# Combining Textual Entailment and Argumentation Theory for Supporting Online Debates Interactions

**Elena Cabrio** and **Serena Villata**
INRIA
2004 Route des Lucioles BP93
06902 Sophia-Antipolis cedex, France.
{elena.cabrio, serena.villata}@inria.fr

## Abstract

Blogs and forums are widely adopted by online communities to debate about various issues. However, a user that wants to cut in on a debate may experience some difficulties in extracting the current accepted positions, and can be discouraged from interacting through these applications. In our paper, we combine textual entailment with argumentation theory to automatically extract the arguments from debates and to evaluate their acceptability.

## 1 Introduction

Online debate platforms, like Debatepedia[1], Twitter[2] and many others, are becoming more and more popular on the Web. In such applications, users are asked to provide their own opinions about selected issues. However, it may happen that the debates become rather complicated, with several arguments supporting and contradicting each others. Thus, it is difficult for potential participants to understand the way the debate is going on, i.e., which are the current accepted arguments in a debate. In this paper, we propose to support participants of online debates with a framework combining Textual Entailment (TE) (Dagan et al., 2009) and abstract argumentation theory (Dung, 1995). In particular, TE is adopted to extract the abstract arguments from natural language debates and to provide the relations among these arguments; argumentation theory is then used to compute the set of accepted arguments among those obtained from the TE module, i.e., the arguments shared by the majority of the participants without being attacked by other accepted arguments. The originality of the proposed framework lies in the combination of two existing approaches with the goal of supporting participants in their interactions with online debates, by automatically detecting the arguments in natural language text, and identifying the accepted ones. We evaluate the feasibility of our combined approach on a set of arguments extracted from a sample of Debatepedia.

[1] http://debatepedia.idebate.org
[2] http://twitter.com/

## 2 First step: textual entailment

TE was proposed as an applied framework to capture major semantic inference needs across applications in NLP, e.g. (Romano et al., 2006; Barzilay and McKeown, 2005; Nielsen et al., 2009). It is defined as a relation between two textual fragments, i.e., the text (*T*) and the hypothesis (*H*). Entailment holds if the meaning of H can be inferred from the meaning of T, as interpreted by a typical language user. Consider the pairs in Example 1 and 2.

**Example 1.**
***T1:*** *Research shows that drivers speaking on a mobile phone have much slower reactions in braking tests than non-users, and are worse even than if they have been drinking.*
***H:****The use of cell-phones while driving is a public hazard.*

**Example 2** (Continued)**.**
***T2:*** *Regulation could negate the safety benefits of having a phone in the car. When you're stuck in traffic, calling to say you'll be late can reduce stress and make you less inclined to drive aggressively to make up lost time.*
***H:****The use of cell-phones while driving is a public hazard.*



*Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 208–212,
Jeju, Republic of Korea, 8-14 July 2012. 

A system aimed at recognizing TE should detect an entailment relation between T1 and H (Example 1), and a contradiction between T2 and H (Example 2).

As introduced before, our paper proposes an approach to support the participants in forums or debates to detect the accepted arguments among those expressed by the other participants on a certain topic. As a first step, we need to *(i)* automatically recognize a participant's opinion on a certain topic as an argument, as well as to *(ii)* detect its relationship with the other arguments. We therefore cast the described problem as a TE problem, where the T-H pair is a pair of arguments expressed by two different participants on a certain topic. For instance, given the argument "The use of cell-phones while driving is a public hazard" (that we consider as H as a starting point), participants can support it expressing arguments from which H can be inferred (Example 1), or can contradict such argument with opinions against it (Example 2). Since in debates arguments come one after the other, we extract and compare them both with respect to the main issue, and with the other participants' arguments (when the new argument entails or contradicts one of the arguments previously expressed by another participant). For instance, given the same debate as before, a new argument T3 may be expressed by a third participant with the goal of contradicting T2 (that becomes the new H (H1) in the pair), as shown in Example 3.

