

二次元ト ラスト 表現に基づく アブセンス・オブ・トラストの3分類

河辺 義信¹ 小泉 佑揮² 大木 哲史³ 西垣 正勝³ 長谷川 享² 五郎丸 秀樹⁴

概要: トラストの有無に関する概念として、アブセンス・オブ・トラスト (absence of trust) がある。我々は、ファジィ多項目並列評定法 (FCR 法) に基づく二次元的なトラスト値の領域内において、アブセンス・オブ・トラストがどの場所にあたるかを検討する。Marshらは、信頼不足 (lack of trust) と呼ばれる概念についての議論を行っているが、アブセンス・オブ・トラストについては議論がない。一方、システムセーフティの分野では「システムを安定運用する上で、ここまでの不具合ならば許容できる」という、不具合の許容範囲が扱われる。この考え方と同様に、我々は、一次元的なトラスト値の上で、不具合が許容できる範囲を信頼不足、許容できない範囲をアブセンス・オブ・トラストと定義する。本研究ではさらに、この結果を拡張して、FCR法に基づく二次元的なトラスト値についても、アブセンス・オブ・トラストの領域を定義する。FCR法では、対象に対する肯定的評価と否定的評価が独立に扱われる。これにより、アブセンス・オブ・トラストの領域を、3種類 (矛盾・混乱型、無関心型、嫌悪型) に分類する。

キーワード: トラスト, ファジィ理論, アブセンス・オブ・トラスト, 信頼不足

On Classifying Absence of Trust Notions Based on 2D Trust Representation

YOSHINOBU KAWABE¹ YUKI KOIZUMI² TETSUSHI OHKI³ MASAKATSU NISHIGAKI³ TORU HASEGAWA²
HIDEKI GOROMARU⁴

Abstract: The concept of the absence of trust is well-known but not well-formalized. This study defines the region of the absence of trust in the two-dimensional trust area. By extending a lack of trust, we define the absence of trust. Even if a trustee does not have enough faith, we may work with the trustee as the shortage of trust is very low. However, if the trust shortage is massive and unacceptable, an observer must give up coping with the trustee. We regard this as the absence of trust. We first define the absence of trust in the one-dimensional trust theory. And then, we extend the result to the case with the two-dimensional trust representation, where we deal with positive and negative evaluations on a trustee independently. We classify the absence of trust into three notions: AoT confusional, AoT ignorant, and distrust.

Keywords: on-line trust, fuzzy logic, absence of trust, lack of trust

1. はじめに

災害時においてメッセージや送信者のトラストを評価することの重要性が増している。たとえば、文献 [6] および [14] では、大阪府北部地震や西日本豪雨のときに Twitter 上で交換されたメッセージ (ハッシュタグ「#災害」「#災害救助」がついたもの) を対象に、どのメッセージが緊急性の高い情報かを検討している。文献 [6] および [14] では、

¹ 愛知工業大学 情報科学部
Department of Information Science, Aichi Institute of Technology
² 大阪大学大学院
Graduate School of Information Science and Technology, Osaka University
³ 静岡大学大学院
Graduate School of Science and Technology, Shizuoka University
⁴ 千葉工業大学 社会システム科学部
Department of Risk Science in Finance and Management, Chiba Institute of Technology

メッセージを信頼できるものとそうでないものに二分しているが、そうした分類が常に適切にできるとは限らない。これに対し Marsh と Dibben は、メッセージのトラストがどの程度かを表す指標（トラスト値）を導入している [9]。この指標は -1 から 1 の間の実数値で、本研究では、一次元的トラスト値と呼ぶ。さらに、この指標を使って Marsh らは、トラストの概念を「トラスト」「ディストラスト」「アントラスト」「ミストラスト」に分類している。

Marsh らのトラスト概念の分類では、トラスト値が最も大きい値（ 1 ）のときは「トラスト」の状態であり、最も小さい値（ -1 ）のときは「ディストラスト」の状態である。これより、トラストの対義語はディストラストであると思われる。しかし一方で、Lewicki らは、トラストとディストラストは独立に扱われるべき概念だと指摘している [7]。また、トラストを量で計れるものと仮定し、トラスト（があること）の対義語は「アブセンス・オブ・トラスト（つまり、トラストがないこと）」だとする議論もある [11], [21]。

アブセンス・オブ・トラストに似た概念として「信頼不足」という考え方があり、これについては、Marsh らの研究（文献 [9] の 3.3 節）にも議論がある。しかしながら、アブセンス・オブ・トラストについては、定式化は行われていない。そこで本研究では、信頼不足やアブセンス・オブ・トラストの定式化を検討する。

本研究では、まず Marsh らの一次元的なトラスト値をもとにトラストの量を考え、アブセンス・オブ・トラストを議論する。一次元的なトラスト値は単純でわかりやすいが、「対象を信頼しているが、じつは同時に不信感もある」という矛盾・混乱したトラストの状態や「会ったことのない人に対しては、信頼も不信もない」といった無関心（無関連）の状態を扱うことができない。そこで本稿では、一次元的なトラスト理論におけるトラスト量の議論を、信頼と不信を同時に扱う二次元的なトラスト表現 [5], [16] に拡張する。二次元的なトラスト表現は、印象形成に関する理論（小田の FCR 法 [17], [19]）に基づいており、矛盾や無関心といった心的状態を扱うことができる。本研究では二次元的トラスト表現の結果に基づき、アブセンス・オブ・トラストを、3 種類に分類できることを示す。

