Argument Mining: a Machine Learning Perspective

Marco Lippi[1] and Paolo Torroni[1]

DISI – Università degli Studi di Bologna
{marco.lippi3,p.torroni}@unibo.it

Abstract. Argument mining has recently become a hot topic, attracting the interests of several and diverse research communities, ranging from artificial intelligence, to computational linguistics, natural language processing, social and philosophical sciences. In this paper, we attempt to describe the problems and challenges of argument mining from a machine learning angle. In particular, we advocate that machine learning techniques so far have been under-exploited, and that a more proper standardization of the problem, also with regards to the underlying argument model, could provide a crucial element to develop better systems.

1 Introduction

Argumentation is a multi-disciplinary research field which studies debate and reasoning processes, and spans across and ties together diverse areas such as logic and philosophy, language, rhetoric and law, psychology and computer science. Over the last decades, *computational* argumentation has come to be increasingly central as a core study within artificial intelligence [3], while some cognitive science theories indicate that the function of human reasoning itself is argumentative [27]. Argumentation started to become known even in the computational social sciences, where agent-based simulation models have been proposed, whose micro-foundation explicitly refers to argumentation theories [26, 15]. This, together with the current hype of big data and tremendous advances in computational linguistics, created fertile ground for the rise of a new area of research called *argumentation* (or *argument*) *mining* (henceforth AM).

The growing excitement in this area is tangible. The initial studies started to appear only a few years ago in specific domains such as legal texts, online reviews and debate [28, 39, 7]. In 2014 alone there have been no less than three international events on argumentation mining.[1] While research on this topic is

[1] The First ACL Workshop on Argumentation Mining, http://www.uncg.edu/cmp/ArgMining2014, SICSA Workshop on Argument Mining: Perspectives from Information Extraction, Information Retrieval and Computational Linguistics http://www.arg-tech.org/index.php/sicsa-workshop-on-argument-mining-2014, and the BiCi Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing, http://www-sop.inria.fr/members/Serena.Villata/BiCi2014/frontiersARG-NLP.html

gaining momentum, major commercial players have also joined in, as IBM recently funded a multi-million cognitive computing project whose core technology is AM.[2] But what is AM and what makes it so popular?

The main goal of AM is to automatically extract arguments from generic textual corpora, in order to provide structured data for computational models of arguments and reasoning engines.

The self-evident application potential of AM is one reason for its growing popularity. From an application perspective, AM could be considered in some respects as an evolution of sentiment analysis. Habernal et al. [20] state that, while the goal of opinion mining is to understand *what people think about something*, the aim of argumentation mining is to understand *why*, thus unveiling reasoning processes, rather than just detecting opinions and sentiment. Besides, more or less abstract computational argumentation models and theories now seem closer than ever to the "real world" and the community seems eager to contribute to the creation of significant domains where very expressive models and efficient algorithms developed in recent years can be tested and applied. Another reason of its rapid expansion is that AM poses a scientifically engaging challenge, especially from a machine learning (ML) perspective. Indeed, AM is a difficult NLP task that merges together many different components, such as information extraction, knowledge representation, and discourse analysis. This is also creating new opportunities in the computational argumentation community. Advanced statistical and subsymbolic reasoning methods have never been so tightly conjugated with a discipline, whose roots are in symbolic artificial intelligence.

Most notably, we see AM as a source of new opportunities for the formal argumentation community, drawing a bridge between formal models and theories and argumentative reasoning as it emerges from everyday life.

Due to the novelty of this research domain, at the present stage AM is not a well-defined problem with clear boundaries. On the contrary, AM is rather a broad umbrella for a new set of challenges where many different understandings coexist and contribute towards a common, if under-specified, objective. However, there are already many interesting results, and we feel that time is ripe for attempting an initial road map.

The aim of this article is thus to discuss achievements and challenges in AM from a ML angle. Our ambition is to help making this new domain accessible to scholars that do not necessarily have a computational argumentation background. For this reason, we will start by introducing models which, although well-known in computational argumentation, are crucial design choices that greatly influence the ML problem formulation. We will then proceed to review relevant ML techniques and discuss challenges that AM poses to ML research.