**Example 3** (Continued).
***T3:*** *If one is late, there is little difference in apologizing while in their car over a cell phone and apologizing in front of their boss at the office. So, they should have the restraint to drive at the speed limit, arriving late, and being willing to apologize then; an apologetic cell phone call in a car to a boss shouldn't be the cause of one being able to then relax, slow-down, and drive the speed-limit.*
***T2*** $\rightarrow$ ***H1:*** *Regulation could negate the safety benefits of having a phone in the car. When you're stuck in [...]*

TE provides us with the techniques to detect both the arguments in a debate, and the kind of relation underlying each couple of arguments. The TE system returns indeed a judgment (entailment or contradiction) on the arguments' pairs, that are used as input to build the argumentation framework, as described in the next Section.

## 3 Second step: argumentation theory

Starting from a set of arguments and the attacks (i.e., conflicts) among them, a (Dung, 1995)-style argumentation framework allows to detect which are the accepted arguments. Such arguments are considered as believable by an external evaluator who has a full knowledge of the argumentation framework, and they are determined through the acceptability semantics (Dung, 1995). Roughly, an argument is accepted, if all the arguments attacking it are rejected, and it is rejected if it has at least an argument attacking it which is accepted. An argument which is not attacked at all is accepted.

**Definition 1.** *An abstract argumentation framework (AF) is a pair* $\langle \mathcal{A}, \rightarrow \rangle$ *where* $\mathcal{A}$ *is a set of arguments and* $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$ *is a binary relation called attack.*

Aim of the argumentation-based reasoning step is to provide the participant with a complete view on the arguments proposed in the debate, and to show which are the accepted ones. In our framework, we first map contradiction with the attack relation in abstract argumentation; second, the entailment relation is viewed as a support relation among abstract arguments. The support relation (Cayrol and Lagasquie-Schiex, 2011) may be represented as: (1) a relation among the arguments which does not affect their acceptability, or (2) a relation among the arguments which leads to the introduction of additional attacks.

Consider a support relation among two arguments, namely $A_i$ and $A_j$. If we choose (1), an attack towards $A_i$ or $A_j$ does not affect the acceptability of $A_j$ or $A_i$, respectively. If we choose (2), we introduce additional attacks, and we have the following two options: [Type 1] $A_i$ *supports* $A_j$ then $A_k$ *attacks* $A_j$, and [Type 2] $A_i$ *supports* $A_j$ then $A_k$ *attacks* $A_i$. The attacks of type 1 are due to inference: $A_i$ entails $A_j$ means that $A_i$ is more specific of $A_j$, thus an attack towards $A_j$ is an attack also towards $A_i$. The attacks of type 2, instead, are more rare, but they may happen in debates: an attack towards the more specific argument $A_i$ is an attack towards the more general argument $A_j$. In Section 4, we will consider only the introduction of attacks of type 1.

For Examples 1, 2, and 3, the TE phase returns the following couples: T1 *entails* H, T2 *attacks* H, T3 *attacks* H1 (i.e. T2). The argumentation module maps each element to its corresponding argument: H $\equiv A_1$, T1 $\equiv A_2$, T2 $\equiv A_3$, and T3 $\equiv A_4$ . The resulting $AF$ (Figure 1) shows that the accepted arguments are $\{A_1, A_2, A_4\}$, meaning that the issue "The use of cell-phones while driving is a public hazard" ($A_1$) is considered as accepted. Figure 2 visualizes the complete framework of the debate "Use of cell phones while driving" on Debatepedia. Accepted arguments are double bordered.

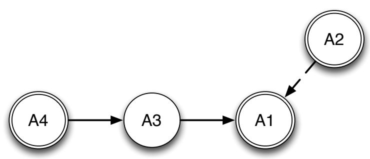


Figure 1: The $AF$ built from the results of the TE module for Example 1, 2 and 3, without introducing additional attacks. Plain arrows represent *attacks*, dashed arrows represent *supports*.

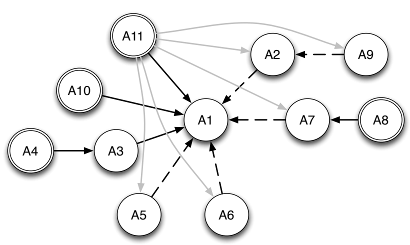


Figure 2: The $AF$ built from the results of the TE module for the entire debate. Grey attacks are of type 1. For picture clarity, we introduce type 1 attacks only from $A_{11}$. The same attacks hold from $A_{10}$ and $A_3$.

## 4 Experimental setting

We experiment the combination of TE and argumentation theory to support the interaction of online debates participants on Debatepedia, an encyclopedia of pro and con arguments on critical issues.