2. 準備: Marsh らの一次元的なトラスト値

2.1 トラスト値の計算

Marsh と Dibben は、文献 [9] において、4 種類のトラストの概念（トラスト、ディストラスト、アントラスト、ミストラスト）を示している。これらの概念を定式化する上で、 $[-1, 1]$ の値をとるふたつの関数

- 一般トラスト $T_x(y)$: 観測者 x の被信頼者 y に対するトラスト値
- 状況トラスト $T_x(y, \alpha)$: 観測者 x の被信頼者 y に対する状況 α に関するトラスト値

が用いられる。ここで、状況トラスト $T_x(y, \alpha)$ は

$$T_x(y, \alpha) = U_x(\alpha) \times I_x(\alpha) \times \widehat{T_x(y)}$$

で定義される。ただし、 $U_x(\alpha)$ (値域は $[-1, 1]$) は観測者 x の状況 α に関する効用関数、 $I_x(\alpha)$ (値域は $[0, 1]$) は観測者 x の状況 α に対する重要度、 $\widehat{T_x(y)}$ (値域は $[-1, 1]$) は観測者 x の被信頼者 y に対する一般トラストの値の「見積り」である（詳しくは、文献 [8] を参照されたい）。 $U_x(\alpha)$ 、 $I_x(\alpha)$ および $\widehat{T_x(y)}$ は、それぞれ $[0, 1]$ 、 $[0, 1]$ および $[-1, 1]$ の値である。

例として、文献 [8] に示されている「二人のエージェントとふたつの家具の例」を紹介する。いま、エージェント A と B がいて、それぞれは、家具一つを運びたいとする。各エージェントは、単独では家具を運ぶことはできない。家具を動かすには二人が必要である。各エージェントは、仕事を完了するためには、もう一方のエージェントと協力することを考えなければならない。しかしながら、あるエージェントが協力しないという選択をしたとしても、制裁などはないとする。この状況を、 α というラベルで表すことにする。ここで、

- エージェント B はエージェント A にとって比較的信頼がおける（その評価値を $\widehat{T_A(B)} = 0.63$ とする）。
- エージェント A は、この状況 α を、ある程度重要だと評価している（ $I_A(\alpha) = 0.5$ で表すとする）。
- エージェント A の家具は自分のものであり、 A は、この部屋から家具を動かす効用はとて高いと考えている（これを $U_A(\alpha) = 0.8$ で表す）。

のとき、状況トラストの値は、 $T_A(B, \alpha) = 0.8 \times 0.5 \times 0.63 = 0.252$ で与えられる。

2.2 トラスト・ディストラスト・アントラスト・ミストラスト

ここでは、Marsh と Dibben によるトラストの 4 分類について述べる。もし、 $T_x(y)$ が協力の閾値と呼ばれる値よりも高ければ、被信頼者 y は観測者 x から信頼されている。また同様に、 $T_x(y, \alpha)$ が閾値よりも高い時、状況 α について被信頼者 y は x から信頼されている。これが「トラスト」の状況である。

ディストラストは、被信頼者が与えられた状況下でどれだけ「わざと足を引っ張るか」を示す指標である。すなわち、観測者 x が被信頼者 y に不信感を持っているというのは、 y が与えられた状況下 α において最悪の（もしくは、少なくとも最善とは言えないような）結果をもたらすだろうと x が予想しているということである。これは、

$$T_x(y, \alpha) < 0 \implies Distrust(x, y, \alpha)$$

として定式化される。つまりディストラストは、 $T_x(y, \alpha)$ が負の場合として定められる。

アントラストは、トラスト値が正ではあるものの、協力の閾値 $CT_x(y, \alpha)$ (「これ以上のトラスト値を持っていないと、このエージェントとは協力したくない」という値) よりも低い状態を指す。これは、

$$(T_x(y, \alpha) > 0 \wedge T_x(y, \alpha) < CT_x(y, \alpha)) \\ \implies Untrust(x, y, \alpha)$$

で表される。アントラストの状態にあるということは、トラストできる状態(トラスト値が1付近)でもなく、ディストラストの状態(トラスト値が負)でもない。すなわちこれは、相手あるいは事柄に対して、信頼して良いのか、決めきれない状態を表している。この状況下では観測者 x は被信頼者 y と協力する状態には至らない。なお閾値 $CT_x(y, \alpha)$ は

$$CT_x(y, \alpha) = \frac{Risk_x(\alpha)}{Comp_x(y, \alpha) + \widehat{T_x(y)}} \times I_x(\alpha) \quad (1)$$

で与えられる。ただし、 $Risk_x(\alpha)$ は x が認識している状況 α に関するリスク値を、 $Comp_x(y, \alpha)$ は x が認識している「 y の α に対する適合度」の値を表す。この式より、リスク ($Risk_x(\alpha)$) やことの重要性 ($I_x(\alpha)$) が大きいときには、被信頼者は高いトラスト値を持たないと協力の閾値に達しない。一方で、被信頼者がうまくことに当たると観測者が信じている ($Comp_x(y, \alpha)$ が大きい) 場合や被信頼者の一般トラストの見積り ($\widehat{T_x(\alpha)}$) が大きい場合は、トラスト状態に至るための閾値は比較的低くなる。以下では、 $CT_x(y, \alpha)$ を CT と略記する。

