Assumption-Based Reasoning for the Next Generation

Alexander Bochman

HIT

LNGAI 2021

What is NMR? (First Approximation)

- NMR $\neq$ (Computable) Logic.

NMR is more than logic; it should always include logic as its essential part, but never reduces to the latter.

- NMR $\neq$ Syntax + Nonmonotonic Semantics. This reduction hides the underlying (deductive, monotonic) Logic.

NMR = Logic + Nonmonotonic Semantics.

- The underlying logic is the main source of variation and quality of nonmonotonic formalisms.

Reiter's default logic

Default theory: (W, D), where W are axioms, D are

default rules — $A : b/C$.

Context-depended inference (Marek & Truscziński 1989):
$\mathcal{D}(s)$ - closure of W wrt classical entailment and the rules:

$$\{A \vdash C \mid A : b/C \in D \;\&\; \neg B \notin s, \text{ for any } B \in b\}.$$

Then s is an extension of the default theory if $s = \mathcal{D}(s)$.

Bisequents — inference rules $a : b \Vdash c : d$.

If no b is assumed, and all d's are assumed, then all a's hold only if one of c's holds.

Representation of default rules $\quad A : b/C \;\equiv\; A : \neg b \Vdash C :$.

A *biconsequence relation* is a set of bisequents satisfying

$$\frac{a : b \Vdash c : d}{a' : b' \Vdash c' : d'} \qquad \text{(Mon)}$$

whenever $a \subseteq a', b \subseteq b', c \subseteq c', d \subseteq d'$.

$$A : \Vdash A : \qquad\qquad : A \Vdash : A \qquad \text{(Ref)}$$

$$\frac{a{:}b \Vdash A, c{:}d \quad A, a{:}b \Vdash c{:}d}{a : b \Vdash c : d} \qquad \frac{a{:}b \Vdash c{:}A, d \quad a{:}A, b \Vdash c{:}d}{a : b \Vdash c : d} \qquad \text{(Cut)}$$

Logical and nonmonotonic semantics

- *Bitheories* of $\Vdash \equiv$ pairs (u, v) such that $u : \overline{v} \nVdash \overline{u} : v$.
- u is a *theory* of $\Vdash \equiv (u, u)$ is a bitheory of $\Vdash$.

Logical binary semantics

A set of *bimodels* — pairs of valuations (sets of propositions).

$a{:}b \Vdash c{:}d$ is *valid* in a binary semantics $\mathcal{B}$, if, for any $(u, v) \in \mathcal{B}$, if $a \subseteq u, b \subseteq \overline{v}$, then either $c \cap u \neq \emptyset$, or $d \cap \overline{v} \neq \emptyset$.

Nonmonotonic semantics of extensions

u is an extension of $\Vdash$ if u is a theory of $\Vdash$, and there is no bitheory (u', u) such that $u' \subset u$.

Classicality

A biconsequence relation is *supraclassical*, if it satisfies

Supraclassicality $\dfrac{a \vDash A}{a : \Vdash A :}$ and $\dfrac{a \vDash A}{: A \Vdash : a}$

Falsity $\mathbf{f} : \Vdash$ and $\Vdash : \mathbf{f}.$

A *classical* binary semantics - pairs of consistent deductive theories - knowledge-belief pairs.

$\Vdash_D$ — the least supraclassical biconsequence relation that includes the rules of a default theory D.

Theorem

Extensions of a default theory D coincide with extensions of $\Vdash_D$.

Assumption-Based Reasoning (ABR)

ABR = Logic + Assumptions

A bipolar reasoning about acceptance and rejection of propositions

Facts: data, evidence, constraints...

Assumptions: laws, defeasible rules, defaults, arguments...

Leibniz's Principle of Sufficient Reason

Any fact should have a reason (cause) for its acceptance.

Principle of default acceptance for assumptions

Any assumption is accepted, unless there is a reason for its rejection (cancellation).

Simple Default Theories

Pair $(D, \mathcal{A})$: D — a set of ordinary inference rules; $\mathcal{A}$ — a set of (default) *assumptions*.
$\mathrm{Cn}_D(u)$ — the closure of a set u wrt D and classical entailment.

A set $\mathcal{A}_0$ of assumptions is stable if $\neg A \in \mathrm{Cn}_D(\mathcal{A}_0)$, for any $A \in \mathcal{A} \backslash \mathcal{A}_0$.

