

フレーム表現における類似性の判定

Similarity Metrics on Frame Knowledge Expressions

王 勇* 犬塚 信博* 石井 直宏*
Yong Wang Nobuhiro Inuzuka Naohiro Ishii

* 名古屋工業大学知能情報システム学科
Dept. of Intelligence and Computer Science, Nagoya Institute of Technology, Nagoya 466, Japan.

1994年7月20日 受理

Keywords: similarity metrics, case-based reasoning, analogical reasoning, fuzzy application.

Summary

Similarity metrics is a central issue in automatic reasoning system, especially in case-based reasoning and analogical reasoning. This paper proposes a method of similarity metrics by focusing the similarity of features of problems which are represented by frame knowledge expressions. For retrieving the most similar previous case and adapting it in solving new problem, similarity metrics are needed to provide the retrieving criterion and the information of adaptation. In our method, we assign a degree of similarity as a retrieving criterion between the features of new problem and past one, and using fuzzy integral to calculate the degree of similarity by comparing target frame which presents the features of new problem with source frame which presents the features of past problem. While calculating the degree of similarity, we also make a comparative frame which includes the information of matching source frame to target frame for adaptation. Finally we show some simulating results.

1. はじめに

事例ベース推論, 類推などの推論方式では, 新しい問題に対して, 過去の経験した類似事例を利用して推論を行うことから, 類似性の判定はこれらの推論方式において重要な役割を果たしている。しかしながら, いままでの開発された数多くの推論システムのなかで, 類似性の判定方法がほとんどおのこの応用分野に限定されており, 移植性が低くて, 一般性が欠如しているのが現状である。その要因は次の点であると考えられる。

1. 類似性の判定方法は問題の表現形式に依存する。
2. 類似性の判定基準は応用分野によって異なる。
3. 視点によって類似性は変わる。

そこで, より通用的柔軟な推論システムを開発するには, 多くの応用分野に対応できる一般的な類似性の判定方法が要求される。我々は与えられた問題と過去事例の特徴の類似性に注目して, フレーム表現を用いた類似性判定の一方法を提案する。

フレームは代表的な知識表現方式の一つで, 問題領域内に関する知識を階層的に記述するにはきわめて適した構造的知識表現である。実際に開発された事例ベース推論システムのなかで, フレームを用いて知識ベースを構築するのが多く見られる。本論文では, 問題と事例の表現形式として, フレームを用いる。また, ファジィ積分はいくつかの属性を持つ評価対象をファジィ測度を用いて総合的に評価するための方法である[菅野 72]ことから, 我々はこれを用いてフレームの要素であるスロットの類似度の総合評価値を測り, それによってフレームの類似度を定義する。類似性判定の目的は現在の問題と類似する過去の事例を推論に利用することから, 類似性判定の段階では, 類似事例の検索基準と事例修正の情報を提供する必要がある。我々の方法では, 現在の問題と過去の事例を表すフレームの比較によって類似性判定の基準となる類似度を, ファジィ積分を用いて計算するとともに, 現在の問題と過去の事例の類似フレームを構成する。類似フレームのなかに, 事例を利用するため, 現在の問題と過去の

事例の間のマッチング情報を格納する。また、事例と問題の各特徴の重視度はフレームの要素であるスロットの重みとして表され、それによってファジィ測度を決定し、ファジィ積分を計算する。同じ問題(または事例)に対しても、視点の違いによって構成されたフレームが違ってくるので、それによって求められた類似度と類似フレームも違ってくるのである(例3・4を参照)。我々の方法を用いて、応用分野に対応したアトムスロットの類似度を定めれば、おのおのの分野の問題の特徴の類似性を判定することができる。

本論文では、2章にフレームとファジィ積分の定義について紹介する。3章に類似度と類似フレームの定義について述べる。4章にシミュレーションの結果について報告する。

2. フレームとファジィ積分の定義

この章では、フレームとファジィ積分の定義について述べる。

【定義2・1】 「属性#値」という形式の知識表現単位はアトムスロット(*atom-slot*)という。

アトムスロットの値として、次の6種類を考える。

- ①実数値 ②実数値の集合 ③文字列 ④文字列の集合 ⑤手続き ⑥変数。

【定義2・2】 属性の異なるアトムスロットの集合は単純フレーム(*simple-frame*)という。

【定義2・3】 スロットとフレームは次のように再帰的に定義される。

1. 空集合または単純フレームはフレームである。
2. $s='F'#F$ は非アトムスロットという。ただし、 F はフレームで、 $'F'$ はフレーム名である。
3. スロットはアトムスロットまたは非アトムスロットである。
4. スロットの集合はフレームである。ただし、同一フレームに属するアトムスロットの属性はそれぞれ異なる。

スロット $s=A#V$ について、 A はスロット名、 V はスロット値といい、それぞれ $Name(s)$ と $Value(s)$ で表される。また、フレーム F に含まれるスロット名の集合は $SN(F)$ で表される。

【定義2・4】 過去の問題の特徴を表すフレームはソースフレーム(*source-frame*)といい、 F_s で表される。

現在の問題の特徴を表すフレームはターゲットフレーム(*target-frame*)といい、 F_t で表される。

【定義2・5】

1. フレーム F の重み W_F は次式で表される。

$$W_F = \sum_{s \in F} W_s \quad \text{ただし,}$$

- (a) $s=a#v$ がアトムスロットである場合、 $W_s = \lambda_a$ 。 λ_a は正実数で、あらかじめ決めておく必要がある。
- (b) $s='f'#f$ は非アトムスロットである場合、 $W_s = W_f$ 。

