Natural Language Arguments: A Combined Approach

Elena Cabrio[1] and **Serena Villata** [2 3]

Abstract. With the growing use of the Social Web, an increasing number of applications for exchanging opinions with other people are becoming available online. These applications are widely adopted with the consequence that the number of opinions about the debated issues increases. In order to cut in on a debate, the participants need first to evaluate the opinions in favour or against the debated issue. Argumentation theory proposes algorithms and semantics to evaluate the set of accepted arguments, given the conflicts among them. The main problem is how to automatically generate the arguments from the natural language formulation of the opinions used in these applications. Our paper addresses this problem by proposing and evaluating the use of natural language techniques to generate the arguments. In particular, we adopt the textual entailment approach, a generic framework for applied semantics, where linguistic objects are mapped by means of semantic inferences at a textual level. We couple textual entailment together with a Dung-like argumentation system which allows us to identify the arguments that are accepted in the considered online debate. The originality of the proposed framework lies in the following point: natural language debates are analyzed and the arguments are automatically extracted.

1 Introduction

In the latest years, the Web has changed in the so called Social Web. The Social Web has seen an increasing number of applications like Twitter[4], Debatepedia[5], Facebook[6] and many others, which allow people to express their opinions about different issues. Consider for instance the following debate published on Debatepedia: the issue of the debate is "Making Internet a right only benefits society". The participants have proposed various pros and cons arguments concerning this issue, e.g., a pro argument claims that the Internet delivers freedom of speech, and a con argument claims that the Internet is not as important as real rights like the freedom from slavery. These kinds of debates are composed by tens of arguments in favour or against a proposed issue. The main difficulty for newcomers is to understand the current holding position in the debate, i.e., to understand which are the arguments that are accepted at a certain moment. This difficulty is twofold: first, the participants have to remember all the different, possibly long, arguments and understand which are the relationships among these arguments, and second they have to be updated with respect to the inner dynamics of such kind of applications.

In this paper, we answer the following research question: how to support the participants in natural language debates to detect which arguments are accepted? The research question breaks down into the following subquestions: (1) How to automatically generate the arguments, as well as their relationships, from natural language debates?, and (2) How to detect which are the accepted arguments?

In this paper, we propose to combine natural language techniques and Dung-like abstract argumentation to generate the arguments from natural language text and then to evaluate this set of arguments to know which are the accepted ones. First, starting from the participants' opinions, we detect which ones imply or contradict, even indirectly, the issue of the debate using the textual entailment approach. Beside formal approaches to semantic inference that rely on logical representation of meaning, the notion of Textual Entailment (TE) has been proposed as an applied framework to capture major semantic inference needs across applications in the Computational Linguistics field [9]. The development of the Web has witnessed a paradigm shift, due to the need to process a huge amount of available (but often noisy) data. TE is a generic framework for applied semantics, where linguistic objects are mapped by means of semantic inferences at a textual level. We use TE to automatically identify, from a natural language text, the arguments. Second, we adopt abstract Dung-like argumentation theory [10] to reason over the set of generated arguments with the aim of deciding which are the accepted ones. Proposals like argumentation schemes [18], Araucaria [19], Carneades [11], and ArguMed [21] use natural language arguments, but they ask the participants to indicate the semantic relationship among the arguments, and the linguistic content remains unanalyzed. As underlined by Reed and Grasso [18], "the goal machinery that leads to arguments being automatically generated has been only briefly touched upon, and yet is clearly fundamental to the endeavor". Finally, we combine the two approaches, i.e., textual entailment and abstract argumentation theory, in a framework whose aim is to *(i)* generate the abstract arguments from the online debates through TE, *(ii)* build the argumentation framework from the arguments and the relationships returned by the TE module, and *(iii)* return the set of accepted arguments. We evaluate the feasibility of our combined approach on a data set extracted from a sample of Debatepedia debates.

The originality of the proposed framework consists in the combination of two techniques which need each other to provide a complete reasoning model: TE has the power to automatically identify the arguments in the text and to specify which kind of relation links each couple of arguments, but it cannot assess which are the arguments that are considered as accepted. This is addressed by argumentation theory which lacks automatic techniques to extract the arguments from free texts. The combination of these two approaches leads to the definition of a powerful tool to reason over online debates.

