

観点を考慮した連想機構の実現

沢田 裕司[†] 大川 剛直^{††} 馬場 口 登^{†††}

連想とは、ある事柄に対し、何らかの点で関連する事柄を想起する能力と捉えることができる。本論文では、連想の一形態である同種連想に着目し、連想機構の一モデル化を提案する。一般に事例間の類似性は、観点到に依存して変化する相対的尺度と考えられ、同種連想を実現する上で、観点を考慮することが必要不可欠と考えられる。そこで、複数の事例を連想のキーとして用いることにより、観点を考慮した連想メカニズムの実現を図る。ここでは、複数のキー事例の類似性を利用し、まず観点を定量的に特定した後、事例間の類似性の尺度である概念的隣接度に対し、属性の重みという形で観点を反映することにより、観点を考慮した連想メカニズムを実現した。また、古建築物を対象としたデータベースに対する検索実験を通じ、本手法の有効性を検討した。

Association Mechanism Based on Viewpoints

YUJI SAWADA,[†] TAKENAO OHKAWA^{††} and NOBORU BABAGUCHI^{†††}

Association is defined as ability to recall a case which is relevant or similar to another case given as key in some points. In this paper, we propose a modeling of association mechanism based on viewpoints. Our method makes use of more than one cases as the key in order to solve the problem that similarities between cases are affected by viewpoints. This method firstly evaluates viewpoints from the similarities between key cases quantitatively. Then, it retrieves similar cases from the casebase under the viewpoints, based on *conceptual contiguity* which is defined as a measure of the similarity between cases. From the experimental results for retrieving the database of old temples in Japan, we verified that our method is sufficiently flexible for viewpoints.

1. ま え が き

ある事柄から何らかの点で関連した別の事柄を想起する能力は、一般に連想 (association) と呼ばれている。想起される事柄がすでに経験したものであることを考慮すると、この連想のメカニズムは、記憶として蓄えられている過去に経験した事柄の中から、現在直面している事柄に関連するものを検索する枠組といえる。例えば、データベース検索や、事例ベース推論^{1),2)} (case based reasoning, CBR) における事例検索などは、連想の工学的な応用例と考えられ、これらの検索の高度化へ向けて、連想機構の実現に対する要求はきわめて高い。

ところで、連想はキーとなるもとの事例 (キー事

例, key case) からどのような事柄を想起するかにより、同種連想と異種連想の2種類に大別できる。同種連想では、「鳩」から「鷹」が想起されるように、キー事例と想起事例は同じ種類に属するもの、すなわち、ある点で類似したものとなる。一方の異種連想では、「鳩」から「平和」を想起するように、キー事例と想起事例は全く異なる種類となり、両者の間に直接的な類似性は存在しない。

前述のデータベース検索や CBR における事例検索などは、同種連想の応用といえ、例えば CBR では、現在の問題に類似した過去の問題の解決過程をもとに、問題解決が図られる。そこで、本論文では、特に同種連想に焦点をあて、そのメカニズムを実現する。

同種連想のメカニズムを実現する上で重要となるのが、事例間の類似性の尺度である。例えば CBR の分野でもさまざまな尺度が用いられており、その主なものとして、属性値の一致する属性数から類似性を判断するもの³⁾、属性値間の類似性をあらかじめシステム構築者が設定しておくもの^{4),5)}等を挙げることができるが、そのいずれもが類似性は不変の絶対的な尺度となっている。しかしながら、一般に事例間の類似性

[†] シャープ株式会社
Sharp Corporation

^{††} 大阪大学工学部情報システム工学科

Department of Information Systems Engineering,
Faculty of Engineering, Osaka University

^{†††} 大阪大学産業科学研究所
Institute of Scientific and Industrial Research,
Osaka University

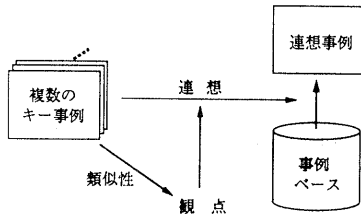


図1 観点を考慮した連想メカニズム
Fig. 1 An association mechanism based on viewpoints.

は、観点や視点に大きく依存する相対的な尺度^{6),7)}と考えられ、常に一定の尺度では類似性を正しく評価することができない。

例えば、“ワープロ”と“鉛筆”について考える。これらを利用者の立場から、「使用目的」という観点を重視して類似性を比較した場合、両者は類似したものと判断できるのに対し、生産者の立場から「製造工程」や「材質」等の観点を重視すると、両者は全く似ておらず、むしろ“ワープロ”は“TV”や“パソコン”等に類似していることになる。このように、観点により類似性は変化し、同一事例に対する検索結果は観点によって変化する。したがって、より柔軟な目的指向の検索を実現するためには、観点を考慮した連想メカニズムが必要不可欠といえる。

本研究では、図1のように連想のキーとして複数の事例を用意することにより、観点 (viewpoint) を考慮した連想メカニズムを実現する。ある観点のもとで類似していると判断された複数のキー事例間で高い類似性が得られるように、類似性の尺度を適応的に修正することにより、観点を特定する。その際、概念的隣接度 (conceptual contiguity) と呼ぶ、事例の一般化に基づく類似性尺度を新たに導入することにより、記号的に表現された事例間の定量的な類似性評価を可能としている。この結果、重み付けという比較的容易な操作により、類似性尺度自身の適応的な修正が実現される。修正された類似性尺度を用いて、キー事例と事例ベース内の事例の間の類似性を評価することにより、特定された観点のもとでの連想を実現している。

2. 事例間の類似度

本研究では、キー事例に類似した事例を既存の事例ベースから検索するという枠組で連想の実現を図る。このとき、事例間の類似性がきわめて重要な役割を果たす。本章では、まず、事例の記述法について定式化し、これをもとに事例間の類似度を定義する。

2.1 事例の記述

ここでは、1つの事例内に複数のオブジェクトが存在する事例を対象とし、各事例をさまざまなオブジェクトの集合という形で捉える。ここで、オブジェクトとは意味的にもしくは物理的に1つのまとまりと見なされる個体である。

