

名詞述語文からの概念学習システム — 継承と経路情報の処理 —

佐々木淳一 富永睦 荒木健治 桃内佳雄

北海学園大学

我々は、従来より、自然言語理解システム構築のための基礎として、名詞述語文からの概念学習システムを作成し、いくつかの実験を行ってきた。今回、これまでの実験の結果を踏まえて、概念空間における概念区分、学習および解析ヒューリスティクスの詳細化、継承処理の追加などのシステムの改良を行った。また、概念関係経路情報という新しい手法を取り込み、より学習能力を強化するようにした。本稿では、新しいシステムにおける学習と解析処理のためのヒューリスティクスと、新たに提案される継承処理および概念関係経路情報の利用処理について述べる。

LEARNING SYSTEM OF CONCEPTS FROM COPULAR SENTENCES - INHERITANCE AND PATH INFORMATION -

Jun-ichi SASAKI Mutumi TOMINAGA Kenji ARAKI Yoshio MOMOUCHI

Hokkai-Gakuen University

S26-W11, Chuo-ku, Sapporo, 064, JAPAN

We presented a system for learning concepts from copular sentences, as a basis for constructing natural language understanding system, and carried out some experiments. On the basis of the experiments, we have improved the system on the division of concept space, the details of heuristics for learning and analysis, and the inheritance processing. We also presents a new method of using the path information of the relation between concepts in order to reinforce the learning capacity of the system. In this paper, we describe the heuristics for learning and analysis, the inheritance processing and the path information of the relation between concepts in the new system.

1 はじめに

我々は、従来より、自然言語理解システム構築のための基礎として、名詞述語文からの概念学習システムを作成し、いくつかの実験を行ってきた^{1,2,3)}。今回、これまでの実験の結果を踏まえて、システムの改良を行った。特に、学習と解析の処理をより詳細に組み立て直した。本稿では、新しいシステムにおける学習と解析処理のためのヒューリスティクスと、新たに提案される継承処理および概念関係経路情報の利用処理について述べる。

システムの主な改良は、①概念空間における概念区分、②学習および解析ヒューリスティクスの詳細化、③継承処理の追加、④概念関係経路情報の利用などに関連して行われた。

継承処理とは、上位概念の属性が下位概念に継承される過程を実現したものである。

概念関係経路情報の利用処理は、従来より行ってきた解析処理に代わるものである。解析処理はヒューリスティクスを用いており、これでは学習範囲の拡張など、さまざまな状況の変化に対処可能な汎用性に欠ける。一方、概念関係経路情報利用処理では、解析ヒューリスティクスに相当する概念関係経路情報を自動的に獲得するため、その文は汎用性が高まると考える。この手法を組み込んだ実験システムを構築し、評価実験を行った。

なお、本稿から、従来の解析処理を概念間の意味関係利用学習処理と名称を変更した。

2 概念区分

本手法における概念区分、つまり概念空間は、横軸を基準概念と複合概念とし、縦軸を既知概念と未知概念とする2次元の4つの区分からなる。基準概念とは、〈花〉や〈太郎〉などの1つの名詞に対応する概念である。複合概念とは、「〈名詞〉の〈名詞〉」、「〈名詞〉と〈名詞〉」に対応する概念のことである。例えば、「菊の花」や「太郎と次郎」などである。既知概念は概念空間に存在している概念であり、未知概念は概念空間に存在していない概念である。また、各々の概念は2種類のマークを有している。1つは、個概念か、集合概念かを印づけるマークである。なお、個概念と集合概念の違いは、その概念が実体か、集合かである。もう1

つは、概念の内容、型による分類で、対象概念か、属性概念かを印づけるマークである。対象概念とは、対象（もの）の概念であり、属性概念とは、事象（できごと）、状態、性質などの概念である。また、属性概念は抽象的であるので集合概念として構成する。

