

Comparing Two-Dimensional Trust Representations : FCR- and Subjective-Logic-Based Approaches

メタデータ	言語: jpn 出版者: 公開日: 2020-06-17 キーワード (Ja): キーワード (En): 作成者: 河辺, 義信, 小泉, 佑揮, 大木, 哲史, 西垣, 正勝, 長谷川, 亨, 小田, 哲久 メールアドレス: 所属:
URL	http://hdl.handle.net/10297/00027525

二次元的トラスト表現法の比較 — FCR法と Subjective Logic

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 小田 哲久⁴

概要: 本稿では、小田によるファジィ多項目並列評定法 (FCR法) と Jøsang の Subjective Logic を、二次元的なトラスト値の表現法としての観点から比較する。両手法とも、「信頼度」と「不信度」の2軸に基づき、トラスト値を二次元的に表現できる。また、両手法とも対象に対する無関心（「知らない人物に対しては、信頼も不信もない」など）を扱える。しかし一方で、矛盾したトラスト（「ある人物に対して、基本的には信頼しているが、じつは心の中では同時に不信感も持っている」など）を扱えるのは、FCR法に基づくアプローチのみである。本稿では、両手法の共通点と相違点について検討する。とくに、FCR法と Subjective Logic では、時系列を追ったトラスト値の計算について、異なる前提をおいていると考えられる。トラスト形成の典型的なシナリオ例を考えながら、これについても議論する。

キーワード: トラスト, FCR法, ファジィ理論, Subjective Logic, 無関心と矛盾

Comparing Two-Dimensional Trust Representations — FCR- and Subjective-Logic-Based Approaches

YOSHINOBU KAWABE¹ YUKI KOIZUMI² TETSUSHI OHKI³ MASAKATSU NISHIGAKI³ TORU HASEGAWA²
TETSUHISA ODA⁴

Abstract: This paper compares two approaches for trust computations — Oda’s fuzzy-set concurrent rating (FCR) method and Jøsang’s subjective logic. In both approaches, a trust value is represented as a pair of the degrees of “trust” and “distrust,” and we can deal with the ignorant trust (e.g. “Since I have never met him, I have no impression of him.”). However, we cannot deal with the confusional trust (e.g. “Although he can basically be trusted, in some cases he is not trustworthy”) in the subjective-logic-based approach, while we can do in the FCR-based approach. In this paper, we discuss the difference of the approaches. Especially, the FCR- and subjective-logic-based approaches have different assumptions about trust computations. Showing typical scenarios about trust computations, we also discuss the assumptions.

Keywords: on-line trust, FCR method, fuzzy logic, subjective logic, irrelevance/contradiction

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 愛知工業大学 経営学部
Department of Business Administration, Aichi Institute of Technology

1. はじめに

インターネット上で交換されるメッセージにはさまざまなものがあり、疑わしい内容を持つものも多く存在する。また、メッセージの送受信者についても、信頼度はさまざまである。そのため、「その送信者からメッセージを受け取ってもよいのか」「その受信者（あるいは中継者）に情報を渡しても差し支えないのか」などを、ユーザは適切に判断する必要がある。

メッセージや送受信者が信頼できるかを数理的に評価するため、Marsh と Dibben [9] はトラスト値と呼ばれる評価値を導入し、トラスト概念を

- トラスト (信頼, trust) : 相手あるいは事柄について、信頼を置いている状態
- ディストラスト (不信, distrust) : 相手や事柄について、信頼できないと思っている状態
- アントラスト (無信頼, untrust) : 相手や事柄について、信頼して良いかわからない状態

のように分類した。トラスト値は一次元的な区間 $[-1, 1]$ 上の 1 点として与えられ、観測者の目線からの効用や重要度などを用いた、具体的なトラスト値の計算方法も提案されている [8]。

一次元的なトラストの評価値は単純でわかりやすいが、「知らない人物に対しては、信頼も不信もない」などの対象に対する無関心や、「ある人物に対して、基本的には信頼しているが、じつは心の中では同時に不信感も持っている」といった対象に対する矛盾した評価を、うまく扱えなかった。そこで、予備研究 [14], [15] において我々は、「信頼と不信は完全に独立である」という文献 [7] の主張に基づき、「信頼度」と「不信度」の 2 軸を用いた新たなトラスト値の表現方法 (二次元的なトラスト表現法) を導入した。さらに、このトラスト値の表現方法が従来の一次元的な表現方法の自然な拡張になっていることを示した。我々のトラスト表現は、小田によるファジィ多項目並列評定法 (FCR 法) [16], [17], [18] を応用したものである。この手法により、ファジィ理論の知見に基づき、無関心の状況を「信頼度・不信度がともに低い状態」、矛盾・混乱した状況を「信頼度・不信度がともに高い状態」として扱えるようになった。

