

自然言語概念の意味構造記述モデルと 会話文データベースへの応用

野崎 剛一† 脇山 正博†† 河口 英二††

† 長崎大学総合情報処理センター

†† 九州工業大学工学部電気工学科

本論文では、自然言語における個々の概念や陳述、知識等の意味構造を記述する新しい方法（SD式）を提案し、その記述法に基づいた知識システム・モデルにおける自然言語概念の認識、理解、学習等を考察する。まず、SD式を基に2つの概念間に構文的な詳述関係と知識データによる詳述関係を定義し、概念間に意味差の尺度を導入する。そして、詳述関係を用いたシステム知識の体系化や、入力概念の認識、理解、学習等を定式化し、意味を定量的に扱う枠組みを示す。最後に、このSD式モデルの英会話文データベースへの応用事例の概要について述べる。

A Semantic Description Model of Natural Language Concept

and its Application to a Database of Conversational English/Japanese Texts

Koichi Nozaki†, Masahiro Wakiyama††, Eiji Kawaguchi††

† Nagasaki University, †† Kyushu Institute of Technology

We present a new concept-data description method (SD-Form) to describe structural patterns of concepts in the world of natural language. The SD-Forms work as an interlingua of natural languages. It can describe most of the natural language statements and emotional utterances as well as the world knowledge in a system. In order to establish a semantic difference measure between two conceptual patterns, we introduce a syntactic and a knowledge-based elaboration relations using the SD-Forms. Concept-pattern recognition and understanding problems of a knowledge system are discussed under the proposed measure. In the final part of the paper we show its application to an English/Japanese conversational text database.

1. まえがき

知識や意味記述に関しては1970年以降、フレーム⁽¹⁾、スクリプト、意味ネットワーク⁽²⁾等の多くの研究が行われてきている。^{(3) - (5)} 本論文では、自然言語における概念パターンや陳述、知識等の意味構造を記述する新しい方法を示し、これを基にして自然言語概念を処理対象とする知識ベース・システムのモデルを提案する。そして、このモデルの英会話文データベースへの応用について述べる。

本論で提案するSD式 (Semantic-Structure Description Form) は自然言語における個々の概念、陳述表現、感情表現、あるいはシステムに与える知識データ等を記述するための中間言語であり、その構文は曖昧さの無い一つの文脈自由文法で規定する。そして、自然言語における認識、理解および学習の動作を定式化するために、SD式で表わされた概念間の二項関係として "詳述関係 (elaboration relation)" を定義し、これに基づく概念パターン間の意味差の尺度を導入する。

意味記述の方法としては従来から種々の提案がなされている^{(1) - (3)} にもかかわらず、本論文でSD式という新たな意味記述法を用いる理由は以下の通りである。

- (1) 意味記述のデータは構文が明確に定義されているべきである。SD式は曖昧さのない文脈自由文法で定義している。
- (2) 意味記述データの構文的な差異がそれで表わされる概念の意味差を定量的に表わすべきである。
- (3) 意味記述データは自然言語 (特に日本語や英語) との相互変換が容易で、かつ覚え易いものでなければならぬ。SD式では英語の基本的な文型に模した主語、述語、目的語や順接、逆接、因果関係、包含関係等の自然言語表現を意識した構文を定めている。

2. SD式の定義と概念の意味構造の記述⁽⁶⁾

本論文では、"自然言語における概念" は構造的に曖昧さのない記号列として記述可能であると仮定し、個々の概念、陳述表現、感情表現、あるいはシステムに与える知識データ等をSD式で記述する。SD式とは、形式的には以下に示す文脈自由文法SDGで生成される任意の記号列のことである。

$$SDG = \{ \Sigma_n, \Sigma, P, \Phi \}$$

各記号の役割は以下の通りである。

Φ : 開始記号

P : 生成規則

Σ_n = { Φ, X, X_n, X_c, X_s, X_o } : 非終端記号 X, X_n, …, X_o 等は変数とも言う。

Σ_o = { f₁, f₂, …, f_i, f_j, …, g₁, g₂, …, g_i, SOME, /, nega, pass, assu, mood, fcus, para, incl, excl, equa, defl, andx, orxx, eorx, caus, indu, bcau, pseq, nseq, inst, addi, plus, than, ofal, exmp, like, such, csol, ptof, prof, thru, s, v, c, i, o, b, a, r, e, [,], (,), ", " } : 終端記号

概念の意味構造記述における各終端記号の役割は以下の通りである。

f₁, f₂, …, f_i, f_j, … : 数値を含む一般概念ラベル
 g₁, g₂, …, g_i, … : 非数値の一般概念ラベル
 / : 修飾子, 左辺を右辺で修飾
 nega, pass, assu, mood, fcus : 規定子, 一般に P_i と表わす
 para, incl, excl, …, thru : 結合子, 一般に C_i と表わす
 s, v, i, o, c, b : 陳述式の機能項目記号, F_s
 a, r, e : 感情式の機能項目記号, F_e
 [,], (,) : 句切り記号としての括弧
 ", " : 句切り記号としてのコンマ

