Argumentative Link Prediction using Residual Networks and Multi-Objective Learning

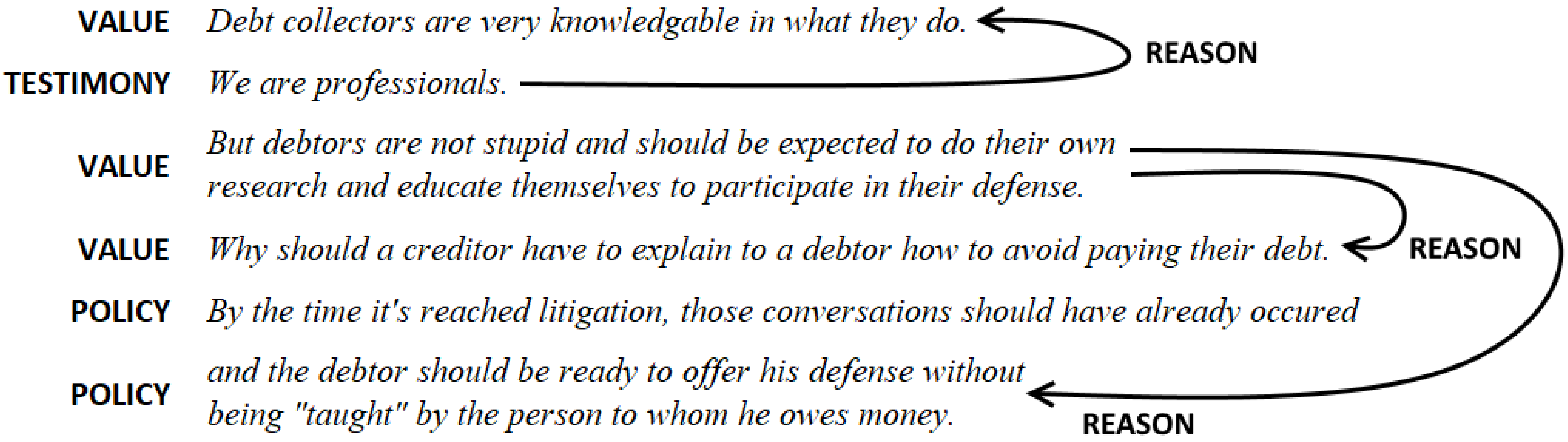
Galassi Andrea[1]
Marco Lippi[2]
Paolo Torroni[1]

Departments of
[1]Computer Science and Engineering (DISI)
University of Bologna, Italy
[2]Sciences and Methods for Engineering (DISMI)
University of Modena and Reggio Emilia, Italy

Argumentation mining tasks:

- *Segmentation*: detect the boundaries of argumentative components
- ***Component Classification***: label the components according to their type (es: claim/premise)
- ***Link prediction***: identify the (pairwise) relations between components
- ***Relation Classification***: label such links (es: support/attack)

Argumentation mining tasks:

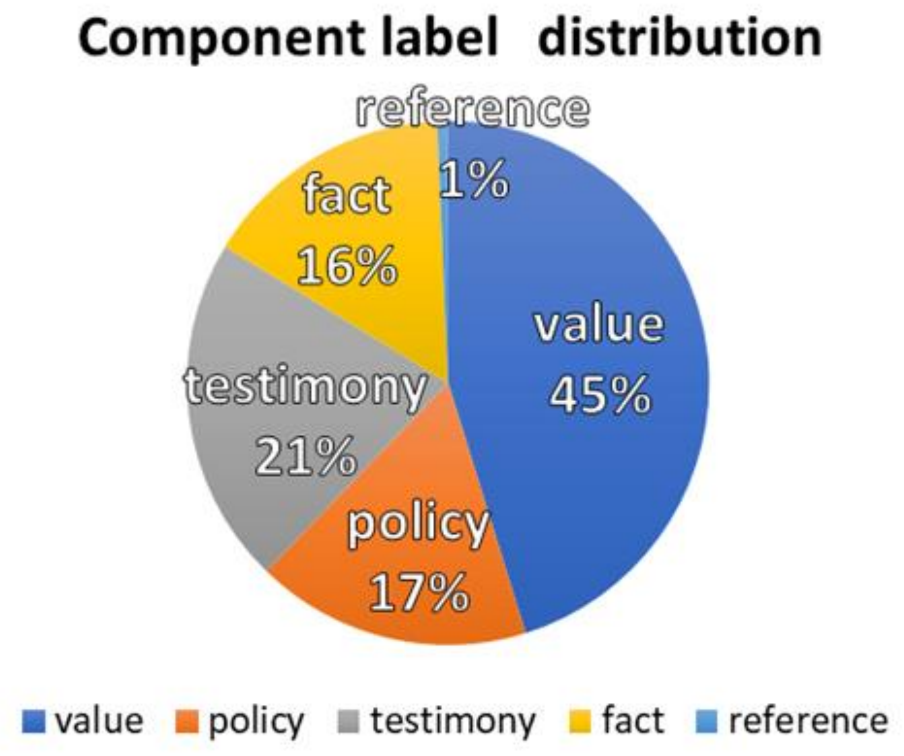


Niculae et al., 2017

Cornell eRulemaking Corpus (CDCP)

- 731 unstructured documents
- 4,779 propositions
  - Avg: 6.5 per document
- 43,384 potential directed links
  - Avg: 59.3 per document
- 1,338 directed links: **3%**
  - Avg: 1.8 per document
  - 97% “reason” labelled links
  - 3% “evidence” labelled links

- Component labels heavily unbalanced:



Niculae et al., 2017

State-of-the-Art: Structured Learning

Structured learning framework that **jointly** classifies all the propositions in a document and determines which ones are linked together

Factor graphs:

- Use first-order and second-order factors
- Relies on a great amount of complex features: lexical, structural, indicators, contextual, syntactic, probability, discourse, embeddings…
- The argumentative model can be imposed

Obtained state-of-the-art results also on another dataset:

UKP Argument Annotated Essays, version 2 (Stab and Gurevych, 2017)

Niculae et al., 2017

Our Approach

Multi-objective learning: all tasks are learnt and performed at the same time

Component Classification, Link prediction, Relation Classification

Local classification: only two propositions are considered at the same time

Minimal set of features, so as to make the approach:

- Domain, model and language agnostic
- Computationally lightweight at pre-process time

Features:

- Pre-trained GloVe embeddings of the words
- Binary encoding of the argumentative distance between pairs of propositions
  - 10 bits to encode positive and negative distances from -5 to +5

