

概念の動的管理システム

白石 誠[†] 納富一宏^{††} 内山明彦[†]

[†]早稲田大学理工学部

^{††}早稲田大学人間科学部

概要

自然言語理解を精密に行なうには、概念の階層構造と相互関係が明確である必要がある。さらに、それらの知識はシステムが自動的に獲得しその過程で随時修正され更新されることが望ましい。そこで、自然言語文の中から概念管理に有用な情報である 1)各名詞の属性、2)概念間の所属関係、3)格構造に関する情報を獲得し、概念の「分化・統合」による動的管理を行なう。

知識表現形式には意味ネットワークを採用した。ノードには1)名詞概念、2)動作概念、3)属性が存在し、リンクには1)所属関係、2)属性関係、3)格構造関係が存在する。今後の展望としては、他の自然言語処理システムへの搭載等が考えられる。

Dynamic Concept Management System

Makoto SHIRAISHI[†] Kazuhiro NOTOMI^{††} Akihiko UCHIYAMA[†]

[†]School of Science and Engineering ^{††}School of Human Sciences

Waseda University

3-4-1 Okubo, Shinjuku-ku, Tokyo, 169, Japan

Abstract

To understand natural language properly, it is necessary that the structures of classes of concepts and their mutual relations be clear. It is hoped that natural language processing systems obtain, modify, and renew these information automatically. Therefore, we have implemented such system. The system finds the attributes of nouns, the relations of the class between two concepts, and case structures in natural language. The system manages concepts dynamically by "specialization and unification". We have used semantic network, in which, the nodes have noun concepts, action concepts, and attributes. The links can represent the relations of classes, attributes, or case structures.

1. はじめに

自然言語理解を精密に行なうためには、その解析のための十分な知識を必要とする。自然言語処理システムを構築するときには、解析のための知識をどのようにして獲得するか、獲得した知識をどのようにして管理するかという問題が生じる。人間が解析に必要な大量の知識を記述する場合、必要な作業にかかる人的・時間的コストは相当なものとなると考えられる。さらに記述のもれや個人差による記述のゆれが生じる恐れがある。知識の管理では、年毎に変化する言葉による新しい単語・用法や、広がりゆくシステムの適用範囲などに対処する必要がある。知識とは常に管理・保守され、更新されなければならない。人間が知識の管理・保守を行なう場合にも、知識の記述の場合と同様に人的・時間的コスト、人為ミスの問題が生じる。

このような問題に対処するためには、自然言語処理システムが入力された自然言語テキストから自動的に知識を獲得し、保守する能力をもつ必要がある。我々は、このようなシステムの構築を目標として処理手法を提案するとともに今回のシステムを試作した。

現在、名詞概念の動的管理システムが実現されている[1]。本稿では、このシステムからの拡張である動作・状態概念の管理について処理手法を提案し、システムでの動作例を示す。

2. 概念の管理

2.1 概念の獲得と管理

概念獲得は品詞同定と概念同定のプロセスからなる。ここでは、概念同定を取り上げる。概念同定とは未知単語がどのような概念に属するかを推論するものである。一般的には、概念同定は知識として与えられている格フレームを用いることによってある程度可能であると言われている[1]。

ここで、意味ネットワークを使用した概念同定を考える。未知語が名詞概念の場合は、意味索性付加を行ない、意味ネットワーク上に名詞概念と意味索性を表現することで概念同定を行なうことができる。未知語が動作・状態概念の場合は、その述語の使用された文より共起パターンを抽出し、意味ネットワーク上に獲得知識を表現することで概念同定を行なうことができる。

次に、概念管理について考える。概念管理の対象

には名詞概念、動作・状態概念が存在する。

名詞概念をターゲットにする場合、意味ネットワークは主として名詞的概念の相互関係、あるいは名詞とそれを規定する属性との関係であり、そのほとんどは2項関係として表現される。

