NoDE: A Benchmark of Natural Language Arguments

Elena CABRIO [a,b,1], Serena VILLATA [a,2]
[a] *INRIA Sophia Antipolis, France*
, [b] *EURECOM, France*

Abstract. In the latest years, natural models of argumentation and argument mining are becoming more and more important topics in the argumentation community. Given this tendency, there is the need to produce standard datasets on which natural language approaches to argumentation can be evaluated. In this paper, we present NoDE, a benchmark of natural language arguments composed of three datasets, built from different textual sources and annotated highlighting positive and negative connections between arguments.

Keywords. computational linguistics, abstract bipolar argumentation, corpora

Benchmark description

The Web contains huge amounts of texts coming from social media websites like Facebook and Twitter, e-commerce websites, scientific papers, online newspapers, etc. Such online texts are used by people to take decisions. The argumentation community has proposed several frameworks to deal with evidences and justifications, to reason about the arguments and their sources to make decisions. Applying such frameworks to this scenario to support users in dealing with the sheer volume of information overwhelming them may seem straightforward. However, the main problem is that this information is distributed and unstructured. A solution to these problems consists in adopting Natural Language Processing (NLP) approaches to structure, classify and summarize arguments in texts, to enable the subsequent application of formal models of argumentation to reason about the arguments' structure, the relations they have with other arguments, and their justification status. However, in order to apply NLP approaches on texts, the data on which automatic systems can be trained and evaluated should be first collected. This step is unavoidable and time-consuming as it requires a deep, manual analysis and annotation of the pieces of texts relevant to the task. Supervised approaches can then be applied on such data to identify and assign the correct relations among the arguments.

Building a benchmark of natural language arguments is not a straightforward task: *i)* the kind of natural language arguments to be collected from texts should be defined (texts can be extracted from online debates, newspaper articles, blogs and forums, etc.),

[1]The work of E. Cabrio was funded by the French Government (ANR-11-LABX-0031-01).
[2]Corresponding Author: Serena Villata, INRIA Sophia Antipolis, 2004 Route des Lucioles, BP 93, Sophia Antipolis Cedex, France; E-mail: serena.villata@inria.fr.

and *ii)* the annotation guidelines should be defined according to the addressed task from the NLP point of view (e.g., classification, textual entailment, discourse, relations extraction). We present NoDE (Natural language arguments in online DEbates), a benchmark of natural language arguments extracted from different kinds of textual sources.[3] It is composed of three datasets of natural language arguments, released in two machine-readable formats, i.e., the standard XML format, and XML/RDF format adopting the vocabulary SIOC-Argumentation extended in [1]. We have identified three different scenarios for data extraction: *(i)* online debate platforms (Debatepedia[4] and ProCon[5]) present a set of topics to be discussed, and participants argue about the related issues risen by the platform, highlighting whether their "arguments" are in favor or against the central issue, or with respect to the other participants' arguments, *(ii)* the script of a play titled "Twelve Angry Men" where the jurors of a trial discuss in order to decide whether a young boy is guilty or not, and *(iii)* the Wikipedia revision history over a four-year period, focusing on the top five most revised articles. These three scenarios lead to three different resources: the online debates resource collects the arguments in favor or against the main issue or the other arguments into small bipolar argumentation graphs. The same happens for the Wikipedia dataset where the revisions of the articles are used to build small bipolar argumentation graphs. The "Twelve Angry Men" resource collects again pro and con arguments, but the resulting three bipolar argumentation graphs present a higher complexity than the debates graphs. It is interesting to note that bipolar argumentation graphs constructed in NoDE result in quite simple structures, where usually arguments are inserted in reinstatement chains, rather than complex structures with several nested cycles.

Dataset	**pairs**	**supports**	**attacks**	**graphs**	**Inter annotator agreement**		
					annotators	**pairs**	**K**
Debatepedia	260	140	120	22	2	100	0.7
12 Angry Men	80	25	55	3	2	40	0.74
Wiki revisions	452	215	237	416	2	140	0.7

Table 1. Summary of the datasets composing NoDE.

This resource (summarized in Table 1) can be exploited by existing argumentation systems as a collection of examples of argumentation frameworks. In NoDE, the pairs of arguments can be linked either by a positive relation (i.e., a *support* relation in bipolar argumentation [2]), or by a negative relation (i.e., an *attack* relation in argumentation [3]). From these pairs, abstract argumentation graphs are then constructed.

References

[1] E. Cabrio, S. Villata, and F. Gandon. A support framework for argumentative discussions management in the web. In *Procs of The Semantic Web: Semantics and Big Data, ESWC 2013*, pages 412–426, 2013.

[2] C. Cayrol and M.C. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *Procs of ECSQARU, LNCS 3571*, pages 378–389, 2005.

[3] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–358, 1995.

[3] `http://www-sop.inria.fr/NoDE/`

[4] `http://idebate.org/debatabase`

[5] `http://www.procon.org/`