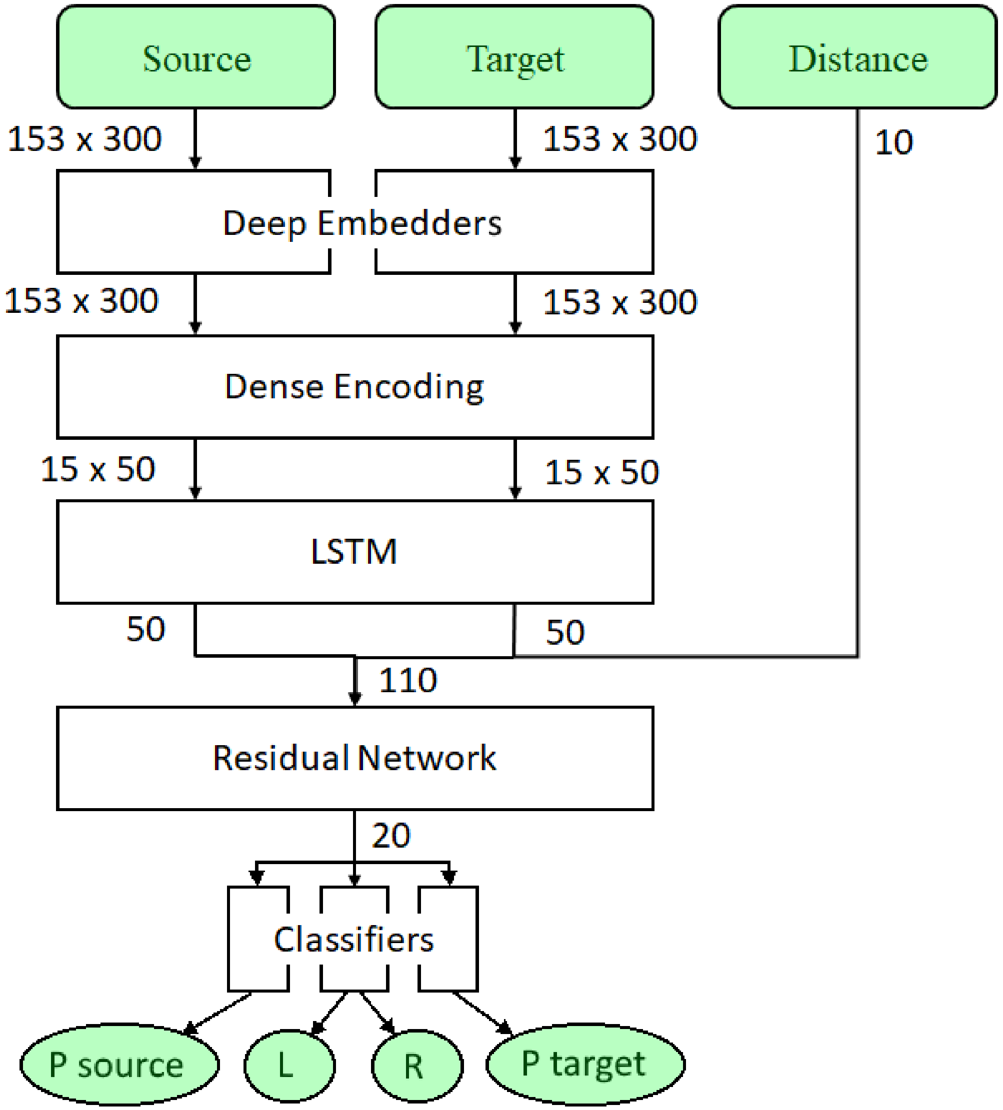
Architecture

Inputs

- GloVe embeddings of two propositions: the source and the target of the potential link
- Encoded distance

Outputs

- Propositions labels
- Link prediction (true/false)
- Link label



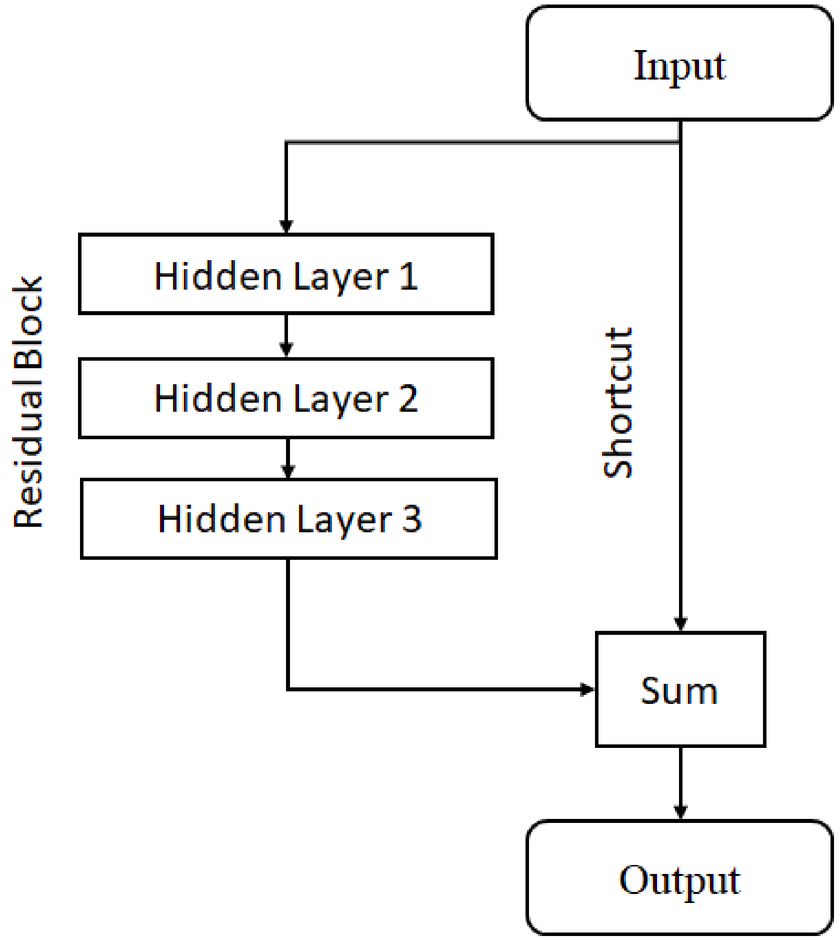
Residual Neural Networks (ResNets)

