

Argumentation Mining: Techniques and Emerging Trends

Shobhit Sinha

Research Scholar, Department of Computer Applications, Shri Ramswaroop Memorial University, Lucknow – Deva Road 225003, Uttar Pradesh, India
Email: sinhashobhit@icloud.com

Bineet Kumar Gupta

Associate Professor, Department of Computer Applications, Shri Ramswaroop Memorial University, Lucknow – Deva Road 225003, Uttar Pradesh, India
Email: hod.ca@srmu.ac.in

Rajat Sharma*

Research Scholar, Department of Computer Applications, Shri Ramswaroop Memorial University, Lucknow – Deva Road 225003, Uttar Pradesh, India
Email: rajatcivilsharma@gmail.com

Abstract. By Argument we mean persuasion of a reason or reasons in support of a claim or evidence. In Artificial Intelligence computational argumentation is the field dealing with computational logic upon which many models of argumentation have been suggested. The goal of Argumentation Mining is to automatically extract structured arguments from the unstructured text. It has the potential of extracting information from web and social media, making it one of the most sought after research area. Some recent advances in computational logic and Machine Learning methods do provide a new insight to the applications for policy making, economic sciences, legal, medical and information technology. Different models have been proposed for argumentation mining with different machine learning methods applied on the argumentation frameworks proposed for this particular mining task. In this survey article we will review the existing systems and applications and will cover the three categories of argumentation models and a comparative table depicting the most frequently applied ML method. This survey paper will also cover the various challenges of the field with the new potential perspectives in this new emerging research area.

Index Terms:: Artificial Intelligence, Argumentation Mining, Computational logic, Machine Learning

1. Introduction

By "Argument," we mean persuasion of a reason or reasons in support of some action or claim to the audience. It refers to the set of sentences or phrases that act as a "premises" and work in conjunction with another set of sentences or phrases that act as "claims" or "conclusions," and when these are combined, they form an argument.

Discourse markers reveal Arguments which are present in the natural language, but sometimes such markers are absent or they give ambiguous meaning. Moreover, the premises in favor of the claim are far away to make a claim in which scenario it becomes difficult to make a conclusion by linking them together. Argumentation structures also take the form of graphs many a times where nested tree structures are present. In such scenarios, we humans recognize such argumentative discourse by our understanding of the world or, in more simple terms, domain knowledge or common sense which a machine does not possess. For example, "Technology influences negatively on how people communicate. Some people use their mobile phones constantly ignoring their environment"

Here, the second sentence acts as a premise for the claim made in the first sentence, but the machine has to make an argumentative relationship that mobile phone is used to communicate and is a communication device using the common sense or knowledge we humans possess.

"Knowledge is a major gridlock to Argument Mining when dealing with a contentious issue in a set of text where arguments can be found." [Gabbriellini and Torroni 2014]

Argumentation is as a discourse activity aimed at increasing or diminishing the worthiness of a disputable claim or some perspective. It is an intelligent task with respect to communication that is inherent to human behavior. In Artificial Intelligence (henceforth, AI), computational argumentation is a new field dealing with Argumentation Mining (AM) which has now turned into a prominent field in the research and study of AI because it has the ability to combine figurative needs with the cognitive models which represent tasks defined by user and computational models based on automated reasoning. AM is becoming one of the core study and research area in the field of cognitive sciences, where some studies have indicated that the functioning of the human brain itself is argumentative. According to P.M. Dung, one of the pioneers in argument computation/computational logic and most of the recent studies on abstract argumentation are based on Dung Framework, "The natural human reasoning is argumentative itself". There have been some other models which do suggest agent-based simulation in the field of computational social-sciences whose micro foundation specifically refers to various argument theories proposed in literature [Gabbriellini and Torroni 2014].

2. Overview of Argumentation Mining

Argumentation is a broad field of research which comprises of studies on debate and reasoning process, and glides across many diverse areas such as rhetoric and law, psychology, logic and philosophy and computer science. Argumentation Mining (henceforth AM) is processing of the natural language of humans by the machine, which in itself is quite a challenging task given from the standpoint of a machine. AM is an offshoot of Artificial Intelligence in which a machine itself extracts arguments occurring in the natural language and constructs inferences from the relations in between the arguments extracted from the given text with the final output being structured data which in turn is machine-processable, used as input for argumentation computational models.

AM from texts approaches started in the early 2000, when methods to mine connotations of argument from natural language texts by different authors began to appear. Teufel [Teufel 1999] introduced the concept of Argument Zoning which deals with argument analysis in AM from rhetorical structure of scientific articles and paper demonstrated that annotations schemes can be applied to life sciences also apart from traditional computational linguistics (CL).

IBM is funding a multi-million cognitive computing project; the Project Debater, which IBM claims is the first AI, based system that can debate with humans on various complex topics.

There are several research areas in AM from Artificial Intelligence perspective: Natural Language Processing (NLP) is the area in which natural language texts are processed through various methods which identify the arguments and their components occurring in the text (i.e., premises and claims) and then the relations among the arguments are predicted. Knowledge Representation and Reasoning (KRR) capabilities are then used upon the arguments and relations that were retrieved through NLP methods so that, for instance, inconsistencies and fallacies can be identified automatically in the texts. This interaction between computer and human helps in designing a quality computer-human digital argumentative software tool.

Current approaches used in AM mainly focus on the current development in the NLP field. Here, we do propose an analysis of the existing work in the current scenario with respect to AM literature and then defining the application of different machine learning algorithm scenarios with respect to the particular argument Model.

3. The Argument Mining Frameworks

Argument(ation) mining can be defined as, "Analysis of the text on the realistic level and then applying an argumentation theory on the model and analyzing the given data" [Mochales and Moens 2011] or it can be thought of as "the analysis of the general task of the discourse on the realistic level and then applying a particular theory of argumentation to the model and analyzing automatically the given data" [Habernal and Gurevych 2017].