生成規則 P の詳細は以下の通りである。単純ラベル g_i に別の単純ラベル f_j を組み合わせたものをパラメータ付きラベルと言い g_i(f_j) と表す。各規則の右に括弧で示した数値については後で言及する (4.1 節 参照)。

- 0. Φ → X (1.0)
- 1. X → X_n, X_c, X_s, X_o (0)
- 2. X → f_i (2.0)
- 3. X → g_i(f_j) (5.0)
- 4. X → g_i/X (4.0)
- 5. X → g_i(f_j)/X (7.0)
- 6. X → (f₁)plus(f₂)…plus(f_n)/X (4n)
- 7. X → g_i/(X)para(X)…para(X) (2n+2)
- 8. X → g_i(f_j)/(X)para(X)…para(X) (2n+5)
- 9. X → (f₁)plus(f₂)…plus(f_n)/
(X)para(X)…para(X) (4m+2n-2)
— n X's (2 ≤ n) —
- 10. X_n → P_i(X) (2.0)
- 11. X_c → (X)C_i(X) (2.0)
ただし、この C_i は plus 以外の結合子
- 12. X_c → (X)plus(X)…plus(X) (2n-2)
— n X's (2 ≤ n) —
- 13. X_c → (X)C_i([(X)…(X)]) (n+1)
— n X's (2 ≤ n) —
ただし、この C_i は incl, excl, ofal, exmp, like, csol, prof の何れかである。
- 14. X_n → [s(X), v(X)] (4.0)
- 15. X_n → [s(X), v(X), o(X)] (6.0)
- 16. X_n → [s(X), v(pass(X)), b(X)] (8.0)
- 17. X_n → [s(X), v(X), c(X)] (6.0)
- 18. X_n → [s(X), v(X), i(X), o(X)] (8.0)
- 19. X_n → [s(X), v(pass(X)), i(X), b(X)] (10.0)
- 20. X_n → [s(X), v(pass(X)), o(X), b(X)] (10.0)

- | | | |
|---|--------------|------------------------|
| 21. $X_n \rightarrow [s(X), v(X), o(X), c(X)]$ (8.0) | (D1)excl(D2) | (D)excl([D1, ..., Dn]) |
| 22. $X_n \rightarrow [s(X), v(pass(X)), c(X), b(X)]$ (10.0) | (D1)equa(D2) | (D)csof([D1, ..., Dn]) |
| 23. $X_n \rightarrow [s(X), v(pass(X)), b(X), c(X)]$ (10.0) | (D1)defi(D2) | (D)prof([D1, ..., Dn]) |
| 24. $X_n \rightarrow [a(X), [r(X)], [e(X)]$ (2.0) | (D1)orxx(D2) | (D1)eorx(D2) |
| | (D1)ptof(D2) | (D1)kdof(D2) |

記号列 D が S D 式であること、すなわち、SDG で生成可能であることを $\Phi \Rightarrow D$ と表記する。生成規則の形から S D 式の形は以下の形式 (f1-f8) の何れかであることが分かる。

- f1. X : S D 式変数
- f2. f_i : 単純概念ラベル
- f3. $g_i(f_i)$: パラメータ付き概念ラベル
- f4. D_1/D_2 : この形式を修飾子形式と呼ぶ。
- f5. $P_i(D)$: この形式を規定子形式と呼ぶ。
- f6. $(D_1)C_i(D_2)$: この形式を結合子形式と呼ぶ。
- f7. $[Fs_i(D_i)]$: 陳述 S D 式 (陳述式項目のリスト)
- f8. $[Fe_i(D_i)]$: 感情 S D 式 (感情式項目のリスト)

なお、S D 式による様々な概念の記述例は文献[6]に示している。

3. 知識記述のための S D 式

本論では、知識データを内蔵し自然言語概念を処理するシステムを“知識ベース・システム”と呼ぶ。ここで、知識ベース・システムは対象世界を記述する世界データ (World Data) と対象世界で有効となる知識データ (Knowledge Data) を持つものとする。これらは何れも対象世界の実事データ (Fact Data) である。図 1 に本モデルにおける事実データの階層を示す。知識データの中の“事実知識 (Fact Knowledge)”とは、対象世界の中で常に真となる個別事実データであり、“規則知識 (Rule Knowledge)”とは、その世界での推論に根拠を与える規則データである。

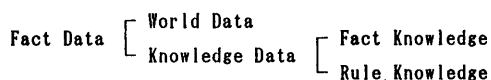


図 1. 事実データの階層

Fig.1 Hierarchy of fact data

システムの中の個々の知識データは、何れも S D 式で記述されたものであるが、任意の S D 式が知識データを表わしうるものではなく、以下の形式のみが知識データとなり得るものと定める。

(1) 事実知識データの形式

(a) 規定子形式

nega(D)

(b) 結合子形式

(D1)incl(D2) (D)incl([D1, ..., Dn])

(c) 陳述式

[s(D1), v(D2)]
[s(D1), v(D2), o(D3)]
[s(D1), v(D2), c(D3)]
[s(D1), v(D2), i(D3), o(D4)]
[s(D1), v(D2), o(D3), c(D4)]

以上の定義から分かるように、事実知識データとなり得る規定子形式は nega (否定概念を表わす知識データ) だけであり、結合子形式では概念の包含関係を示すものだけである。

(2) 規則知識データの形式

(a) 結合 S D 式

(assu(D1))caus(D2) および (assu(D1))indu(D2)

(b) 陳述 S D 式

[S(α), V(β), ...]

