

3H-6

視点を考慮した概念の同一化

徳永 健伸 田中 穂積
東京工業大学

1 はじめに

計算機に自然言語を理解させるためには、人間が持つような常識的知識を計算機上に実現することが必要である。常識的知識の中でも特に概念を体系的に整理した概念体系の重要性はすでに多くの研究者によって指摘されている[5,1,6]。

概念体系を自然言語処理に利用する場合、処理のどこかの段階で、表層の文に現れる語と概念体系中の概念との対応付けをおこなわなければならないが、語と概念の対応関係は十分な情報がないと一意には決まらない。処理の途中で、語が複数の概念と対応付けられる可能性がある場合は、すべての可能性を表現できなければならない。本稿では特に視点の違いによる語のあいまい性を取り上げ、視点を表現するための道具として概念のパス式を提案する。

2 概念階層

本稿では、上位/下位関係を唯一の概念間の関係とする概念階層について議論する。準備としていくつかの定義をおこなう。

定義 1 (概念) 概念は性質の集合である。ただし、性質は属性と属性値の対である。

定義 2 (複合概念) 概念あるいは、複合概念を“,”(カンマ)で区切って並べ、“{”と“}”で括弧つめたものを複合概念という。複合概念は、要素の持つ性質を兼ね備えた概念を表わす。複合概念の要素として複合概念を持つ場合、内側の複合概念の要素を外側の複合概念の要素としてよい。

定義 3 (概念階層) 概念階層は概念の集合 C と概念間を結ぶ有向リンクの集合 L からなるループを含まない有向グラフ $\langle C, L \rangle$ である。有向リンクは、概念 a のインスタンスが概念 b の性質を継承できるときに概念 a から概念 b に向かって張られる。この時、概念 b を概念 a の上位概念、概念 a を概念 b の下位概念という。

定義 4 (直接リンク, 間接リンク) 概念 a と b を直接結ぶリンクがあるとき、これを直接リンクと呼び、 $a < b$ と表わす。また、概念 a と b が 0 個以上のリンクを介して結ばれているとき、このリンクの系列を間接リンクと呼び、 $a \ll b$ と表わす。いずれの場合も、 b が a の上位概念である[2]。

定義 5 (排他的概念) 複数の概念が共通の属性を持ち、その属性の値が異なる場合には、これらの概念の間には排他性があるといい、 $x \otimes y$ と表わす。

3 語のあいまい性と視点

自然言語の意味処理を、文章が伝えようとしている概念を決定することであると考えると、語のあいまい性は、語が指し示す

概念が複数あることに起因している。本稿では、語のあいまい性の中でも視点の違いによるあいまい性を取り上げる。

これまでの概念体系に関する研究の中にも視点を取り扱ったものがいくつかあるが[4,7,3,6]、これらの視点は、ある概念をその概念の持つある性質をもとに下位分類する時に、どのような性質を分類の基準とするか、を指していた。本稿ではこれを「下位方向の視点」(downward view)と呼ぶ。

これに対して本稿では、ある概念が複数の上位概念を多重継承している場合に、どの上位概念によってその概念を特徴付けるか、という意味での視点を考える。これを特に「上位方向の視点」(upward view)と呼んで下位方向の視点と区別する。以下、単に視点といえば上位方向の視点を指すものとする。

本稿で考える視点の違いによるあいまい性とは、ひとつのインスタンスを異なった視点から眺めることによって、異なる概念のインスタンスに見える、というあいまい性である。このようなあいまい性は、ある概念が互いに性質の異なる上位概念を多重継承しているような場合に起こる。

1. この野菜は害虫に強い。(植物としての野菜)
2. 野菜をたくさん食べなさい。(食物としての野菜)

たとえば、例文 1 の“野菜”は*(植物としての野菜)¹を指しており、例文 2 の“野菜”は*(食物としての野菜)を指している。ひとつの方法として*(野菜)を「植物としての野菜」という概念(野菜 1)と「食物としての野菜」という概念(野菜 2)に分けることが考えられる。しかしながら、概念を分割するとその下位概念も分割しなければならなくなり、概念の数が下位概念を分割するたびに指数関数的に増加するという問題が生じる。また、概念を分割すると、*(野菜 1)、*(野菜 2)が共通に持っている性質を両方の概念に記述しておく必要があり、不経済である。