我々の研究と類似のアプローチとして、Jøsang の Subjective Logic [5] が挙げられる。Subjective Logic を用いても、「信頼度」と「不信度」に基づく二次元的なトラスト表現が可能である。そこで本稿では、FCR 法と Subjective Logic を、トラスト値の表現法としての観点から比較したい。本稿では、両手法の共通点と相違点について検討する。両手法とも対象に対する無関心を扱えるが、一方で、矛盾したトラストを扱えるのは FCR 法に基づくアプローチのみである。また、FCR 法と Subjective Logic では、時系列を追ったトラスト値の計算について、異なる前提をおいていると考えられる。本稿では、トラスト形成の典型的なシナリオ例を考えながら、こうした違いについて議論する。

本稿は、以下のように構成される。まず 2 節では、従来の一次元的なトラスト値とトラスト概念の分類について紹介する。さらに、FCR 法に基づく二次元的なトラスト値の表現法について述べる。次に、3 節で、FCR 法と Subjective Logic を比較する。Subjective Logic の概要を述べた後、FCR 法との共通点と相違点について議論する。

2. 準備: 信頼と不信に基づくトラストの表現法

本節では、Marsh らによる従来の (1 次元的な) トラスト表現を述べ、さらにトラストの分類について述べる。その後、小田の FCR 法 [16], [17], [18] の考え方に基づく二次元的なトラスト値 [14], [15] について述べる。

2.1 一次元的なトラスト値とトラスト概念の分類

Marsh と Dibben [9] は、「エージェント x の対象 y に対するトラスト値 (一般トラスト, $T_x(y)$)」や「エージェント x の対象 y に対する状況 α におけるトラスト値 (状況トラスト, $T_x(y, \alpha)$)」を、値域 $[-1, 1]$ の関数として与えた。つまり、トラスト値は、 -1 から 1 までの実数値として与えられる。さらに、この (一次元的な) トラスト値に基づき、トラストの概念を

- トラスト (信頼, trust) : 相手あるいは事柄について、信頼を置いている状態。トラスト値が 1 に近い状態として定式化する。より具体的には、「これ以上のトラスト値を持っていないと、この行為者とは協力したくない」という基準値 (信頼の閾値) $CT_x(y, \alpha)$ を考え^{*1}、トラスト値がその信頼の閾値を越える場合にトラストすると考える。
- ディストラスト (不信, distrust) : 相手や事柄について、信頼できないと思っている状態。評価値が負の状態、すなわち、 $T_x(y) < 0$ や $T_x(y, \alpha) < 0$ の場合として定式化する。ディストラストは、被信頼者 (trustee) y 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、エージェント x が対象 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。
- アントラスト (無信頼, untrust) : トラスト値が正であるものの、信頼の閾値 $CT_x(y, \alpha)$ よりも低い状態を指す。すなわち、

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

のときとして定式化する。アントラストの状態にある

^{*1} 閾値 $CT_x(y, \alpha)$ は

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

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。また、 $I_x(\alpha)$ (値域は $[0, 1]$) は、 x の状況 α に対する重要度である。この式より、リスク ($Risk_x(\alpha)$) やことの重要性 ($I_x(\alpha)$) が大きければ、閾値は高いレベルに設定されることになり、一方で、被信頼者がうまくことに当たると信じられる ($Comp_x(y, \alpha)$ 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。

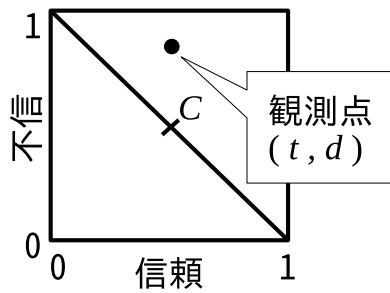


図 1 トラストの二次元的表現

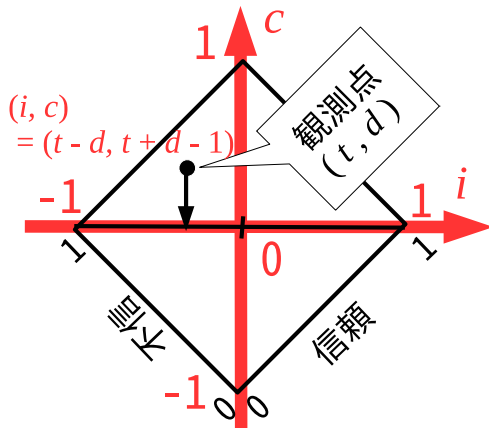


図 2 2 次元的なトラスト値の表現から 1 次元的な表現への変換

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ついて、信頼して良いか決めきれない状態を表している。この状況下では、エージェント x は対象 y と協力する状態には至らない。

に分類した*2。

2.2 信頼と不信に基づくトラストの二次元表現

Marsh と Dibben によるトラスト概念の分類は、単純でわかりやすいが、トラスト値が一次元的（すなわち、 -1 から 1 の間の数直線上で表される）であることから、次のような疑問が浮かぶ。たとえば、

- トラスト値が 0.9
- 信頼の閾値が 0.85

の場合を考えてみよう。このときは、トラスト値が信頼の閾値を越えており、定義から、信頼している状態である。しかしながら、このとき被信頼者に対する不信が全くないとまで言えるだろうか。トラスト値の満点は1であるから、トラスト値が 0.9 ということは、その満点までは 0.1 だけ足りないことになる。その足りない分だけ、対象を信頼し