動作・状態概念をターゲットにする場合は、動作・状態に多くの対象概念がかかわりを持つ。格フレームは動作・状態概念を中心とした語の結合関係を示す。例えば、動作概念「勉強する」に関して以下の関連する概念が存在する。

《時間、場所、行為者、対象、道具、目的、方法等》

本稿での概念管理に関する提案点は、動作・状態概念の管理手法が中心となる。この提案点については4章で説明するとともに6章で動作例を報告する。

2.2 動的な概念の獲得・管理システム

動的な概念管理システムの概略図を図1に示す。

概念抽出部は入力される自然言語文を獲得し、情報を辞書に登録する。概念管理部は、既存の概念ネットワークと更新された辞書に基づき、概念ネットワークの更新を行なう。

図に示されるように本システムは入力される自然言語文に依存している。

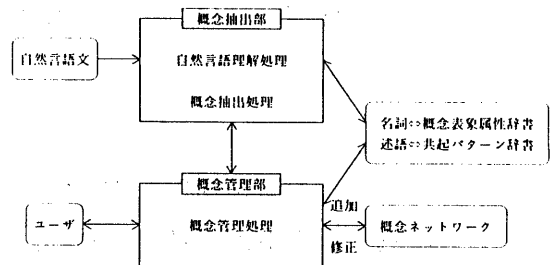


図1 システム概略図

3. 名詞概念の管理

名詞概念の管理は、既の実現されている[1]。動作・状態概念の管理についても基本的に同様の手法をとる。そのため、本章では名詞概念を自然言語文から抽出した情報を利用して、「分化・統合」により動的に管理する手法について説明する。

3.1 概念分化と概念統合

自然言語文をもとにして、名詞概念を獲得し調整するための手がかりとなる情報を考えた場合、概念

および名詞インスタンスに関して、以下のようなものがあげられる。

・従属関係表現

Instance は Concept である。

[例] ツバメは鳥である。

Concept は Concept である。

[例] 人間は動物である。

・動作可能表現

Instance は Act する。

[例] ツバメは飛ぶ。

Concept は Act する。

[例] 鳥は飛ぶ。

概念分化とは、ある概念に属するインスタンス群の中から、特定の属性を持つものを分離し、新たな概念とすることである（図2参照）。

概念統合とは、分化した概念中にインスタンスの大部分が入る場合、および、概念に対して新たなデフォルト属性が付加された場合に、インスタンスを統合するものである（図3参照）。

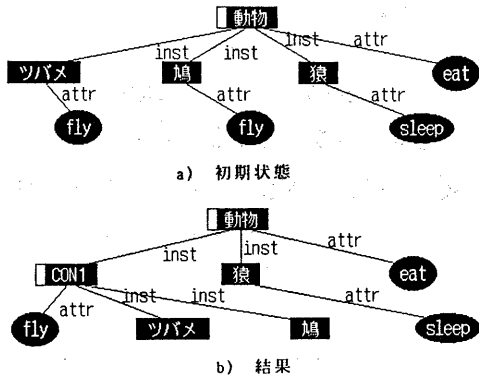
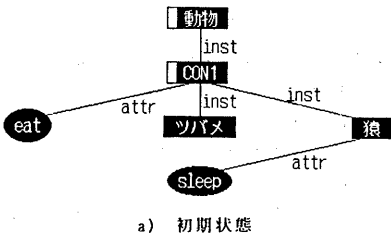


図2 概念分化の例



a) 初期状態

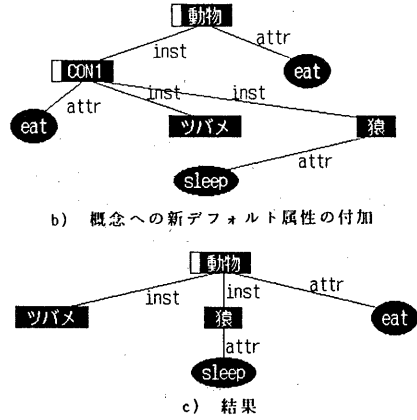


図3 概念統合の例

3.2 名詞概念の管理手法

本手法では、次の処理を行なう。

a) 概念に対する属性付加

はじめに、システムに使用される述語およびその格情報に対応する意味素性を登録しておく。入力テキスト中でシステムの持つ述語-格情報と同じ入力があった場合に、名詞概念に述語-格情報に対応する意味素性を付加する。格情報と意味素性情報との対応を次に示す。