[2] More about IBM Debating Technologies at `http://researcher.watson.ibm.com/researcher/view_group.php?id=5443`

2 Problem formulation

The discipline of argumentation has ancient roots in dialectics and philosophy, as that branch of knowledge dedicated to the study and analysis of how statements and assertions are proposed and debated, and conflicts between diverging opinions are resolved [3]. Starting from the pioneering works by Pollock [33], Simari and Loui [40], and Dung [12], among others, models of argumentation have also spread in the area of AI, especially in connection with knowledge representation, non-monotonic reasoning, and multi-agent systems research, giving rise to a new field named "computational argumentation."

The two main approaches in computational argumentation are called *abstract* argumentation, and *structured* argumentation. The former is rooted in Dung's work, and it considers each argument as an atomic entity without internal structure. It thus provides a very powerful framework to model and analyze "attack" relations between arguments, or sets of them, which may or may not be *justified* according to some semantics. The latter proposes an internal structure for each argument, described in terms of some knowledge representation formalism. Structured argumentation models are those typically employed in AM, as defining the structure of an argument becomes crucial, when the goal is to extract portions of arguments from natural language.

Because there are many significant proposals for structured argumentation [4], it is impossible to give a single formal, universally accepted definition of structured argument. A rather comprehensive account of argumentation models under an argument analysis perspective is given by Peldszus and Stede [32]. A simple and intuitive characterization is given by Walton, who describes an argument as a set of statements consisting in three parts: a conclusion, a set of premises, and an inference from the premises to the conclusion [45]. In the literature, conclusions are sometimes referred to as *claims*, premises are often called *evidence* or *reasons*, and the link between the two, i.e., the inference, is sometimes called the *argument* itself. Besides this basic *claim/premise* argument model, other noteworthy models are due to Tuolmin [44] and Freeman [14].

Here is an example of a sentence containing a claim:[3]

> *Health risks can be produced by long-term use or excessive doses of anabolic steroids.*

The following sentence instead contains a premise, supporting the previous claim:

> *A recent study has also shown that long term AAS users were more likely to have symptoms of muscle dysmorphia.*

The term *argumentation* has historically referred to the process of constructing arguments and, since the advent of computational argumentation, to the process of determining the set of justified conclusions of a set of arguments. However, *argumentation mining* and *argument mining* are often used interchangeably and in a broad sense, as the field yet retains a strong element of conceptual exploration.

[3] All examples in this paper are taken from the IBM corpus, described in Section 4.

The task of detecting the premises and conclusion of an argument, as found in a text of discourse, is typically referred to as *detection* or *identification* [45]. More specific sub-tasks are *claim detection* [24] and *evidence detection* [36].

Being this a young research domain, not only its definitions but also its approaches and targets vary widely. Some research aims at extracting the arguments from generic unstructured documents, which is a fundamental step in practical applications [24], whereas other starts from a given set of arguments and focuses on aspects such as the identification of attack/support [10] or entailment [8] relations between them, or on the classification of argument schemes [13] in the sense of Walton et al. [46].

In the next section we will review ML methods for the task of automatically extracting arguments from text.

3 Methods

Argument mining is a complex, multi-faceted problem, which embraces many different concepts from various disciplines. For this reason, addressing AM usually requires dealing with a variety of intertwined sub-tasks. This intrinsic heterogeneity makes AM an extremely engaging application for machine learning, by involving aspects of natural language processing and understanding, information extraction, feature discovery and discourse analysis. All the argument mining frameworks proposed so far can be described as multi-stage pipeline systems, whose input is natural, free text document, and whose output is a mark-up document, where arguments (or parts of arguments) are annotated. Each stage addresses a sub-task of the whole AM problem, by employing one or more machine learning and natural language processing methodologies and techniques.

3.1 Argumentative sentence detection

A first stage usually consists of detecting which sentences in the input document are argumentative, which means that they contain an argument, or part thereof. This task is typically implemented by a machine learning classifier. A common implementation consists of training a binary classifier, with the goal of simply discarding propositions that are not argumentative, while a second classifier at a later stage in the pipeline will subsequently be trained to distinguish among various argument components (e.g., claims from premises). Alternatively, a single multi-class predictor could be employed to discriminate between all the possible categories of argument components.