*2 このほか文献 [9] では、当初の信頼を（いい方向への裏切りも含め）裏切られた状態である「ミストラスト（誤信頼, mistrust）」についても扱っている。これは、時間経過に伴うトラスト値の変化に関する性質である。

きれていないとの解釈も可能に見える。

予備研究 [14], [15] において我々は、こうした解釈ができるのは、一次元的な基準に基づくトラスト値の表現力の不足によるものと考えた。そこで、信頼の度合いの集合 $Trust$ と不信の度合いの集合 $DisTrust$ を

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

と定め、小田の FCR 法の考え方を適用して、トラスト値を $Trust \times DisTrust$ の要素（以下では、FCR 法での用語と同じく、観測点と呼ぶことにする）として二次元的に定めた (図 1)。ここでは、信頼度と不信度は独立に扱われる*3。なお、小田のモデルと類似の考え方で、真偽値を「真」と「偽」の対を用いて表すモデルには、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 [19] などが知られている。

観測点の集合は、4 点 $(0,0)$, $(1,0)$, $(1,1)$ および $(0,1)$ を頂点とする正方形の周および内部の点の集合である。ここで、 $(1,0)$ 付近にある観測点 $(t,d) \in Trust \times DisTrust$ (図 1 の右下付近の点) はトラスト値 t が高くディストラスト値 d が低い「トラスト」の状態を、 $(0,1)$ 付近の観測点 (図 1 の左上付近の点) は信頼が低く不信が高い「ディストラスト」の状態を表すとみなせる。また、観測点 $(0.5,0.5)$ は、従来のトラスト値 0 に対応する。我々は、従来型の一次元的なトラスト値は、 $(1,0)$ と $(0,1)$ を結ぶ線分（すなわち、図 1 に示される対角線）上の点に対応すると解釈した。この対角線上では、信頼と不信がつねに定和尺度 $(t+d=1)$ を満たす意味で、理想化されている。すなわち、従来のトラスト値は、暗黙的に「不信度は、信頼度が定まれば一意に決まる ($d=1-t$)」という仮定をおいていると言える。一方、我々のトラスト値は、そうした仮定・制限を設けずに正方形全体の観測点を考える意味で、一般化されている。

$(1,0)$ と $(0,1)$ を結ぶ対角線を区間 $[-1,1]$ に対応させるため、観測点 $(t,d) \in Trust \times DisTrust$ に関する変換

$$\begin{aligned} & \left[\begin{pmatrix} \cos \frac{\pi}{4} & -\sin \frac{\pi}{4} \\ \sin \frac{\pi}{4} & \cos \frac{\pi}{4} \end{pmatrix} \left\{ \begin{pmatrix} t \\ d \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} + \begin{pmatrix} \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix} \right] \times \frac{1}{\frac{\sqrt{2}}{2}} \\ & = \begin{pmatrix} t-d \\ t+d-1 \end{pmatrix} \end{aligned}$$

*3 ここで、なぜ FCR 法が相反する二つの指標（「真」と「偽」、「好き」と「嫌い」など）を独立に扱うかを補足しておきたい。アンケートでは「良い」「悪い」「どちらとも言えない」などの複数の項目から一つを選ぶ形式の評定尺度法がよく用いられる。しかしこの手法には、尺度の中間項目に回答が集中するという問題点がある。たとえば、被験者が「良い」とも「悪い」とも思っている場合や、質問の内容がわからなかったりするとき、中間項目（「ふつう」や「どちらとも言えない」）を選ぶ傾向にある。こうした回答は被験者の真の回答ではないため、分析が困難となる。この問題に対処するため、FCR 法では、全項目を独立に扱い、「そう言える度合い」すなわち確信度（該当度）を各度合い毎に測定する。これにより、従来の評定尺度法で測定できない種類の情報（「良いとも悪いとも思っている」など）を取り出すことができる。

を考えよう。この変換の図的表現を、図2に示す。移った先の点 $(t-d, t+d-1)$ を、 (i, c) と呼ぶ。変換後の第1要素 i は、 $-1 \leq i \leq 1$ を満たし、従来型の一次元的なトラスト値に対応するとみなせる。 $i = t-d$ であり、この式は「信頼できると考える度合い t から不信の度合い d を引いた差分が、正味のトラスト値である」ことを示している。

値 $i = t-d$ は、FCR 法において「統合値」と呼ばれる値に対応する*4。これは、「真」と「偽」などの相反する指標から求めた「本当の真の値」である。統合値は、0から1の値である。つまり、値 i は「信頼度」と「不信度」という指標から統合値を求め、さらにその値の範囲が -1 から 1 の範囲になるように正規化した値と言え。ただし、以下の説明では便宜上、値 i も統合値と呼ぶことにする。

統合値 i を用いることで、一次元的なトラスト値に対するトラスト概念の分類を、二次元的トラスト値に拡張できる。具体的には、信頼の閾値 CT に対して $i = t-d \geq CT$ であればトラスト、 i の値が負であればディストラスト、 $0 \leq i < CT$ であればアントラストである。ディストラストの条件については、 $i = t-d < 0$ より、 $t < d$ と言い換えられる。これは、「信頼の度合い t よりも不信の度合い d の方が上回っている」ときにディストラストの状態であることを表しており、直感に合致すると言える。