各事例の特徴は、次に定義する属性記述と関係記述を用いて記述する。

定義1 (属性記述)

a を属性名、 v を a の属性値、 o をオブジェクト名とする。このとき、 $\langle a, v, o \rangle$ を属性記述と呼ぶ。

定義2 (関係記述)

r を関係名、 v を r の関係値、 o_1, o_2 をそれぞれオブジェクト名とする。このとき、 $\langle r, v, o_1, o_2 \rangle$ をオブジェクト o_1, o_2 の関係記述と呼ぶ。

属性記述と関係記述を用いて事例の特徴を記述したものを事例記述と呼ぶ。

例えば、「白く大きい三角形 o_1 が赤く小さい正方形 o_2 の上に存在する」事例の事例記述は

$\{ \langle \text{色}, \text{白}, o_1 \rangle, \langle \text{大きさ}, \text{大}, o_1 \rangle, \langle \text{形状}, \text{三角形}, o_1 \rangle, \langle \text{色}, \text{赤}, o_2 \rangle, \langle \text{大きさ}, \text{小}, o_2 \rangle, \langle \text{形状}, \text{正方形}, o_2 \rangle, \langle \text{相対的位置}, \text{上}, o_1, o_2 \rangle \}$

となる。

2.2 概念的隣接度に基づく事例間の類似度

2.2.1 オブジェクト間の対応

各事例の特徴は、これを構成するオブジェクトの特徴から求められるため、事例の類似性はそれらを構成しているオブジェクトの特徴の類似性に帰着できる。そこで、本研究では以下で述べるオブジェクト間の類似度と、オブジェクト間の関係の類似度を求めることにより、事例間の類似性を評価する。

このことを模式的に示したのが図2である。2つの事例間の類似度は、オブジェクト o_{11} と o_{21} の間の類似度 S_1 、 o_{12} と o_{22} の間の類似度 S_2 、および o_{11} と o_{12} の関係と o_{21} と o_{22} の関係の類似度 S_r をもとに評価される。

さて、このように事例間の類似性を評価する場合、以下の点に注意しなければならない。すなわち、図2

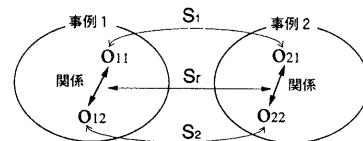


図2 事例間の類似性
Fig. 2 The similarity between two cases.

のように評価する場合、暗黙的にオブジェクト o_{11} と o_{21} が対応し、オブジェクト o_{12} と o_{22} が対応していることを仮定しているが、もし、 o_{11} と o_{22} が対応し、 o_{12} と o_{21} が対応するというように対応関係を変更した場合、これに伴い、事例間の類似性は変化する。したがって、複数のオブジェクトから成る事例間の類似性を評価する際には、オブジェクトの対応づけがきわめて重要となる。そこで、オブジェクトの対応を変化させながら、最も事例間の類似性が高くなるような対応を真の対応と考え、これに従い、事例間の類似性を評価する。以下では、オブジェクトの対応について定義する。

定義3 (事例間のオブジェクトの対応)

事例 c_1, \dots, c_n に存在するオブジェクトの集合をそれぞれ、

$$O_{c_1} = \{\dots, o_{1i}, \dots\}, \dots, O_{c_n} = \{\dots, o_{ni}, \dots\}$$

とする。このとき、 c_1, \dots, c_n 間のオブジェクトの対応 θ とは、 O_{c_1}, \dots, O_{c_n} の直積

$$O_{c_1} \times \dots \times O_{c_n}$$

の部分集合である。ただし、 θ 内のオブジェクトは重複しない。このとき、 $(o_1, \dots, o_n) \in \theta$ ならば、オブジェクト o_1, \dots, o_n が対応 θ のもとで対応しているという。

例えば、事例 c_1, c_2, c_3 に存在するオブジェクトを

$$O_{c_1} = \{o_{11}, o_{12}\}$$

$$O_{c_2} = \{o_{21}, o_{22}, o_{23}\}$$

$$O_{c_3} = \{o_{31}, o_{32}, o_{33}, o_{34}\}$$

オブジェクトの対応 θ を、

$$\theta = \{(o_{11}, o_{21}, o_{33}), (o_{12}, o_{22}, o_{31})\}$$

とすると、対応 θ のもとでオブジェクト o_{11}, o_{21}, o_{33} が対応、オブジェクト o_{12}, o_{22}, o_{31} が対応していることを意味する。なお、定義3の「 θ 内のオブジェクトは重複しない」という条件は、あるオブジェクトが他の同一事例内の複数のオブジェクトに対応しないことを保証するためのものである。

2.2.2 オブジェクト間の類似度

ここでは、比較対象となるオブジェクトの共通点を属性ごとに抽出し、属性値間の類似性をもとにオブジェクト間の類似性を評価するといったアプローチをとる。

まず、オブジェクト記述を定義する。これを用いて定義した一般化オブジェクトを求めることにより、比較対象となるオブジェクトの共通点抽出が可能となる。

定義4 (オブジェクト記述)

オブジェクト名が o である属性記述の集合

$$\{\langle a_1, v_1, o \rangle, \dots, \langle a_i, v_i, o \rangle, \dots, \langle a_p, v_p, o \rangle\}$$

をオブジェクト o のオブジェクト記述と呼ぶ。

定義5 (最小一般化)

属性値 v_1, \dots, v_k の最小一般化 (least generalization) とは、 v_1, \dots, v_k のどの属性値に対しても、等しいか、もしくは一般的となる属性値の中で、最も特殊な属性値のことを指し⁸⁾、 $Ig(v_1, \dots, v_k)$ と表記される。

定義6 (一般化オブジェクト)

オブジェクト o_1, \dots, o_k のオブジェクト記述を、それぞれ

$$\{\dots, \langle a_i, v_{1i}, o_1 \rangle, \dots\}, \dots, \{\dots, \langle a_i, v_{ki}, o_k \rangle, \dots\}$$