2. フレーム F 上のファジィ測度 g_F は次のように定める。

- (a) $g_F(\phi) = 0$,
- (b) $g_F(F) = 1$,
- (c) $g_F(F') = \frac{W_{F'}}{W_F}$ 。ただし、 $F' \subset F$ 。

3. g_F による F のファジィ積分 $\int F \circ g$ は次式で定義する。

$$\int F \circ g = \bigvee_{F' \subset F} \{ [\bigwedge_{A, \#V_i \in F'} V_i] \wedge g_F(F') \}$$

ただし、 \bigvee と \bigwedge はそれぞれ上限と下限を表す記号である。

この定義はファジィ集合におけるファジィ積分の定義をフレームに拡張したものである。

明らかに、上のように定められたファジィ測度 g_F は単調性、有界性と加法性を持ち、したがって g_F によるファジィ積分は総合評価としての意味を持つ[菅野72, 菅野73]。本論文では、ファジィ積分を比較フレームに適用し、比較フレーム内に、おのおのの類似スロットの値——つまり、比較されるスロットの類似度(定義3・5と定義3・8を参照)——の総合評価値を測る。比較の方法によって比較フレームが違ってくるので、構成された比較フレームのなかで、最大ファジィ積分を類似度とする(3・3節を参照)。

3. 類似度と類似フレーム

類似度と類似フレームを算出するためのアルゴリズムを次に示す。

1. 応用分野に応じてアトムスロットの類似度を定める。それに基づく類似アトムスロットを求める(3・1節, 定義3・11~3・13を参照)。
2. スロットの比較によって、フレームの比較フレームを構成する。比較の方法によって、二つのフレーム F_t と F_s に対して比較フレームが三つある。 $F_t : F_s, F_t : \{ 'F_s' \# F_s \}$ と $\{ 'F_t' \# F_t \} : F_s$ (定義3・10を参照)。
3. 要素の重みによって決められたファジィ測度に基づく比較フレームのファジィ積分を計算する。

最大ファジィ積分は類似度で、その類似度を持つ比較フレームは類似フレームとする(定義3・10を参照)。

4. フレームにフレームを含んでいる場合、ステップ2と3を繰り返して、再帰的に計算する。

この章では、我々は定義と例を用いて、上記の方法について説明する。

3・1 アトムスロットの類似度と類似アトムスロット

フレームの類似度を計算するため、まず、フレームの基本要素であるアトムスロットの類似度を決定しなければならない。二つのフレームを比較するとき、アトムスロットについて、明らかに同一属性を持つものだけ比較すればよい。この節では、アトムスロットの類似度を決定する方法について述べる。

【定義3・1】 アトムスロット $s_t = a\#v_t$ と $s_s = a\#v_s$ の類似度は $DS_{s_t \sim s_s}$ のように表され、これは値 v_t と v_s の類似度である $DS_{v_t \sim v_s}$ と等しい。

【定義3・2】 同一属性の値 v_t と v_s について、その二つの値が等しいことだけに注目する場合、 v_t と v_s の類似度を次のように定義する。

$$DS_{v_t \sim v_s} = \begin{cases} 1 & (v_t = v_s) \\ 0 & (v_t \neq v_s) \end{cases}$$

この定義では、アトムスロットの類似度は0と1しかない。それだけでは不十分な場合について、次の定義を与える。

【定義3・3】 v_t と v_s は属性 a の値で、数値または文字列に限られているとする。 v_t と v_s の相対距離を $RD(v_t, v_s)$ のように表す。

1. v_t と v_s は数値である場合

v_t と v_s の相対距離を次のように定義する。

$$RD(v_t, v_s) = \frac{v_t - v_s}{R_a}$$

ただし、 R_a は a の値の範囲、つまり、 a のすべての値のなかで、最大値と最小値の差である。

2. v_t と v_s は文字列である場合

- (a) 文字列の順序の類似性に注目する場合。

順序によって a のすべての値に順位をつける。その結果、各文字列はそれぞれ一つの自然数(順位)と対応する。この場合、 v_t と v_s の相対距離を $RD_{v_t \sim v_s} = RD_{v'_t \sim v'_s}$ のように定義する。ただし、 v'_t と v'_s は文字列 v_t と v_s に対応する自然数である。

- (b) 文字列の形状または意味の類似性に注目する場合。

a のすべての値は a を根とする重みつき無向

木 G_a によって表され、 G_a のうち、 a のすべての値は葉の位置に置かれているとする。葉 v_x と v_y との距離 $D(v_x, v_y)$ は v_x と v_y の間の最短道の長さとして定義される。この場合、 v_t と v_s の相対距離を次のように定義する。

$$RD(v_t, v_s) = \frac{D(v_t, v_s)}{D_{\max}}$$

ただし、 $D_{\max} = \max_{v_x, v_y \in V_a} D(v_x, v_y)$ 。 V_a は G_a の節点の集合である。

【定義3・4】 $s_t = a\#v_t$ と $s_s = a\#v_s$ は同一属性 a を持ち、値が数値または文字列であるアトムスロットとする。 v_t と v_s の類似度を $DS_{v_t \sim v_s}$ のように表し、これを次のように定義する。

$$DS_{v_t \sim v_s} = e^{-\mu RD^2(v_t, v_s)}$$

ただし、 $\mu > 0$ で、 μ をあらかじめ定めておく必要がある。

μ を適当に選択することによって、応用分野に適する類似度が得られる。

$x = RD(v_t, v_s)$ とすると、 $DS_{v_t \sim v_s} = e^{-\mu x^2}$ ($x \in [0, 1]$)。いくつかの μ に対して、関数 $e^{-\mu x^2}$ のグラフを図1に示す。例えば、 $x = 0.1$ のとき、類似度がほぼ0.9になるのが望ましい場合、 $\mu = 10$ を選ばばよい。

