

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

6H-1

沢田 裕司¹ 大川 剛直² 馬場口 登¹ 手塚 慶一³

¹大阪大学工学部 ²大阪大学情報処理教育センター ³関西大学工学部

1. はじめに

ある事例から何らかの点で関連した別の事例を想起する能力は、一般に連想と呼ばれている。事例に基づく推論(CBR)^[1]における類似事例の探索、データベースにおける検索機構などは、この連想を工学的に実現した一例と捉えることができる。一般に事例間の類似性は観点や視点といったものに大きく依存するため、高度かつ柔軟な連想メカニズムを実現するためには、観点を考慮に入れることが必要不可欠になると考えられる。

本稿では、連想のキーとなる複数の事例を用意し、これらの類似性を利用して観点を特定したのち、事例間の類似性を表わす概念的隣接度(conceptual contiguity)を用いて事例を想起する連想メカニズムを提案する。

2. 観点と連想事例

人間がある事例から類似した別の事例を想起するとき、一般に連想される事例は、連想を行う人により、また状況によっても異なる。これは、一つの事例にも様々な特徴が存在するため、どのような特徴を重視するか、すなわちどのような観点で連想を行うかにより、類似する事例が異なるためと考えられる。従って、連想メカニズムのモデル化を行う場合、観点を考慮したメカニズムを考慮していく必要がある。そこで、複数の事例をキーとして用い、これらの類似性から観点を特定したのち連想を行うことにより、観点を考慮した連想メカニズムの実現を図る。図1にその概略を示す。

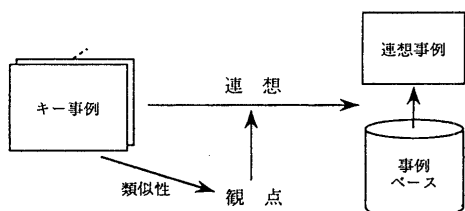


図1 観点を考慮した連想メカニズム

3. 概念的隣接度

3.1. 事例の記述

本研究では事例を様々なオブジェクトの集合と捉え、個々のオブジェクトの特徴を記述することにより、事例を表現する。ここでオブジェクトは、意味的にもしくは物理的に1つのまとまりと見なし得る個体である。

オブジェクト、事例の特徴は、次に定義する単位記述を用いて、定義1、定義2に従い記述する。

【定義1】オブジェクト記述

aを属性名、oをオブジェクト名、vをaの属性値とする。このとき、 $\langle a, o, v \rangle$ を単位記述と呼び、オブジェクト名としてoをもつ単位記述の集合

$$\{ \langle a_1, o, v_1 \rangle, \dots, \langle a_i, o, v_i \rangle, \dots, \langle a_n, o, v_n \rangle \}$$

をoのオブジェクト記述と呼ぶ。 □

【定義2】事例記述

事例c中に存在するオブジェクトを o_1, \dots, o_k とする。

o_1, \dots, o_k のオブジェクト記述の要素の集合

$$\{ \langle a_i, o, v_i \rangle \mid o \in \{ o_1, \dots, o_k \} \}$$

をcの事例記述と呼ぶ。 □

3.2. 概念的隣接度

ここでは事例間の類似性を表わす尺度、概念的隣接度を定義する。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、事例間の類似性はそれらを構成しているオブジェクト間の類似性に帰着する。そこで、まずオブジェクト間の類似度を定義する。

オブジェクト間の類似度は、次に定義する一般化オブジェクトを基準に求める。

【定義3】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

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

とする。このとき、

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

により記述さ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。但し、 $lg(v_{1i}, \dots, v_{ki})$ は v_{1i}, \dots, v_{ki} の最小一般化である。 □

ここで、属性値 v_1, \dots, v_k の最小一般化とは、 v_1, \dots, v_k のどの属性値に対しても、等しいか、もしくは一般的な属性値(一般化と呼ぶ)の中で、最も特殊な属性値のことを指す^[2]。例えば、属性値“三角形”“四角形”の最小一般化は“多角形”になる。

さて、この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るという特徴を持つため、属性値の一般性を示す尺度を定義することにより、属性値の類似性を評価できると考えられる。そこで、属性値の一般度を次のように定義し、オブジェクト間の類似度を定義する。

【定義4】属性値vの一般度g(v)

$$g(v) = \frac{\text{属性値}v\text{を満たすオブジェクト数}}{\text{事例ベース内の全オブジェクト数}}$$

【定義5】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 δ のオブジェクト記述を

$$\{ \langle a_1, \delta, v_{11} \rangle, \dots, \langle a_i, \delta, v_{i1} \rangle, \dots, \langle a_n, \delta, v_{n1} \rangle \}$$

