# COMBINATION OF FORMAL LOGIC AND HEDGE ALGEBRA TO ESTIMATE THE DEGREE OF TRUST

NGUYEN MANH HUNG[1,2]

[1]*Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam*
[2]*UMI UMMISCO 209 (IRD/UPMC), Hanoi, Vietnam*
*nmhufng@yahoo.com, mhnguyen@ptit.edu.vn*

**Abstract.** Trust plays a more and more important role in many fields of computer sciences such as ambient intelligence, interaction human - machine, affective computing, etc. This paper firstly constructs a logical framework to represent the qualitative aspect of trust/distrust based on a logic of belief, a logic of time, and dynamic logic. Secondly, the paper constructs a hedge algebra and its extension with fuzzy logic to represent the quantitative aspect of trust/distrust and that of their cognitive factors. Then these two aspects are combined to estimate the degree of trust/distrust based on the degree of their cognitive factors.

**Keywords.** Modal logic, hedge algebra, fuzzy logic, trust, distrust

## 1. INTRODUCTION

Trust plays a more and more important role in many fields of computer sciences such as ambient intelligence, interaction human - machine, affective computing, etc. In this paper, an approach is introduced to combine the qualitative and the quantitative aspects of trust/distrust. At the qualitative cognitive level, some emotions based on the cognitive definition of trust [3] and distrust [4] are represented in a formal logic based on the logic of beliefs and choices as the one of Herzig and Longin [12] (a refinement from Cohen and Levesque [7]), the logic of time (introduced by Arthur Prior [18]), and dynamic logic introduced by Fischer and Ladner [8] and Harel et al. [11]. This part is closed to the work of Goudou et al. [10, 16], Bonnefon et al. [1, 2], and Nguyen [17]. At the quantitative level, the qualitative aspects of trust/distrust are represented by introducing a hedge algebra to represent several different (and fuzzy) degrees of trust/distrust as well as that of their cognitive factors. This hedge algebra is based on the hedge algebra of Nguyen and Wechler [5, 15], and its extension with fuzzy logic (Nguyen et al. [13, 14]).

As being discussed in our previous works [1,2,17], one of the main limitations of the logic proposed in these works is that they can only present the concept of trust/distrust in two bipolar values (*yes* or *no*). The main objective of this paper is to extend these logics by combining them with a hedge algebra and its extension with a fuzzy logic to present several degrees of trust/distrust.

This paper is organized as follows: Session 2 introduces the logical framework to represent trust/distrust based on their cognitive factors. Session 3 introduces the hedge algebra to represent many degree of trust/distrust and that of their cognitive factors. Session 4 combines the cognitive and qualitative aspects of trust/distrust. The final session is a conclusion and future works.

## 2. LOGICAL FRAMEWORK

In this section, a logical framework is introduced to cognitively represent the concept of trust. This logic is mainly based on the logic Bonnefon et al. [1,2] and Nguyen [17], the logic of beliefs and choices as the one of Herzig and Longin [12] (a refinement from Cohen and Levesque [7]), the logic of time (introduced by Arthur Prior [18]), and dynamic logic introduced by Fischer and Ladner [8] and Harel et al. [11].

### 2.1. Syntax

The syntactic primitives of our logic are as follows: a nonempty finite set of agents $AGT = \{i_1, i_2, \ldots, i_n\}$, a nonempty finite set of atomic events $EVT = \{e_1, e_2, \ldots, e_p\}$, and a nonempty set of atomic propositions $ATM = \{p_1, p_2, \ldots\}$. The variables $i, j, k \ldots$ denote agents. The expression $i_1{:}e_1 \in AGT \times EVT$ denotes an event $e_1$ intentionally caused by agent $i_1$ and $e_1$ is thus called an " action ". The variables $\alpha, \beta \ldots$ denote such actions. The language of our logic is defined by the following BNF :

$$\varphi := p \mid i{:}\alpha\text{-}do \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathtt{X}\varphi \mid \mathtt{X}^{-1}\varphi \mid \mathtt{G}\varphi \mid \mathtt{Bel}_i\varphi \mid \mathtt{Choice}_i\varphi \mid \mathtt{Grd}_I\varphi$$

where $p$ ranges over $ATM$, $i{:}\alpha$ ranges over $AGT \times EVT$, $i{:}\alpha\text{-}do$ ranges over $ATM$ for each $i{:}\alpha \in AGT \times EVT$, and $I \subseteq AGT$. The classical boolean connectives $\wedge$ (conjunction), $\rightarrow$ (material implication), $\leftrightarrow$ (material equivalence), $\top$ (tautology) and $\bot$ (contradiction) are defined from $\neg$ (negation) and $\vee$ (disjunction).