次に、上述の変換による結果 (i, c) の第2要素 $c = t+d-1$ を考える。これは、ファジィ理論で「矛盾度」と呼ばれる値である。FCR 法では、人格心理学の分野でしばしば扱われる「好きだが嫌い」といった矛盾した応答や「好きでも嫌いでもない」のような無関心も許容している。そうした応答の内部矛盾は、観測点の矛盾度で測られる。矛盾度の定義には数種類が知られているが[20]、本稿では観測点 (t, f) に関する定和からのずれ $C(t, d) = t+d-1$ を用いる。ここでは、完全な矛盾の場合は $C(t, d) = 1$ 、 $t+d=1$ が成り立つ「無矛盾の場合」は $C(t, d) = 0$ 、好きでも嫌いでもないという無関心の場合は $C(t, d) = -1$ となる。

2.3 例

3つの国で、各国政府の支持率を調べるための世論調査を行ったとしよう。この調査では「あなたは、政府を信頼していますか?」という質問を用い、解答として、

- (1) 興味がない
- (2) はい
- (3) いいえ
- (4) 「はい」の時も「いいえ」の時もある

*4 統合値の求め方には、いくつかの方法が提案されている。本稿では、「逆転項目平均法」と呼ばれる求め方を採用している。これは、観測点 (t, d) から $(1, 0)$ と $(0, 1)$ を結ぶ直線に下ろした垂線の足元の点（つまり、横軸の次元の値）を求め、さらに、とりうる値の区間を $[-1, 1]$ に正規化した値として求められる。

を用いた。国 c における、上述の各項目に関する回答数を a_1^c, \dots, a_4^c であらわす。また、 $s^c = a_1^c + a_2^c + a_3^c + a_4^c$ とする。この例において、政府に対する信頼度と不信度の値 t_c および d_c を、それぞれ $t_c = \frac{a_2^c + a_4^c}{s^c}$ および $d_c = \frac{a_3^c + a_4^c}{s^c}$ で計算する。

三つの国 X, Y および Z のそれぞれで100人の住民に対してアンケートを行い、

$$\begin{aligned} (a_1^X, a_2^X, a_3^X, a_4^X) &= (10, 20, 30, 40) \\ (a_1^Y, a_2^Y, a_3^Y, a_4^Y) &= (50, 30, 10, 10) \\ (a_1^Z, a_2^Z, a_3^Z, a_4^Z) &= (20, 25, 5, 50) \end{aligned}$$

という結果が得られたとしよう。各国について、信頼度の値 t_c と不信度の値 d_c の組は、上述の定義より

$$\begin{aligned} (t_X, d_X) &= (0.6, 0.7) \\ (t_Y, d_Y) &= (0.4, 0.2) \\ (t_Z, d_Z) &= (0.75, 0.55) \end{aligned}$$

である。さらに、各国の統合値 i_c および矛盾度 c_c の組は、

$$\begin{aligned} (i_X, c_X) &= (-0.1, 0.3) \\ (i_Y, c_Y) &= (0.2, -0.4) \\ (i_Z, c_Z) &= (0.2, 0.3) \end{aligned}$$

と計算できる。この結果から、次のような分析ができる。

- X 国については、政府に対する一定の不信度が見られる。しかし X 国においては矛盾度の値が正、すなわち混乱を表しており、住民の間に意見の対立がいくらか見られる。
- Y 国においては、トラスト値がディストラスト値を上回っているが、矛盾度がマイナスである。これは、 Y 国では政府に対する関心が低いことを示している。
- 最後に、 Y 国と Z 国を比較する。両国は同じ統合値 (0.2) を持つが、矛盾度の値が異なる。具体的には、 Y 国では矛盾度が負（無関心）なのに対して、 Z 国は正（賛否両論）であり、異なる特徴を示している。

最後の分析例は、一次元的な指標（統合値のみ）では比較できない2国も矛盾度を使うことで比較できることを示している。

3. FCR 法 vs. Subjective Logic

FCR 法と類似の考え方として Jøsang の Subjective Logic [5] が知られている。本節では、Subjective Logic に基づく二次元的なトラスト値の表現法について述べ、さらに、FCR 法に基づく表現法との共通点や相違点を議論する。

3.1 Subjective Logic の概要

不確実性を扱う枠組みとして、古くから確率論理 [12] が知られている。確率論理では、頻度（ランダム性）に基づく議論が行われ、結果への確信度は試行回数に依存する。

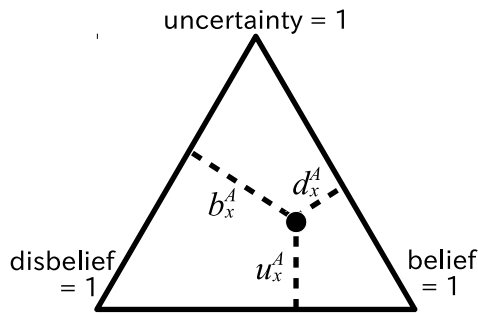


図 3 Subjective Logic による二項対立型意見の表現

すなわち、十分に多くの試行ができるときには結果への確信度は高くなり、できないときには低くなる。一方で Subjective Logic [5] では、頻度に加えて主観（認識性）に基づく議論を行う。命題に対する確信度は、その命題の示す内容に関する「ことの経緯」が明らかになっているかに依存する。つまり、「ことの経緯」が明らかなきは、命題に関する確信度は高くなり、そうでないときは低くなる。確率論理と Subjective Logic で扱われる典型的な命題としては、それぞれ、