次に、いくつかの例を示す。

【例3・1】 人の年齢の類似度を定義してみよう。

$R_{\text{年齢}} = 100$ とし、 $\mu = 20$ を選ぶ。 v_1 と v_2 は年齢で、 $D(v_1, v_2) = |v_1 - v_2|$ とすると、次の類似度が得られる。

$$DS_{v_1 \sim v_2} = \begin{cases} 1 & (D(v_1, v_2) = 0, \text{ i. e. } v_1 = v_2) \\ 0.9512 & (D(v_1, v_2) = 5) \\ 0.8187 & (D(v_1, v_2) = 10) \\ 0.4493 & (D(v_1, v_2) = 20) \end{cases}$$

【例3・2】 属性名は最終学歴である場合について考える。その可能な値は次に示すものとする。

小学校中退、小学校卒業、中学校中退、中学校卒業、高校中退、高校卒業、大学中退、大学卒業、学士、修

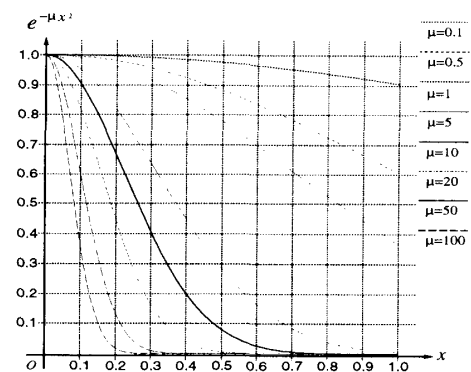


図1 それぞれの μ に対して $e^{-\mu x^2}$ のグラフ

士課程中退, 修士課程修了, 修士, 博士課程中退, 博士課程修了, 博士

以上の順序に従って, 1 から 15 までの順位をつけ, $R_{最終学歴} = 15 - 1 = 14$ である. $\mu = 20$ を選び, N_{LSC} は最終学歴 LSC と対応する順位とすると, 次の類似度が得られる.

$$DS_{LSC_1 \sim LSC_2} = \begin{cases} 0.9030 & (|N_{LSC_1} - N_{LSC_2}| = 1) \\ 0.6649 & (|N_{LSC_1} - N_{LSC_2}| = 2) \\ 0.3992 & (|N_{LSC_1} - N_{LSC_2}| = 3) \end{cases}$$

【例 3・3】 属性名は英小文字である場合を考える. 我々は英小文字の形の類似性に注目する. その類似性は図 2 のような重みつき無向木で表されているとする.

この場合, $D_{max} = 20$ である. $\mu = 20$ とすると, $DS_{letter_1 \sim letter_2} = 0.8187$. ただし, $letter_1$ と $letter_2$ は木の葉で, 兄弟(親節点と同じ)である. このようなペアが八つ存在する. $v-y, z-s, i-j, t-l, u-n, b-d, q-p$ と $c-o$ である. その他の結果をいくつか示す.

$$DS_{e \sim c} = 0.6376, \quad DS_{b \sim p} = 0.4493, \\ DS_{i \sim t} = 0.2865, \quad DS_{r \sim m} = 0.0002.$$

3・3 節では, 属性の値が他のケースである場合について議論する.

【定義 3・5】 同一属性 a を持つアトムスロット $s_t = a\#v_t$ と $s_s = a\#v_s$ の類似アトムスロットを $s_t \sim s_s$ のように表し, $s_t \sim s_s = a \sim a\#DS_{s_t \sim s_s}$ として定義する.

次節では, この節の定義を用いて, 類似単純フレームと単純フレームの類似度を定義する.

3・2 類似単純フレームと単純フレームの類似度

【定義 3・6】 単純フレーム f_t と f_s の類似単純フレームを $f_t \sim f_s$ のように表し, これを次のように定義する.

$$f_t \sim f_s = \bigcup_{a \in SN(f_t) \cup SN(f_s)} s(a)$$

ただし

$$s(a) = \begin{cases} s_t \sim s_s & (a \in SN(f_t) \cap SN(f_s), \\ & s_t \text{ と } s_s \text{ の属性名} = a) \\ a \sim \phi \# 0 & (a \in SN(f_t) \cap \overline{SN(f_s)}) \\ \phi \sim a \# 0 & (a \in \overline{SN(f_t)} \cap SN(f_s)) \end{cases}$$

