

通信プロトコルの自然語仕様から代数的仕様への変換における文章間の依存関係の解析

4H-3

石原 靖哲[†]八木 鉄也[†]島袋 潤[‡]関 浩之[†]嵩 忠雄[†][†]大阪大学基礎工学部情報工学科[‡]日立製作所

1 まえがき

筆者らは、通信プロトコル仕様のような‘動的’な自然語仕様を代数的仕様に変換する手法の考察を行なっている。‘動的’な仕様とは、定義しようとする機械(プログラム)への外部からの入力によって、内部変数の値の更新やデータ出力などの動作がどのように起こるかを記述した仕様をさす。一般に、このような自然語仕様では、入力に対する出力などの個々の動作が記述されているだけで、その動作の前提となる機械の状態については陽に表現されていないことが多い。

本手法では、各自然語文を、その構文木と、局所的な構文によって定まる情報とから、中間表現と呼ばれる形式に変換する。各文に対する中間表現は、その文が記述している動作の前提となる機械の状態を値としてとるような属性をもつ。そして、文章間の依存関係を解析することにより、その属性値を具体的に決定する。

本稿では、まず、各自然語文に記述されている動作の前提となる機械の状態を形式的に表現する手段として、状況という概念を導入する。そして、状況という概念によって、文章間の依存関係の解析法を簡潔に定義できることを示す。

2 状況

‘動的’な自然語仕様における各文について、その文に記述されている動作を機械が行なう直前の状態と直後の状態を、それぞれその文の直前の状況、直後の状況と呼ぶ。各文は、その直前と直後の状況の関係を規定していると考えられる。例えば、文献[1]において連続する次のような2文

s_1 : A valid incoming MAP results in an SSYNMind.

s_2 : If Vsc is false, V(A) is set equal to V(M).

を考える。ここで、文 s_1 は、“ s_1 の直後の状況は、 s_1 の直前の状況のもとで、正常な MAP の入力があり、かつその入力の直後に SSYNMind を出力した機械の状態と等しい”ということ述べている。

各文が規定している関係、すなわち、直前と直後の状況の関係を形式的に表現するために、状況を表すデータ型を、入出力や内部変数の値の更新などの各動作を表す式の有限長の系列型(以下、動作系列型と呼ぶ)と定義する。こうすれば、例えば、文 s_1 の直前の状況 S_{1pre} と直後の状況 S_{1post} の関係は、動

作を表す式 $in(MAP)$, $out(SSYNMind)$ (それぞれ、‘正常な MAP の入力’ ‘SSYNMind の出力’を表す)を用いて、

$$S_{1post} = S_{1pre} \cdot in(MAP) \cdot out(SSYNMind)$$

(‘ $\cdot$ ’は動作系列の連接を行なう演算子)

のように形式的かつ簡潔に表現できる。さらに、文 s_2 の直前の状況 S_{2pre} が S_{1post} と等しいことがわかれば、

$$S_{2pre} = S_{1pre} \cdot in(MAP) \cdot out(SSYNMind)$$

と表現できる。

このように、動作系列型の式である状況という概念を導入することによって、各自然語文が規定している直前と直後の状況の関係を形式的に表現できる。

3 中間表現

3.1 中間表現の定義

中間表現^{[3][5]}とは、自然語の文や語句について、局所的な構文によって定まる情報を一括して表現する形式であり、HPSG^[2]の語彙構造を参考にして定義したものである。形式的には、属性 f とその値 v との組 $[f, v]$ の集合(ただし、属性が等しく値が異なるような組は含まない)を中間表現と呼ぶ。現在、属性は18種類あり、例えば、属性 $type$ は、対応する語句の意味を表す式のデータ型名を値にとる。また、データ型が整数の集合であるような語句に対する中間表現は、 $[type, set(integer)]$ という組を含む。

自然語文の構文木と、その文に現れる各単語の中間表現とから、その文に対する中間表現は、構文木の構造に基づいて再帰的に(ボトムアップに)定義されている^[3]。

3.2 属性 event

中間表現の属性 $event$ は、その中間表現が対応づけられた文や語句の、直前の状況 v_{pre} 、直後の状況 v_{post} 、直前と直後の状況の関係 v_{relate} を値としてとる。実際には、 $event$ の値は、それ自身、

$$\{[pre, v_{pre}], [post, v_{post}], [relate, v_{relate}]\}$$

という中間表現の形式をしており、以下ではこれを

$$\begin{bmatrix} pre & v_{pre} \\ post & v_{post} \\ relate & v_{relate} \end{bmatrix}$$

のように表記する。

例えば、文

A Translation from Natural Language Specifications of Communication Protocols into Formal Specifications

Yasunori Ishihara[†], Tetsuya Yagi[†], Jun Shimabukuro[‡], Hiroyuki Seki[†] and Tadao Kasami[†]

[†]Osaka University [‡]HITACHI, Ltd.

s_3 : The SPM starts the timer.