3.1 Argumentation Models

Argumentation has its ancient roots in philosophy and dialects whose study can be traced back to 2000 B.C, from Socrates dialogue in rhetoric whose work was taken ahead by Plato in the form of Platonic dialogue. Having such a long tradition of over the centuries, it may come as no surprise that argumentation has pervaded many knowledge based areas such as logic, rhetoric, computer science, law, medicine, communication, and language. Literature is abundant on argumentation models, but it cannot be said that a perfect model has been achieved, however, research is taking new dimensions, and many models have been proposed. In this field different models have been developed during the past years which can be categorized mainly into three different categories as shown in the fig. 1 –

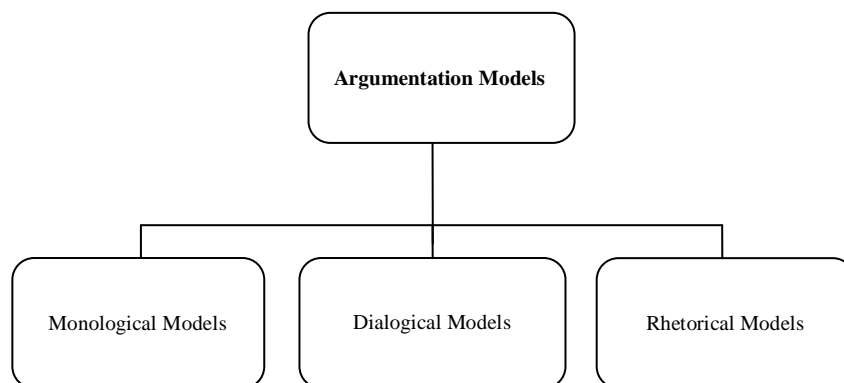


Fig 1: The three categories of Argumentation models

3.1.1 Monological Models

This model assumes a tentative proof of a given argument and then applying a set of rules on its internal structure. Several models based on this approach which addresses the internal structure have been developed [Habernal and Gurevych 2017; Farley and Freeman 1995; Reed and Walton 2003]. These models try to establish a link between the different components of the arguments and how the conclusion relates to the given premises or a set of premises. Their main focus is on the relationships which can exist in between the different components of the argument in a monological structure. Therefore, these types of models are known as Monological Models.

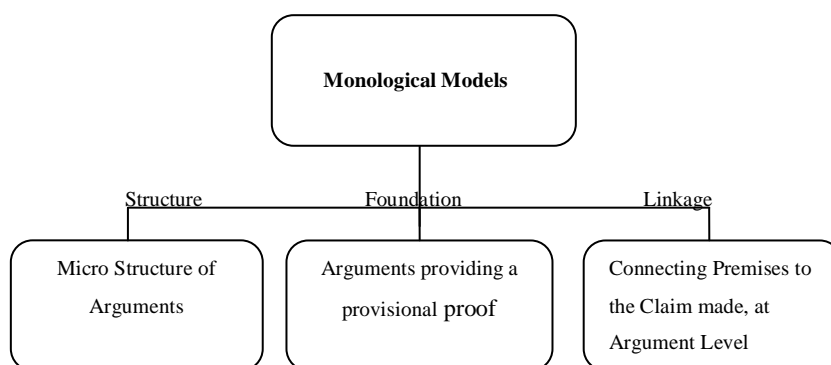


Fig 2: A conceptual framework of Monological Argumentation Model

One of the best monological model is that of Toulmin. [Toulmin 1958]. Toulmin method is an informal reasoning method proposed by Stephen Toulmin which assumes that the human argumentation and reasoning logic microstructure can be further broken down into six basic categories :

- 1) Fact/Data – Facts or data which are incontrovertible or evidences which are used to prove the given argument.
- 2) Claim/Qualifier/Conclusion – It is a subjective statement or inference drawn in support of the possibly controversial fact which is argued about.

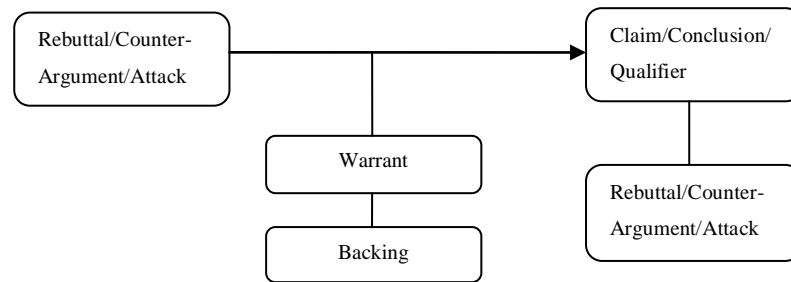


Fig 3: Toulmin Model

3) Qualifiers – Statements that make the arguments weaker which are made in support of the claim, which holds the validity of the argument to be true.

4) Warrants - It is the link between the claim and evidence .In support of the fact, which helps in drawing the inference and making a claim.

5) Backing – Statements made to tell as why the warrant is a rationale one, supporting the Validity of the Warrants only and not necessarily proving the main point which is being argued.

6) Rebuttal/Counter-arguments/Attack – It attacks the validity of the inference drawn or addressing the potential objections raised against the evidence drawn.

Toulmin's model has largely been influential, but one of the criticisms it faces is that it does not considers the implicit questions made in the argumentation. It assumes that an argumentation starts with a claim or a fact and ends up with a conclusion, ignoring the implicit questions arising. For e.g. "X was born in Assam, so he must be an Indian subject", the question arises, "Is X an Indian subject?" which is totally ignored, neglecting which questions should be asked and others not. Therefore, Toulmin's model use in diverse fields such as philosophy, science, law, user-generated content and design becomes a matter of discussion[Habernal et al. 2014].Many extensions to Toulmin have been proposed in the literature.

3.1.1.1 Reed and Walton Monological Model:

Reed and Walton pioneer work on AM is one of the most important literatures found in Monological models. They proposed argumentation schemes which take the form of arguments, defining structure of inference. It describes the model taking notions of argument as product. This notion enables to evaluate and identify common argument types used in everyday discourse. Walton has proposed 25 such schemes in his extended work [Reed and Walton 2003; Walton 2009].