とする。このとき、

$$\{\dots, \langle a_i, Ig(v_{1i}, \dots, v_{ki}), \delta \rangle, \dots\}$$

により得られる概念上のオブジェクト δ を o_1, \dots, o_k の一般化オブジェクトと呼ぶ。

本手法では、最小一般化を得るために、属性ごとに図3のような一般-特殊階層木を用意している。例えば、属性値「三角形」と「四角形」の最小一般化は、「多角形」となる。また、次の3つのオブジェクト

$$o_1: \{\langle \text{色}, \text{黄色}, o_1 \rangle, \langle \text{形状}, \text{正方形}, o_1 \rangle\}$$

$$o_2: \{\langle \text{色}, \text{クリーム色}, o_2 \rangle, \langle \text{形状}, \text{四角形}, o_2 \rangle\}$$

$$o_3: \{\langle \text{色}, \text{黄色}, o_3 \rangle, \langle \text{形状}, \text{長方形}, o_3 \rangle\}$$

から生成される一般化オブジェクト δ は、属性ごとに属性値の最小一般化を求めることにより、

$$\{\langle \text{色}, \text{純黄系}, \delta \rangle, \langle \text{形状}, \text{四角形}, \delta \rangle\}$$

となる。

一般化オブジェクトの属性値は、もとの属性値が類似しているほど特殊なものになり、逆に類似していない場合にはより一般的な属性値になることがわかる。したがって、属性値の一般性を示す尺度を導入することにより、属性値間の類似性が評価できる。そこで、次のように属性値の一般度を定義する。

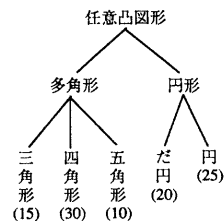


図3 形状に関する階層木

Fig. 3 An example of hierarchy about shapes.

定義7 (属性値の一般度)

$$g_a(a, v)$$

$$= \frac{\text{属性値が } v \text{ を満たす属性名が } a \text{ の属性記述数}}{\text{属性名が } a \text{ の事例ベース内の全属性記述数}}$$

例えば、属性値間に図3に示す一般-特殊階層木を用いた場合、属性値「四角形」、「円」の一般度は次のように計算される。

$$g_a(\text{形状, 四角形}) = 30/100 = 0.3$$

$$g_a(\text{形状, 円}) = 25/100 = 0.25$$

ただし、図中の括弧内の数字は、属性名が「形状」である属性記述のうち、その属性値を用いて記述されている事例ベース内の属性記述の総数を示している。

同様に、属性値「多角形」の場合、それを満たす属性記述数は、三角形の15、四角形の30、五角形の10の和から、一般度は

$$g_a(\text{形状, 多角形}) = 55/100 = 0.55$$

となる。

なお、一般-特殊階層が木構造ではなく通常のグラフ構造をなす場合、すなわち、ある属性値に対して、より一般的な値が複数種類存在する場合には、各属性値に対して可能なすべての最小一般化を求め、その属性値の一般度が最小となるものをもとに一般化オブジェクトを構成する。

さて、一般化オブジェクトの属性値に対して一般度を求めると、その逆数はもとの属性値の類似性を表す尺度になっている。ここで定義から明らかなように、類似性が事例ベース全体に依存して評価されることに注意されたい。これは類似性が対象となる事例ベースにおける属性値の生起確率に基づく相対的な尺度であるという考察に基づいている。

例えば、「オブジェクトのほとんどが四角形である」ような事例を集めた事例ベースを考える。このとき、ある「正方形」オブジェクトと別の「台形」オブジェクトの類似性を評価する際、どちらも「四角形」であるという特徴はあまり意味をなさず、類似性にはほとんど寄与しない。一方、同じ事例ベースにおいて、「楕円」オブジェクトと「真円」オブジェクトは「楕円」と「真円」という相違はあるものの、ともに「円形」という点で、きわめて類似したものと捉えられる。すなわち、事例ベース全体の中で、その属性値を有するオブジェクトの占める割合が高ければ、それは類似性にあまり寄与しないと考え、逆に割合が低ければ少々値が異なっても類似していると思える。

妥当と思われる。

次に、一般度をもとに、オブジェクト間の類似度を定義する。ここでは、オブジェクト間の類似性を、おのおの共通した特徴を数え上げることにより、評価することを基本方針とする。ここで、オブジェクトの記述に用いた各属性は独立であることを仮定すると、共通する特徴の数え上げは、属性ごとの類似性評価により実現できる。共通する特徴、すなわち、類似した属性値を持つ属性が多いほどオブジェクト全体としての類似性が高くなると考えられるため、オブジェクト間の類似性は、一般化オブジェクトの属性値ごとに得られた一般度の逆数の和から求めることが可能となる。ただし、一般にすべての属性が同等の重要度を示すとは限らないため、加重和により定義する。

また、一般度の逆数の平均や分散が属性によって変化する場合、一般度の逆数を正規化する必要がある。すなわち、ある属性 a_i の一般度の逆数の取りうる値が他の属性に比べて大きい場合、 a_i に関してあまり類似していない場合でも、他の類似している属性の値より大きくなる恐れがある。よって、属性ごとに一般度の逆数の平均 m_a と分散 σ_a^2 を求め、

$$1/\hat{g}_a(a, v) = \begin{cases} \frac{1/g_a(a, v) - m_a}{\sigma_a} & (1/g_a(a, v) \geq m_a) \\ 0 & (1/g_a(a, v) < m_a) \end{cases}$$

により正規化した一般度の逆数値 $1/\hat{g}_a(a, v)$ を利用する。

ここで、一般度の逆数が平均値以下となるとき、 $1/\hat{g}_a(a, v) = 0$ としている理由は次のとおりである。前述のように、ここではオブジェクト間の類似性を評価する際に、共通した特徴を数え上げることを基本原理としている。したがって、類似性評価時には、対応するオブジェクトが、どの程度類似しているかのみが重要となり、どの程度非類似であるかについては評価の対象とすることができない*。一方、一般度の逆数の平均値は、無作為に属性値を取り出した際の類似度の期待値を表しており、これを下回る値は、類似性に対しては何ら寄与せず、非類似性に対してのみ意味を持つことになる。したがって、共通する特徴の数え上げという側面から捉えた場合、この値を無視することが適切といえる。

定義8 (オブジェクト間の類似度)

オブジェクト o_1, \dots, o_k の一般化オブジェクト \bar{o} の

* 実際には、非類似に関する要因についても評価対象とするためには、対応がとれなかったオブジェクトについて考慮することが必要となる。

オブジェクト記述を

$$\{ \langle a_1, v_1, \delta \rangle, \dots, \langle a_i, v_i, \delta \rangle, \dots, \langle a_p, v_p, \delta \rangle \}$$