(a)の形式は、IF D1 THEN D2 の形式であり、通常は D1, D2 の中に変数を含み、推論の根拠となる知識である。(b)の形式は一つの陳述として表される概念である。

知識データは適用範囲や無矛盾性に注意して設定しなければならない。例えば、修飾子“/”の後には正確な修飾情報を与える。そのため特別なラベル“ANY”, “SOME”, “MOST”を設定し、“全ての”, “ある”, “大方の”を区別する。修飾子を持たないラベルには通常“SOME”が省略されているものと見なす。

本モデルでは、知識処理の対象世界は外部から指定するものとしている。そのため全ての事実データには世界名を付けておくものとする。

4. S D 式による概念相互の意味差の尺度

概念パターンに関する認識、理解、学習等の動作を定式化するために、それぞれの概念の意味的な差異に関する定量的な尺度が必要となるので、ここで次の 2 種の詳述関係 (elaboration relation) を定義する。

(1) S D 式の構文構造に基づく詳述関係

(2) 知識データに基づく詳述関係

4.1 構文構造に基づく詳述関係と意味差の尺度

本論文では生成規則の適用の度に加えられる情報の増分を生成規則の“詳述量 (elaboration score)”と呼ぶ。 $\Phi \Rightarrow X$ (開始記号から一変数の生成) は、“ある概念”即ち最小情報単位を生成しているのので、この生成規則の詳述量を“1 semit (semantic unit)”とする。そして、 $X \Rightarrow f_i$ (変数から一ラベルの生成) を 2 semit

とする。その他の詳述量については、前述の各生成規則の右端の()内に示した数値として定めた。

D の意味量 (Semantic Information score) を、D を得るために用いた生成規則の詳述量の累計で、 $SI(D)$ と定義する。

定理 1 $SI(D)$ の値は、一意的に定まる。

いま、2つのSD式 $D1, D2$ に関して、 $\Phi \Rightarrow D1$ であり、 $D1$ に生成規則を何度か適用して $D2$ となるとき、 $D1 \Rightarrow D2$ と表わす。 $D1$ から $D2$ に至る詳述量は、 $D1$ から $D2$ を得るために用いた生成規則の詳述量の累計であり、 $ELAB_{r.o.}(D1, D2)$ と表す。

$ELAB_{r.o.}(\Phi, D1) = SI(D1)$, $ELAB_{r.o.}(\Phi, D2) = SI(D2)$ であるので、

$$ELAB_{r.o.}(D1, D2) = SI(D2) - SI(D1)$$

と表せる

<例> $SI([s(\text{人}), v(\text{ほめる}), o(\text{彼}), c(\text{勇敢})]) = 17.0$

次に、 $\Phi \Rightarrow D$, $D \Rightarrow D1$, $D \Rightarrow D2$ が成立する場合、 D は $D1$ と $D2$ の共通先祖であり、

$$ELAB_{r.o.}(D, D1) = m_1, \quad ELAB_{r.o.}(D, D2) = m_2 \\ m^* = \min(m_1, m_2) = ELAB_{r.o.}(D^*, D1) + ELAB_{r.o.}(D^*, D2)$$

となる D^* を $D1, D2$ の最近共通先祖と呼び、これらの関係を次のように表す。

$$NCOA_{r.o.}(D1, D2, D^*)$$

また、この m^* を $D1$ と $D2$ の生成規則に基づく意味差 $DIFF_{r.o.}(D1, D2)$ と定義する。

<例> $NCOA_{r.o.}([s(\text{人間}), v(\text{歩く})], [s(\text{鳥}), v(\text{飛ぶ})], [s(X), v(X)])$
 $DIFF_{r.o.}([s(\text{人間}), v(\text{歩く})], [s(\text{鳥}), v(\text{飛ぶ})])$
 $= ELAB_{r.o.}([s(X), v(X)], [s(\text{人間}), v(\text{歩く})])$
 $+ ELAB_{r.o.}([s(X), v(X)], [s(\text{鳥}), v(\text{飛ぶ})])$
 $= 8.0$

次に、生成規則に基づく詳述関係を拡張し、構文構造に基づく詳述関係を以下のように定める。

$D1$ と $D2$ に関して、次の(1)-(6)の何れかが成立するとき、これを $D1$ から $D2$ への構文的な詳述 (syntactic elaboration) という。 $ELAB_{s.o.}(D1, D2)$ は、 $D1, D2$ の構文的な詳述関係およびそのときの詳述量を表すものとする。

<構文的な詳述関係>

(1) $D2$ は $D1$ から生成規則により生成されたものである。

$$ELAB_{s.o.}(D1, D2) = ELAB_{r.o.}(D1, D2)$$

(2) $D1$ がラベルで、 $D2$ が特別なラベル "SOME" 以外の任意のSD式 D により修飾された $D1/D$ の形式のものである。

$$ELAB_{s.o.}(D1, D2) = SI(D) + 1.0$$

(3) 次の各条件が成立する $D1$ と $D2$.