In both cases, two crucial issues within this step involve (1) the choice of the classifier, and (2) the features to be used to describe the sentences. As for the adopted machine learning classifiers, many works in the literature so far have made attempts to compare several approaches, including Naïve Bayes classifiers [28, 30], Support Vector Machines (SVM) [28, 30], Maximum Entropy classifiers [28], Logistic Regression [24], Decision Trees and Random Forests [42]. The obtained results are in some cases conflicting, as for example in [28] the SVM

model performs worse than Naïve Bayes, while in [42] the opposite happens. As a matter of fact, the vast majority of the aforementioned approaches have been based on classic, off-the-shelf classifiers, while all the effort has been focused on the creation of a set of highly engineered features, sometimes also obtained as the outcome of other external predictors [24]. It is therefore not surprising that the key element for achieving good performance has been shown to be the choice of the features, rather than the machine learning algorithm. Indeed, in several cases, different classifiers trained with the same feature sets lead to very similar performance.

Many works employ classical features for text representation, including bag-of-words representations of sentences, word bigrams and trigrams, part-of-speech information obtained with some statistical parser, information on punctuation, verb tenses and the use of some pre-determined list of key phrases [28, 42]. An example fed to the machine learning classifier is therefore a sentence, typically represented as a vector x of k features $x = \{x_1, \ldots, x_k\}$, where x_j indicates the value of the j-th feature. In the formalism of bag-of-words, also extended to bigrams and trigrams, the j-th feature can indicate, for example, the presence, within the sentence, of the j-th word (or bigram, or trigram) of the dictionary. Yet, this classic and still widely used approach has a limitation: it does not capture the semantic similarity between different words, but only counts common terms in order to measure the similarity between two sentences. In this sense, for example, two terms such as *argue* and *believe* are orthogonal, and therefore they are as different as *argue* and *eat*. More advanced features try to incorporate linguistic and semantic information on the most informative words (typically verbs and nouns) in order to capture such similarities, by employing onthologies such as WordNet [24]. Some additional features are also used to mark the presence of certain syntactical descriptors, with the aim to detect recurrent structural patterns, but these methods are prone to overfitting, as they are typically well-suited for the corpus they have been constructed on. Even more sophisticated features include sentiment analysis indicators, subjectivity scores of sentences, dictionaries of keywords or keyphrases that may be highly informative of the presence of an argument [24]. Also in this case, the risk of obtaining methods that cannot generalize to different corpora is certainly not negligible, and, as a matter of fact, we are not aware of any method so far that has been extensively tested on a variety of different corpora.

Another key problem within this context is whether it is convenient to build systems that need to employ contextual information to detect argumentative sentences. The approach developed at IBM Research in Haifa, as a part of The Debater™ project, makes a strong use of the topic information (given in advance) when attempting to extract arguments [24]. Also in other specific applicative scenarios, as in the case of legal documents [28], features are very often highly dependent on the domain. While the use of contextual information can no doubt significantly boost the performance of the system in a given context, we remark that this is another element that could greatly limit the general applicability of the system across different contexts.

In a recent work [25] we propose to overcome these issues by employing an SVM based on structured kernels built upon constituency parse trees to identify sentences containing claims. Basically, the similarity between the structure of the parse trees is used in order to measure the similarity between sentences. In this way, the rhetorical structure of sentences is automatically captured by the implicit feature space, without the need of manually specifying the feature set, and without resorting to explicit contextual information.

Previous work by Rooney et al. [37] also considers kernel methods for an AM task. However, it only uses the sequence of parts-of-speech tags without exploiting the powerful representation of parse trees. The authors use their own tagging of the AraucariaDB (see Section 4).

3.2 Argument boundaries detection

Once the non-argumentative sentences have been discarded by the first stage of the pipeline, it is necessary to determine the exact boundaries of argument components, or "argumentative discourse units" [32]. Clearly, this phase greatly depends on the underlying argument model, since the AM system must be able to discriminate between all the possible components defined by the model of choice: claims and premises, for example, if we adopt the claim/premise model, or warrants, backings, qualifiers and rebuttals if we adopt the Toulmin model. Existing AM systems usually adopt the claim/premise model, because of its simplicity and generality. Yet, a recent work by Harbenal et al. [20] argues that different argumentation models could be better suitable for different application domains. For this reason, they employ the Toulmin model for the annotation of a corpus of web documents collected from blogs, forums, and news.

Regardless of the considered argument model, in addition to discriminating between all the possible argument components, this stage of the AM pipeline also needs to address a so-called *segmentation problem*, since a whole sentence does not necessarily correspond exactly to an argument component. In fact, we can distinguish between three different cases:

1. only a portion of the sentence coincides with an argument component;
2. two or more argument components can be present within the same sentence;
3. an argument component can span across multiple sentences.