- 「コイン投げをした時に表が出る確率は 2 分の 1」
 - 確率論理で扱われる典型的な命題である。この命題への確信度は、十分な回数のコイン投げができるときに高くなる。
- 「リー・ハーヴェイ・オズワルドがジョン・F・ケネディを殺した確率は 2 分の 1」
 - Subjective Logic で扱うことが想定される命題の例である。この事件は、経緯に多くの謎があるとされ、命題の確信度は高くはない。

が挙げられる。

Subjective Logic では、主観性を定式化する際にドメインと呼ばれる集合を用いる。ドメインの異なる要素は、互いに素な事象を表す。以下の説明では、意見 x と反対意見 \bar{x} の対による、大きさ 2 のドメイン $\{x, \bar{x}\}$ を考えることにする。これは、二項対立型の意見を考えることにあたる。

(二項対立型の) 意見は、

- b_x^A : 観測者 A の x についての確信度 (belief)
- d_x^A : A の x への懐疑度 (disbelief, \bar{x} の度合い)
- u_x^A : A の x についての不確かさ (uncertainty)
- a_x^A : x のベースレート (確信の度合いや懐疑の度合いを無視して考えるときの事前確率)

による 4 項組 $\omega_x^A = (b_x^A, d_x^A, u_x^A, a_x^A)$ として表される。値 b_x^A, d_x^A, u_x^A については、

- $0 \leq b_x^A, d_x^A, u_x^A, a_x^A \leq 1$
- $b_x^A + d_x^A + u_x^A = 1$

が成り立つとする。値 b_x^A, d_x^A, u_x^A の関係は、しばしば、図

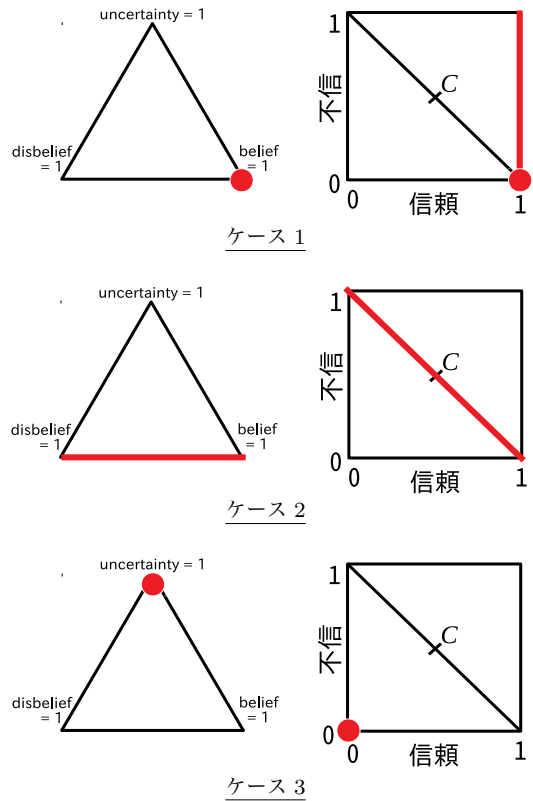


図 4 観測点 (FCR 法) と意見 (Subjective Logic) の対応

3 に示すような高さ 1 の正三角形で表される*5。正三角形の右下の頂点が $b_x^A = 1$ 、左下の頂点が $d_x^A = 1$ 、中央上の頂点が $u_x^A = 1$ の状態を表す。

なお、本稿では要素 a_x^A は扱わないが、これは事象 x に関する確率 (projected probability) $P(x)$ を

$$P(x) = b_x^A + a_x^A \cdot u_x^A$$

のように与える際に用いられる。たとえば $\omega_x^A = (0.4, 0.2, 0.4, 0.9)$ のとき $P(x) = 0.76$ である。

3.2 FCR 法と Subjective Logic の共通点・相違点

本節では、Subjective Logic による二次元的トラスト表現と FCR 法によるトラスト表現の対応を考える。

3.2.1 両者の共通点

Subjective Logic では、確信度 b_x^A を肯定的度合い（本研究では、対象 x に対して観測者 A がトラストしている度合い）と、懐疑度 d_x^A を否定的度合い（ディストラストの度合い）と考える。つまり、Subjective Logic における値の組 (b_x^A, d_x^A) は、FCR 法での組 (t, d) に対応する。意見 ω_x^A が与えられると、図 3 の正三角形の 1 点が決まる。以下では、三種類の特徴的な点について、考えたい。

*5 この図では、二項対立型の意見を表す点と正三角形の各辺からの距離が、値 b_x^A, d_x^A, u_x^A に対応するように描かれている。しかし実際には、そうした距離の和がつねに 1 になるとは限らない。その意味では不正確な面もあるが、概念を表す模式図と思って参照されたい。