**Data set.** To create the data set of arguments pairs to evaluate our task[3], we randomly selected a set of topics (reported in column *Topics*, Table 1) of Debatepedia debates, and for each topic we coupled all the pros and cons arguments both with the main argument (the issue of the debate, as in Example 1 and 2) and/or with other arguments to which the most recent argument refers, e.g., Example 3. Using Debatepedia as case study provides us with already annotated arguments (*pro* ⇒ *entailment*[4], and *cons* ⇒ *contradiction*), and casts our task as a yes/no entailment task. As shown in Table 1, we collected 200 T-H pairs, 100 used to train the TE system, and 100 to test it (each data set is composed by 55 entailment and 45 contradiction pairs).[5] Test set pairs concern completely new topics, never seen by the system.

**TE system.** To detect which kind of relation underlies each couple of arguments, we used the EDITS system (Edit Distance Textual Entailment Suite), an open-source software package for recognizing TE[6] (Kouylekov and Negri, 2010). EDITS implements a distance-based framework which assumes that the probability of an entailment relation between a given T-H pair is inversely proportional to the distance between T and H. Within this framework, the system implements different approaches to distance computation, providing both edit distance algorithms and similarity algorithms.

**Evaluation.** To evaluate our combined approach, we carry out a two-step evaluation: we assess *(i)* the performances of the TE system to correctly assign the entailment/contradiction relations to the pairs of arguments in the Debatepedia data set; *(ii)* how much such performances impact on the goals of the argumentation module, i.e. how much a wrong assignment of a relation between two arguments leads to an incorrect evaluation of the accepted arguments.

For the first evaluation, we run the EDITS system off-the-shelf on the Debatepedia data set, applying one of its basic configurations (i.e. the distance entailment engine combines cosine similarity as the core distance algorithm; distance calculated on lemmas; stopword list included). EDITS accuracy on the training set is 0.69, on the test set 0.67 (a baseline applying a Word Overlap algorithm on tokenized text is also considered, and obtains an accuracy of 0.61 on the training set and 0.62 on the test set). Even using a basic configuration of EDITS, and a small data set (100 pairs for training) performances

[3]Data available for the RTE challenges are not suitable for our goal, since the pairs are extracted from news and are not linked among each other (they do not report opinions on a certain topic). `http://www.nist.gov/tac/2010/RTE/`

[4]Arguments "supporting" another argument without inference are left for future work.

[5]Available at `http://bit.ly/debatepedia_ds`

[6]Version 3.0 available at `http://edits.fbk.eu/`

| Training set | | | | | Test set | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Topic** | **#argum** | **#pairs** | | | **Topic** | **#argum** | **#pairs** | | |
| | | **TOT.** | **yes** | **no** | | | **TOT.** | **yes** | **no** |
| *Violent games boost aggressiveness* | 16 | 15 | 8 | 7 | *Ground zero mosque* | 9 | 8 | 3 | 5 |
| *China one-child policy* | 11 | 10 | 6 | 4 | *Mandatory military service* | 11 | 10 | 3 | 7 |
| *Consider coca as a narcotic* | 15 | 14 | 7 | 7 | *No fly zone over Libya* | 11 | 10 | 6 | 4 |
| *Child beauty contests* | 12 | 11 | 7 | 4 | *Airport security profiling* | 9 | 8 | 4 | 4 |
| *Arming Libyan rebels* | 10 | 9 | 4 | 5 | *Solar energy* | 16 | 15 | 11 | 4 |
| *Random alcohol breath tests* | 8 | 7 | 4 | 3 | *Natural gas vehicles* | 12 | 11 | 5 | 6 |
| *Osama death photo* | 11 | 10 | 5 | 5 | *Use of cell phones while driving* | 11 | 10 | 5 | 5 |
| *Privatizing social security* | 11 | 10 | 5 | 5 | *Marijuana legalization* | 17 | 16 | 10 | 6 |
| *Internet access as a right* | 15 | 14 | 9 | 5 | *Gay marriage as a right* | 7 | 6 | 4 | 2 |
| | | | | | *Vegetarianism* | 7 | 6 | 4 | 2 |
| **TOTAL** | 109 | **100** | **55** | **45** | **TOTAL** | 110 | **100** | **55** | **45** |

Table 1: The Debatepedia data set.

on Debatepedia test set are promising, and in line with performances of TE systems on RTE data sets.

