Title	A study of modeling 2-tier argumentation framework with modified Toulmin structure
Author(s)	閻, 真竺
Citation	
Issue Date	2022-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17627
Rights	
Description	Supervisor:東条　敏, 先端科学技術研究科, 修士（情報科学）

Master's Thesis

A study of modeling 2-tier argumentation framework with modified
Toulmin structure

2010019 Yan Zhenzhu

Supervisor Satoshi Tojo

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Information Science)

February, 2022

Abstract

Argumentation, which exists everywhere, is a manifestation of human intelligence. Arguments arise when there is a conflict between the statements (arguments) that one wants to express. When we are faced with a large amount of information with opinions, such as legal regulations, user-generated texts in social media, scientific articles, etc., obtaining and analyzing these arguments is an important research direction in artificial intelligence, urged to emerge the development that focuses on analyzing arguments from auto-generated data (a.k.a. argument mining).

Dung's pioneering theory of abstract argumentation explains its wide application as a general framework for all kinds of non-monotonic reasoning, and more generally, reasoning in the presence of conflict. Dung's argumentation framework is instantiated by arguments and binary conflict-based attack relations, based on some underlying logical theoretical definitions. However, the description of each argument itself in Dung's proposal is not detailed enough, we hope to provide a structure for each argument. Toulmin proposes an argumentation model in which claims against challenges can be defended. This model replaces the traditional concepts of "claim" and "premise" with new concepts such as "claim", "evidence" and "warrant".

In our study, the design of Toulmin was adjusted to propose a simplified Toulmin's structure. We show that Toulmin's idea provides a visual interpretation of logic-based arguments and produces a human-comprehensible form. Then, this thesis implements the findings and proposes a novel 2-tier Argumentation Framework that combines the advantages of Toulmin's model and Dung's Framework. To demonstrate its effectiveness, we consider the available datasets annotated in Toulmin's method. We develop machine learning models by using Argument Mining techniques to automatically indicate each components of Toulmin's scheme from text.

Finally, combining the results of the machine learning model and the 2-tier Argumentation Framework we just mentioned, we built a 2-tier Argumentation Framework system to achieve the purpose of allowing people to better understand the arguments that appear in the text.

Keywords: Argumentation Framework, Toulmin Model, Argument Mining, Project Debater Datasets

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Argumentation is an important area of artificial intelligence. Argumentation has a long history, dating back to Aristotle in the 4Th century BC, and it has been fully examined from the different fields of logic, philosophy, law, linguistics, computer science, and some others. Argumentation is a process of debate and inference, In our daily conversations, human being search for premises from different stances to discuss their arguments, depending on the topics. We can refer to the deduction of logic language to express this argumentative reasoning process. This kind of argumentative reasoning can be formalized by using the logical language of the premises and the consequence relationship to prove that the claim logically follows the premises. Artificial Intelligence (AI), which has a long history and a wide range of applications. In the research scope of artificial intelligence, one of the important areas is Natural Language Processing (NLP), which is dedicated to realizing the goal of realizing communication between humans and machines in natural language. In recent years, with the popularization of smartphones, etc. and the acquisition of a large amount of data, great breakthroughs have been made in machine learning and neural network technologies

There are a large number of proposals for argumentative processes based on the language of logic(logic-based formalization of argumentation). In these works, argumentation often come in pairs (e.g. Dung's Abstract Argumentation Framework [7]). These works allow representations of arguments for claims, representations of their counterarguments depend on their stances (Pros and Cons), and showing relationships between arguments. The relationship between arguments can be simply divided into attack and support.

Here is a simple example about Dung's Abstract Argumentation Frame-

work. To illustrate, let us consider two individuals argument $a1$ and $a2$ representing the sale of violent video games to minors.

> a1: *Exposure to violent video games causes at least a temporary increase in aggression and this exposure correlates with aggression in the real world.*
>
> a2: *Video game violence is not related to serious aggressive behavior in real life.*

$a1$ and $a2$ are two arguments connected by contradictory conclusions and so attack each other. Although these works can represent arguments and their relationships well, they do not allow humans to visualize elements of natural language sentences constituting an argument with structure.

One of the most influential models of argumentation, the Toulmin model, was first proposed in 1958 by the British philosopher Stephen Toulmin [33]. The model takes a single logical view of the argument and focuses on the internal structure of the argument, with a detailed internal structure that we will focus on in Chapter 2. From this one of the research goals of this work: to make Toulmin arguments computable.

Regarding how to analyze and obtain arguments, in recent years, there is a new field called Argument Mining (AM) based on Natural Language Processing(NLP). One of the main tasks of argument mining technology is how to automatically detect and obtain structured arguments from articles. With the spread of e-commerce and intelligent analytic and the growing user-generated data generated by social media such as Twitter, Facebook and various forums. The need to discover, identify, and analyze arguments in a structured way from large-scale information flows highlights the importance of argumentative mining. You can see a simple example from Wikipedia articles in Figure 1.1: Wikipedia is a great source for argumentative texts, automatically fetching arguments from web texts.

Currently, the techniques used to extract information from the above data sources are mainly based on statistics and network analysis, such as opinion mining and social network analysis. The Argumentation Mining system enables massive qualitative analysis of arguments, decisions, reviews, etc. In professional newspaper articles, government reports, court judgment records, online social network content, and more, providing unprecedented automation for policymakers and researchers in the social and political sciences tools to create new prospects for corporate marketing.

· Original Text

A 2001 study found that exposure to violent video games causes at least a temporary increase in aggression and that this exposure correlates with aggression in the real world. A decrease in prosocial behavior (caring about the welfare and rights of others) was also noted.
<Effects on aggressive behavior >
Anderson, C.A., & Bushman, B.J. (2001).
The most recent large scale meta-analysis examining 130 studies with over 130,000 subjects worldwide-- concluded that exposure to violent video games causes both short-term and long-term aggression in players and decreases empathy and prosocial behavior.

· Claim:

Exposure to violent video games causes at least a temporary increase in aggression and this exposure correlates with aggression in the real world.

· Evidence:

The most recent large scale meta-analysis -- examining 130 studies with over 130,000 subjects worldwide: concluded that exposure to violent video games causes both short-term and long-term aggression in players.

Figure 1.1: Example of text-based argument mining, text from the IBM's data sets

AM can be used in many domains; for instance, legal documents, instructional texts, student essays, User—Generated Web texts, social-media texts, Twitter articles, etc. In particular, Wikipedia articles represents a wealth of controversial data.

In the face of complex data information, a suitable and well-labeled data set is not only for the argument mining, but also is particularly important for the entire research about argumentation study. In this work, we mainly focus on the IBM Data-sets which you can find at this URL.

The IBM company used Wikipedia and other data extensively to create the first AI system that could debate complex topics with humans for their extensive work on Project Debater [15] ,. Debater can respond to a given topic by automatically constructing a set of relevant pro/con arguments expressed in natural language.

After Dung's Argumentation Framework's theory, plenty of structural frameworks have been proposed. In particular, these contents can have a logical formalization, for example, Dung proposed Assumption-Based Argumentation [8], ASPIC+ Framework [23]. Given that the existing structural framework for arguments focuses more on the arguments themselves, and although arguments can be formalized based on logic, this idea has been qualified well. Hence all of related works to Dung's Argumentation Framework does not allow arguments to be computed; and the definition of arguments in Dung's Abstract Argumentation Framework is not comprehensive enough, we hope to combine some Argument structure models such as Toulmin's model to it. Based on Dung's Abstract Argumentation Framework, we are going

to propose a application with two-tier argumentative model, which we call it two-tier Argumentation Framework (2-tier AF). Our main goal is to exploit the inter-pretability and readability of Toulmin's to human being.

Here, we suppose that the knowledge base set(S) is represented by classical propositional logic (PL); thus studying proof theory in PL to construct arguments and counterarguments from S.Define a subset of the knowledge base by judging whether argument components(such as claims) in the subset can derive the arguments or not.Furthermore, by applying the idea of argument structure, (Toulmin's structure will be used in this paper) for the structured argument framework (2tier-AF), we can not only give structure to arguments , but also express relationships between those structure arguments. In addition to this, we can use logic programming to show the relationship between these argument components, that is, this is why arguments can be computed, which makes human beings could understand the content in the clutter of information in the daily life much more better.

1.2   Background

The answer to the question of what is an argument has been of great interest in philosophy. The better known among them is the thought of Aristotle, which he mentions in his book *Rhetorics*: He tried to develop a general theory of persuasion and applied it to three types of speech:

- the deliberative speech , which advises on future courses of action;

- the judicial speech, which accuses or defends a person whose past activities will be tried;

- the epideictic speech (or epideictic oratory), which praises or accuses a person of public interest.

There have also been some definitions of argumentation in recent years, one of them is following. This definition nicely combines important aspects of the argumentation. van Eemeren and Grootendorst came up with [37]:

> "Argumentation is a verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint."

- verbal activity People can do this by gesturing or frowning. This may be a way to handle some differences. However, in our notion here, discussion, whether verbal or written, is essentially an oral activity. This should be distinguished from combat, for example, which could be another way to resolve a dispute.

- Social activity Society emphasizes that discussion is an interaction between multiple people, at least two people. Of course, many of us make difficult decisions for some time and then consider different outcomes, but the actual discussions need to be done individually.

- Reasonable activity Obviously, we can engage in a variety of verbal and social ways. Among them, the discussion is especially directed to the dimension of rationality. When one person "reasonably" reminds another person, it finds a useful way of speaking in gentle communication, or rational discussion, rather than emotional outbursts, power struggles or personal attacks.

Hence, in addition to the Aristotle mentioned above, one question to ask at this point about standpoint concerns the range of possible positions, which Eggs proposes are the following three basic categories [9]:

- the epistemic argument: some proposition is true or false;

- the ethical (or esthetical) argument: something is good or bad (or: beautiful or ugly)

- the deontic argument: some action should be done or not done.

In the usual case, we can judge the standpoints of an argument by the vocabulary in the argument.

The 1990s saw some of the foundational work in argumentation theory. This includes the work of Simari and Loui [32] that later evolved into Defeasible Logic Programming. Then the ground-breaking work's [38] way of constructing arguments has subsequently been applied in the various versions of the ASPIC+ formalism [22].

There are two approaches stand out for their ability to model a wide range of existing formalisms for non-monotonic inference. First of all, there is the Dund's Abstract Argumentation Framework, which is shown to be able to model formalisms like default logic, logic programming under stable and well-founded model semantics, as well as Nute's defeasible logic [11] and logic programming under the 3-valued stable model semantics [42]. Secondly, the Assumption-based Argumentation Framework of Dung models formalisms

like default logic, logic programming under stable model semantics, auto epistemic logic, and circumscription [4]. One of the essential differences between these two approaches is that Abstract Arguments Framework are based on arguments. People use the information in the knowledge base to build arguments and examine how those arguments attack each other. The definition of different semantics is then based on the generated framework. On the other hand, in the Assumption-based Argumentation Framework, the definition of different semantics is not based on arguments, but on sets of assumption that attack each other based on their possible inferences.

1.3   Objective

As far as this research is concerned, the research objectives can be divided as follows:

1 . Get arguments from massive amounts of data from the IBM Data sets.

2 . Enrich the description of arguments in Dung's Argumentation Framework, based on the Toulmin's theory.

3 . Define a new attack relation between structured arguments, called undercutting.

4 . Propose the knowledge-based 2tier AF so that arguments can be computed by Logic Programming.

5 . Visualization, to show that 2-tier Argumentation Framework enable users to understand human being's argument.

1.4   Contribution

We summarize the contributions of our work in the following areas:

1 .The Dung's AF does not provide a structure for each argument. Hence, we want to provide the structure using the Toulmin's formalism and study how to automatically build up this extended framework from text.

2 .To introduce a new attack method as a 2-tier Argument Framework system based on the Toulmin's model structure. Since there is no description of the undercut relationship in IBM's data set, in order to

completely construct our 2-tier AF structure, we marked the undercut relationship between the Claim and the Evidence components in the IBM data set.

3 . we proposed a system 2-tier Argumentation Framework to show how arguments can be computed

4 . To show that the 2-tier Argumentation Framework enable users to understand human being's argument.

1.5   Thesis structurtion

The reminder of the structure of this paper is as follows: Chapter 2 provides general concepts, definitions, and semantics of argumentation theory. We also introduce the structure of arguments, including linkages in structures and types of structures, and some argument models. and some datasets in the domain of argumentation. The third chapter introduces the formalization of 2-tier Argumentation Framework and related research: related works to Argumentation Framework and Argument Mining research. Chapter 4 presents the design and structure of our proposed framework and briefly introduces the data and models used in the machine learning framework. In Chapter 5, we present the evaluation results of the machine learning framework and assess whether people can better understand arguments through 2-tier Argumentation Framework. Chapter 6 summarizes our work and discusses potential future directions

Chapter 2

Preliminary

2.1 Abstract Argumentation Framework

2.1.1 Basic

In 1995 Dung proposed a new formalization of argument called Argumentation Framework(AF) or Abstract Argumentation Framework(AAF) . Compared with previous argumentation works, Dung's AF only focuses on the relationship between arguments and does not focus on the internal structure and relationship of arguments.

Abstract Argumentation Framework consists of two kinds of directed graphs: circles and arrows: the nodes represented by the circles represent an argument, and the arrows represent the relationship between the two arguments. A binary relationship expressed as attack relationship. Relationships with other arguments can represent the standpoint represented by the argument.

More formally, Dung's AF is defined as a pair $< S, attacks >$ where S is a set of arguments and attacks is a binary relation of the arguments in the form of $S \times S$. As shown in Figure 2.1 where a, b, and c represent three arguments in the S , $S = a, b, c$, the attack relationship between them can be written as $attacks = (a, b), (b, c)$
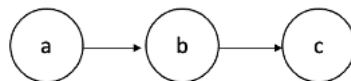
Figure 2.1: In this sample Dung's argumentation framework, A attacks B and B attacks C.

Definition 2.1. *An argumentation framework is a pair*

$$AF =< AR, attacks >$$

where AR represents a set of arguments and attack is a binary relation on AR, i.e. $attack \in AR \times AR$. For two arguments A and B, attacks(A, B) means that A represents an attack on B.

Definition 2.2. *Let $AF = (AR, R)$, $S \subseteq AR$ and let A,B,C,...denote arguments in s .Then:*

A set of $S \subseteq AR$ is conflict-free iff there are no arguments $A, B, c, .. \in S$ not the case that $(A, B), (B, C)... \in R$.

Definition 2.3. *Acceptability for Dung's Argumentation Frameworks*

Let $AF = (AR, R)$, $S \subseteq AR$,S defends $A \in AR$. The set of S is acceptable,iff, for any argument $B \in AR$, $(B, A) \in R$, there is $C \in S$ such that $(C, B) \in R$.

Argument acceptability in stand of the evaluation of argument status. S is said to be acceptable if S has conflicts-free (arguments of S do not attack each other) and all arguments of S are acceptable. Second, assess the state of the discussion in terms of credible or not. An extension of the framework defined under various semantics.

All arguments are considered to have the same strength. Each attack is considered successful.In addition to it is attacked in reverse. For example, in Figure 2.1, A attacks B, and B attacks C. In this case, accepting B and its attacker (A) or what it attacks (B) is incoherent. More formally, acceptability means finding a subset that is conflict-free and all of them defend themselves. A set s defends any arguments b of attack a if s contains the arguments of attack b. An argument a is acceptable for defending the set it belongs to.

Definition 2.4. *Extensions of a Dung's Argumentation Framework*

Let $AF = (AR, R)$, $S \subseteq AR$,S defends $A \in AR$ .Then

- $Defended(S) := \{A|SdefendsA\}$ iff set s defends $A \in AR$.

- Set S is a complete extension of AF iff Set S is conflict-free, and$S = Defended(S)$.

- Set S is a grounded extension of AF iff Set S is the minimal completeextension.

- Set S is a preferred extension of AF iff Set S is the maximal completeextension.

- Set S is a stable extension of AF iff Set S is conflict-free and set S attacks every argument which is not belongs to S.

The idea of $Defendis$ mentioned in the semantics before, and the statement of Acceptability is used in the Dung's Argumentation Framework.

Definition 2.5. *Arguments' status in the Dung' Argumentation Framework*
Let $AF = (AR, R)$, $S \subseteq AR$, S is a

- At least one S extension of AF has $A \in AR$ in it, then A is justified credulously in the S semantics.

- All of the S extension of AF has $A \in AR$ in it, then A is justified sceptically in the S semantics.