Deep neural network architecture

Core idea: create shortcuts that link neurons belonging to distant layers

Results:

- speedier training phase
- train networks with a very large number of layers



He et al., 2016

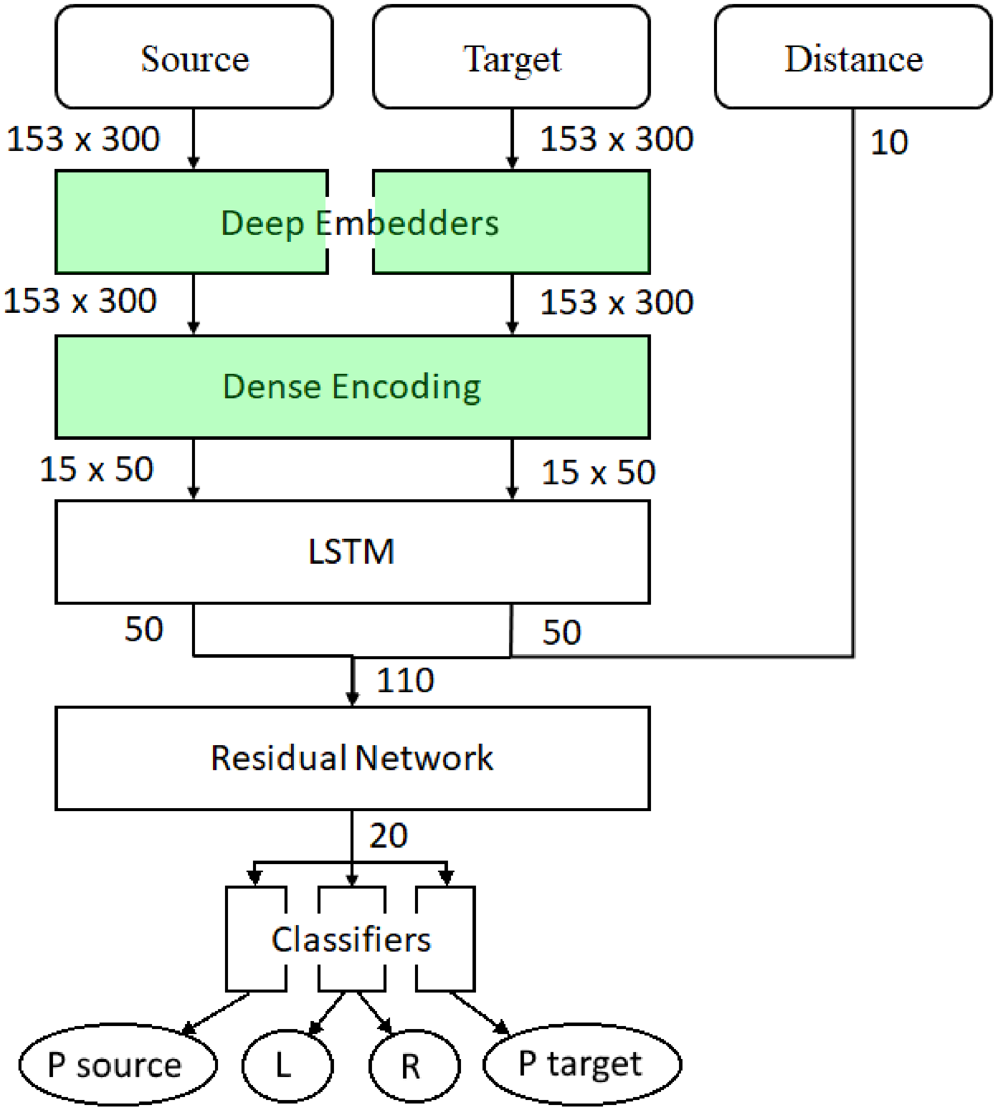
Architecture

2 Deep Embedders: train new embeddings

Residual networks that apply the same transformation to each GloVe embedding, mapping each embedding in a new one

