

確率文脈自由文法を用いた漸進的構文解析における出力タイミング決定手法

加藤 芳秀†

松原 茂樹†

外山 勝彦‡§

稲垣 康善†

†名古屋大学大学院工学研究科

‡名古屋大学言語文化部

§名古屋大学統合音響情報研究拠点

1 はじめに

実時間対話処理システムを実現するためには、自然言語文をその出現順序に従って順次解釈する枠組，すなわち，漸進的解釈手法が不可欠となる。相手の発話を漸進的に解釈することにより、発話途中での割り込みの生成や即時的な応答が可能となり、自然で円滑な対話の実現が期待できる。

漸進的に意味解析や文脈解析を行うためには、入力途中の段階でそれまでの入力に対する構文構造を作成する必要がある。それを実現する手法として、漸進的チャート解析が提案されている [1]。この手法は、語が入力されるごとにそれまでの入力に対する構文構造を随時作成する。しかし、作成される構造は、以後の入力に対する解析に用いられることのない不適切な構造である可能性がある。そのような構造をもとに、意味解析や文脈解析を行い、発話内容を定めると、対話システムは相手の発話に対して不適切な応答をすることになる。それを解決する方法として、語単位で構文解析結果を出力するという制約を緩和することが考えられる。

そこで本稿では、漸進的な構文解析において出力タイミングを動的に決定する手法を提案する。本手法では、確率文脈自由文法 (PCFG) を用いて、入力途中の段階で随時、構文構造が適切である可能性を評価し、評価値が閾値を超えた構造を出力する。これにより、適切である可能性が高い構文構造をできるかぎり早い段階で出力することができる。

2 漸進的チャート解析における構文構造の決定

漸進的チャート解析 [1] は、従来の上昇型チャート解析に、活性弧に文法規則を適用する操作、及び活性弧の項の最左未決定項を別の活性弧の項で置き換える操作を導入した枠組である。この手法では、語が入力されるごとにそれまでの入力に対する構文構造を作成する。例えば、図 1 に示す文法と辞書を用いるとき、英語文

(1) I saw the girl with the telescope.

における “I” が入力された段階で、“I” に対する項

(2) $[[I]_{np}[?]_{vp}]_s$

が作成される。次に “saw” が入力された段階で、“saw” に対する項 $[[saw]_{vi}]_{vp}$, $[[saw]_{vt}[?]_{np}[?]_{pp}]_{vp}$, $[[saw]_{vt}[?]_{np}]_{vp}$ が作成され、(2) の最左未決定項 $[?]_{vp}$ をこれらの項で置き換えることにより、“I saw” に対する項

(3) $[[I]_{np}[[saw]_{vi}]_{vp}]_s$

(4) $[[I]_{np}[[saw]_{vt}[?]_{np}[?]_{pp}]_{vp}]_s$

(5) $[[I]_{np}[[saw]_{vt}[?]_{np}]_{vp}]_s$

をそれぞれ作成することができる。これらの項より、この段

(g1) s → np vp	(d1) np → I
(g2) np → det n	(d2) det → the
(g3) np → det n pp	(d3) n → girl
(g4) pp → p np	(d4) n → telescope
(g5) vp → vi	(d5) p → with
(g6) vp → vt np	(d6) vt → saw
(g7) vp → vt np pp	(d7) vi → saw

図 1: 構文解析のための文法と辞書

階で動詞 “saw” の主語は “I” であるなどといった文の構成要素間の関係を捉えることができる。

漸進的チャート解析では、以後の入力を予測して項を作成するため、その項はそれ以後の解析に用いられることのない不適切な項である可能性がある。例えば、英語文 (1) において、(3) は “saw” 以後の入力に対して解析を続行できない不適切な項である。

項が適切であるかどうかは以後の入力に応じて定まる。したがって、上の問題を解決する方法として、適切な項の特定に十分な情報が得られるまで出力を遅延することが考えられる。それを実現する手法として、これまでに著者らは、項が適切であることが保証されるまで出力を遅延する手法を提案している [2]。この手法では、項の包含関係、及び項の未決定範疇列を用いて、項が適切であるかどうかを判別する。