- Not any S extension of AF has $A \in AR$ in it, then A is rejected in the S semantics.

Definition 2.6. *The characteristic function F_{AF}, $AF = (AR, R)$ is defined as follows:* $\quad \big| \; F_{AF} = 2^A F \rightarrow 2^A F$

2.1.2 Labelling-based approach

Let's see the definition of approach to labelling. There is a lot of work on how to define labelling for Argumentation Framework, you can find it here

Definition 2.7. *Labelling Function*

Given an Argumentation Framework AF, Let AF=(AR,R). label $\in \{IN, OUT, UNDEC\}$, and Labelling ($\mathcal{L}$)is a total function that help $A \in AR$ with label $\in \{IN, OUT, UNDEC\}$ to AF=(AR,R).

- $\mathcal{L}_{IN} = \{A | \mathcal{L}(A) = IN\}$

- $\mathcal{L}_{OUT} = \{A | \mathcal{L}(A) = OUT\}$

- $\mathcal{L}_{UNDEC} = \{A | \mathcal{L}(A) = UNDEC\}$

2.1.3 Extension-based approach

Definition 2.8. *Given an Argumentation Framework AF, Let AF=(AR,R). label $\in \{IN, OUT, UNDEC\}$, a argument A is with label.*

- *An admissible labelling $\mathcal{L}$ is a labelling without arguments that are IN and OUT.*

- *A complete labelling $\mathcal{L}$ is an admissible labelling without arguments that are UNDEC*

If $\mathcal{L}$ is with a complete labelling. Then:

- *$\mathcal{L}$ is a grounded labelling iff there is no complete labelling $\mathcal{L}' \quad \subset \mathcal{L}$*

- *$\mathcal{L}$ is a preferred labelling iff there is no a complete labelling $\mathcal{L}' \quad$ such that $\mathcal{L} \subset \mathcal{L}'$*

- *$\mathcal{L}$ is a stable labelling iff $UNDEC(L) = \emptyset$*

11

2.1.4   Semantics

In the definition of Dung, we can find the definitions of these semantics: grounded semantics, preferred semantics, stable semantics, complete semantics. We will briefly introduce in this section.

Definition 2.9. *Let labelling ($\mathcal{L}$)is a total function that help $A \in AR$ of an Argumentation Framework AF=(AR,R).*

- *A labelled argument with IN is legally iff all its attackers are not labelled.*

- *A labelled argument with OUT is legally iff it has at least one attacker that is labelled IN.*

Naive semantics

Naive semantics (denoted as $\mathcal{NA}$) correspond to choosing as many arguments as possible, unless there is an attack between them. This is the sole criterion of avoiding collisions. Formally, it addresses the requirements of conflict-free and maximizing properties, and can be easily expressed in label-based and extension-based approach.

Complete Semantics

Full semantics (denoted as $\mathcal{CO}$) can be seen as an enhancement of the basic requirements enforced by the concept of admissibility. Permissive should be able to indicate the reason for the accepted and rejected arguments, but you are conflict-free from the arguments, but with full semantics, the other You should only abstain if there is no reason. That is, if you refrain from giving an opinion as to whether an argument is accepted or rejected, you reject the argument for reasons that are insufficient to accept the argument (that is, not all attackers are rejected). Need an inadequate reason for. (Meaning that no attacker is acceptable). In particular, keep in mind that while the trivial solution of leaving nothing undecided is always acceptable, it is not always perfect, as there may be good reasons not to give up.

Grounded semantics

If we consider each complete labeling (or complete extension) to be a reasonable position that can be taken in the presence of contradictory information

expressed in the framework of the discussion, the most possible question is the most "can be taken". Finding a "founded" position. In other words, it is the most suspicious position. The idea is to accept only those arguments that are inevitable to accept, reject only those arguments that are inevitable to reject, and abstain as much as possible. This results in the most skeptical (or least committed) semantics of full extension-based semantics: grounded semantics ($\mathcal{GR}$).

Preferred Semantics

While evidenced semantics takes a skeptical or minimal commitment perspective, you can also consider alternative perspectives aimed at accepting as much discussion as reasonably possible. This can lead to mutually exclusive acceptance alternatives. For example, a mutual attack can be reasonably resolved by accepting one of the contradictory arguments, but obviously not both (these alternatives are the non-skeptical solution in the example below). Is called). The idea of maximizing the accepted argument is represented by the Preferred Semantics ($\mathcal{PR}$) given in the definition below, which is explained in the labeling-based and extension-based approaches.

Stable semantics

So far, we've described semantics with the intuitive idea that you can accept, reject, or leave arguments undecided. However, there is no room for neutrality or shades of gray, and you may prefer a more committed rating that is all black or white. This simply means that undecided discussions are simply "banned", such as statements such as "Are you with us or are you against us?" This clear and powerful view supports stable semantics ($\mathcal{ST}$) and is directly formulated with both labeling-based and extension-based approaches.

Semi-Stable Semantics

As explained in the previous section, the requirement to "forbid" undetermined arguments turned out to have no consequences in some cases. A more sophisticated idea is to express a clear opinion on as many sets of discussions as possible, while limiting (but not necessarily avoiding) the discussions that are left undecided. It is in. This intuition is based on Semi-Stable Semantics ($\mathcal{SST}$).

ideal semantics

The concept of ideal semantics ($\mathscr{ID}$) is best explained using a description of the aggregate context of judgment [6]. Here, different people have different opinions about a set of arguments, and each opinion is expressed as a label. We need to create an aggregated opinion, that is, an aggregated labeling. In particular, ideal labeling / expansion arises from the following assumptions about the aggregation procedure:

- Each participant tries to accept as much discussion as possible. That is, the aggregated opinions correspond to the preferred set of labels / extensions.

- Each discussion is (provisionally) accepted or rejected only if all participants are in agreement. Otherwise, it is considered undecided.

- The resulting labeling / expansion may or may not correspond to a position that can be protected. That is, it may or may not be acceptable.

If not, refrain from tentatively accepted or rejected discussions in aggregated judgment until it is defensive.

2.2 Argument Structures

Argumentation, a research field that dating back to the phase of Aristotle, has played an important role in philosophy, forensic, and more. For Argument Model, it is not surprising that a variety of structures existing for arguments. Argument Models is a standpoint centring definition of the relationship between Argument Components. Depending on the structure of the argument, there are many kinds of expressions used to express *standpoints* such as conclusions. In our research, we will choose *claim* to represent the *standpoints*.

2.2.1 Components and Structure

Argument Components

An argument is considered to consist of three parts: the speaker's claim, and the data provided to *support* their claim, and possible counterclaims, rebuttals, and supplementary data. We can further define certain terms used for argument structure. An argument structure consists of a series of propositions related to the claim, which are the propositions the argument

is trying to build. In the micro-level models, other propositions are used as premises, used to *support* the claim, and sometimes to *attack* it. Any claim may serve as the final conclusion of an argument, and any claim may serve as a premise to support further reasoning.

Structures

Once the Components of the argument have been detected in our study, the method of Argument Mining was used to achieve it. And the next step is to examine their relational connections. Between propositions and claims, the relational distinction between propositions and propositions can be divided into relations of support and attack. Support and attack relationships between propositions sometimes exist only as support or attack relationships. Therefore, different structures can also be considered as different propositions, and the relationship between the propositions, the following are some simple examples:

- Convergent Arguments In Convergent Arguments, different premises are directly supporting the argument. The premises are independent, and there is no connection between the arguments.

- Linked Arguments In Linked Arguments, multiple premises together support the conclusion, each of which is indispensable and mutually supports the conclusion. It can be considered that it is a collection of premises to support the conclusion.

- Divergent Arguments In Divergent Arguments, a premise can support multiple conclusions. A common example is the dispute tree

- Sequential Arguments In Sequential Arguments, most of the premises can be ordered in order to support the conclusion. One of the premises leads to the other, and this process leads to the conclusion($p \rightarrow q, q \rightarrow r$,r is the conclusion).

Then you can see the diagramming of structures in Figure 2.2 [28].

2.2.2 Argument Models' Type

Argumentation schemes are patterns of inference, connecting a set of premises to a conclusion, that represent stereotypical patterns of human reasoning. Such schemes were originally viewed as rhetorical methods by which a speaker could influence their audience; later, they have also been adopted as a way to distinguish good arguments from bad. Different people have different schemes

A | A B | A B | A | A

B | C | C | B → C | C B

(i) Single   (ii) Linked   (iii) Convergent   (iv) Serial   (v) Divergent
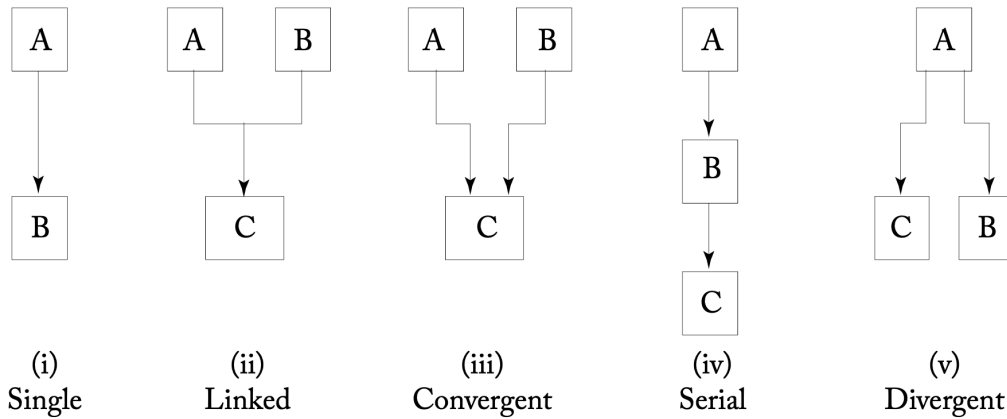
Figure 2.2: Common basic argument structures.

about what an argumentative model is : Vreeswijk [39] porposed Argumentation models abstract from the language level to a concept level that stresses the links between the different components of an argument or how arguments relate to each other ; Bentahar, Moulin, and Belanger [3] propose a classification of argumentation models that is divided into three categories: micro-level models(monological models), macro-level models(dialogical models), and rhetorical models.

- In a monological models: An argument is a tentative proof that describes the internal reasoning structure (microstructure) of each argument by linking premises and claims.

- In the rhetorical model: The argument is perceived by the audience, emphasizing the persuasive structure of the argument by linking the entire argument and rhetorical structure.

- In the dialogue model, arguments are viewed through defeasible reasoning, and by connecting the entire argument into a macrostructure (dialogue structure).

Above we briefly introduced the classification of types of argument models. Each of these models is a complete argument model for analyzing arguments from their own perspective. Next we will provide a concrete instance based on each model type. Dung's Argumentation Framework mentioned in the previous section looks at argument from the point of view of the macro-structural dialogue, it does not concern about the internal structure of the argument and the supporting relationships.

Model Type	Argument Evaluation Based On	What is Linked	How They Are Linked	Structure
Monological	Tentative proofs	Premises, claims	Internal inference structure	Microstructure
Rhetorical	Audience's perception	Whole arguments	Persuasion structure	Rhetorical structure
Dialogical	Defeasible reasoning	Whole arguments	Dialogical structure	Macrostructure

Figure 2.3: Bentahar, Moulin, and Bélanger's taxonomy of argumentation models, modified from Bentahar et al.

Toulmin's Model

As a basis for generating logically structured output, we need to use a structured framework for arguments, which is ideal for arguments. A promising option to achieve this involves integrating the Toulmin's model. One of the most influential models of argumentation, the Toulmin model, was first proposed by British philosopher Stephen Toulmin in 1958. This model takes a monological view of the argument, focusing on the internal structure of the argument. In Toulmin's opinion, he was the first to believe that most arguments can be modeled simply using the following six components: *claim*, *data*, *warrant*, *backing*, *qualifiers* and *rebuttals*.

In the model, according to some warrants, claims have irrefutable backing to their ground. Qualifiers or rebuttals may also be displayed but are not required. A Toulmin's model example of an argument rendered is illustrated in Figure. The lines and arrows represent relations between the components. Then you can see an Toulmin's Model example in Figure 2.4

In Figure 2.4, the claim is "Exposure to violent video games causes at least a temporary increase in aggression, and this exposure correlates with aggression in the real world." Its data is "Anderson and Bushman claim violent video games promote violent behavior, attitudes, and beliefs by desensitizing an individual to aggression". Its warrant is "If violent video games promote violent behavior, attitudes, and beliefs by desensitizing an individual to aggression. Then games cause at least a temporary increase in aggression, and this exposure correlates with aggression in the real world", which has the backing, "exposure to violent video games causes both short term and long term aggression in players and decreases empathy and pro-social behavior" Its qualifier is "at least". Its rebuttal is "video game violence is not related to serious aggressive behavior in real life."

Exposure to violent video games causes at least a temporary increase in aggression, and this exposure correlates with aggression in the real world.

Anderson and Bushman claim violent video games promote violent behavior, attitudes, and beliefs by desensitizing an individual to aggression.

at least

If violent video games promote violent behavior, attitudes, and beliefs by desensitizing an individual to aggression, then games cause at least a temporary increase in aggression, and this exposure correlates with aggression in the real world.

video game violence is not related to serious aggressive behavior in real life.

exposure to violent video games causes both short-term and long-term aggression in players and decreases empathy and prosocial behavior.

Figure 2.4: Sample Toulmin diagram based on Toulmin [2008].

The New Rhetoric

In the New Rhetoric [27]: In debates, it has persuasion and audiences as central concerns: in their view, arguments cannot be reduced to formally infeasible logic. Instead, arguments try to convince a specific audience, or to convince an audience in general. Arguing, therefore, is a persuasive act involving two or more interlocutors and "relative to the affected audience." From this perspective, "the goal of all arguments...is to create or increase adherence to the argument presented." We can describe it as an argument of rhetorical structure.

2.3   Existing Data-sets

Arguments are applied to many fields, and argument mining is an important branch of natural language processing. Whether it is machine learning technology or artificial intelligence technology, most of them are based on a large amount of data analysis, so a good corpus becomes particularly important. Moreover, with the rapid development of Internet technology in recent years, we can obtain a large number of natural language containing arguments in various fields. In this section we will do some brief descriptions about the corpus of some arguments

2.3.1   Annotation for argument

In the previous section, we have briefly introduced the scheme of argument structures that have been transformed in various ways into argument instances in annotated texts. Below, we present some schemes that have been used for corpus annotation.

Microtext scheme

Microtext scheme Following an early pilot study by [36], [26] proposed a tree-oriented annotation scheme that argues for the idea of a dialectical exchange of hypotheses between proponents. Presents and defends his claims, as well as opponents, who critically question ("attack") them in a controlled manner. Each step in this exchange corresponds to a structural element in the argument tree.
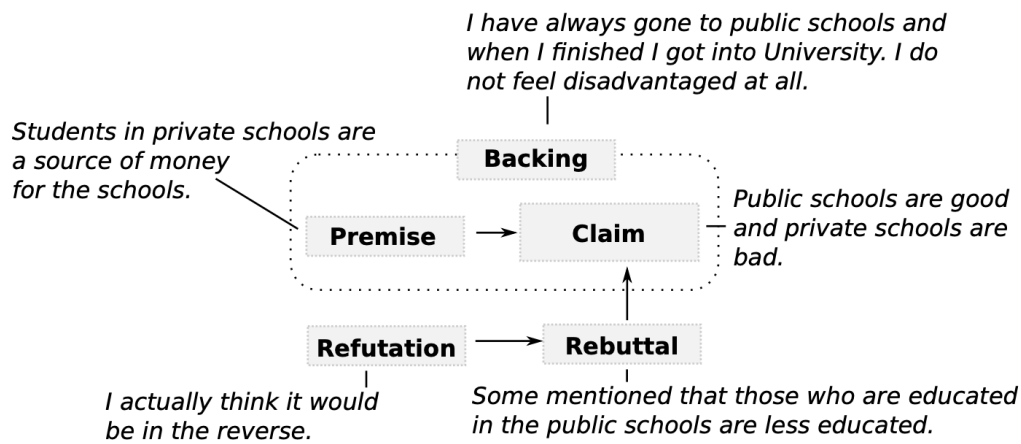
Figure 2.5: Diagrammatic representation of annotated argument in the Modified Toulmin Scheme based on Habernal and Gurevych

The Modified Toulmin Scheme

Habernal and Gurevych [13] set out to choose a scheme to annotate their user-generated web utterance corpora based on Toulmin's model. They retain their general premise that the argument role completely determines its position in the argument representation—thus, the scheme does not need to explicitly annotate any relationships between units. Therefore, the annotation is a flat dimension; an example is shown in the figure 2.5.