このような問題は、概念を考える時には様々な視点がある、ということに起因している。たとえば、*(野菜)も植物としてみた野菜と食物としてみた野菜を考えることができ、文脈によってその視点が変わってくる。動的な視点の変化を静的な概念階層で表現しようとする問題が生じる。したがって、概念階層中では様々な視点を含む概念を静的な形で表現し、概念階層を解釈する時に視点を含めた表現ができるような枠組を与えることが必要である。概念が様々な視点を持つことは、概念階層中では概念が複数の上位概念を多重継承することに対応する。また、特定の視点からみた概念は、上位概念から注目している概念へ至るパスを含めて概念を表現することに対応する。

4 概念のパス式

この節では、概念階層において視点を表現するための道具としてパス式を提案する。パス式は、概念を上位概念からの継承経路を含む形で表現する方法であり、注目している概念が多重継承をしているときに意味を持つ。パス式を表現するために“\”(バックslash)を使う。たとえば、図 1 において、*(野菜)はすべての上位概念の性質を継承するが、パス式で表現された概念*(野菜\植物)は、上位概念*(植物)、*(生物)、*(具体物)の性質しか継承しない概念を意味する。すなわち、植物という視点からみた野菜という概念を表現していることになる。

¹以下、語を表わす時は 2 重引用符で囲み、概念を表わす時は*を前に付ける。

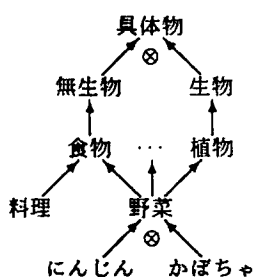


図 1: 概念階層の例

パスの途中には複合概念を含んでもよい。その場合、複合概念を展開できる。たとえば、 $*(野菜 \setminus \{食物, 植物\})$ は $*(食物)$ と $*(植物)$ の性質しか継承しない概念 $*(野菜)$ を表わす。また、 $*(野菜 \setminus \{食物, 野菜\})$ と $*(\{野菜 \setminus 食物, 野菜 \setminus 植物\})$ は等価である。パス式で表現された概念の持つ性質を以下のように定義する。

定義 6 (パス式で表現された概念の持つ性質)

$$Prop(x_1 \setminus x_2 \setminus \dots \setminus x_n) = \bigcup_{i=1}^{n-1} SelfProp(x_i) \cup Prop(x_n)$$

ただし、

$Prop(x)$: 概念 x がもつすべての性質

$SelfProp(x)$: 上位概念からの継承なしに、 x がもつ性質

$InheritProp(x)$: 概念 x が継承によってのみもつ性質

上位概念を多重継承をしない概念については、パス式で表現しても意味は変わらない。すなわち、

定義 7 (パス式の縮退) パス式 $x_1 \setminus x_2 \setminus \dots \setminus x_n$ ただし、 $x_i (i = 1 \dots n)$ において、 $x_i (i = 1 \dots n)$ が多重継承をしていなければ、パス式 $x_1 \setminus x_2 \setminus \dots \setminus x_n$ と x_1 は等価である。

5 パス式で表現された概念の同一化

パス式を含む概念同志の同一化を以下のように定義する。前提として、パス中に複合概念がある場合は、複合概念に展開できるので、以下の定義ではパス中には複合概念がないものとする。また、複合概念同志の同一化については、複合概念の各要素同志を同一化して得られた結果によって構成される複合概念を同一化の結果とする。

定義 8 (同一化) 概念 $x = x_1 \setminus x_2 \setminus \dots \setminus x_m$ と概念 $y = y_1 \setminus y_2 \setminus \dots \setminus y_n (m, n > 0)$ の同一化を次のように定義する。

1. $x_m \ll y_1 \Rightarrow x_1 \setminus \dots \setminus x_n \setminus path(x_m \ll y_1) \setminus y_1 \setminus \dots \setminus y_n$
2. $\exists k(x_{m-k+1} \setminus \dots \setminus x_m = y_1 \setminus \dots \setminus y_k) \Rightarrow x_1 \setminus \dots \setminus x_m \setminus y_{k+1} \setminus \dots \setminus y_n$
3. $\neg(x \otimes y) \Rightarrow \{x, y\}$
4. $x \otimes y \wedge \Rightarrow FAIL$