《述語-格-意味素性の関係*の例》

述語	格	意味素性
飛ぶ	Agent	fly

*これを「RACS Set(The Relation of Action, Case and Semantic Primitive)」と呼ぶことにする。

入力例 : 「ツバメは飛ぶ」

解析結果 : [ツバメ : Agent]

属性付加 : [ツバメ ← fly]

b) 共起情報を利用した副概念(sub-concept)生成

入力テキストで用いられる名詞インスタンス群を、述語と格情報から付加された属性情報を基準として、新たなより細かい概念(副概念)へと概念分化させる。副概念は、システムが自動的に生成したものであり、ラベル付けされていない、実際に概念として

使用するためにはラベルの同定処理が必要となる。

c) 副概念のラベル同定処理

獲得した名詞概念は、はじめインスタンスとして扱われる。新たなインスタンスが、その下層概念中に属する時に、インスタンスの概念化を行なう。この時に属性を参照し既に存在する副概念とのマッチングを行なう。マッチングできた場合、副概念にラベル付けを行なったことになる。

d) インスタンスの所属管理

ラベル付けされた下層の概念への所属が明らかとなった場合には所属の変更を行なう。

e) 概念統合処理

概念に新たなデフォルト属性が付加された場合および分化された概念中にインスタンスの大部分が入る場合には、分化した属性値に属さないもののみを分離し、他のインスタンスを統合する。

f) 矛盾整合処理

獲得したインスタンスの属性が上位概念のデフォルト属性と矛盾した場合には、矛盾している属性を分化の基準として、矛盾インスタンスを分化させる。

また、獲得した概念の属性が、その概念に所属するインスタンスの一部と矛盾した場合についても同様に分化させる。

知識利用時には、下層概念の属性を優先することで矛盾知識によるシステムの破綻を避ける。

4. 動作・状態概念

本章では、今回新たに提案する動作・状態概念の管理手法について述べる。

4.1 動作・状態概念の管理の考え方

自然言語文を基にして、動作・状態概念を獲得し管理するための情報を考える。情報には、以下のようなものあげられる。

《動作・状態概念管理のための情報例》

[ツバメは飛ぶ]
[鯛は泳ぐ]

これらの例に対して次の情報が存在すれば、動作・状態概念に関してRACS Setを作成することができる。

《概念の上下関係》

[ツバメ→鳥]
[鯛 →魚]

《RACS Set》

飛ぶ [agt, 鳥]
泳ぐ [agt, 魚]

このようにして次のRACS Setが存在すると仮定する。

《RACS Setの例》

愛する [agt, 人, obj, 人]
諦める [agt, 人, obj, 物事]
あきらめる [agt, 人, obj, 物事]
集める [agt, 人, obj, 具象物, spa, 場所]
開く [agt, 具象物]

この時、格-意味素性ペアを属性とみなすことにより、次のような意味ネットワークが構成される。

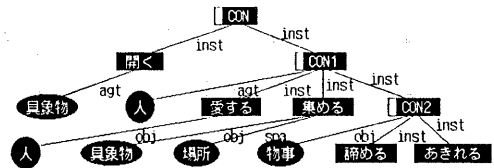


図4 格-意味素性ペアを属性とする意味ネットワーク

このような手法をとれば動作・状態概念を分類することができると思われる。

4.2 管理上の問題点

前項の考え方は動作・状態概念の獲得より前に名詞概念の概念ネットワークが構成されている時のみ有効な手法である。なぜなら、RACS Set中の格については入力文より獲得することができるが、RACS Set中の意味素性については入力文から獲得することができず、各名詞について名詞概念の概念ネットワークから意味素性を獲得せざるを得ないからである。