3.1.2. Dialogical Models

A second stream of research in AI has stressed on the existing relationships between the arguments, which at times is considered as abstract entities and totally discarding their internal structure. Since, these types of models emphasize on the argument structure similar to a dialogical framework, hence the name Dialogical Model. Many dialogical models have been proposed, to name a few are that of, Dung[Dung 1995], Bentahar[Bentahar et al. 2010], Hamblin[Hamblin 1970]; MacKenzie[MacKenzie 1979].

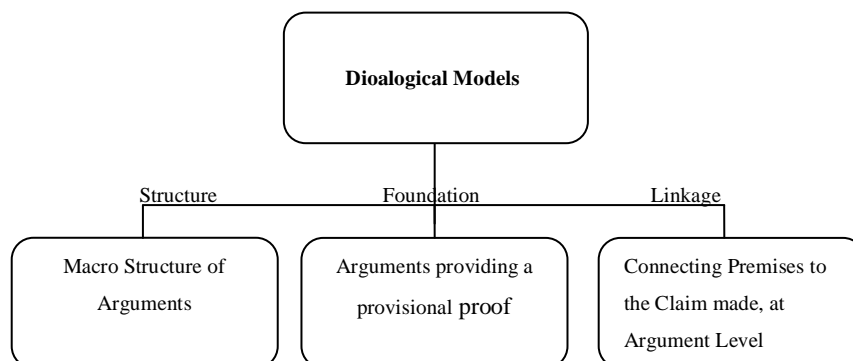


Fig 4: A conceptual framework of Dialogical Argumentation Model**3.1.2.1 Dung Model:**

PM Dung model is one of the most prominent abstract model using semantics of various types for evaluation of arguments. It takes an abstract approach which takes argument as atomic entity, where the entire focus is on the relevant relations in between them. It is known as Abstract Argumentation Framework (AFs for short). AFs are directed graphs with the vertices representing abstract argument and edges representing attacks between them. Conflicts arising are then resolved using semantics and different semantics reflect different perceptions as what can be considered to be reasonable, providing a calculi of opposition. The output is acceptable subsets of arguments, which are called as extensions, each corresponding to various scenarios given the available arguments. Admissibility and conflict-freeness are quite important factors, and based on these more advanced semantics have been proposed – right from PM Dung's original work of stable grounded and preferred semantics to the recent approaches of ideal, semi-stable, cf2-semantics.

Definition:

IF (AF) is a pair (A, R) in Dung abstract argument framework, where A is a set of arguments and R is such that,

$$R \subseteq A \times A$$

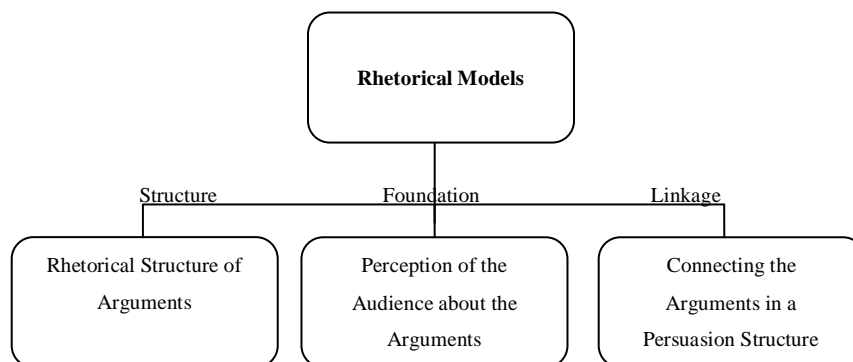
Representing the relation as attack then we say that a_1 attacks a_2 iff,

$$(a_1, a_2) \in R$$

There have been several extensions of Dung's abstract argumentation theory original work, defining and extending rules of his original work, like ASAF (Attack-Support Argument Framework [Cohen et al. 2005], and other extensions [Cayrol et al. 2018].

3.1.3. Rhetorical Model

In general, dialogical and monological models consider the macro (external) and micro (internal) structure of the arguments. Some of the models do not follow both of these two approaches. These

**Fig 5:** A conceptual framework of Rhetorical Argumentation Model

models are called rhetorical models which follow the rhetorical structure of arguments (schemas or rhetorical patterns). In these models, the aim is to take into account the way of using the arguments for the purpose of persuasion. A basic characteristic of these models is that they consider the perception of the audience about a particular argument. Rhetorical models are based on arguments that take into account the world's perception from the viewpoint of the audience along with evaluative judgments, and they do not try to establish the validity of the conclusion.

3.2 Argumentation Methods

An AM system has to perform many strictly interrelated tasks, so before discussing methods, first, we define taxonomy in order to organize different tasks which go under one roof. Next, we will examine Natural Language Processing and machine learning methods which have been applied by the existing systems based on the role they play in the architecture of a typical AM system. The existing systems which have been developed till now implements a pipeline architecture using which they take unstructured text documents as input and produce structured document as an output, where the relations detected in the argument are annotated to construct an argument graph. Each of the stages in the pipeline method corresponds to one subtask in the whole AM problem, which we will discuss briefly in the next section.

The challenges in the AM field share many important similarities in the other subset of AI fields like Natural Language Processing, discourse analysis, machine language, information extraction, knowledge representation and computer linguistics.