このほか文献 [9] では、当初の信頼を(良い方向へのそれも含め)裏切られた状態である「ミストラスト」を扱っている。これは、上述の三つの性質と異なり、時間経過に伴うトラスト値の変化に関する性質である。トラストの遷移については、swift trust [10], [15] も知られている。これらの性質に対する詳しい分析は、Primiero の文献 [12] (ディストラストとミストラスト) や文献 [13] (トラストとディストラスト) が詳しい。

3. 信頼不足とアブセンス・オブ・トラスト

3.1 トラストの量

「トラストがあること」の対義語が「トラストがないこと」だと考えるとき、我々は暗黙的にトラストを量で計っていることになる。つまり、被信頼者が十分な量の信頼度を持っている場合には、その被信頼者はトラストの状態にあり、持っているトラストの量が不十分なときには、信頼不足の状態にある。そして、もし被信頼者がトラスト量を全く持っていないならば、その被信頼者は、アブセンス・オブ・トラストの状態にある。このことから、トラスト量について考えれば、

トラスト > 信頼不足 > アブセンス・オブ・トラスト

といった関係にある。

本研究では、トラストの量について、次の三つの条件を仮定する。

- (1) トラスト量は、0 から 1 の間の実数で計るとする。
- (2) Marshらによるトラスト値 v がトラストの領域にある、つまり、協力の閾値よりも v が高いならば、対応するトラスト量 $q(v)$ も決められた閾値よりも高いとする。
- (3) トラスト値 v がアブセンス・オブ・トラストの領域にあるときには、対応するトラスト量 $q(v)$ は 0 である。この条件を満たす簡単な例として、Marshらの一次元的なトラスト値 $v \in [-1, 1]$ に関して $q(v) = \max(0, v)$ と定める定義が考えられる。ただし、協力の閾値 CT に対応するトラスト量に関する閾値としては、 CT 自身を使うことにする。

3.2 信頼不足

被信頼者が十分なトラスト量を持っていないとき、その信頼者は信頼不足の状態にある。この場合では、対応する (Marshらの) トラスト値は、協力の閾値よりも低い値である。また、不足分のトラスト量は、比較的小さいと考えられる。

ここでは、信頼の閾値 $CT_x(y, \alpha)$ の定義式 (1) から、信頼不足の状態にある被信頼者が、どのようなときに信頼できるようになるのかを考えたい。式 (1) の右辺を見ると、

- リスク $Risk_x(\alpha)$ の値が、現状のそれよりも、低い値であったとき、
- ことの重要度の見積もり $I_x(\alpha)$ が、より低いとき、
- 被信頼者 y が状況 α に対して、より上手に対応できるとみなせるとき、すなわち、値 $Comp_x(y, \alpha)$ をより高い値に設定できるとき、もしくは、
- 観測者 x から見た被信頼者 y のトラスト $\widehat{T_x(y)}$ が、より高く見積もれるとき

には閾値が低くなり、相対的にトラスト量の不足分が小さくなる。つまり、トラストの量が不足しているために、その時点では協力するほどにはない状態であっても、不足分がわずかであれば、上述の条件を改善することでトラストできるようになる可能性がある。トラストの不足量が十分小さく、改善できる範囲内にあるとき、我々は、これを信頼不足の状態と考えることにする。

上述に基づき、協力の閾値までのトラスト量不足の許容範囲を考え、(十分小さな) 正の値 ϵ で表すことにしよう。以下では値 ϵ を最大許容量と呼ぶ。また、信頼不足のトラスト値の集合を $LoT(\epsilon)$ とする。ただし、

$$LoT(x) = \{v | v \in T \wedge 0 < CT - q(v) < \epsilon\} \quad (2)$$

であり、 T は Marshらの一次元的なトラスト値の集合 $[-1, 1]$ である。4 節では、 $LoT(x)$ の定義を、二次元的なトラスト値の集合を用いて再定義する。

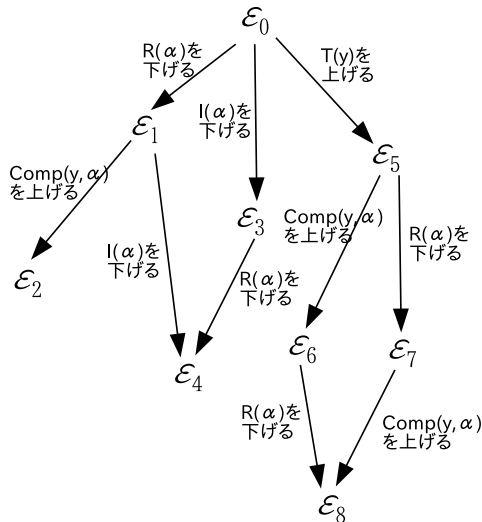


図1 順序 \preceq の一例

3.3 アブセンス・オブ・トラスト

前節では、トラストの不足量が改善できる範囲内か、という考え方に基づいて、信頼不足を考えた。これに似た考え方は、システムセーフティの分野にも見られる。コンピュータシステムに、たとえ小さな問題点があったとしても、安定運用の上では支障のない場合がある。言い換えれば、運用者は問題の発生を許容する場合がある、ということである。しかしながら、故障が多くなりすぎると、安定運用できなくなる。つまり、許容範囲を超える不具合は受け入れられない、ということである。