- $i{:}\alpha\text{-}do$ reads " agent $i$ is just about to perform the action $\alpha$ ";

- $\mathtt{X}\varphi$ reads " $\varphi$ will be true next instant ";

- $\mathtt{X}^{-1}\varphi$ reads " $\varphi$ was true at the previous instant ";

- $\mathtt{G}\varphi$ reads " henceforth, $\varphi$ is true ";

- $\mathtt{Bel}_i\varphi$ reads " agent $i$ believes that $\varphi$ is true ";

- $\mathtt{Choice}_i\varphi$ reads " agent $i$ prefers that $\varphi$ be true ";

- $\mathtt{Grd}_I\varphi$ reads " $\varphi$ is publicly grounded between the agents in group I ".

Abbreviations are defined as follows:

$$i{:}\alpha\text{-}done \stackrel{def}{=} \mathtt{X}^{-1}i{:}\alpha\text{-}do$$

$$\mathtt{Happens}_{i{:}\alpha}\varphi \stackrel{def}{=} i{:}\alpha\text{-}do \wedge \mathtt{X}\varphi$$

$$\mathtt{After}_{i{:}\alpha}\varphi \stackrel{def}{=} i{:}\alpha\text{-}do \rightarrow \mathtt{X}\varphi$$

$$\mathtt{Done}_{i{:}\alpha}\varphi \stackrel{def}{=} i{:}\alpha\text{-}done \wedge \mathtt{X}^{-1}\varphi$$

$$\mathtt{F}\varphi \stackrel{def}{=} \neg\mathtt{G}\neg\varphi$$

$$\mathtt{Goal}_i\varphi \stackrel{def}{=} \mathtt{Choice}_i\mathtt{FBel}_i\varphi$$

$$\mathtt{Intend}_i\alpha \stackrel{def}{=} \mathtt{Choice}_i\mathtt{F}i{:}\alpha\text{-}do$$

$$\mathtt{Able}_i\alpha \stackrel{def}{=} \neg\mathtt{After}_{i{:}\alpha}\bot$$

$$\texttt{Capacity}_i(\alpha,\varphi) \stackrel{def}{=} \texttt{After}_{j:\alpha}\varphi \wedge \texttt{Able}_j\alpha$$

$$\texttt{Possible}_i\varphi \stackrel{def}{=} \neg\texttt{Bel}_i\neg\varphi$$

$$\texttt{Awareness}_i\varphi \stackrel{def}{=} \texttt{X}^{-1}\neg\texttt{Bel}_i\varphi \wedge \texttt{Bel}_i\varphi$$

- $i{:}\alpha\text{-}done$ reads " agent $i$ has done action $\alpha$ ";

- $\texttt{Happens}_{i:\alpha}\varphi$ reads " agent $i$ is doing action $\alpha$ and $\varphi$ will be true next instant ";

- $\texttt{After}_{i:\alpha}\varphi$ reads " $\varphi$ is true after any execution of $\alpha$ by $i$ ";

- $\texttt{Done}_{i:\alpha}\varphi$ reads " agent $i$ has done action $\alpha$ and $\varphi$ was true at previous instant ";

- $\texttt{F}\varphi$ reads " $\varphi$ will be true in some future instants ";

- $\texttt{Goal}_i\varphi$ reads " agent $i$ has the goal (chosen preference) that $\varphi$ be true ";

- $\texttt{Intend}_i\alpha$ reads " agent $i$ intends to do $\alpha$ ";

- $\texttt{Able}_i\alpha$ reads " agent $i$ is capable to do $\alpha$ ";

- $\texttt{Capacity}_i(\alpha,\varphi)$ reads " agent $i$ is capable to do $\alpha$ to bring $\varphi$ ";

- $\texttt{Possible}_i\varphi$ reads " agent $i$ believes that it is possible $\varphi$ ";

- $\texttt{Awareness}_i\varphi$ reads " agent $i$ has just experienced that $\varphi$ is true ".

## 2.2.   Semantics

For temporal operators, we use a semantics based on linear time described by a sequence (or story) of time points. (This semantics is very close to CTL* [6]) A frame $\mathbb{F}$ is a 4-tuples $\{\mathbb{H}, \mathbb{B}, \mathbb{C}, \mathbb{G}\}$ where:

- $H$ is a set of stories that are represented as sequences of time points, where each time point is identified by an integer $z \in \mathbb{Z}$, a time point $z$ in a story $h$ is called a situation $< h, z >$;

- $\mathbb{B}$ is the set of all $\mathbb{B}[i]$ such that $\mathbb{B}[i](h, z)$ denotes the set of stories believed as being possible by the agent $i$ in the situation $< h, z >$;

- $\mathbb{C}$ is the set of all $\mathbb{C}[i]$ such that $\mathbb{C}[i](h, z)$ denotes the set of stories chosen by the agent $i$ in the situation $< h, z >$;

- $\mathbb{G}$ is the set of all $\mathbb{G}[I]$ such that $\mathbb{G}[I](h, z)$ denotes the set of stories which are publicly grounded in the group I of agents, in the situation $< h, z >$.

