

SD式意味処理モデルによる自然言語概念の意味処理

Semantic Processing Using SD-Form Semantic Model on Natural Language Concepts

脇山 正博*¹
Masahiro Wakiyama

日高 康展*¹
Yasunobu Hitaka

吉原将太*²
Shota Yoshihara

河口英二*³
Eiji Kawaguchi

*¹ 北九州工業高等専門学校
Kitakyushu National College of Technology

*² 長崎純心大学
Nagasaki Junshin Catholic University

*³ 九州工業大学名誉教授
Kyushu Institute of Technology

Abstract

In this paper we adopt a context-free language (SD-Forms) to deal with meaning representation of a natural language. We introduce the basic idea of the model. The most fundamental idea in the model is the elaboration relation between two concepts. A semantic difference measure of concepts or sentence meaning is presented by using elaboration relations. It is not a logical formulation of meaning, instead, it is a data structure of concepts and sentence meaning when they are processed in a knowledge system. We propose the formalization of intelligent operations such as recognition, understanding and etc.

1. はじめに

自然言語処理の研究はAI研究の中の一つの典型的な分野であり、機械翻訳に関しては、既に多くの商用システムがある。しかしこれらに関しては、システムに柔軟性が無いこと、意味処理が取り入れられていないこと等が問題である。自然言語の研究において従来から提案されてきた知識や意味記述の手段としては、フレームや意味ネットワークが知られている。その他にも意味表現に関しては多くの試みがなされているが、それらが本格的なシステムの中で稼働しているとの報告は無い。

まず、自然言語概念の意味構造を表すための一つの記述式としてSD式モデル (Semantic-Structure Description Form Model) を採用する⁽¹⁾。SD式は具体的なデータとして記述する際の構文を規定し、多様な意味内容を計算機データとして記述できる。本論の目的は、自然言語における概念パターンを記号列として記述し、これを基にして工学的なシステムにおける概念の認識、理解等の知識処理を定式化するものである。

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“知識システム”と呼ぶ。知識システムにおいて個々の意味記述データ間、あるいは意味記述データの集合(例えば一つの物語データや限定された対象世界のデータ)の“意味距離”の尺度を計算するために、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、これに基づく意味差を導入する。詳述関係としては、SD式の構文構造に基づくものと、知識データに基づくもの2種を導入する。本論では、それをを用いて知識システムにおける認識、理解等の動作を定式化する。

2. SD式モデル⁽¹⁾

2.1 SD式の定義

SD式とは自然言語における個々の概念、陳述表現、感情表現、或はシステムに与える事実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ない一つの文脈自由文法SDGで規定されている。SD式は文脈自由文法SDGで生成される任意の記号列のことである。

2.4 SD式による概念間の詳述関係と詳述量

SD式として記述された2つの概念間の D_1, D_2 について、 D_2 は D_1 を“詳しく記述(elaborate)”したものであるとき、

$$ELAB(D_1, D_2) = n$$

と表す。ただし n は詳述の程度を示す“詳述スコア”である。

子の場合、 D_1 は D_2 の先祖、 D_2 は D_1 の子孫という。詳述の概念は、従来のIS-A, PART-OF, IF-THEN等を一般化した考え方である。このようなelaborationは、構文的な場合($ELAB_{\text{synt}}$)と知識システムで利用可能な知識データによる場合($ELAB_{\text{know}}$)の何れかであり、以下のように最小な値をとるように定めている。

$$ELAB(D_1, D_2) = \min\{ELAB_{\text{synt}}(D_1, D_2), ELAB_{\text{know}}(D_1, D_2)\}$$

そして2つの概念 D_1, D_2 について、 $ELAB_{\text{synt}}$ も $ELAB_{\text{know}}$ も成立しないとき D_1, D_2 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、 $D \ D$ である。

2.5 詳述関係の連結

与えられた概念 D_1, D_2 について、直接 $D_1 \ D_2$ が成立しないときでも、或る媒介概念 D に関して、 $D_1 \ D$ および $D \ D_2$ が成立することがある。この場合、これらを連結することによって $D_1 \ D_2$ が成立するものとする。詳述量に関しては

$$ELAB(D_1, D_2) = \min\{ELAB(D_1, D) + ELAB(D, D_2)\}$$

と定義する。

2.6 コンパクト概念体系

知識システム内の世界データや、知識データは一つ一つの概念として孤立して蓄えられているのではなく、全体を有機的に関連づけて蓄積しておく。この場合、各種の知的処理の効率を高めるための工夫と、知識の蓄積に要する記憶量を出来るだけ少なくすることが重要である。本論では、システムに具体的に与えられた事実データとしての概念集合 \mathcal{C}_0 を基にして創り出される一つの概念体系を以下のように定義する。