3 知識表現と概念の生成

本手法では、知識表現としてS. C. Shapiroらによって提唱された意味ネットワークSNePS^{6, 8)}を基に作成した意味ネットワークを用いている。この意味ネットワークでは、ノード（節）が、命題、対象、属性、関係などの概念を表し、アーク（弧）は、それらの概念間の構造的、意味的つながりを表す。2つの概念間の関係は2つの概念間を結ぶノードのアークに意味関係を付加して表現する。2つの概念間を結ぶノードとアークの連なりをパスと呼ぶ。このパスには、直接のパスと冗長なパスの2種類がある。直接のパスは、1つのノードで2つの概念間を結ぶ経路のことであり、冗長なパスは、2つ以上のノードで結ぶ経路のことである。

概念生成は、未知概念のとき起こり、概念空間における未知の基準概念か複合概念として生成される。また、未知概念は、1文入力処理終了後には既知概念に組み込まれる。

4 名詞句概念間の意味関係

我々が対象にした名詞述語文とは、主語と述語の対立の中で、述語が名詞句で作られている文のこと⁴⁾で、「〈名詞句1〉は〈名詞句2〉です。」と表される。

本稿が対象とする名詞述語文における名詞句概念間の基本的な意味関係を以下に示す。

(1) 下位・上位関係: $X \leftarrow \text{sub-} : -\text{su} \rightarrow Y$

XとYは集合概念であり、YがXの上位概念を表す。

(2) 要素・集合概念: $X \leftarrow \text{mem-} : -\text{class} \rightarrow Y$

Xが個概念で、Yが集合概念であり、XがYの要素を表す。

(3) 同一関係: $X \leftarrow \text{eq-} : -\text{eq} \rightarrow Y$

XをYと名付けること、または、XとYが同一概念であることを表す。

(4) 対象・属性関係: $X \leftarrow \text{ob-} : -\text{prop} \rightarrow Y$

Yが集合概念であり、Xの付属物がYで

あることを表す。

なお、属性には、性質、状態、部分、事象などがある。これらのより詳細な分類は、今後の課題とし、今回、対象とする属性は、性質、状態、部分の3つである。

5 学習システムの概要

本システムの処理の流れを図1に示す。

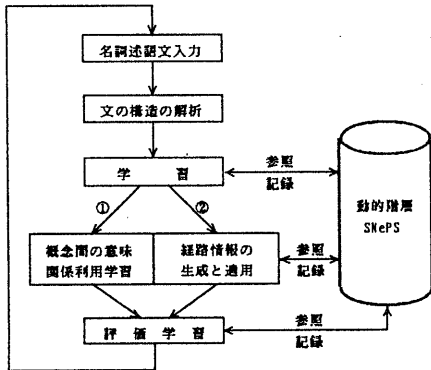


図1 処理過程

本システムでは、名詞述語文が入力されると、はじめに文の構造を解析する。これは、〈名詞句1〉と〈名詞句2〉の抽出および名詞句内の構造を解析する処理である。なお、今回対象にしている名詞句の構造は「〈名詞〉」、「〈名詞〉の〈名詞〉」、「〈名詞〉と〈名詞〉」であり、名詞述語文は形態素単位に「/」によって分割されているものとする。次の学習処理は、名詞句間の意味関係の曖昧さを減少させる処理であり、学習ヒューリスティクスを用いて行う。学習ヒューリスティクスは、概念区分と概念間の意味関係を同定するためのヒューリスティクスである。「AはBである」における〈A〉と〈B〉の概念区分と〈A〉と〈B〉の間の意味関係の同定、既存の知識の融合、知識の変換、継承を行う。なお、今回は知識の変換については行っていない。また、A、Bの構造については、次の3つの場合がある。一つは、「N」の場合、〈N〉の概念区分と同定を行う。もう一つは、「N1のN2」の場合、〈N1〉と〈N2〉と〈N1のN2〉の概念区分と同定、〈N1〉と〈N2〉、〈N1〉と〈N1のN2〉、〈N2〉と〈N1のN2〉の間の意味関係の同