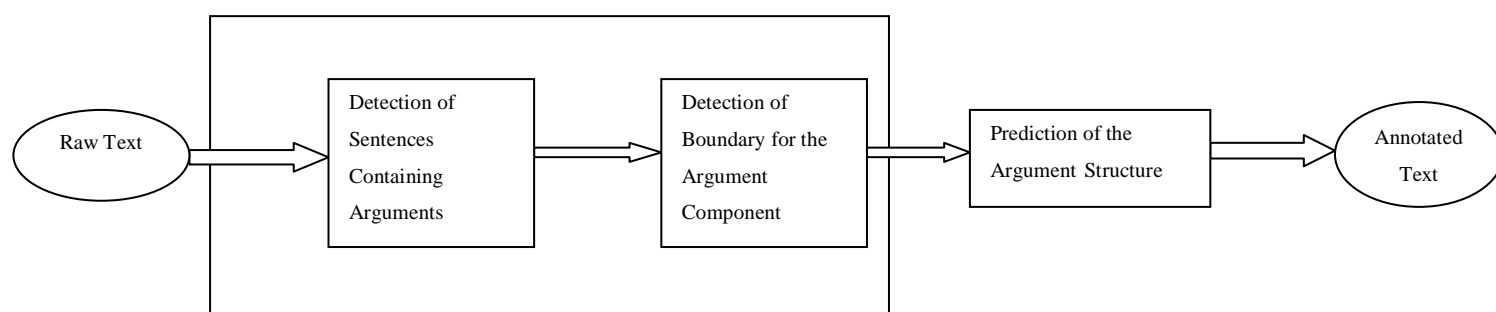


Fig 6: Pipeline Architecture of Argumentation Model

3.2.1. Taxonomy of the AM Problem

The problem of AM can be addressed along five dimensions –

- i) Granularity of the input
- ii) Type/Genre of input
- iii) Argument Model
- iv) Granularity of the target
- v) Goal of Analysis

The granularity of the text which is being processed indicates the level of the details which are searched in an argument (or parts of the argument). Some approaches examine a portion of the text at paragraph level; like the work done by Teufel [Teufel 1999], whose work is considered to be the forerunner in the field of AM. Most of the researches are focused at text on the sentence level whereas; in some other fine grained intra sentence argument component boundaries have been considered. The genre of the text defines what type of data is being input. With type of data it is meant like law/legal, news, essays, online discussions, medical etc.

Each AM system has to be associated with a specific argument model; most of the present existing systems do follow a claim/premise model. With the granularity of the target, it is meant that the aim of the approach is only at some specific components in the text such as claims or the target is the full argument. Thus, the target of AM systems varies in terms of granularity.

The goal of the argument varies on a lot of possibilities depending upon the approaches followed such as detection of arguments, prediction of relations in between arguments, classification problem of arguments, attribution, and completion.

3.3. Component Detection in Arguments-

Argument Component Boundary Detection (ACBD) is one of the key goals in the AM systems first stage, which is to detect arguments or argument components depending upon the desired

granularity of the target. It is considered as the first sub-task in the argument pipeline model. The entities so retrieved will represent a particular node in the argument graph. In AM systems, the approach to this problem is to divide this sub-task into two stages:

3.3.1 Argumentative Sentence Detection

It quite is important to mention that not all AM systems necessarily follow this pipeline model. Some had a different approach[Teufel et al. 2009;Bar-Haim et al. 2017] considering that boundaries of the components in the argument are already detected and they concentrated on a single goal i.e. classification task. In the detection of argument components, the first step is to extract sentences in the text that do contain argument or part of the argument and which can be classified as argumentative. This problem is a classification task which can be addressed by any machine learning classifier, in principle. For this particular task, many different solutions can be proposed, which depends mainly on the argument model which has been adopted and the final goal of the AM system.

Training a binary classifier to differentiate between argumentative and non-argumentative sentences, leaving the task of identifying the argument component type (claim or premises) for the second stage. Training a multi-class classifier to discriminate among all type of argument components existing in the argument model that is being adopted by the AM system. This type of approach assumes that at most there is one argument component in the sentence and then training a set of binary classifiers in the considered model, one for each existing argument component.

In a nutshell, no matter whichever approach we take among the three, we have to choose a classifier type and features to employ. Existing AM systems use a variety of classic machine algorithms such as Logistic Regression [Teufel et al. 2009; Stab and Gurevych 2017], Maximum Entropy classifiers[Lippi and Torroni 2016], Decision Trees and Random Forests[Stab and Gurevych 2014b, Eckle-Kohler et al. 2015], Support Vector Machines (SVM)[Mochales and Moens 2011, Stab and Gurevych 2014b , Eckle-Kohler et al. 2015,Park and cardie 2014].

Table 1 compares a comprehensive comparison of different approaches which are applied to AM tasks by different Authors

MACHINE LEARNING APPROACHES APPLIED	COMPONENT DETECTION		PREDICTION OF RELATIONS STRUCTURE
	SENTENCE CLASSIFICATION	DETECTION OF BOUNDARIES	
Support Vector Machine(SVM)	[Habernal and Gurevych, 2017], [Mochales and Moens, 2011], [Eckle-Kohler et al. 2015], [Duthie et al. 2016], [Lippi and Torroni 2016], [Bar-Haim et al. 2017]	[Mochales and Moens, 2011], [Lippi and Torroni, 2016],	[Naderi and Hirst 2015], [Niculae et al. 2017], [Stab and Gurevych 2017] [Menini et al. 2018]
Parsing Algorithm (P) – using a Context Free Grammar	[Villalba and Saint-Dizier, 2012], [Peldszus and Stede 2015], [Eger et al. 2017]	[Eger et al. 2017]	[Villalba and Saint-Dizier 2012], [Peldszus and Stede 2015], [Eger et al. 2017]
Logistic Regression (LR)	[Levy et al. 2014], [Rinott et al. 2015], [Nguyen and Litman 2018]	[Dusmanu et al. 2017], [Ibeke et al. 2017], [Nguyen and Litman, 2018]	[Nguyen and Litman, 2018]
Maximum Entropy models (ME)	[Mochales and Moens, 2011], [Duthie et al. 2016]	[Mochales and Moens, 2011],	---

Conditional Random Forests (CRF)	[Stab and Gurevych 2017]	---	---
Naive Bayes (NB)	[Duthie et al. 2016], [Eckle-Kohler et al. 2015]	---	---
Random Forest (RF)	---	[Dusmanu et al. 2017], [Eckle-Kohler et al. 2015]	---
Maximum Likelihood (ML)	---	[Levy et al. 2014],	---
Deep Neural Networks with Weak Supervision (DNN)	[Levy et al. 2018],	---	---

Table 1 compares a comprehensive comparison of different approaches which are applied to AM tasks. These are ordered starting from the most frequently applied methods. As for other tasks in NLP, SVMs have proved to be the most performing algorithms in different settings, and for different AM sub-tasks. The acronyms stand for: Support Vector Machine (SVM), Parsing algorithms (P), Logistic Regression (LR), Recurrent Neural Networks for language models (RNN), Maximum Entropy models (ME), Conditional Random Fields (CRF), Naive Bayes (NB), Random Forests (RF), Textual Entailment Suites (TES) and Maximum Likelihood (ML).