とし、属性 a_1, \dots, a_p の重みを $weight = (\omega_{a_1}, \dots, \omega_{a_p})$ とする。このとき、オブジェクト o_1, \dots, o_k 間の類似度 $S(o_1, \dots, o_k, weight)$ は

$$S(o_1, \dots, o_k, weight) = \sum_{i=1}^p \omega_{a_i} \cdot (1/\hat{g}_a(a_i, v_i))$$

で与えられる。 $S(o_1, \dots, o_k, weight)$ は $S(\delta, weight)$ と書かれることもある。ただし、属性の重みは、各属性の重要性を表す*。

2.2.3 オブジェクト間の関係の類似度

次に、オブジェクト間の関係の類似度を定義する。基本的には、オブジェクト間の類似度と同様に、次に示す関係値の一般度をもとに関係値の類似性から評価する。

定義 9 (関係値の一般度)

$$g_r(r, v) = \frac{\text{関係値が } v \text{ を満たす関係名が } r \text{ の関係記述数}}{\text{関係名が } r \text{ の事例ベース内の全関係記述数}}$$

オブジェクト間の類似度と異なり、オブジェクト間の関係の類似度を評価する場合にはオブジェクトの対応関係に注意を払わなくてはならない。

例えば、図 4 に示す 2 つの事例

- 事例 1 : 机 (オブジェクト o_{12}) の上りにんご (o_{11}) が存在する事例
 - 事例 2 : 2 つのテーブル (o_{22}, o_{23}) があり、片方のテーブル (o_{22}) の上にみかん (o_{21}) が存在する事例
- に対し、事例 1 と事例 2 の類似点を抽出することを考える。このとき、図の破線のように、オブジェクト o_{11} と o_{21} 、 o_{12} と o_{22} をそれぞれ対応づけている場合 (対応 1) を考えよう。このとき、事例 1 のオブジェクト o_{12} の上には o_{11} がのっているのに対し、 o_{12} に対応する事例 2 のオブジェクト o_{23} の上には何のオブ

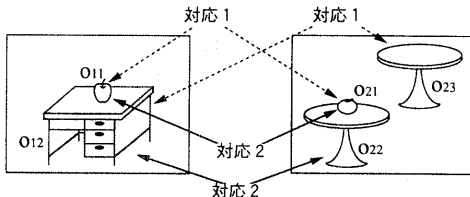


図 4 関係記述間の類似性

Fig. 4 The similarity between descriptions for relations.

ジェクトものっていないため、この場合には事例 1, 2 の共通点として「あるオブジェクトの上に別のオブジェクトがある」という特徴を挙げるのは妥当ではない。一方、図の実線のように、 o_{11} と o_{21} 、 o_{12} と o_{22} が対応している場合 (対応 2) には、このようなオブジェクト間の関係を共通点として抽出すべきであると考えられる。

そこで、関係記述間の類似度を、定義 3 に示したオブジェクトの対応 θ を用い、次のように定義する。

定義 10 (関係記述間の類似度)

事例 c_1, \dots, c_n 間のオブジェクトの対応を θ 、関係 r_1, \dots, r_q の重みを $weight = (\omega_{r_1}, \dots, \omega_{r_q})$ とする。このとき、 c_1, \dots, c_n の各事例記述内に存在する関係記述をそれぞれ、

$$d_1 = \langle r_{1i}, v_{1i}, o_{1s}, o_{1t} \rangle; o_{1s}, o_{1t} \text{ は } c_1 \text{ 内のオブジェクト, } r_{1i} \in \{r_1, \dots, r_q\}$$

$$\vdots$$

$$d_n = \langle r_{ni}, v_{ni}, o_{ns}, o_{nt} \rangle; o_{ns}, o_{nt} \text{ は } c_n \text{ 内のオブジェクト, } r_{ni} \in \{r_1, \dots, r_q\}$$

とすると、関係記述 d_1, \dots, d_n 間の類似度 $s_r(\theta, r_1, \dots, r_n, weight)$ を、 $r_{1i} = \dots = r_{ni}$ かつ、 $\{(o_{1s}, \dots, o_{ns}), (o_{1t}, \dots, o_{nt})\} \subseteq \theta$ に限り、

$$s_r(\theta, d_1, \dots, d_n, weight) = \omega_{r_{1i}} \cdot \frac{1}{\hat{g}_r(r_{1i}, lg(v_{1i}, \dots, v_{ni}))}$$

により評価する。このとき、 $\langle r_{1i}, lg(v_{1i}, \dots, v_{ni}), \delta_s, \delta_t \rangle$ を関係記述 d_1, \dots, d_n の一般化記述と呼び、関係記述 d_1, \dots, d_n は対応 θ のもとで対応しているという。

次に、事例間における関係の類似度を定義する。各事例記述には通常、複数の関係記述が存在するため、次のように関係記述のすべての組合せに対し類似度を求めることにより定義する。

定義 11 (事例間における関係の類似度)

事例 c_1, \dots, c_n 間のオブジェクトの対応を θ 、事例 c_i の関係記述の集合を d_{c_i} 、関係 r_1, \dots, r_q の重みを $weight = (\omega_{r_1}, \dots, \omega_{r_q})$ とする。このとき、 d_{c_1}, \dots, d_{c_n} の各要素のすべての組合せのうちで θ のもとで対応している関係記述間の類似度の和

$$S_r(\theta, c_1, \dots, c_n, weight) = \sum_{\substack{d_{1i} \in d_{c_1} \\ \vdots \\ d_{ni} \in d_{c_n}}} s_r(\theta, d_{1i}, \dots, d_{ni}, weight)$$

を、事例 c_1, \dots, c_n 間における関係の類似度と定義する。

* 重みは後述の観点の特定の結果、自動的に決定される。

2.2.4 概念的隣接度

ここでは、2.2.2項と2.2.3項で定義したオブジェクト間の類似度と、オブジェクト間の関係の類似度を用いて、事例間の類似度である概念的隣接度を定義する。

2.2.1項で説明したように、複数のオブジェクトが存在する事例の類似性を比較する場合、事例間のオブジェクトの対応関係が重要であり、対応関係が変化すれば、事例間の類似性も変化する。したがって、オブジェクト間の対応関係を変化させながら、類似性が最も高くなるときの値を概念的隣接度として定義する。