定を行う。最後は、〈N1とN2〉の場合、〈N1〉と〈N2〉と〈N1とN2〉の概念区分と同定、〈N1〉と〈N2〉、〈N1〉と〈N1とN2〉、〈N2〉と〈N1とN2〉の間の意味関係の同定を行う。ここで、AとBは名詞句を表し、NとN1とN2は名詞を表す。ただし、A、Bが共に名詞句の構造は対象外とする。学習ヒューリスティクスは、3つに分けることができる。一つは、入力文によるヒューリスティクス、もう一つは、つながる概念との意味関係を利用したヒューリスティクス、最後は、未知概念を既知概念に組み込むヒューリスティクスである。入力文による学習ヒューリスティクスについては表1に示す。なお、番号の抜けている所に、つながる概念との意味関係を利用したヒューリスティクスが入り、適用順序は番号順である。次の概念間の意味関係利用学習処理は、既に存在している知識(冗長なパス)を用い〈名詞句1〉と〈名詞句2〉の直接のパスを制約する処理とその制約された直接のパスを用いた冗長なパスを制約する処理から成り立っている。直接のパスの制約は、冗長なパスから、使用方法に関するヒューリスティクス、概念間の関係を1つにする計算法、状況を考慮した計算法を用い意味関係を求め、この関係と直接のパスに付加されている意味関係との共通項を求めて行う。冗長なパスの制約は、共通項として求めた関係以外の関係を算出するときその関係を削除することにより行う。なお、概念間の関係を1つにする計算法と状況を考慮した計算法を表2、表3に示す。この概念間の意味関係利用学習に代わるものとしての経路情報の生成と適用については7章で述べる。最後の評価学習²⁾は、意味ネットワーク上で一意に決定されたもので、その後情報(文)が入力され誤りであることが判明したものを正しい関係に復元する処理である。なお、今回、行った継承処理は学習処理で行われ、第6章で詳細に述べる。

6 継承処理

今回行った継承処理は、名詞句の構造が「N1のN2」の場合のみを対象とした。これは、〈N1〉で〈N2〉の属性を制約して生成される属性〈N1のN2〉は、〈N1〉のみの属性を表しているため、継承を行っても不備が起

らないものと考えたためである。継承を行う学習ヒューリスティクスは表1の12番である。これを、簡単な例を付加し、図示したものが図2と図3である。図2は継承のヒューリスティクスが適用可能な1つの状況であり、〈人間〉と〈花子〉が集合・要素関係（〈花子〉はただ一つの実体とする）にあり、〈人間〉の属性に〈人間の性格〉があり、その〈人間の性格〉の上位概念に〈性格〉があることを表している。継承のヒューリスティクスは、このような状況下で動作する。その結果が図3である。これは、学習ヒューリスティクス12番のある概念 α に

〈花子〉が当てはまり適用された結果である。つまり、概念 α である〈花子〉と要素・集合関係にある〈人間〉が概念 β であり、その〈人間〉と対象・属性関係にある〈人間の性格〉が概念 γ で「の」の構造をしている。この〈人間の性格〉と下位・上位関係にある〈性格〉が概念 δ である。これから、概念 ε が、〈花子の性格〉になり、空間内に存在しないので生成され、〈花子〉と属性・対象関係、〈人間の性格〉と下位・上位関係になる。