Definition

s is an extension of $(D, \mathcal{A})$ if $s = \mathrm{Cn}_D(\mathcal{A}_0)$, for some stable set $\mathcal{A}_0$ of assumptions.

Theorem

s is an extension of a simple default theory $(D, \mathcal{A})$ iff

- $s = \mathrm{Cn}_D(s \cap \mathcal{A})$;
- *s decides $\mathcal{A}$: for any $A \in \mathcal{A}$, either $A \in s$, or $\neg A \in s$;*

The reduction

New propositional atoms A°, for any classical proposition A.

If D is a default theory, then a simple default theory is

$$D^\circ = \{A, b^\circ \vdash C \mid A : b/C \in D\} \cup \{\neg A \vdash \neg A^\circ\}$$

$$\mathcal{A} = \{A^\circ \mid A \text{ appears as a justification in } D\}$$

Polynomial, modular and faithful translation (PFM).

Theorem

u is an extension of a default theory D iff there is a unique extension u_0 of $(D^\circ, \mathcal{A})$ such that $u = u_0 \cap \mathcal{L}$.

History

- Lin & Shoham 1989: Reiter's DL as an argument system:

$$A{:}B_1, \ldots, B_n/C \;\equiv\; A, \neg ab(B_1), \ldots, \neg ab(B_n) \vdash C.$$

  plus $\neg A/ab(A)$ and nonmonotonic rules $\mathbf{t} \Rightarrow \neg ab(B)$.

- Bondarenko et al. 1997: ABA — an assumption-based framework. Deductive system of rules, plus a set of *assumptions*.

$$A{:}B_1, \ldots, B_n/C \;\equiv\; A, \mathbf{M}B_1, \ldots, \mathbf{M}B_n \vdash C,$$

  where $\mathbf{M}B$ are assumptions.

Argumentation Theory

- Dung has shown a fundamental and unifying role of argumentation in nonmonotonic formalisms and logic programming.
- Key novel features of Dung's formalism: default acceptance of arguments and directionality of the attack relation.

> *Dung's argumentation is an abstract theory of assumptions.*

Full-fledged argumentation *reasoning* requires exploring logical systems that could host Dung's argumentation frameworks.

Collective argumentation

A generalization of Dung's argumentation theory: $a \hookrightarrow b$ says that a set a of arguments attacks a set of arguments b.

An informal meaning of an attack $a \hookrightarrow b$:

If all arguments in a are accepted, then at least one of the arguments in b is rejected.

Any argument A is assigned a *subset* $\nu(A) \subseteq \{t, f\}$, where t denotes acceptance, while f denotes rejection — the *Belnap's interpretation* of four-valued logic.

Definition

$a \hookrightarrow b$ *holds* in a four-valued interpretation ν if either $t \notin \nu(A)$, for some $A \in a$, or $f \in \nu(B)$, for some $B \in b$.

Nonmonotonic Semantics of Arguments

Basic Argumentation Principle

An argument is accepted unless there is a reason for its rejection.

An argument is refuted, if it is attacked by an accepted argument set.

An argument is accepted iff it is not refuted.

For the 'classical' understanding of acceptance and rejection, this gives the stable semantics.

Relaxed Argumentation Principles

An argument is rejected iff it is refuted. (Jakobovich & Vermeir)

An argument is accepted iff its attackers are rejected. (Caminada & Gabbay)

Global negation

$\sim A$ is accepted iff A is rejected

$\sim A$ is rejected iff A is accepted.

N-attack relations

$$A \hookrightarrow \sim A \qquad \sim A \hookrightarrow A$$

$$\text{If } a \hookrightarrow A, b \text{ and } a, \sim A \hookrightarrow b, \text{ then } a \hookrightarrow b \tag{AN}$$

$$\text{If } a, A \hookrightarrow b \text{ and } a \hookrightarrow b, \sim A, \text{ then } a \hookrightarrow b$$

Belnap consequence relation

A *Belnap consequence relation* in a language with $\sim$ is a Scott consequence relation $\Vdash$ satisfying the Double Negation rules for $\sim$:

$$A \Vdash \sim\sim A \qquad \sim\sim A \Vdash A.$$

For any set u of propositions, $\sim u = \{\sim A \mid A \in u\}$.

Any N-attack relation generates the corresponding Belnap consequence relation:

$$a \Vdash b \equiv \ a \hookrightarrow \sim b \tag{CA}$$

and vice versa:

$$a \hookrightarrow b \equiv \ a \Vdash \sim b \tag{AC}$$

Global negation switches between arguments (assumptions) and facts!