The reminder of the paper is as follows. Sections 2 and 3 present the fundamentals of textual entailment and argumentation theory. In Section 4, we describe the experimental setting as well as its evaluation. Section 5 compares the proposed approach to the related work.

[1] INRIA Sophia Antipolis, France, email: elena.cabrio@inria.fr
[2] INRIA Sophia Antipolis, France, email: serena.villata@inria.fr
[3] Acknowledges support of the DataLift Project ANR-10-CORD-09.
[4] `http://twitter.com/`
[5] `http://www.debatepedia.org/`
[6] `http://www.facebook.com/`

2 NLP approaches to semantic inference

Classical approaches to semantic inference rely on logical representations of meaning that are external to the language itself, and are typically independent of the structure of any particular natural language. Texts are first translated, or interpreted, into some logical form and then new propositions are inferred from interpreted texts by a logical theorem prover. But, especially after the development of the Web, we have witnessed a paradigm shift, due to the need to process a huge amount of available (but often noisy) data. Addressing the inference task by means of logical theorem provers in automated applications aimed at natural language understanding has shown several intrinsic limitations [2]. As highlighted in Monz and de Rijke [15], in formal approaches semanticists generally opt for rich (i.e. including at least first order logic) representation formalisms to capture as many relevant aspects of the meaning as possible, but practicable methods for generating such representations are very rare. The translation of real-world sentences into logic is difficult because of issues such as ambiguity or vagueness [16]. Moreover, the computational costs of deploying first-order logic theorem prover tools in real world situations may be prohibitive, and huge amounts of additional linguistic and background knowledge are required. Formal approaches address forms of deductive reasoning, and therefore often exhibit a too high level of precision and strictness as compared to human judgments, that allow for uncertainties typical of inductive reasoning [4]. While it is possible to model elementary inferences on the precise level allowed by deductive systems, many pragmatic aspects that play a role in everyday inference cannot be accounted for. Inferences that are plausible but not logically stringent cannot be modeled in a straightforward way, but in NLP applications approximate reasoning should be preferred in some cases to having no answers at all.

Especially in data-driven approaches, like the one sought in this work, where patterns are learnt from large-scale naturally-occurring data, we can settle for approximate answers provided by efficient and robust systems, even at the price of logic unsoundness or incompleteness. Starting from these considerations, Monz and de Rijke [15] propose to address the inference task directly at the textual level instead, exploiting currently available NLP techniques. While methods for automated deduction assume that the arguments in input are already expressed in some formal meaning representation (e.g. first order logic), addressing the inference task at a textual level opens different and new challenges from those encountered in formal deduction. Indeed, more emphasis is put on informal reasoning, lexical semantic knowledge, and variability of linguistic expressions.

The notion of Textual Entailment has been proposed as an applied framework to capture major semantic inference needs across applications in NLP [9]. It is defined as a relation between a coherent textual fragment (the Text T) and a language expression, which is considered as the Hypothesis (H). Entailment holds (i.e. $T \Rightarrow H$) if the meaning of H can be inferred from the meaning of T, as interpreted by a typical language user. The TE relationship is directional, since the meaning of one expression may usually entail the other, while the opposite is much less certain. Consider the pairs in Example 1 and 2.

Example 1

T1*: Internet access is essential now; must be a right. The internet is only that wire that delivers freedom of speech, freedom of assembly, and freedom of the press in a single connection.*

H*: Making Internet a right only benefits society.*

Example 2 (Continued)

T2*: Internet not as important as real rights. We may think of such trivial things as a fundamental right, but consider the truly impoverished and what is most important to them. The right to vote, the right to liberty and freedom from slavery or the right to elementary education.*

H*: Making Internet a right only benefits society.*

A system aimed at recognizing textual entailment should detect an inference relation between T1 and H (i.e. the meaning of H can be derived from the meaning of T) in Example 1, while it should detect a contradiction between T2 and H in Example 2. In this applied framework, inferences are performed directly over lexical-syntactic representations of the texts [9]. Such definition of TE captures quite broadly the reasoning about language variability needed by different applications aimed at natural language understanding and processing, e.g. information extraction [20] or text summarization [1].