Habernal and Gurevych did make some significant modifications to the original Toulmin model:

- They removed Qualifier and Warrant as they did not play a role in their corpus data;

- redefines the concept of support and is now used as "additional evidence", which is distinct from the premise but still provides overall support for the argument.

- To explain counterattack, they introduced the unit type Refutation.

- Ambiguous statements were frequently found in their corpora, and annotators were instructed to add a detailed statement when they inferred the author's position from the text.

2.3.2   Examples in different domains

scientific articles

Preliminary work on manual identification guidelines for ten custom argumentation scenarios for genetics research articles is presented in [12], which highlights his aim to create a verifiable a freely available open access corpus of full-text scientific articles, annotated to support the challenges encountered in the idea of argument mining research.

In [14] work, much of the data used was unlabeled data from online debate portals. [2] also utilize an online debate portal to generate annotations by automatically mapping source data.

Essay

The Argument-Annotated Essay Corpus (AAEC), proposed by [34] and updated in [13], consists of argument-annotated persuasive essays with topic and position identification, argument-component annotations, and argument-relations. The final corpus comes from 402 English papers and contains 751 main claims, 1,506 claims, and 3,832 premises, connected by 3,613 supports and 219 attack relations.

Legal document

[41] in Legal Texts are developing a type system for marking successful and unsuccessful argument patterns in US judicial decisions. Another example of legal material is the ECHR corpus [21], a set of documents extracted from European Court of Human Rights (ECHR) legal texts. Although the content of the argument is not clearly stated. In ECHR material, standard reasoning and argument structures are included, so corpora can easily be used as data for argument mining.

Debate

The Internet Argument Corpus (IAC) [40] is a corpus for studying political debate in Internet forums. It contains about 11,000 discussions, 390,000 posts, and about 73,000,000 words. The data are annotated for topics, positions, agreements, sarcasm, and dirty. After that [1], there was a further extension to IAC, a corpus for studying political debate in Internet forums. It consists of three datasets: 4forums (414K posts), ConvinceMe (65K posts), and samples from CreateDebate (3K posts). It includes topic notes, response features (4forums), and positions.

Argument datasets based on languages such as debates, legal texts, etc., are structured argument data, and some unstructured argument data, such as Wikipedia, Twitter, or reviews of shopping sites. The IBM dataset used in this study is based on Wikipedia. We will briefly introduce it next.

Wikipedia

Extensive use of Wikipedia and other data in IBM Project Debater to create the first artificial intelligence system that can debate complex topics with humans. Debater can respond to a given topic by automatically constructing a set of related pro/con arguments expressed in natural language. For example, on the topic "The sale of violent video games to minors should be banned," an early prototype of Debater identified the 10 most relevant articles from scanning some 4 million Wikipedia articles, reviewing all 3,000 of those articles. Sentences, detect sentences containing candidate statements, assess their pro and con polarity, and then present three appropriate pro and con arguments [19], first address the challenge of detecting context-sensitive claims (CDCs) in Wikipedia articles, showing how, given a choice of topics and related articles, the option "directly supports or questions a given A general, concise statement "theme" can be found [30]. Follow these works. Extracting supporting evidence for a given CDC from Wikipedia data, introduces the stance classification task, which detects a target for a given CDC and determines a stance on that target. These data were then used for our research.

2.4 The gap to develop 2 tier-Argumentation Framework

Dung's Argumentation Framework only shows the relationship between the arguments in the macro structure. The description of the inside of the arguments is not varied enough, to exploit argument inter-pretability and readability. Arguments and content of arguments cannot be computed. No existing approaches can offer computational content of arguments in a form understandable by naive users.

Chapter 3

2 tier-Argumentation Framework

3.1 Overview

In this chapter, we will introduce the 2-tier Argumentation Framework(2-tier AF). It is a renewed instance of Dung's Abstract Argumentation Framework based on Toulmin's structure to provide an implementation of the argument reasoning approach with a knowledge base based on propositional logic. We expanded the approach of machine learning technology in the field of natural language processing, which can take arguments from natural language and can automatically present argument reasoning.

In this chapter, we will want to give a definition of the 2-tier AF and then introduce some related works.

3.2 Modified Toulmin's model

In this study, we simplify the elements in the Toulmin model, and we call the simplified model the Modified Toulmin Model. Modified Toulmin Model is a simplified argumentation structure that contains only two argument components: claim and evidence. The example of the original Toulmin Model mentioned in the previous chapter is shown in Figure 3.1 in the Modified Toulmin Model.

As shown in the figure 3.1.

- Claim: Exposure to violent video games causes at least a temporary increase in aggression, and this exposure correlates with aggression in the real world.
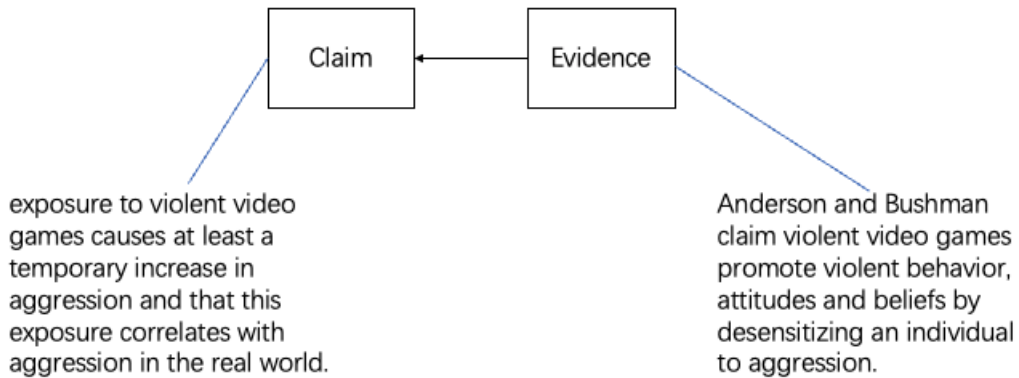
Figure 3.1: modified toulmin

- Evidence(Data): Anderson and Bushman claim violent video games promote violent behavior, attitudes, and beliefs by desensitizing an individual to aggression.

3.3 Concepts

Motivated by the gap mentioned in the last chapter investigated previously, this subsection presents a further refinement based on the Toulmin's structure, called a 2-tier argumentation framework (2-Tier AF). Our main goal is to combine Toulmin's structure and Dung's framework to make arguments can be computed so that lay people can understand the contents and relationship between arguments

Definition 3.1 (Modified Toulmin). *An Modified Toulmin's diagram(MT) is tuple*

$$MT = < \mathcal{C}, \mathcal{E}, \mathcal{C} \leftarrow \mathcal{E} >$$

in which Claim $C \in \mathcal{C}$, Evidence $E \in \mathcal{E}$ are argument components of the Modified Toulmin's diagram. And, $\mathcal{C} \leftarrow \mathcal{E}$ is a binary relationship, in which $\mathcal{E}$ is the basic condition for the existence of $\mathcal{C}$.

Definition 3.2. *The notion of support between claim and evidence in each argument: in each Modified Toulmin's diagram, the argument component is connected with the support relationship.*

Definition 3.3. *Each Modified Toulmin's diagram is an argument in the 2-Tier AF.*

Definition 3.4 (The contrary of argument). *Given a notion of the contrary of argument:*

Let $\mathcal{MT} = <\mathcal{C}, \mathcal{E}, \mathcal{C} \leftarrow \mathcal{E}>$, $A \in \mathcal{MT}$, $B \in \mathcal{MT}$. Then argument A attacks argument B iff the Claim of argument A is the contrary of Evidence of argument B and argument B is arguable.

Definition 3.5. *The notion of attack between arguments: the 2-Tier AF depends only on attacking ('undercutting') uncancelled hypotheses.*

Definition 3.6 (2-Tier argumentation framework). *A 2-Tier argumentation framework is quadruple in which:*

$$2\mathcal{TAF} = <\mathcal{C}, \mathcal{E}, \mathcal{C} \leftarrow \mathcal{E}, ->$$

Claim $C \in \mathcal{C}$, Evidence $E \in \mathcal{E}$ are argument components of the Toulmin's diagram.
$-$ is a partial mapping from the set $\mathcal{E}$ of Evidence into the set $\mathcal{C}$ of Claim. Where $\overline{E}$ is called the contrary of Evidence $E \in \mathcal{E}$.
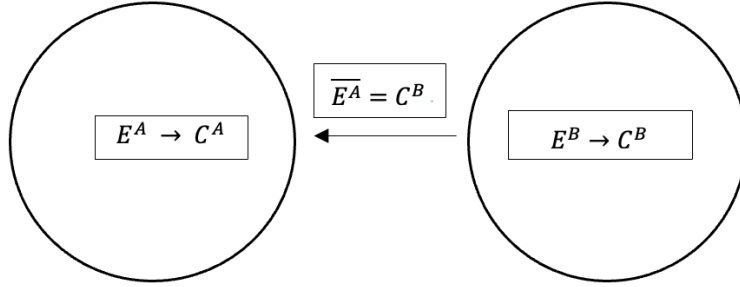
Figure 3.2: 2tieraf

Note that $\overline{E}$ is defined as a partial mapping due to the fact that not every argument can be attacked. An argument is open for attack if it involves uncancelled hypotheses for deriving the claim. Figure 3.2 illustrates an example of an argumentation framework instantiated by a 2-Tier AF, where an attack is formalized by the contrary of *Evidence E* of argument *A* and the contrary is derivable on argument *B*, i.e. $\overline{E_A} = C_B$.

Obviously, the introduced 2-Tier AF is an instance of the Dung's abstract argumentation Framework (AF). Thus, all semantic notions for determining 'acceptability' of arguments in AF are also applied to arguments in 2-Tier AF.

Definition 3.7. $2\mathcal{T}\mathcal{AF}$ *is the extension of AF built with* $2\mathcal{T}\mathcal{AF} = <\mathcal{C}, \mathcal{E}, \mathcal{C} \leftarrow \mathcal{E}, - >$. $Con_{2\mathcal{T}\mathcal{AF}}$ *is a set of consequences that built upon the Modified Toulmin structure MT.* $Con_{2\mathcal{T}\mathcal{AF}} = \{c_B, e_A | (c_A, e_A, c_A \leftarrow e_A), (c_B, e_B, c_B \leftarrow c_B) \in \mathcal{MT}, (c_2, e_1) \in 2\mathcal{T}\mathcal{AF}\}$

This connection does not only provide benefits on the interpretability of logic-based arguments computed but also gives potential for AI adoptions especially argument mining.

3.4 Related Works to AF

3.4.1 Assumption-Based Arguments

Assumption-Based Arguments (ABA) are a form of structured argumentation, and there have been versions over the years that have a slightly different form. Like most forms of structured arguments, its concepts of argument and attack relation are somewhat similar but different from those in abstract

arguments (AA). However, due to the flexibility of ABA, other forms of non-monotonic reasoning are allowed to use its tools without any advertising mechanism.

In this section, we will introduce some details of ABA structures.

Definition 3.8. *Given a deductive system $< \mathcal{L}, \mathcal{R} >$ in which*

$\mathcal{L}$ *is a logical language, and $\mathcal{R}$ is a set of inference rules on logical language. A set of assumptions $\mathcal{A} \subseteq \mathcal{L}$ is non-empty.*

Definition 3.9. *An ABA framework is a tuple $< \mathcal{L}, \mathcal{R}, \mathcal{A}, - >$ in which*

- *$< \mathcal{L}, \mathcal{R} >$ is a deductive system.*

- *A set of assumptions $\mathcal{A} \subseteq \mathcal{L}$ is non-empty.*

- *is a total mapping from $\mathcal{A}$ into $\mathcal{L}$, in which $\bar{a}$ is called the contrary of a.*

An argument claim $c \in \mathcal{L}$ supported by $S \subseteq \mathcal{L}$ is a tree labelled $\mathcal{L}$ with nodes(N). S is the set of all assumptions labelling the leaves. We say that a set of assumptions $\mathcal{A}sms \subseteq \mathcal{A}$ enables the construction of an argument A if A is supported by a subset of $\mathcal{A}sms$. We limited ourselves to the flat ABA framework, which means that is no assumptions are at the heart of inference rules.

Definition 3.10. *Let $< \mathcal{L}, \mathcal{R}, \mathcal{A}, - >$ is an ABA framework, and let the set of arguments $\mathcal{A}r$ can be constructed from the ABA framework, then $\mathcal{A}rgs \subseteq \mathcal{A}r$ is:*

- *a complete argument extension iff $\mathcal{A}rgs$ is conflict-free and $\mathcal{A}rgs = \{A \in Ar - \mathcal{A}rgs defends A\}$*

Note that ground argument expansion is unique, just as ideal argument expansion and eager argument expansion are unique. Each stable argument extension is a semi-stable argument extension, and each semi-stable argument extension is a preferred argument extension. Furthermore, every semi-stable argument extension is a stable extension if there is at least one stable argument extension. It also considers ground argument expansion to be a subset of ideal argument expansion, which is a subset of eager argument expansion.

27

There is a clear correspondence between AF and ABA approaches, that allows one to convert ABA-extensions to abstract argumentation extensions, and vice versa. To formalise this correspondence we first define a function Asms2Args that maps assumptions extensions to argument extensions and a function Args2Asms that maps argument extensions to assumptions extensions.

Definition 3.11. *Let $< \mathcal{L}, \mathcal{R}, \mathcal{A}, - >$ is an ABA framework, let $\mathcal{A}r$ be the set of all arguments that can be constructed using this ABA framework.*

- *Asms2Args: $2^{Ar} \to 2^{A}$ is a function then: Asms2Args(Asms) = $\{A \in Ar$ — A can be constructed based on $\mathcal{A}sms\}$*

- *Asms2Args: $2^{A} \to 2^{Ar}$ is a function then: Asms2Args(Args) = $\{a \in A$ — a is an assumption occurring in an $A \in \mathcal{A}rgs\}$.*

3.4.2 ASPIC+

This section presents the ASPIC+ framework for structured arguments. ASPIC+ aims to generate Abstract Argument Frameworks in the sense of Dung. While Dung's abstract calculus is an integral part of argument theory, it does not account for the structure of arguments or the nature of attack or failure, and thus provides no guidance for modeling practical argument problems. That is, how an argument's constituent reasons should be modeled, and the reasoning steps from those reasons to the argument's claims, how these structural properties of an argument determine attacks between arguments, and, in turn, how preferences should be used to determine whether an argument succeeds to attack (defeat) another argument? The ASPIC+ framework will attempt to answer this question.

ASPIC+ is explicitly designed as a mid-level abstraction between concrete logic and dung abstraction AF; the idea is that logic conforms to the ASPIC+ framework that defines the arguments and defeat of ASPIC+, which then constitute a dung framework.We already briefly review the main concepts from Dung's abstract argumentation theory in the chapter 2.

To use ASPIC, we need to provide the following information. We assume a logical language $\mathcal{L}$ composed of atoms n, closed under negation $\neg$ (a notion of conflict): $\mathcal{L} \to 2^{\mathcal{L}}$, denoting contraries of $x \in \mathcal{L}$ by $\neg x \subseteq \mathcal{L}$.

We must then provide two (possibly empty) sets of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules belong to one part of an ASPIC framework knowledge base $\mathcal{R} \subseteq \mathcal{L}$. If you provide a non-empty set of defeasible rules,

you then need to also specify which wellformed formulas in $\mathcal{L}$ correspond to (i.e. name) which defeasible rule in $\mathcal{R}_d$.

The above is summarised in the following formal definition:

Definition 3.12. *An argumentation system is a triple $AS = (\mathcal{L}, \mathcal{R}, n)$ in which*

- *$\mathcal{L}$ is a logical language closed under conflict ($\neg$).*

- *$\mathcal{R}$ is consisting of a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form $\varphi_1, ..., \varphi_n \to \varphi$ and $\varphi_1, ..., \varphi_n \to \varphi$, and $\mathcal{R}_s \bigcap \mathcal{R}_d =$.*

- *n is a partial function: $\mathcal{R}_d \to \mathcal{L}$ names defeasible relus.*

Definition 3.13. *A knowledge base in an $AS = (\mathcal{L}, \mathcal{R}, n)$ is a set $\mathcal{K}_n \cup \mathcal{K}_p = \mathcal{K} \subseteq \mathcal{L}$ consisting of two disjoint subsets $\mathcal{K}_n$ (the axioms) and $\mathcal{K}_p$ (the ordinary premises)*

The core definition of argumentation theory is based on the knowledge base and rules.