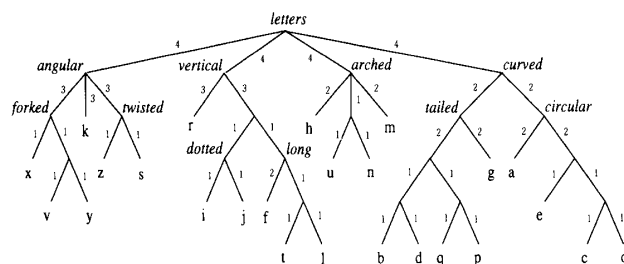


図 2 英小文字の類似性を表す重みつき無向木

【定義 3・7】 単純フレーム f_t と f_s の類似度は $DS_{f_t \sim f_s}$ のように表し, これを次のように定義する.

$$DS_{f_t \sim f_s} = \int f_t \sim f_s \circ g$$

ファジィ測度 $g_{f_t \sim f_s}(F)$ を決定するために, 類似アトムスロット $s_t \sim s_s$ の重みである $W_{s_t \sim s_s}$ を次のように計算する.

$$W_{a \sim \phi \# 0} = W_{\phi \sim a \# 0} = \lambda_a, \quad W_{a \sim a \# v} = 2\lambda_a.$$

次に示される例のなかで, すべてのアトムスロットの重みは 1 で, 任意な属性の値は $[0, 1]$ に属し, $\mu = 10$ とする.

【例 3・4】 直方体₁ = {長# 0.1, 幅# 0.2, 高# 0.3}, 直方体₂ = {長# 0.25, 幅# 0.5, 高# 0.75} とすると

$$\text{直方体}_1 \sim \text{直方体}_2 \\ = \{\text{長} \sim \text{長} \# 0.7985, \text{幅} \sim \text{幅} \# 0.4066, \\ \text{高} \sim \text{高} \# 0.1320\}$$

$$DS_{\text{直方体}_1 \sim \text{直方体}_2} = 0.4066$$

直方体の特徴の表現形式を次のように変える.

$$\text{直方体}_1 = \{\text{長} \# 0.1, \text{幅/長} \# 2, \text{高/長} \# 3\} \\ \text{直方体}_2 = \{\text{長} \# 0.25, \text{幅/長} \# 2, \text{高/長} \# 3\}$$

$R_{幅/長} = R_{高/長} = 10\ 000$ とすると

$$\text{直方体}_1 \sim \text{直方体}_2 = \{\text{長} \sim \text{長} \# 0.7985, \text{幅/長} \sim \\ \text{幅/長} \# 1, \text{高/長} \sim \text{高/長} \# 1\}$$

$$DS_{\text{直方体}_1 \sim \text{直方体}_2} = 0.7985$$

後述の表現形式を用いて, 相似の関係にある直方体の類似度は $2/3$ 以上である.

この例では, 視点によって類似性が変わることを示す.

【例 3・5】 色リンゴ = {赤# 0.8, 緑# 0.15, 茶色# 0.05}, 色ブドウ = {赤# 0.1, 緑# 0.2, すみれ色# 0.7} とすると

$$\text{色リンゴ} \sim \text{色ブドウ} \\ = \{\text{赤} \sim \text{赤} \# 0.0074, \text{緑} \sim \text{緑} \# 0.9753, \\ \text{茶色} \sim \phi \# 0, \phi \sim \text{すみれ色} \# 0\}$$

$$DS_{\text{色リンゴ} \sim \text{色ブドウ}} = 0.3333$$

【定義 3・8】 フレーム F_t と F_s の類似スロットは $S_{F_t \sim F_s}$ のように表し, これを次のように定義する.

$$S_{F_t \sim F_s} = 'F_t' \sim 'F_s' \# DS_{F_t \sim F_s}$$

次節では, 3・1 節と 3・2 節の定義を用いて, 任意のフレームに対して, 類似度と類似フレームを定義する.

3・3 類似フレームとフレームの類似度

すべてのフレーム名はクラスごとに分けられていると仮定する. この仮定のもとで, 類似度と類似フレームを求めるときに, 同一クラスに属さないフレームの比較が不要となり, 同一クラスに属するフレームどうしだけの比較により, 計算速度を上げることができる.

例えば、著者、第一著者、第二著者などのフレーム名は同一フレーム名クラスに属すべきである。

【定義3・9】 F_t と F_s は非アトムスロットからなるフレームとする。 F_t と F_s の比較フレームは $F_t:F_s$ のように表し、これを次のように定義する。

1. $F_t=F_s=\phi$ ならば

$$F_t:F_s=\phi$$
2. $F_t\neq\phi, F_s=\phi$ ならば

$$F_t:F_s=\bigcup_{f\#f\in F_t}\{f\sim\phi\#0\}$$
3. $F_t=\phi, F_s\neq\phi$ ならば

$$F_t:F_s=\bigcup_{f\#f\in F_s}\{\phi\sim f\#0\}$$
4. $F_t\neq\phi, F_s\neq\phi$ ならば
 F_t から任意の一つの非アトムスロット ' $f_t\#f_t$ ' を選び、' $f_t\#f_t\in FNC$ ' (FNC は一つのフレーム名クラス) とする。
 (a) $\forall f\#f\in F_s, f\#f\notin FNC$ ならば

$$F_t:F_s=\bigcup_{\substack{f\#f\in F_t \\ f\#f\in FNC}}\{f\sim\phi\#0\}\cup F'_t:F_s$$
 ただし、 $F'_t=F_t-\bigcup_{\substack{f\#f\in F_t \\ f\#f\in FNC}}\{f\#f\}$
 (b) $\exists f_s\#f_s\in F_s, f_s\#f_s\in FNC$ ならば

$$F_t:F_s=\{f_{t0}\sim f_{s0}\#DS_{f_{t0}\sim f_{s0}}\}\cup F'_t:F'_s$$
 ただし、

$$DS_{f_{t0}\sim f_{s0}}=\max_{\substack{f_1\#f_1\in FNC \\ f_1\#f_1\in F_t, f_2\#f_2\in F_s}} DS_{f_1\sim f_2}$$

$$F'_t=F_t-\{f_{t0}\#f_{t0}\}, F'_s=F_s-\{f_{s0}\#f_{s0}\}$$

【定義3・10】 $F_t=f_t\cup F'_t$ と $F_s=f_s\cup F'_s$ はフレームとする。ただし、 f_t と f_s は単純フレームで、 F'_t と F'_s は非アトムスロットからなるフレームである。