The goal of our paper is to propose an approach to support the participants in forums or debates (e.g. Debatepedia, Twitter) to detect which arguments among the ones expressed by the other participants on a certain topic are accepted. As a first step, we need to *(i)* automatically generate the arguments (i.e. recognize a participant opinion on a certain topic as an argument), as well as *(ii)* detect their relation with respect to 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 in a debate on a certain topic. For instance, given the argument "Making Internet a right only benefits society" (that we consider as H as a starting point), participants can be in favor of it (expressing arguments from which H can be inferred, as in Example 1), or can contradict such argument (expressing an opinion against it, as in Example 2). Since in debates one participant's argument comes after the other, we can extract such arguments and compare them both w.r.t. the main issue, and w.r.t. 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 to contradict T2 (that becomes the new H (H1) in the pair), as shown in Example 3.

Example 3 (Continued)

T3*: I've seen the growing awareness within the developing world that computers and connectivity matter and can be useful. It's not that computers matter more than water, food, shelter and healthcare, but that the network and PCs can be used to ensure that those other things are available. Satellite imagery sent to a local computer can help villages find fresh water, mobile phones can tell farmers the prices at market so they know when to harvest.*

T2 $\equiv$ ***H1****: Internet not as important as real rights. We may think of such trivial things as a fundamental right, but consider the truly impoverished and what is most important to them. The right to vote, the right to liberty and freedom from slavery or the right to elementary education.*

With respect to the goal of our work, TE provides us with the techniques to identify the arguments in a debate, and to detect which kind of relationship underlies each couple of arguments. A TE system returns indeed a judgment (entailment or contradiction) on the arguments' pairs related to a certain topic, that are used as input to build the argumentation framework, as described in the next Section.

3 Argumentation theory

This section provides the basic concepts and insights of Dung's abstract argumentation [10].

Definition 1 *(Abstract argumentation framework AF) An abstract argumentation framework is a pair $\langle \mathcal{A}, \rightarrow \rangle$ where $\mathcal{A}$ is a set of elements called arguments and $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation called attack. We say that an argument A_i attacks an argument A_j if and only if $(A_i, A_j) \in \rightarrow$.*

Dung [10] presents several acceptability semantics that produce zero, one, or several sets of accepted arguments. Such semantics are grounded on two main concepts called conflict-freeness and defence.

Definition 2 *(Conflict-free, Defence) Let $C \subseteq \mathcal{A}$. A set C is conflict-free if and only if there exist no $A_i, A_j \in C$ such that $A_i \rightarrow A_j$. A set C defends an argument A_i if and only if for each argument $A_j \in \mathcal{A}$ if A_j attacks A_i then there exists $A_k \in C$ such that A_k attacks A_j.*

Definition 3 *(Acceptability semantics) Let C be a conflict-free set of arguments, and let $\mathcal{D} : 2^{\mathcal{A}} \mapsto 2^{\mathcal{A}}$ be a function such that $\mathcal{D}(C) = \{A | C\ defends\ A\}$.*

- *C is admissible if and only if $C \subseteq \mathcal{D}(C)$.*
- *C is a complete extension if and only if $C = \mathcal{D}(C)$.*
- *C is a grounded extension if and only if it is the smallest (w.r.t. set inclusion) complete extension.*
- *C is a preferred extension if and only if it is a maximal (w.r.t. set inclusion) complete extension.*
- *C is a stable extension if and only if it is a preferred extension that attacks all arguments in $\mathcal{A} \setminus C$.*

