Notes on Abstract Argumentation Theory

Xin Chen (chenxin_hello@outlook.com)

Latest edit: November 3, 2024

This note gives a concise introduction to Dung's abstract argumentation theory [4] and some extensions (probabilistic extension, in particular). We mainly investigate Dung's original notions of complete, grounded, preferred, and stable semantics. However, for other semantics which are available in the literature since Ding's seminal work [4], for instance, semi-stable, ideal, eager, stage, CF2 and stage2 semantics [1, 3], will be considered in the next version of this note.

Contents

1 Abstract Argumentation Theory

This section is devoted to the *abstract argumentation theory* introduced in the seminal paper by Dung [4]. This formalism is based on the idea that arguments are defeasible entities which may attack each other and whose acceptance is subject to a given reasonable criterion (called *semantics*). Formally, an argumentation framework is represented as a directed graph in which the arguments are represented as nodes and the attack relations as edges.

Given such a graph, a key question is which set(s) of arguments can be accepted. That is, once the argumentation framework has been constructed, how to determine which arguments to accept or reject? To answer this kind of questions corresponding to define an *argumentation semantics*. By and large, two kinds of approaches for argumentation semantics are available in the literature:

(1) the *extension*-based approach (proposed in Dung's original paper [4]), and

(2) the *labelling*-based approach.

In this note, however, for the sake of simplicity, we only focus on the first one.

1.1 Argumentation frameworks

An (abstract) argumentation framework is nothing, mathematically speaking, but a pair consists of a set of arguments and a binary relation representing the attack relationship between arguments, that is, a *directed graph* in which the arguments as nodes and the attack relations as arrows. An argument is an abstract entity whose role is solely determined by its relations to other arguments.

Definition 1.1 (Argumentation frameworks) An *argumentation framework* (AF) is a pair

$$AF = (Ar, \rightarrow)$$

where Ar is a non-empty set of arguments, and $\rightarrow$ is a binary relation on Ar, i.e., $\rightarrow \subseteq A \times A$. $\dashv$

Given an AF $AF = (Ar, \rightarrow)$, let $a, b \in Ar$, we say that a *attacks/defeats* b (accordingly, a is an *attacker* of b) iff $a \rightarrow b$ holds. For any $X \cup \{a\} \subseteq Ar$, then we say that X *attacks* a, denoted as $X \rightarrow a$, if there exists $b \in X$ such that $b \rightarrow a$. Likewise, we say that a *attacks* X, written as $a \rightarrow X$, if there is $b \in X$ such that $a \rightarrow b$. For $a \in Ar$, we define a^- and a^+ as follows:

$$\begin{array}{rcl} a^- & \coloneqq & \{b \in Ar \mid b \rightarrow a\}, \\ a^+ & \coloneqq & \{b \in Ar \mid a \rightarrow b\}. \end{array}$$

Those two operations can be extended to any subset $X \subseteq Ar$ by

$$X^\circ = \bigcup_{a \in X} a^\circ \qquad \circ \in \{+, -\}.$$

The attack relations between argument(s) aforementioned are summarized by Table 1.

—

Definition 1.2 (Defense) Let $AF = (Ar, \rightarrow)$ be an AF and $X \subseteq Ar$. The set X *defends* $a \in Ar$ (or, a *acceptable* w.r.t. X) iff $\forall b \in Ar : b \rightarrow a \Rightarrow X \rightarrow b$ (or equivalently, $\forall b \in a^- : X \rightarrow b$, where a^- is the set of attackers of a). [1] $\dashv$

[1] The original terminology in Dung's paper [4] was that "a is *acceptable* w.r.t. X".

Table 1: The attack relations between argument(s)

notation	meaning
$a \rightarrow b$	a attacks b
$a \rightarrow X$	a attacks X $(\exists b \in X : a \rightarrow b)$
$X \rightarrow a$	X attacks a $(\exists b \in X : b \rightarrow a)$
$X \rightarrow Y$	X attacks Y $(\exists a \in X, \exists b \in Y : a \rightarrow b)$

According to vacuous truth, if $a^- = \emptyset$ (such a are called *initial arguments*, the set of initial arguments of an AF AF denoted as $\mathcal{NI}(AF)$, i.e., $\mathcal{IN}(AF) = \{a \in Ar \mid a^- = \emptyset\}$) then a defended by any set of arguments, including $\emptyset$ of course.

