Mining arguments in scientific abstracts with discourse-level embeddings.

(Accuosto and Saggion, 2020)

Laurine Huber

LORIA, Université de Lorraine, France

Café TAL
April 27, 2021

Argumentation: a definition (Argumentation mining, Stede, 2018)

"Argumentation is a verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint."
(van Eemeren and Grootendorst, 2004)

Argumentation: a definition (Argumentation mining, Stede, 2018)

"Argumentation is a verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint."
(van Eemeren and Grootendorst, 2004)

Argumentation: a definition (Argumentation mining, Stede, 2018)

"Argumentation is a verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint."
(van Eemeren and Grootendorst, 2004)

Argumentation: a definition (Argumentation mining, Stede, 2018)

"Argumentation is a verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint."
(van Eemeren and Grootendorst, 2004)

Argumentation: a definition (Argumentation mining, Stede, 2018)

"Argumentation is a verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint."
(van Eemeren and Grootendorst, 2004)

Argumentation: a definition (Argumentation mining, Stede, 2018)

"Argumentation is a verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint."
(van Eemeren and Grootendorst, 2004)

Argumentation in scientific abstracts

"State-of-art systems for grammar error correction often correct errors based on word sequences or phrases. In this paper , we describe a grammar error correction system which corrects grammatical errors at tree level directly. [...] Experiments show that our system outperforms the state-of-art systems and improves the F1 score." [1]

[Assertion] – *support* ⟶ [Proposal] ⟵ *support* – [Result]

[1] Simplified abstract from SciDTB (*D14-1033*)

Argumentation in scientific abstracts

"State-of-art systems for grammar error correction often correct errors based on word sequences or phrases. In this paper , we describe a grammar error correction system which corrects grammatical errors at tree level directly. [...] Experiments show that our system outperforms the state-of-art systems and improves the F1 score." [1]

[Assertion] – *support* ⟶ [Proposal] ⟵ *support* – [Result]

[1] Simplified abstract from SciDTB (*D14-1033*)

Argumentation in scientific abstracts

"State-of-art systems for grammar error correction often correct errors based on word sequences or phrases. In this paper , we describe a grammar error correction system which corrects grammatical errors at tree level directly. [...] Experiments show that our system outperforms the state-of-art systems and improves the F1 score." [1]

[Assertion] – *support* ⟶ [Proposal] ⟵ *support* – [Result]

[1] Simplified abstract from SciDTB (*D14-1033*)

Argumentation in scientific abstracts

"State-of-art systems for grammar error correction often correct errors based on word sequences or phrases. In this paper , we describe a grammar error correction system which corrects grammatical errors at tree level directly. [...] Experiments show that our system outperforms the state-of-art systems and improves the F1 score." [1]

[Assertion] – *support* ⟶ [Proposal] ⟵ *support* – [Result]

[1] Simplified abstract from SciDTB (*D14-1033*)

Argumentation mining

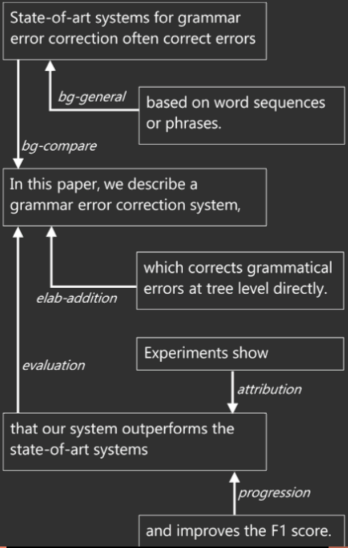
1. Segment the text into argumentative units
2. Identify the central claim
3. Identify the type of argumentative units
4. Identify the function of argumentative units
5. Identify the relations between argumentative units

SciDTB

- 798 abstracts from ACL Anthology
- NLP conferences (ACL, EMNLP, ...)
- Annotated in dependency (almost like RST[2])
- Extension: 60 documents annotated in argumentation