事例間の類似度を定義する際にも、事例間で共通する特徴を数え上げることを基本方針とする。本研究では、各事例には同様の重要性を有するオブジェクトがほぼ均質に存在するという前提を設ける。すると、対応付けられたオブジェクトをその類似の度合に応じて数え上げることに、事例全体の中の類似性評価が実現されると考えられる。そこで、事例を構成するオブジェクト間の類似度の和により、事例全体の類似度を評価するというアプローチを採用する。また、オブジェクト間の関係も事例を特徴づける重要な特徴の1つと見なされるため、事例間の類似度評価の対象として考慮する。ただし、属性を用いて規定されたオブジェクトそのものとオブジェクト間の関係を対等に扱うことはできないため、各属性、各関係の重要度に応じて、重みを独立に設定し、おのおの類似度の加重和により、概念的隣接度を定義する。

定義12 (概念的隣接度)

事例 c_1, \dots, c_n 間のオブジェクトの対応を θ , 対応するオブジェクトごとに求めた一般化オブジェクトを $\bar{o}_1, \dots, \bar{o}_l$ とする。また、属性および関係の重みをそれぞれ $w_a = (w_{a_1}, \dots, w_{a_p})$, $w_r = (w_{r_1}, \dots, w_{r_q})$ とし、 $weight = (w_{a_1}, \dots, w_{a_p}, w_{r_1}, \dots, w_{r_q})$ とする。このとき、 c_1, \dots, c_n 間の概念的隣接度 $cc(c_1, \dots, c_n, weight)$ を

$$cc(c_1, \dots, c_n, weight) = \max_{\theta} \left\{ \sum_{i=1}^l S(\bar{o}_i, w_a) + S_r(\theta, c_1, \dots, c_n, w_r) \right\}$$

と定義する。 ■

3. 観点を考慮した連想メカニズム

先に述べたように、類似事例を検索する際には観点が大きく関与するため、キーとなる事例を複数用いし、これらから観点を特定する。ここで、複数のキー事例とは、何らかの規準に従い、互いに類似していると判断され、与えられた事例である。この「何らかの

規準」が、利用者の観点到相当する。ここでは、事例を記述する際に用いた各属性が、おのおの、どの程度類似性に対して寄与しているかを定量的に評価することにより、観点を特定するというアプローチを採用する。すなわち、キー事例を記述している各属性に対して、各キー事例で用いられた属性値間の類似の度合を、一般度の逆数をもとに評価する。この結果、例えば、ある属性に関して属性値の類似の度合が平均的に高ければ、その属性はキー事例が類似していることに対する正当性を与える成分となり、逆に、これが平均的に低ければ、その属性はキー事例の類似にあまり貢献していないと考えられる。このように、キー事例の記述に用いられた属性値の類似の度合は、キー事例を互いに類似していると見なす際の、属性の重要度を与えるものとなる。したがって、この属性の重要度を重みとして概念的隣接度に反映させることにより、観点を考慮した連想メカニズムが実現される。

このような考え方に基づき、以下に、観点の定義を示す。

定義13 (観点)

連想のキーとなる事例を c_1, \dots, c_n とする。また、属性の重み $weight = (1, \dots, 1)$ に対して求めた事例 c_1, \dots, c_n の概念的隣接度を属性および関係ごとの和の形式を用いて表現し、

$$\begin{aligned} cc(c_1, \dots, c_n, weight) &= \sum_j 1/\hat{q}_a(a_i, v_{1j}) + \dots + \sum_j 1/\hat{q}_a(a_p, v_{pj}) \\ &\quad + \sum_j 1/\hat{q}_r(r_1, v_{1j}) + \dots + \sum_j 1/\hat{q}_r(r_q, v_{qj}) \end{aligned}$$

とする。このとき、属性 a_i , 関係 r_i の事例 c_1, \dots, c_n 間の類似性に対する寄与 $cont(a_i), cont(r_i)$ はおのおの、

$$\begin{aligned} cont(a_i) &= \frac{1}{l} \sum_{j=1}^l \frac{1/\hat{q}_a(a_i, v_{ij})}{l} \\ cont(r_i) &= \frac{1}{l} \sum_{j=1}^l \frac{1/\hat{q}_r(r_i, v_{ij})}{l} \end{aligned}$$

で与えられ、

$view$

$$= (cont(a_1), \dots, cont(a_p), cont(r_1), \dots, cont(r_q))$$

を観点と定義する。 ■

このようにして求めた観点 $view$ を属性の重み $weight$ として用いることにより、概念的隣接度に観点を反映させることが可能である。すなわち、属性 a_i の類似性に関する寄与 $cont(a_i)$ が大きいとき、事例がその属性に関して類似している（つまり事例を構成する個々のオブジェクトがその属性に関して類似して

いる)ほど概念的隣接度は大きい値を取る。したがって、この尺度を用いて連想を行うことにより、観点のもとで類似している事例ほど検索されやすくなる。また、観点にはほとんど寄与していないと考えられる特徴、すなわちキー事例間の類似性にあまり寄与していない属性の特徴に対しては属性の重みが0に近づき、あまり考慮されなくなる。

以上より、キー事例をもとに特定された観点、ならびに得られた観点を反映させた概念的隣接度を用いて、次のように連想が定義できる。

定義 14 (連想)

連想のキーとなる事例を c_1, \dots, c_n , 事例ベース内の事例の集合を C とし, c_1, \dots, c_n から特定される観点を $view$ とする. このとき, 各 c_i に対して, $cc(c_i, x, view)$ を最大にする事例 $x(x \in C)$ を想起することを連想と呼ぶ. ■

4. データベース検索への適用

提案した連想機構の有効性を確認するため、このメカニズムを LISP を用いてワークステーション上に実装し、神社仏閣を対象としたデータベースに対する検索実験を行った。

