Cross-lingual RST discourse parsing

EACL 2017

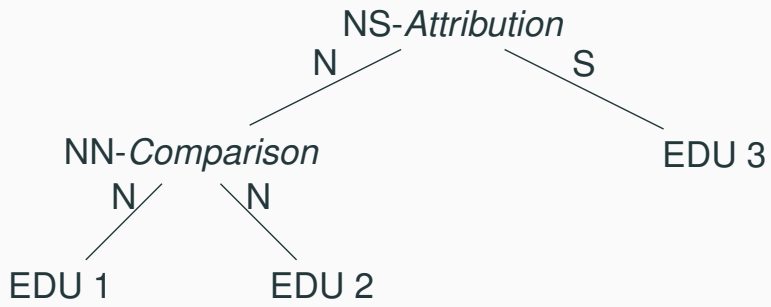
Chloé Braud, Maximin Coavoux and Anders Søgaard

TABLE DES MATIÈRES

Rhetorical Structure Theory [Mann and Thompson, 1988]

- A tree covering a document:
  - Elementary Discourse Unit (EDU): mostly clauses
  - Discourse relations: *Comparison*, *Attribution*, *Result* . . .
  - Nuclearity: Nucleus-Satellite (NS), Nucleus-Nucleus (NN)
- Useful for many tasks:
  - summarization, sentiment analysis, question-answering ...

[Consumer spending in Britain rose 0.1% in the third quarter from the second quarter]$_1$ [and was up 3.8% from a year ago,]$_2$ [the Central Statistical Office estimated.]$_3$



RST Discourse parsing

- Many systems for English (RST DT)

  [Feng and Hirst, 2014; Li et al., 2014; Ji and Eisenstein, 2014]

- Corpora for other languages, but a few systems:
  - Spanish: rule-based [Maziero et al., 2011]
  - Brazilian Portuguese: rule-based [Pardo and Nunes, 2008], statistical intra-sentential [Maziero et al., 2015]
  - German
  - Dutch
  - Basque

  $\rightarrow$ Develop full statistical systems for other languages

- Corpora limited in size (RST DT 385 documents)

  $\rightarrow$ Combine the corpora

RST Corpora

- Small corpora: monolingual systems if > 100 trees
- Different annotations schemes: harmonization

Corpus	#Trees	#Words	#Rel	#EDU
En-DT	385	206,300	56	21,789
Pt-DT	329	135,820	32	12,573
Es-DT	266	69,787	29	4,019
De-DT	173	32,274	30	2,790
Nl-DT	80	27,920	31	2,345
Eu-DT	85	27,982	31	2,396

RST Corpora

- Small corpora: monolingual systems if > 100 trees
- Different annotations schemes: harmonization

Corpus	#Trees	#Words	#Rel	#EDU
En-DT	385	206,300	56	21,789
Pt-DT	329	135,820	32	12,573
Es-DT	266	69,787	29	4,019
De-DT	173	32,274	30	2,790
Nl-DT	80	27,920	31	2,345
Eu-DT	85	27,982	31	2,396

Harmonization

Pre-processing + Binarization of the trees + Relation sets

- 18 coarse grained relations [Carlson and Marcu, 2001]
  1. Relations annotated in the En-DT
  2. Relations similar to a class
     e.g. *explanation* $\rightarrow$ *Explanation*
     reformulation $\rightarrow$ *Restatement*
  3. Remaining cases: relations that fit the properties of a class
     e.g. *parenthetical* $\rightarrow$ *Elaboration*
     justify, motivation $\rightarrow$ *Explanation*
- Relation sets still different
  e.g. *Attribution* only in the En-DT and Pt-DT
- Full mapping in the paper
  Code: `https://bitbucket.org/chloebt/discourse/`

APPROACH: CROSS-LINGUAL SYSTEMS

Systems

- Monolingual systems:
  - Fully supervised, state-of-the-art performance
  - En, De, Es, Pt: > 100 trees (38 kept for test)
- Cross-lingual source only:
  - Performance when no data for the target language?
  - Train and optimize on all the source languages
- Cross-lingual source+target:
  - Improvements by combining the corpora?
  - Train on source(+target) and optimize on target

APPROACH: DISCOURSE PARSER

Constituent parser [Coavoux and Crabbé, 2016]

- Lexicalized shift-reduce transition system
- Scoring system: feed-forward neural network
- Any features, mapped to real-valued vectors
- Pre-trained embeddings
- Beam search

Information from:

- the 2 EDUs on the top of the stack + EDU on the queue
- the left and right children of the 2 elements on the stack
- representation built for the 2 top elements on the stack

APPROACH: REPRESENTATION

Features types:

- First 3 and last words + POS, "head set" [Sagae, 2009]
  vs all words [Li et al., 2014; Ji and Eisenstein, 2014]
- Position of the EDU and length, position of the head
- Number/date/percent/money