- (a) $D2$ が $D1$ から生成規則により詳述されていない。
- (b) $D1$ と $D2$ が同一形式 (2章f1-f8) のSD式である。
- (c) 記号列 $D1$ が、記号列 $D2$ の一部である。
- (d) $D1$ と $D2$ を対比させた場合、 $D1$ 中の $D1'$ と $D2$ 中の $D2'$ 部分が対応するものとする。このような任意の $D1'$ と $D2'$ について $ELAB_{s.o.}(D1', D2')$ が成立する。

以上の条件が成立するとき、

$$ELAB_{s.o.}(D1, D2) = SI(D2) - SI(D1)$$

この条件(3)は、構文的な詳述関係が再帰的に定義されるものであることを示している。

(4) $D1 = (D) \text{incl}(f1)$ かつ

$D2 = (D) \text{incl}([f1, f2, \dots, fi, \dots, fn])$ であるとき、

$$ELAB_{s.o.}(D1, D2) = 3(n-1)$$

(5) $D1 = D2$ のとき、

$$ELAB_{s.o.}(D1, D2) = 0.0$$

(6) $D2 = D1/SOME$ のとき、

$$ELAB_{s.o.}(D1, D2) = 2.0$$

即ち、ラベル SOME と同等である。

定理 2 $D2$ が $D1$ と詳述関係にあるとき、

$ELAB_{s.o.}(D1, D2)$ の値は、一意的に定まる。

定理 3 $ELAB_{s.o.}(D1, D2)$

$$= ELAB_{s.o.}(D1, D) + ELAB_{s.o.}(D, D2)$$

4.2 知識データに基づく詳述関係と意味差の尺度

構文的な詳述関係はなくとも、" [s(JACK), v(HAVE), o(HOUSE) / (BIG)para(FANCY)] ." は、" [S(JACK), v(WEALTHY)] ." よりも詳しい概念のように思える。このような状況を考慮し、新たに知識データに基づく詳述関係を導入する。

3章に示した形式の知識データの中で、次の(a)-(g)の結合子形式のデータのみに対して、 $D1, D2$ には知識データに基づく詳述関係 (knowledge-based elaboration) があるといひ、その詳述量 $ELAB_{k.o.}(D1, D2)$ を以下のように定義する。

(a) $(D1) \text{equa}(D2) \rightarrow ELAB_{k.o.}(D1, D2) = 0.0$

(b) $(D1) \text{defi}(D2) \rightarrow ELAB_{k.o.}(D1, D2) = 0.0$

(c) $(D1) \text{csof}([D_{11}, \dots, D_{1n}], [D_{21}, \dots, D_{2n}])$
 $\rightarrow ELAB_{k.o.}(D1, D_{2n}) = 1.0$

(d) $(\text{assu}(D2)) \text{caus}(D1)$

$\rightarrow ELAB_{k.o.}(D1, D2) = 1.0$

(e) $(D1) \text{incl}(D2) \rightarrow ELAB_{k.o.}(D1, D2) = 1.5$

(f) $(D2) \text{ptof}(D1) \rightarrow ELAB_{k.o.}(D1, D2) = 1.5$

(g) $(D2) \text{kdof}(D1) \rightarrow ELAB_{k.o.}(D1, D2) = 1.5$

<例> 知識データを次の通りとする。

K1: (MOON)defi(SATELLITE/EARTH)

K2: (HUMAN)csof([MAN, WOMAN])

K3: (FISH)incl(SALMON)

この知識データに基づく詳述量は次の通りである。

ELAB_{know}(MOON, SATELLITE/EARTH)=0.0
 ELAB_{know}(HUMAN, WOMAN)=1.0
 ELAB_{know}(FISH, SALMON)=1.5

以上の知識データに基づく詳述関係の他に、以下の特別な詳述関係を定義する。ここで D_i はラベルで、D は任意の S D 式である。

- (1) ELAB_{know}(D_i/MOST, D_i/ANY)=0.5
- (2) ELAB_{know}(D_i/D, D_i/ANY)=1.0
- (3) ELAB_{know}(D_i/SOME, D_i/D)=1.0
- (4) ELAB_{know}(D_i/SOME, D_i/MOST)=1.5
- (5) 次の知識データに関して、(a)~(d)を定義する。
 (assu(D_j))caus(D_i) or (D_i)csof([...D_j...])
 and
 (D_i)incl(D_k) or (D_k)ptof(D_i) or (D_k)kdof(D_i)
 (a) ELAB_{know}(D_i/SOME, D_j)=0.5
 (b) ELAB_{know}(D_i/SOME, D_k)=0.5
 (c) ELAB_{know}(D_j, D_i/ANY)=1.5
 (d) ELAB_{know}(D_k, D_i/ANY)=1.5