These classifiers are trained under supervised settings on a collection of labeled examples from a data set. For each particular example, some part of the text which has to be classified is provided (for e.g., a form of feature vector) associated together with the class (label). This supervised training thus produces a model which is then used to predict new text. Although there are many studies done which can be found in the existing literature which have compared existing approaches in order to tell that which particular classifier should be used, but there has been no clear success. Frequently employed approaches are punctuation and verb tenses information [Mochales and Moens 2011; Eckle-Kohler et al. 2015] and discourse markers, named entity recognition systems [Levy et al. 2018; Rinott et al. 2015].

Table 2: A list of features most frequently computed for AM tasks, from the most frequently used ones.

FEATURES
1. Syntactic and Positional
2. Lexicon
3. Topic Relatedness/Semantic similarity
4. Sentiment
5. Embeddings
6. Pattern(regex)
7. Discourse
8. Bag – of – words
9. Subjectivity classifier
10. NER
11. Vocal (speech)
12. Wikipedia – based
13. PMI
14. Emoticons

It is one of the most advanced AM systems. This system [Levy et al. 2014] addresses the task of building classifiers for the purpose of detection of argumentative sentences, one crucial choice has to be considered, which is whether and how to employ contextual information. Many approaches use domain-specific knowledge. IBM project debator developed in Haifa, Israel is able in detecting context dependent claims automatically from Wikipedia extracted based on this information which, has been provided previously. These type of tasks are articles. In this, the system developed is designed specifically in a work setting where the topic is given in advance and the argument components detected have to be known as Context-Dependent Claim Detection (CDCD) and

Context-Dependent Evidence Detection (CDED).Ruty Rinott[Rinott et al. 2014] gave a follow- up and addressed the task of (CDED). In 2017, Bar-Haim[Bar-Haim et al. 2017] addressed the task of claim stance classification, which can be decomposed into detection of –

- i) Targets of the topic given and the claim
- ii) Sentiment or the polarity towards each of the targets
- iii) Whether targets are consistent or contrastive.

If contextual information is provided, it serves as a decisive factor in the accuracy of the AM system, but on the contrary, it hinders the generalization capability of the AM system. Domain specific features do help in constructing an argument, but they are most likely to overwrite the very data upon which they have been constructed; therefore, in order to build a generic AM system, it should be tested across different application scenarios and corpora.

In order to address the problem, Lippi and Torroni[Lippi and Torroni 2015] proposed Context Independent Claim Detection (CICD), based on Support Vector Machine (SVM) ML algorithm which is based on structured kernels on the constituency parse trees, which in turn was more or less based on the Partial Tree Kernel proposed by Moschitti[Moschitti et al. 2006]. Partial Tree Kernel approach was used for measuring the similarity between the given sentences. The constituency parse tree generally reflects the presence of a claim. Tree kernels construct an implicit feature space automatically and hence does not needs to be defined manually neither there is the need for domain specific, context dependent information.

There are some cases however, in which context dependent information becomes a necessity for example Twitter data, which is used for opinion mining[Pan and Yang. 2010; Pang and Lee. 2008; Grosse et al. 2015] and microblogs data which contains jargons, short hand typing, smileys etc which require specific ML methodologies and representations.

3.3.2 Argument Component Boundary Detection

This stage is the second stage which deals with detecting the exact boundaries of argument component[Stab and Gurevych 2014b], which is also called as an argument discourse unit[Peldszus and Stede 2013] or argument unit[Eckle-Kohler et al. 2016]. This is a segmentation problem dealing with the beginning and end of argument components, which becomes necessary as the whole sentence cannot be taken as a single argument component[Habernal et al. 2014].

There are three cases which needed to be examined while dealing with the input granularity of the given sentence:

- (i) The whole sentence or a part of it becomes the argument component itself;
- (ii) A single sentence may contain multiple arguments;
- (iii) Multiple sentences may connect in support or attack of a single sentence.

The prevailing methods consider any one of the above three possibilities [Levy et al. 2014; Mochales and Moens 2011]. Different types of argument components have quite different and specific characteristics hence the problem of boundary detection is heavily dependent upon the argument model which is chosen. For example, report about an annotation study of ca. 4,000[Habernal et al. 2014] where the sentences are based upon the claim/premise model and the average length of a claim sentence is 1.1 and an average length of of a premise was found to be 2.2 sentences.

Aharoni and Levy[Levy et al. 2014; Habernal et al. 2014] performed AM tasks on the IBM corpus and considered part of the text from the sentence as claims, while the premises stretched across multiple sentences or paragraphs. Maximum likelihood is used to identify probable boundaries in the context dependent claims[Levy et al. 2014; Rinott et al. 2015] used a different approach considering segments of text up to three sentences contained within a single paragraph as evidence supporting/attacking the claim. Some of the research work done totally ignores the boundary detection approach, instead they considered identification of clauses or sub sentences originating from a parsed tree build from the argument component[Mochales and Moens 2011], whereas some work reflects another approach assuming that sentences are already segmented and they focused mainly on their characteristics subdivided into following four types;

- i) premise
- ii) claim
- iii) major claim

iv) none

Some other works explored structured-output and relational classifiers using the sequence labelling approach[Nguyen et al. 2007], where a class or a tag is assigned to each word in a given sentence so that classes were able to differentiate words in a sentence classified as the argument component from the other words. Hidden Markov models, Conditional Random Fields and the likewise algorithms have been applied successfully in this category of approach, for example, the recognition of named entities in tweets[Ritter et al. 2011], extraction of information from the social media data[Imran et al. 2013]. Works by Goudas[Goudas et al. 2014], Sardianos[Sardianos et al. 2015], and Park [Park et al. 2014] contributed significantly in this segmentation problem. Sardianos[Sardianos et al. 2015] also used Recursive Neural Networks algorithms to represent words.