とする。このとき、オブジェクト o_1, \dots, o_k 間の類似度 $s(o_1, \dots, o_k)$ は

$$s(o_1, \dots, o_k) = \sum_{i=1}^m (1/g(v_i)) / m$$

で与えられる。 $s(o_1, \dots, o_k)$ は $s(\delta)$ と書かれることもある。 □

次に事例間の類似性を示す概念的隣接度を定義する。概念的隣接度は、オブジェクト間の類似度より求められるが、このとき重要となるのがオブジェクトの対応関係である。例えば、図2において、 o_{11} と o_{22} 、 o_{12} と o_{21} を対応づけて事例間の類似性を評価すると、たとえ同じ事例であっても事例間の類似性は低くなる。すなわち、事例間のオブジェクトの対応を変化させると、事例間の類似性も変化する。そこで、オブジェクトの対応を次のように定義し、オブジェクト間の類似度の平均を最大にするようなオブジェクトの対応を用いて概念的隣接度を定義する。

case1	case2
o_{11} : <shape O11 triangle> <color O11 red> <size O11 small> ⋮	o_{21} : <shape O21 triangle> <color O21 red> <size O21 small> ⋮
o_{12} : <shape O12 oval> <color O12 white> <size O12 big> ⋮	o_{22} : <shape O22 circle> <color O22 white> <size O22 small> ⋮

図2 オブジェクトの対応と事例間の類似性

【定義6】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をそれぞれ、 $o_{c_1} = \{\dots, o_{1i}, \dots\}, \dots, o_{c_n} = \{\dots, o_{nj}, \dots\}$

とする。このとき、 c_1, \dots, c_n 間のオブジェクトの対応 θ 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。但し、 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 が対応しているという。 □

【定義7】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 とし、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 $\delta_1, \dots, \delta_\theta$ とする。このとき、 c_1, \dots, c_n 間の概念的隣接度 $CC(c_1, \dots, c_n)$ を、

$$CC(c_1, \dots, c_n) = \max_{i=1}^{\theta} s(\delta_i) / \theta$$

と定義する。 □

4. 観点を考慮した連想メカニズム

先に述べたように、連想を行う際には観点が大きく関与する。従って、連想を行う前に観点を特定する必要がある。そこで、連想のキーとなる複数の事例から次の定義9に従い、まず観点を特定する。これはキーとなる事例がどの様な特徴に関して類似しているかを評価することにより、どの特徴を重視して連想を行えばよいか把握できるという考えに基づいている。

【定義8】観点

連想のキーとなる事例 c_1, \dots, c_n に対し、 $CC(c_1, \dots, c_n)$ を与えるオブジェクトの対応 θ をもとに、対応するオブジェクト毎に求めた一般化オブジェクト $\delta_1, \dots, \delta_\theta$ のオブジェクト記述を

$$\langle a_i, \delta_j, v_j \rangle \quad (1 \leq j \leq \theta)$$

とする。このとき、

$$cont(a_i) = \sum_{j=1}^{\theta} (1/g(v_j)) / \theta$$

を属性 a_i の c_1, \dots, c_n 間の類似性に対する寄与とい

い、

$$view = \{cont(a_1), \dots, cont(a_m)\}$$

を観点と定義する。 □

このようにして求めた観点をを用いて、オブジェクト間の類似度、概念的隣接度を次のように再定義し、キーとなる事例に最も類似した事例を想起する。これにより、観点を考慮した連想メカニズムが実現できる。

【定義9】観点を考慮した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 δ のオブジェクト記述を

$$\langle a_1, \delta, v_1 \rangle, \dots, \langle a_i, \delta, v_i \rangle, \dots, \langle a_m, \delta, v_m \rangle$$

とし、観点を $view = \{cont(a_1), \dots, cont(a_m)\}$ とする。このとき、観点を考慮したオブジェクト間の類似度 $S(o_1, \dots, o_k, view)$ は

$$S(o_1, \dots, o_k, view) = \sum_{i=1}^m (cont(a_i) \cdot (1/g(v_i))) / m$$

で与えられる。 $S(o_1, \dots, o_k, view)$ は $S(\delta, view)$ と書かれることもある。 □

【定義10】観点を考慮した概念的隣接度

事例 c_1, \dots, c_n 間のオブジェクトの対応を θ とし、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 $\delta_1, \dots, \delta_\theta$ とする。このとき、観点を考慮した c_1, \dots, c_n 間の概念的隣接度 $CC(c_1, \dots, c_n, view)$ は、

$$CC(c_1, \dots, c_n, view) = \max_{i=1}^{\theta} S(\delta_i, view) / \theta$$

と定義される。 □

以上より、連想は次のように定義できる。

【定義11】連想

連想のキーとなる事例を c_1, \dots, c_n 、事例ベースに蓄えられている個々の事例を X_i とし、 c_1, \dots, c_n から求められる観点を $view$ とする。このとき、

$$\max_i \{CC(c_1, \dots, c_n, X_i, view)\}$$

となる事例 X を想起することを、連想と呼ぶ。 □

5. おわりに

概念的隣接度に基づく連想メカニズムを提案した。本手法は、観点により事例間の類似性が変化することに着眼し、複数のキーとなる事例の類似性から観点を特定したのち連想を行うことにより、観点を考慮した連想メカニズムを実現している。また、帰納学習的な考え方を導入することにより、nominalに記述された事例間の類似性を評価することも可能となっている。

最後に、この連想メカニズムのデータベース検索への応用について触れておく。提案した連想メカニズムはある事例に類似した事例を想起するものであるが、一つの事例を1個のデータと捉えれば、あるデータに類似したデータを取り出すという、高度なデータ検索が可能となる。また、検索者の観点により検索すべきデータが変化してしまうという類似検索における問題点も、本手法のメカニズムを採用することにより解決可能となる。

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
 [2] Yves Kodratoff : "Introduction to Machine Learning", Morgan Kaufmann, 1988