対象とする事例は、「興福寺」のような神社仏閣であり、オブジェクトはその中の国宝を中心とした建物である。例えば、事例「興福寺」においては「五重塔」「東金堂」がオブジェクトに相当する。各事例は、「所在地 (属性値数は 65)」、「型 (32)」、「様式 (17)」、「屋根 (29)」、「建立年代 (25)」、「指定 (5)」の 6 属性、および「隣接性 (関係値数は 4)」の 1 つの関係により記述され、97 事例が事例ベース内に蓄えられている*。事例「興福寺」の記述例の一部を図 5 に示す。

4.1 基本特性に関する検討

上で述べた事例ベースに対し、表 1 の実験 (1)~

興福寺: {<所在地, 奈良, 五重塔>, <型, 五重塔, 五重塔>, <様式, 和様, 五重塔>, <屋根, 方形造り, 五重塔>, <時代, 室町初期, 五重塔>, <指定, 国宝, 五重塔>, <所在地, 奈良, 東金堂>, ..., <隣接性, 横並び, 五重塔, 東金堂>}

図 5 “興福寺”の事例記述

Fig. 5 A part of description of Kofukuji-temple.

* なお、これらの属性は互いに全く無関係な属性であるとはいえないが、少なくとも閏数従属性が成立しない程度の独立性は存在し、手法の評価には十分である。

(6) の 6 通りのキー事例を用意し、各ペアごとにそれらに類似する事例を検索する実験を行った。各ペアのそれぞれの事例に対し、第 10 位まで候補事例を検索し、順位点 (第 1 位—10 点, 第 2 位—9 点, ..., 第 10 位—1 点) の合計により検索事例を決定した。特定された観点と検索事例を表 2 に示す。表は検索事例を上記の順位点の合計が高い順に第 3 位まで示したものである。なお、実験 (0) は、重みを (1, ..., 1) とした概念的隣接度に従い、単に「銀閣寺」に類似している事例を検索したものである。

以下では実験結果に基づき、提案手法の有効性を検討し、基本特性を明らかにする。

[1] 複数キー事例による観点特定の有効性

実験 (0) と実験 (1)、および実験 (0) と実験 (2) の検索結果を比較することにより、連想のキーとして複数の事例を用いる本手法の枠組の有効性を検討する。実験 (0) は従来の類似事例検索の手法を想定し、観点を考慮せずに検索した場合である。

表 2 の検索事例からわかるように、実験 (0) の第 1 位には「時代」に関して類似している事例が検索されたのに対し、第 2, 3 位には「所在地」と「型」に関して類似している事例が検索されており、検索事例に統一性が見られないという結果が得られた。これは、「銀閣寺」という 1 つの事例にさまざまな特徴が存在することに起因し、検索者にとり重要でない事例が検索される可能性が多分にある。一方、複数のキー事例を用いた実験 (1) と実験 (2) の場合には、それぞれのキー事例の類似点が属性の重みとして強調されることにより、上述のような問題は生じておらず、観点を考慮した本手法の枠組の有効性が確認できる。

[2] 観点に対する柔軟性

実験 (1) と実験 (2)、実験 (4) と実験 (5) の検索結果から、観点に対する柔軟性を検討する。これらは共に片方のキー事例を固定し、もう一方を変化させる

表 1 キー事例
Table 1 Key cases.

実験	キーとする事例	類似点
(0)	「銀閣寺」のみ	
(1)	「銀閣寺」と「清水寺」	京都府に存在する
(2)	「銀閣寺」と「金峯山寺」	室町後期の建物を含む
(3)	「功山寺」と「瑠璃光寺」	山口県に存在する
(4)	「円覚寺」と「正福寺」	関東地方にあり、 禅宗様式の建物を含む
(5)	「円覚寺」と「功山寺」	南北朝時代の 禅宗様式の建物を含む
(6)	「興福寺」と「唐招提寺」	横並びの建物を含む

表 2 特定された観点と検索事例
Table 2 Viewpoints and retrieved cases.

実験	観 点							検 索 事 例
	所在地	型	様式	尾根	時代	指定	隣接性	
(0)								1. 金峯山寺 (室町後期) 2. 醍醐寺 (京都府, 書院) 3. 高山寺 (京都府, 書院)
(1)	0.87	0.27	0.00	0.02	0.00	0.00	0.00	1. 醍醐寺 (京都府) 2. 浄瑠璃寺 (京都府) 3. 東寺 (京都府)
(2)	0.00	0.27	0.00	0.07	5.75	0.00	0.00	1. 不動院 (室町後期) 2. 大笠原神社 (室町後期) 3. 根来寺 (室町後期)
(3)	9.66	0.00	0.00	0.00	0.00	0.00	0.00	1. 住吉神社 (山口県) 2. 神魂神社, 出雲大社, 三仏寺 (島根県) (鳥取県) (鳥取県) 3. 厳島神社, 明王院, 他 4 事例 (すべて, 山陽地方)
(4)	7.14	0.00	3.65	0.13	0.00	0.00	0.00	1. 清白寺 (山梨県, 禅宗様) 2. 大善寺 (山梨県, 三種混入) 3. 永保寺, 安国寺 (岐阜県, 禅宗様)
(5)	0.00	0.00	3.65	0.13	2.31	0.00	0.00	1. 永保寺 (禅宗様, 南北朝) 2. 安楽寺 (禅宗様, 南北朝) 3. 明王院 (禅宗様系, 南北朝)
(6)	0.09	0.22	0.00	0.39	0.32	0.00	2.50	1. 法隆寺 (横並び) 2. 東大寺 (縦並び) 3. 室生寺 (近接)

ことにより、意図的に観点を変化させて実験した例である。

表からわかるように、それぞれのキー事例の類似点が観点の重みとして強調され、検索事例もそれぞれの観点に応じた結果になっている。特に、実験(1)と実験(2)のように1つの属性のもとで類似している場合だけでなく、実験(4)と実験(5)のように複数の属性に関して類似している場合にも、本手法が柔軟に対応できることが実験結果から確認できる。