Dense Encoding: reduce dimensionality

Reduce both spatial and temporal dimension through a dense layer and a time average-pooling layers



Architecture

Bi-LSTM

Creates an embedding of the propositions

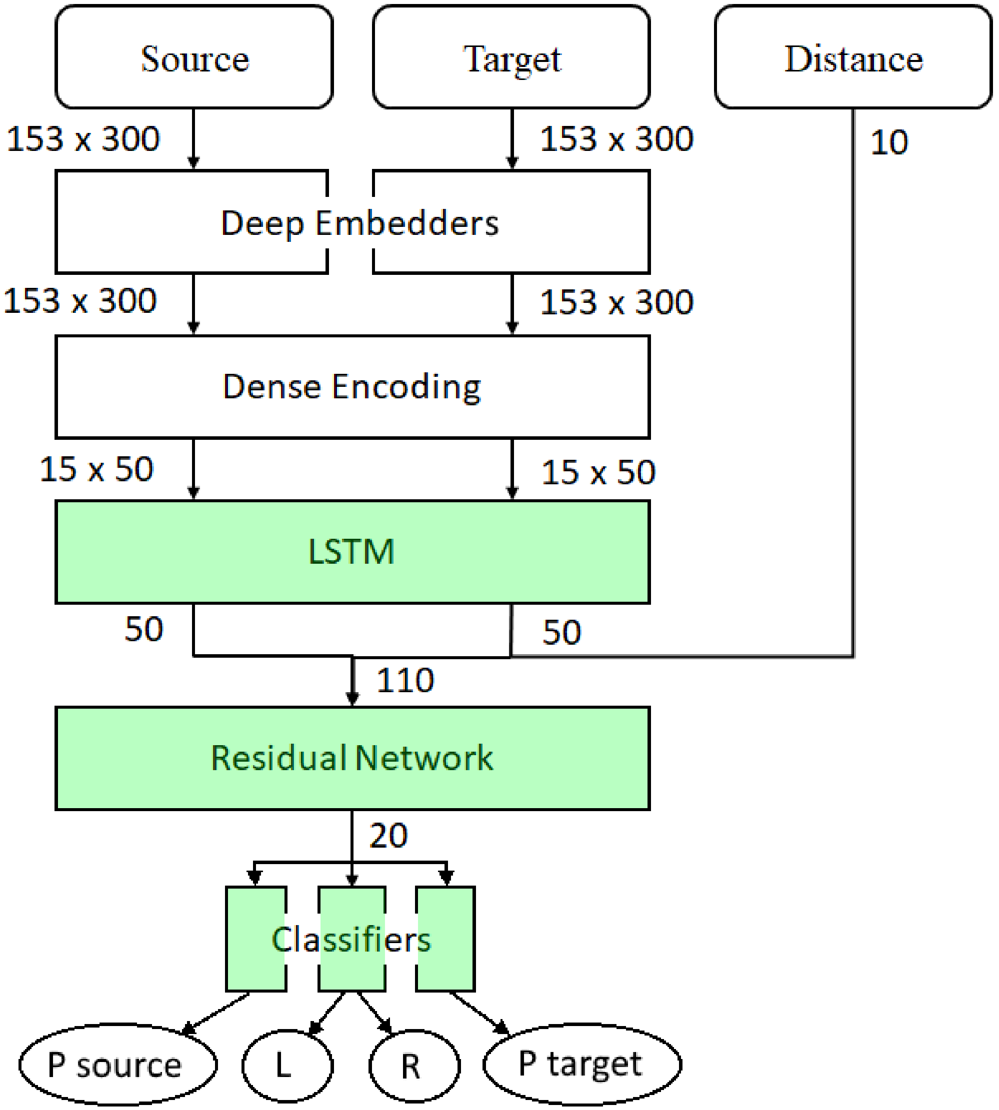
Residual Networks

Elaborates the propositions embeddings and the distance encoding

3 Classifiers

Softmax layers that act in parallel, providing the probability distribution among the classes

