A NON–STANDARD APPROACH
TO AUTOEPISTEMIC LOGIC

Abstract

The aim of this paper is to present certain approach to autoepistemic logic (also called ae logic) is presented. The knowledge base of an ideal agent is set up in the standard manner with respect to a set of initial facts (formulas) which are the only facts primarily accepted by the agent. The knowledge base is determined by autoepistemic expansions which are to be sets of all formulas accepted by the agent. In the paper the knowledge base of an agent is characterized by a pair of theories where the first one based on a set of accepted premises is to contain all facts accepted by the agent as earlier. The second one based on an initial set of rejected facts is intended to contain all facts rejected by the agent. A semantical description is provided by means of a Kripke–style structure of a special kind.

1. Syntax

Language $L_L$ is the language of the classical propositional logic augmented by a modal operator $L$.
Small letters $p, q, r, \ldots$ (with indices if necessary) are for propositional variables. At is the set of all propositional variables. Formulas are denoted by Greek letters, with indices if necessary. Logical connectives $\lor, \land, \Rightarrow, \Leftrightarrow$, are interpreted in a standard way, and $L\alpha$ is understood as “the agent believes $\alpha$”.

The set of all well–formed formulas $F_L$ of $L_L$ is the least set of logical expressions closed under the conditions:

1. $\text{At} \subseteq F_L$
2. If $\alpha, \beta \in F_L$ then $\alpha \land \beta, \alpha \Rightarrow \beta, \alpha \Leftrightarrow \beta, \neg \alpha, L\alpha \in F_L$.

Let $R$ be a set of inference rules and $X$ be a set of axioms. The well known definition of the consequence operator $Cn$ is

**Definition 1.**

- $Cn^0(R, X) = X$
- $Cn^{n+1}(R, X) = Cn^n(R, X) \cup \{\alpha \in F_L : \exists r \in R \exists \pi \subseteq Cn^n(R, X) ((\pi\alpha) \in r)\}$
- $Cn(R, X) = \bigcup_{n \in \mathbb{N}} Cn^n(R, X)$

we shall use $\alpha - \beta$ instead of $\neg (\alpha \Rightarrow \beta)$.

The inference rules are modus ponens (MP): $\alpha, \alpha \Rightarrow \beta, /\beta,\ Nece : \alpha/L\alpha$
- Rej : $\alpha - \beta, \beta/\alpha$

Ax denotes the set of all substitutions of axioms of the classical propositional logic in language $L_L$. We write $CnX$ instead of $Cn(\{\text{MP}\}, Ax \cup X)$ and, accordingly, $Cn\emptyset = Cn(\{\text{MP}\}, Ax)$ is the set of all classical propositional tautologies.

Thus all substitutions of these tautologies in the language $L_L$ are considered as classical tautologies. In this way we assume that formulas of the form $L\alpha$ may be treated as variables and valuated by 0, 1 as usual.

The set of all co- tautologies $Co$ in the language $L_L$ is $\{\alpha \in F_L : \neg \alpha \in Cn\emptyset\}$.

Operator $Cn'$ describes an inference process under sets of rejected formulas. Let $Cn'X = Cn(\{\text{Rej}\}, Co \cup X)$ and, consequently, $Cn'\emptyset = Cn(\{\text{Rej}\}, Co)$.

For instance, $(p \land q) \Rightarrow p \in Cn\emptyset$, hence, $(p \land q) - p \in Co \subseteq Cn'\emptyset \subseteq Cn'X$ by definition. Now if $p \in Cn'X$ then $p \land q \in Cn'X$ by Rej.

Let $-X = \{\alpha \in F_L : \alpha \notin X\}$, $\neg X = \{\neg \alpha \in F_L : \alpha \in X\}$, $LX = \{L\alpha \in F_L : \alpha \in X\}$, $\neg L - X = \{\neg L\alpha \in F_L : \alpha \notin X\}$.

Definitions 2 and 3 are standard definitions of an ae expansion and a stable set (cf. [2]).
Definition 2. The set of formulas $T$ is called an \textit{ae expansion} of the set of formulas $A$ iff $T = Cn(A \cup LT \cup \neg L - T)$.