For example, in the case of claims, the following sentence falls into the first category:

> *A significant number of republicans assert that hereditary monarchy is unfair and elitist.*

where the annotated claim is underlined. An example of a premise spanning more than a single sentence is the one below:

> *When New Hampshire authorized a state lottery in 1963, it represented a major shift in social policy. No state governments had previously directly*

run gambling operations to raise money. Other states followed suit, and now the majority of the states run some type of lottery to raise funds for state operations.

Most of the existing methods assume only one of the above possibilities, and they address the segmentation problem as a separate stage from the extraction of argumentative sentences [28, 24].

However, different solutions could in principle be exploited, for example resorting to structured output classifiers or to statistical relational learning models, which are capable of performing *collective classification* on a set of examples, rather than considering each of them independently. This framework allows to consider relationships and dependencies between examples and has shown to be a crucial element in many machine learning tasks on structured data [16]. A first step in this direction is observed in [18] and [31], where conditional random fields are used to perform the segmentation task for argument components.

Multi-class classification systems similar to the ones described in the previous section are typically employed to discriminate amongst different argument components, but sometimes they do not properly address the segmentation task [42]. In other cases, clauses (sub-sentences) resulting from the parsing of a sentence are considered as boundaries [28], or maximum likelihood systems are employed to identify the most probable boundaries of the argument components [24].

3.3 Argument structure prediction

Following the detection of argument components, a further stage in the pipeline has the aim to predict links between arguments or their components. As customary in machine learning, we speak in this case of *prediction* rather than *detection*, because the target of the classification is not a specific portion of the input document, but rather a connection (or link) between them. If the desired output consists in only relations between argument components, then the system will produce a sort of map of the arguments retrieved from the input text. Another possibility is to infer the connections between arguments, in which case support and attack relations have to be distinguished. This second point is a very important step, as the output of the AM system could be used as an input to a formal argumentation framework, so that different semantics could be applied to identify sets of arguments with desired characteristics.

As in the previous steps of the AM pipeline, even for structure prediction the implementation choices strongly depend on the underlying argument model. When considering a claim/premise model, for example, the task of inferring connections between claims and premises can be seen as a link prediction problem within a bipartite graph. With a more complex model, such as the Toulmin model, the link categories that can be predicted clearly grow, and more fine-grained predictors have to been designed, in order to correctly predict the connections between all the components. It is also worth noticing that some argument components can also be *implicit* within the original textual document: this is the case, for example, of *enthymemes*, or even of implicit warrants in the

Toulmin model, corresponding to unsaid assumptions. Therefore, the argument structure prediction phase should, in principle, be able also to detect such implicit components and add them to the model: from a machine learning point of view, this is a highly challenging task, and currently no attempt has been made in this direction. A possible reference model for constructing enthymemes was proposed in [5].

In some cases, further simplifications can be modeled: in the work developed at the IBM Haifa Research Group, for example, premises (which they call *evidence*) are labeled given a certain claim [1]. In this way, the information regarding the claim can be used when detecting the evidence, and therefore there is no need to further predict the structure links, which are obtained (by definition) when predicting the evidence. In [42], a claim/premise model based on work by Freeman [14] is adopted, and thus attack/support links between argument components are predicted using a plain SVM binary classifier. In the context of legal documents, [28] adopt a manually-constructed context-free grammar to predict relations between argument components: this is a strongly domain-specific approach, based on the common structures of legal texts, which could hardly be applied to different application scenarios. Another quite popular approach is based on Textual Entailment (TE) [6] and aims to understand whether there exists an entailment relation between two given argument components.

4 Corpora

It is a fact that the whole AM process, based on ML and AI techniques, requires a collection of annotated documents, to be used as a training set for any kind of predictor. Constructing annotated corpora is, in general, a complex and time-consuming task, which requires to commit costly resources such as teams of experts, so that homogeneous and consistent annotations can be obtained. This is particularly true for AM, as the identification of argument components, their exact boundaries, and how they relate to each other can be quite complicated (and controversial!) even for humans. Moreover, very often the existing data sets have been built with slightly different goals or for some specific aim, and therefore they cannot always be used within all machine learning approaches.