これらの(1)~(5)の関係を図2に示す。これらは、知識システムの推論機構の中で重要になってくる。図2中の SPEC 関係については、4.6節で述べる。

- (6) 次の同一形式の陳述 S D 式に関して、

$$D1=[s(\alpha), v(\beta), \dots]$$

$$D2=[s(\alpha), v(\text{nega}(\beta)), \dots]$$

この場合、

$$\text{ELAB}_{\text{know}}(\text{nega}(D1), D2)=1.0$$

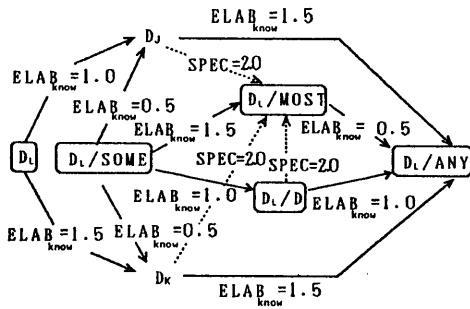


図2. ELAB_{know}の関係の詳細
 Fig.2 Specific ELAB_{know} relations

4.3 2種の詳述関係の統合

(ELAB(D1, D2)とDIFF(D1, D2)の定義)

これまでに概念間に構文構造に基づく詳述関係と知識データに基づく詳述関係を定義した。以下ではこれら2種を統合したものをあらためて“詳述関係”と定義し、記号 \supseteq で表す。また詳述量は ELAB_{know} と ELAB_{know} をまとめて ELAB_{know} (一段) または ELAB_{know} (多段) と表記する。

$$D1 \supseteq D2$$

$$\text{ELAB}_{\text{know}}(D1, D2) = \min \{ \text{ELAB}_{\text{know}}(D1, D2), \text{ELAB}_{\text{know}}(D1, D2) \}$$

$$\text{ELAB}_{\text{know}}(D1, D2) = \text{ELAB}_{\text{know}}(D1, D) + \text{ELAB}_{\text{know}}(D, D2)$$

(任意の D に対して、ELAB_{know}(D, D)=0)

この ELAB_{know} は、多段の詳述関係を定義している。一般に概念 D2 が、概念 D1 から直接的に詳述されていない場合にも、他の媒介概念 D を連結して、概念 D1 から詳述されていることもある。この場合、少なくとも一つの知識データに基づく詳述関係が ELAB_{know} 関係の連結の中に含まれている。そうでなければ、概念 D2 は概念 D1 から構文的に直接詳述されている。(定理3参照)。詳述量に関しては、次のように定義する。

$$\text{ELAB}(D1, D2) = \min \{ \text{ELAB}(D1, D) + \text{ELAB}(D, D2) \}$$

概念 D1 から概念 D2 への複数の異なる概念の連結パス (ELAB_{know}) は多段階になり得る。この場合、詳述量が最小となるパスを ELAB(D1, D2) と定義する。知識データが大量になるとこの連結パスを探し出すことが難しくなる。この事情は人間が多段推論を行なうときの難しさに似ている。

2つの概念 D1, D2 について、ELAB_{know} も ELAB_{know} も成立しないとき D1, D2 には詳述関係がないという。一般に知識データによる詳述量は構文的な詳述量よりも小さく定めているので、この定義は“意味の近さとしては、概念の構文的パターンよりも知識データを優先させる”ことを意味している。任意の概念 D は自分自身を詳述する。すなわち、D \supseteq D である。

D1 と D2 が詳述関係にある (D1 \supseteq D2) ととき、生成規則による詳述関係の (ELAB_{know}) の場合と同様に、D1, D2 および最近共通先祖 D* との関係

$$\text{NCOA}(D1, D2, D^*)$$

と表記する。これにより、S D 式で表された異なる2つの概念の意味差の尺度を次のように定義する。

$$\text{DIFF}(D1, D2) = \text{ELAB}(D^*, D1) + \text{ELAB}(D^*, D2)$$

4.4 知識の体系化

以上のように定めた意味差の尺度は距離の公理を満たさない。このことは、意味空間は距離空間ではないとすることであり、距離空間を対象とする従来からのパターン認識の手法は利用できないことになる。

詳述関係 \supseteq は半順序関係である。従ってシステム内の全ての知識データをこの半順序関係を用いて体系化することが可能である。以下に簡単な例を示す。

<例>

ある一つの世界における知識データが次のように与え

られているとする。()内の数値は各意味量である。

- [0]: X (変数) (1.0)
- [1]:(assu((X2)incl(X1)))caus (15.0)
- [2]:[s(人間),v(である),c(動物)] (13.0)
- [3]:[s(太郎),v(である),c(生徒)] (13.0)
- [4]:(生徒/賢い)incl(太郎) (11.0)
- [5]:(動物)incl(人間) (7.0)
- [6]:(人間)incl(太郎) (7.0)
- [7]:(動物)incl(生徒) (7.0)