The textual entailment step returns a set of couples of the kind: argument A_i is in contradiction with argument A_j, or argument A_i entails argument A_j. The aim of the argumentation-based reasoning step is to provide to the participant the whole view on the arguments, i.e., opinions, proposed in the debate, and to show which are the accepted ones, w.r.t. a particular Dung's semantics. To this aim, we have to represent the two relations among arguments in our argumentation-based setting. First, we map the contradiction with the attack relation in abstract argumentation. Second, the entailment relation may be viewed as a sort of "support" relation among abstract arguments. Due to the various conflicting views on the meaning of support in argumentation theory [6], in this paper we propose two possible solutions to represent the entailment relation, without taking a position in this debate. The entailment relation can be represented as: (1) a relationship among the arguments which does not effect their acceptability, or (2) a relationship among the arguments which leads to the introduction of additional attacks among the arguments involved in this relation, thus effecting their acceptability. In particular, following the different proposals in the argumentation literature concerning the support relation [7, 3], the entailment relation may lead to the introduction of the following two kinds of attacks:

Definition 4 (Entailment attacks) *Let $A_j, A_k \in \mathcal{A}$. An entailment attack of type 1 for A_j by A_k is a sequence $A_1 R_1 \ldots R_{n-2} A_{n-1}$ and $A_n R_{n-1} A_{n-1}$, $n \geqslant 3$, with $A_1 = A_j, A_n = A_k$, such that $R_{n-1} =$ attack and $\forall i =$ 1 ... n-2, $R_i =$ entailment. An entailment attack of type 2 for A_j by A_k is a sequence $A_1 R_1 \ldots R_{n-1} A_n$, $n \geqslant 3$, with $A_1 = A_k, A_n = A_j$, such that $R_1 =$ attack and $\forall i =$ 2 ... n-1, $R_i =$ entailment.*

The idea behind entailment attacks is the following: given an entailment relationship among two arguments, namely A_i and A_j, what happens when there is an attack against A_i or A_j? If we choose solution 1, an attack towards A_i or A_j does not effect the acceptability of A_j or A_i, respectively. If we choose solution 2, i.e., we choose to introduce additional attacks, then we have the following two options: [Type 1] A_i *entails* A_j and A_k *attacks* A_j, and [Type 2] A_i *entails* A_j and A_k *attacks* A_i. The attacks of type 1 are due to the inference relation: the fact that A_i entails A_j means that A_i is more specific than A_j, thus an attack towards A_j is an attack also towards A_i. The attacks of type 2, instead, are more rare, even if they may happen in the debates: an attack towards the more specific argument A_i is an attack towards the more general argument A_j. Attacks of type 1 are equivalent to mediated attacks [3], and attacks of type 2 are equivalent to secondary attacks [7]. In the experimental setting of Section 5, we will consider only the introduction of entailment attacks of type 1. Entailment attacks of type 2 may be introduced, dependently on the participant's needs and on the specific debate. Note that the goodness of the introduction of type 2 attacks may be automatically evaluated by the TE step. Enhancing our framework to include this kind of evaluation is left as future work.

Example 4 (Continued) *The textual entailment phase returns the following couples for the natural language opinions detailed in Example 1, 2, and 3:*

- *T1 entails H*
- *T2 attacks H*
- *T3 attacks H1 (i.e., T2)*

Given this result, the argumentation module of our framework 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 argumentation framework, visualized in Figure 1, shows that the accepted arguments using complete semantics[7] are $\{A_1, A_2, A_4\}$. This means that the issue "Making Internet a right only benefits society" A_1 is considered as accepted. Figure 2 visualizes the complete argumentation framework, from the debate about the subject "Internet access as a right" on Debatepedia, as it is returned by the TE module. Double bordered arguments are the accepted ones.

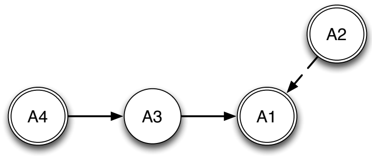

Figure 1. The AF built from the results of the TE module for Example 1, 2 and 3, without introducing entailment attacks. Plain arrows represent *attacks*, dashed arrows represent *entails*.

4 Experimental setting

As a case study to experiment the combination of TE and argumentation theory approaches to support the interaction of participants in online debates, we select Debatepedia, an encyclopedia of pro and

[7] The proposed framework can adopt all admissibility-based semantics.

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.

