santa Barbara	Factbank	Timebank	Timebank	Timebank	Timebank	Ontonotes	PDTB
Sent 1	turn1	Certain	Progressive	Must	Positive	Past	Root	Root
Sent 2	turn2	Probable	Perfective	Could	Negative	Future	Coreferent	Contrast

Experiments

System	Fine	Fact	Speech	Asp	Dep	Nuc+lab	Mod	Pol	PDTB	Coref	Ten	Span	Nuclearity	Relation
Prior work														
DPLP concat	-	-	-	-	-	-	-	-	-	-	-	82.08	**71.13**	61.63
DPLP general	-	-	-	-	-	-	-	-	-	-	-	81.60	70.95	**61.75**
Our work														
Hier-LSTM	-	-	-	-	-	-	-	-	-	-	-	81.39	64.54	49.15
MTL-Hier-LSTM	✓	-	-	-	-	-	-	-	-	-	-	82.88	67.46	53.25
MTL-Hier-LSTM	-	✓	-	-	-	-	-	-	-	-	-	83.40	67.16	52.10
MTL-Hier-LSTM	-	-	✓	-	-	-	-	-	-	-	-	83.26	67.51	51.75
MTL-Hier-LSTM	-	-	-	✓	-	-	-	-	-	-	-	83.69	66.25	51.25
MTL-Hier-LSTM	-	-	-	-	✓	-	-	-	-	-	-	81.25	65.34	51.24
MTL-Hier-LSTM	-	-	-	-	-	✓	-	-	-	-	-	82.09	65.68	51.12
MTL-Hier-LSTM	-	-	-	-	-	-	✓	-	-	-	-	81.66	65.31	50.58
MTL-Hier-LSTM	-	-	-	-	-	-	-	✓	-	-	-	82.01	65.29	50.11
MTL-Hier-LSTM	-	-	-	-	-	-	-	-	✓	-	-	81.61	63.10	48.89
MTL-Hier-LSTM	-	-	-	-	-	-	-	-	-	✓	-	80.26	63.35	47.70
MTL-Hier-LSTM	-	-	-	-	-	-	-	-	-	-	✓	81.33	62.34	47.57
Best combination	-	-	-	-	✓	✓	✓	-	✓	-	-	**83.62**	69.77	55.11
Human annotation	-	-	-	-	-	-	-	-	-	-	-	88.70	77.72	65.75

- DPLP: scores reproduced from [Ji and Eisenstein 2014]
- Hier-LSTM: baseline system
- MTL-Hier-LSTM: multi-task learning

Results

- The architecture captures some of the **syntactic and contextual information** needed
  → **MTL improves** over STL for 8/11 tasks
- **New interesting sources of information** for the task
  → **Alternate views** are the most beneficial, especially Fine-grained
  → **Speech** is the most beneficial auxiliary task
- **Best system: Task combinations**
  → Based on different views (Nuclearity + Labels + Dependency) and different tasks (Modality + PDTB)
- No improvement with **Tense and Coreference** known as crucial information for the task
  → Calls for a **finer grained encoding**
- **Low scores on Relation**, future work:
  - Using a finer grained encoding for the auxiliary tasks
  - Adding syntactic information [Lin et al. 2009]
  - Using different combination schemes between the arguments [Ji and Eisenstein 2014]

Further information

Code available at `http://bitbucket.org/chloebt/discourse`

Contacts:
chloe.braud@gmail.com
soegaard@hum.ku.dk
bplank@gmail.com

Acknowledgements

Chloé Braud and Anders Søgaard were funded by the ERC Starting Grant LOWLANDS No. 313695.