The link-prediction is obtained from the relation classification



F1 scores and F1 macro averaged scores	Deep Baseline		Deep Residual		Structured	
Metric	LG	PG	LG	PG	SVM	RNN
Average (Link and Proposition)	33.18	42.88	47.28	46.37	**50.0**	43.5
Link (272)	22.56	22.45	**29.29**	20.76	26.7	14.6
Proposition (973)	43.79	63.31	65.28	71.99	**73.5**	72.7
VALUE (491)	73.77	74.45	72.19	73.24	76.4	73.7
POLICY (153)	73.85	76.09	74.36	76.43	77.3	76.8
TESTIMONY (204)	71.36	65.98	72.86	68.63	71.7	75.8
FACT (124)	0	0	40.31	41.64	42.5	42.2
REFERENCE (1)	0	100	66.67	100	100	100
Relation (272)	11.68	11.52	**15.01**	10.31		
REASON (265)	23.35	23.04	30.02	20.62		
EVIDENCE (7)	0	0	0	0		

Results

The residual architecture outperforms the baseline

Our approach outperforms the state-of-the-art in the link prediction task

The Structured SVM is still better at joint tasks of Component Labelling and Link Prediction

The performance for Relation Classification is poor

Error Analysis

Our misclassification errors for Components Labelling are similar to the state-of-the-art Structured SVM.

ResNet LG		Predicted				
		P	F	T	V	R
True	P	0.76	0.06	0.01	0.17	0.00
	F	0.06	0.42	0.08	0.44	0.00
	T	0.00	0.06	0.75	0.18	0.00
	V	0.07	0.12	0.10	0.70	0.00
	R	0.00	0.00	0.00	0.00	1.00

Structured SVM full		Predicted				
		P	F	T	V	R
True	P	0.76	0.05	0.04	0.16	0.00
	F	0.04	0.44	0.10	0.42	0.00
	T	0.01	0.06	0.72	0.21	0.00
	V	0.05	0.11	0.08	0.76	0.00
	R	0.00	0.00	0.00	0.00	1.00

Conclusion

Our architecture outperforms the non-residual baseline and the state-of-the-art on a difficult dataset

- Without relying on any complex feature or on the document context

Hopefully, it would be easy to integrate this architecture in a more structured and constrained framework

We plan to extend the analysis to other datasets, and integrate other neural architecture components (such as attention)

Thank you for your attention

Details: Experimental setup

Loss Function:

- Misclassification error on Source, Target and Link labels
- L2 regularization factor

Early stopping:

- Validation split: randomly chosen 10% of training documents
- Stopping criterion: no improvement on macro F1 score for 200 epochs
- Two trainings: Link Prediction guided (LG) and Proposition Classification guided (PG)

Baseline: similar architecture without residual connections in its final part

Details: Argumentative Distance

Position of the source proposition relatively to the target proposition, in terms of number of propositions (capped at +5 and -5)

5 bits to indicate positive argumentative distances and 5 to indicate negative ones

The number of consecutive bits is the absolute value of the argumentative distance

- The Hamming distance between two encodings is the absolute value of the difference between two argumentative distances

Proposition	P1	**P2 (source)**	P3	P4	P5
Argumentative Distance	-1	0	1	2	3
Encoding	00001 00000	00000 00000	00000 10000	00000 11000	00000 11100

Details: Component Classification

Proposition are classified multiple times, both as source and target

To classify a proposition, the average score for any possible label is considered

Example:

In a document that contains just two propositions P1 and P2, P1 is classified as follows:

Subject	Role	Destination	p(V)	p(P)	P(T)	P(F)	P(R)
P1	Source of	P2	0.2	0.1	0.5	0.1	0.1
P1	Target of	P2	0.2	0.2	0.2	0.2	0.2
P1			0.2	0.15	**0.35**	0.15	0.15

Details: Link Prediction and Relation Classification

In order to make the class distribution for Relation Classification less unbalanced, the inverse relations are considered. So the classes are:

None (93.8%), Reason (3.0%), inv_Reason (3.0%), Evidence (0.1%) , inv_Evidence (0.1%)

The probability scores for the Link Prediction are derived as the sum of the Relation Classification probability scores

Relation Classification	Reason	Evidence	inv_Reason	inv_Evidence	None
Link Prediction	True		False		

References

- Vlad Niculae, Joonsuk Park, and Claire Cardie. 2017. Argument mining with structured SVMs and RNNs. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*.

- Christian Stab and Iryna Gurevych. 2017. Parsing argumentation structures in persuasive essays. In *Computational Linguistics*.

- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.