All the accessibility relations $\mathbb{B}[i]$ are serial. All the accessibility relations $\mathbb{G}[I]$ are serial, transitive and Euclidean (This is similar to the operator group grounding introduced by Gaudou et al. [9]). All the accessibility $\mathbb{C}[i]$ are serial. Moreover, we impose for every $z \in \mathbb{Z}$ that: if $h' \in \mathbb{B}[i](h, z)$ then $\mathbb{C}[i](h, z) = \mathbb{C}[i](h', z)$. It means that if an agent believes that the world h' is possible from the world h, then the set of his/her preference worlds from h and h' are the same. In other terms, the worlds an agent prefers and the ones that agent believes that s/he prefers are the same (briefly, the agent is conscious about his/her preferences, and s/he prefers what s/he believes that s/he prefers).

A model $\mathbb{M}$ is a couple $\{\mathbb{F}, \mathbb{V}\}$ where $\mathbb{F}$ is a frame and $\mathbb{V}$ is a function associating each atomic proposition $p$ with the set $\mathbb{V}(p)$ of couple $(h, z)$ where $p$ is true. Truth conditions are defined as follows:

$$\mathbb{M}, h, z \models p \text{ iff } (h, z) \in \mathbb{V}(p)$$
$$\mathbb{M}, h, z \models \mathtt{X}\varphi \text{ iff } \mathbb{M}, h, z+1 \models \varphi$$
$$\mathbb{M}, h, z \models \mathtt{X}^{-1}\varphi \text{ iff } \mathbb{M}, h, z-1 \models \varphi$$
$$\mathbb{M}, h, z \models \mathtt{G}\varphi \text{ iff } \mathbb{M}, h, z' \models \varphi \text{ for every } z' \geq z$$
$$\mathbb{M}, h, z \models \mathtt{Bel}_i\varphi \text{ iff } \mathbb{M}, h', z \models \varphi \text{ for every } (h', z) \in \mathbb{B}[i](h, z)$$
$$\mathbb{M}, h, z \models \mathtt{Choice}_i\varphi \text{ iff } \mathbb{M}, h', z \models \varphi \text{ for every } (h', z) \in \mathbb{C}[i](h, z)$$
$$\mathbb{M}, h, z \models \mathtt{Grd}_I\varphi \text{ iff } \mathbb{M}, h', z \models \varphi \text{ for every } (h', z) \in \mathbb{G}[I](h, z)$$

Other truth conditions are defined as usual.

## 2.3. Axiomatics

Due to our linear time semantics, the temporal operators satisfy the following principles:

$$i{:}\alpha\text{-}do \leftrightarrow \mathtt{X}i{:}\alpha\text{-}done$$
$$\mathtt{X}\varphi \leftrightarrow \neg\mathtt{X}\neg\varphi$$
$$\varphi \leftrightarrow \mathtt{X}\mathtt{X}^{-1}\varphi$$
$$\varphi \leftrightarrow \mathtt{X}^{-1}\mathtt{X}\varphi$$
$$\mathtt{G}\varphi \leftrightarrow \varphi \wedge \mathtt{X}\mathtt{G}\varphi$$
$$\mathtt{G}(\varphi \rightarrow \mathtt{X}\varphi) \rightarrow (\varphi \rightarrow \mathtt{G}\varphi)$$

$\mathtt{Bel}_i$ and $\mathtt{Choice}_i$ operators are defined in a normal modal logic plus (D) axioms. Thus, if $\square$ represents a $\mathtt{Bel}_i$ operator or $\mathtt{Choice}_i$ operator:

$$\frac{\varphi}{\square\varphi} \qquad\qquad (\mathrm{RN}_\square)$$

$$\square(\varphi \rightarrow \psi) \rightarrow (\square\varphi \rightarrow \square\psi) \qquad\qquad (\mathrm{K}_\square)$$

$$\square\varphi \rightarrow \neg\square\neg\varphi \qquad\qquad (\mathrm{D}_\square)$$

For example, axiom $\mathrm{D}_\square$ applied to operator $\mathtt{Bel}_i$ is $D_{\mathtt{Bel}_i}$, which is described as: $\mathtt{Bel}_i\varphi \rightarrow \neg\mathtt{Bel}_i\neg\varphi$.

$(\mathrm{RN}_\square)$ means that all theorems are believed (respectively: chosen) by every agent $i$; $(\mathrm{K}_\square)$ means that beliefs (respectively: choices) are closed under material implication for every agent $i$; $(\mathrm{D}_\square)$ means that beliefs (respectively: choices) of every agent $i$ are rational: they cannot be contradictory.

The $\mathtt{Bel}_i$ operators satisfy the following principles of introspection:

$$\mathtt{Bel}_i\varphi \leftrightarrow \mathtt{Bel}_i\mathtt{Bel}_i\varphi$$
$$\neg\mathtt{Bel}_i\varphi \leftrightarrow \mathtt{Bel}_i\neg\mathtt{Bel}_i\varphi$$

that means that agent $i$ is conscious of its beliefs and of its disbeliefs.