Definition 1.3 (Characteristic function) Let $AF = (Ar, \rightarrow)$ be an AF. The *characteristic function* of AF is mapping $C_{AF} \colon \wp(Ar) \rightarrow \wp(Ar)$ such that

$$C_{AF}(X) = \{a \in Ar \mid X \text{ defends } a\}$$

for each $X \subseteq Ar$. $\dashv$

Proposition 1.4 For any AF $AF = (Ar, \rightarrow)$, then:

(1) $C_{AF}(\emptyset) = \mathcal{IN}(AF) = \{a \in Ar \mid a^- = \emptyset\}$.

(2) $\mathcal{IN}(AF) \subseteq C_{AF}(X)$.

(3) $C_{AF}(X) = (X^+)^+$.

where $X \subseteq Ar$. $\dashv$

Definition 1.5 (Conflict-freeness) Let $AF = (Ar, \rightarrow)$ be an AF and $X \subseteq Ar$. X is said to be *conflict-free* iff $\neg\exists a, b \in X : a \rightarrow b$. $\dashv$

Clearly, the emptyset $\emptyset$ is conflict–free. Furthermore, if each argument has at least one attacker in an AF, i.e. $a^- \neq \emptyset$ for every argument a, then $\emptyset$ is a (conflict–free) fixed point of the characteristic function.

After introduced the notion of *defense*, a basic requirement for a set of arguments is the capability to defend all its elements.

Definition 1.6 (Admissibility) Let $AF = (Ar, \rightarrow)$ be an AF. A set $X \subseteq Ar$ is called an *admissible set* iff (1) X is conflict-free and (2) $X \subseteq C_{AF}(X)$. $\dashv$

Thus, an admissible set is required to be both internally coherent, that is, conflict-free and able to defend its elements. Admissible sets always exist. A very trivial case is that the empty set $\emptyset$ is admissible for any argumentation framework.

Example 1.7

$$AF: \qquad a \longrightarrow b \longrightarrow c \rightleftarrows d$$

- non-empty conflict-free sets: $\{a\}, \{b\}, \{c\}, \{d\}, \{a, c\}, \{a, d\}, \{b, d\}$.

- non-empty admissible sets: $\{a\}, \{d\}, \{a, c\}, \{a, d\}$.

- $C_{AF}(\{a\}) = \{a\}$, $C_{AF}(\{d\}) = \{d, a\}$, $C_{AF}(\{a, c\}) = \{a, c\}$, $C_{AF}(\{a, d\}) = \{a, c, d\}$. ⊣

Finally, it is worth recalling that admissibility and defense are related by a basic property. In terms of extensions, if an admissible set defends an argument, it is possible to add the argument to the set while preserving its admissibility and its capability to defend any other argument. This was proved in the so-called Dung's Fundamental Lemma.

Lemma 1.8 (Dung's Fundamental Lemma [4]) For any AF $AF = (Ar, \rightarrow)$, let X be an admissible set and a, b be arguments defended by X. Then

(1) $X' = X \cup \{a\}$ is an admissible set;

(2) X' defends b.

⊣

1.2 An overview of (extension-based) argumentation semantics

In this subsection we provide an overview of some well-known argumentation semantics, starting from the very basic notion of "naïve semantics" [1, § 3.2] and then discussing Dung's original concepts of *complete*, *stable*, *preferred* and *grounded* semantics [4].

Remark 1.9 In the literature, *admissibility* and *conflict–freeness* mentioned in § 1.1 are sometimes viewed as semantics, and sometimes as properties. We choose to treat them as properties in this note. ⊣

Definition 1.10 (Extension-based semantics) An *extension-base semantics* (*semantics*, for short) σ associates with any argumentation framework $AF = (Ar, \rightarrow)$ is a subset of $\wp(Ar)$, denoted as $\sigma(AF)$. The elements in $\sigma(AF)$ are called *extensions* (under semantics σ).