As an example, several annotated corpora have been constructed with the goal of analyzing arguments and their relations. Among them, we mention the collections maintained by the University of Dundee,[4] which aggregate many datasets—including, notably, AraucariaDB—with annotated argument maps, in a variety of standardized formats. Another collection is the NoDE benchmark data base [9] which contains arguments obtained from a variety of sources, including Debatepedia[5] and ProCon[6]. Yet, due to the goal they were built for, these corpora do not put an emphasis on—and they often lack—the non-argumentative

[4] `http://corpora.aifdb.org`
[5] `http://www.debatepedia.com`
[6] `http://www.procon.org`

parts, which are necessary as negative examples for the training of some kind of discriminative machine learning classifier.

Furthermore, most of the AM systems proposed so far have been mainly used in pilot applications in specific domains only, where a few annotated corpora exist. Law has been the pioneering application domain for AM, and certainly among the most successful ones, with the work by Mochales Palau and Moens [28] on the European Court of Human Rights (ECHR) dataset for the extraction of claims and their supporting premises from a collection of structured legal documents. More recently, also the Vaccine/Injury Project (V/IP) [2] was carried out, with the goal of extracting arguments from a set of juridical cases involving vaccine regulations. Unfortunately, these corpora are not publicly available.

A new trend which is recently gaining attention is that of creating annotated data sets from biology and medicine texts [19, 21]. This could be an extremely important step towards building ontologies and knowledge bases describing the links between either symptoms and diseases, or between genes and diseases, or even to assist personalized medicine prescriptions.

Rhetorical, philosophical and persuasive essays represent another interesting case study. The creation of a corpus from a collection of 19th century philosophical essays was proposed in [22]. A limited-scope but well-documented data set was proposed by Stab and Gurevych [41] as a collection of 90 persuasive essays. The topics covered are very heterogeneous. Due to the nature of the data, and to the annotation guidelines, only a few sentences in this corpus are non-argumentative. Being specifically designed for the analysis of persuasive essays, this corpus would likely not be the most appropriate choice for a training set, if the goal were to generalize to other kinds of data sources. In fact, these essays are annotated with claims, premises, and "major claims" (one per essay), these being highly domain-specific tags, often detected thanks to dedicated features, such as the position of the sentence within the essay.

A much larger data set is currently being developed at IBM Research,[7] starting from plain text in Wikipedia pages [1, 36]. The purpose of this corpus is to collect context-dependent claims and evidence facts (i.e., premises), which are relevant to a given topic. At the time of writing, the data set covers 58 topics, for a total of 547 Wikipedia articles. The data set contains about 7,000 argumentative entities (claims or evidence), and is an extremely challenging benchmark. An approach to context-dependent claim detection on this corpus was proposed in [24], while a context-independent approach was applied in [25] for the same dataset.

Additional datasets were recently collected from online resources, including online reviews, blogs, and newspapers. Two of them have been developed by [38], for the task of extracting so-called *opinionated claims*: they consist in 285 LiveJournal blogposts and 51 Wikipedia discussion forums, respectively. Each dataset consists of 2,000 sentences. Another well-annotated corpus was developed by Habernal et al. [20], to model arguments following a variant of the Toulmin model. This dataset includes 990 instances, 524 of which are labeled as

[7] `https://www.research.ibm.com/haifa/dept/vst/mlta_data.shtml`

argumentative. A smaller corpus of 345 examples is annotated with finer tags. The authors report the annotation procedure in detail, together with a review of the inter-agreement evaluation procedures of other existing corpora. Finally, data collected by web sources have been used also in [18], but unfortunately they are not publicly available.

5 Challenges

From the point of view ML, this blossoming research field poses new challenges and paves the way to unprecedented opportunities. We discuss them here.

Owing to the only recent development of the area, there is still a lack of general agreement regarding the models which should be adopted to build an AM system. Although one could argue that the intrinsic heterogeneous nature of data sources and application domains makes it difficult to propose a single and general model to be adopted in many contexts, yet we believe that some clarifications should be made in order to pose guidelines for the constructions of corpora. An attempt in this direction has certainly been made by several authors (e.g., see [24, 20]). This process would bring a twofold benefit also on the ML side. First of all, it would allow more appropriate comparisons between different algorithms and techniques, as the same performance measurements could be applied to compare different approaches. Secondly, such a framework would also help the development of more general and context-independent methodologies, capable of performing AM on different kinds of data sources, since a novel system could be applied across different domains, exploiting what in ML is typically referred to as *transfer learning* [29].