The following principle follows from the semantical constraint between belief accessibility relation and choice accessibility relation, and from axiom $(\mathrm{D}_\square)$ for $\mathtt{Bel}_i$:

$$\mathtt{Choice}_i\varphi \leftrightarrow \mathtt{Bel}_i\mathtt{Choice}_i\varphi$$
$$\neg\mathtt{Choice}_i\varphi \leftrightarrow \mathtt{Bel}_i\neg\mathtt{Choice}_i\varphi$$

that means that agent $i$ is conscious of its choices and of its dischoices.

The sound and complete axiomatization of $\mathtt{Grd}_I$ operator is defined as the one of common belief operator (also called mutual belief), which is closed to the operator described in Walton and Krabbe [19], also introduced by Gaudou et al. [9]:

$$\frac{\varphi}{\mathtt{Grd}_I\varphi} \tag{$\mathrm{RN}_{\mathtt{Grd}_I}$}$$

$$\mathtt{Grd}_I(\varphi \to \psi) \to (\mathtt{Grd}_I\varphi \to \mathtt{Grd}_I\psi) \tag{$\mathrm{K}_{\mathtt{Grd}_I}$}$$

$$\mathtt{Grd}_I\varphi \to \neg\mathtt{Grd}_I\neg\varphi \tag{$\mathrm{D}_{\mathtt{Grd}_I}$}$$

$$\mathtt{Grd}_I\varphi \to \mathtt{Grd}_I\mathtt{Grd}_I\varphi \tag{$4_{\mathtt{Grd}_I}$}$$

$$\neg\mathtt{Grd}_I\varphi \to \mathtt{Grd}_I\neg\mathtt{Grd}_I\varphi \tag{$5_{\mathtt{Grd}_I}$}$$

Axiom ($\mathrm{RN}_{\mathtt{Grd}_I}$) means that every tautology is public ground. Axiom ($\mathrm{K}_{\mathtt{Grd}_I}$) means that if $\varphi$ is publicly grounded in $I$ and that $\varphi$ implies $\psi$ then $\psi$ is also publicly grounded in $I$. Axiom ($\mathrm{D}_{\mathtt{Grd}_I}$) means that the set of grounded informations is consistent: it can not be the case that both $\varphi$ and $\neg\varphi$ are simultaneously grounded. The positive introspection axiom ($4_{\mathtt{Grd}_I}$) and negative introspection axiom ($5_{\mathtt{Grd}_I}$) account for the public character of $\mathtt{Grd}_I$. From these collective awareness results: if $\varphi$ has (resp. has not) been grounded then it is established that $\varphi$ has (resp. has not) been grounded.

Linear time semantics entail the following principles:

$$\mathtt{G}\varphi \to \mathtt{After}_{i:\alpha}\varphi \tag{$\mathrm{G}_\varphi$}$$

$$\mathtt{Happens}_{i:\alpha}\varphi \to \mathtt{After}_{j:\beta}\varphi \tag{$\mathrm{Happens}_\varphi$}$$

$$\mathtt{After}_{i:\alpha}\varphi \leftrightarrow \neg\mathtt{Happens}_{i:\alpha}\neg\varphi \tag{$\mathrm{After}_\varphi$}$$

Axiom ($\mathrm{G}_\varphi$) describe the relationship between time and action: if henceforth $\varphi$ is true then after every action $\alpha$ of every agent $i$, $\varphi$ will be true. (Note that the converse is not valid: it is possible that $\varphi$ be true after every action $\alpha$ of every agent $i$ performed in a situation $< h, z >$, and that $\varphi$ be false at time $z' > z$.)

As time is linear, actions are deterministic on a given history. Thus, axiom ($\mathrm{Happens}_\varphi$) reads: if agent $i$ is just about to perform $\alpha$ after what $\varphi$ will be true, then after every performance of every action $\beta$ by every agent $j$, $\varphi$ will be true. In other words, if action $\alpha$ leads to a time point where $\varphi$ is true, then every action performed by every agent leads to this time point.

Finally, axiom ($\mathrm{After}_\varphi$) means that $\mathtt{After}_{i:\alpha}$ and $\mathtt{Happens}_{i:\alpha}$ operators are dual operators. This property is fair with respect to dynamic logic [11].

## 3.   HEDGE ALGEBRA

In this section, a hedge algebra is introduced to represent the degree of trust. This hedge algebra is based on the hedge algebra theory of Nguyen and Wechler [5, 15] and its extension with fuzzy logic [13, 14].