Assumption-based argumentation (ABA)

- An underlying deductive system;
- a set *Ab* of *assumptions*;
- a *'contrary'* mapping from *Ab* to (factual) propositions.

The representation with N-attack relations:

- The deductive system:

$$a \vdash A \quad \equiv \quad a \hookrightarrow \sim A.$$

- Global negation $\sim$ formalizes both the assumptions and the contrary mapping: for a *factual language* $\mathcal{L}$ without $\sim$,

$$Ab \quad =_{df} \{\sim A \mid A \in \mathcal{L}\}$$

Conjunction

$A \wedge B$ is accepted iff A is accepted and B is accepted

$A \wedge B$ is rejected iff A is rejected or B is rejected

$$a \hookrightarrow b \quad \text{iff} \quad \bigwedge a \hookrightarrow \bigwedge b.$$

An axiomatization:

$$a, A \wedge B \hookrightarrow b \;\text{ iff }\; a, A, B \hookrightarrow b$$
$$a \hookrightarrow A \wedge B, b \;\text{ iff }\; a \hookrightarrow A, B, b \qquad (A_\wedge)$$

Propositional Argumentation

A *propositional attack relation* is a relation $\hookrightarrow$ on the set of classical propositions satisfying the postulates:

(Left Strengthening) If $A \vDash B$ and $B \hookrightarrow C$, then $A \hookrightarrow C$;

(Right Strengthening) If $A \hookrightarrow B$ and $C \vDash B$, then $A \hookrightarrow C$;

(Truth and Falsity) $\mathbf{t} \hookrightarrow \mathbf{f}$ and $\mathbf{f} \hookrightarrow \mathbf{t}$.

An extension to arbitrary sets of propositions:

$$u \hookrightarrow v \;\equiv\; \bigwedge a \hookrightarrow \bigwedge b, \text{ for some finite } a \subseteq u \text{ and } b \subseteq v$$

will satisfy the properties of collective argumentation, and the postulates $(A_\wedge)$ for conjunction.

Adding the global negation $\sim$: postulates AN for $\sim$, plus the rule

Classicality If $a \vDash A$, then $a \hookrightarrow \sim A$ and $\sim A \hookrightarrow a$.

Logical Semantics

Bimodels - pairs (u, v) of deductively closed theories, where u is the set of accepted propositions, while v is the set of non-rejected propositions. A set of bimodels determines a *binary semantics*.

Definition

An attack $A \hookrightarrow B$ is *valid* in a binary semantics $\mathcal{B}$ if there is no bimodel (u, v) from $\mathcal{B}$ such that $A \in u$ and $B \in v$.

$\hookrightarrow_{\mathcal{B}}$ is the set of attacks that are valid in a semantics $\mathcal{B}$.

Theorem

$\hookrightarrow$ is a propositional attack relation iff it coincides with $\hookrightarrow_{\mathcal{B}}$, for some binary semantics $\mathcal{B}$.

Default Logic as Argumentation

Default rule $a{:}b/A$ can be interpreted as an attack

$$a, \sim\neg b \hookrightarrow \sim A.$$

$\hat{u}$ denotes the set of assumptions $\{\sim A \mid A \notin u\}$.

A set w of assumptions is *stable* in an argument theory Δ if, for any assumption A, $A \in w$ iff $w \not\hookrightarrow_{\Delta} A$.

Theorem

A set u is an extension of a default theory Δ iff $\hat{u}$ is a stable set of assumptions in $tr(\Delta)$.

Causal Reasoning

- Causal reasoning is a historical precursor of NMR, but it has been eschewed at the beginning of 20th century by an overly ambitious logical (deductive) approach to reasoning.
- The Causal Calculus (McCain&Turner 1997) can be viewed as an instantiation of ABR in which causal rules $A \Rightarrow B$ (A *causes* B) play the role of assumptions.
- The nonmonotonic semantics of the causal calculus is based on Leibniz's Principle of Sufficient Reason (the Law of Causality).

A detailed and systematic picture can be found in

A. Bochman *A Logical Theory of Causality.* Forthcoming in the MIT Press, 2021.

Conclusions and Prospects

- Logic should still play a crucial (though already not exclusive) role in NMR.
- Assumption-based reasoning (ABR) constitutes the core of NMR.
- ABR and causal reasoning suggest themselves as primary instantiations of NMR due to their potential (yet to be explored) of capturing commonsense reasoning in AI.