Definition 3.14. *An argumentation theory (AT) is a tuple $AT = (\mathcal{L}, \mathcal{R}, n, \mathcal{K})$, with a knowledge base $\mathcal{K} \subseteq \mathcal{L}$, rules $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ over $\mathcal{L}$, a contrary function $\neg : \mathcal{L} \to 2^{\mathcal{L}}$ and a partial function n: $\mathcal{R}_d \to \mathcal{L}$.*

We restrict each part of the AT to be limited. Arguments consists of part of AT.

Definition 3.15. *An argument A on the basis of an argumentation theory with a knowledge base $\mathcal{K}$ and an argumentation system $(\mathcal{L}, \mathcal{R}, n)$ is*

- *If $x \in \mathcal{K}$, then an argument $A = x$ with $Conc(A) = x$.*

- *If $A_1, ..., A_n$ are arguments, $x_i = Conc(A_i) \to \mathcal{R}$ ($1 \le i \le n$), then x is an argument with $Conc(A) = x$.*

Recall that ASPIC+ is meant to generate a set of arguments with a binary relation of defeat in Dung's styles. With the defined argument above, we define undercutting attack , the function n of an argumentation system is used, which assigns to elements of R_d, $n(r)(r \in R_d$) means that r is applicable. The conclusion $\overline{n(r)}$ undercut any arguments r.

Definition 3.16. *A attacks B iff A undercuts B, in which: A undercuts argument B(on $\acute{B}$) iff $Conc(A) = \overline{n(r)}$ for some $\acute{B} \in Sub(B)$ such that $\acute{B}$'s top rule r is defeasible.*

The attack relation tells us which arguments conflict with each other: if two arguments are conflict, then they cannot be justified. Users give a binary ordering $\leq$ on the set of all arguments that can be constructed based on an argumentation theory. Then, as usual, if $A \leq B$ and $B \nleq A$, B is strictly preferred to A (denoted $A < B$). Also, if $A \leq B$ and $B \leq A$.

Definition 3.17. *Rebuts undermines:*

- *Argument A successfully rebuts argument B if A rebuts B on $\acute{B}$ and either A contrary-rebuts $\acute{B}$ or $A \nprec \acute{B}$.*

- *Argument A successfully undermines argument B if A undermines B on $\emptyset$ and either A contrary-undermines B or $A \nprec \emptyset$.*

3.5   Related Works to AM

Argument Mining is an essential field of Natural Language Processing technology. In this section, we will briefly introduce what Argument Mining is, and some related works to the Argument Mining.

3.5.1   What is argument mining

Argument Mining is essential for a comprehensive range of technologies. In the previous chapter, we introduced a lot of the structure of the argument displayed in the graph, which contains the argument components and the relationship between them. This form of the graph can also be seen as the result of the Argument Mining. The content of Argument Mining task can be simply divided into the following parts:

1 Identify argumentative text.

2 Segment the argumentative text into units.

3 Identify the central claim.

4 Identify the role of units.

5 Identify relations between units.

6 Build a visual representation of the overall structure.

In some cases the type of text under consideration has already been defined as controversial. The first step can be omitted. Argument classification is required in cases where the type of text is unclear, and if the document can contain both argumentative and non-argumentative parts, it can be applied to the full text or to a partial text. Step 2 can be combined with some other tasks or performed as a separate program that specifies clauses, sentences, or their sequences as units of argument structure. A fundamental task of any argument mining scenario is step 3, which attempts to separate the central claim being argued. Likewise, Step 4 categorizes the role the unit plays in the argument, such as whether it is evidence or a possible objection. Step 5 Establish relationships between units ("support" and "attack").

Of course,step 6 there are also argument mining tasks that aim to build a full-fledged structured application that shows the relationship between arguments

The ability to automatically detect arguments, especially support in argumentation. So far, most argument analysis has relied on painstaking manual reconstruction and annotation, significantly limiting the scalability and speed of argument analysis. This limits the ability to build a system that eliminates the use of argument structures.

You can use (manual) argument analysis to validate certain arguments. Fallacy in discussions has been studied for thousands of years and has been done to determine under what conditions a discussion that normally falls into the category of fallacy is sound. Therefore, manual analysis allows you to check and validate or discard arguments. In particular, if the arguments are fully formalized and represented by logic, the arguments can be thought of as a messy where only the overall structure is calculated, not internally.

You can use the automatic inference feature to determine if the argument is appropriate. Given inconsistent information, the inferrer can also choose the most consistent subset. However, the applications on such systems are, of course, limited by the inputs they can use. Argument mining can fill this gap by finding an instance of the "real" argument in the text. Arguments play an important role in human meaning construction and decision making,

and the task of detecting and extracting arguments from written texts based on argument structures in argument mining can be supported by computers. Thus, online discussions can visually summarize the current state of the discussion by computer, without limiting individual responses or requiring humans, showing each person's point of view and how it relates to the overall discussion. Currently, human sense-making and decision-making are easily overwhelmed, as the work of understanding debate grows with each new contribution, limiting the potential for debate in large groups.

3.5.2 Argument Component Detection

Argumentative or Not?

Depending on the scenario, the task of classifying whether the text is argumentative or not is first performed before mining the argument components. In the document level, Earlier work by [17] and [16], on the Brown corpus, aimed to distinguish various categories. The researchers experimented with part-of-speech (PoS) tags, document and sentence lengths, punctuation, and the frequency of specific words that the authors considered relevant.

Later work replaced language analysis with bag-of-words [10] or bag-of-character-n-grams [31] models. This knowledge-free approach produces excellent results in the domain of training data, however, this idea is susceptible to changes in subject matter and discipline. Working with the New York Times Annotated Corpus (NYTAC) 1, Petrenz and Webber reimplemented an earlier approach, constructing subsets of articles covering different topics, and demonstrating earlier work using powerful linguistic features as the question of the article changes.

In principle, the attribute of being 'argumentative' can also be applied to paragraphs or other stretches of text. The task is commonly tackled on the level of sentences or clauses.

We start with some seminal work on argument mining by [24]. They use text from AraucariaDB [29] and court decisions of the European Court of Human Rights (ECHR) as their first analytical step to classify sentences as argumentative or non-argumentative. As features, they use lexical (token n-grams, adverbs, verbs, modality, argument tokens) and syntactic (punctuation marks, parse tree depth, number of clauses) features and some textual statistics such as the length of sentences and the number of clauses in the text . Using multinomial naive Bayes and maximum entropy models, they achieved 73% and 80% accuracy on the two datasets, respectively.

In[2]. propose a distantly supervised method for categorizing argumentative units: they automatically label text from idebate.org as either argumen-

tative or non-argumentative, depending on which predicate in the forum the text belongs to. Define components. On this data, they trained a classifier using n-grams, constituency grammar generation rules, and various morpho-syntactic features. They evaluated the model on the same type of data (obtaining an accuracy of 0.92) and on two other corpora: the persuasive paper corpus by Stab and Gurevych in 2014a and the web text corpus by Habernal and Gurevych in 2017. On these corpora, the accuracies are 0.67 and 0.88, respectively. Then, the authors also report the results of cross-domain training/testing experiments.

While the approaches discussed thus far performed binary classification (argumentative or not), it is also possible to solve the problem as a byproduct of running a multi-class classifier that detects the specific argumentative role of a sentence and allows for a 'none' value. An example of this technique is Stab and G[35], which is also trained and tested on student essays. Using a very rich set of features (structural, lexical, syntactic, and specific linguistic cues), they experimented with several classification techniques, leading to the best result of 77.3% accuracy (macro-F 0.73) obtained by an SVM. The F-score for the class 'none' was 0.88. It turned out that the text statistics and structural features are most helpful, which points to the role of conventionalized writing patterns in the genre of student essays.

Identifying Claim

One of the crucial tasks of any argument mining system is identifying claims. The other is identifying premises (or so-called evidence), which we will cover in the next section. But it has to be admitted that not all textual content is directly formulated to claim what it is, and the reader needs to judge for himself. There are many obscure expressions. For example, in social media comments, people often choose more ambiguous language to express their claims. The problem of inferring actual claims for the work of argumentative mining is undoubtedly complicated and has not received much attention. Instead, research at this stage focuses on the more straightforward task of identifying only explicit claims.

The task of classification requires that given a minimum unit of analysis (sentences are chosen in our study), it can be classified in different ways.

- Binary classification: claim or no claim.

- Binary classification: claim/premise.

- Multi-class classification: When there are a lot of argument components that need to be identified by a single classifier.

> { [**SUPPORT:** The Court recalls that the rule of exhaustion of domestic remedies referred to in Article x of the Convention art. x obliges those seeking to bring their case against the State before an international judicial or arbitral organ to use first the remedies provided by the national legal system.
> **CONCLUSION:** Consequently, States are dispensed from answering before an international body for their acts before they have had an opportunity to put matters right through their own legal systems.]
>
> [**SUPPORT:** The Court considers that, even if it were accepted that the applicant made no complaint to the public prosecutor of ill-treatment in police custody, the injuries he had sustained must have been clearly visible during their meeting.
> **AGAINST:** However, the prosecutor chose to make no enquiry as to the nature, extent and cause of these injuries, despite the fact that in Turkish law he was under a duty to investigate see paragraph above.
> **SUPPORT:** It must be recalled that this omission on the part of the prosecutor took place after Mr Aksoy had been detained in police custody for at least fourteen days without access to legal or medical assistance or support.
> **SUPPORT:** During this time he had sustained severe injuries requiring hospital treatment see paragraph above.
> **CONCLUSION:** These circumstances alone would have given him cause to feel vulnerable, powerless and apprehensive of the representatives of the State.]
> **CONCLUSION:** The Court therefore concludes that there existed special circumstances which absolved the applicant from his obligation to exhaust domestic remedies. }

Figure 3.3: Annotated text from an ECHR decision based on Palau and Moens.

Identifying Evidence

Detecting claims is the core task of argument mining, but in order to find a complete argument, there is at least one more component that needs to be identified: the statement to be supported by the author's presentation of the claim. In argumentation literature, this is often referred to as the "premise" of an argument, and in some cases it is the "evidence" or "reason". In addition to the claims and evidence, it may also be important to look for counterarguments, in case the text also mentions a "contrary point" and the application tries to identify it.

Support or evidence, by definition, is evidence for something and therefore stands about a claim. As long as there is only a single claim in our text, there is no need to establish and represent this relation explicitly. Instead, it can just be left implicit. In these cases, sentence classification or sequence labeling, as described in the previous chapter for claim detection, is also sufficient for identifying evidence.

Next, let's learn some examples of argument mining experiments. I will introduce them according to the domains of the database.

Legal documents

A prominent example of contention in legal texts is when a court justifies its decision in a decision document. As shown in figure 3.3 "Conclusion" in their terminology corresponds to what we call "claims." The simple form "court concluded that X" or some interpretation could be considered typical for the central findings.

The SVM model for premise/conclusion classification of [25] takes input sentences predicted as controversial. The resulting premises and conclusions

> "You can't always do what you want to do!," my mother said. She scolded me for doing what I thought was best for me. It is very difficult to do something that I do not want to do. **<Thesis>**But now that I am mature enough to take responsibility for my actions, I understand that many times in our lives we have to do what we should do. However, making important decisions, like determining your goal for the future, should be something that you want to do and enjoy doing.**<Thesis>**
>
> I've seen many successful people who are doctors, artists, teachers, designers, etc. In my opinion they were considered successful people because they were able to find what they enjoy doing and worked hard for it. It is easy to determine that he/she is successful, not becaue it's what others think, but because he/she have succeed in what he/she wanted to do.
>
> In Korea, where I grew up, many parents seem to push their children into being doctors, lawyers, engineer, etc. Parents believe that their kids should become what they believe is right for them, but most kids have their own choice and often doesn't choose the same career as their parent's. I've seen a doctor who wasn't happy at all with her job because she thought that becoming doctor is what she should do. That person later had to switch her job to what she really wanted to do since she was a little girl, which was teaching.
>
> **<Conclusion>** Parents might know what's best for their own children in daily base; but deciding a long term goal for them should be one's own decision of what he/she likes to do and want to do. **<Conclusion>**

Figure 3.4: Annotated essay from Burstein and Marcu , used with permission.

have F-scores of 0.68 and 0.74, respectively. Their features include syntactic features, subject type, primary verb tense, domain-specific cues, token counts, sentence position, and contextual features with predicted previous and following segments. With proposed an SVM sequence kernel classifier as an alternative to the type of feature engineering done by Palau and Moens. The kernel compares subsequences of sentences, where a word is labeled with its root form and PoS label. These authors use the Araucaria DB dataset, which contains arguments from newspapers, advertisements, and several other types in addition to arguments from legal documents. The task is to mark sentences as containing a premise, including a statement, being part of a premise and a statement, or not containing. They used 10-fold cross-validation. The overall accuracy of 0.65 is reported, which these authors consider promising as it does not require complex functions. However, for claims, the results are meager (around 0.3).

Students essays

One of NLP targets is a paper written by a student in response to a given prompt, primarily to apply automatic grading or to support human graders or peer reviewers by automatically adding useful marks.

The figure 3.4 is an essay annotated example[5]. Their annotation rules state that both components can be one or more sentences, but no subsentences. The output of early thread-based utterance parsers [3] was mapped

to two sentence features: the core state and the coherent relationship of sentence participation (according to rhetorical structure theory. No functional utility tests reported. A key question for a vocabulary-based approach like this is how to transfer to new content domains, namely new paper prompts. The method achieves an average F-score of 0.54 for the paper and 0.8 for the conclusion section when evaluating cues not present in the training set.

Much research in recent years started with corpora of persuasive essays, and the detection of argumentative components was implemented as a four-way classifier (the main claim, claim, premise, non-argument) proposed by Stab and Gurevych. Among various classifiers, SVM performs the best. The F-scores are: Main Claim 0.63, Claim 0.54, Premise 0.83, Non-Argument 0.88. While premise can be identified very reliably using only one set of features (F between 0.65 for syntactic features and 0.78 for structural features), the corresponding values for declarations and major declarations are much lower, with a maximum of 0.48 (major declarations) and structural features of 0.42 (claims).

WIKIPEDIA

[2] and [19] take different perspectives on claim detection. These researchers work on the IBM Debater project, whose overall goal is to scour the web for pro and con arguments on a given controversial topic. Working on a slightly bigger version of the IBM dataset, [20] started by observing that common 'rhetorical structures often characterize argumentative sentences' and operationalized that intuition as the similarity between syntactic (constituency) parse trees. Their goal thus is to account for argumentative language irrespective of the domain and do so based on syntactic patterns. They obtained the Stanford CoreNLP suite parses and trained an SVM classifier that exploits a partial tree kernel to compute similarities. The authors conclude that even without considering topic similarity. When adding a single context feature, the F-measure increases by a further 1.2 points.

Now I am going to present an example of where the IBM dataset was used,[18] was used in experiments with RNNs and CNNs to detect claims and evidence. Their goal is to establish the first deep learning baselines for these tasks and report a comparison of various architectural variants. They used the Context Dependent Claim Detection dataset[19] to train the Claim Sentence(Claim detection) task, and the Claim detection task in our experiment uses the same data as the test set. They used other additional metrics like P@200, R@200, F1@200, P@50, R@50 and F1@50 [19] in addition to reporting AUC and Average Precision.

Task	P@200	R@200	F1@200	P@50	R@50	F1@50	P@20	P@10	P@5	AVGP	AUC
Claim Sentence Task	9.64	61.5	15.8	17.1	27.7	19.2	22.4	27.9	28.5	0.173	0.812
EXPERT Evidence Task	9.53	64.0	14.5	14.5	35.0	15.7	18.6	21.1	22.5	0.160	0.750
STUDY Evidence Task	8.33	79.5	13.5	15.5	53.9	18.9	20.8	25.3	31.8	0.298	0.836

Figure 3.5: The results from the Laha and Raykar's experiment

Chapter 4

System Design

4.1 Overview

We define our work as an automated argument reasoning system based on the two-tier Argumentation Framework(2-tier AF).