表1 入力文によるヒューリスティクス

番号	内 容
0 1	<A>との関係は、下位・上位関係、要素・集合関係、同一関係、対象・属性関係である。
0 2	<名詞句>の構造が「N1のN2」の場合
2.1	<N1>と<N2>の関係は、対象・属性関係である。
2.2	<N2>と<N1のN2>の関係は、上位・下位関係、集合・要素関係である。
0 3	<名詞句>の構造が「N1とN2」の場合
3.1	<N1>と<N2>の間には、排他性が存在する。
3.2	<N1>と<N1とN2>の関係は、下位・上位関係、要素・集合関係である。
3.3	<N2>と<N1とN2>の関係は、下位・上位関係、要素・集合関係である。
0 4	「N1のN2はN3です」の場合
4.1	<N1のN2>と<N3>の関係は、同一関係である。
0 5	「N1とN2はN3です」の場合
5.1	<N1とN2>と<N3>の関係は、下位・上位関係、同一関係である。
5.2	<N1>と<N3>の関係は、下位・上位関係、要素・集合関係である。
5.3	<N2>と<N3>の関係は、下位・上位関係、要素・集合関係である。
0 6	「N1はN2のN3です」の場合
6.1	<N1>と<N2のN3>の関係は、下位・上位関係、要素・集合関係、同一関係である。
6.2	<N1>と<N3>の関係は、下位・上位関係、要素・集合関係である。
0 7	「N1はN2とN3です」の場合
7.1	<N1>と<N2とN3>の関係は、同一関係である。
7.2	<N1>と<N2>の関係は、上位・下位関係、集合・要素関係である。
7.3	<N1>と<N3>の関係は、上位・下位関係、集合・要素関係である。
1 1	排他性は、<名詞句A>と<名詞句B>の関係が下位・上位関係、あるいは要素・集合関係であり、<名詞句B>と<名詞句C>の関係が上位・下位関係、あるいは集合・要素関係のとき、<名詞句A>と<名詞句C>の間に存在する。ただし、<名詞句A>と<名詞句C>の関係が下位・上位関係、あるいは要素・集合関係、またはその反対の関係なら排他性は存在しない。
1 2	ある概念 α を中心に以下の3つの条件が成立するとき、概念 α と概念 ε ^{*1)} の関係は、対象・属性関係である。また、概念 δ と概念 ε の関係は、上位・下位関係である。概念 ε が存在しないときは、生成する。
* 1 条件	概念 ε とは、概念 α と概念 δ を助詞「の」で関係づけた名詞句概念である。 ①概念 α と下位・上位関係、あるいは要素・集合関係にある概念 β が存在する。 ②その概念 β と対象・属性関係にある概念 γ が存在する。ただし、概念 γ の構造は、「 β の δ 」である。 ③その概念 γ と下位・上位関係にある概念 δ が存在する。

注意) AとBは「AはBです」のAとBである。また、N1, N2, N3は名詞を表している。

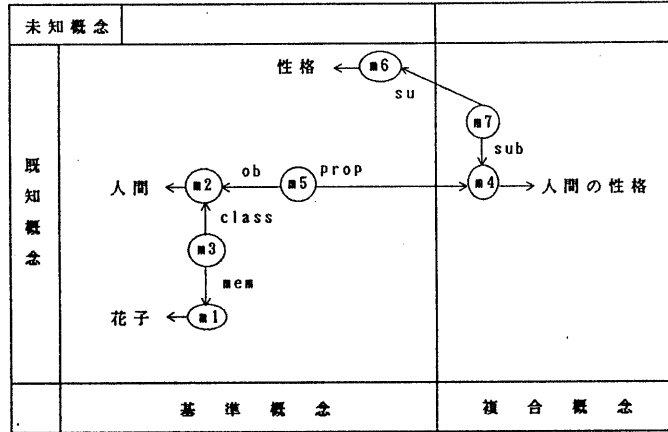


図2 継承が起こる状況

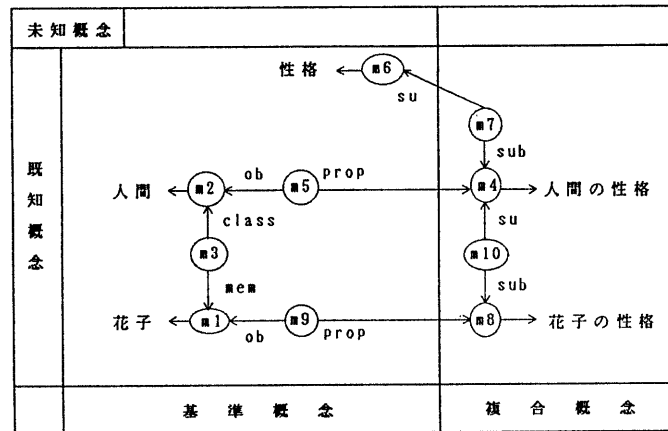


図3 継承が行われた後の状況

7 概念関係経路情報の生成および適用

学習処理，またはメタ知識によって直接のパスに曖昧さが存在しない場合には概念関係経路情報を生成し，直接のパスに曖昧さが存在する場合には直接のパスを一意に決定するために，概念関係経路情報を適用する。