This hedge algebra is a set $AX = \{X, G, H, \leqslant\}$, where:

- $X$: the basic set of $AX$ $X = \{$0 (not at all), very low, low, little low, average, little high, high, very high, 1 (absolute)$\}$

- $G$: the set of generators $G = \{little, moreorless(\text{neutral element}), much\}$

- $H$: the set of linguistic hedges (also called *unary operators*, or *hedge operations*) $h = \{very\}$

- $\leqslant$: is an ordered relation between any two elements of $X$: $0 \leqslant$ *very little* $\leqslant$ *little* $\leqslant$ *more or less* $\leqslant$ *much* $\leqslant$ *very much* $\leqslant 1$

This algebra provides a set of hedges which could be associated with any cognitive factor of trust/distrust. For instance, with the concept of *trust*, we could have many variations:

- Agent $i$ does *not* trust agent $j$ *at all*

- Agent $i$ trusts *very little j*

- Agent $i$ trusts *little j*

- Agent $i$ trusts *more or less j*

- Agent $i$ trusts *much j*

- Agent $i$ trusts *very much j*

- Agent $i$ trusts *absolutely j*

These variations have different degrees. The degree of each depends on the *fuzziness measure* [14] of the linguistic associated with it. In this algebra, the fuzziness of each element is defined as indicated in Figure 1.
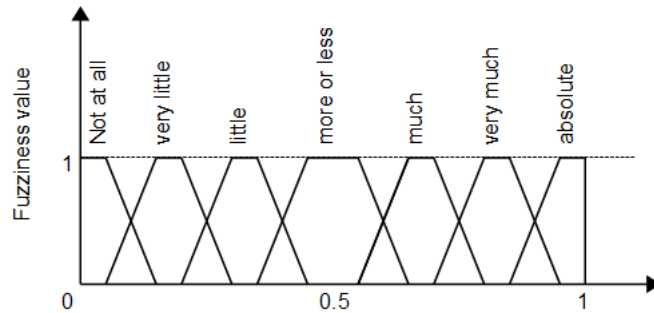


Figure 1: The fuzziness of linguistic hedges associated with trust/distrust and their cognitive factors

In this model, it defines how the degree of a variation of trust/distrust (or its cognitive factors) is identical to the fuzziness measure of the linguistic hedge associated with it

$$d(< hedge >< object >) = fm(< hedge >) \tag{1}$$

where $< hedge > \in AX$ is any linguistic hedge in our hedge algebra. $< object >$ could be trust/distrust or a cognitive factor of them. $fm(x)$ is the *fuzziness measure* of the linguistic hedge $x$.

For instance, the degree of *trust much* is the same as the fuzziness measure of the hedge *much* ($d(\text{trust much}) = fm(much)$), or the degree of *trust very much* is the same as the fuzziness measure of the hedge *very much* ($d(\text{trust very much}) = fm(\text{very much})$).

## 4.  COMBINATION OF COGNITIVE AND QUALITATIVE ASPECTS OF TRUST

In this section, the formalization of trust and distrust is presented basing on the cognitive definition of Castelfranchi and colleagues [3, 4].

### 4.1.  Trust

The concept of trust is formalized basing on Castelfranchi and Falcone's definition [3] of trust in action which says that agent $i$ trusts agent $j$ to ensure $\varphi$ by performing action $\alpha$ if and only if (i) agent $i$ desires to achieve $\varphi$, and agent $i$ expects that: (ii) agent $j$ is able to perform action $\alpha$ to achieve $\varphi$; and (iii) agent $j$ has the intention to do action $\alpha$.

However, these three factors are only necessary conditions, but not sufficient ones. So here it is needed to add another condition for trust: an agreement between trustor and trustee that the trustee will perform such an action ($\texttt{Grd}_I trustee : \alpha\text{-}do$), where $I = \{trustor, trustee\}$. Accordingly the concept of trust is formalized as:

$$\texttt{Trust}_{i,j}(\alpha, \varphi) = \texttt{Goal}_i\varphi \wedge \texttt{Bel}_i\texttt{Capacity}_j(\alpha, \varphi)$$
$$\texttt{Bel}_i\texttt{Intend}_j\alpha \wedge \texttt{Bel}_i\texttt{Grd}_{\{i,j\}}j{:}\alpha\text{-}do$$

The effects of each cognitive factor on trust are under consideration:

- The first factor, $\texttt{Goal}_i\varphi$, has many possible values corresponding to different degree of agent $i$'s preference for $\varphi$: the more agent $i$ prefers for $\varphi$, the more agent $i$ trusts $j$ in doing $\alpha$ to achieve $\varphi$, and vice versa.

- The second factor, $\texttt{Capacity}_j(\alpha, \varphi)$, has many possible values corresponding to different degree of agent $i$'s belief in the ability of $j$ to do $\alpha$ to achieve $\varphi$: the more agent $i$ believes that $j$ can do $\alpha$ to achieve $\varphi$, the more agent $i$ trusts $j$ in doing $\alpha$ to achieve $\varphi$, and vice versa.

- The third factor, $\texttt{Intend}_j\alpha$, also has many possible values corresponding to different degree of agent $i$'s belief in the intention of $j$ to do $\alpha$: the more agent $i$ believes that $j$ intents to do $\alpha$, the more agent $i$ trusts $j$ in doing $\alpha$ to achieve $\varphi$, and vice versa.