The definition of the 2 tier-AF system, takes into account automatically indicates each component of Toulmin's argument from text, as figure 4.1 shown, will include two parts:

- Machine Learning Framework

- 2-tier Argumentation Framework

The structure of our system is:

- Argument Component: – Claims and evidence base on the Modified Toulmin Model from texts.
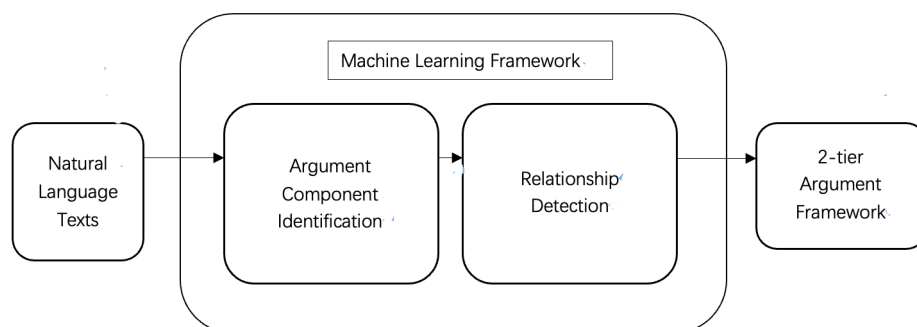
Figure 4.1: Image of the 2-tier Argumentation Framework system

- Support Relationship: – The relationship connects Claim and evidence in each Modified Toulmin Diagram.

- Undercut Relationship: –The notion of attack between arguments: the 2-Tier AF depends only on undercutting.

4.2   ML experience

Since the purpose of our work is to develop a system for automatically reasoning arguments, we employ argument mining techniques in the field of natural language processing to obtain relationships between argument machines; we break this task into four main tasks in this problem defined as:

- Claims detection

- Evidence detection

- Identifying the relationship of a claim and a evidence(undercut or not?).

- Identifying the relationship of a claim and a evidence(support or not?).

By combining these tasks, we will be able to tell Which claims in the long text are supported by the evidence and which arguments have an undercut relationship.

4.2.1   Data

In terms of data, we mainly selected three different data sets (Claim Stance Dataset(2017), Evidence Sentence(2018), Claim and Evidence(2014)) from IBM data sets for Claims and Evidence detection tasks. Due to the lack of relation on undercut in the IBM dataset, we annotate this aspect based on the content of the claims and evidence(2014) data set, which we will introduce next.

Claim and Evidence(2014)

The teat sets I used the for the Claims and Evidence from IBM Debater project. It contains 1,392 labeled claims for 33 different topics, and 1,291 labeled evidence for 350 distinct claims in 12 different topics. Each article we are given the following points:

- Topic: A short phrase that frames the discussion.

- Article: The title of the Wikipedia article

- Claim:A general, typically concise statement that directly supports or contests the topic.

- CDE: A text segment that directly supports a claim in the context of the topic

Claim Stance Dataset(2017)

The train sets for the Claim detection from IBM Debater project Claim Stance Dataset (2017). Manually identified and annotated claims from Wikipedia; it contains 2,394 labeled claims for 55 topics pulled from 1,065 Wikipedia articles. For each article, we are given the following data points:

- Topic Target (text): the sale of violent video games to minors

- Full text from Wikipedia (text): 44,000 word plain text Wikipedia article

- Claim Text (text): They increase the violent tendencies among youth

- Stance: (Pro or Con)

- Claim Start Index (integer): 8119

- Claim End Index (integer):8167

Evidence Sentence(2018)

The train sets for the Evidence detection from IBM Debater project Evidence Sentence (2018), automatically retrieved Wikipedia sentences. It contains 118 topics for 5,785 annotated evidence pairs.

- Topic: The debating topic serving as context for the sentence candidate.

- The concept of the topic: The main concept featured in the topic, as obtained through wikipedia.

- Candidate: The sentence candidate.

- Candidate masked

- label (1 for evidence , 0 for non-evidence)

- Wikipedia article name: The title of the Wikipedia article

- Wikipedia URL

These three datasets are used in two argument mining tasks, claim identification and evidence identification. It has to be noted that the 2014 dataset is used as the training set for the two tasks.

Undercut Relationship Dataset

The Undercut dataset's annotation the text content based on the Claim and Evidence(2014) from the IBMDebater Project, which contains 1040 claims-evidence relationships, of which 714 pairs are connected with Undercut relationships.

Labeling this dataset mainly relies on human annotation, the specific labeling method mainly goes through the following three steps:

1 . Dividing the Claims data sets into two parts as the Requires correction.

2 . Finding the corresponding claims and evidence.

3 . Based on the topic, finding the Evidence of being attacked by Claims.

Then, we can get a pair of claims and evidence based on the same topic and have Undercut relationship through the above steps. A concrete example with an undercut relationship in the dataset is

Topic:	the sale of violent video games to minors
Claim:	exposure to violent video games causes at least a temporary increase in aggression and that this exposure correlates with aggression in the real world.
Evidence:	John Sherry has said (2007) "Further, why do some researchers (e.g. Gentile and Anderson, 2003) continue to argue that video games are dangerous despite evidence to the contrary.
Undercut Relationship:	True

For this example, we can say that regarding the topic, "the sale of violent video games to minors," it is true that there is a relationship undercut between "exposure to violent video games causes at least a temporary increase

in aggression and that this exposure correlates with aggression in the real world." and "John Sherry has said (2007) "Further, why do some researchers (e.g. Gentile and Anderson, 2003) continue to argue that video games are dangerous despite evidence to the contrary." Among them, the direction of undercut is from claims to evidence.

Support Relationship Dataset

The Supports Relationships dataset combines Claim and Evidence 2014 and Claim and Evidence 2015. And we propose to define the relationship between the Claim and the corresponding Context Dependent Evidence(CDE) as a supporting relationship. The dataset contains 2170 pairs of claims and evidence, of which 1636 pairs are in Support Relationship. This is an example of the Support Relationship Dataset.

Topic:	the sale of violent video games to minors
Claim:	exposure to violent video games causes at least a temporary increase in aggression, and this exposure correlates with aggression in the real world.
Evidence:	The most recent large scale meta-analysis: examining 130 studies with over 130,000 subjects worldwide– concluded that exposure to violent video games causes both short term and long term aggression in players.
Support Relationship:	True

In our study, these two datasets are used in the task of the Relationship Identification.

4.2.2 Methodology

We divided this research problem into two parts, Argument Component Detection and Relationship Identification.

Claims and Evidence Detection

The table 4.1 represents the number of Argument Components in the dataset. To balance the data set,we randomly upsampled the train sets, with replacement to 10000 observations, and also randomly sampled 3500 observations from the test set, by using re-sample function.

Table 4.1: The number of Argument Components in the Datasets

claim	count	evidence	count
label=1(train)	2223	label=1(train)	1499
label=0(train)	88273	label=0(train)	2566
label=1(test)	1232	label=1(test)	404
label=0(test)	46514	label=0(test)	17370

We experimented with a simple fully connected Feedforward Neural Network model(FNN) and an LSTM model.

- Feedforward Neural Networks: The simplest and most widely used neural network models, where information is processed in only one direction. They have many different names, such as Multi-layer Perceptron(MLP).His model structure is simple, with only three parts: input layer, hidden layer, and output layer. Although the data may pass through multiple hidden nodes, it always moves in one direction and never moves backward.

- LSTM: Long Short-Term Memory Networks - often referred to as LSTM, are a special kind of RNN(Recurrent Neural Network) that can learn long-term dependencies and are often used for sentence classification. Proposed by Hochreiter and Schmidhuber (1997) and improved and extended by many others in subsequent work. LSTMs have performed exceptionally well on a wide variety of problems and are now widely used.

Relationship Identification

Relationship classification focuses on the problem of determining if given a claim and a evidence, can we figure out if the claim is for or undercut/support the evidence?

$$Undercut(Claim, Evidence) = True, False$$
$$Support(Claim, Evidence) = True, False$$

We used the mean embedding vector of each sentence from labeled data sets described in the previous section and trained the model using Logistic Regression and Random Forests.

- Logistic regression is a machine learning algorithm that is often used for classification tasks, such as identifying spam. It is a probability-based analysis algorithm.

```
supported(X) :- claim(X), evidence(Y), inferredBy(X,Y), \+ undercut(Y,_).
undercut(X,Y) :- contrary(X,Y), supported(Y).
```

Figure 4.2: Implementing 2-tier Argumentation Framework with Prolog

- Random Forest is a machine learning technique. Ho first proposed in 1995 to construct Random Forest by Decision Tree algorithm to solve regression and classification problems.

The above is the method used in the Machine Learning part, and the content of the 2-tier Argumentation Framework will be introduced in the next section.

4.3 Demo of 2-tier AF

We capture all the elements that make up a 2-tier AF from natural language through the Machine Learning Framework described in the previous section: claims, evidence, supporting relations, attacking relations. In this section, we show how the results are realized, compute the argumentative relationship between them through the 2-tier Argumentation Framework.

The specific method we used was Prolog to implement our 2-tier Argumentation Framework. Prolog is a logic programming language applied to the computational linguistics field. Prolog is a language based on logical design, which contains logical calculation relationships. Therefore, we can use the logical relationship contained in the prolog, which will make the implementation. Figure 4.2 show the code of prolog.

4.3.1 The sample 2-tier Argumentation Framework

Next, we can feed the predictions from the machine learning framework as facts into this 2-tier Argumentation Framework. All facts entered into 2-tier AF will be represented as index numbers. To better understand this framework, the following is a simple example.

The elements in the sample 2-tier Argumentation Framework

We obtained four claims, three evidences, and connections between them through a machine learning framework. In the table 4.2 The claim and the evidence are the results predicted by the model; Inferredby refers to the relationship between the claim and the evidence, and the evidence can infer the claim; contrary refers to the opposition between the claim and the evidence.

44

Table 4.2: The facts in the sample 2-tier Argumentation Framework

Claim :	C1, C2, C3, C4
Evidence:	E1, E2, E3
Inferred:	(C1, E1), (C2, E2), (C3. E1), (C4, E3)
Contrary:	(E1, C2), (E3, C3)
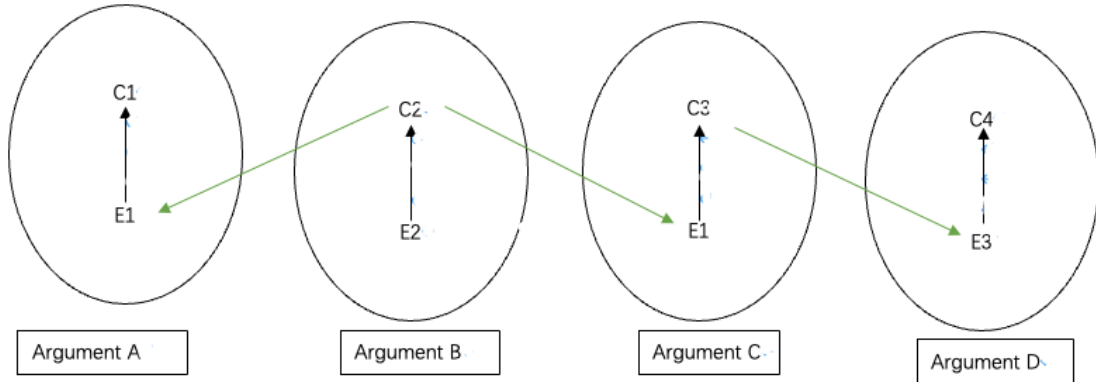
Figure 4.3: An example of the 2-tier Argumentation Frameowork

Outputs from the sample 2-tier Argumentation Framework

As shown in the figure, the black arrows represent the inferred relationship, and the green arrows are the opposite relationship. Argument A contains elements (C1, E1); Argument B contains elements (C2, E2); Argument C contains (C3, E1); Argument D contains (C4, E3).

We entered the facts from the resulting Table 4.2 into 2tier AF implemented with Prolog and got the following results:

1. Claim C2 and Claim C4 is Supported.
2. There is only one Undercut Relationship between Evidence E1 and Calim C2.

E1 exists in Argument A and Argument B; C2 exists in Argument B, so we can say Argument A is attacked by Argument B and Argument C is attacked by Argument B. Although there is a contrary relationship between C3 and E3, but the claim C3 is not supported, so in the four Arguments of A, B, C, and D, there is only Argument A, and Argument C is attacked(Undercut) by Argument B. The above is a simple example of 2-tier Argumentation Framework that shows how we can use 2-tier AF to compute the undercutting relationship between Arguments, and the Supporting relationship between

the claims and the evidences of the Arguments.

Through the two frameworks introduced in this chapter: the machine learning framework and the two-tier argument framework, we can achieve the purpose of computing arguments in natural language. In the end, we hope that we can better show the relationship between sentences like lay human through the simplified Toulmin model, which will be described in Chapter 5.

Chapter 5

Evaluation

The Evaluation of our work will be divided into two aspects, and we will use traditional methods for evaluating machine learning and present our final results by building a server.

5.1 The result of ML framework

In the following, we summarize the results of the various learners described above.

5.1.1 Claims and Evidence Detection Model

We use grid search for each Argument Component Detection model to find the optimized combination of the number of hidden layers, layer width, and activation function based on model accuracy on the test set.

Claim Detection

The optimized FNN model has three layers(two hidden layers and one output layer) of width 512 and tanh activation function, and the optimized LSTM model is with one four layers(one LSTM + two hidden layers + one output layer) of width 1024 and identity activation function.

Performance of the model of the Claim Detection task is summarized in table 5.1. It turns out that both models achieve almost the same highest accuracy for this task, since the LSTM model is better for processing longer texts, the accuracy is slightly higher

Figure 5.1 show the Confusion Matrix of two Claim Detection model.

Table 5.1: Claims Detection

score	Accuracy	F1 score	Precision	Recall
Feed NN	0.73400	0.7646612	0.6856301	0.8642857
LSTM	0.7440000	0.765445	0.7062802	0.8354286
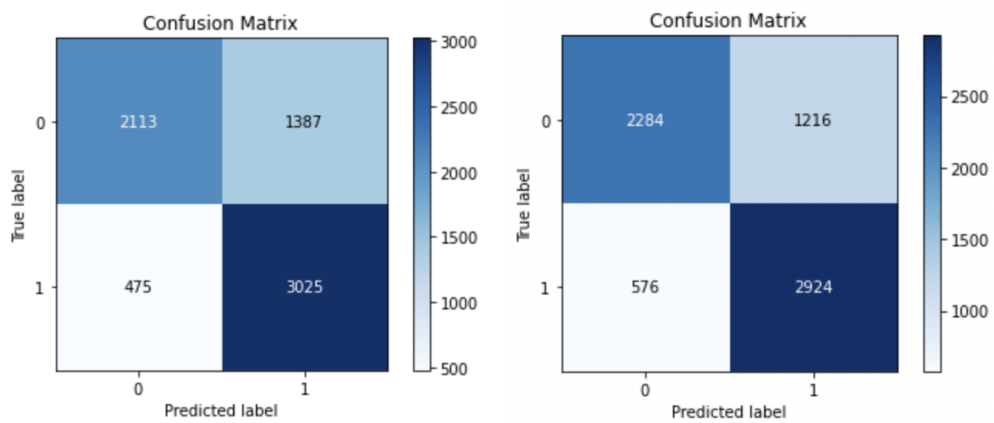
(a) Claim Detection in FNN model   (b) Claim Detection in LSTM Model

Figure 5.1: Claim Detection
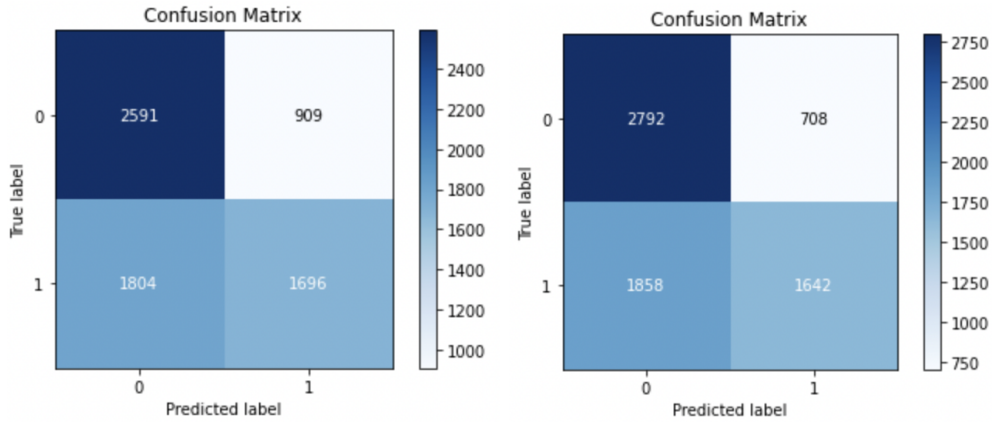
(a) Evidence Detection in FNN model   (b) Evidence Detection in LSTM Model

Figure 5.2: Evidence Detection

Table 5.2: Evidence Detection

score	Accuracy	F1 score	Precision	Recall
Feed NN(E)	0.612429	0.55561	0.65106	0.48457
LSTM(E)	0.63343	0.56137	0.698723	0.469142

Evidence Detection

The optimized FNN model is with four layers(three hidden layers and one output layer) of width 256 and tanh activation function, and the optimized LSTM model is with one four layers(one LSTM + two hidden layers + one output layer) of width 128 and relu activation function.Performance of the model of the Claim Detection task is summarized in table 5.2. Figure 5.2 show the Confusion Matrix of two Evidence Detection model.