の starts の中間表現における属性 event の値は,

$$\begin{bmatrix} \text{pre} & S_1 \\ \text{post} & S_2 \\ \text{relate} & S_2 = S_1 \cdot \text{start}(X) \end{bmatrix}$$

と定義される。ここで X は, starts の中間表現において starts の目的語を表している。そして, starts の中間表現と the timer の中間表現とから句 starts the timer の中間表現が構成されると, その event の値は

$$\begin{bmatrix} \text{pre} & S_1 \\ \text{post} & S_2 \\ \text{relate} & S_2 = S_1 \cdot \text{start}(tim) \end{bmatrix}$$

となる。ここで tim は, the timer の意味を表す式である。すなわち, 機械の動作を表し得る単語について属性 event の値を定義することにより, その単語を含む文の直前と直後の状況の関係を一意に決定することができる。しかし, これだけでは, 直前と直後の状況の具体的な値を一意に決定することができない。これらの値は, 次節で述べる, 文章間の依存関係の解析を行なうことにより, 決定される。

4 文章間の依存関係の解析

4.1 依存グラフ

自然語仕様の各文の依存関係を, 依存グラフ^[4]と呼ばれる有向無閉路グラフで表現する。依存グラフにおいて, 頂点 s_i から頂点 s_j への有向枝が存在するのは, 文 s_i の直後の状況が文 s_j の直前の状況となるとき, かつそのときのみである。例えば, 2 節で示した自然語文 s_1, s_2 において, s_1 の直後の状況が s_2 の直前の状況となるとき, 依存グラフには有向枝 $s_1 \rightarrow s_2$ が存在する。以下では, 自然語文あるいはそれに対応する中間表現と依存グラフの頂点とを同一視して扱う。

依存関係で閉じた文の系列を段落と呼ぶ。 n 個の文の系列 $s_1, s_2, \dots, s_n$ からなる段落 P に対する依存グラフは, 以下のようにして求められる。

1. P に対する依存グラフの根を s_1 とする。
2. 文の系列 $s_1, \dots, s_{i-1}$ に対する依存グラフ G_{i-1} と文 s_i に対し,
 - (a) s_i が場合分けを表す語 (if, otherwise など) を含むかどうかにしたがって, s_i の親頂点の候補を決定する (4.2 節参照)。
 - (b) (a) で得られた候補のなかで, 前方照応による制約条件 (4.3 節参照) を満たしている頂点を s_i の親頂点とし, 文の系列 $s_1, \dots, s_i$ に対する依存グラフ G_i を得る。

4.2 構文による場合分け

文 s_i がどのような場合分けを表す語を含むか (あるいはどれも含まないか) によって, s_i の親頂点の候補を $s_1, \dots, s_{i-1}$ の中から選び出す^[3]。例えば, s_i が語 otherwise を含むときは,

$s_1, \dots, s_{i-1}$ において if を含んでいる頂点のうち, 添字が最大の頂点 s_j の親頂点

を s_i の親頂点の候補とする。つまり, もとの段落において, 文 s_j は文 s_i に最も近い if を含む文であり, この if を s_i の otherwise に対応させるわけである。

4.3 前方照応による制約条件

正しい依存グラフであるための必要条件として,

s_i に前方照応句が含まれている場合, それに対する先行句が, 依存グラフの根から s_i への道上のある頂点 (文) に含まれなければならない

を用いる。したがって, 前方照応句と先行句との対応関係が与えられているならば, この条件を用いることにより, s_i の親頂点の候補を絞りこむことができる。

逆に, 構文による場合分けだけで s_i の親頂点が一意に決定された場合, この条件により, s_i 中の前方照応句とその先行句に関する曖昧性の減少が期待できる。

5 あとがき

OSI セッションプロトコル仕様^[1]の主要部 29 段落約 100 文に対して本解析法を適用した結果, 一つの例を除いて, 依存関係を一意に正しく決定することができた。なお, 依存関係を一意に決定できなかった例についても, 例えば, “タイマの停止” が意味を持つためには, タイマが動作していなければならない— といったような, 特定のデータ型に依存した性質を用いることにより, 文章間の依存関係を正しく決定できる場合がある。この点については, 現在検討中である。

また, ‘動的’ な自然語仕様を本手法に基づいて代数的仕様に変換するシステムを現在試作中である。

参考文献

- [1] ISO: “Basic connection oriented session protocol specification”, ISO 8327.
- [2] C.J.Pollard: “Lectures on HPSG”, unpublished manuscript, Stanford University (1985-02).
- [3] 石原: “通信プロトコルの自然語仕様から代数的仕様への変換システムにおける中間表現構成部および文章間の依存関係解析部の実現”, 大阪大学基礎工学部情報工学科特別研究報告 (1990-02).
- [4] 島袋, 萩本, 関, 藤井, 嵩: “自然語仕様から代数的仕様への文脈を考慮した変換”, 信学技報, SS89-3 (1989-07).
- [5] 並河, 松村, 関, 藤井, 嵩: “自然語によるプログラム仕様の, 抽象的データタイプを用いた構文及び意味解析”, 情報学研報, NL-65-5 (1988-03).