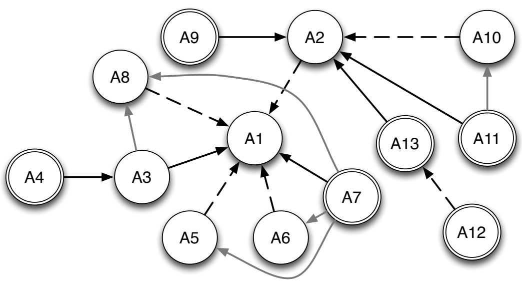

Figure 2. The AF built from the results of the TE module for the debate about "Internet access as a right". Grey attacks are type 1 attacks introduced to compute the set of accepted arguments. For clarity of the picture, we visualize only a subset of type 1 attacks.

con arguments on critical issues. In Section 4.1 we describe the creation of the data set of T-H pairs extracted from a sample of Debatepedia debate topics, in Section 4.2 we present the TE system we used, and in Section 4.3 we report on obtained results.

4.1 Data set

To create the data set of arguments pairs to evaluate our task, we follow the criteria defined and used by the organizers of the Recognizing Textual Entailment Challenges (RTE).[8] To test the progress of TE systems in a comparable setting, the participants to RTE are provided with data sets composed of T-H pairs involving various levels of entailment reasoning (e.g. lexical, syntactic), and TE systems are required to produce a correct judgment on the given pairs (i.e. to say if the meaning of one text snippet can be inferred from the other). The 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 (i.e. they do not report opinions on a certain topic).

For this reason, we created a data set to evaluate our combined approach focusing on Debatepedia[9], as reported in Table 1. We manually selected a set of topics (reported in column *Topics*) of Debatepedia debates, and for each topic we coupled all the pro and con arguments both with the main argument (the title 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*[10], and *con* ⇒ *contradiction*), and casts our task as a yes/no entailment task. We collected 200 T-H pairs (Table 1), 100 to train the TE system, and 100 to test it (each data set is composed by 55 entailment and 45 contradiction pairs). The pairs considered for the test set concern completely new topics, never seen by the system.

[8] http://www.nist.gov/tac/2010/RTE/

[9] The data set is freely available at http://bit.ly/debatepedia_ds.

4.2 TE system

To detect which kind of relation underlies each couple of arguments, we take advantage of the modular architecture of the EDITS system (Edit Distance Textual Entailment Suite) version 3.0, an open-source software package for recognizing TE[11] [13]. 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 (i.e. the higher the distance, the lower is the probability of entailment).[12] Within this framework the system implements different approaches to distance computation, i.e. both edit distance algorithms (that calculate the T-H distance as the cost of the edit operations, i.e. insertion, deletion and substitution that are necessary to transform T into H), and similarity algorithms. Each algorithm returns a normalized distance score (a number between 0 and 1). At a training stage, distance scores calculated over annotated T-H pairs are used to estimate a threshold that best separates positive from negative examples. Such threshold is then used at a test stage to assign a judgment and a confidence score to each test pair.

4.3 Evaluation

To evaluate our combined approach, we carry out a two-step evaluation: first, we assess the performances of the TE system to correctly assign the entailment and the contradiction relations to the pairs of arguments in the Debatepedia data set. Then, we evaluate how much such performances impact on the application of the argumentation theory module, i.e. how much a wrong assignment of a relation to a pair of arguments is propagated in the argumentation framework.

[10] We consider only arguments implying another argument, leaving for future work arguments "supporting" another argument, but not inferring it.

[11] http://edits.fbk.eu/

[12] In previous RTE challenges, EDITS always ranked among the 5 best participating systems out of an average of 25 systems, and is one of the two RTE systems available as open source http://aclweb.org/aclwiki/index.php?title=Textual_Entailment_Resource_Pool

For the first evaluation, we run EDITS on the Debatepedia training set to learn the model, and we tested it on the test set. In the configuration of EDITS we used, the distance entailment engine combines cosine similarity as the core distance algorithm. Furthermore, distance is calculated on lemmas, and a stopword list is defined to have no distance value between stopwords. We use the system off-the-shelf, applying one of its basic configurations. As future work, we plan to fully exploit EDITS features, integrating background and linguistic knowledge in the form of entailment rules, and to calculate the distance between T and H basing on their syntactic structure.