これらの知識は図3のように体系化できる。楕円で囲んだものは与えられたデータで、四角で囲んだものは与えられたデータから得られた媒介概念である。矢印に付した数値は概念間の詳述量である。

各概念間の意味差の尺度は、すべてこの体系図に表されている。

- <例> DIFF([3],[4])=5.0 (semit)
- DIFF([4],[5])=9.0 (semit)
- DIFF([5],[7])=5.0 (semit)

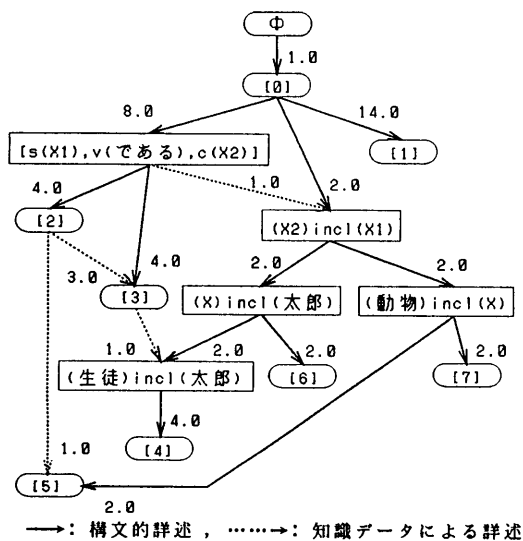


図3 体系化された知識データ
Fig.3 Organized knowledge data

SD式モデルでは知識ベース・システムに対して、与えられた知識データを自ら体系化(概念ネットワークの構築)できる能力を持たせることが可能である。このことは、従来から提案されている概念のネットワークが人手で作成することを前提としていたのに対し、本モデルでは機械的なアルゴリズムで実現できる特徴がある。

4.5 システムの知識量と記憶量

対象世界における知識データの"実知識量"(NK:

Net Knowledge)とは知識データを体系化した際の概念間の詳述量の総和である。一方、知識データの"総事実記憶量"(TM: Total fact knowledge-Memory)とは個々の知識データの意味量の総和である。図3の例では、

$$\text{実知識量 (NK)} = 55.0(\text{semit})$$

$$\text{総事実記憶量 (TM)} = 74.0(\text{semit})$$

となる。これらは知識ベース・システムにおける知識データの効率を定量的に評価するための指標となる。

4.6 SPEC関係の定義

ここまで、詳述関係に基づいて述べてきたが、次にSPEC関係を次のように定義する。

- (1) ELAB(D1,D2) が成り立つとき,
SPEC(D1,D2)=ELAB(D1,D2)
- (2) (assu(D2))indu(D1) が事実のとき,
SPEC(D1,D2)=1.5
- (3) SPEC(D_L/D_L/MOST)=2.0
- (4) ELAB_{know}(D_L,D_K)=1.0 または 1.5 のとき,
SPEC(D_K,D_L/MOST)=2.0

図2では、このSPEC関係の(3)および(4)を点線で示している。

SPEC関係を導入したので、DIFF関数を次のように拡張する。

$$\text{DIFF}(D1,D2)=\text{SPEC}(D*,D1)+\text{SPEC}(D*,D2)$$

5. 知識ベースにおける推論機構

知識ベース・システムは、事実データを基にした知識ベースと推論機構をもつ。

以下、対象世界での概念の真偽(true/false)、尤もらしさ(plausibility)、可能(possibility)、不明(uncertainty)判定を行う推論関数に関する定義を行い、認識、理解、学習等の機構について述べる。

5.1 真偽(true/false)関数

知識処理には概念の真偽や、尤もらしさ(plausible)の判定を伴うことが多い。本モデルではD₁,...をある事実データであるとするとき、詳述関係に基づき、D₁ ⊃ D₁,... なる全てのD₁を真とする。すなわち、true(D₁)とする。これは、与えられたD₁そのものがシステムの事実として存在するか、あるいは、D₁よりも詳しい概念が事実として存在すれば、当然D₁も正しいはずであるからである。これは事実データに基づく演繹的推論と言える。

さらにtrue(D)なるDに関して、D₂ ⊃ Dであればtrue(D₂)とする。これもまた演繹的推論である。

またC₁=caus, indu, およびandxのとき、D₁が結合子形式D₁=(assu(D_{1,1}))C₁(D_{1,2})であり、かつ

$D_{1,1} \supset D_{r,\dots,1}$, $D_{1,2} \supset D_{r,\dots,2}$ なる事実データ $D_{r,\dots,1}$, $D_{r,\dots,2}$ があるときも $\text{true}(D_i)$ とする。これは、 caus , indu , および andx の用法に照らして、 $D_{1,1}$ および $D_{1,2}$ が共に true であればこれらで結合された概念もまた正しいと考えられるからである。これは事実データに基づく帰納的推論に相当する。以上は詳述関係に基づく推論である。