ひとつめは、正三角形の右下の頂点、すなわち $b_x^A = 1$ の場合である (図4の「ケース1」の左側)。これは、事象 x について確信を持っている状態である。実際、 $b_x^A = 1$ のとき、不確かさ u_x^A の値は0である。 $b_x^A = 1$ の場合に対応するFCR法の観測点の集合は、 $t = 1$ となる点の集合、すなわち図1の平面の右端の一辺である (図4の「ケース1」の右側)。ただし実際には、 $b_x^A = 1$ の状態は、FCR法の観測点 $(1, 0)$ にもみ対応する。これは、Subjective Logic が、FCR法の無関心領域 (図1の、 $(1, 0)$ と $(0, 1)$ を結ぶ対角線の左下の領域) のみを扱うためである。詳しくは、3.2.2節で述べる。観測点 (t, d) について、 t と d をそれぞれ真偽の度合いと考えたとき、観測点 $(1, 0)$ は「完全に真となる状態」とみなせる。Subjective Logic においても、 $b_x^A = 1$ のときは $d_x^A = 0$ であり、完全に真の状態と解釈される。

次に、図4の「ケース2」を考えよう。図の左側は、正三角形の底边上の点の集合を示している。これは、 $u_x^A = 0$ を満たす、不確かさのない意見の集合に対応する。このとき、 $b_x^A + d_x^A + u_x^A = 1$ より $b_x^A + d_x^A = 1$ となり、確信度と懐疑度の和は1になる。FCR法との対応で考えれば、定和尺度 ($t + d = 1$) にあたる。つまり、 $u_x^A = 0$ のとき矛盾度は0である。またトラスト表現の見地からは、 $u_x^A = 0$ のときは、2.2節での議論により、 ω_x^A は従来型の一次元的なトラスト値に対応すると言える。

最後に、図3の正三角形の中央上の頂点を考えよう (図4の「ケース3」の左側)。これは、 $u_x^A = 1$ の場合である。さらに、 $b_x^A + d_x^A + u_x^A = 1$ かつ $0 \leq b_x^A, d_x^A, u_x^A \leq 1$ より $(b_x^A, d_x^A) = (0, 0)$ が導かれる。Subjective Logic では、この状態は、真偽が決まらない「未定義 (undefined)」の状態と呼ばれる。FCR法においては、図1の平面の左下の1点、すなわち $(0, 0)$ が $u_x^A = 1$ に対応する観測点 (のひとつ) である。

3.2.2 相違点：ふたつのシナリオ例による比較

前節では、FCR法の観測点 $(0, 0)$ が、 $u_x^A = 1$ を満たす意見に対応すると述べた。しかしFCR法では、未定義の状態を表す典型的な観測点として、 $(0, 0)$ のほかに $(1, 1)$ を考えることができる。

Subjective Logic では、 $b_x^A + d_x^A + u_x^A = 1$ が成り立ち、さらにこの等式を変形すると

$$-u_x^A = b_x^A + d_x^A - 1$$

が得られる。値の組 (b_x^A, d_x^A) はFCR法の観測点に対応し、さらに、この等式の右辺は矛盾度の定義と一致している。つまり、左辺 $-u_x^A$ は矛盾度である。さらに $0 \leq u_x^A \leq 1$ より $-1 \leq b_x^A + d_x^A - 1 \leq 0$ が成り立つ。

いま、 $-u_x^A = b_x^A + d_x^A - 1 \leq 0$ であるから、Subjective Logic では、矛盾度の値が0以下の場合のみを扱うことになる。FCR法では、集合 $\{(t, d) | t, d \in [0, 1] \wedge t + d > 1\}$ を矛盾領域と、集合 $\{(t, d) | t, d \in [0, 1] \wedge t + d < 1\}$

を無関心領域もしくは無関連領域と呼ぶ。また、集合 $\{(t, d) | t, d \in [0, 1] \wedge t + d = 1\}$ は、数値真理値空間あるいは無矛盾領域と呼ばれる。この定義を当てはめると、Subjective Logic では、無矛盾領域と無関心領域のみを扱い、矛盾領域は考慮していないとみなせる。なお、同様の矛盾領域を持たない論理モデルとしては、AtanassovのA-IFS model [1] などもある。実際には、矛盾領域を持つモデルの方が少数派のようである。

ファジィ理論の多くの研究では、矛盾領域上の点は情報過多の状況に、無関心領域上の点は情報不足の状況に対応するとみなされる。この知見をもとにすると、Subjective Logic とFCR法のそれぞれに基づくトラスト表現についての違いを見ることが出来る。以下では、Subjective Logic とFCR法のそれぞれにおける、トラスト形成の典型的なシナリオ例を考えよう。