$\rightarrow$ Concatenation

Multi-lingual:

- Universal Dependencies and UDPipe
- Bi-lingual Wiktionaries
- Cross-lingual embeddings [Levy et al., 2017]

EXPERIMENTS: SYSTEMS

- Gold segmentation
- MFS: Most Frequent Baseline
- Mono: monolingual systems, fully supervised
- Source only: train on source languages
- Source+target: train on source(+target), optimize on target

System	En-DT			Pt-DT			Es-DT			De-DT			Nl-DT			Eu-DT		
	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel
MFS	58.2	33.4	22.1	57.3	33.9	23.23	82.0	51.5	17.7	61.3	37.8	13.2	57.9	35.5	22.0	63.2	34.9	18.8
Li et al.	85.0	70.8	58.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ji and Eisenstein	82.1	71.1	61.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mono	85.0	72.3	60.1	82.0	65.1	49.9	89.7	72.7	54.4	80.2	53.9	35.0	-	-	-	-	-	-
+ emb.	83.5	68.5	55.9	81.3	62.9	48.8	89.3	72.4	51.4	77.7	51.6	31.1	-	-	-	-	-	-
Source only	76.3	50.5	31.3	76.5	54.6	35.5	78.1	45.4	27.0	76.0	46.0	26.1	69.5	42.1	25.3	78.6	53.0	26.4
Source+Target	85.1	73.1	61.4	81.9	65.1	49.8	88.8	68.0	50.4	79.6	53.6	34.1	69.2	43.4	28.3	76.7	50.5	29.5

EXPERIMENTS: RESULTS

- Monolingual systems for languages with a large corpus

System	En-DT			Pt-DT			Es-DT			De-DT			Nl-DT			Eu-DT		
	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel
MFS	58.2	33.4	22.1	57.3	33.9	23.2	82.0	51.5	17.7	61.3	37.8	13.2	57.9	35.5	22.0	63.2	34.9	18.8
Li et al.	**85.0**	70.8	58.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ji and Eisenstein	82.1	**71.1**	**61.6**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mono	85.0	72.3	60.1	82.0	65.1	49.9	89.7	72.7	54.4	80.2	53.9	35.0	-	-	-	-	-	-
+ emb.	83.5	68.5	55.9	81.3	62.9	48.8	89.3	72.4	51.4	77.7	51.6	31.1	-	-	-	-	-	-
Source only	76.3	50.5	31.3	76.5	54.6	35.5	78.1	45.4	27.0	76.0	46.0	26.1	69.5	42.1	25.3	78.6	53.0	26.4
Source+Target	85.1	73.1	61.4	81.9	65.1	49.8	88.8	68.0	50.4	79.6	53.6	34.1	69.2	43.4	28.3	76.7	50.5	29.5

EXPERIMENTS: RESULTS

- Competitive on English: less words, more context

System	En-DT			Pt-DT			Es-DT			De-DT			Nl-DT			Eu-DT		
	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel
MFS	58.2	33.4	22.1	57.3	33.9	23.2	82.0	51.5	17.7	61.3	37.8	13.2	57.9	35.5	22.0	63.2	34.9	18.8
Li et al.	**85.0**	70.8	58.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ji and Eisenstein	82.1	**71.1**	**61.6**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mono	85.0	72.3	60.1	82.0	65.1	49.9	89.7	72.7	54.4	80.2	53.9	35.0	-	-	-	-	-	-
+ emb.	83.5	68.5	55.9	81.3	62.9	48.8	89.3	72.4	51.4	77.7	51.6	31.1	-	-	-	-	-	-
Source only	76.3	50.5	31.3	76.5	54.6	35.5	78.1	45.4	27.0	76.0	46.0	26.1	69.5	42.1	25.3	78.6	53.0	26.4
Source+Target	85.1	73.1	61.4	81.9	65.1	49.8	88.8	68.0	50.4	79.6	53.6	34.1	69.2	43.4	28.3	76.7	50.5	29.5

EXPERIMENTS: RESULTS

- Cross-lingual: first systems on Dutch and Basque > MFS

System	En-DT			Pt-DT			Es-DT			De-DT			Nl-DT			Eu-DT		
	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel
MFS	58.2	33.4	22.1	57.3	33.9	23.2	82.0	51.5	17.7	61.3	37.8	13.2	57.9	35.5	22.0	63.2	34.9	18.8
Li et al.	**85.0**	70.8	58.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ji and Eisenstein	82.1	**71.1**	**61.6**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mono	85.0	72.3	60.1	82.0	65.1	49.9	89.7	72.7	54.4	80.2	53.9	35.0	-	-	-	-	-	-
+ emb.	83.5	68.5	55.9	81.3	62.9	48.8	89.3	72.4	51.4	77.7	51.6	31.1	-	-	-	-	-	-
Source only	76.3	50.5	31.3	76.5	54.6	35.5	78.1	45.4	27.0	76.0	46.0	26.1	69.5	42.1	25.3	78.6	53.0	26.4
Source+Target	85.1	73.1	61.4	81.9	65.1	49.8	88.8	68.0	50.4	79.6	53.6	34.1	69.2	43.4	28.3	76.7	50.5	29.5