トラスト量の不足があまりに大きいとき、もはや被信頼者は、多少の条件を改善する程度では「トラスト」の状態には達しえない。これが、アブセンス・オブ・トラストの状態だと考えるのが自然である。以下では、トラスト不足の量が最大許容量を超えるような、トラスト値の集合を考える。「トラスト量の不足分がこの値よりも大きいときは、もはやトラスト量についての改善は行えない」という値を ζ とする。この値 ζ を回復不可能量と呼び、 $\zeta \geq \epsilon$ とする。また、アブセンス・オブ・トラストの領域を

$$AoT(x) = \{v \mid v \in T \wedge q(v) < CT - x\} \quad (3)$$

を用いて、 $AoT(\zeta)$ とする。これより、許容不可能なトラスト値 $v \in T$ については $CT - q(v) > \zeta \geq \epsilon > 0$ である。

3.4 閾値に対する段階的操作

3.2 節では、協力の閾値 $CT_x(y, \alpha)$ を低減するため、 $Risk_x(\alpha)$ 、 $I_x(\alpha)$ 、 $Comp_x(y, \alpha)$ および $\widehat{T_x}(y)$ の値を変更した。変更の際には、これらの設定値をすべて同時には変えず、段階を追って協力の閾値を低減することが考えられる。このような段階的な操作を行うときには、いくつかの仮の信頼不足量の列 $\epsilon_1, \epsilon_2, \dots$ と、それに対応するトラスト値の集合の列 $LoT(\epsilon_1), LoT(\epsilon_2), \dots$ を考えることができ

る。値 $\epsilon_1, \epsilon_2, \dots$ は、最大許容量と回復不可能量の候補となる。以下では、これらの候補となる値を比較し、最大許容量と回復不可能量の定義を検討する。

仮の信頼不足量 $\epsilon_1, \epsilon_2, \dots$ に 0 を加えた集合を DC とする。また、 DC 上の半順序 \preceq を定める*1。 DC および \preceq について、次の条件を仮定する。

- (1) 任意の $a \in DC$ について $0 \preceq a$ である。
- (2) 任意の $a, b \in DC$ について、 $a \preceq b$ ならば $a \leq b$ である (ただし、 \leq は通常の実数間の順序)。

図1は、順序 \preceq の一例である ($\epsilon_0 = 0$ とする)。ここでは、順序関係 $\epsilon_0 \preceq \epsilon_1 \preceq \epsilon_4$ に注目する。 ϵ_0 は、協力の閾値 CT からの設定値の変更がないことを表す。 ϵ_1 は、閾値が CT から $CT - \epsilon_1$ に低減した状態を表す。図1にあるように、この低減は、リスクに関する設定値の操作によるものとする。さらに ϵ_4 は、上述のリスクの設定値の操作に加えて、重要度を調整して閾値を $CT - \epsilon_4$ まで下げることによる。

閾値の低減を行って本来トラストの状態にないトラスト値を許容する操作は、コンピュータシステムの例で言うところの、安定運用のために不具合を許容範囲内で受け入れることに対応する。値 ϵ_2 は、設定値 $Risk_x(\alpha)$ と $Comp_x(y, \alpha)$ に関する低減を行った結果の信頼不足量で、これ以上の許容はできない限界値である (そのため、 $\epsilon_2 \preceq x$ となる x は ϵ_2 のみである)。値 ϵ_2 は $Risk$ と $Comp$ を操作して得たが、別の操作をした場合には、許容できる限界値が変わることも考えられる。たとえば、 ϵ_4 は $Risk$ と I を、 ϵ_8 は $Risk$ 、 $Comp$ および \widehat{T} を操作することによる、閾値の低減の許容限界である。

DC の要素のうち、閾値の低減の許容限界となる値を集めた集合を $UbDC$ とする。つまり、

$$UbDC = \{b \mid a \preceq b \implies a = b\}$$

である。3.2 節の最大許容量 ϵ と 3.3 節の回復不可能量 ζ を定めるとき、 $\epsilon = \zeta = \max UbDC$ とする定義が考えられる。ただし \max は、 $UbDC$ を通常の実数上の大小関係 \leq で比べたときの最大値である。この考え方では、 $\epsilon = \zeta = \max UbDC = \max DC$ であり、すべての DC の要素は許容可能である。つまり、閾値に対するいくつかの低減基準のうち、ある基準を満たしていなくても、より緩やかな別の基準を満たしていれば許容する。この意味で、この定義は楽観的と言える。また、トラスト量 $q(v)$ が $CT - \max UbDC$ を下回るトラスト値 v は、つねに受け入れ不可能であり、アブセンス・オブ・トラストの領域にある。

一方で、 $\epsilon = \zeta = \min UbDC$ とする、より厳しい基準に

*1 関係 \preceq は DC 上で定義され、設定値 $Risk_x(\alpha)$ 、 $I_x(\alpha)$ 、 $Comp_x(y, \alpha)$ および $\widehat{T_x}(y)$ の集合を比較するものではないが、 \preceq はこれらの設定値に対する段階的な変更結果を反映するように意図して定めるものとする。

よる定義も考えられる。この場合は、閾値に対する低減基準のうち、もっとも厳しい基準に合わせている。この場合は、 $CT - \min UbDC$ を下回るトラスト値はアブセンス・オブ・トラストとなる。