3.2.2.1 シナリオ例1：Subjective Logic の場合

計算の初期段階において、観測者は、対象 x がトラストできるかを判断する証拠を何も持っていない。このときは $(b_x^A, d_x^A) = (0, 0)$ であり、未定義の状態 ($u_x^A = 1$) である。本研究では、計算を進めて証拠 (対象が信頼できると思わせる証拠もあれば、不信を抱かせる証拠もあるだろう) が集まるにつれて、トラストの度合い b_x^A やディストラストの度合い d_x^A が増えてゆくと仮定する。それに伴い、 u_x^A の値は0に近づいてゆく。最終的に $u_x^A = 0$ になったとき、これは、定和尺度 ($b_x^A + d_x^A = 1$) を満たす無矛盾な状態である。無矛盾な状態にあるということは、対象に対するトラストとディストラストの割合を、観測者が合理的に評価できるということである。実際、ディストラスト値 d_x^A は、トラスト値 b_x^A から計算できる ($d_x^A = 1 - b_x^A$)。このことから、Subjective Logic では、「十分な証拠が集まるにつれて、 x のトラスト値に関する曖昧さは下がってゆき、最終的には観測者 A が x のトラスト値を合理的に計算できるようになる」という暗黙的な前提をおいているとみなせる。

3.2.2.2 シナリオ例2：FCR法の場合

一方で、FCR法によるトラスト値のモデリングの下では、トラスト形成の典型的なシナリオは、次のようなものになる。トラスト値を決めるための証拠を何も持たない初期段階では、Subjective Logic の場合と同様、 $(t, d) = (0, 0)$ である。このとき、矛盾度 $C(0, 0)$ の値は -1 で、観測者は対象 x に対して無関心の状態にある。また、矛盾度の絶対値 $|C(0, 0)| = 1$ は、とりうるなかで最も高い値である。矛盾度の絶対値は、観測点の無矛盾領域からの距離を表しており、これは Subjective Logic の値 u_x^A に対応している。計算がすすみ、トラスト値を決定するための証拠が集まるにつれて、トラストの度合い t やディストラストの度合い d が増えてゆき、矛盾度の絶対値 $|C(t, d)|$ は0に近づいてゆく。そして、矛盾度の値が0になると、無矛盾な状態になり、観測者は、対象に対するトラストとディストラ

トの割合を合理的に評価できようになる。ここまでのシナリオは Subjective Logic の場合と同様であるが、FCR 法の場合では矛盾領域が存在するため、さらに続きがある。得られる証拠がさらに増え続けると、矛盾度 $C(t, d)$ の値は増加を続け、1 に近づいてゆく。これにともない、減少し続けて 0 になった矛盾度の絶対値は、今度は増加に転ずる。最終的には「対象 x を信頼できると判断される証拠も、信頼できないと判断される証拠も、どちらも多すぎる」といった、証拠の多すぎる混乱した状況に達すると考えられる。つまり、FCR 法では「多すぎる証拠の下では、混乱のため、観測者 A は x のトラスト値を正しく計算できない」場合があることを仮定しているといえる。

3.3 考察

上述のように、Subjective Logic と FCR 法では、

- Subjective Logic : 十分な証拠が集まったとき、観測者は対象のトラスト値を合理的に計算できるようになる
- FCR 法 : 多すぎる証拠の下では、混乱が生まれ、観測者は対象のトラスト値を正しく計算できない

という異なる前提に基づいていると考えられる。我々は、FCR 法に基づくトラスト表現に関する前提の方が合理的であり、Subjective Logic に基づくトラスト表現の前提は強すぎると考える。実際、2.3 節の例について、矛盾度が正の値になる X 国と Z 国の場合 ($c_X = 0.3$ および $c_Z = 0.3$ である) を、Subjective Logic に基づくトラスト表現ではうまく扱えない。

しかし、政府に対するトラストの度合い t_c とディストラストの度合い d_c の定義を変更し、 $t_c = \frac{a_2^c + 0.5 \times a_4^c}{s^c}$ および $d_c = \frac{a_3^c + 0.5 \times a_4^c}{s^c}$ のようにすれば、つねに $t_c + d_c - 1 \leq 0$ が満たされるようになり、Subjective Logic に基づくトラスト値においても、 X 国と Z 国の場合を扱えるようになる。この修正においては、変数 a_4^c に関して、重み 0.5 が使われている。この重みの導入は、回答項目

(4) 「はい」の時も「いいえ」の時もある

を選ぶ回答者個人の政府に対するトラスト値とディストラスト値を 0.5 ずつにすることと解釈できる。言い換えれば、項目 (4) を選んだ回答者は、「はい」と「いいえ」にそれぞれ 0.5 票ずつ投票している、ということである。

t_c と d_c の見積もりで重み 0.5 を用いるモデリングにおいては、項目 (4) を選ぶ回答者は、政府に対する信頼を定和尺度を満たすように評価していることになる。実際、「はい」への 0.5 票と「いいえ」への 0.5 票を合計すると、1 である。しかしながら、2.3 節では、我々はそのような仮定を採用していない。これは、本稿では「トラストとディストラストは完全に独立である」という主張 [7] に基づくためである。トラストの度合い (t_c) とディストラストの