Table 2 reports on the obtained results both using EDITS and using a baseline that applies a Word Overlap algorithm on tokenized text. Even using a basic configuration of EDITS, and a small data set (100 pairs for training) performances on Debatepedia test set are promising, and in line with performances of TE systems on RTE data sets (usually containing about 1000 pairs for training and 1000 for test). In order to understand if increasing the number of argument pairs in the training set could bring to an improvement in the system performances, the EDITS learning curve is visualized in Figure 3. Note that augmenting the number of training pairs actually improves EDITS accuracy on the test set, meaning that we should consider extending the Debatepedia data set for future work.

		Train			**Test**		
	rel	*Pr.*	*Rec.*	*Acc.*	*Pr.*	*Rec.*	*Acc.*
EDITS	*yes*	0.71	0.73	**0.69**	0.69	0.72	**0.67**
	no	0.66	0.64		0.64	0.6	
WordOverl.	*yes*	0.64	0.65	0.61	0.64	0.67	0.62
	no	0.56	0.55		0.58	0.55	

Table 2. Systems performances on the Debatepedia data set (precision, recall and accuracy)

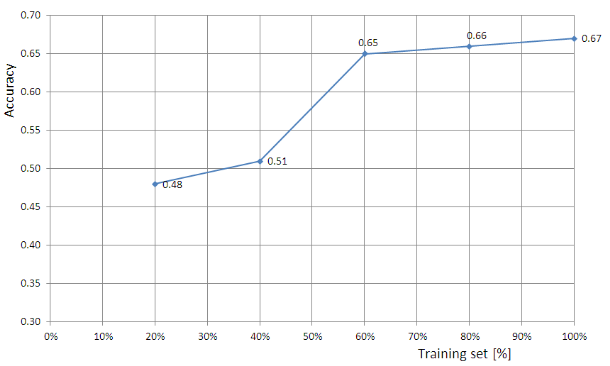

Figure 3. EDITS learning curve on Debatepedia data set

As a second step in our evaluation phase, we consider the impact of EDITS performances on the acceptability of the arguments, i.e. how much a wrong assignment of a relation to a pair of arguments affects the acceptability of the arguments in the argumentation framework. We use admissibility-based semantics to identify the accepted arguments both on the correct argumentation frameworks of each Debatepedia topic (where entailment/contradiction relations are correctly assigned, i.e. the goldstandard), and on the frameworks generated assigning the relations resulted from the TE system judgments. The precision of the combined approach we propose in the identification of the accepted arguments is on average 0.74 (i.e. arguments accepted by the combined system and by the goldstandard w.r.t. a certain Debatepedia topic), and the recall is 0.76 (i.e. arguments accepted in the goldstandard and retrieved as accepted by the combined system). Its accuracy (i.e. ability of the combined system to accept some arguments and discard some others) is 0.75, meaning that the TE system mistakes in relation assignment propagate in the argumentation framework, but results are still satisfying and foster further research in this direction.

5 Related work

Reed and Grasso [18] underline the need for a machinery which leads to arguments being automatically generated.

Debategraph[13] is an online system for debates supporting the incremental development of argument structures, but it is not grounded on argument theory to decide the accepted arguments.

Chasnevar and Maguitman [8] use defeasible argumentation to assist the language usage assessment. Their system provides recommendations on language patterns using indices (computed from Web corpora) and defeasible argumentation, where the preference criteria for language usage are formalized as defeasible and strict argumentation rules. The aim of the paper is different from ours. No NL techniques are used to automatically detect and generate the arguments.

Carenini and Moore [5] present a complete computational framework for generating evaluative arguments. The framework, based on the user's preferences, produces the arguments following the guidelines of argumentation theory to structure and select evaluative arguments. Then, a natural language processing step returns the argument in natural language. The output of the argumentation strategy is a text plan indicating the propositions to include in the argument and its overall structure. The aim of the paper is different from the aim of this paper: we do not use natural language generation to produce the arguments, but we use textual entailment to detect the arguments in natural language text. We use the word "generation" with the meaning of generation of the abstract arguments from the text, and not with the meaning of natural language generation. Concerning the argumentation part, we use the computational abstract model proposed by Dung to reason over the arguments to identify the accepted ones. We do not address argumentation-based persuasion or planning.