1. F_t と F_s の比較フレームを

$$F_t:F_s=f_t\sim f_s\cup F'_t:F'_s$$
 のように定義する。
2. F_t と F_s の類似度を次のように定義する。

$$DS_{F_t\sim F_s}=\max\left\{\int F_t:F_s\circ g, \int\{F'_t\#F_t\}:F_s\circ g, \int F_t:\{F'_s\#F_s\}\circ g\right\}$$
3. F_t と F_s の類似フレーム $F_t\sim F_s$ は類似度 $DS_{F_t\sim F_s}$ と対応する比較フレームである。単純フレームの比較フレームは類似フレームと等しいことに注意してほしい。

上記の定義を説明するために、次の例をあげる。
 【例3・6】

$$F_t=\{a_1\#0.1, a_2\#0.2, f\#f, F'_s\#F_s\}$$

$$F_s=\{a_2\#0.3, a_3\#0.3, f\#f\}$$

$$f=\{a_4\#0.4, a_5\#0.5, a_6\#0.6\}$$

とし、フレーム名 ' F_t ', ' F_s ', ' f ' は同一フレーム名クラスに属しているとする。

まず、フレーム F_t と F_s との直接比較によって、比較フレーム $F_t:F_s$ を求める。 F_t と F_s 内の非アトムスロットについて、' f ' と ' F_s ' は同一フレーム名クラスに属しているので、定義3・9の4(b)を適用する。 F_t と F_s 内の非アトムスロットから、それぞれ1個を選んで、類似スロットを作る組合せは2通りである：' $f\sim f\#1$ ' と ' $F_s\sim f\#0.75$ ' (その計算は定義3・10を用いて再帰的に行う)。 $DS_{f\sim f}=1>DS_{F_s\sim f}=0.75$ から、類似スロット ' $f\sim f\#1$ ' は比較フレーム $F_t:F_s$ のスロットに選ばれる。結局、 $F_t:F_s$ は次のようになる。

$$F_t:F_s=\{a_1\sim\phi\#0, a_2\sim a_2\#0.9048, \phi\sim a_3\#0, f\sim f\#1, F'_s\sim\phi\#0\}$$

$$\int F_t:F_s\circ g=0.5333$$

しかし、フレーム F_t 内に、非アトムスロットの値としてフレーム f と F_s が含まれていることから、 F_s は F_t 内の非アトムスロットの値と比較する必要がある。このため、 F_s を $F'_s=\{F'_s\#F_s\}$ のように書き換え、 F_t との比較フレーム $F_t:F'_s$ を求める。その結果を次に示す。

$$F_t:\{F'_s\#F_s\}=\{a_1\sim\phi\#0, a_2\sim\phi\#0, f\sim\phi\#0, F'_s\sim F'_s\#1\}$$

$$\int F_t:\{F'_s\#F_s\}\circ g=0.6667$$

一般に、 F_t は F_s 内の非アトムスロットの値と比較する必要がある。このため、 F_t を $F'_t=\{F'_t\#F_t\}$ のように書き換え、 F_s との比較フレーム $F'_t:F_s$ を求める。その結果を次に示す。

$$\{F'_t\#F_t\}:F_s=\{\phi\sim a_2\#0, \phi\sim a_3\#0, F'_t\sim f\#0.4615\}$$

$$F_t\sim f=\{a_1\sim\phi\#0, a_2\sim\phi\#0, f\sim f\#1, F'_s\sim\phi\#0\}$$

$$\int\{F'_t\#F_t\}:F_s\circ g=0.4615$$

この三つの比較フレームのファジィ積分のなかで、最大値を類似度として定義するのが妥当であろう。ゆえに

$$DS_{F_t\sim F_s}=\int F_t:\{F'_s\#F_s\}\circ g=0.6667$$

$$F_t\sim F_s=F_t:\{F'_s\#F_s\}.$$

【例3・7】
 リンゴ={色#{赤#0.8, 緑#0.15, 茶色#0.05},
 形状#{円#0.9, 円筒#0.1},
 表面#{滑らか#0.9, ざらざら#0.1}
 モモ={色#{赤#0.45, 白#0.45, 黄#0.1},

形状#{円# 0.8, 円錐# 0.2},
表面#{滑らか# 0.5, ざらざら# 0.5}}

とすると

リング～モモ={色～色# 0.2938,
形状～形状# 0.5,
表面～表面# 0.2019}

$DS_{\text{リング～モモ}}=0.2938$

ちなみに,

色～色={赤～赤# 0.2938, 緑～ ϕ # 0,
茶色～ ϕ # 0, ϕ ～白# 0, ϕ ～黄# 0}

形状～形状={円～円# 0.9048,
円筒～ ϕ # 0, ϕ ～円錐# 0}

表面～表面={滑らか～滑らか# 0.2019,
ざらざら～ざらざら# 0.2019}

明らかに, 色, 形状, 表面というフレーム名はそれぞれ異なるフレーム名クラスに属すべきである. ゆえに, リング～モモを求めるには, 色～形状, 形状～表面のような類似フレームを求める必要はない. したがって, フレーム名クラスを用いて, 計算速度を上げることができる.