5.1.2   Relationship Identification

We used both random forest and logistic regression models to train this Relationship Identification task. We used the *GridSearchCV* method through 10-fold cross-validation to find the optimal model parameters.

In detail, each parameter of the model has many candidate values. We do a 10-fold cross-validation for each parameter combination, and finally get the highest score, which is our optimal parameter

(a) Support for Logistic regression (b) Support for Random Forests
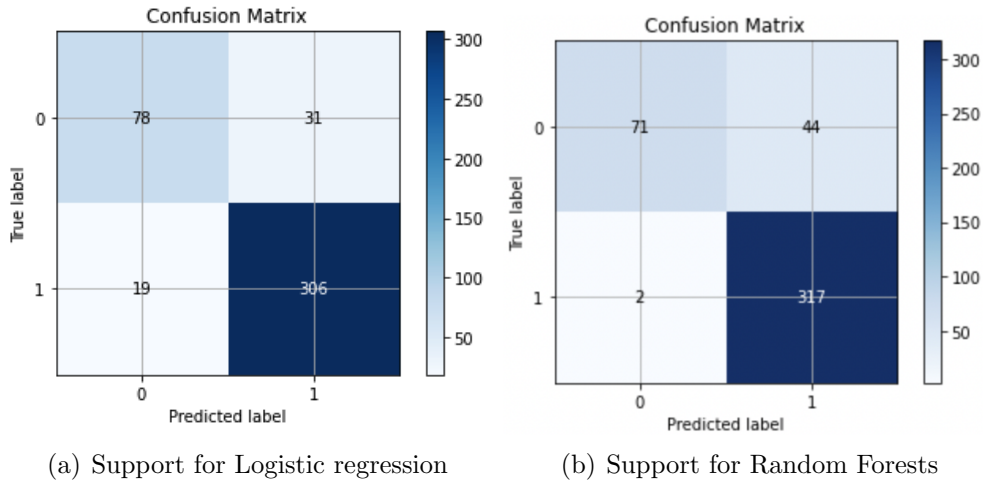
Figure 5.3: Support Relationship Identification

Table 5.3: Support Relationship Identification

score	Accuracy	F1 score	Precision	Recall
Random Forest(s)	0.89401	0.93235	0.87812	0.99373
Logistic(s)	0.76497	0.85593	0.787012	0.938080

Support Relationship Identification

In the Support Relationship Identification task, We adjust its parameters to be able to fit and predict better.

For the Random Forest model, n-estimators is the number of trees to be used in the forest. Since the Random Forest is an ensemble method comprising of multiple decision trees, this parameter is used to control the number of trees used in the process. The max-features, on the other hand, determines the maximum number of features to consider while looking for a split. The optimized Hyper-parameter values in Random Forest model are: n-estimators is 81, max-features is 9.

For the Logistic Regression model, the parameter max-iter is the maximum number of iterations for the solver to converge. The parameter C is the inverse of the regularization strength in logistic regression. The optimized Hyper-parameter values in the Logistic Regression model are: max-iter is 10, C is 80. Figure 5.3 show the Confusion Matrix of two Support Relationship Identification models. Performance of the model of Support Relationship Identification task is summarized in table 5.3

Table 5.4: Undercut Relationship Identification

score	Accuracy	F1 score	Precision	Recall
Random Forest(U)	0.87981	0.91582	0.88312	0.95104
Logistic(U)	0.884615	0.919463	0.913333	0.925676
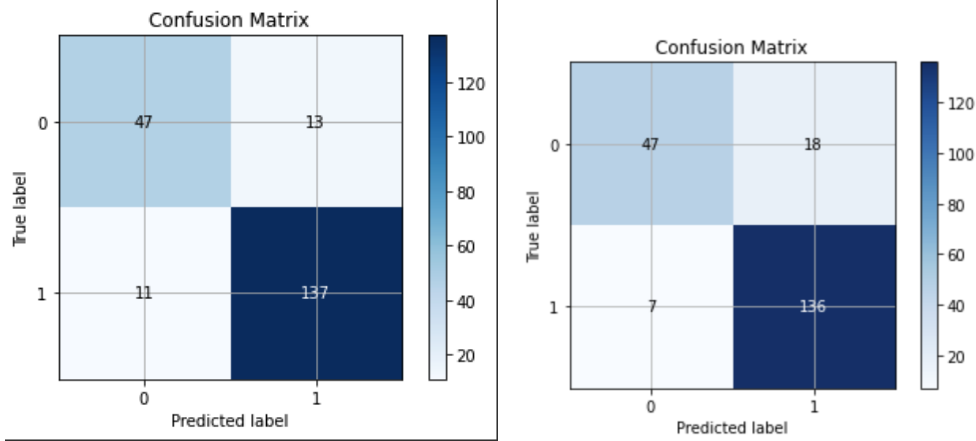
(a) Undercut Relationship Identification with Logistic regression model

(b) Undercut Relationship Identification with Random Forests model

Figure 5.4: Undercut Relationship Identification

Undercut Relationship Identification

In the Undercut Relationship Identification task:

- The optimized Hyper-parameter values in Random Forest model are: n-estimators is 61, max-features is 10.

- The optimized Hyper-parameter values in the Logistic Regression model are: max-iter is 100, C is 10.

Figure 5.4 show the Confusion Matrix of two Undercut Relationship Identification models. Performance of the model of the Undercut Relationship Identification task is summarized in table 5.4

5.1.3 Examples of bad machine learning results

Next, we need to input the prediction results of the above trained machine learning model into the 2-tier AF system:

- By default, we select the data where the prediction results and the real results are true.

- We trained two models for each task, we will select the results of the models with higher scores.

Next, we briefly discuss these bad, inconsistent predictions:

Claim Detection: There were 1773 mis-predictions in this task, of which 1215 sentences were misidentified as claims. Here is an example:

c1	Even in 1971, a Time Magazine essayist wondered why there were 375 major foreign military bases around the world with 3,000 lesser military facilities and concluded "there is no question that the U.S.
c2	Students that attend year round schooling may miss out on experiences such as summer camps.

Where c1 was mistaken for a claim, we assume that the presence of "concluded" in the sentence is what makes the model misjudgment, and c2 was a claim but was not detected.

Evidence Detection: There were 1114 mis-predictions in this task, of which 907 sentences were misidentified as evidences. Here is an example:

e1	The ethicist Margaret Somerville,[REF] suggested that Dawkins "overstates the case against religion",[REF] particularly its role in human conflict
e2	According to the Los Angeles Times, many babies put up for adoption had not been abandoned by their parents but confiscated by family planning officials.

Where e1 was mistaken for a evidence, and e2 was a evidence but was not detected. The evidence components in the evidence data set used in this experiment are not sufficient, which will greatly affect the prediction results of the model.

Support Identification: where 46 relationships were wrongly judged, only two of them were wrong because the supporting relationship was not identified, and the rest were misjudged as supporting relationships:

Claim	It creates dependence to the state.
Evidence	Some British Conservatives, such as Conservative Party co-chair Sayeeda Warsi, have also criticized the welfare state's "gain for nothing" culture, claiming that the high level of the welfare state "prevents the unemployed from finding work"

It is possible to judge semantically that there is a supporting relationship between the claim and the evidence by human, however, the lack of similarity between the two sentences, we assume this is the cause of the misjudgment.

Undercut Identification: where 25 relationships were wrongly judged, only 7 of them were wrong because the supporting relationship was not identified, and the rest were misjudged.

Claim	no long-term relationship between playing violent video game and youth violence or bullying.
Evidence	On April 20, 1999, 18-year-old Eric Harris and 17-year-old Dylan Klebold killed 12 students and a teacher in the Columbine High School massacre. The two were allegedly obsessed with the video game Doom.

As far as semantic judgment is concerned, there is a u relationship between these two sentences, but it is not predicted.

5.2 The result of 2-tier Argumentation Framework

In order to better display the results of 2-tier Argumentation Framework, we built a simple server to display the results of the experiment.

5.2.1 Knowledge base description

Through the machine learning framework system mentioned in the previous section, we get the components that need to be put into 2-tier AF. The knowledge base is stored as four excel files, as shown in the table5.5. The serial number of each proposition finally entered into 2-tier AF begins with a lowercase letter followed by a number. The letter is defined as follows:

- c stand for Claims.

- e stand for Evidence.

It is worth noting that because the results of all the constituent components are taken from the results predicted by the model (Prediction = 1, Truth=

Claim	It creates dependence to the state.
Evidence	Some British Conservatives, such as Conservative Party co-chair Sayeeda Warsi, have also criticized the welfare state's "gain for nothing" culture, claiming that the high level of the welfare state "prevents the unemployed from finding work"

It is possible to judge semantically that there is a supporting relationship between the claim and the evidence by human, however, the lack of similarity between the two sentences, we assume this is the cause of the misjudgment.

Undercut Identification: where 25 relationships were wrongly judged, only 7 of them were wrong because the supporting relationship was not identified, and the rest were misjudged.

Claim	no long-term relationship between playing violent video game and youth violence or bullying.
Evidence	On April 20, 1999, 18-year-old Eric Harris and 17-year-old Dylan Klebold killed 12 students and a teacher in the Columbine High School massacre. The two were allegedly obsessed with the video game Doom.

As far as semantic judgment is concerned, there is a u relationship between these two sentences, but it is not predicted.

5.2 The result of 2-tier Argumentation Framework

In order to better display the results of 2-tier Argumentation Framework, we built a simple server to display the results of the experiment.

5.2.1 Knowledge base description

Through the machine learning framework system mentioned in the previous section, we get the components that need to be put into 2-tier AF. The knowledge base is stored as four excel files, as shown in the table5.5. The serial number of each proposition finally entered into 2-tier AF begins with a lowercase letter followed by a number. The letter is defined as follows:

- c stand for Claims.

- e stand for Evidence.

It is worth noting that because the results of all the constituent components are taken from the results predicted by the model (Prediction = 1, Truth=

Table 5.5: Components to the 2-tier AF

Component	Claim	Evidence	Support	Undercut
Quantity	994	196	322	137

1), and because our implementation of the Prolog is also correct, we judge that the results of 2-tier AF are also accurcy.

5.2.2 2-tier AF serve

The figure shows a schematic diagram of a 2-tier Argumentation Framework server, which can implement the functions of the 2-tier Argumentation Framework system based on the knowledge base.

According to our previous definition of a 2-tier AF system, this server can achieve the following two functions:

- In the first-tier Argumentation Framework, in each Modified Toulmin's argument, when we choose a pair of claim and evidence that have an inferred relationship to each other, we can judge whether the claim in the argument is supported by the relevant evidence.

- In the second-tier, the relationship between the arguments of two Toulmin's structures is represented: whether there is a relationship of Undercut. After we have obtained two arguments that satisfy the Toulmin's structure through the first tier, according to the calculation of the 2-tier AF system, we can judge which of the arguments they belong to based on a pair of evidence and claims (Evidence, Claim) that have a contrary relationship. Whether there is an Undercut relationship between or not.

Through the above steps, the results of the arguments for the existence of Toulmin's structure computed using 2-tier AF are presented in a straightforward approach.
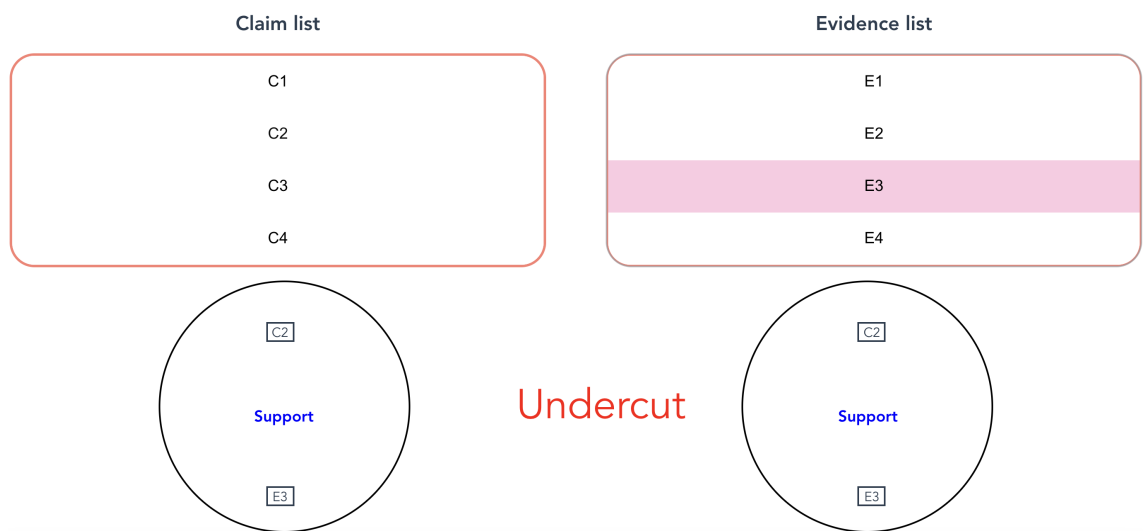
Figure 5.5: 2-tier Argumentation Framework serve

Chapter 6

Conclusion

6.1 Conclusion

We demonstrate the proposed 2-tier Argumentation Framework (2-tier AF) System as an instance of the well-known Dung's Abstract Argumentation Framework, which presented the method of making the Modified Toulmin's arguments computable. We develop Machine Learning Frameworks to automatically indicate each component of the Modified Toulmin's argument from text for the realization of automatic argument reasoning.

6.1.1 Machine Learning Framework

Given a Knowledge Base:

1) Obtain the required Claim and Evidence components from the Knowledge Base through two task models: Claim Detection and Evidence Detection.

2) Use the Relationship Identification task model to determine whether there is an undercutting, supporting relationship between the Claim and Evidence components obtained from 1).

From the machine learning framework we get all the required components and relationships.

6.1.2 2-tier Argumentation Framework

3) We compute the Modified Toulmin's arguments by feeding all the resulting components and relations into a 2-tier Argumentation Framework implemented with the Prolog.

4) We visualized the results of 2-tier AF so that the lay human can better understand the arguments through our 2-tier Argumentation Framework System.

We used Wikipedia articles to demonstrate the analytical computing performance of our framework system for arguments. Through visualization, we not only clearly observe the various arguments in the article, but also compute the relationship between the various arguments.

6.2　Future Works

The future direction of our system has two aspects: theoretical aspect, application aspect:

6.2.1　Theoretical aspect

- In this study, We use the Modified Toulmin's model, which contains only claim and evidence components. The actual Toulmin's model also contains the warrant component, which will be added in the future research. In addition, the warrant component can exist in the form of multiple propositions with an inference relationship.

- In this study, the structure of the 2-tier AF we experiment is mainly based on the attack relationship between the two arguments. Given that our framework is an instance of Dung's framework, the semantics of being satisfied with Dung's framework are equally satisfied with our proposal. From this we can expand the number of connected arguments to three or four or even more, and the relationship between arguments is not limited to attacking this one.

- In the 2-tier AF claims and evidence presented this time, we assume that the evidence supports the claim 100%, but there are cases where the evidence cannot fully support the definition of the claim. For Example:

  Claim: Schools should ban the sale of carbonated drinks for the sake of students' health. Evidence: Studies have shown that drinking carbonated beverages for a long time is harmful to human health.In this case, the evidence limits long-term consumption of carbonated beverages, but the claims do not. That is to say, the evidence can support the claim to a certain extent. For this case, we will mention in future research.

6.2.2 Application aspect

- The machine learning framework designed in this study is four separate task models, and future research will link the model of element recognition with the model of relation Identification.

- The data annotated in this experiment are all performed by one person, and the number of data annotators (at least three) will be increased in future research to improve the quality of the annotated data.

- The texts used in this experiment are all from the field of online text Wikipedia. There are limitations in the visualization results. We can apply the proposed 2-tier AF to the other available datasets [13] annotated in Toulmin's method.