From a more technical point of view, it is clear that, up to now, ML methodologies so far have been applied in AM pipelines only as off-the-shelf black boxes, while very often devolving the performance of the whole systems upon the sophistication of features employed. We believe that the time is ripe to move the ML contributions to AM a step forward, by trying more advanced algorithms, or even by developing specific approaches. Within this context, a crucial contribution will likely come from statistical relational learning, a recent area of ML dedicated to handling relational and structured data. The idea driving this research field is that relations between patterns often represent crucial information to build classifiers with high performance. When data is represented in a structured form, as it happens with the sequentiality of text, or with the graphical structure of argument maps, the potential of this kind of methodology is evident. Many of the approaches developed within this field also exploit logic formalisms to describe the domain of interest, thus allowing the embedding of background knowledge in the form of predicates and logic clauses. The success of statistical relational learning in relevant tasks somehow related to AM, such as link discovery in social and biological networks [17], information extraction and entity resolution in textual corpora [11, 34], sequence tagging and sentence parsing [35] offers an additional very strong motivation. Another area of machine learning which may contribute to AM is active learning, where the learning sys-

tems actively ask for supervisions rather than being given in advance a fixed, static batch of supervised data. Active learning approaches have scored interesting results in several natural language processing applications [43] and thus they could be successfully applied also to some steps in the AM pipeline, being particularly useful when annotated data are hard to collect.

Last but not least, the AM community should certainly not ignore the huge impact that deep learning is bringing within artificial intelligence. Models based on deep architectures have obtained breakthrough results on a wide variety of applications, ranging from speech recognition and computer vision to natural language processing and understanding (e.g., see [23] and references therein). By dominating the ML scene in the last years, deep learning approaches are with no doubt among the novel methodologies which could bring decisive contributions to AM systems.

6 Conclusions

Argumentation mining represents a novel, exciting application domain for machine learning. Nevertheless, in spite of some promising initial results, there is still a lot of work to be done, in order to exploit all the potential of ML approaches within the AM community, and to build successful applications to be employed as an input to formal argumentation frameworks.

While other surveys have been dedicated to the modeling aspect of the AM tasks [32], this is the first step towards a more principled formulation of the problem from the ML point of view. In particular, this paper is a first attempt to highlight challenges and opportunities for ML systems in this area.

We argue that current approaches too often rely on methodologies that demand a great deal of effort in the development of powerful but highly domain-dependent features, and are thus difficult to generalize.

Moreover, we believe that a major obstacle to progress in AM is the lack of a standardized methodology for annotating relevant corpora. We find that most works define their own labeled corpora, hindering comparison between various approaches on the same dataset and between the performance of approaches across datasets.

We thus argue that a major effort should be put into the construction of annotated corpora that meet the needs of ML algorithms. In particular, if (as we believe) identifying relations between different arguments and between different argument components is a valuable output of prospective AM applications, then corpora should contain all the necessary annotations. As a matter of fact, argument structure prediction is the stage in the AM pipeline that has produced less results so far.

Finally, the methods we reviewed mostly target homogeneous and domain-specific data sources. An interesting direction could be developing AM techniques capable of handling heterogeneous data sources, as well as relational and structured data.