ここで, 定義 3・10 を用いて, アトムスロットの値は集合, 手続き, 変数である場合について, アトムスロットの類似度を定義する.

【定義 3・11】 $s_t = a \# v_t$ と $s_s = a \# v_s$ は同一属性 a を持つアトムスロットとする.

1. v_t と v_s は手続きである場合, それらの手続きの実行結果の類似度を $DS_{v_t \sim v_s}$ として定義する.

2. v_t と v_s は実数または文字列の集合である場合
 $v_t = \{v_{t1}, \dots, v_{tn}\}$, $v_s = \{v_{s1}, \dots, v_{sm}\}$
とする(これはフレームのなかに, 複数のアトムスロットが同一属性名を持つ場合である).

s_t と s_s を次のようにフレームに書き換える.

$s_t \rightarrow F_t \{a_i \# f_{t1}, \dots, a_n \# f_{tn}\}$

$s_s \rightarrow F_s \{a_1 \# f_{s1}, \dots, a_m \# f_{sm}\}$

ただし, $f_{ti} = \{a \# v_{ti}\}$, $f_{sj} = \{a \# v_{sj}\}$ ($i=1, 2, \dots, n; j=1, 2, \dots, m$) はただ一つのアトムスロットを要素とする単純フレームである. この場合, $DS_{v_t \sim v_s} = DS_{F_t \sim F_s}$ のように定義する.

【例 3・8】 図 2 に示す英文字の形の類似性を表す重みつき無向木のなかで, 葉でない節点はその節点を親とする葉の集合と考えることができる. 例えば

$tailed = \{b, d, q, p, g\}$

$curved = \{b, d, q, p, g, a, e, c, o\}$

と表すことができる. それをスロットの形に書き換えると, 次のようになる.

'tailed'#tailed='tailed'#{b, d, q, p, g},

'curved'#curved='curved'#{b, d, q, p, g, a, e, c, o}

上の定義を適用すると, $DS_{tailed \sim curved} = 0.7143$.

ターゲットフレームは新しい問題の特徴を示しているため, ターゲットフレームのなかに, 値の未知または部分的に未知であるスロットが存在する可能性がある. この場合, 次の定義を与える.

【定義 3・12】 未知または部分的に未知である属性値は変数のペア (KV, PV) を用いて表される. ただし, KV は属性値の既知値の集合で, PV は属性値の可能値の集合である.

【定義 3・13】 $v_t = (KV_t, PV_t)$ は属性 a の未知または部分的に未知である値で, v_s は a の既知の値とする.

$$V_s = \begin{cases} \{v_s\} & (v_s \text{ は数値または文字列}) \\ v_s & (v_s \text{ は集合}) \end{cases}$$

とする. KV_t, PV_t と V_s は次のようにフレームに書き換える.

$KV_t \rightarrow KF_t = \{a_{k1} \# f_{tk1}, \dots, a_{kn} \# f_{tkn}\}$

$PV_t \rightarrow PF_t = \{a_{p1} \# f_{tp1}, \dots, a_{pm} \# f_{tpm}\}$

$V_s \rightarrow F_s = \{a_i \# f_{s1}, \dots, a_l \# f_{sl}\}$

ただし, $f_{tki} = \{a \# v_{tki}\}$, $v_{tki} \in KV_t$ ($i=1, 2, \dots, n$), $f_{tpj} = \{a \# v_{tpj}\}$, $v_{tpj} \in PV_t$ ($j=1, 2, \dots, m$), $f_{sk} = \{a \# v_{sk}\}$, $v_{sk} \in V_s$ ($k=1, 2, \dots, l$).

$FX = KF_t : F_s - \bigcup_{\phi \sim a_k \# 0 \in KF_t : F_s} \{\phi \sim a_k \# 0\}$,

$F'_s = F_s - \bigcup_{a_i \sim a_k \# DS_{a_i \sim a_k} \in FX} \{a_k \# f_{sk}\}$,

$FY = PF_t : F'_s - \bigcup_{a_i \sim \phi \# 0 \in PF_t : F'_s} \{a_i \sim \phi \# 0\}$.

とする. $DS_{v_t \sim v_s}$ は次のように定義される.

$$DS_{v_t \sim v_s} = \int (FX \cup FY) \circ g.$$

PV_t は属性 a のすべての値の集合である場合, 明らかに, $FY = \bigcup_{a_i \# f_{s_i} \in F_s} \{a_i \sim a_i \# 1\}$ である. ゆえに, FY は F'_s から直接得られる. さらに, $KV_t = \phi$ という条件を加えたとき, $DS_{v_t \sim v_s} = 1$.