[3] 概念的隣接度に基づく事例間類似度評価の有用性

実験(3)の検索結果より、類似性の定量的な評価尺度である概念的隣接度の有用性を検討する。

表に示すように、キー事例をもとに観点を特定した結果、属性「所在地」が重視されている。検索結果の第1位には、キー事例と同じ「所在地」である山口県の神社が検索されている。さて、この事例ベースには「所在地」が山口県である事例は他に存在しない。ここで、第2、3位の事例に注目されたい。これらは、重視した属性「所在地」に関して、必ずしもキー事例に一致しているわけではなく、比較的類似した事例となっている。これは、本来、半順序構造を成すため比較が困難な事例間の類似性に対し、概念的隣接度という一般度に基づいた定量的な尺度を導入し、任意の事例間の類似性を評価可能とすることにより成しえたものであり、その有用性が確認できる。

[4] 関係を考慮した連想の妥当性

実験(6)では、「隣接性」という関係を考慮した検索が実現されている。関係「隣接性」は、各事例内のオブジェクトが点在しているか、あるいは隣接しているか、そして、もし隣接しているならば、どのように並んでいるかについて、オブジェクト間の相対的な位置関係の形式で記述したものである。キー事例としては、オブジェクトが「横並び」に隣接したものが与えられており、その結果、第1位として、キー事例の1つである「興福寺」と同様に、「五重塔」と「金堂」が横並びの「法隆寺」が得られている。また、第2、3位にもおのおの国宝級の建築物が隣接している事例が検索されており、関係を考慮した連想の妥当性が示されている。

[5] 一般度の妥当性

以上の実験全体から、事例間の類似性の基本となる一般度の妥当性を検討する。すべての実験に対して、属性「指定」の重みが0となっていることに注意されたい。属性「指定」は、国宝や重要文化財などの指定の有無を記述する属性である。キーとして用いた事例内のオブジェクトのほとんどが国宝に指定されており、キー事例は属性「指定」に関して直観的にはきわめて類似しているにもかかわらず、キー事例の共通点として強調されていない。これは、一般度の定義が事例ベース全体を考慮していることに起因する。実際、事例ベース内の各事例のオブジェクトのほとんどが国宝である場合に、国宝という特徴をキー事例の特徴と

して強調するのは適切でなく、一般度の定義の妥当性が示されている。

4.2 キー事例の増加に対する連想特性の検討

同じデータベースに対して、キー事例を順次追加しながら検索し、キー事例の増加に対する本方式の特性を調べた。奈良県に所在する事例をキー事例として、2つから5つまで増加させながら与えることにより得られた検索事例を表3に示す。また、その際に求められた各属性の寄与の推移を図6に示す。なお、検索結果の順位の設定法は、上記の実験と同じである。

図6からわかるように、最初に与えられた2つのキー事例は「所在地」よりは、むしろ「屋根」、「様式」などの属性に関して共通した特徴を有すると判断

されており、類似した「屋根」造りを持つ事例が上位の検索結果として得られている。この結果は、奈良県に存在する事例がデータベース内に多数存在することから、わずか2つのキー事例の「所在地」に関する一致が偶然性に依るものと解釈できる。

これに対して、順次、奈良県に存在するキー事例を増加させることにより、次第に、「所在地」に関する寄与が向上していることがわかる。検索結果にもこれが反映されており、第1位として、「所在地」が奈良県である事例がすべて検索されている。また、第2位以降も、「所在地」が近畿地方の事例、「所在地」が西日本の事例というように、すべて、属性「所在地」を基準とした検索が実現されている。

このように、提案した連想メカニズムでは、キー事

表3 キー事例の数と検索事例

Table 3 Number of key cases and retrieved cases.

キー事例の数	キー事例	検索事例
2	円成寺 宇太水分神社	1. 春日大社 (春日造り, 奈良県) 2. 神魂神社 (大社造り, 島根県) 3. 宇治上神社 (流造り, 京都府)
3	円成寺 宇太水分神社 石上神宮	1. 春日大社 (春日造り, 本殿) 2. 宇治上神社 (流造り, 本殿) 3. 伊勢神宮 (神明造り, 本殿)
4	円成寺 宇太水分神社 石上神宮 室生寺	1. 靈山寺 (奈良県, 鎌倉後期) 2. 元興寺, 興福寺, 東大寺 (奈良県, 鎌倉初期) 3. 唐招提寺 (奈良県, 天平)
5	円成寺 宇太水分神社 石上神宮 室生寺 法輪寺	1. 靈山寺, 他 17 事例 (全て奈良県) 2. 銀閣寺, 他 39 事例 (全て近畿地方) 3. 出雲大社, 他 18 事例 (全て西日本)

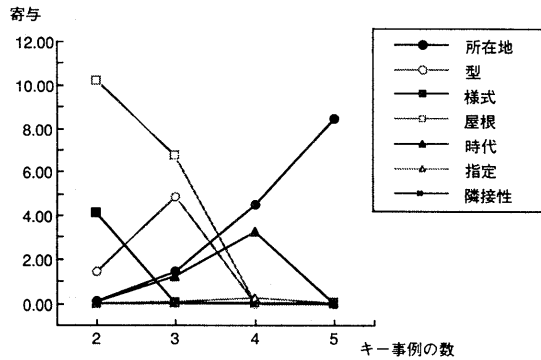


図6 キー事例の数と観点
Fig. 6 Number of key cases and viewpoints.

表4 重みの変化と検索事例
Table 4 Weight and retrieved cases.

実験	$\omega = cont$	$\omega = (cont)^2$	$\omega = (cont)^{1/2}$
(1)	1. 醍醐寺 2. 浄瑠璃寺 3. 東寺	1. 醍醐寺, 浄瑠璃寺 2. 東寺 3. 宇治上神社	1. 醍醐寺 2. 浄瑠璃寺, 東寺 3. 宇治上神社
(2)	1. 不動院 2. 大笠原神社 3. 根来寺	1. 不動院 2. 大笠原神社, 根来寺 3. 興福寺	1. 不動院 2. 根来寺 3. 大笠原神社
(3)	1. 住吉神社 2. 神魂神社, 他 2 事例 3. 厳島神社, 他 5 事例	1. 住吉神社 2. 神魂神社, 他 2 事例 3. 厳島神社, 他 5 事例	1. 住吉神社 2. 神魂神社, 他 2 事例 3. 厳島神社, 他 5 事例
(4)	1. 清白寺 2. 大善寺 3. 永保寺, 安国寺	1. 清白寺 2. 大善寺 3. 永保寺, 安国寺	1. 清白寺 2. 大善寺 3. 永保寺, 安国寺
(5)	1. 永保寺 2. 安楽寺 3. 明王院	1. 永保寺 2. 安楽寺 3. 明王院	1. 永保寺 2. 安楽寺 3. 明王院
(6)	1. 法隆寺 2. 東大寺 3. 室生寺	1. 法隆寺 2. 東大寺 3. 室生寺	1. 法隆寺 2. 東大寺 3. 室生寺