Wyner and van Engers [22] present a policy making support tool based on forums. They propose to couple natural language processing and argumentation to provide the set of well structured statements that underlie a policy. Apart from the different goal of this work, there are several points which distinguish our proposal from this one. First, their NLP module guides the participant in writing the input text using Attempt to Controlled English which allows the usage of a restricted grammar and vocabulary. After parsing the text, the sentences are translated to FOL. We do not have any kind of lexicon or grammar restriction, and we do not support the participant in writing the text, but we automatically extract the arguments from the debates. Second, the inserted statements are associated with a mode which indicates the relationship between the existing statements and the input statement. We do not ask the participants to explicit the relationship among the arguments, we infer them using TE. Moreover, no evaluation of their framework is provided.

Heras et al. [12] show how to model the opinions put forward on business oriented websites using argumentation schemes. The idea is to associate a scheme to each argument to have a formal structure

[13] http://debategraph.org

which makes the reasoning explicit. We share the same goal, that is providing a formal structure to on-line dialogues to evaluate them, but, differently from [12], in our proposal we achieve this issue using an automatic technique to generate the arguments from natural language texts as well as their relationships.

Rahwan et al. [17] present Avicenna, a Web-based system used to reason about arguments, ranging from automatic argument classication to reasoning about chained argument structures. In Avicenna, the arguments are inserted by participants through a form, and the participants can decide to attack or support existing arguments, while in our framework participants do not enter arguments, it returns completely automatically the abstract arguments, the relationships among them highlighting in the AF the accepted arguments.

Moens et al. [14] experiment ML approaches to recognize features characterizing legal arguments. We adopt a more general framework, i.e. TE (implementable also using ML techniques) to extract open-domain arguments, and automatically assign their relations.

6 Conclusions

The research presented in this paper is interdisciplinary. We have integrated in a combined framework a technique from computational linguistics and a technique for non-monotonic reasoning. The aim of this research is to provide the participants of online debates and forums with a framework supporting their interaction with the application. In particular, the proposed framework helps the participants to have an overview of the debates, understanding which are the accepted arguments at time being. The key contribution of our research is to allow the automatic detection and generation of the abstract arguments from natural language texts. We adopt a TE approach to inference because of the kind of (noisy) data present on the Web. TE is used to retrieve and identify the arguments, together with the relation relating them: the entailment relation (i.e. inference among two arguments), and the attack relation (i.e. contradiction among two arguments). The arguments and their relationships are then sent to the argumentation module which introduces the additional conflicts due to the entailment relation. The argumentation module returns the set of acceptable arguments w.r.t. the chosen semantics. We experimented the combined approach on a sample of topics extracted from Debatepedia. We created a data set of 200 pairs of arguments, and we tested an off-the-shelf open source TE system (i.e. EDITS) on it. The argumentation frameworks built using the relations assigned by the TE system have been evaluated to select the accepted arguments. The accuracy of the combined approach in identifying the arguments in a debate, and to correctly propose to the participant the accepted arguments is about 75%.

Several research lines have to be considered as future research. In particular, the combination of two different techniques will address many open issues in both the research fields of computational linguistics and non-monotonic reasoning. First, the use of natural language processing 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 argumentation theory. Second, given the promising results we obtained, we plan to extend the experimentation setting both increasing the pairs of arguments in the Debatepedia data set and to improve the TE system performances to apply our combined approach in other real applications. Third, we plan to integrate our combined framework together with a notification module able to send advises to the participants of online debates when a new argument is introduced in the debate against or in favour of her arguments.

REFERENCES

[1] R. Barzilay and K.R. McKeown, 'Sentence fusion for multidocument news summarization.', *Computational Linguistics*, **31(3)**, 297–327, (2005).

[2] P. Blackburn, J. Bos, M. Kohlhase, and H. de Nivelle., 'Inference and computational semantics.', *Studies in Linguistics and Philosophy, Computing Meaning*, **77(2)**, 1128, (2001).