MARGOT (Mining ARGuments frOm Text)[Lippi and Torroni 2016], is a web based tool designed for component classification and segmentation problem and has been tested on the IBM corpus[Bar-Haim et al. 2017, Rinott et al. 2015]. Habernal and Gurevych[Habernal and Gurevych 2017] modified Toulmin's model to identify argument components using a sequence labelling approach in Web discourses on user generated data(sample of controversial topics about education).

3.3.3 Argument Structure Prediction

This is the most complex and final stage of AM system and it deals with predicting the links in between the argumentative sentences or components, depending upon the granularity of the target. This is an extremely challenging task involving the relationships and the connections in between the detected arguments, involving high-level representation of knowledge and reasoning issues. This task is known as 'prediction' rather than 'detection', as the final goal is not a portion of a single sentence but a link between portions of sentences.

The output of this prediction stage is a graph, connecting the arguments retrieved in the process where edges of the graph depict relations among the arguments retrieved like entailment and support/attack on the claim in the sentence. Graph representation is invaluable with respect to retrieving information from social media and web documents in many applications such as analysis of debates, social media network influence etc.

4. Conclusions

Many models have been developed and tested[Walton 2009; P.M. Dung 1995; Bentahar et al. 2010; Teufel et al. 2009] over different corpora and using different Machine Learning algorithms, with the goal of predicting the argumentative sentences and their links and providing optimum answer with high level of accuracy.

For AM using ML and AI techniques requires a collection of annotated documents also known as corpus, which are used as training set for any kind of predictor. Constructing a corpus is a complex and time consuming task but is required for obtaining homogenous and consistent annotations. Argument mining in a particular domain needs the corpus related to that domain so as to identify the argument components, their exact boundaries and the existing relations between them[Mochales and Ieven 2009; Habernal et al. 2014]. Different datasets have to be constructed for particular domain with specific objectives because every corpora does not fits to every approach taken for all the stages of the argumentation pipeline.

Finally, we summarize on the performance achieved by the various argumentative systems taken in this survey paper as they are supposed to have achieved high level of accuracy. These systems have been tested on various corpora using different machine learning methods with the single objective of getting high level of accuracy with respect to mining of arguments and claims.[Mochales and Moens 2011] reported an accuracy of 73% and 80% on argumentative sentences classification task on the AraucariaDB datasets and European Court of Human Rights(ECHR) datasets respectively. It is one of the most successful extraction of claims and he supporting premises in the Legal domain. With respect to the identification of the argument components and classification task they achieved a precise/recall of 77%/61% for conclusions and 70%/66% of accuracy for premises.[Rooney et al. 2012] achieved an accuracy of 65% on AraucariaDB dataset with respect to argument classification task.[Lippi and Torroni 2015; Levy et al. 2018] achieved an accuracy rate of 17% F_1 and 18% F_1 with respect to claim mining on the same IBM corpus.[Lippi and Torroni 2015] achieved a F_1 -Score of 71% on the persuasive essay corpus, however, [Stab and Gurevych 2014b] achieved F_1 -Score of 63% for detection of major

claims, 54% on detection of minor claims and 83% on the premise classification on the same dataset.[Goudas et al. 2014] reported a 52% F_1 -Score on the argumentative sentence classification task on a Greek corpus social media dataset.[Sardianos et al. 2015] reported F_1 between 15% and 20% for boundary detection problem.

5. Future Scope

AM is one of the most prominent field of research in AI at present which can greatly benefit the customer profiling and market analysis through mining data in social media and web. It could provide development of reasoning engines for the arguments originating from the web or social media. AM may be the next step after opinion mining and sentiment analysis in the field of Artificial Intelligence[Habernal et al. 2014]. While opinion mining deals with ‘what people think about someone or something’, AM helps to understand ‘why’ they think so with the help of reasoning and causes as to understand the real cause of the present mindset of the people. The goal of AM is to analyse the process of ‘human reasoning’ through which humans rationally accept or reject an argument, opinion or a theory.

Argumentation tools pave the way to a lot of new futuristic applications cutting across many disciplines like humanities, social sciences, engineering. Decision makes and policy learning[Stab et al. 2018], could be employed automatically for extracting arguments in order to improve models and support strategic choices. Engineering workflow processes have already exploited argumentation models for the automated evaluation of alternative design solutions[Moschitti et al. 2006], and argumentation mining could be an additional asset in the process.

Customer profiling and market analysis could greatly benefit from the analysis of arguments provided by users over the Web. Here, argumentation mining might unlock innovative ways of organizing, supporting, and visualizing online debates, by clustering posts, and proposing new rankings. Finally, AM has the potential of dealing with the building of sophisticated AI systems capable of maintaining structured knowledge representation in open domains from the unstructured data.