References

1. Aharoni, E., Polnarov, A., Lavee, T., Hershcovich, D., Levy, R., Rinott, R., Gutfreund, D., Slonim, N.: A benchmark dataset for automatic detection of claims and evidence in the context of controversial topics. In: Proceedings of the First Workshop on Argumentation Mining. pp. 64–68. Association for Computational Linguistics (2014)
2. Ashley, K.D., Walker, V.R.: Toward constructing evidence-based legal arguments using legal decision documents and machine learning. In: Francesconi, E., Verheij, B. (eds.) ICAIL 2013, Rome, Italy. pp. 176–180. ACM (2013)
3. Bench-Capon, T.J.M., Dunne, P.E.: Argumentation in artificial intelligence. Artificial Intelligence 171(10-15), 619–641 (2007)
4. Besnard, P., García, A.J., Hunter, A., Modgil, S., Prakken, H., Simari, G.R., Toni, F.: Introduction to structured argumentation. Argument & Computation 5(1), 1–4 (2014)
5. Black, E., Hunter, A.: A relevance-theoretic framework for constructing and deconstructing enthymemes. J. Log. Comput. 22(1), 55–78 (2012)
6. Cabrio, E., Villata, S.: Combining textual entailment and argumentation theory for supporting online debates interactions. In: Proceedings of the 50th annual meeting of the Association for Computational Linguistics (ACL 2012). pp. 208–212. Association for Computational Linguistics, Jeju, Korea (2012)
7. Cabrio, E., Villata, S.: Natural language arguments: A combined approach. In: Raedt, L.D., Bessière, C., Dubois, D., Doherty, P., Frasconi, P., Heintz, F., Lucas, P.J.F. (eds.) ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track, Montpellier, France, August 27-31 , 2012. vol. 242, pp. 205–210. IOS Press (2012)
8. Cabrio, E., Villata, S.: A natural language bipolar argumentation approach to support users in online debate interactions. Argument & Computation 4(3), 209–230 (2013)
9. Cabrio, E., Villata, S.: NoDE: A benchmark of natural language arguments. In: Parsons, S., Oren, N., Reed, C., Cerutti, F. (eds.) COMMA 2014. Frontiers in Artificial Intelligence and Applications, vol. 266, pp. 449–450. IOS Press (2014)
10. Chesñevar, C.I., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G.R., South, M., Vreeswijk, G., Willmott, S.: Towards an argument interchange format. The Knowledge Engineering Review 21(4), 293–316 (2006)
11. Culotta, A., McCallum, A., Betz, J.: Integrating probabilistic extraction models and data mining to discover relations and patterns in text. In: Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. pp. 296–303. Association for Computational Linguistics (2006)
12. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence 77(2), 321–358 (1995)
13. Feng, V.W., Hirst, G.: Classifying arguments by scheme. In: Lin, D., Matsumoto, Y., Mihalcea, R. (eds.) The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA. pp. 987–996. ACL (2011)
14. Freeman, J.B.: Dialectics and the macrostructure of arguments: A theory of argument structure, vol. 10. Walter de Gruyter (1991)

15. Gabbriellini, S., Torroni, P.: A new framework for ABMs based on argumentative reasoning. In: Kaminski, B., Koloch, G. (eds.) Advances in Social Simulation - Proceedings of the 9th Conference of the European Social Simulation Association, ESSA 2013, Warsaw, Poland, September 16-20, 2013. Advances in Intelligent Systems and Computing, vol. 229, pp. 25–36. Springer (2014)
16. Getoor, L.: Tutorial on statistical relational learning. In: Kramer, S., Pfahringer, B. (eds.) ILP. Lecture Notes in Computer Science, vol. 3625, p. 415. Springer (2005)
17. Getoor, L., Diehl, C.P.: Link mining: a survey. ACM SIGKDD Explorations Newsletter 7(2), 3–12 (2005)
18. Goudas, T., Louizos, C., Petasis, G., Karkaletsis, V.: Argument extraction from news, blogs, and social media. In: Likas, A., Blekas, K., Kalles, D. (eds.) Artificial Intelligence: Methods and Applications, LNCS, vol. 8445, pp. 287–299. Springer International Publishing (2014)
19. Green, N.: Towards creation of a corpus for argumentation mining the biomedical genetics research literature. In: Proceedings of the First Workshop on Argumentation Mining. pp. 11–18. Association for Computational Linguistics (2014)
20. Habernal, I., Eckle-Kohler, J., Gurevych, I.: Argumentation mining on the web from information seeking perspective. In: Cabrio, E., Villata, S., Wyner, A. (eds.) Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing. Forlì-Cesena, Italy, July 21-25, 2014. CEUR Workshop Proceedings, vol. 1341. CEUR-WS.org (2014)
21. Houngbo, H., Mercer, R.: An automated method to build a corpus of rhetorically-classified sentences in biomedical texts. In: Proceedings of the First Workshop on Argumentation Mining. pp. 19–23. Association for Computational Linguistics (2014)
22. Lawrence, J., Reed, C., Allen, C., McAlister, S., Ravenscroft, A.: Mining arguments from 19th century philosophical texts using topic based modelling. In: Proceedings of the First Workshop on Argumentation Mining. pp. 79–87. Association for Computational Linguistics (2014)
23. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 531, 436–444 (2015)
24. Levy, R., Bilu, Y., Hershcovich, D., Aharoni, E., Slonim, N.: Context dependent claim detection. In: Hajic, J., Tsujii, J. (eds.) COLING 2014, Dublin, Ireland. pp. 1489–1500. ACL (2014)
25. Lippi, M., Torroni, P.: Context-independent claim detection for argument mining. In: Yang, Q., Wooldridge, M. (eds.) Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015. pp. 185–191. AAAI Press (2015)
26. Mäs, M., Flache, A.: Differentiation without distancing. explaining bi-polarization of opinions without negative influence. PLoS ONE 8(11), e74516 (11 2013)
27. Mercier, H., Sperber, D.: Why do humans reason? arguments for an argumentative theory. Behavioral and Brain Sciences 34, 57–74 (4 2011)
28. Mochales Palau, R., Moens, M.F.: Argumentation mining. Artificial Intelligence and Law 19(1), 1–22 (2011)
29. Pan, S.J., Yang, Q.: A survey on transfer learning. Knowledge and Data Engineering, IEEE Transactions on 22(10), 1345–1359 (2010)
30. Park, J., Cardie, C.: Identifying appropriate support for propositions in online user comments. In: Proceedings of the First Workshop on Argumentation Mining. pp. 29–38. Association for Computational Linguistics, Baltimore, Maryland (June 2014)