[2] Rhetorical Structure Theory, Mann and Thompson, 1988

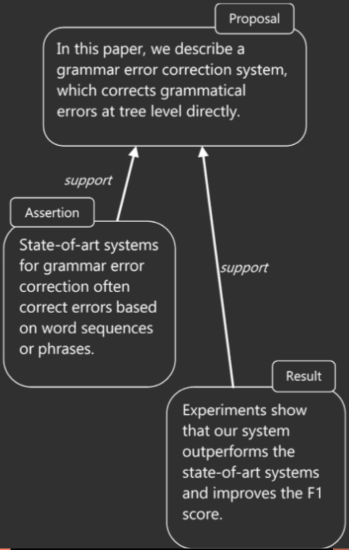
Discourse annotation (in RST)



Discourse relations

- 17 coarse grained relations
- 26 fine grained relations
  Attribution, Background (Related, Goal, General), Cause-effect (Cause, Result), Comparison, Condition, Contrast, Elaboration (Addition, Aspect, Process-step, Definition, Enumerate, Example), Enablement, Evaluation, Explain (Evidence, Reason)...

Argumentation annotation

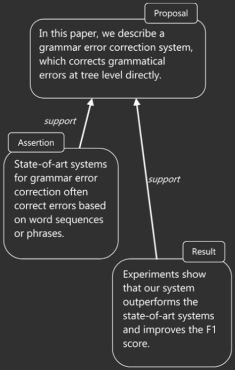


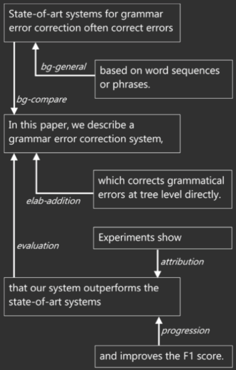
Argumentation annotation

- **6 argument types**
  proposal, assertion, result, observation, means, description
- **5 argumentative functions**
  support, attack, detail, additional, sequence

Leveraging discourse structures for AM

Do discourse structures help to improve argument mining ?





Leveraging discourse structures for AM

How to leverage discourse structures to improve argument mining ?

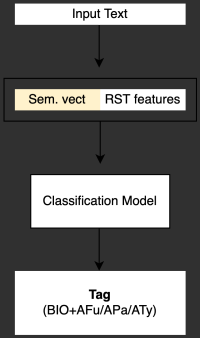
- **Discourse features as input**: The AM model takes as input a concatenation of linguistic features and specific discourse features.
- **Multi-task learning**: The AM model and the discourse parsing model have a shared a layer.
- **Transfer learning**: The AM model takes as input vectors learnt from a discourse parsing model.

Experiments overview

- Argumentation structure prediction:
  - Boundaries
  - Types (**ATy**)
  - Function (**AFu**)
  - Attachment (**APa**)
- Discourse structure prediction
  - Function (**DFu**)
  - Attachment (**DPa**)
- Token labeling ($f : tok \rightarrow tag$)

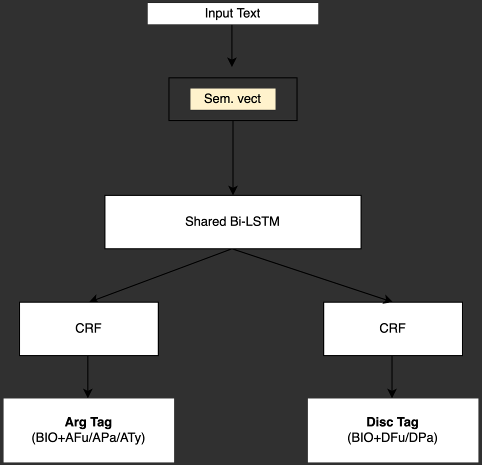
Token	ABound	ATy	AFu	APa
We	B	Proposal	ROOT	0
describe	I	Proposal	ROOT	0
a	I	Proposal	ROOT	0
grammar	I	Proposal	ROOT	0
error	I	Proposal	ROOT	0
correction	I	Proposal	ROOT	0
system	I	Proposal	ROOT	0
...	I	Proposal	ROOT	0
at	I	Proposal	ROOT	0
tree	I	Proposal	ROOT	0
level	I	Proposal	ROOT	0
directly.	I	Proposal	ROOT	0
Experiments	B	Result	Support	-1
show	I	Result	Support	-1
that	I	Result	Support	-1
our	I	Result	Support	-1
system	I	Result	Support	-1
...	I	Result	Support	-1
improves	I	Result	Support	-1
F1-score.	I	Result	Support	-1