References

- [1] Simone Gabbriellini and Paolo Torroni. Advances in Intelligent Systems and Computing. A New Framework for ABMs Based on Argumentative Reasoning, 2014, Volume 229:25-3, ISBN : 978-3-642-39828-5.
- [2] Simone Teufel. 1999. Argumentative Zoning. PhD thesis. University of Edinburgh.
- [3] Raquel Mochales Palau and Marie-Francine Moens. 2011. Argumentation mining. Artificial Intelligence and Law 19, 1 (2011), 1–22.
- [4] I. Habernal and I. Gurevych. Argumentation mining in user-generated web discourse. Computational Linguistics, Volume 43 Issue 1(125-179), April 2017, MIT Press Cambridge
- [5] Arthur M. Farley Kathleen Freeman, Burden of proof in legal argumentation, ICAIL '95: Proceedings of the 5th international conference on Artificial intelligence and law May 1995 Pages 156–164.
- [6] Chris Reed Douglas Walton, Argumentation schemes in argument-as-process and argument-as-product. 2003, https://www.researchgate.net/publication/255630807_Argumentation_schemes_in_argument-as-process_and_argument-as-product.
- [7] Stephen E. Toulmin. The Uses of Argument: Cambridge University Press (1958).
- [8] Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, Benno Stein. 2014. The Argument Reasoning Comprehension Task: Identification and Reconstruction of Implicit Warrants.
- [9] Douglas Walton. 2009. Argumentation theory: A very short introduction In Argumentation in Artificial Intelligence, Guillermo Simari and Iyad Rahwan (Eds.). Springer, 1–22.
- [10] Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. 77(2) (1995) 321–358.
- [11] Jamal Bentahar, Bernard Moulin, and Micheline Belanger. 2010. A taxonomy of argumentation models used for knowledge representation. Artificial Intelligence Review 33, 3 (2010), 211–259.
- [12] C. L. Hamblin, Fallacies, Vale Press (1970).
- [13] R.A.F. MacKenzie, Jesuit School of Theology Berkeley, CA., The Transformation of Job, 1979
- [14] Andrea Cohen, Sebastian Gottifredi, Alejandro J. García, Guillermo, R. Simari. An approach to abstract argumentation with recursive attack and support. Journal of Applied Logic 13 (2015) 509-533.
- [15] Claudette Cayrol, Marie-Christine, Lagasque-Schiex. Argumentation Frameworks with Recursive Attacks and Evidence-Based Supports. Foundations of Information and Knowledge Systems 2018 (pp.150-169)
- [16] Simone Teufel, Advait Siddharthan, Colin R Batchelor. Towards discipline-independent argumentative zoning: evidence from chemistry and computational linguistics. EMNLP '09: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3 August 2009 Pages 1493–1502.
- [17] Christian Stab and Iryna Gurevych. 2014b. Identifying argumentative discourse structures in persuasive

- essays. In EMNLP 2014, Alessandro Moschitti, Bo Pang, and Walter Daelemans (Eds.). ACL, 46–56.
- [18] Ran Levy, Ben Bogin, Shai Gretz, Ranit Aharonov, Noam Slonim, 2018 IBM Research.
- [19] Milan Li, Yang Gao, Hui Wen, Yang Du, Haijing Liu, Hao Wang Joint RNN Model for Argument Component Boundary Detection. In IEEE SMC 2017.
- [20] Ruty Rinott, Lena Dankin, Carlos Alzate Perez, Mitesh M. Khapra, Ehud Aharoni, and Noam Slonim. 2015. Show me your evidence—An automatic method for context-dependent evidence detection. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP'15), Lluís M. Marquez, Chris Callison-Burch, Jian Su, Daniele Pighin, and Yuval Marton (Eds.). The Association for Computational Linguistics, 440–450.
- [21] Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. Stance Classification of context-dependent claims. In EACL, pages 251–261, 2017.
- [22] Christian Stab and Iryna Gurevych. Parsing argumentation structures in persuasive essays. *Comput. Linguist.*, 43(3):619–659, 2017.
- [23] Marco Lippi and Paolo Torroni. Argument mining from speech: Detecting claims in political debates. In AAAI, pages 2979–2985, 2016.
- [24] Christian Stab and Iryna Gurevych. 2014b. Identifying argumentative discourse structures in persuasive essays. In EMNLP 2014, Alessandro Moschitti, Bo Pang, and Walter Daelemans (Eds.). ACL, 46–56.
- [25] Judith Eckle-Köhler, Roland Kluge, and Iryna Gurevych. 2015. On the role of discourse markers for discriminating claims and premises in argumentative discourse. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, 2236–2242.
- [26] Rory Duthie, Katarzyna Budzyska, and Chris Reed. Mining ethos in political debate. In COMMA, pages 299–310, 2016.
- [27] Joonsuk Park and Claire Cardie. 2014. Identifying appropriate support for propositions in online user comments. In Proceedings of the 1st Workshop on Argumentation Mining. Association for Computational Linguistics, 29–38. <http://www.aclweb.org/anthology/W/W14/W14-2105>.
- [28] Vlad Niculae, Joonsuk Park, and Claire Cardie. Argument mining with structured svms and runs. In ACL, pages 985–995, 2017.
- [29] Nona Naderi, Henning Wachsmuth, Ivan Habernal, Yufang Hou, Graeme Hirst, Iryna Gurevych, and Benno Stein. Argumentation quality assessment: Theory vs. practice. In ACL, pages 250–255, 2017.
- [30] Christian Stab, Iryna Gurevych. Parsing Argumentation Structures in Persuasive Essays, *Computational Linguistics*, Volume 43, Issue 3 - September 2017
- [31] Stefano Menini, Elena Cabrio, Sara Tonelli, and Serena Villata. Never retreat, never retract: Argumentation analysis for political speeches. In AAAI, pages 4889–4896, 2018.
- [32] Marco Lippi and Paolo Torroni. Margot: A web server for argumentation mining. *Expert Systems with Applications*, 65:292–303, 12 2016.
- [33] Christian Stab and Iryna Gurevych. 2014a. Annotating argument components and relations in persuasive essays. In COLING 2014, Jan Hajic and Junichi Tsujii (Eds.). ACL, 1501–1510.
- [34] [Peldszus and Stede 2013] Andreas Peldszus, Manfred Stede, From Argument Diagrams to Argumentation Mining in Texts: *International Journal of Cognitive Informatics and Natural Intelligence* 7(1):1-31.
- [35] Maria Paz Garcia Villalba and Patrick Saint-Dizier. 2012. Some facets of argument mining for opinion analysis. In COMMA 2012 (Frontiers in Artificial Intelligence and Applications), Bart Verheij, Stefan Szeider, and Stefan Woltran (Eds.), Vol. 245. IOS Press, 23–34.
- [36] Steffen Eger, Johannes Daxenberger, Iryna Gurevych, 2017, Neural End-to-End Learning for Computational Argumentation Mining Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).
- [37] Nam Nguyen and Yunsong Guo. 2007. Comparisons of sequence labeling algorithms and extensions. In Proceedings of the 24th International Conference on Machine Learning. ACM, 681–688.
- [38] Mihai Dusanu, Elena Cabrio, and Serena Villata. Argument mining on twitter: Arguments, facts and sources. In EMNLP, pages 2317–2322, 2017.
- [39] Ebuka Ikebe, Chenghua Lin, Adam Z. Wyner, and Mohamad Hardyman Barawi. Extracting and understanding contrastive opinion through topic relevant sentences. In IJCNLP, pages 395–400, 2017.
- [40] Ran Levy, Ben Bogin, Shai Gretz, Ranit Aharonov, Noam Slonim. 2018. Towards an argumentative content search engine using weak supervision. Proceedings of the 27th International Conference on Computational Linguistics, pages 2066–2081. Santa Fe, New Mexico, USA, August 20-26, 2018
- [41] Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In COLING 2014, Jan Hajic and Junichi Tsujii (Eds.). ACL, 1489–1500. <http://www.aclweb.org/anthology/C14-1141>.
- [42] Marco Lippi and Paolo Torroni. 2015. Context-independent claim detection for argument mining. In Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI'15), Qiang Yang and Michael Wooldridge (Eds.). AAAI Press, 185–191.
- [43] Alessandro Moschitti. 2006. Efficient convolution kernels for dependency and constituent syntactic trees. In Machine Learning: ECML 2006, Johannes Frnkranz, Tobias Scheffer, and Myra Spiliopoulou (Eds.). Lecture Notes in Computer Science, Vol. 4212. Springer, Berlin, 318–329.
- [44] Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22, 10 (Oct. 2010), 1345–1359
- [45] Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2, 1–2 (Jan. 2008), 1–135.