ただし、 $path(x \ll y)$ は両端に x, y を含まない x から y へのパスを表わす。

図 1 について同一化の例を説明する。 \sim は同一化を表わす。

- $*(にんじん) \sim *(野菜) \Rightarrow *(にんじん \setminus 野菜)$
通常の同一化と同じで、概念は下位概念の方向に特定化される[6]。この例では、 $*(にんじん)$ が上位概念を多重継承していないので、定義 7 より、同一化の結果、 $*(にんじん \setminus 野菜)$ は $*(にんじん)$ と等価である。
- $*(野菜) \sim *(植物) \Rightarrow *(野菜 \setminus 植物)$
(野菜)が上位概念を多重継承しているため、同一化の結果上位概念の方向に特定化される。この場合の同一化の結果のパス式は意味を持つ。

- $*(植物) \sim *(食物) \Rightarrow FAIL$

同一化される概念間に上位/下位関係がなく、両概念が排他的なので同一化は失敗する。

- $*(野菜 \setminus 食物) \sim *(\{食物, 植物\}) \Rightarrow *(野菜 \setminus 食物)$
($\{食物, 植物\}$)は互いに排他的な関係にある概念を要素として持っているが、 $*(野菜 \setminus 食物)$ と同一化することにより、概念が $*(野菜 \setminus 食物)$ に特定化され、あいまい性が解消される。

6 おわりに

本稿では、概念階層における概念に関して、これまでの下位方向の視点とは異なる上位方向の視点の存在を明らかにし、その表現方法として概念のパス式を提案した。また、パス式で表現された概念同志の同一化も定義した。我々の同一化の定義では、同一化される概念が互いに排他的な場合、同一化は失敗するが、必ずしもこれは適切でない。

3. 太郎は庭で野菜を作っている。(植物としての野菜)
4. 太郎は毎日野菜を食べている。(食物としての野菜)

たとえば、例文 3 の“野菜”は“作っている”との共起関係から、 $*(野菜 \setminus 植物)$ のインスタンスに対応し、例文 4 の“野菜”は“食べる”との共起関係から、 $*(野菜 \setminus 食物)$ のインスタンスに対応すると考えるのが自然である。しかしながら、適当な文脈があれば、両例文の“野菜”は同一のインスタンスに対応してもおかしくはない。我々の同一化の定義ではこれらのインスタンスの同一化は失敗する。しかしながら、“作って”、“食べる”間には、同一のインスタンスといえどもその野菜には、なんらかの変化が起こっているはずである。我々は、同一化の失敗をこのような変化を検出するきっかけとして扱うべきで、これを同一化のレベルで扱うべきではないと考えている。今後、このようなきっかけを受けて、インスタンスの変化を扱うメカニズムを考える必要がある。

参考文献

- [1] 荻野綱男. シソーラスの作成の問題点. 言語, 6(5):64-71, 1987.
- [2] 堺和宏, 徳永健伸, 田中穂積. シソーラス作成支援ツールに関する基礎的考察. 情報処理学会第 35 回全国大会予稿集, 1801-1802 ページ, 1987.
- [3] 石崎俊, 井佐原均, 橋田浩一, 内田ユリ子, 横山晶一. 文脈理解のための概念記述法. 情報処理学会自然言語処理研究会, 53-59 ページ, November 1987.
- [4] 鶴丸弘昭, 藤田毅, 首藤公昭, 吉田将. 日本語の機械処理. 電子通信学会オートマトンと言語研究会, 41-50 ページ, 1976.
- [5] 田中穂積, 仁科喜久子. 上位下位関係シソーラス ISAMAP1 の作成. 情報処理学会自然言語処理研究会, 25-44 ページ, November 1987.
- [6] 木下聡, 佐野洋, 浮田一男, 天野真家. 文脈理解のための知識の表現と推論. Logic Programming Conference '88, 205-212 ページ, ICOT, 1988.
- [7] D. G. Bobrow and T. Winograd. An overview of KRL. Cognitive Science, 1:3-46, 1977.