- In order to better verify the results of the model, we will test with real users.

We hope that in the near future, with the implementation of the above aspects, a fully automatic argument reasoning system can be built. It can automatically reason and summarize the structural information in the article while the user is browsing the text, helping the user to better understand the text.

Appendix A

Use case

All textual data for our experiments are sourced from IBM Project Debater Debater Datasets and can be found on https://research.ibm.com/haifa/dept/vst/debating_data.shtml.

Regarding knowledge base annotated by the experiments mentioned in Chapter 4, the code that implements the 2-tier Argumentation Framework and Argument Mining tasks can be found on https://github.com/carol1521/2-tier-Argumentation-Framework.git.

The server's Knowledge base in Appendix B is about the server for 2-tier Argumentation Framework visualization, the URL of the server is http://localhost:8080/. And the relevant code for the server can be found on https://github.com/carol1521/2-tier-Argumentation-Framework.git.

Appendix B

Serve's knowledge base

Table B.1: The claims knowledge base to serve

index_claim	claim
c226	violent games cause youth violence
c100	a large increase in population would bring, "certain poverty on the citizenry
c101	Everyone has the right to seek and to enjoy in other countries asylum from persecution
c124	Atheism has been criticized as a faith in itself
c130	atheism is a superior basis for ethics
c142	It is the only sport where the intention is to inflict serious injury on your opponent
c154	The selection process should not be based on some arbitrary or irrelevant criterion
c160	Substance abuse can be harmful to your health and may even be deadly in certain scenarios
c18	Americans have an individual right described in the Second Amendment to possess firearms
c182	no one has a legal right to have any demographic characteristic they possess be considered a favorable point on their behalf
c196	redistribution of legitimately obtained property cannot ever be just
c199	Control over reproduction is a basic need and a basic right for all women
c20	exposure to violent video games causes at least a temporary increase in aggression and that this exposure correlates with aggression in the real world
c208	religion as a human invention used to frighten people into following moral order

c214	no connection between exposure to media violence and real life violence
c215	it leads to increased political corruption, compulsive gambling and higher crime rates
c22	immigrants are thought to compete with employees who are already in the country
c221	a creator of a universe with such complexity would have to be complex and improbable
c234	the likelihood that a death will result is significantly increased when either the victim or the attacker has a firearm
c238	Earth's surviving biodiversity provides resources for increasing the range of food and other products suitable for human use
c246	all people should be treated similarly
c246	jobs should go to those "most qualified
c247	content-based restrictions on games are unconstitutional
c247	parents, not government bureaucrats, have the right to decide what is appropriate for their children
c269	Frequently, these countries of asylum are some of the world's poorest nations and cannot handle the large influx of persons
c281	Gambling is now a diverse, vibrant and innovative industry
c284	not all depictions of violence are even bad to witness
c298	more guns can reduce crime
c299	the majority of illegal guns in Mexico really come from the United States
c305	it creates "a cult of victimization
c31	parents should make the decision" about what video games they purchase for their children, and what constitutes "too violent
c31	states do not have the right to decide that some video games are too violent for [minors
c310	religious belief is a delusion
c311	video game publishers unethically train children in the use of weapons and, more importantly, harden them emotionally to the act of murder
c313	human beings are the ultimate resource
c326	Higher levels of economic inequality tend to intensify social hierarchies and generally degrade the quality of social relations
c327	Health risks can be produced by long-term use or excessive doses of anabolic steroids
c331	Historical racism continues to be reflected in socio-economic inequality
c333	media influences are too weak and distant to have much influence
c340	a certain amount of redistribution would be justified
c344	American guns are arming the Mexican drug cartels
c356	violent video games can increase children's aggression

c36	no long-term relationship between playing violent video game and youth violence or bullying
c370	poverty, once started, is likely to continue unless there is outside intervention
c374	higher population density leads to more specialization and technological innovation, which in turn leads to a higher standard of living
c383	most of the financing for the Mexican traffickers comes from American drug consumers
c390	there was no direct link between violent video games and their influence on children
c397	the high-speed instant gratification of Internet games and the high level of privacy they offer may exacerbate problem and pathological gambling
c40	race and sex shouldn't matter when getting a job
c411	atheism systematically influences people to do bad things
c427	around 10,000 former athletes bear the physical and mental scars of years of drug abuse
c429	to ensure that minority groups within a society are included in all programs
c434	It may increase racial tension and benefit the more privileged people within minority groups at the expense of the least fortunate within majority groups
c437	everyone can exercise the right to seek asylum and find safe refuge in another state
c439	the "freedom of association and the effective recognition of the right to collective bargaining" as an essential right of workers
c445	American) government has been sending weapons to Mexico in a premeditated and systematic manner, knowing that their destinations were Mexican criminal organizations
c449	many skills can be learned from the gaming experience, it builds practical and intellectual skills
c472	Many former athletes suffer from health problems related to steroid consumption
c473	the people have a right to "keep and bear arms" as a protection from the government
c474	God's existence can be demonstrated
c488	gun possession is a civil right
c497	societal media consumption and violent crime rates are not well associated
c509	God cannot exist with, or would want to prevent, all evils
c510	violence in video games is not causally linked with aggressive tendencies

c520	the groups are considered hostile or alien to the natural culture
c522	Natural selection and similar scientific theories are superior to a "God hypothesis"—the illusion of intelligent design—in explaining the living world and the cosmos
c525	social and political issues surrounding the issue of immigration
c54	the existence of the Universe requires an explanation, and the creation of the Universe by a First Cause, generally assumed to be God, is that explanation
c542	violent games affect students positively and not negatively
c549	any equalities achieved will entail future inequalities
c550	One area of controversy surrounding casinos is their relationship to crime rates
c559	gun possession is a fundamental civil right
c575	affirmative action has undesirable side-effects in addition to failing to achieve its goals
c575	affirmative action is "reverse discrimination
c575	affirmative action has undesirable side-effects
c577	economists doubt that a correlation between population reduction and economic growth exists
c58	it is often contested on constitutional grounds
c582	gambling is a type of regressive tax on the individuals
c59	countering the effects of a history of discrimination
c590	Pathological gambling is a common disorder that is associated with social costs
c598	Physical exercise is important for maintaining physical fitness
c601	Physical inactivity is increasing or high among many groups in the population
c603	breeds dependence on government aid
c603	welfare empowers individuals
c604	God is improbable
c611	leads to adverse consequences for the gambler, others, or for the community
c616	religion does more harm than good
c620	sustains or even creates poverty
c623	everyone has the right to leave any country, including his own, and to return to his country
c623	the freedom of movement both within and between countries is a basic human right
c631	Racial prejudice and discrimination no longer exists

c632	Rapid fertility reduction associated with the one-child policy has potentially negative results
c639	Reinforcement of sexist stereotypes has also been claimed as an effect of violent video games
c64	efforts to sanction Burma were useless
c643	religion is a positive civilizing influence on society
c648	Parents have a basic right to decide freely and responsibly on the number and spacing of their children
c656	violent video games may increase mild forms of aggressive behavior in children and young adults
c676	God and evil are logically incompatible
c678	it is economically beneficial for a state to allow and support gambling institutions
c686	video games have many healthy and positive aspects
c690	the welfare state has produced a generation of dependents who, instead of working, rely solely upon the state for income
c701	affirmative action requires the very discrimination it is seeking to eliminate
c702	it fails to achieve its goals
c705	the existences of such a god and of evil are logically incompatible
c710	a correlation between children playing violent video games and suffering psychological effects
c713	it creates dependence to the state
c714	are logically contradictory
c714	the existence of some or all gods is logically impossible
c72	no evidence linked video games to youth violence
c720	foreign aid is seen to be serving the interests of the donor more than the recipient
c722	Sports related to combat skills have been a part of human culture for thousands of years
c727	the overall idea is to give children from less fortunate backgrounds more of a chance
c743	merit" itself should be a primary consideration during evaluation
c75	freedom of movement is often recognized as a civil right
c751	all phenomena could be understood as resulting from purely natural causes
c764	something caused the Universe to exist, and this First Cause must be God
c768	God's existence "can be known with certainty from the created world by the natural light of human reason

c769	internet gambling as a legitimate activity that citizens have the right to engage in
c77	exposure alone does not cause a child to commit crimes
c771	gambling in some form or another has been seen in almost every society in history
c773	immigrants can "swamp" a local population
c777	the identification of oppressed classes was difficult to carry out
c78	evidence for harmful effects were inconclusive
c781	there was a "first cause", or "prime mover" who is identified as God
c807	The impetus towards affirmative action is to redress the disadvantages[REF][REF][REF][REF][REF] associated with overt historical discrimination
c809	legal gambling provides significant government revenue
c812	it helps to compensate for past discrimination, persecution or exploitation by the ruling class of a culture
c822	exposure to violent video games causes both short term and long term aggression in players and decreases empathy and prosocial behavior
c830	The one-child policy is challenged in principle and in practice for violating a human right to determine the size of one's own family
c833	The policy is controversial both within and outside China because of the manner in which the policy has been implemented, and because of concerns about negative social consequences
c835	inequality is harmful for economic development
c838	The possibility of getting shot by an armed victim is a substantial deterrent to crime
c840	neurological link between playing violent video games and aggressive behaviour in children and teenagers
c846	a high degree of relationship between violent games and youth violence
c853	Parents have a basic human right to determine freely and responsibly the number and the spacing of their children
c857	humanity would be better off without religion or belief in God
c863	correlations between violent gameplay and some common childhood problems
c87	exposure to violent video games results in increased physiological arousal, aggression-related thoughts and feelings as well as decreased prosocial behavior
c879	the social and economic consequences of casino gambling outweigh the initial revenue that may be generated
c902	there is no convincing evidence that prove that media violence cause violent crime or any type of real life violence

c92	violent video games promote violent behavior, attitudes and beliefs by desensitizing an individual to aggression
c926	Excessive exposure to violent video games and other violent media has been linked to aggressive behaviour
c929	children may imitate aggressive behaviors witnessed in media
c93	video game violence is not related to serious aggressive behavior in real life
c936	money goes directly towards stimulating the economy
c937	members of minorities require specific provisions and rights to ensure that they are not marginalised within society
c940	no evidence violent games are psychologically harmful to minors
c945	gun control laws are effective in reducing gun-related accidents and crime
c951	affirmative action hurts its intended beneficiaries
c967	there is little danger from anabolica, as they call it, when the athletes are kept on strictly monitored programmes
c976	reduction of the population is a key to economic growth
c978	a 'significant' percentage of their firearms originate from gun stores and other sources in the U.S
c981	poverty and famine are caused by bad government and bad economic policies, not by overpopulation
c988	application for asylum could be denied regardless of the legitimacy of their claim
c991	there was a connection between video games and violence
c992	it is impossible to favor somebody without discriminating against others

Table B.2: The evidences knowledge base to serve

index_evidence	evidence
e10	John Roemer who used the term nondiscrimination principle to mean that "all individuals who possess the attributes relevant for the performance of the duties of the position in question be included in the pool of eligible candidates, and that an individual's possible occupancy of the position be judged only with respect to those relevant attributes."[REF
e101	Incidents such as the Columbine High School massacre in 1999 have heightened concerns of a potential connection between video games and violent actions [REF
e107	Justice Thomas, in his dissent, considered that historically, the Founding Fathers "believed parents to have complete authority over their minor children and expected parents to direct the development of those children

e108   Karl Marx and Friedrich Engels, influenced by the work of Feuerbach, argued that belief in God and religion are social functions, used by those in power to oppress the working class

e111   Legal representatives of asylum seekers, such as the UNHCR and the Committee for Civic Assistance, complain that individuals with legitimate claims are being denied asylum without just cause – and where a showing of persecution upon return to their home country is clear and apparent

e128   Opponents of racial affirmative action argue that the program actually benefits middle- and upper-class African Americans and Hispanic Americans at the expense of lower class European Americans and Asian Americans

e129   government agencies, including the Government Accountability Office and the National Drug Intelligence Center, have estimated that Mexico's cartels earn upwards of $23 billion per year in illicit drug revenue from the United States [REF

e13   A study published in the American Journal of Public Health estimated that: "over 886,000 deaths could have been prevented from 1991 to 2000 if African Americans had received the same care as whites

e131   Paik and Comstock note that when aggression toward another person, and particularly actual violent crime is considered, the relationship between media violence and these outcomes is near zero

e133   according to Article 13 of the Universal Declaration of Human Rights, everyone has the right to leave or enter a country, along with movement within it

e135   Research at University of Rochester suggest that playing video games that contain high levels of action can improve eyesight [REF

e137   Research has shown that most weapons and arms trafficked into Mexico are from gun dealers in the United States [REF

e14   Academic studies have attempted to find a connection between violent video games and the rate of violence and crimes from those that play them; some have stated a connection exists,[REF][REF][REF

e142   Several studies show a correlation between violent content conveyed through media (including videogames) and violent or aggressive behavior

e145   Simon also claimed that if you considered a list of countries ranked in order by population density, there is no correlation between population density and poverty and starvation

e145	Simon also claimed that if you considered a list of countries ranked in order by population density, there is no correlation between population density and poverty and starvation, and instead, if you considered a list of countries ranked in order by corruption within their respective governments, there is a significant correlation between government corruption and poverty and famine [REF
e146	Since 1996, the ATF has traced more than 62,000 firearms smuggled into Mexico from the United States [REF
e148	Some Brazilian Universities (State and Federal) have created systems of preferred admissions (quotas) for racial minorities (blacks and native Brazilians), the poor and people with disabilities
e149	Conservative Party co-chairman Sayeeda Warsi, also criticise the "'something for nothing' culture" of the welfare state, claiming that the high extent of the welfare state "discourages the unemployed from finding jobs" [REF
e151	A meta-analysis by psychologist Jonathan Freedman, who reviewed over 200 published studies and found that the "vast and overwhelming majority" did not find a causal link, also reached this conclusion [REF
e153	Some opponents of affirmative action view the greater access by women and minority groups to be at the expense of groups considered dominant (typically white men
e157	Some scientific studies show that the degree of religiosity is generally found to be associated with higher ethical attitudes[REF][REF][REF][REF] — for example, surveys suggesting a positive connection between faith and altruism [REF
e161	Sowell writes that affirmative action policies encourage non-preferred groups to designate themselves as members of preferred groups (i.e., primary beneficiaries of affirmative action) to take advantage of group preference policies; that they tend to benefit primarily the most fortunate among the preferred group (e.g., upper and middle class blacks), often to the detriment of the least fortunate among the non-preferred groups (e.g., poor whites or Asians); that they reduce the incentives of both the preferred and non-preferred to perform at their best — the former because doing so is unnecessary and the latter because it can prove futile — thereby resulting in net losses for society as a whole; and that they increase animosity toward preferred groups
e162	Stanford computer scientist John McCarthy states, "We also have no need for [the hypothesis of God], because science has been successful, and science is the best approach to solving the mysteries that remain."[REF

e165	Studies have shown that exercising in middle age leads to better physical ability later in life [REF
e166	Studies show[REF][REF] that in welfare states poverty decreases after countries adapt welfare programs
e167	Surveys by Gallup, the National Opinion Research Center and the Pew Organization conclude that spiritually committed people are twice as likely to report being "very happy" than the least religiously committed people [REF
e170	The Chinese government estimated that it had three to four hundred million fewer people in 2008 with the one-child policy, than it would have had otherwise [REF
e170	The Chinese government estimated that it had three to four hundred million fewer people in 2008 with the one-child policy, than it would have had otherwise [REF
e175	The Employment Equity Act and the Broad Based Black Economic Empowerment Act aim to promote and achieve equality in the workplace (in South Africa termed "equity"), by advancing people from designated groups
e179	The Labour Party passed the Sex Discrimination (Election Candidates) Act 2002, allowing them to use all-women shortlists to select more women as election candidates
e180	The most recent large scale meta-anlysis– examining 130 studies with over 130,000 subjects worldwide– concluded that exposure to violent video games causes both short term and long term aggression in players and decreases empathy and prosocial behavior
e185	Proclamation of Teheran was the first international document to recognize one of these rights when it stated that: "Parents have a basic human right to determine freely and responsibly the number and the spacing of their children."[REF][REF
e186	The Sankhya- tattva-kaumudi, commenting on Karika 57, argues that a perfect God can have no need to create a world
e187	According to John Hills, children of wealthy and well-connected parents usually have a decisive advantage over other types of children, and he notes that "advantage and disadvantage reinforce themselves over the life cycle, and often on to the next generation" so that successful parents pass along their wealth and education to succeeding generations, making it difficult for others to climb up a social ladder [REF
e19	According to Stefan Zillich, quotas should be "a possibility" to help working class children who did not do well in school gain access to a Gymnasium (University-preparatory school