- The last factor, $\texttt{Grd}_{\{i,j\}}j{:}\alpha\text{-}do$, has only two possible values of *true* or *false* (or *yes/no, 1/0*).

Let $d(x) \in [0, 1]$, $x$ could be trust, distrust, or one of their cognitive factor, be the degree of $x$. So the degree of *trust* is estimated as:

$$d(\texttt{Trust}_{i,j}(\alpha, \varphi)) = f_{taf}(d(\texttt{Goal}_i\varphi), d(\texttt{Bel}_i\texttt{Capacity}_j(\alpha, \varphi)), \qquad (2)$$
$$d(\texttt{Bel}_i\texttt{Intend}_j\alpha), d(\texttt{Bel}_i\texttt{Grd}_{\{i,j\}}j{:}\alpha\text{-}do))$$

where $f_{taf}$ is a *trust-aggregation-function (TAF)* mapping: $f_{taf} : [0, 1]^4 \rightarrow [0, 1]$ which satisfies the following conditions:

$$(1). f_{taf}(a, b, c, 0) = 0$$
$$(2). 0 \leq f_{taf}(a, b, c, 1) \leq 1$$
$$(3). f_{taf}(a_1, b, c, 1) \leq f_{taf}(a_2, b, c, 1) \text{ if } a_1 \leq a_2$$
$$(4). f_{taf}(a, b_1, c, 1) \leq f_{taf}(a, b_2, c, 1) \text{ if } b_1 \leq b_2$$
$$(5). f_{taf}(a, b, c_1, 1) \leq f_{taf}(a, b, c_2, 1) \text{ if } c_1 \leq c_2$$

Table 1: Some variations of trust and distrust degree based on their contribution factors

| $\mathtt{Goal}_i\varphi$ | $\mathtt{Capacity}_j(\alpha,\varphi)$ | $\mathtt{Intend}_j\alpha$ | $\mathtt{Grd}_{\{i,j\},j:\alpha\text{-}do}$ | $\mathtt{Trust}_{i,j}(\alpha,\varphi)$ | $\mathtt{DisTrust}_{i,j}(\alpha,\varphi)$ |
|---|---|---|---|---|---|
| not at all | not at all | not at all | yes | not at all | more or less |
| not at all | absolute | not at all | yes | little | not at all |
| not at all | absolute | absolute | yes | much | not at all |
| ... | ... | ... | ... | ... | ... |
| very little | very little | very little | yes | very little | more or less |
| very little | very much | very little | yes | little | little |
| very little | very much | very much | yes | more or less | little |
| ... | ... | ... | ... | ... | ... |
| little | little | little | yes | little | more or less |
| little | much | little | yes | more or less | more or less |
| little | much | much | yes | more or less | little |
| ... | ... | ... | ... | ... | ... |
| more or less | more or less | more or less | yes | more or less | more or less |
| more or less | very little | very little | yes | little | much |
| more or less | very much | very little | yes | more or less | little |
| more or less | very much | very much | yes | much | little |
| ... | ... | ... | ... | ... | ... |
| much | much | much | yes | much | more or less |
| much | much | little | yes | more or less | more or less |
| much | little | little | yes | more or less | much |
| ... | ... | ... | ... | ... | ... |
| very much | very much | very much | yes | very much | more or less |
| very much | very little | very little | yes | little | very much |
| very much | very much | very little | yes | much | more or less |
| ... | ... | ... | ... | ... | ... |
| absolute | not at all | not at all | yes | little | absolute |
| absolute | very much | very little | yes | much | much |
| absolute | much | little | yes | much | much |
| absolute | absolute | absolute | yes | absolute | little |

For instance, when using $f_{taf}(a,b,c,d) = d * \dfrac{a+b+c}{3}$, some degree of *trust* depends on that its cognitive factors are indicated in Table 1. For example, Federer is a professional tennis player, his coach at the U.S. Open Champions is Stefen. Let's consider the trust of Stefen on Federer before the final match of the U.S. Open Champions with the action is $\alpha = $ *"win the final match"*, and the desired status is $\varphi = $ *"the champion of the U.S. Open 2015"*:

- Stefen desires *absolutely* that Federer becomes the champion of U.S. Open 2015:

  $d(\mathtt{Goal}_{Stefen}\varphi) = absolute.$

- There is a common ground between Stefen and Federer that Federer will play the final match:

  $d(\mathtt{Bel}_{Stefen}\mathtt{Grd}_{\{Stefen,Federer\}}Federer{:}\alpha\text{-}do) = yes.$

- Stefen believes *absolutely* that Federer also intends to win the final match:

  $d(\mathtt{Bel}_{Stefen}\mathtt{Intend}_{Federer}\alpha) = absolutely.$

- But Stefen believes that the ability to win the final match for Federer is *50-50*:

  $d(\texttt{Bel}_{Stefen}\texttt{Capacity}_{Federer}(\alpha, \varphi)) = $ *more or less.*

- So, Stefen trusts *very much* Federer to win the final match to becomes the U.S. Open champion:

  $d(\texttt{Trust}_{Stefen,Federer}(\alpha, \varphi)) = $ *very much.*

## 4.2. Distrust

The definition of distrust given by Castelfranchi et al. [4] also adopted. This definition says that agent $i$ distrusts agent $j$ to ensure $\varphi$ by performing action $\alpha$ if and only if (i) agent $i$ desires to achieve $\varphi$, and agent $i$ believes that at least one of these conditions is fulfilled: (ii) agent $j$ is not in the capacity to do action $\alpha$, or (iii) agent $j$ is not able to do $\alpha$ to achieve $\varphi$. Accordingly this concept is formalized as:

$$\texttt{DisTrust}_{i,j}(\alpha, \varphi) = \texttt{Goal}_i\varphi \wedge (\texttt{Bel}_i\neg\texttt{Capacity}_j(\alpha, \varphi) \vee \texttt{Bel}_i\neg\texttt{Intend}_j\alpha))$$

The effects of each cognitive factor on distrust are under consideration:

- The first factor, $\texttt{Goal}_i\varphi$, is the same in the concept of trust: It has many possible values corresponding to different degree of agent $i$'s preference for $\varphi$: the more agent $i$ prefers for $\varphi$, the more agent $i$ distrusts $j$ in doing $\alpha$ to achieve $\varphi$, and vice versa.

- The second factor, $\neg\texttt{Capacity}_j(\alpha, \varphi)$, has many possible values corresponding to different degree of agent $i$'s belief in the ability of $j$ to do $\alpha$ to achieve $\varphi$: the more agent $i$ believes that $j$ can not do $\alpha$ to achieve $\varphi$, the more agent $i$ distrusts $j$ in doing $\alpha$ to achieve $\varphi$, and vice versa.

- The third factor, $\neg\texttt{Intend}_j\alpha$, also has many possible values corresponding to different degree of agent $i$'s belief in the intention of $j$ to do $\alpha$: the more agent $i$ believes that $j$ does not intent to do $\alpha$, the more agent $i$ distrusts $j$ in doing $\alpha$ to achieve $\varphi$, and vice versa.

So the degree of *distrust* is estimated as:

$$
\begin{aligned}
d(\texttt{DisTrust}_{i,j}(\alpha, \varphi)) &= f_{daf}(d(\texttt{Goal}_i\varphi), d(\neg\texttt{Capacity}_j(\alpha, \varphi)), d(\neg\texttt{Intend}_j\alpha)) \qquad (3)\\
&= f_{daf}(d(\texttt{Goal}_i\varphi), 1 - d(\texttt{Capacity}_j(\alpha, \varphi)), 1 - d(\texttt{Intend}_j\alpha))
\end{aligned}
$$

where $f_{daf}$ is a *distrust-aggregation-function (DAF)* mapping: $f_{daf} : [0,1]^3 \rightarrow [0,1]$ which satisfies the following conditions:

$$
\begin{aligned}
&(1). 0 \leq f_{daf}(a,b,c) \leq 1\\
&(2). f_{daf}(a_1,b,c) \leq f_{daf}(a_2,b,c) \text{ if } a_1 \leq a_2\\
&(3). f_{daf}(a,b_1,c) \leq f_{daf}(a,b_2,c) \text{ if } b_1 \leq b_2\\
&(4). f_{daf}(a,b,c_1) \leq f_{daf}(a,b,c_2) \text{ if } c_1 \leq c_2\\
&(5). f_{daf}(a,b_1,c_1) \leq f_{daf}(a,b_2,c_2) \text{ if } min(b_1,c_1) \leq min(b_2,c_2)
\end{aligned}
$$

For instance, when using $f_{daf}(a,b,c) = \dfrac{a + min(b,c)}{2}$, some degree of *distrust* depends on that its cognitive factors are indicated in Table 1. For example, let's consider a situation in which, a boss asks

his new secretary to prepare an important report for the meeting today ($\alpha$ = "*prepare the report*", and $\varphi$ = "*having the good report*"):

- A boss desires *absolutely* that he has the report for the meeting:

  $d(\texttt{Goal}_{boss}\varphi) = absolute.$

- In spite the boss believes that the new secretary intends to write the report, he believes *more or less* that the she is not able to write such the important report:

  $d(\neg\texttt{Capacity}_{secrectary}(\alpha, \varphi)) = more\ or\ less.$

- So, the boss distrusts *very much* his new secretary in writing the report to have a good report:

  $d(\texttt{DisTrust}_{boss,secretary}(\alpha, \varphi)) = very\ much.$

The results extracted in the Table 1 also indicate that it is not the case that $d(\texttt{Trust}_{i,j}(\alpha, \varphi)) = 1 - d(\texttt{DisTrust}_{i,j}(\alpha, \varphi))$. In other words, These results indicate that *trust* and *distrust* are not a polar pair of concepts. This is consistent with the argumentation of Castelfranchi et al. [3, 4].