Let $\mathcal{D}^\sigma$ be the class of AFs where a semantic σ is defined, that is,

$$\mathcal{D}^\sigma = \{AF \mid \sigma(AF) \neq \emptyset\}.$$

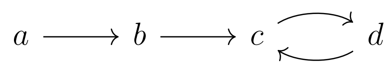
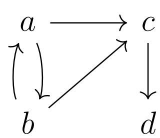
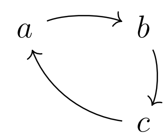
A semantics σ is called *universally defined* if $\mathcal{D}^\sigma$ includes all AFs. If for all $AF \in \mathcal{D}^\sigma$ we have that $|\sigma(AF)| = 1$, i.e. the semantics σ always prescribes exactly one extension, then σ is said to belong to the *unique-status approach*, it is said to belong to the *multiple-status approach*, otherwise. ⊣

1.2.1 Naïve Semantics

Naïve semantics (denoted as $\mathcal{NA}$) corresponds to selecting as many arguments as possible, provided that there are no conflicts among them. It is a sort of greedy strategy, driven by the only criterion of avoiding conflicts. Formally it corresponds to requiring conflict–freeness together with a maximality property.

Definition 1.11 (Naïve extensions) Let $(Ar, \rightarrow)$ be an AF. A set $X \subseteq Ar$ is called a *naïve extension* iff X is a maximal conflict-free set. $\dashv$

Example 1.12

AF	naïve extensions $\mathcal{NA}(AF)$
	$\{a, c\}, \{a, d\}, \{b, d\}$
	$\{a, d\}, \{b, d\}, \{c\}$
	$\{a\}, \{b\}, \{c\}$

$\dashv$

1.2.2 Complete Semantics

Complete semantics (denoted as $\mathcal{CO}$) can be regarded as a strengthening of the basic requirements enforced by the idea of admissibility, it lies at the heart of all Dung's semantics (see Fig. 1).

A complete extension is a conflict–free set which includes precisely those arguments it defends. That is, if an argument is defended by the set it should be in the set, and if an argument is not defended by the set, it should not be in the set.

Definition 1.13 (Complete extensions) Let $AF = (Ar, \rightarrow)$ be an AF. A set $X \subseteq Ar$ is called a *complete extension* iff X is conflict–free and $X = C_{AF}(X)$. $\dashv$

Technically this means that a complete extension is a *conflict–free fixed point* of the characteristic function. It is clear that every complete extension is an admissible set, but the reverse does not hold in general.

Proposition 1.14 Let $AF = (Ar, \rightarrow)$ be any AF.

- $\mathcal{CO}(AF) \neq \emptyset$, that is, complete semantics is universally defined (see Definition 1.10).
- $\emptyset \in \mathcal{CO}(AF)$ iff $\mathcal{NI}(AF) = \emptyset$, where $\mathcal{NI}(AF) := \{a \mid a^- = \emptyset\}$.
- $\forall E \in \mathcal{CO}(AF) : \mathcal{NI}(AF) \subseteq E$.

$\dashv$

Example 1.15

$\mathcal{AF}$	complete extensions $\mathcal{E}_{\mathcal{CO}}(\mathcal{AF})$
$a \longrightarrow b \longrightarrow c \rightleftarrows d$	$\{a\}, \{a, d\}, \{a, c\}$
$\begin{tikzcd} a \arrow[r] \arrow[d, bend left] & c \arrow[d] \\ b \arrow[u, bend left] \arrow[ur] & d \end{tikzcd}$	$\emptyset, \{a, d\}, \{b, d\}$
$\begin{tikzcd} a \arrow[r, bend left] & b \arrow[d, bend left] \\ & c \arrow[ul, bend left] \end{tikzcd}$	$\emptyset$

$\dashv$

1.2.3 Grounded Semantics

To accept only the arguments that one cannot avoid accepting, to reject only the arguments that one cannot avoid rejecting, and abstaining as much as possible. This gives rise to the most skeptical semantics among those based on complete extensions, namely the *grounded semantics* (denoted as $\mathcal{GR}$).

Definition 1.16 (The grounded extension) Let $AF = (Ar, \rightarrow)$ be an AF. The *grounded extension* of AF is a minimal conflict–free fixed point of the characteristic function. $\dashv$

Notice that the uniqueness of the grounded extension. Since the characteristic function C_{AF} is monotonic, it follows from the Knaster-Tarski Theorem that C_{AF} has a unique smallest fixed point, it can then be proved that this fixed point is also conflict–free.