e192	social justice requires redistribution of high incomes and large concentrations of wealth in a way that spreads it more widely, in order to "recognise the contribution made by all sections of the community to building the nation's wealth." (Patrick Diamond and Anthony Giddens, 27 June 2005, New Statesman)[REF
e193	USC Professor Henry Jenkins, for instance, suggested in his speech to congress that The Basketball Diaries utilizes violence in a form of social commentary that provides clear social benefit [REF
e20	According to the article 13 on the Universal Declaration of Human Rights, fundamental human rights are violated when citizens are forbidden to leave their country
e21	Account Services lawyer Jeff Ifrah said that the government "has never seized an account that belongs to players who are engaged in what [Ifrah] would contend is a lawful act of playing peer-to-peer poker online."[REF
e22	Affirmative action in the United States began as a tool to address the persisting inequalities for African Americans in the 1960s
e23	After conducting a two-year study of more than 1,200 Middle School children about their attitudes towards video games, Harvard Medical School researchers Lawrence Kutner and Cheryl Olson found that playing video games did not have a particularly negative effect on the researched group [REF
e26	Although some scholars had claimed media violence may increase aggression,[REF] this view is coming increasingly in doubt both in the scholarly community[REF] and was rejected by the US Supreme Court in the Brown v EMA case
e27	a review of 498 studies published in peer-reviewed journals concluded that a large majority of them showed a positive correlation between religious commitment and higher levels of perceived well-being and self-esteem and lower levels of hypertension, depression, and clinical delinquency [REF].[REF
e27	a meta-analysis of 34 recent studies published between 1990 and 2001 also found that religiosity has a salutary relationship with psychological adjustment, being related to less psychological distress, more life satisfaction, and better self-actualization [REF
e27	Studies by Keith Ward show that overall religion is a positive contributor to mental health,[REF
e28	An article from the World Health Organization calls safe, legal abortion a "fundamental right of women, irrespective of where they live

e29	according to Hans Breiter, MD, co-director of the motivation and Emotion Neuroscience Centre at the Massachusetts General Hospital, "Monetary reward in a gambling-like experiment produces brain activation very similar to that observed in a cocaine addict receiving an infusion of cocaine."[REF][REF
e30	Anderson and Bushman claim violent video games promote violent behavior, attitudes and beliefs by desensitizing an individual to aggression
e31	Another 2001 meta-analyses using similar methods[REF] and a more recent 2009 study focusing specifically on serious aggressive behavior[REF] concluded that video game violence is not related to serious aggressive behavior in real life
e31	a more recent 2009 study focusing specifically on serious aggressive behavior[REF] concluded that video game violence is not related to serious aggressive behavior in real life
e33	Article 23 of the Universal Declaration of Human Rights identifies the ability to organize trade unions as a fundamental human right [REF
e35	research supporting the view that video game violence leads to youth violence has been produced
e36	the formerly privileged white minority was compelled by law to employ previously disenfranchised groups (blacks, Indians, and Coloureds), collectively referred to as "blacks
e37	The basic argument against empirical theism dates back at least to David Hume, whose objection can be popularly stated as "Who designed the designer
e39	Barker also suggests 'that we must look beyond a specific film to think about the specific context in which it has been consumed, and the wider social background of the people'.,[REF
e41	Beale also cites studies that conclude that religiousity correlates with better mental health and less likelihood of suicide [REF
e44	In 2007, one study of amateur boxers showed that protective headgear did not prevent brain damage,[REF] and another found that amateur boxers faced a high risk of brain damage [REF
e45	By 2001, the estimated number of people who had participated in online gambling rose to 8 million and growth would continue
e45	By 2001, the estimated number of people who had participated in online gambling rose to 8 million and growth would continue
e48	China now has an increasingly aging population; it is projected that 11.8% of the population in 2020 will be 65 years of age and older

e5	A 2006 study of two pairs of identical twins, in which one twin used anabolic steroids and the other did not, found that in both cases the steroid-using twin exhibited high levels of aggressiveness, hostility, anxiety, and paranoid ideation not found in the "control" twin [REF
e56	Richard Dawkins argue that religious belief often involves delusional behavior [REF
e6	A 2010 Australian hospital study found that 17% of suicidal patients admitted to the Alfred Hospital's emergency department were problem gamblers [REF
e62	Drug cartels in Mexico control approximately 70% of the foreign narcotics that flow into the United States [REF
e63	Empirical evidence suggests that taxes and transfers considerably reduce poverty in most countries whose welfare states commonly constitute at least a fifth of GDP
e71	a recent long-term outcome study of youth found no long-term relationship between watching violent television and youth violence or bullying [REF
e71	a recent long-term outcome study of youth found no long-term relationship between watching violent television and youth violence or bullying [REF
e74	From the 1840s to 1920 German Americans were distrusted because of their separatist social structure, their love of beer, their German-language schools, their attachment to their native tongue over English, and their neutrality in World War I
e75	studies seem to show positive links in the relationship between religiosity and moral behavior[REF][REF][REF] — for example, surveys suggesting a positive connection between faith and altruism [REF
e77	One study did find an increase in reports of bullying, noting, "Our research found that certain patterns of video game play were much more likely to be associated with these types of behavioral problems than with major violent crime such as school shootings [REF
e77	One study did find an increase in reports of bullying, noting, "Our research found that certain patterns of video game play were much more likely to be associated with these types of behavioral problems than with major violent crime such as school shootings [REF
e77	several major studies by groups such as The Harvard Medical School Center for Mental Health, The Journal of Adolescent Health, and The British Medical Journal have shown no conclusive link between video game usage and violent activity."[REF][REF][REF

e79	In 1976, a group of Italian-American professors at City University of New York asked to be added as an affirmative action category for promotion and hiring [REF
e83	In 2004 the American Psychological Association summarized the issue as "Psychological research confirms that violent video games can increase children's aggression, but that parents moderate the negative effects."[REF
e83	In 2004 the American Psychological Association summarized the issue as "Psychological research confirms that violent video games can increase children's aggression
e86	In a 1995 survey of 184 Gamblers Anonymous members in Illinois, Illinois State Professor Henry Lesieur found that 56 percent admitted to some illegal act to obtain money to gamble
e88	In a 2004 report by the US Department of Justice, researchers interviewed people who had been arrested in Las Vegas and Des Moines and found that the percentage of problem or pathological gamblers among the arrestees was three to five times higher than in the general population [REF
e88	In a 2004 report by the US Department of Justice, researchers interviewed people who had been arrested in Las Vegas and Des Moines and found that the percentage of problem or pathological gamblers among the arrestees was three to five times higher than in the general population [REF
e91	In attempt to close the gap between Arab and Jewish education sectors, the Israeli education minister announced an affirmative action policy, promising that Arabs would be granted 25% of the education budget, more than their proportional share in the population (18
e91	In attempt to close the gap between Arab and Jewish education sectors, the Israeli education minister announced an affirmative action policy, promising that Arabs would be granted 25% of the education budget, more than their proportional share in the population (18
e92	In Buddhism, the problem of evil, or the related problem of dukkha, is one argument against a benevolent, omnipotent creator god, identifying such a notion as attachment to a false concept [REF
e95	the American Civil Rights Institute's Ward Connerly stated, "There is nothing positive, affirmative, or equal about 'affirmative action' programs that give preference to some groups based on race."[REF

Bibliography

[1] Benjamin P Abbott et al. "Astrophysical implications of the binary black hole merger GW150914". In: *The Astrophysical Journal Letters* 818.2 (2016), p. L22.

[2] Khalid Al Khatib et al. "A news editorial corpus for mining argumentation strategies". In: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2016, pp. 3433–3443.

[3] Jamal Bentahar, Bernard Moulin, and Micheline Bélanger. "A taxonomy of argumentation models used for knowledge representation". In: *Artificial Intelligence Review* 33.3 (2010), pp. 211–259. DOI: 10.1007/s10462-010-9154-1. URL: https://doi.org/10.1007/s10462-010-9154-1.

[4] Andrei Bondarenko et al. "An abstract, argumentation-theoretic approach to default reasoning". In: *Artificial intelligence* 93.1-2 (1997), pp. 63–101.

[5] Jill Burstein and Daniel Marcu. "A machine learning approach for identification thesis and conclusion statements in student essays". In: *Computers and the Humanities* 37.4 (2003), pp. 455–467.

[6] Martin Caminada and Gabriella Pigozzi. "On judgment aggregation in abstract argumentation". In: *Autonomous Agents and Multi-Agent Systems* 22.1 (2011), pp. 64–102.

[7] Phan Minh Dung. "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games". In: *Artificial Intelligence* 77.2 (1995), pp. 321–357. ISSN: 0004-3702. DOI: https://doi.org/10.1016/0004-3702(94)00041-X. URL: https://www.sciencedirect.com/science/article/pii/000437029400041X.

[8]   Phan Minh Dung, Robert A Kowalski, and Francesca Toni. "Assumption-based argumentation". In: *Argumentation in artificial intelligence.* Springer, 2009, pp. 199–218.

[9]   Ekkehard Eggs. "39. Vertextungsmuster Argumentation: Logische Grundlagen". In: *Handbücher zur Sprach-und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (HSK).* De Gruyter Mouton, 2008, pp. 397–414.

[10] Luanne Freund, Charles L. A. Clarke, and Elaine G. Toms. "Towards Genre Classification for IR in the Workplace". In: *Proceedings of the 1st International Conference on Information Interaction in Context.* IIiX. Copenhagen, Denmark: Association for Computing Machinery, 2006, pp. 30–36. ISBN: 1595934820. DOI: 10.1145/1164820.1164829. URL: https://doi.org/10.1145/1164820.1164829.

[11] Guido Governatori et al. "Argumentation semantics for defeasible logic". In: *Journal of Logic and Computation* 14.5 (2004), pp. 675–702.

[12] Nancy Green. "Identifying argumentation schemes in genetics research articles". In: *Proceedings of the 2nd Workshop on Argumentation Mining.* 2015, pp. 12–21.

[13] Ivan Habernal and Iryna Gurevych. "Argumentation mining in user-generated web discourse". In: *Computational Linguistics* 43.1 (2017), pp. 125–179.

[14] Ivan Habernal and Iryna Gurevych. "Exploiting debate portals for semi-supervised argumentation mining in user-generated web discourse". In: *Proceedings of the 2015 conference on empirical methods in natural language processing.* 2015, pp. 2127–2137.

[15] *IBM Project Debater.* Website. https://research.ibm.com/haifa/dept/vst/debating_data.shtml/.

[16] Jussi Karlgren and Douglass Cutting. "Recognizing Text Genres with Simple Metrics Using Discriminant Analysis". In: *Proceedings of the 15th Conference on Computational Linguistics - Volume 2.* COLING '94. Kyoto, Japan: Association for Computational Linguistics, 1994, pp. 1071–1075. DOI: 10.3115/991250.991324. URL: https://doi.org/10.3115/991250.991324.

[17] Brett Kessler, Geoffrey Numberg, and Hinrich Schütze. "Automatic Detection of Text Genre". In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational*

Linguistics. ACL '98/EACL '98. Madrid, Spain: Association for Computational Linguistics, 1997, pp. 32–38. DOI: 10.3115/976909.979622. URL: https://doi.org/10.3115/976909.979622.

[18] Anirban Laha and Vikas Raykar. "An empirical evaluation of various deep learning architectures for bi-sequence classification tasks". In: *arXiv preprint arXiv:1607.04853* (2016).

[19] Ran Levy et al. "Context dependent claim detection". In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. 2014, pp. 1489–1500.

[20] Marco Lippi and Paolo Torroni. "Context-independent claim detection for argument mining". In: *Twenty-Fourth International Joint Conference on Artificial Intelligence*. 2015.

[21] Raquel Mochales and Aagje Ieven. "Creating an argumentation corpus: do theories apply to real arguments? A case study on the legal argumentation of the ECHR". In: *Proceedings of the 12th international conference on artificial intelligence and law*. 2009, pp. 21–30.

[22] Sanjay Modgil and Henry Prakken. "A general account of argumentation with preferences". In: *Artificial Intelligence* 195 (2013), pp. 361–397.

[23] Sanjay Modgil and Henry Prakken. "The ASPIC+ framework for structured argumentation: a tutorial". In: *Argument & Computation* 5.1 (2014), pp. 31–62.

[24] Marie-Francine Moens et al. "Automatic Detection of Arguments in Legal Texts". In: *Proceedings of the 11th International Conference on Artificial Intelligence and Law*. ICAIL '07. Stanford, California: Association for Computing Machinery, 2007, pp. 225–230. ISBN: 9781595936806. DOI: 10.1145/1276318.1276362. URL: https://doi.org/10.1145/1276318.1276362.

[25] Raquel Mochales Palau and Marie-Francine Moens. "Argumentation Mining: The Detection, Classification and Structure of Arguments in Text". In: *Proceedings of the 12th International Conference on Artificial Intelligence and Law*. ICAIL '09. Barcelona, Spain: Association for Computing Machinery, 2009, pp. 98–107. ISBN: 9781605585970. DOI: 10.1145/1568234.1568246. URL: https://doi.org/10.1145/1568234.1568246.

[26] Andreas Peldszus and Manfred Stede. "From argument diagrams to argumentation mining in texts: A survey". In: *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)* 7.1 (2013), pp. 1–31.

[27] Chaim Perelman and Lucie Olbrechts-Tyteca. "The new rhetoric: a treatise on argumentation, trans". In: *John Wilkinson and Purcell Weaver (Notre Dame, IN: University of Notre Dame Press, 1969)* 19 (1969).

[28] Iyad Rahwan. "Mass argumentation and the semantic web". In: *Journal of Web Semantics* 6.1 (2008), pp. 29–37.

[29] Chris Reed et al. "Language resources for studying argument". In: *Proceedings of the 6th conference on language resources and evaluation-LREC 2008*. ELRA. 2008, pp. 2613–2618.

[30] Ruty Rinott et al. "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*. 2015, pp. 440–450.

[31] Eyal Shnarch et al. "GRASP: Rich Patterns for Argumentation Mining". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics, Sept. 2017, pp. 1345–1350. DOI: 10.18653/v1/D17-1140. URL: https://aclanthology.org/D17-1140.

[32] Guillermo R Simari and Ronald P Loui. "A mathematical treatment of defeasible reasoning and its implementation". In: *Artificial intelligence* 53.2-3 (1992), pp. 125–157.

[33] T. J. Smiley. "The Uses of Argument. By S. E. Toulmin, Professor of Philosophy, University of Leeds. [Cambridge: at the University Press. 1958. vii, 261 and (index) 2 pp. 22s. 6d. net.]" In: *The Cambridge Law Journal* 16.2 (1958), pp. 251–252. DOI: 10.1017/S0008197300003937.

[34] Christian Stab and Iryna Gurevych. "Annotating argument components and relations in persuasive essays". In: *Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical papers*. 2014, pp. 1501–1510.

[35] Christian Stab and Iryna Gurevych. "Identifying argumentative discourse structures in persuasive essays". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014, pp. 46–56.

[36] Manfred Stede and Antje Sauermann. "Linearization of arguments in commentary text". In: *Proceedings of the Workshop on Multidisciplinary Approaches to Discourse, Oslo.* 2008, pp. 242–243.

[37] Frans Van Eemeren, Rob Grootendorst, and Frans H van Eemeren. *A systematic theory of argumentation: The pragma-dialectical approach.* Cambridge University Press, 2004.

[38] Gerard AW Vreeswijk. "Abstract argumentation systems". In: *Artificial intelligence* 90.1-2 (1997), pp. 225–279.

[39] Gerard AW Vreeswik and Henry Prakken. "Credulous and sceptical argument games for preferred semantics". In: *European Workshop on Logics in Artificial Intelligence.* Springer. 2000, pp. 239–253.

[40] Marilyn Walker et al. "A corpus for research on deliberation and debate". In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12).* 2012, pp. 812–817.

[41] Vern Walker, Karina Vazirova, and Cass Sanford. "Annotating patterns of reasoning about medical theories of causation in vaccine cases: Toward a type system for arguments". In: *Proceedings of the First Workshop on Argumentation Mining.* 2014, pp. 1–10.

[42] Yining Wu, Martin Caminada, and Dov M Gabbay. "Complete extensions in argumentation coincide with 3-valued stable models in logic programming". In: *Studia logica* 93.2-3 (2009), p. 383.