概念関係経路情報とは，以下の2つの条件を満たす冗長なパスを左辺，一意に決定された直接のパスの意味関係を右辺とするルールのことである。

- (1) 2つの概念を結ぶ最短の冗長なパスであること。
- (2) 曖昧さを含んだ意味関係を示す概念を経由しないか，あるいは1ヶ所のみ経由すること。

直接のパスが一意に決定された場合には，上の条件を満たす特定のパスを抽出し経路情報を生成する。また，直接のパスに曖昧さが存在する場合には，特定のパスを抽出し，それと同じ左辺を持つ経路情報が存在する場合，それを適用して直接のパスの意味関係を経路情報の右辺に決定する。

7.1 入力文

以下に示す文は，実験システムにおける入力文の一例である。

1. 象／は／ほ乳類／です。
2. 人間／と／象／は／動物／です。

3. ほ乳類／は／動物／です。
4. 人間／は／動物／です。
5. 人間／と／象／は／ほ乳類／です。
6. 動物／は／生物／です。
7. 人間／は／生物／です。

7. 2 概念関係経路情報の生成例

図7は、文4「人間／は／動物／です。」の学習処理後の世界である。ここで、学習処理において直接のパスの意味関係が一意に決定されたので、先の2つの条件を満たす冗長なパスを抽出し、経路情報を生成する。このSNePSにおいて、条件を満たす冗長なパスは太線で示したものである。これにより生成される経路情報は図4に示したものとなる。ここで、概念1、概念2はそれぞれ概念間の意味関係を示す概念である。

このような処理により経路情報が生成されるが、図4の経路情報が生成された後に図5のような経路情報が生成された場合には、2つの経路情報を比較し、論理積をとるようにして図6に示すような曖昧さを含まない経路情報を生成した。

概念および概念間の意味関係を学習・解析する場合、学習処理のみでは曖昧さが多く残るために、これまでのシステムでは曖昧さを含まない経路情報を生成するためには多くのメタ知識の入力を必要としたが、この手法を用いることにより少ないメタ知識で曖昧さを含まない経路情報を生成することが可能になった。

7. 3 概念関係経路情報の適用例

図8は、文7「人間／は／生物／です。」の学習処理後の世界である。ここで、学習処理において直接のパスの意味関係に<下位・上位関係><同一関係><対象・属性関係>という3つの曖昧さが残ったので、経路情報を生成する場合と同様に、2つの条件を満たす冗長なパスを抽出し、適用する。このSNePSにおいて、条件を満たす冗長なパスは太線で示したものである。これにより文4で生成された経路情報を適用することにより、人間と生物の意味関係を下位・上位関係と決定することができる。