最後に、最大許容量 ϵ と回復不可能量 ζ を異なる値とする（つまり $\zeta > \epsilon$ とする）方法も考えられる。たとえば、 $\epsilon = \min UbDC$ および $\zeta = \max UbDC$ とする定義などが考えられる。より詳細な検討は、今後の課題である。

4. 二次元的トラストモデルへの拡張

4.1 FCR 型トラストモデル

$[-1, 1]$ の要素を扱う Marsh らのトラスト値（一次元的なトラスト値の表現法）は単純でわかりやすいが、「相手を基本的に信頼しているが、じつは同時に不信感も抱いている」といった矛盾・混乱したトラストの状態や「知らない人に対しては信頼も不信もない」といった無関心（無関連）の状態を扱うことができない。これに対して我々は、文献 [5] において、二次元的トラスト表現法（FCR 型モデル）を導入した。この表現法では、信頼の度合い $Trust$ と不信の度合い $DisTrust$ を

$$Trust = DisTrust = \{v \mid 0 \leq v \leq 1\}$$

と定め、二次元的なトラスト値（「観測点」とも呼ぶ）を、 $Trust \times DisTrust$ の要素として与える。

図 2 に、二次元的なトラスト表現を示す。この図では、観測点 $(1, 0) \in Trust \times DisTrust$ は信頼の度合いが高く（1）、一方で不審の度合いは低い（0）。つまり、 $(1, 0)$ は、完全に信頼されている、トラストの状態を表す。この状態は、従来のトラスト値 1 に対応する。一方で、観測点 $(0, 1)$ は完全なディストラストの状態、従来型のトラスト値では -1 にあたる。我々は、図 2 内の観測点 $(1, 0)$ と $(0, 1)$ を結ぶ線が、従来型のトラスト値の集合に対応すると考えた。この集合は、ファジィ理論 [3], [17] では無矛盾集合と呼ばれる。

無矛盾集合上の任意の点 (t, d) では、 $t + d = 1$ が成り立つ。このとき、矛盾度（または、矛盾・無関連指数）と呼ばれる値 $c = t + d - 1$ は、つねに 0 である。任意の観測点 (t, d) および矛盾度 c について、 $c = 1$ は完全な矛盾を、 $c = -1$ は全くの無関心を表す。また、 $c = 0$ は無矛盾な状況である。これより、つねに矛盾度が 0 の場合を扱う従来型の一次元的なトラスト値の理論では、矛盾や無関心（または無関連）といった心的状態を考慮していないことがわかる。

矛盾や無関心を扱うため、以下では、3.2 節や 3.3 節で用いた集合 T （以前の節では、 T は Marsh らの一次元的なトラスト値の集合 $[-1, 1]$ だった）として、二次元的トラスト値の集合 $T = Trust \times Distrust$ を扱うことにする。また、トラスト量を求める関数として、 $q(t, d) = \max(0, t - d)$

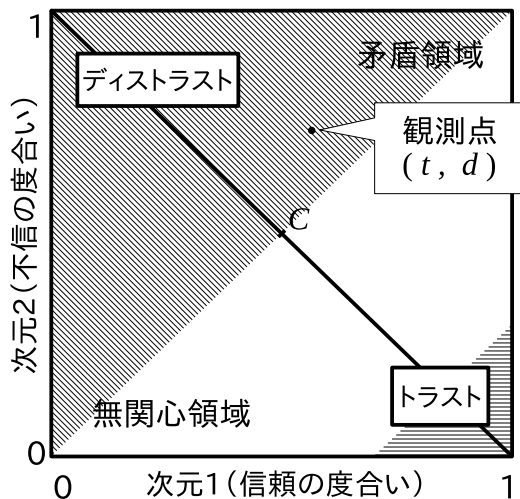


図 2 二次元的トラスト表現

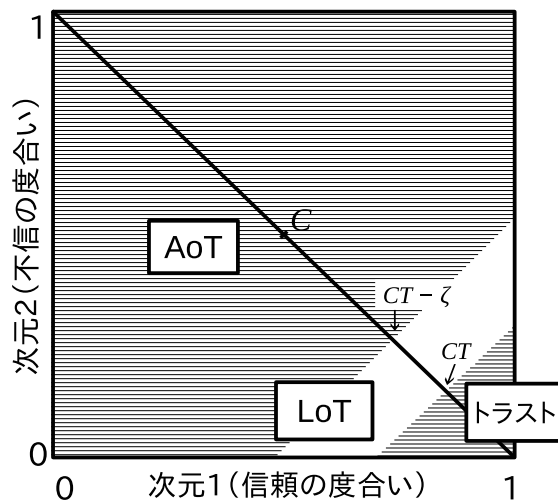


図 3 トラスト，信頼不足，アブセンス・オブ・トラスト

を用いる。ただし、 $(t, d) \in T (= Trust \times Distrust)$ は任意の観測点である。再定義された関数 $q(t, d)$ について、 $0 \leq q(t, d) \leq 1$ である。また、

- (t, d) が図 2 のディストラストの領域にあるならば、 $q(t, d) = 0$ 。
- (t, d) がディストラストの領域にないならば、値 $q(t, d)$ は、 (t, d) から図 2 の対角線に推薦を引いた交点の「次元 1」の値に対応する（付録の A.1 節も参照されたい）が成り立つ。もし $q(t, d)$ の値が正ならば、「正味の」トラスト値 $t - d$ は、信頼の度合い t から不信の度合い d を引くことで計算できる。これは直感にも合致すると考える。