RST as input features

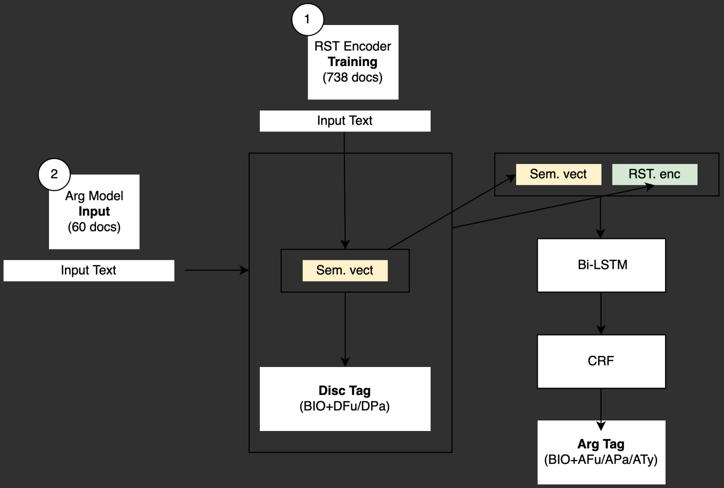


- Majority-class Classifier
- CRF Classifier
- Bi-LSTM + CRF

Multi-task architecture



Transfer learning architecture



Results - Discourse features

Average F1-measures and standard deviations obtained with and without discourse (Disc), syntactic (Syn) and positional (Pos) features for the different types of classifiers in 10-fold cross-validation settings.

Algorithm	Features	AFu		ATy		APa	
		Avg. F1	σ	Avg. F1	σ	Avg. F1	σ
Majority	Syn, Pos	46.52	3.54	43.84	10.69	31.26	6.29
Majority	Syn, Pos, Disc	**57.04**	7.87	**56.03**	8.14	**47.06**	9.89
CRF	Syn, Pos	53.33	17.53	61.62	12.21	39.81	15.42
CRF	Syn, Pos, Disc	**62.51**	10.54	**66.04**	15.42	**44.96**	7.61
BiLSTM	No Feat.	69.94	6.30	66.94	8.82	41.74	10.43
BiLSTM	Disc	**71.07**	8.51	**69.72**	8.70	43.23	10.17
BiLSTM	Syn, Pos	68.45	4.22	67.91	9.84	42.95	9.06
BiLSTM	Syn, Pos, Disc	70.02	5.40	69.67	9.07	**43.39**	10.66

Results - Discourse features

- Incorporating discourse information contributes to better performances in all cases
- Deep learning models $>$ Classical classifiers ($\textit{no optimization} + \textit{few data}$)
- Syn and Pos: no significant improvement
- Argumentative Parent: large number of categories + unbalanced data

Distance to parent	
adjacent	167
1 arg. unit	55
2 arg. units	36
3 arg. units	17
4 arg. units	11
5 arg. units	5
6 arg. units	1

Results - TL

Average F1-measures and standard deviations with and without information transferred from discourse parsing tasks.

Method	Input	AFu		ATy		APa	
		Avg. F1	σ	Avg. F1	σ	Avg. F1	σ
Single-task	$[\vec{k}, \vec{e}]$	69.94	6.30	66.94	8.82	41.74	10.43
Multi-task	$[\vec{k}, \vec{e}]$	67.38	6.90	65.65	12.39	40.69	9.98
Seq. transfer	$[\vec{k}, \vec{e}, \vec{d}]$	**70.98**	7.17	**70.38**	8.39	**43.44**	11.16

Results - TL

- Seq. transfer > Multi-task
- Seq. transfer advantage : can be applied to texts where no discourse level annotation is provided
- Multi-task : *negative transfer effect* (regularization effect introduced by aux. tasks too strong → affects performances of the main task)

Conclusion

- Including discourse data improves AM, independently of the learning algorithm
- Use other deep learning archi. (e.g. attention-based)
- Working on optimization could yield better performances

Other directions

- Transition-based parsing
- Using other corpora in a Multi-Task setting:
  - BioDRB (Prasad et. al., 2011)
  - Argument Annotated Scientific Publications (Lauscher, 2018)
  - Argumentative Zoning corpus (Teufel, 2014)
  - ArgMicroTexts (Peldsusz and Stede, 2016)
- Other tree properties (depth, width)

Thank you!