次に名詞概念ネットワークの獲得手法について考える。この手法は、あらかじめRACS Setをシステムに登録しておき、入力される名詞に意味素性を属性として付加するものである[1]ことは既に述べた。つまり、名詞概念ネットワークを構成するためには、あらかじめRACS Setをシステムが獲得している必要がある。

これらのことは、次のことを意味する。すなわち、名詞概念ネットワークの獲得は動作・状態概念ネットワークの獲得と互いに依存関係にあることである。

4.3 名詞概念への属性付加の変更

前項の問題点に対処するために、次の手法を新しく提案する。動作・状態概念1つにつき属性を1つ作り、入力文中の行為者格(Agent)の名詞にその属性を与えることにする。動作・状態概念の格共起パターンの他の格に現われる名詞については、その格が所持する属性情報を入力文中に見つけることができないと考えられるため、属性を付加しない。例を次に示す。

《名詞概念への属性付加の例(変更後)》

《入力文》

[人がお金を銀行に預ける]

《格文法解析結果》

「預ける」

[agt, 人]

[obj, お金]

[sto, 銀行]

《属性の付加》

人 ← f(預ける) 《属性》

この例では、動詞「預ける」の行為者格(Agent)は「人」であるので、「預ける」の属性を「f(預ける)」と表現し、「人」にf(預ける)を付加する。

4.4 動作・状態概念への属性付加

動作・状態概念の獲得では、前項の属性付加手法により獲得された名詞概念ネットワークを使用し、RACS Set中の意味素性には入力文中の名詞インスタンスの上位概念を登録する。例を次に示す。

《動作・状態概念への属性付加の例》

《入力文》

[白石がA円をB銀行に預ける]

《格文法解析結果》

「預ける」

[agt, 白石]

[obj, A円]

[sto, B銀行]

《概念の上下関係》

白石 → 人間

A円 → 金額

B銀行 → 銀行

《獲得した意味ネットワーク》



図5 獲得した意味ネットワーク

4.5 動作・状態概念の獲得・管理

動作・状態概念の共起パターンを解析し、意味素性と共に格により意味ネットワークを構成し、動作・状態概念の意味表現とする。アルゴリズムを図6に示す。

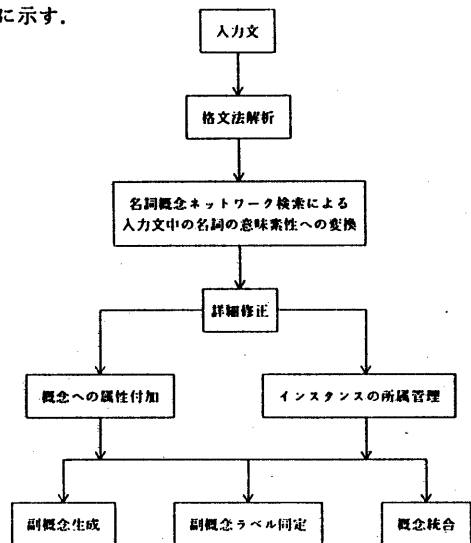


図6 アルゴリズム

4.5.1 動作・状態概念の意味表現

a) ノード(node)とリンク(link)

意味ネットワーク上に表現されるノードとリンクの種類を表1に示す。

表1 ノード(node)とリンク(link)

a) ノード(node)	
ノードは集合を表わし、集合には集合名としてラベルが付けられている。ラベルの付された集合は概念を表わす。集合中の要素は、概念、もしくはインスタンス、あるいは動作・状態である。	
1	インスタンス集合(instance set) インスタンスを要素とする集合。概念として扱われる。
2	動作・状態集合(action/state set) 動作もしくは状態を要素として持つ集合。
3	属性集合(feature/attribute set) 特徴もしくは属性を要素として持つ集合。スキーマ(schema)もしくはフレーム(frame)として扱われる。
b) リンク(link)	
ノードとノードの関係をアーク(arc)で結んだものをリンク(link)と呼ぶ。リンクには関係ラベル(relational label)が付されている。	
1	関係ラベルの分類
①	概念の上位-下位関係 概念と概念の従属関係を示す。
②	特徴・属性の所有関係 インスタンス、または概念の所有する特徴・属性を表わす。
③	格構造関係 動作・状態と共起する名詞概念の格構造関係を示す。