例を増やすことにより、利用者の観点をより明確化することが可能である。

一方、例えば、10個のキー事例のうち、9つは「所在地」が奈良県で、1つが京都府である場合と、奈良県の事例と京都府の事例がともに5つずつ存在する場合のように、キー事例における属性値の多少に対して、有意な結果の差異は得られない。これは、上に挙げた例のいずれに対しても、奈良県と京都府の最小一般化である近畿地方を基準としてキー事例間の類似性が評価されることに起因する。これに対処するためには、類似性評価の方式をさらに拡張する必要があり、今後の課題の1つと考えられる。

4.3 観点による重み付けの妥当性に関する検討

最後に、概念的隣接度の定義において、属性の重要度を示す尺度として用いた重み付けの妥当性について評価する。ここでは、4.1節における実験(1)~(6)と同じキー事例を対象として、重みの作用度合を変化させ検索した。具体的には、特定された各属性の寄与の自乗、あるいはその平方根を求め、これを概念的隣接度を計算する際の重み ω として用いた。

検索結果を表4に示す。表4からわかるように、実験(3)~(6)に関しては、検索結果の上位3位に対し、すべて同様の結果が得られている。また、実験(1)、(2)に関しても、若干、順位の変動はあるものの、それほど大きな結果の差異は導かれなかった。以上の実験から、概念的隣接度は重みの作用度合の変化に対して、敏感でないことがわかる。すなわち、提案メカニズムの枠組において、概念的隣接度の重み付け方法はそれほど重大な問題でなく、逆にいえば、定義13に従って求められた属性の寄与をそのまま重みとして利用することの妥当性が示唆される。

5. む す び

複数の事例から類似事例の検索が可能な連想機構を実現した。また、古建築物を対象としたデータベースに対する検索実験により、提案手法の種々の特徴を確認した。提案手法の特徴を列挙すると以下のとおりである。

- 複数のキー事例を用いることにより、観点を考慮した連想が実現できる。
- 複数のオブジェクトから構成される事例間の類似性評価が可能である。
- オブジェクト間の関係を考慮した事例間の類似性評価が可能である。

- 事例ベース内の属性値の存在の割合を考慮した事例間の類似性評価が可能である。
- 半順序構造を成すため直接的な比較が困難な事例間の類似性を定量的に評価可能である。

事例間の類似性の評価尺度となっている概念的隣接度は、事例を構成するオブジェクト間の類似性のみを考慮したものとなっている。このため、オブジェクトの個数が増えるに従い、事例間の類似性は単調に増大するという問題がある。これに対しては、オブジェクト間の非類似性を考慮することにより解決できると考えられ、類似性、非類似性の両者に基づく新しい概念的隣接度の設定が残された課題である。

謝辞 本研究に関してご指導を頂いた、故手塚慶一大阪大学名誉教授に心より御礼申し上げる。なお、本研究の一部は、文部省科学研究費、電気通信普及財団の補助による。

参 考 文 献

- 1) Kolodner, J. L. et al.: A Process Model of Case-Based Reasoning in Problem Solving, *IJCAI-85*, pp. 284-290 (1985).
- 2) 小林重信: 事例ベース推論の現状と展望, *人工知能学会誌*, Vol. 7, No. 4, pp. 559-566 (1992).
- 3) 中村孝太郎: シガレット製造機の故障診断支援, *人工知能学会誌*, Vol. 7, No. 4, pp. 576-581 (1992).
- 4) 島津秀雄, 柴田晃宏, 高島洋典: 関係データベースを使った事例ベース検索(1)アルゴリズム, 第45回情報処理学会全国大会論文集, 6H-5, pp. 2-175-2-176 (1992).
- 5) 黄 錦法, 吉村 晋, 白鳥則郎: 事例ベースを適用した HSC に基づいた通信ソフトウェア開発環境, *信学技報*, AI-92-94, pp. 33-40 (1993).
- 6) 沢田裕司, 大川剛直, 馬場口登, 手塚慶一: 概念的隣接度に基づく連想メカニズムのモデル化の一提案, 第45回情報処理学会全国大会論文集, 6H-1, pp. 2-167-2-168 (1992).
- 7) 沢田裕司, 大川剛直, 馬場口登, 手塚慶一: 観点を考慮した連想機構の一モデル化, *情報処理学会研究会報告*, 92-FI-28-2, pp. 9-16 (1992).
- 8) Kodratoff, Y.: *Introduction to Machine Learning*, Morgan Kaufman (1988).

(平成5年4月14日受付)

(平成6年1月13日採録)



沢田 裕司 (正会員)

昭和 43 年生. 平成 3 年大阪大学工学部通信工学科卒業. 平成 5 年同大学院工学研究科通信工学専攻前期課程修了. 現在, シャープ(株)に勤務. 在学中は帰納学習の研究に従事. 人工知能学会会員.



大川 剛直 (正会員)

昭和 38 年生. 昭和 61 年大阪大学工学部通信工学科卒業. 昭和 63 年同大学院工学研究科前期課程修了. 平成 2 年同後期課程退学. 同年大阪大学工学部通信工学科助手. 現在, 大阪大学工学部情報システム工学科助手. 博士(工学). 機械学習, 知識獲得, 定性シミュレーションなどの研究に従事. 電子情報通信学会, 人工知能学会, 電気学会などの会員.



馬場口 登 (正会員)

昭和 32 年生. 昭和 54 年大阪大学工学部通信工学科卒業. 昭和 56 年同大学院前期課程修了. 昭和 57 年愛媛大学助手, 大阪大学工学部助手, 講師を経て, 現在, 大阪大学産業科学研究所助教授. 工学博士. 人工知能, パターン認識, 画像処理の研究に従事. IEEE, 電子情報通信学会, 人工知能学会各会員.