【例 3・9】 本の情報はフレームで表されているものとする. そのようなソースフレームのなかに, 書名, 著者, 出版社などのスロット名が考えられる. 例えば, ソースフレームから, 著者の一人が King である本の情報を検索したい. ここで, 定義 3・2 を用いることにする. ターゲットフレームのなかで, 著者というスロットは

「著者#($KV_{\text{著者}}, PV_{\text{著者}})$ 」

のように表す. ただし, $KV_{\text{著者}} = \{King\}$ で, $PV_{\text{著者}}$ はデータベースにあるすべての本の著者の集合である. その他のスロットは, すべて

「属性#($KV_{\text{属性}}, PV_{\text{属性}})$ 」

のように表す。ただし、 $KV_{属性} = \phi$ で、 $PV_{属性}$ はその属性のすべての値の集合である。そうすると、上の定義を用いて、ターゲットフレームとソースフレームの類似度は、ソースフレームのなか、著者の一人が *King* である場合 1 で、それ以外の場合 0 である。

この例は、我々の方法が情報検索へ応用できることを示す。

この節では、フレームの類似度と類似フレームを定義した。定義 3・10 は最も一般的な定義である。

次のような条件が成立すれば、我々が定義した類似度と類似フレームはただ一つだけ存在する。

1. アトムスロットの値は集合ではない。または定義 3・2 を用いてすべてのアトムスロットの類似度を計算する。
2. 同一フレームに含まれる任意の二つの非アトムスロット名は同一フレーム名クラスに属しない。
3. 類似度は 0 ではない。

この結果は、この章の定義とファジィ積分の性質から容易に得られる。他のケースでは、比較フレーム F_i : F_s は唯一ではないことから、これに基づいて得られた類似フレームと類似度も異なる可能性がある。類似度を唯一にするには、すべての可能な比較フレームから最も類似度の高いものを選ぶように定義 3・10 を修正すればよいが、必ずしもその必要がない。その理由を次に示す。

1. 我々の目的は類似度を厳密に定義することではなく、与えられた問題と最も類似する事例を検索する基準と事例修正の情報を提供することである。類似性が高ければ高いほど、異なる比較フレームから得た類似度の誤差が少なくなり、どの類似度を使っても、最も類似する事例を検索することができる。次章のシミュレーションの結果から、それを検証できる。
2. このような類似度の定義の修正に伴う計算量が増加することになる。さらに、類似度を唯一に定義しても、類似フレームが異なる可能性が生じることから、必ずしも類似度を唯一にするのは好ましくない。

4. シミュレーション

我々は上の章に示したアルゴリズムに従って、AZ-Prolog 言語で類似性判定のプログラムを作成し、Sun-4 ワークステーション上でシミュレーションを行った。

シミュレーションのために、次のような条件を設定した。

1. すべてのアトムスロットの値は実数値で、 $[0, 1]$ に属する。ゆえに、 $R_{任意の属性} = 1$ 。
2. すべてのアトムスロットの重みは 1 である。
3. $\mu = 10$ とする。
4. フレームに含まれるアトムスロットと非アトムスロットの数はそれぞれ 10 と 3 以下である。
5. すべてのフレーム名は同一フレーム名クラスに属する。
6. フレームの深さは 4 以下である。

フレーム F の深さを $Depth(F)$ のように表し、これを次のように定義する。

- (a) F は単純フレームである場合

$$Depth(F) = 1$$

- (b) $Depth(F) = \max_{\substack{f \in F \\ f \text{ is a frame}}} Depth(f) + 1$

シミュレーションでは、このような条件のもとで、200 個フレームをソースフレームとしてランダムに作る。このなかから、ランダムに 100 個フレームを選ぶ。選ばれたそれぞれのフレームについて、そのフレームの n 変更フレーム ($n=1, 2, 3, 4, 5$) を作る。 F の n 変更フレームとは、 F を次のように変更して得たフレームである。 F のアトムスロットについて、ランダムに n_1 個アトムスロットを削除する、あるいは n_2 個アトムスロットの値を変更する。あるいは新しく n_3 個アトムスロットを追加する。ただし、 $n_1 + n_2 + n_3 = n$ である。さらに、 F の非アトムスロットに対して、同じように変更する。

まず、フレームとその n 変更フレームとの間の類似度を計算した。そして、選ばれた 100 個のフレームについて、次の条件を満たすかどうかを調べた。

$$\forall i \in \{1, 2, 3, 4, 5\} \quad (i \neq n) \quad (1)$$

$$\begin{cases} DS_{RF_n(F) \sim F} \geq DS_{RF_i(F) \sim F} & \text{if } n < i \\ DS_{RF_n(F) \sim F} \leq DS_{RF_i(F) \sim F} & \text{if } n > i \end{cases}$$

この条件を満たすことは、変更数の小さい変更フレームのほうが、他の変更フレームより、もとのフレームとの類似度が高いことを意味する。

F の n 変更フレームの i に対する条件 1 の満足度は、 i に対して、条件 1 を満たす場合は 0.25、満たさない場合は 0 とすると、 F の n 変更フレームの条件 1 の満足度は、すべての i (四つある) に対する満足度の和として定義される。それぞれの変更数 n に対して、選ばれた 100 個のフレームの n 変更フレームについて条件 1 の満足度を計算し、その平均を求めた。その結果を表 1 に示す。