b) 意味素性

本手法では、あらかじめシステムに意味素性を登録しない。名詞の上位概念を意味素性として用いて、動作・状態と共起する名詞を意味素性に変換する。

c) 格

システムで使用する格について、表2に示す。

表2 格の種類

格	格ラベル	格	格ラベル
動作格	act	場所経過点格	sth
行為者格	agl	時間格	dur
「が」格	ga	条件格	con
対象格	obj	時点格	tim
変化の起点格	sou	時の起点格	tfr
変化の終点格	ooa	時の終点格	tto
場所の起点格	sfr	原因理由格	au
場所の終点格	sto	結果格	res
場所格	spa	譲歩格	coc
道具格	too		

4.5.2 入力文中の名詞の意味素性への変換

入力文中の名詞と動作・状態の関係をそのまま意味ネットワークに登録すると、名詞のインスタンスは膨大に存在するために、リンクの組み合わせ爆発が起こり管理が困難となると考えられる。そこで、入力文中の名詞の汎化処理を考える。本システムでは、入力文中の名詞をインスタンスとして捉え、そのインスタンスを上位概念で置き換えることにする。具体的には、入力インスタンスを既存ネットワーク上のより上位のノードに付されたラベルで代用する。

4.5.3 詳細修正

この処理は、随意格集合と必須格集合の区別をつけるものである。動作・状態概念の共起パターンは極めて自由度が高く、これをすぐに獲得しては、組み合わせ爆発を起こす。故に、人間との対話形式で、学習を進めて行く。この補助情報として、共起した格とその意味素性、および表層パターンを記録しておき、頻度計算を行なう。この頻度がユーザの設定した閾値を越えた時点で、対話学習が開始される。頻度が閾値を越えた格を必須格、越えない格を随意格として学習する。意味ネットワークには、必須格集合と随意格集合とをフレームとしてリンクさせる。

4.5.4 動作・状態概念と属性の関係

動作・状態概念の属性は、動作・状態のRACS Setとする。RACS Set中の格-意味素性ペアを1属性とする。

表3 各情報の詳細

1	<p>概念の上位・下位関係</p> <p>ある自立名詞をラベルインスタンスとして検索することで、これに対する階層情報として、それぞれ上位概念、下位概念集合、同列概念集合を取り出すことができる。</p>
2	<p>基底素[3](primitive)</p> <p>特徴(feature)、属性(attribute)のうち、最初から定義されているものを基底素と呼ぶ。あるいは、自動獲得後に、人間によって細部修正がなされているものを指す。基底素は、完全にシステムが把握可能な情報であり、かつ個々の概念を完全に規定するためのプリミティブである。基底素情報は無矛盾性と信頼性が保証されている。</p>
3	<p>スキーマ(schema)</p> <p>個々の概念を規定するための基底素集合をスキーマと呼ぶ。スキーマ情報のインヘリタンス(inheritance)を概念階層において行なうことができる。即ち、ある概念の上位概念の有するスキーマは、下位概念においても成立する。</p>
4	<p>フレーム(frame)</p> <p>スキーマと同義に使われるが、ここでは、概念の基底素インスタンス集合をフレームと呼び区別する。フレームにおけるスロットが基底素であり、そのフィルラーが基底素インスタンスである。一対のスロットとフィルラーにより基底素の表現する特徴・属性の内容にアクセスすることができる。属性に付加された重みへのアクセスを可能にする。</p>
5	<p>格フレーム(case frame)</p> <p>ある述語に共起する格の集合を格フレームと呼ぶ。ここでは、格と格の意味素性とが対になっているため、動作・状態概念辞書としても利用することができる。</p>