一方、真 (true) に対する偽 (false) は、 $\text{true}(\text{nega}(D))$ の場合、 $\text{false}(D)$ と定義する。

5.2 尤もらしさ (plausible), 可能 (possible), 不明 (uncertain) 関数

変数 X や Y を含む概念 $D_i(X), D_j(X), D_o(Y), D_v(Y)$ に関する規則知識データ

$(\text{assu}(D_i(X)))\text{caus}(D_j(X))$

$(\text{assu}(D_o(Y)))\text{indu}(D_v(Y))$

があり、ある事実データと $D_i(X)$ や $D_o(Y)$ との単一化の結果が $X=\alpha$ や $Y=\gamma$ であるとき、

$\text{true}(D_j(\alpha))$ plausible($D_v(\gamma)$) とする。

これらは規則データによる演繹的推論である。逆に、ある事実データと $D_i(X)$ との単一化の結果が $X=\beta$ であるとき、

plausible($D_v(\beta)$) とする。

これらは事実データを基にしたアブダクションである。このような推論は、さらに多変数を含む規則知識データに基づく場合も可能である。

次に、可能 possible(D) とは、概念 D が真なるデータを基にした "indu" 関係による演繹的推論であるか、または plausible(D') なる概念 D' からの "caus" 関係によるアブダクションであるときと定義する。もちろん true と plausible は、possible に含まれる。

そして、不明 uncertain (D) とは、possible (D) でも false (D) でもない場合とする。

以上の推論関数をまとめると、次の図4のようになる。

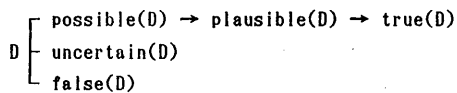


図4. 推論関数の階層

Fig.4 The hierarchy of inference functions

5.3 外部入力 の 認 識

外部入力 $D_{1,n}$ の認識動作を、システムが知識ベースの体系で、 $D_{1,n}$ に最も近い先祖 $D_{r,\dots}$ を見いだすことであると定義する。すなわち、

$$\text{ELAB}(D_{r,\dots}, D_{1,n}) = \min \text{ELAB}(D, D_{1,n})$$

である。このことを $D_{r,\dots}(D_{1,n})$ と表記する。変数ラベル X が事実データとして登録されており、

$$D_{r,\dots}(D_{1,n})=X$$

となった場合、その D は事実上システムにとって新しい概念である。

5.4 外部入力 の 理 解

外部入力 $D_{1,n}$ の理解動作を、以下の (A) または (B) の何れかであると定義する。

(A) $D_{1,n}$ が真なることを確認する。すなわち $D_{1,n}$ に対して $\text{true}(D_{1,n})$ を確認することである。

(B) $D_{1,n}$ と知識データからの帰結が真であることを確認すること。すなわち、 $\text{true}(D_{\dots}(D_{1,n}))$ を確認することである。

(B) は帰結が真であることから結局入力 $D_{1,n}$ も真であると認めることである。

しかし、我々が自然言語で "理解する / understand" という言葉を用いる場合は、(A), (B) の他に、

(C) $D_{\dots}(D_{\dots}(D_{1,n})) (\neq X)$

を得る動作も含めていると思われる。すなわち、帰結の真偽は定かではないが、既知概念の枠の中に在ることだけは認識出来る場合である。

5.5 外部入力 の 解 釈

外部入力 $D_{1,n}$ の解釈動作を、以下のような関係

$\text{INTP}(D_{1,n}, D)$ を満たす D をその解釈と定義する。

(1) $\text{SPEC}(D_{1,n}, D)$ または $\text{SPEC}(D, D_{1,n})$ が成立する。

(2) 次の (a) および (b) の条件を満たすとき。

(a) $D_{1,n}$, D は、同一形式の SD 式である。

(b) $D_{1,n}$ と D の互いに対応する $D_{1,n}'$ と D' が、

$\text{INTP}(D_{1,n}', D')$ を満たすとき。

5.6 外部入力 の 学 習 と 自 己 学 習

学習とは、認識や理解の能力を高めるための新たな知識データの獲得や、既知知識の再編成動作である。一般に知識データの記憶量が増えると応答の質 (正しい応答) は向上するが、システムの処理速度は低下する。従って知識の意味量をなるべく減らさずに、知識データの記憶量の削減と処理速度の向上が学習のねらいである。

新しい知識の獲得

外部からの入力 $D_{1,n}$ は知識データとの関係で次のように分類できる。

(1) $D_{r,\dots}(D_{1,n})=X$ (全く未知の概念)

(2) $D_{r,\dots}(D_{1,n})=D_{r,\dots}$ (既知の概念枠内での新しい概念であることを認識)

(3) $\text{true}(D_{1,n})$ または $\text{true}(D_{\dots}(D_{1,n}))$

(理解するが新しい情報はない)