Definition 3. The set of formulas $T$ is called a \textit{stable set} based on the set of formulas $A$ iff

1. $CnA \subseteq T$
2. $T$ is closed under $MP$ and $Nec$
3. if $\alpha \notin T$ then $\neg L\alpha \in T$.

It is known that every ae expansion of a set of premises $A$ is a stable set based on $A$.

In our approach let $T$ be a set of all facts accepted by an agent and $T'$ be a set of all facts rejected by her. Which facts are believed and which are disbelieved if there exist facts neither accepted nor rejected i.e. $T \cup T' \neq F_L$?

One of the possible approaches is:

- if $\alpha \in T$ then $L\alpha \in T$ and
- if $\alpha \in T'$ then $\neg L\alpha \in T$.

The latter is a modification of the commonly accepted condition, if $\alpha \notin T$ then $\neg L\alpha \in T$.

Let $A$ be an initial set of accepted formulas, and $B$–an initial set of rejected ones. The pair $(T, T')$ of suitable candidates for sets of all accepted and rejected facts, respectively, is searched for.

Definition 4. A pair of theories $(T, T')$ is called \textit{ae expansion} of initial sets of accepted and rejected premises $(A, B)$ iff

$$T = Cn(A \cup Am \cup LT \cup \neg LT')$$
$$T' = Cn'(B \cup \neg T \cup L\neg(\neg T'))$$

where $Am = \{L(\alpha \Rightarrow \beta) \Rightarrow (L\alpha \Rightarrow L\beta), \ L\alpha \Rightarrow LL\alpha, \ \neg L\alpha \Rightarrow L\neg L\alpha, \ L\alpha \Rightarrow \neg L\neg\alpha, \ L(L\alpha \Rightarrow \beta) \lor L(L\beta \Rightarrow \alpha)\}$. 

23
The shortening $\neg T, LT, \neg LT'$ have been already explained. Accordingly to the notion of a stable set a stable theory is defined.

**Definition 5.** A *stable theory* based on a pair of sets of premises $(A, B)$ is a pair of theories $(T, T')$ such that

1. $(CnA, Cn'B) \subseteq (T, T')$, $\subseteq$ is an usual inclusion under coordinates
2. $T$ is closed under $MP$ and $Nec$
3. $T'$ is closed under $Rej$
4. $Am \subseteq T$

and for any formula $\alpha \in F_L$

5. if $\alpha \in T$ then $\neg \alpha \in T'$
6. if $\alpha \in T'$ then $\neg L\alpha \in T$
7. if $\alpha \not\in T'$ then $L\neg \alpha \in T'$.

As in the previous case every ae expansion is a stable theory.

**2. Semantics**

Stable theories $(T, T')$ can be described in terms of Kripke–style structures. Let $\forall \alpha s \phi(s)$ state that for almost all $s \phi(s)$ holds (i.e. there holds $\phi(s)$ for all $s$ except a finite number at most).

**Definition 6.** A Kripke structure $\mathcal{M}$ is a relational structure $\mathcal{M} = (M, R, \models, V)$ where $M$ is a set of worlds. $R \subseteq M \times M$ is a binary relation. $V : M \to P(At)$ ($P(At)$ denotes the power set of $At$) is a function such that for each $s \in M$, $V(s) = \{p \in At : s \models p\}$. $\models \subseteq M \times F_L$ is a satisfiability relation such that for all $s, t \in M, p \in At, \alpha, \beta \in F_L$
\[ s \models p \iff p \in V(s) \]
\[ s \models \alpha \land \beta \iff (s \models \alpha \land s \models \beta) \]
\[ s \models \alpha \lor \beta \iff (s \models \alpha \lor s \models \beta) \]
\[ s \models \alpha \Rightarrow \beta \iff (\neg (s \models \alpha) \lor s \models \beta) \]
\[ s \models \neg \alpha \iff \neg (s \models \alpha) \]
\[ s \models L\alpha \iff \forall t \in M (sRt \Rightarrow t \models \alpha). \]

**Definition 7.** A *model* of a formula \( \alpha \in F_L \) is the set \( Mod(\alpha) = \{ s \in M : s \models \alpha \} \).