この結果から、条件 1 の満足度は非常に高いことは明らかである。

次に、 n に対して、選ばれた 100 個のフレームとその

n 変更フレームの類似度の平均値, 最大値と最小値を求めた。その結果を表 2 に示す。

n は小さければ小さいほど, 類似度の平均値は大きくなることに注意してほしい。

最後に, 選ばれたフレーム F の n 変更フレームをターゲットフレームとして使い, そのターゲットフレームとそれぞれのソースフレームとの類似度を計算する。検索条件として, ターゲットフレームとの類似度が最も高い i ($i=1, 2, 3, 4$) 個ソースフレームを検索す

表 1 条件 1 の満足度

	n 変更					全ての n
	1	2	3	4	5	
平均満足度	0.9875	0.9350	0.9100	0.9050	0.9225	0.9320

表 2 変更数に対する類似度の平均値, 最大値と最小値

DS	n 変更				
	1	2	3	4	5
平均値	0.8408	0.7480	0.6932	0.6422	0.6017
最大値	0.9565	0.9000	0.8889	0.8000	0.8235
最小値	0.6923	0.5714	0.3571	0.4800	0.3333

表 3 変更数に対する検索率

i	n 変更				
	1	2	3	4	5
1	63%	63%	57%	47%	27%
2	100%	96%	89%	64%	35%
3	100%	97%	90%	66%	41%
4	100%	98%	94%	71%	47%

るものとし, F がソースフレームから検索された確率を計算する。その結果を表 3 に示す。

n は小さければ小さいほど, F の検索された確率は大きくなることに注意してほしい。

n はより小さいというのは, 変更数がより少ない, つまり F とその n 変更フレームがより類似することである。このシミュレーションの結果は, より類似するフレームの類似度はより 1 に近づき, その結果, ターゲットフレームと最も類似するソースフレームが検索される。このことは, 我々の方法が有効であることを示す。

5. む す び

本論文では, ファジィ積分を用いて, フレーム表現に基づく類似性判定の一方法を提案した。我々の方法を用いて, 検索の基準となる類似度と, 事例修正のマッチング情報を格納する類似フレームが得られる。類似度と類似フレーム両方を用いて, 幅広い類似性判定の要求を満たすことができる。今後の課題として, 本論文で定義した類似度と類似フレームについてより詳しい解析を進め, この方法を実際の推論システムの開発に応用すること, さらに, 他の知識表現方式における類似性の判定方法を研究することである。

◇ 参 考 文 献 ◇

- [Bareiss 89] Bareiss, R. and King, J. A.: Similarity Assessment in Case-Based Reasoning, *Proc. in Case-Based Reasoning Workshop (DARPA)*, pp. 67-71 (1989).
- [Iwayama 90] Iwayama, M., Tokunaga, T. and Tanaka, H.: A Method of Calculating the Measure of Salience in Understanding Metaphors, *Proc. AAAI*, pp. 298-303 (1990).
- [Kibler 87] Kibler, D. and Aha, D. W.: Learning Representative Exemplars of Concepts: An Initial Case Study, *Proc. 4th Int. Workshop on Machine Learning*, pp. 24-30 (1987).
- [Kobayashi 91] Kobayashi, K.: On a Searching Method of the Most Similar String in the File of a Document Retrieval System, *Trans. of IEICE*, J74-D-I, No. 1, pp. 39-49 (1991).
- [Kolodner 92] Kolodner, J. L.: An Introduction to Case-Based Reasoning, *Artificial Intelligence Review*, Vol. 6, pp. 3-34 (1992).
- [Sattath 77] Sattath, S. and Tversky, A.: Additive Similarity Trees, *Psychometrika*, Vol. 42, pp. 319-345 (1977).
- [菅野 72] 菅野道夫: Fuzzy 測度と Fuzzy 積分, 計測自動制御, Vol. 8, No. 2, pp. 218-226 (1972).
- [菅野 73] 菅野道夫: Fuzzy 測度の構成と Fuzzy 積分によるパターンの類似度評価, 計測自動制御, Vol. 9, No. 3, pp. 361-368 (1973).
- [Tversky 77] Tversky, A.: Features of Similarity, *Psychological Review*, Vol. 84, pp. 327-352 (1977).

[担当編集委員・査読者: 松本裕治]

著 者 紹 介



王 勇 (学生会員)

1983 年中国湖南師範大学数学科卒業。1990 年来日。1995 年名古屋工業大学大学院工学研究科博士前期課程電気情報工学専攻修了。現在, 同大学院工学研究科博士後期課程電気情報工学専攻在学中。人工知能, 特に推論, ファジィ理論, 遺伝的アルゴリズムに関する研究に従事。情報処理学会学生会員。



犬塚 信博 (正会員)

1987 年名古屋工業大学工学部情報工学科卒業。1992 年同大学院工学研究科博士課程電気情報工学専攻修了。工学博士。同年, 同大学電気情報工学科助手。1993 年同大学知能情報システム学科助手。現在に至る。1994 年より学術振興会海外特別研究員として英国インペリアルカレッジを訪問中。人工知能, 特に機械学習, 知識表現に関する研究に従事。



石井 直宏 (正会員)

1968 年東北大学工学部電気工学科卒業。1973 年同大学院工学研究科博士課程修了。工学博士。東北大学医学部助手, 名古屋工業大学情報工学科助教授を経て, 現在, 名古屋工業大学知能情報システム学科教授。この間, しきい値論理, 生体情報処理, 非線形システム解析, ニューラルネットワークの研究に従事。電子情報通信学会, 神経回路学会, IEEE 各会員。