As a second step of the evaluation, we consider the impact of EDITS performances on arguments acceptability, i.e., how much a wrong assignment of a relation to a pair of arguments affects the computation of the set of accepted arguments. We identify the accepted arguments both in the correct $AF$ of each Debatepedia debate of the data set (the gold-standard, where relations are correctly assigned), and on the $AF$ generated basing on the relations assigned by EDITS. Our combined approach obtained the following performances: precision 0.74, recall 0.76, accuracy 0.75, meaning that the TE system mistakes in relation assignment propagate in the $AF$, but results are still satisfying and foster further research in this direction.

## 5 Related work

DebateGraph[7] is an online system for debates, but it is not grounded on argument theory to decide the accepted arguments. Chasnevar and Maguitman's (2004) system provides recommendations on language patterns using indices computed from Web corpora and defeasible argumentation. No NLP is used for automatic arguments detection. Carenini and Moore (2006) present a computational framework to generate evaluative arguments. Based on users' preferences, arguments are produced following argumentation guidelines to structure evaluative arguments. Then, NL Generation techniques are applied to return the argument in natural language. Unlike them, we do not create the arguments, but we use TE to detect them in texts, and we use Dung's model to identify the accepted ones. Wyner and van Engers (2010) present a policy making support tool based on forums, where NLP and argumentation are coupled to provide well structured statements. Beside the goal, several points distinguish our proposal from this one: *(i)* the user is asked to write the input text using Attempt to Controlled English, with a restricted grammar and vocabulary, while we do not support the participant in writing the text, but we automatically detect the arguments (no language restriction); *(ii)* a mode indicates the relations between the statements, while we infer them using TE; *(iii)* no evaluation of their framework is provided.

## 6 Future challenges

Several research lines are considered to improve the proposed framework: first, the use of NLP to detect the arguments from text will make argumentation theory applicable to reason in real scenarios. We plan to use the TE module to reason on the introduction of the support relation in abstract argumentation theory. We plan to extend our model by considering also other kinds of relationships among the arguments. Moreover, given the promising results we obtained, we plan to extend the experimentation setting both increasing the size of the Debatepedia data set, and to improve the TE system performances to apply our combined approach in other real applications (considering for instance the presence of unrelated arguments, e.g. texts that do not entail nor contradict).

[7] `http://debategraph.org`

## References


Barzilay R. and McKeown K.R. 2005. *Sentence fusion for multidocument news summarization*. Computational Linguistics, 31(3). pp. 297-327.

Carenini G. and Moore J.D. 2006. *Generating and evaluating evaluative arguments*. Artificial Intelligence, volume 170, n. 11. pp. 925-952.

Cayrol C. and Lagasquie-Schiex M.C. 2011. *Bipolarity in Argumentation Graphs: Towards a Better Understanding*. Proceedings of SUM 2011. pp.137-148

Chesñevar C.I. and Maguitman A.G. 2004. *An Argumentative Approach to Assessing Natural Language Usage based on the Web Corpus*. Proceedings of ECAI. pp.581-585.

Dagan I. and Dolan B. and Magnini B. and Roth D. 2009. *Recognizing textual entailment: Rational, evaluation and approaches*. Natural Language Engineering (JNLE), Special Issue 04, volume 15. pp. i-xvii. Cambridge University Press.

Dung P.M. 1995. *On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games*. Artificial Intelligence, volume 77, n.2. pp.321-358.

Kouylekov M. and Negri M. 2010. *An Open-Source Package for Recognizing Textual Entailment*. Proceedings of ACL 2010 System Demonstrations. pp.42-47.

Nielsen R.D. and Ward W. and Martin J.H. 2009. *Recognizing entailment in intelligent tutoring systems*. The Journal of Natural Language Engineering, (JNLE), volume 15. pp. 479-501. Cambridge University Press.

Romano L. and Kouylekov M. O. and Szpektor I. and Dagan I. and Lavelli A. 2006. *Investigating a Generic Paraphrase-Based Approach for Relation Extraction*. Proceedings of EACL 2006. pp. 409-416.

Wyner A. and van Engers T. 2010. *A framework for enriched, controlled on-line discussion forums for e-government policy-making*. Proceedings of eGov 2010.