従って、(1), (2) の場合が知識獲得動作の対象となる。

<知識獲得動作>

(1) の場合: D をそのまま取り入れる。

(2) の場合: あらかじめ定めた閾値 θ により、

- (a) $ELAB(D, \dots, D) < \theta$ のとき何もしない
 (b) $ELAB(D, \dots, D) \geq \theta$ のとき D を知識データとする。

(2)の(a)は知識データの記憶量を増加させないための入力の無視である。

自己学習機能

知識システムにおける自己学習とは、既にシステムに蓄えられている知識データを再編成して知識データの記憶量を減らすことと、処理速度の向上を計ることである。記憶量の削減の原理は次の通りである。

いま事実知識データの中の

$$D_1(\alpha_1), D_2(\alpha_2), \dots, D_n(\alpha_n)$$

は、もし $\alpha_1 = \alpha_2 = \dots = \alpha_n$ であれば全てが同一になるものとする。このとき $D_1(\alpha_1), D_2(\alpha_2), \dots, D_n(\alpha_n)$ を全て抹消し、その代わりに

$$D(\alpha) \text{ および } (\alpha) \text{csof}([\alpha_1, \alpha_2, \dots, \alpha_n])$$

を登録する。このことにより一般的には記憶量の削減が期待できる。これは一群の概念に新しい概念ラベルを与え、複数の個別事実知識データを一個のデータとして抽象化することに相当する。

6. SD式の会話文データベースへの応用

日本人が英語で話そうとするとき、まず英語の文法の知識をかりて日本語を英語に直訳しようとする。しかし、この文法だけに頼るやり方では必ずしも適切な会話表現とはならない。即ち、会話で使う口語体英語は文語体英語とは相当違って、この口語体英語の知識がなければ、会話の際に話したり聴きとることが困難になり、会話内容の理解は浅くなる。口語体英語の実例に即した表現の学習には、自然言語理解のための知識として必要な知識の中でも特に発話意図や状況に関する知識が重要である。

著者等は「NHKラジオ/テレビ英会話テキスト」を基にして会話例文を収集し、本モデルのSD式データを作成し「会話文データベース」の構築中である。

このSD式モデルの英会話文データベースへの応用例では、会話知識ベース内の会話環境、知識、心理・言語行動規則等を参照して、入力文の発話意図や状況に基づく実会話例文の検索出力が可能となる。

この会話知識ベースは、次のSD式データで構成し、実会話例文については、英語会話原文とそれに対応する日本語訳文を別のデータとしている。

- (1) 対象となる世界データからなるSD式
- (2) 発話意図を示す、陳述/感情式からなるSD式
各実会話例文の発話意図を、挨拶、呼びかけ、応答、問投詞、感嘆、感謝、謝罪、提案、依頼、命令、質問等に分類している。

- (3) 心理・言語行動を示すSD式
- (4) 会話環境を示すSD式
- (5) 話題関連知識を示すSD式
- (6) 一般知識を示すSD式

<例> 英会話の対話例文とSD式

Emi : Do you mind if I take a picture of them?

SD式1 : [s(自分), v(質問/求了解), o(相手),
c([s(自分), v(写真撮影), o(彼等)])]

訳文1 : この子たちと一緒に写真を撮ってもいいですか?

Woman : Oh, that would be lovely.

SD式2 : [s((自分)equa(女性)), v(応答/(肯定)para
(感嘆)), o(相手),

c([s(事柄), v(素敵)])]

訳文2 : まあ、それは素敵ね。

7. むすび

本論文は自然言語における概念の意味構造記述法の一つとしてSD式を提案し、それに基づくいくつかの知識処理をモデル化したものである。このSD式モデルは、会話文や物語文等の記述例を通して、自然言語の概念をかなり詳細に記述できることが確かめられた。

本SD式モデルの最大の特徴は従来からの意味記述モデルでは扱い難かった2つの概念間の意味差を定量的に扱えるようにした点である。今後の課題としては、詳述量の再検討、SD式実験システムSDENVの高速化、規模の大きい対象世界での動作実験およびSD式-自然言語間のインターフェイス問題等が残されている。

参考文献

- (1) Minsky, M. : "A Framework for Representing Knowledge." in The Psychology of Computer Vision, edited by Winston, P.H., McGraw-Hill Book Company (1975).
- (2) Quillian, M.R. : "Semantic memory, in Semantic Information Processing", edited by Minsky, M.L., pp.227-270, MIT Press(1968).
- (3) Montague, Richard : "Proper Treatment of Quantification" in Ordinary English, in Formal philosophy, edited by Thomason, R.H., pp.247-270. Yale University Press (1974).
- (4) Schank, R.C. and Abelson, R.P. : "Scripts, Plans, Goals and Understanding", Lawrence Erlbaum(1977).
- (5) 野村 : "自然言語理解の構造-理解の表現", 情報処理, Vol.30, No.10, 1989
- (6) 河口他 : "自然言語における概念パターンの記述モデルと認識理解過程の定式化" 電子情報通信学会技術研究報告, PRU90-28, pp.1-8, 1990