31. Park, J., Katiyar, A., Yang, B.: Conditional random fields for identifying appropriate types of support for propositions in online user comments. In: Proceedings of the Second Workshop on Argumentation Mining. Association for Computational Linguistics (2015)
32. Peldszus, A., Stede, M.: From argument diagrams to argumentation mining in texts: A survey. International Journal of Cognitive Informatics and Natural Intelligence (IJCINI) 7(1), 1–31 (2013)
33. Pollock, J.L.: Defeasible reasoning. Cognitive Science 11(4), 481–518 (1987)
34. Poon, H., Domingos, P.: Joint inference in information extraction. In: Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, 2007, Vancouver, Canada. pp. 913–918. AAAI Press (2007)
35. Poon, H., Domingos, P.: Unsupervised semantic parsing. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1. pp. 1–10. EMNLP '09, Association for Computational Linguistics, Stroudsburg, PA, USA (2009)
36. Rinott, R., Khapra, M., Alzate, C., Dankin, L., Aharoni, E., Slonim, N.: Show me your evidence – an automatic method for context dependent evidence detection. In: Proceedings of the 2015 Conference on Empirical Methods in NLP (EMNLP), Lisbon, Portugal, 17-21 September 2015. pp. 440–450. Association for Computational Linguistics (2015)
37. Rooney, N., Wang, H., Browne, F.: Applying kernel methods to argumentation mining. In: Youngblood, G.M., McCarthy, P.M. (eds.) Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference, Marco Island, Florida. May 23-25, 2012. AAAI Press (2012)
38. Rosenthal, S., McKeown, K.: Detecting opinionated claims in online discussions. In: Sixth IEEE International Conference on Semantic Computing, ICSC 2012, Palermo, Italy, September 19-21, 2012. pp. 30–37. IEEE Computer Society (2012)
39. Saint-Dizier, P.: Processing natural language arguments with the <TextCoop> platform. Argument & Computation 3(1), 49–82 (2012)
40. Simari, G.R., Loui, R.P.: A mathematical treatment of defeasible reasoning and its implementation. Artificial Intelligence 53(23), 125 – 157 (1992)
41. Stab, C., Gurevych, I.: Annotating argument components and relations in persuasive essays. In: Hajic, J., Tsujii, J. (eds.) COLING 2014, Dublin, Ireland. pp. 1501–1510. ACL (2014)
42. Stab, C., Gurevych, I.: Identifying argumentative discourse structures in persuasive essays. In: Moschitti, A., Pang, B., Daelemans, W. (eds.) EMNLP 2014, Doha, Qatar. pp. 46–56. ACL (2014)
43. Thompson, C.A., Califf, M.E., Mooney, R.J.: Active learning for natural language parsing and information extraction. In: Bratko, I., Dzeroski, S. (eds.) Proceedings of the Sixteenth International Conference on Machine Learning (ICML 1999), Bled, Slovenia, June 27 - 30, 1999. pp. 406–414. Morgan Kaufmann (1999)
44. Toulmin, S.E.: The Uses of Argument. Cambridge University Press (1958)
45. Walton, D.: Argumentation theory: A very short introduction. In: Simari, G., Rahwan, I. (eds.) Argumentation in Artificial Intelligence, pp. 1–22. Springer US (2009)
46. Walton, D., Reed, C., Macagno, F.: Argumentation Schemes. Cambridge University Press (2008)