Let \( Ac(s) = \{ t \in M : sRt \} \), \( N(\alpha) = \{ s \in M : Mod(\alpha) \cap Ac(s) = \emptyset \} \).

**Definition 8.** \( M \) is the same Kripke structure.

a. \( \text{Th}(M) = \{ \alpha \in F_L : \forall s \ s \models \alpha \} \) is the set of all true formulas in \( M \).

b. \( \text{Th}'(M) = \{ \alpha \in F_L : \forall \alpha s \not\models (s \models \alpha) \} \) is the set of all rejected formulas in \( M \).

In the sequel we shall consider Kripke structures satisfying the following

**Assumption 9.**

1. \( |M| \geq \aleph_0 \)
2. \( \forall s | Ac(s)| \geq \aleph_0 \)
3. \( R \) is transitive i.e. \( \forall s, t, u \in M (sRt \land tRu \Rightarrow sRu) \)
4. \( R \) is Euclidean (cf. [1]) i.e. \( \forall s, t, u \in M (sRt \land sRu \Rightarrow tRu) \)
5. \( \exists \alpha \in At \exists s, t \in M (sRt \land s \models \alpha \land \not\models (t \models \alpha)) \)
6. \( \forall \alpha \in F_L (|Mod(\alpha)| \geq \aleph_0 \Rightarrow |n(\alpha)| < \aleph_0) \)
Definition 10.

a. A theory $T$ is consistent iff $T \neq F_L$.
b. A pair of theories $(T_1, T_2)$ is consistent iff both of them are consistent.

Now let us formulate

Theorem 11. Let $(A, B)$ be a pair of sets of premises and $\mathcal{M}$ be a Kripke structure satisfying Ass 9. If $(A, B) \subseteq (\text{Th}(\mathcal{M}), \text{Th}'(\mathcal{M}))$, then $\text{Th}(\mathcal{M}), \text{Th}'(\mathcal{M}))$ is a consistent stable theory based on $(A, B)$.

Remark. The case when $T$ and $T'$ are not “mirror images” of each other seems to be of great interest. Given “if $\alpha \in T'$ then $\neg \alpha \in T$”, then $T$ and $T'$ are just “mirror images”. And an additional theory $T'$ is not worth being defined nor searched. $T'$ is to help to formalize all the situations when an agent cannot make any clear decision: whether she believes some fact or not. The following example shows that there exists a stable theory $(T, T')$ consistent, and $T, T'$ are not “mirror images” of each other.

Example. The main concept is to set up a Kripke structure $\mathcal{M}$ such that $(\text{Th}(\mathcal{M}), \text{Th}'(\mathcal{M}))$ is not a pair of “mirror images” i.e. there is a formula $\alpha \in F_L$ such that $\alpha \in \text{Th}'(\mathcal{M})$ and $\neg \alpha \not\in \text{Th}(\mathcal{M})$.

Consider the set of all 0–1 valuations denoted by $2^{At}$ and its power set $P(2^{At})$. A variable $p \in At$ is fixed. Define $\models$ for all $q \in At$, $s \in P(2^{At})$ $s \models q$ iff $\forall v \in s \ v(q) = 1$.

Let $X = \{ s \in P(2^{At}) : s \models p \}$. $X \neq \emptyset$ since for $v' \in 2^{At}$ such that $v'(p) = 1$, and $v'$ is arbitrary for other variables, we have $s' = \{ v' \} \in X$. $s'' \in X$ is chosen.

The set $M = (P(2^{At}) - X) \cup \{ s'' \}$ is a carrier of $\mathcal{M}$.

For exactly one world $s''$ $s'' \models p$, that is $\forall_s s \not\models \neg p$, so $p \in \text{Th}'(\mathcal{M})$. Obviously, $\neg p \not\in \text{Th}(\mathcal{M})$ since not($s'' \models \neg p$) and $\neg p \not\in \text{Th}(\mathcal{M}) \cup \text{Th}'(\mathcal{M})$ is established.
Let \( R \subseteq M \times M \) be a binary relation satisfying Ass 9.

**References**


Anna Gomolińska
Institute of Mathematics
University of Warsaw
00–913 Warsaw 59, Banacha 2