度合い (d_c) が独立ならば、その両者がともに大きな値である場合も考えられる。よって、 $t_c + d_c \leq 1$ を満たす場合に限定せず、 $t_c + d_c > 1$ の場合も扱う必要がある。したがって、矛盾領域 $\{(t, d) | t + d > 1\}$ が必要である。実際、2.3 節でのモデル化においては、項目 (4) を選ぶ回答者の政府に対するトラスト値とディストラスト値はともに 1 であり、それらの和はとりうる最大値 (2) となっている。本稿ではこうした理由により、矛盾領域を扱った。

上述の修正版の例では、項目 (4) を選んだ回答者は「はい」と「いいえ」にそれぞれ 0.5 票ずつ投票するという、やや大雑把なモデル化をしている。これは議論の簡単化のためであり、より精密な評価も可能であると考えられる。たとえば、

$$a_{4pos}^c = \sum_{i \in S_{4th}} \text{回答者 } i \text{ のトラスト値}$$

$$a_{4neg}^c = \sum_{i \in S_{4th}} \text{回答者 } i \text{ のディストラスト値}$$

として (ただし、 S_{4th} は、項目 (4) を選んだ回答者の集合)、 $t_c = \frac{a_2^c + a_{4pos}^c}{s^c}$ および $d_c = \frac{a_3^c + a_{4neg}^c}{s^c}$ のように定義する方法が考えられる。

4. むすび

本稿では、小田によるファジィ多項目並列評定法 (FCR 法) と Jøssang の Subjective Logic を、二次元的なトラスト値の表現法としての観点から比較した。両手法とも、「信頼度」と「不信度」の 2 軸に基づき、トラスト値を表現できる。本稿では、「トラスト値が最大の場合」「定和尺度を満たす場合」「トラスト値に関する曖昧さが最大の場合 (すなわち、未定義の場合)」の三つの場合を対象に、両手法による二次元的トラスト値 (観測点と意見) の対応を検討した。また本稿では、両手法の違いについても議論した。両手法とも無関心領域を持つため、「知らない人物に対しては、信頼も不信もない」といった対象に対する無関心を扱える。しかし、Subjective Logic では矛盾領域を扱っておらず、これを持つのは FCR 法のみである。そのため、FCR 法に基づくトラスト表現においてのみ、「ある人物に対して、基本的には信頼しているが、じつは心の中では同時に不信感も持っている」といった矛盾したトラストを扱える。このほか本稿では、FCR 法と Subjective Logic のそれぞれについてトラスト形成の典型的なシナリオ例を考えながら、時系列を追ったトラスト値の計算に関する前提条件の違いについても議論した。

本研究では、トラスト形成における時系列的なトラスト値の遷移に関する考察を行ったが、予備研究 [14], [15] においても、トラストに関する安全性などを題材に、時系列的なトラスト値の変化に関する形式検証法を検討している。トラスト値の時系列的な変化に関する性質を明らかにし、検証法を確立することは重要である。実際、swift trust

[10], [13] のような, 短期間のうちに構築する集団のトラストに関する概念も知られており, 大規模災害時におけるトラストや協力関係の構築 [2], [6], [11] を行う際にも, 時系列的なトラスト値の遷移の分析は重要だと考えられる. 今後の課題として, 検証技術を確立し, さらに実システムへの適用可能性を確保することが挙げられる. 本研究はソーシャルメディアを一部に含む災害時通信システムの開発に関する研究プロジェクトの一部として行われており, 適用先の一例として, そうした災害時通信システム ([3][4]) を考えている.

このほか, 別の課題として, 矛盾度の利用方法を検討することが挙げられる. 実際の通信システムでは, 分析者はさまざまな情報源から矛盾する証拠を受け取る場合があると考えられる. つまり, 一部の情報源は, 間違った証拠を提供するかもしれない. 我々は, こうした状況をうまく扱うのに, 矛盾度が利用できると考えている. たとえば, 間違った情報源が多数ある場合, 矛盾の度合いが高くなる. こうしたときに, それまでの証拠を破棄して情報を集め直す必要があるかどうかを判断するのに, 矛盾度が利用できる. 詳細についての検討は, 今後の課題である.

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] Maria Grazia Busa, Maria Teresa Musacchio, Shane Finnan, and Cilian Fennell. Trust-building through social media communications in disaster management. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pp. 1179–1184, New York, NY, USA, 2015. ACM.
- [3] 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.
- [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] Audun Jøsang. *Subjective Logic: A Formalism for Reasoning Under Uncertainty*. Springer Publishing Company, Incorporated, 1st edition, 2016.
- [6] F. Lemieux. The impact of a natural disaster on altruistic behaviour and crime. *Disasters*, Vol. 38, No. 3, pp. 483–499, July 2014.
- [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, 2014.
- [12] Nils J. Nilsson. Probabilistic logic. *Artif. Intell.*, Vol. 28, No. 1, pp. 71–88, February 1986.
- [13] 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.
- [14] 河辺 義信, 小泉 佑揮, 大木 哲史, 西垣 正勝, 長谷川 享, 小田 哲久. 矛盾-無関心を考慮したトラスト値の遷移について. CSS 2018 コンピュータセキュリティシンポジウム 予稿集, 第 2 巻, pp. 1194–1201, 2018.
- [15] 河辺 義信, 小田 哲久. FCR 法を用いたト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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.