4.5.5 動作・状態インスタンスの所属管理
 動作・状態インスタンスについて、格-意味素性の属性を既存のネットワークに存在する属性と共起させ、一致する属性の最も多い動作・状態概念の要素とする。

4.5.6 共起情報を利用した副概念の生成
 動作・状態概念についてある属性数が閾値を越える場合、その概念に注目した副概念を作成し、その属性を持つ物すべてをその副概念の要素とする。

4.5.7 副概念のラベル同定処理
 システムの生成した副概念(ラベルなし)とラベルが付された動作・状態概念について属性のマッチングを取る。マッチングができた場合、ラベル付けを行なったこととなる。

4.5.8 概念統合処理
 名詞概念での処理と同様、概念に対して新たな属性が付加された場合および分化された概念中にインスタンスの大部分が入り、概念のデフォルト属性と矛盾しない場合、分化した属性集合に属さないものを分離した副概念として、それ以外のインスタンスを統合する。

5. 概念運用

本章では、獲得・管理される概念の運用について述べる。

5.1 概念辞書

意味ネットワークの構成をそのまま利用することで、汎用概念辞書としての応用が考えられる。すなわち、名詞概念に関して、そのインスタンスをキーにして検索することにより、a)概念の上位・下位関係、b)概念を規定する全ての基底素[3]、c)基底素の集合であるスキーマ、d)概念の基底素インスタンス集合であるフレームの各情報にアクセスすることが可能である。

また、動作・状態概念のネットワークからは、格フレーム情報が利用できる。

表3に利用できる情報の詳細を述べる。

6. 動作例

初期状態が図7の時の動作例を図8に示す。ここでは、副概念の生成が行なわれている。

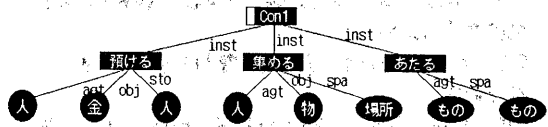


図7 初期状態

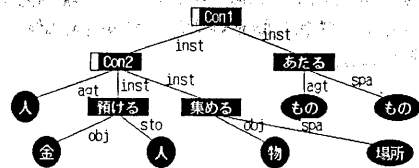


図8 副概念生成

7. おわりに

本稿では、動的な概念管理システムの動作・状態概念への拡張を行なった。本手法の問題点としては、1)入力文に依存したシステムであるため、入力文の品質を管理する必要がある、2)ラベル同定の際に属性が完全に一致しない場合の処理を決定することが困難である、3)分化した結果の意味ネットワークが最適でない場合が存在する、4)入力文が否定文であった時の処理が困難である、等が考えられる。

現在、これらの問題点について改善手法を検討している。

今後の展望としては、他の自然言語システムへの登載等が考えられる。

8. 参考文献

- [1] 木子：「自然言語処理における基本文型の学習に関する研究 ～自然言語文からの動的な知識獲得手法～」,平成元年度修士論文,早稲田大学(1990).
- [2] 木子, 納富, 内山：「自然言語解析における概念分化の手法」,早大情科センター紀要, Vol. 11, Spring, pp. 24-27(1990).
- [3] 納富, 木子, 内山：「日本語文からの概念獲得とイベント系列の学習」,早大情科センター紀要, Vol. 11, Spring, pp. 28-31(1990).
- [4] 木子, 納富, 内山：「共起パターンと名詞意味組成のテキストからの抽出」,情処第39全大7G-6, pp. 752-753(1989).
- [5] 岡田：「語の概念の表現と蓄積」,コロナ社(1991).
- [6] 水谷, 石綿, 荻野, 他：「文法と意味1, 2」, 朝倉書店(1988).
- [7] 大須賀, 佐伯：「知識の獲得と学習」,オーム社(1987).
- [8] 上野, 石塚：「知識の表現と利用」,オーム社(1987).
- [9] R. S. Michalski他編, 電総研人工知能研究グループ他訳：「概念と規則の学習 例からの学習」,共立出版(1988).