- [46] Kathrin Grosse, Maria P. Gonzalez, Carlos Iván Chesnevar, and Ana Gabriela Maguitman. 2015. Integrating argumentation and sentiment analysis for mining opinions from Twitter. *AI Communications* 28, 3(2015), 387–401.
- [47] Ivan Habernal, Iryna Gurevych, Henning Wachsmuth, Benno Stein. 2014. The Argument Reasoning Comprehension Task: Identification and Reconstruction of Implicit Warrants.
- [48] Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In *COLING 2014*, Jan Hajic and Junichi Tsujii (Eds.). ACL, 1489–1500. <http://www.aclweb.org/anthology/C14-1141>.
- [49] Alan Ritter, Sam Clark, Mausam, and Oren Etzioni. 2011. Named entity recognition in tweets: An experimental study. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'11)*. Association for Computational Linguistics, Stroudsburg, PA, 1524–1534. <http://dl.acm.org/citation.cfm?id=2145432.2145595>.
- [50] Muhammad Imran, Shady Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. 2013. Practical extraction of disaster-relevant information from social media. In *Proceedings of the 22nd International Conference on World Wide Web Companion (WWW'13 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 1021–1024.
- [51] Goudas T., Louizos C., Petasis G., Karkaletsis V. (2014) Argument Extraction from News, Blogs, and Social Media. In: Likas A., Blekas K., Kalles D. (eds) *Artificial Intelligence: Methods and Applications. SETN 2014*. Lecture Notes in Computer Science, vol 8445. Springer, Cham. https://doi.org/10.1007/978-3-319-07064-3_23
- [52] Christos Sardianos, Ioannis Manousos Katakis and Georgios Petasis Vangelis Karkaletsis Argument Extraction from News. *Proceedings of the 2nd Workshop on Argumentation Mining(2015)*. DOI: 10.3115/v1/W15-0508.
- [53] Michela Milano, Barry O'Sullivan, and Marco Gavanelli. 2014. Sustainable policy making: A strategic challenge for artificial intelligence. *AI Magazine* 35, 3 (2014), 22–35.
- [54] Pietro Baroni, Marco Romano, Francesca Toni, Marco Aurisicchio, and Giorgio Bertanza. 2015. Automatic evaluation of design alternatives with quantitative argumentation. *Argument and Computation* 6, 1(2015), 24–49.
- [55] Macagno, F., Walton, D. & Reed, C. (2017). Argumentation Schemes. History, Classifications, and Computational Applications. *Journal of Logics and their Applications* 4(8): 2493-2556
- [56] Marco Lippi and Paolo Torroni. Argumentation mining: State of the art and emerging trends. *ACM Trans. Internet Techn.*, 16(2):10, 2016.
- [57] Christian Stab and Iryna Gurevych. Parsing argumentation structures in persuasive essays. *Comput. Linguist.*, 43(3):619–659, 2017.
- [58] Christian Stab, Tristan Miller, and Iryna Gurevych. Cross-topic argument mining from heterogeneous sources using attention-based neural networks. *CoRR*, abs/1802.05758, 2018.
- [59] P.M. Dung, F. Toni, P. Mancarella, Some design guidelines for practical argumentation systems, in: *Proc. Computational Models of Argument, COMMA'10*, 2010, pp.183–194.
- [60] P.M. Dung, P.M. Thang, Closure and consistency in logic-associated argumentation, *J. Artif. Intell. Res.* 49(1) (2014) 79–109.
- [61] Trevor J. M. Bench-Capon and Paul E. Dunne. 2007. Argumentation in artificial intelligence. *Artificial Intelligence* 171, 10–15 (2007), 619–641.
- [62] Christian Stab, Tristan Miller, and Iryna Gurevych. Cross-topic argument mining from heterogeneous sources using attention-based neural networks. *CoRR*, abs/1802.05758, 2018.
- [63] Ehud Aharoni, Anatoly Polnarov, Tamar Lavee, Daniel Hershcovich, Ran Levy, Ruty Rinott, Dan Gutfreund, and Noam Slonim. 2014. A benchmark dataset for automatic detection of claims and evidence in the context of controversial topics. In *Proceedings of the 1st Workshop on Argumentation Mining*. Association for Computational Linguistics, 64–68. <http://acl2014.org/acl2014/W14-21/pdf/W14-2109.pdf>.
- [64] Macagno, F., Walton, D. & Reed, C. (2017). Argumentation Schemes. History, Classifications, and Computational Applications. *Journal of Logics and their Applications*.