4.2 アブセンス・オブ・トラストの分類

上述では、集合 T を二次元的トラスト値の集合 $Trust \times DisTrust$ として扱い、式 (2) や式 (3) の集合 LoT および AoT を定め直した。図 3 は、二次元的なトラスト値の集合をもとに、信頼不足の領域とアブセンス・オブ・トラ

ストの領域を示したものである（ここでは、 $\epsilon = \zeta$ としている）。トラストの領域が図の右下の三角形の領域で、アブセンス・オブ・トラストは左上の領域である。そのふたつの間にあるのが、信頼不足の領域である。図3から、矛盾度の値などを用いて、アブセンス・オブ・トラストを3種類に分けることができる。

- (1,1) 付近の観測点: この観測点は、アブセンス・オブ・トラストの領域にあるが、同時に矛盾の領域にもある。実際、矛盾・無関連指数 $c = t + d - 1$ の値は正の高い値を示す。これは、信頼の度合い t と不信の度合い d の両方が高いときである。ファジィ理論 [3], [17] の結果によれば、情報過多が、こうした矛盾した状況を引き起こす。ここで言う情報過多とは、被信頼者に関する判断材料があまりにも多い状態を指す。そうした状況下では、ある情報によって被信頼者への信頼の度合いが上がるかもしれないが、別の情報によっては不信の度合いが上がるかもしれない。こうした「賛否両論」の状況によって矛盾・混乱が起こり、観測者は、被信頼者を本当にトラストしてよいのか、わからなくなる。この種のアブセンス・オブ・トラストは、矛盾・混乱型と呼ぶべきものである。
- (0,0) 付近の観測点: この観測点は、やはりアブセンス・オブ・トラストの領域にあるが、こちらは矛盾・無関連指数 c の値が負で低い値のときである。このときは、信頼の度合い t と不信の度合い d の両方が低い。この状況は、情報不足により起こる。つまり、被信頼者についての十分な判断材料がないため、トラストできるかわからない、ということである。これは、無関心によるアブセンス・オブ・トラストと呼ぶべきものである。
- (0,1) 付近の観測点: これは、典型的なディストラストである。被信頼者に対する肯定的評価（信頼できる、好き）が低く、否定的評価（信頼できない、嫌い）が高いため、トラスト量はゼロである。これは、嫌悪型のアブセンス・オブ・トラストと言える。

矛盾度 $c = t + d - 1$ の絶対値 $|c|$ は、評価の曖昧さを表す指標である。 $|c|$ の値が0に近いとき（つまり、図2において、観測点対角線付近にあるとき）は信頼の度合いの評価と不信の度合いの評価に一貫性があり、 $|c|$ の値が1に近いとき（観測点対角線の距離が大きいとき）は評価は曖昧である。アブセンス・オブ・トラストの「矛盾・混乱型」と「無関心」については、観測点対角線から遠く、値 $|c|$ は高くなる。つまり、トラスト量が少ない原因は、被信頼者に対する評価を適切に行えないためと考えられる。一方で、「嫌悪型」のときは、値 $|c|$ は低い（図2の左上にある観測点は、必然的に、対角線との距離が小さくなる）。そのため、被信頼者に対する評価を適切に行った上で、それでもなおかつ、トラスト量がない状態と言える。

5. むすび

本稿では、トラスト量とアブセンス・オブ・トラストの定義について検討した。具体的には、まずトラストの量についての定義を与え、さらに信頼不足の考え方をもとに、アブセンス・オブ・トラストを定めた。また二次元的なトラスト値を用いてこれを拡張し、ファジィ理論の矛盾度などに基いてアブセンス・オブ・トラストの概念を、「矛盾・混乱型」「無関心型」および「嫌悪型」の三つに分類した。

本研究の二次元的なトラストモデルでは、トラスト量の定義にファジィ理論の「逆転項目平均法」と呼ばれる手法 [3], [17] を用いている。これは、統合値と呼ばれる値（これは、信頼の度合い t と不信の度合い d から求められる「正味のトラスト値」 $t - d$ に対応）を計算する方法のひとつである。定義の単純さから、本研究では逆転項目平均法を使ったが、文献 [3], [17] には、ほかにもいくつかの統合値の求め方が示されている。これらをもとに、より優れたトラスト量の定義を検討することは、今後の課題である。なお、付録A.1節に統合値の例を示す。また本稿では、アブセンス・オブ・トラストを議論するため、最大許容量 ϵ と回復不可能量 ζ を扱った。本研究では $\epsilon = \zeta$ の場合を考えたが、 $\zeta > \epsilon$ とする場合についても、今後検討したい。