Proposition 1.17 For any AF $AF = (Ar, \rightarrow)$, the following statements are equivalent:

(1) X is a minimal conflict–free fixed point of C_{AF}.

(2) X is the smallest fixed point of C_{AF}.

$\dashv$

It follows that:

- the grounded extension is unique (i.e. grounded semantics belongs to the unique-status approach);
- the grounded extension is the least complete extension, in particular it is included in any complete extension.

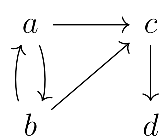
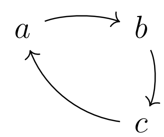
The grounded extension of an argumentation framework AF will be denoted as $GR(AF)$. By definition, the grounded extension coincides with the intersection of all complete extensions, that is

$$GR(AF) = \bigcap \mathcal{CO}(AF).$$

Hence, a unique grounded extension always exists, although it may be the empty set.

Proposition 1.18 If there are no initial arguments in an AF, then the grounded extension is the empty set. That is, $\mathcal{IN}(AF) = \emptyset \Rightarrow GR(AF) = \emptyset$. $\dashv$

Example 1.19 (Grounded extension)

AF	grounded extensions $\mathcal{GR}(AF)$
$a \longrightarrow b \longrightarrow c \rightleftarrows d$	$\{a\}$
	$\emptyset$
	$\emptyset$

$\dashv$

1.2.4 Preferred Semantics

The idea of maximizing accepted arguments is expressed by *preferred semantics* ($\mathcal{PR}$).

Definition 1.20 (Preferred extensions) Let $AF = (Ar, \rightarrow)$ be an AF. A *preferred extension* is a maximal (w.r.t. $\subseteq$) admissible set of AF. $\dashv$

Every argumentation framework possesses at least one preferred extension. Hence, preferred extension semantics is always defined for any argumentation framework. Relationships of preferred extensions with other semantics notions have been analyzed in Dung's [4]. Preferred extensions can for instance equivalently be characterized as maximal complete extensions.

Proposition 1.21 Let $AF = (Ar, \rightarrow)$ be an AF and $X \subseteq Ar$. The following two statements are equivalent:

(1) X is a maximal admissible set of AF.

(2) X is a maximal complete extension of AF.

$\dashv$

This in particular implies that the grounded extension is included in any preferred extension, as it is in any complete extension.

Preferred semantics has often been regarded as the most satisfactory semantics in the context of Dung's framework.

Example 1.22 (Preferred extensions)

AF	preferred extensions $\mathcal{PR}(AF)$
$\begin{tikzcd} a \arrow[r] & b \arrow[r] & c \arrow[r, bend left] & d \arrow[l, bend left] \end{tikzcd}$	$\{a, d\}, \{a, c\}$
$\begin{tikzcd} a \arrow[r] \arrow[d, bend left] & c \arrow[d] \\ b \arrow[u, bend left] \arrow[ur] & d \end{tikzcd}$	$\{a, d\}, \{b, d\}$
$\begin{tikzcd} a \arrow[r, bend left] & b \arrow[d, bend left] \\ & c \arrow[ul, bend left] \end{tikzcd}$	$\emptyset$

$\dashv$

1.2.5 Stable Semantics

Stable semantics (denoted as $\mathcal{ST}$) relies on a very simple intuition: an extension should be able to attack all arguments not included in it.

Definition 1.23 (Stable extensions) Let $AF = (Ar, \rightarrow)$ be an AF. A *stable extension* of AF is a conflict–free set X such that $X \cup X^+ = Ar$. $\dashv$

A stable extension is an admissible set. In the context of game theory, the notion of stable extension coincides with the notion of *stable solution* of n-person games [4]. Every stable extension is a preferred extension, but not vice versa.

Proposition 1.24 Let $AF = (Ar, \rightarrow)$ be an AF and $X \subseteq Ar$. The following statements are equivalent:

(1) X is a stable extension.

(2) X is an admissible set s.t. $X \cup X^+ = Ar$.

(3) X is a complete extension s.t. $X \cup X^+ = Ar$.

(4) X is a preferred extension s.t. $X \cup X^+ = Ar$.