8 おわりに

従来の手法の枠組みの中で、継承処理を実現し、考察した。また、概念関係経路情報では、経路情報の生成と利用について実験し、本手法の有効性について考察した。

今後の課題として、多くの入力文を用いた実験による本システムの評価がある。

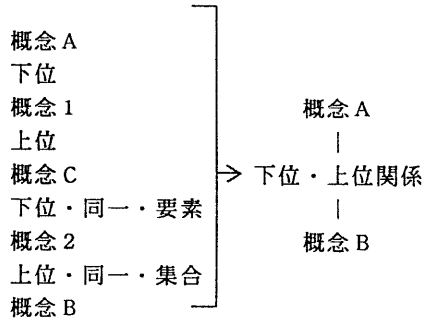


図4 文4により生成された経路情報

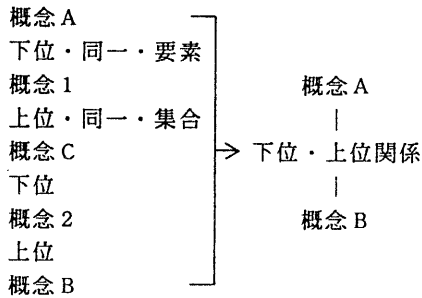


図5

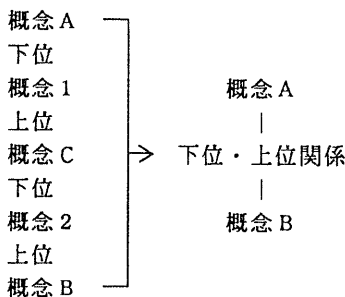
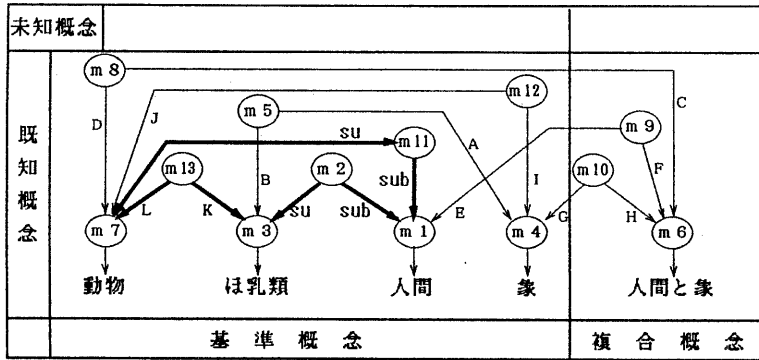
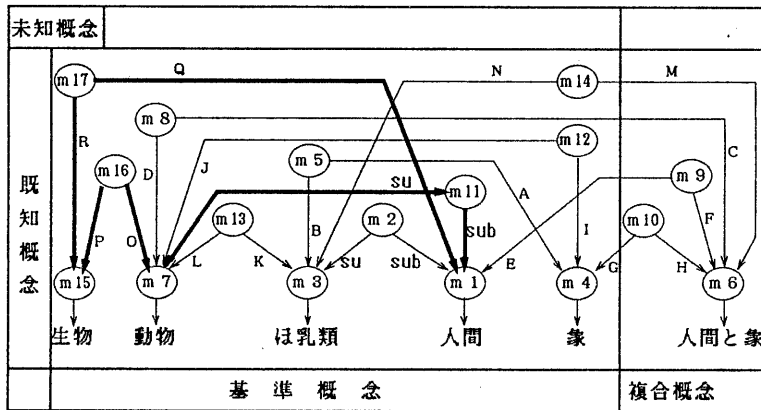


図6 曖昧さを含まない経路情報



[A, B] : <sub--su>, <mem--class>, <eq--eq>, <ob--prop>
[C, D] : <sub--su>, <eq--eq>
[E, F], [G, H], [I, J] : <sub--su>, <mem--class>
[K, L] : <sub--su>, <eq--eq>, <ob--prop>

図7 「人間は動物です」の学習処理後の世界



[C, D], [M, N] : <sub--su>, <eq--eq>
[A, B], [E, F], [G, H], [I, J] : <sub--su>, <mem--class>
[K, L], [O, P], [Q, R] : <sub--su>, <eq--eq>, <ob--prop>

図8 「人間は生物です」の学習処理後の世界

参考文献

- 1) 荒木健治, 桃内佳雄: 名詞述語文における意味概念の学習, 情報研報, Vol. 90, No. 7, pp. 1-8 (1990).
- 2) 荒木健治, 佐々木淳一, 桃内佳雄: 自然言語インタフェースのための未知概念の学習方法, 情報研報, Vol. 91, No. 18, pp. 99-106 (1991).
- 3) 佐々木淳一, 荒木健治, 桃内佳雄: 名詞述語文からの概念学習システム- 比喩の理解と学習-, 情報研報, Vol. 92, No. 33, pp. 17-24.
- 4) 高橋太郎: 名詞述語文における主語と述語

の意味的な関係, 日本語学, 第3巻, 第12号, pp. 18-39 (1984).

- 5) S.C. Shapiro: "The SNePS Semantic Network Processing System," in N.V. Findler (ed.), Associative Networks (New York: Academic Press), pp. 179-203 (1979).
- 6) S.C. Shapiro and W.J. Rapaport: SNePS Considered as A Fully Intentional Propositional Semantic Network, Proc. AAAI86, Vol. 1, pp. 278-283 (1986).