最後に、本研究の結果をソーシャルメディアを用いた災害時通信システム [2], [4] に応用することを挙げておきたい。二次元的トラスト値を用いれば、ソーシャルメディア上のメッセージの内容や送受信者の行動について、矛盾や無関心まで扱って分析できるようになると期待される。しかし、実問題に二次元的トラスト値を適用するには、特有の難しさもある。具体的には、評価者がメッセージや送受信者などの対象に対して、どれだけの評価（肯定的評価および否定的評価）を持っているかを、観測点 (t, d) として正確に表現できなければならない。しかしこれは、人の持つ印象についての情報であり、扱いに難しさがある。二次元的トラスト値を議論するベースとなったファジィ理論の手法である小田のFCR法 [17], [19] の研究でも、肯定的基準と否定的基準のそれぞれを正確に把握することの難しさが指摘されている。また、被験者からより正確な心的情報を得るための新たな手法として、IR法 [18] と呼ばれる手法が研究されている（FCR法とIR法は、一定の条件のもと、相互に観測点の変換が可能である）。IR法では、真偽の度合いと、その真偽値に対する評価者の確信度（すなわち、曖昧さの度合い）を指定する。IR法で扱う「真偽の度合い」が、本研究におけるトラストの量に対応すると考えている。IR法で得られている結果をもとに、トラスト量を正確に計測する手法を考案し、二次元的トラスト値に変換することで、実問題に対する適用が可能になると考えている。これも、今後の課題に挙げられる。

謝辞 本研究は、NICT 委託研究「スマートコミュニティを支える高信頼ネットワーク構成技術の研究開発(副題: ソーシャルメディア時代の高信頼災害時通信の研究開発, 管理番号 19303)」により行われた。

参考文献

- [1] Krassimir T. Atanassov. *Intuitionistic Fuzzy Sets: Theory and Applications*. Physica-Verlag GmbH, Heidelberg, Germany, Germany, 1st edition, 2010.
- [2] Jiachen Chen, Mayutan Arumathurai, Xiaoming Fu, and K. K. Ramakrishnan. CNS: Content-oriented notification service for managing disasters. In *Proceedings of ACM Conference on Information-Centric Networking*, pp. 122–131. ACM, 2016.
- [3] Jianming Deng, Tetsuhisa Oda, and Motohide Umano. Fuzzy logical operations in the two-dimensional hyper logic space concerning the fuzzy-set concurrent rating method. *Journal of Japan Association for Management Systems*, Vol. 17, No. 2, pp. 33–42, 2001.
- [4] Mohammad Jahanian, Yuxuan Xing, Jiachen Chen, K. K. Ramakrishnan, Hulya Seferoglu, and Murat Yuksel. The evolving nature of disaster management in the internet and social media era. In *2018 IEEE International Symposium on Local and Metropolitan Area Networks, LANMAN 2018, Washington, DC, USA, June 25-27, 2018*, pp. 79–84, 2018.
- [5] Yoshinobu Kawabe, Yuki Koizumi, Tetsushi Ohki, Masakatsu Nishigaki, Toru Hasegawa, and Tetsuhisa Oda. On trust confusional, trust ignorant, and trust transitions. In *Trust Management XIII - 13th IFIP WG 11.11 International Conference, IFIPTM 2019, Copenhagen, Denmark, July 17-19, 2019, Proceedings*, pp. 178–195, 2019.
- [6] Megumi Kondo, Shuji Nishikawa, Osamu Uchida, and Keisuke Utsu. Text analysis of tweets with rescue request hashtags posted during 2018 Japan floods. In *6th International Conference on Information and Communication Technologies for Disaster Management, ICT-DM 2019, Paris, France, December 18-20, 2019*, 2019.
- [7] R. J. Lewicki, D. J. B. McAllister, and R. J. Bies. Trust and distrust: New relationships and realities. *Academy of Management Review*, Vol. 23, pp. 438–458, 1998.
- [8] Stephen Paul Marsh. Formalising trust as a computational concept. Technical report, University of Stirling, 1994.
- [9] Stephen Marsh and Mark R. Dibben. Trust, untrust, distrust and mistrust – an exploration of the dark(er) side. In *Proceedings of the Third International Conference on Trust Management, iTrust'05*, pp. 17–33, Berlin, Heidelberg, 2005. Springer-Verlag.
- [10] Debra Meyerson, Karl E. Weick, and Roderick M. Kramer. *Swift Trust and Temporary Groups in Trust in Organizations: Frontiers of Theory and Research*. SAGE, 1995.
- [11] Yuko Murayama. Issues in disaster communications. *Journal of Information Processing*, Vol. 22, No. 4, pp. 558–565, October 2014.
- [12] Giuseppe Primiero. A calculus for distrust and mistrust. In Sheikh Mahbub Habib, Julita Vassileva, Sjouke Mauw, and Max Mühlhäuser, editors, *Trust Management X*, pp. 183–190, Cham, 2016. Springer International Publishing.
- [13] Giuseppe Primiero, Franco Raimondi, Michele Bottone, and Jacopo Tagliabue. Trust and distrust in contradictory information transmission. *Applied Network Science*, Vol. 2, p. 12, 2017.
- [14] Chenjie Song and Hiroyuki Fujishiro. Toward the automatic detection of rescue-request tweets: analyzing the features of

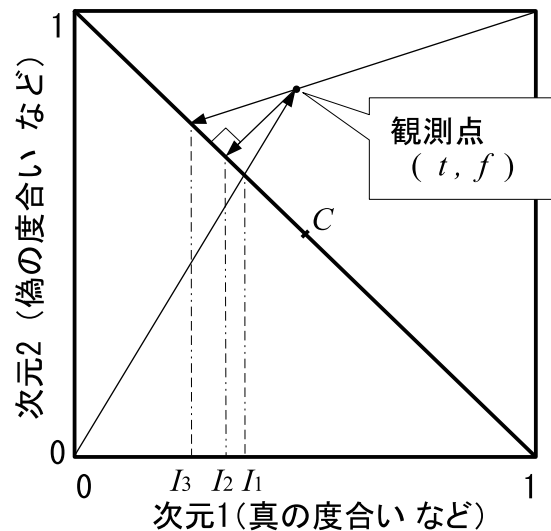


図 A.1 統合値の計算

data verified by the press. In *6th International Conference on Information and Communication Technologies for Disaster Management, ICT-DM 2019, Paris, France, December 18-20, 2019*, 2019.