概念体系 $\{ / \mathcal{C}_0, \}$ とは、概念集合 \mathcal{C} と詳述量の集合 \mathcal{C} からなる意味ネットワークである。ここで概念集合 \mathcal{C} は変数 X と与えられた事実データの集合 \mathcal{C}_0 を含むものである。概念は意味ネットワークにおける節点であり、詳述関係は意味ネットワークの枝によって表される。

連絡先: 脇山正博, wakiyama@kct.ac.jp

3. 知識システムにおける知識処理

3.1 対象世界での真偽の判定

知識処理には概念の真偽の判定を伴うことが多い。本モデルでは D_{fact} を或る事実データであるとするとき、詳述関係に基づき、

D_1, D_{fact}
なる全ての D_1 を真とする。すなわち、
 $true(D_1)$

とする。これは、与えられた D_1 そのものがシステムの事実として存在するか、或いは、 D_1 よりも詳しい概念が事実として存在すれば、当然 D_1 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。

さらにまた $true(D)$ なる D に関して、 $D_2 \supset D$ であればやはり
 $true(D_2)$

とする。これもまた演繹的推論である。

また $C_1 = \text{caus, indu, および andx}$ のとき、 D_1 が

結合子形式 $D_1 = (\text{assu}(D_{11}))C_1(D_{12})$

であり、かつ

$D_{11}, D_{fact1}, D_{12}, D_{fact2}$
なる事実データ D_{fact1}, D_{fact2} があるときも
 $true(D_1)$

とする。これは、 $\text{caus, indu, および andx}$ の用法に照らして、 D_{11} および D_{12} 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 ($false$) の判定は、
 $true(\text{nega}(D))$ の場合 $false(D)$ と定義する。

3.2 対象世界での推論と尤もらしさ(plausible) や可能(possible) や不明(uncertain) の判定

変数 X や Y を含む概念 $D_i(X), D_j(X), D_u(Y), D_v(Y)$

に関する規則知識データ

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

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

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

$true(D_j())$

$\text{plausible}(D_j())$

とする。これらは規則データによる演繹的推論である。逆に或る事実データと $D_j(X)$ や $D_v(Y)$ との単一化の結果が $X=$ であるとき、

$\text{plausible}(D_j()), \text{possible}(D_u())$

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。

以上の推論を連結する多段推論も可能ではあるが、処理効率の点で困難を伴う。また任意の D に基づく推論の帰結概念を $D_{conc}(D)$ と表わす。

次に、可能 $\text{possible}(D)$ の判定は、概念 D' が $\text{plausible}(D')$ 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 $\text{plausible}(D')$ 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定義する。もちろん、 $true$ と plausible は possible に含まれる。

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

以上の概念の信頼性の関係をまとめると次の図1のようになる。

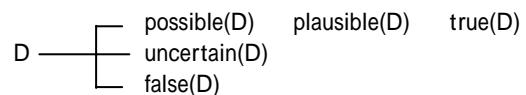


図1 概念の信頼性の階層

3.3 外部入力への認識

外部から概念 D_{in} が与えられたとする。知識システムが、この D_{in} を認識することとは、知識システム内のコンパクト概念体系 $\{ / 0, \}$ において D_{in} に最も近い先祖 D_{rec} を見いだすことである。

すなわち、

$$ELAB(D_{rec}, D_{in}) = \min_D ELAB(D, D_{in})$$

である。このことを $D_{rec}(D_{in})$ 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、

$$D_{rec}(D_{in})=X$$

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。

3.4 外部入力の理解

外部入力 D_{in} の理解動作とは、知識システムのコンパクト概念体系 $\{ / 0, \}$ において、 D なる任意の D に対して以下の (A) または (B) の何れかであると定義する。

(A) D_{in} が真なることを確認する。すなわち $D_{in} \supset D$ に対して

$$true(D_{in})$$

を確認することである。

(B) D_{in} 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、

$$true(D_{conc}(D_{in}))$$

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_{in} 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) $D_{rec}(D_{conc}(D_{in})) (X)$

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。

4. おわりに

SD式モデルを用いて、従来からの意味記述モデルでは扱い難かった2つの概念の意味差を定量的に扱えるようにした。それを用いて自然言語概念の意味処理の基本的な処理を定式化した。今後の課題としては、SD式の実験システムを作成し意味処理実験を行う予定である。

参考文献

[脇山正博 2005] SD式意味処理モデルにおける音声応答システム, 人工知能学会第19回全国大会(JSAI2005), 3F2-01, 6月15-17日, 2005