## 5.   CONCLUSION

This paper firstly introduces a logical framework for the representation of the cognitive concept of trust/distrust based on a logic of belief, a logic of time, and a dynamic logic. Secondly, it introduces a hedge algebra and its extension with fuzzy logic to represent the degree (qualitative aspect) of trust/distrust and that of their cognitive factors. These two aspects of trust/distrust are then combined to estimate the degree of emotions based on the degree of their cognitive factors. The paper estimates seven different degrees (*not at all, very little, little, more or less, much, very much, absolute*) of the two concepts (*trust, distrust*) based on the degree of their cognitive factors.

For future research, the representation, in both cognitive and qualitative aspects, of some emotions and the relation between emotions and trust/distrust is some perspectives in future works.

## REFERENCES

[1] J.-F. Bonnefon, D. Longin, and M. H. Nguyen, "A logical framework for trust-related emotions," *Electronic Communications of the EASST*, vol. 22, pp. 1–16, 2009.

[2] ——, "Relation of trust and social emotions: A logical approach," in *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on*, vol. 2.   IET, 2009, pp. 289–292.

[3] C. Castelfranchi and R. Falcone, "Social trust: A cognitive approach," in *Trust and Deception in Virtual Societies*, C. Castelfranchi and Y. H. Tan, Eds.   Dordrecht: Kluwer Academic Publishers, 2001, pp. 55–90.

[4] C. Castelfranchi, R. Falcone, and E. Lorini, "A non-reductionist approach to trust," in *Computing with Social Trust*, J. Goldbeck, Ed.   Berlin: Springer, 2008, pp. 45–72.

[5] N. Cat Ho and W. Wechler, "Hedge algebras: An algebraic approach to structure of sets of linguistic truth values," *Fuzzy Sets Systems*, vol. 35, no. 3, pp. 281–293, May 1990.

[6] E. M. Clarke, E. A. Emerson, and A. P. Sistla, "Automatic verification of finite-state concurrent systems using temporal logic specifications," *ACM Transactions on Programming Languages and Systems*, vol. 8, no. 2, pp. 244–263, 1986.

[7] P. R. Cohen and H. J. Levesque, "Intention is choice with commitment," *Artificial Intelligence*, vol. 42, pp. 213–261, 1990.

[8] M. Fischer and R. Ladner, "Propositional dynamic logic of regular programs," *Journal of Computer and System Sciences*, vol. 18, no. 2, pp. 194–211, 1979.

[9] B. Gaudou, A. Herzig, and D. Longin, "A Logical Framework for Grounding-based Dialogue Analysis," in *International Workshop on Logic and Communication in Multi-Agent Systems (LCMAS), Edinburgh, Scotland, UK, 01/08/2005*, ser. Electronic Notes in Theoretical Computer Science (ENTCS), W. van der Hoek, A. Lomuscio, E. de Vink, and M. Wooldridge, Eds., vol. 157, no. 4. Elsevier, 2006, pp. 117–137.

[10] ——, "Group belief and grounding in conversation," in *Language, Cognition, Interaction*, A. Trognon, Ed. Presses Universitaires de Nancy, 2009.

[11] D. Harel, D. Kozen, and J. Tiuryn, *Dynamic Logic.* MIT Press, 2000.

[12] A. Herzig and D. Longin, "C&L intention revisited," in *Proceedings of Int. Conf. of knowledge representation and reasoning KR'04.* Morgan Kaufmann, 2004, pp. 527–535.

[13] N. C. Ho and H. V. Nam, "An algebraic approach to linguistic hedges in zadehś fuzzy logic," *Fuzzy Sets Systems*, vol. 129, no. 2, pp. 229–254, Jul. 2002.

[14] N. C. Ho and N. V. Long, "Fuzziness measure on complete hedge algebras and quantifying semantics of terms in linear hedge algebras," *Fuzzy Sets Syst.*, vol. 158, no. 4, pp. 452–471, Feb. 2007.

[15] N. C. Ho and W. Wechler, "Extended hedge algebras and their application to fuzzy logic," *Fuzzy Sets Systems*, vol. 52, no. 3, pp. 259–281, Dec. 1992.

[16] E. Lorini and A. Herzig, "A logic of intention and attempt," *Synthese (2008)*, vol. 163, no. 1, pp. 45–77, 2008.

[17] M. H. Nguyen, "A logical framework for trust-related emotions: formal and behavioral results," Ph.D. dissertation, University of Toulouse III, 2010.

[18] A. N. Prior, *Time and Modality.* Oxford: Clarendon Press, 1957.

[19] D. N. Walton and E. C. Krabbe, *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning.* NY: State University of New-York Press, 1995.