[3] Guido Boella, Dov M. Gabbay, Leendert W. N. van der Torre, and Serena Villata, 'Support in abstract argumentation', in *Proc. of Computational Models of Argument (COMMA), Frontiers in Artificial Intelligence and Applications 216*, pp. 111–122, (2010).

[4] J. Bos and K. Markert, 'When logical inference helps determining textual entailment (and when it doesn't)', in *Proc. of the 2nd PASCAL Workshop on Recognizing Textual Entailment*, (2006).

[5] G. Carenini and J. D. Moore, 'Generating and evaluating evaluative arguments', *Artif. Intell.*, **170**(11), 925–952, (2006).

[6] C. Cayrol and M.C. Lagasquie-Schiex, 'Bipolarity in argumentation graphs: Towards a better understanding', in *Proc. of the 5th International Conference on Scalable Uncertainty Management (SUM), LNCS 6929*, pp. 137–148, (2011).

[7] Claudette Cayrol and Marie-Christine Lagasquie-Schiex, 'On the acceptability of arguments in bipolar argumentation frameworks', in *Proc. of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU), LNCS 3571*, pp. 378–389, (2005).

[8] C. I. Chesñevar and A.G. Maguitman, 'An argumentative approach to assessing natural language usage based on the web corpus', in *Proc. of the 16th European Conference on Artificial Intelligence (ECAI), IOS Press*, pp. 581–585, (2004).

[9] I. Dagan, B. Dolan, B. Magnini, and D. Roth, 'Recognizing textual entailment: Rational, evaluation and approaches', *Natural Language Engineering Journal*, **15**(04), i–xvii, (2009).

[10] 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).

[11] T. F. Gordon, H. Prakken, and D. Walton, 'The carneades model of argument and burden of proof', *Artif. Intell.*, **171**(10-15), 875–896, (2007).

[12] Stella Heras, Katie Atkinson, Vicente J. Botti, Floriana Grasso, Vicente Julián, and Peter McBurney, 'How argumentation can enhance dialogues in social networks', in *Proc. of Computational Models of Argument (COMMA), Frontiers in Artificial Intelligence and Applications 216*, pp. 267–274, (2010).

[13] M. Kouylekov and M. Negri, 'An open-source package for recognizing textual entailment', in *Proc. of the Association for Computational Linguistics, System Demonstrations (ACL)*, pp. 42–47, (2010).

[14] M.F. Moens, E. Boiy, R. Mochales Palau, and C. Reed, 'Automatic detection of arguments in legal texts', *ACM. Proc. Conference on Artificial Intelligence and Law*, 225–230, (2007).

[15] C. Monz and M. de Rijke, 'Light-weight entailment checking for computational semantics', in *Proc. Inference in Computational Semantics (ICoS-3)*, pp. 59–72, (2001).

[16] M. Pinkal, 'Logic and lexicon: the semantics of the indenite', *Studies in linguistics and philosophy*, **56**, (1995).

[17] Iyad Rahwan, Bita Banihashemi, Chris Reed, Douglas Walton, and Sherief Abdallah, 'Representing and classifying arguments on the semantic web', *Knowledge Eng. Review*, **26**(4), 487–511, (2011).

[18] C. Reed and F. Grasso, 'Recent advances in computational models of natural argument', *Int. J. Intell. Syst.*, **22**(1), 1–15, (2007).

[19] C. Reed and G. Rowe, 'Araucaria: Software for argument analysis, diagramming and representation', *Int. Journal on Artificial Intelligence Tools*, **13**(4), 961–980, (2004).

[20] L. Romano, M. O. Kouylekov, I. Szpektor, I. Dagan, and A. Lavelli, 'Investigating a generic paraphrase-based approach for relation extraction', in *Proc. of the 11st Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 409–416, (2006).

[21] B. Verheij, 'Argumed - a template-based argument mediation system for lawyers and legal knowledge based systems', in *Proc. of the 11th International Conference on Legal Knowledge and Information Systems (JURIX)*, pp. 113–130, (1998).

[22] A. Wyner and T. van Engers, 'A framework for enriched, controlled on-line discussion forums for e-government policy-making', in *Proc. of the Annual International e-government Conference (eGov)*, (2010).