- [15] Jessica Wildman, Marissa Shuffler, Elizabeth Lazzara, Stephen Fiore, and Shawn Burke. Trust development in swift starting action teams: A multilevel framework. *Group & organization management*, Vol. 37, No. 2, pp. 137–170, 2012.
- [16] 河辺 義信, 小田 哲久. FCR 法を用いたラスト値の二次元表現について. 第 28 回ソフトサイエンス・ワークショップ 講演論文集, pp. 35–38, 2018.
- [17] 小田 哲久. ファジィ多項目並列評定法の特徴について. 日本経営システム学会誌, Vol. 12, No. 1, pp. 23–32, 1995.
- [18] 小田 哲久. 評定尺度法へのファジィ論的アプローチ — FCR 法, IR 法について —. *心理学評論*, Vol. 56, No. 1, pp. 67–83, 2013.
- [19] 小田 哲久. 人間工学のための計測手法 第 3 部: 心理計測と解析 (3). *人間工学*, Vol. 51, No. 5, pp. 293–303, 2015.
- [20] 向殿政男, 菊池浩明. ファジィ・インターバル論理の提案. *日本ファジィ学会誌*, Vol. 2, No. 2, pp. 209–222, 1990.
- [21] 村山 優子. 情報科学におけるトラストの研究課題 (招待講演). CSS 2019 コンピュータセキュリティシンポジウム 予稿集, 2019.
- [22] 日本ファジィ学会. *ファジィとソフトコンピューティングハンドブック*. 共立出版, September 2000.

付 録

A.1 FCR 法の統合値について

以下では、二次元的トラスト値の議論のベースとなった、小田の FCR 法について、統合値と呼ばれる値の計算方法に絞って説明する。ファジィ推論で統合値を求める操作は、二次元的トラスト値を Marsh らの一次元的なトラスト値に変換する操作に対応している [5]。

複数の評定尺度の値からファジィ推論を行うために、FCR 法では、多次元多値論理空間としての HLS モデル (Hyper Logic Space model) が用いられる。たとえば、図 A.1 は、「真 (横軸. 以下 T)」と「偽 (縦軸. 以下 F)」による 2 次元空間を表している。また、観測点 (真理値) は

点 $(t, f) \in T \times F$ として表される。ここで $t, f \in [0, 1]$ であり、 t と f は独立である。なお、小田のモデルと類似の考え方で、観測点を「真の度合い」と「偽の度合い」の対として表すモデルには、Atanassov の A-IFS モデル [1] や向殿と菊池の拡張区間真理値モデル [20] などが知られている。

HLS 上の観測点 (t, f) が与えられたとき、並列的に与えられた尺度 (t および f) から「統合値」と呼ぶ代表的な値が算出される。統合値の計算方法は何種類かが考案されており、解析対象に応じて使い分けられる。主なものを以下に示す。

- 単純スコア法: t に 1 点, f に 0 点を割り当て、重み付け平均を計算する。ただし、分母が 0 のときは別に定める。統合値 I_1 は

$$I_1(t, f) = \begin{cases} 0.5 & t + f = 0 \text{ のとき} \\ \frac{t}{t + f} & \text{それ以外} \end{cases}$$

で与えられる。

- 逆転項目平均法: 否定的尺度 f の否定を $1 - f$ とし、肯定的尺度 t との平均値をとる。すなわち統合値 I_2 を

$$I_2(t, f) = \frac{t + (1 - f)}{2}$$

と定める。

- 逆スコア法: 肯定的尺度 t の否定 $1 - t$ と否定的尺度 f の否定 $1 - f$ に対して、単純スコア法の式を適用する。すなわち、統合値 I_3 は

$$I_3(t, f) = \begin{cases} 0.5 & t + f = 2 \text{ のとき} \\ \frac{1 - f}{2 - t - f} & \text{それ以外} \end{cases}$$

で与えられる。

各統合値は、図 A.1 に示すように、図的な計算によって

- $I_1(t, f)$: 観測点 (t, f) と点 $(0, 0)$ を結ぶ直線と、図中の対角線の交点の「次元 1」の値
- $I_2(t, f)$: 観測点 (t, f) から図中の対角線に垂線を下ろした点における「次元 1」の値
- $I_3(t, f)$: 観測点 (t, f) と点 $(1, 1)$ を結ぶ直線と、図中の対角線の交点の「次元 1」の値

として求められる。本稿では、逆転項目平均法に基づく統合値 $I_2(t, f)$ を用いた。

なお、統合値と並ぶ重要な値が、本稿でも用いた矛盾度 (矛盾・無関連指数) である。矛盾度の定義についても、数種類が知られている。詳しくは、文献 [22] を参照されたい。