(5) $X^+ = Ar \setminus X$.

$\dashv$

No stable extension is empty, and not every argument framework has stable extensions.

Example 1.25 (Stable extensions)

AF	stable extensions $\mathcal{ST}(AF)$
$\begin{tikzcd} a \arrow[r] & b \arrow[r] & c \arrow[r, bend left] & d \arrow[l, bend left] \end{tikzcd}$	$\{a, d\}, \{a, c\}$
$\begin{tikzcd} a \arrow[r] \arrow[d, bend left] & c \arrow[d] \\ b \arrow[u, bend left] \arrow[ur] & d \end{tikzcd}$	$\{a, d\}, \{b, d\}$
$\begin{tikzcd} a \arrow[r, bend left] & b \arrow[d, bend left] \\ & c \arrow[ul, bend left] \end{tikzcd}$	–

$\dashv$

1.3 Computational problems and Complexity

Given an AF $AF = (Ar, \rightarrow)$ and a semantics σ, the *verification problem*, denoted as Var_σ, is deciding whether a set $X \subseteq Ar$ is a σ-extension of AF.

For $a \in Ar$, the *credulous acceptance problem*, denoted as CA_σ, is deciding whether a is credulously accepted, that is deciding whether a belongs to a σ–extension of AF. Moreover, the *skeptical acceptance problem*, denoted as SA_σ, is deciding whether a is skeptically accepted, that is deciding whether a belongs to every σ–extension of AF.

Clearly, $CA_{\mathcal{GR}}$ and $SA_{\mathcal{GR}}$ are identical problems. The computational complexity of those problems are summarized in table 2 [5, 6].

Table 2: Complexity of Ver, CA and SA problems

semantics	Ver_σ	CA_σ	SA_σ
$\mathcal{CO}$	in P	NP–c	P–c
$\mathcal{GR}$	in P	in P	in P
$\mathcal{PR}$	coNP–c	NP–c	$\prod_2^p$–c
$\mathcal{ST}$	in P	NP–c	coNP–c

1.4 A summary

Table 3: Basic notions of AAT

$AF = (Ar, \to)$ is an AF, $X \subseteq Ar$ and $a, b \in Ar$.

Notions	Definition or equivalent descriptions
$a \to b$	a attacks/defeats b (a is an attacker/counterargument of b)
$X \to a$ (X attacks a)	$\exists b \in X : b \to a$
a^-	$a^- = \{b \in Ar \mid b \to a\}$
a^+	$a^+ = \{b \in Ar \mid a \to b\}$
X^-	$X^- = \{b \mid \exists a \in X : b \to a\} = \bigcup_{a \in X} a^-$
X^+	$X^+ = \{b \mid \exists a \in X : a \to b\} = \bigcup_{a \in X} a^+$
X defends a / a is admissible w.r.t. X	$\forall b : b \to a \Rightarrow X \to b$ [X defends a for any X if $a^- = \emptyset$]
$\mathcal{IN}(AF)$	$\mathcal{IN}(AF) = \{a \in Ar \mid a^- = \emptyset\}$ the set of initial arguments
C_{AF}: the characteristic function	$C_{AF}(X) = \{a \mid X$ defends $a\}$ $C_{AF}X = (X^+)^+$ [$\emptyset$ is a fixed point if $\mathcal{IN}(AF) = \emptyset$]
X is conflict–free	$\nexists a, b \in X : a \to b$ [$\emptyset$ is conflict–free] $\nexists a, b \in X : a \in b^-$ $X \cap X^+ = \emptyset$
X is admissible	X is conflict–free and $X \subseteq C_{\mathcal{AF}}(X)$ [$\emptyset$ is admissible]
Semantics	
$\mathcal{NA}(AF)$: X is a *naïve extension*	X is a maximal conflict–free set
$\mathcal{CO}(AF)$: X is a *complete extension*	X is a conflict–free and $C_{AF}(X) = X$ (conflict–free fixed point)
$\mathcal{GR}(AF)$: X is the *grounded extension*	X is the least fixed point of C_{AF} X is the minimal complete extension $X = GR(AF) = \bigcap \mathcal{CO}(AF)$
$\mathcal{PR}(AF)$: X is a *preferred extension*	X is a maximal admissible set X is a maximal complete extension
$\mathcal{ST}(AF)$: X is a *stable extension*	X is conflict–free and $X \cup X^+ = Ar$ $X^+ = Ar \setminus X$ X is a preferred extension and $X \cup X^+ = Ar$ X is a complete extension and $X \cup X^+ = Ar$ X is admissible and $X \cup X^+ = Ar$ $X = \{a \in Ar \mid X \not\to a\}$ X is conflict–free and attacks each argument that is not in X