EXPERIMENTS: RESULTS

- Cross-lingual Source+Target: improvement on English

System	En-DT			Pt-DT			Es-DT			De-DT			Nl-DT			Eu-DT		
	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel
MFS	58.2	33.4	22.1	57.3	33.9	23.2	82.0	51.5	17.7	61.3	37.8	13.2	57.9	35.5	22.0	63.2	34.9	18.8
Li et al.	**85.0**	70.8	58.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ji and Eisenstein	82.1	**71.1**	**61.6**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mono	85.0	72.3	60.1	82.0	65.1	49.9	89.7	72.7	54.4	80.2	53.9	35.0	-	-	-	-	-	-
+ emb.	83.5	68.5	55.9	81.3	62.9	48.8	89.3	72.4	51.4	77.7	51.6	31.1	-	-	-	-	-	-
Source only	76.3	50.5	31.3	76.5	54.6	35.5	78.1	45.4	27.0	76.0	46.0	26.1	69.5	42.1	25.3	78.6	53.0	26.4
Source+Target	85.1	73.1	61.4	81.9	65.1	49.8	88.8	68.0	50.4	79.6	53.6	34.1	69.2	43.4	28.3	76.7	50.5	29.5

EXPERIMENTS: RESULTS

- Cross-lingual Source only: drop in performance, esp. for "relation"
  - Issue with the representation
  - Bias: drop for *Attribution* in the En-DT: 90% $\rightarrow$ 30% F_1
  - But improvement for others: *Joint* +3%, *Condition* +3%

System	En-DT			Pt-DT			Es-DT			De-DT			Nl-DT			Eu-DT		
	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel
MFS	58.2	33.4	22.1	57.3	33.9	23.2	82.0	51.5	17.7	61.3	37.8	13.2	57.9	35.5	22.0	63.2	34.9	18.8
Li et al.	**85.0**	70.8	58.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ji and Eisenstein	82.1	**71.1**	**61.6**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mono	85.0	72.3	60.1	82.0	65.1	49.9	89.7	72.7	54.4	80.2	53.9	35.0	-	-	-	-	-	-
+ emb.	83.5	68.5	55.9	81.3	62.9	48.8	89.3	72.4	51.4	77.7	51.6	31.1	-	-	-	-	-	-
Source only	76.3	50.5	31.3	76.5	54.6	35.5	78.1	45.4	27.0	76.0	46.0	26.1	69.5	42.1	25.3	78.6	53.0	26.4
Source+Target	85.1	73.1	61.4	81.9	65.1	49.8	88.8	68.0	50.4	79.6	53.6	34.1	69.2	43.4	28.3	76.7	50.5	29.5

EXPERIMENTS: RESULTS

- Pre-trained embeddings do not provide the information needed

System	En-DT			Pt-DT			Es-DT			De-DT			Nl-DT			Eu-DT		
	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel	Sp	Nuc	Rel
MFS	58.2	33.4	22.1	57.3	33.9	23.2	82.0	51.5	17.7	61.3	37.8	13.2	57.9	35.5	22.0	63.2	34.9	18.8
Li et al.	**85.0**	70.8	58.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ji and Eisenstein	82.1	**71.1**	**61.6**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mono	85.0	72.3	60.1	82.0	65.1	49.9	89.7	72.7	54.4	80.2	53.9	35.0	-	-	-	-	-	-
+ emb.	83.5	68.5	55.9	81.3	62.9	48.8	89.3	72.4	51.4	77.7	51.6	31.1	-	-	-	-	-	-
Source only	76.3	50.5	31.3	76.5	54.6	35.5	78.1	45.4	27.0	76.0	46.0	26.1	69.5	42.1	25.3	78.6	53.0	26.4
Source+Target	85.1	73.1	61.4	81.9	65.1	49.8	88.8	68.0	50.4	79.6	53.6	34.1	69.2	43.4	28.3	76.7	50.5	29.5

Conclusion

- New discourse parser competitive on English
- First discourse parsers evaluated on 6 languages
- First cross-lingual experiments for discourse parsing
- Code for the discourse parser on github
- Code for harmonizing the corpora: test your parser!
- ACL17: discourse segmenters for all the languages

`https://bitbucket.org/chloebt/discourse/`

Thanks!

REFERENCES I