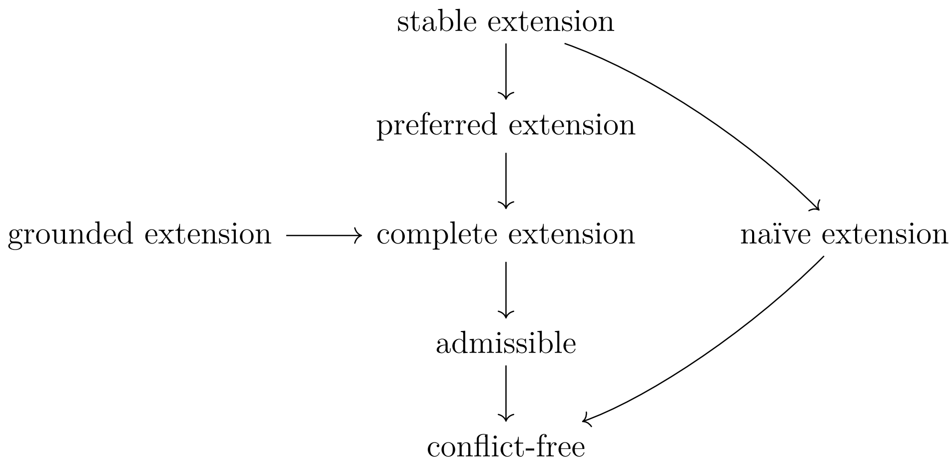
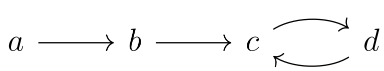
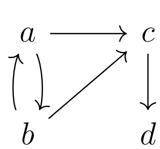
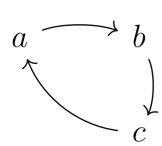


Figure 1: Relations among extension–based semantics

Table 4: A summary of the examples mentioned

AF	naïve	complete	grounded	preferred	stable
	$\{a, c\}$ $\{a, d\}$ $\{b, d\}$	$\{a\}$ $\{a, c\}$ $\{a, d\}$	$\{a\}$	$\{a, c\}$ $\{a, d\}$	$\{a, c\}$ $\{a, d\}$
	$\{a, d\}$ $\{b, d\}$ $\{c\}$	$\emptyset$ $\{a, d\}$ $\{b, d\}$	$\emptyset$	$\{a, d\}$ $\{b, d\}$	$\{a, d\}$ $\{b, d\}$
	$\{a\}$ $\{b\}$ $\{c\}$	$\emptyset$	$\emptyset$	$\emptyset$	–

References

[1] Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. Abstract argumentation frameworks and their semantics. In Baroni et al. [2]. URL: https://www.collegepublications.co.uk/handbooks/?00003.

[2] Pietro Baroni, Dov M. Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors. *Handbook of Formal Argumentation*, volume 1. College Publications, 2018. URL: https://www.collegepublications.co.uk/handbooks/?00003.

[3] Pietro Baroni and Massimiliano Giacomin. Semantics of abstract argument systems. In Rahwan and Simari [7], pages 25–44. doi:10.1007/978-0-387-98197-0_2.

[4] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995. doi:10.1016/0004-3702(94)00041-X.

[5] Paul E. Dunne and Michael Wooldridge. Complexity of abstract argumentation. In Rahwan and Simari [7], pages 85–104. doi:10.1007/978-0-387-98197-0_5.

[6] W. Dvorak and Paul E. Dunne. Computational problems in formal argumentation and their complexity. In Baroni et al. [2]. URL: http://hdl.handle.net/20.500.12708/29850.

[7] Iyad Rahwan and Guillermo R. Simari, editors. *Argumentation in Artificial Intelligence*. Springer, New York, 2009. doi:10.1007/978-0-387-98197-0.