項 σ に対して最左未決定項を別の項で置き換える操作を有限回実行することにより項 τ が得られるとき、 σ は τ を包含するといひ、 $\sigma \triangleright^* \tau$ と書く。これは、 τ が σ を具体化した項であることを意味する。例えば、(2) は (3) ~ (5) を包含している。次に、 τ を入力文 $w_1 \dots w_n$ に対する項とするとき、 $\sigma \triangleright^* \tau$ ならば、 $w_1 \dots w_n$ に対して σ は適切であるという。例えば、英語文 (1) に対して (4), (5) は適切である。

項の未決定範疇列は、項に出現する未決定項の範疇を左から順に並べたものであるが、これは、適切さの一致する項をまとめるために用いられる。例えば、英語文 (1) に対して、“the” が入力された段階で、“I saw the” に対する項

(6) $[[I]_{np}[[saw]_{vt}[[the]_{det}[?]_n[?]_{pp}]_{np}[?]_{pp}]_{vp}]_s$

(7) $[[I]_{np}[[saw]_{vt}[[the]_{det}[?]_n[?]_{pp}]_{vp}]_s$

(8) $[[I]_{np}[[saw]_{vt}[[the]_{det}[?]_n[?]_{pp}]_{vp}]_s$

(9) $[[I]_{np}[[saw]_{vt}[[the]_{det}[?]_n[?]_{pp}]_{vp}]_s$

が作成されるが、(7) と (8) の未決定範疇列はいずれも n, pp である。これらの項は、以後の入力に範疇がそれぞれ n, pp であるような構成要素が入力されたとき、かつそのときに限り、適切となる。

この手法では、 i 番目の語 w_i が入力されたとき、入力文 $w_1 \dots w_n$ に対して項が適切であるかどうかを次のように判別する。

- 1) 項の集合の分割 入力文 $w_1 \dots w_n$ に対して、チャートの節点 0 と節点 i を結ぶ弧の項全体からなる集合を T_i とする。さらに、 T_i 中の未決定範疇列が一致する項を一つのブロック $T_{i,k}$ にまとめる。
- 2) 適切さの判別 σ をチャートの節点 0 と節点 $j (j \leq i)$ を結ぶ弧の項とする。任意のブロック $T_{i,k}$ が σ に包含される項を要素にもつならば、 $w_1 \dots w_n$ に対して σ は適切である。

未決定範疇列の一致に基づいて T_i を分割した各ブロックのいずれかが適切な項の集合であるが、それがいずれであっても、適切さの判別の条件を満たす σ は各ブロックのある項を包含するので、適切な項の集合のある項を包含する。すなわち σ は適切であることが保証される。項が適切さを判別することにより、漸進的チャート解析において適切な項のみを出力することができる。

3 構文構造の適切さの可能性の評価

2節で述べた手法は、項が適切であることが保証された段階で構文構造を出力するため、漸進性が損なわれる場合がある。漸進性を高めるために、適切である可能性が高い項を出力することが考えられる。そこで本節では、2節で述べた手法を拡張し、項が適切である可能性を評価し、それが一定の高さになるまで出力を遅延する手法を提案する。

2節で述べた手法では、それまでの入力に対する項の集合を、未決定範疇列の一致する項をまとめたブロックに分割するが、項 σ が適切となるのは、 σ に包含される項を要素にもつブロックのいずれかが適切な項の集合となるときである。つまり、 σ に包含される項を要素にもつブロックが適切な項の集合となる可能性が高いほど、 σ が適切である可能性が高いと考えられる。そこで本手法では、 σ が適切となる可能性を、 σ に包含される項を要素にもつブロックが適切な項の集合となる可能性により評価する。

3.1 PCFG による適切さの可能性の評価

本手法では PCFG を用いて構文構造が適切である可能性を評価する。PCFG は、構文構造の生起する確率を計算するモデルである。文法規則に対して確率値が与えられており、構文構造の生起確率は、その構造の構成に用いられた文法規則の確率値の積により与えられる。例えば、図 1 の文法に対して、表 1 のように確率値を与える場合、項 (3) の構成に用いられた文法規則は (g1), (g5), (d1), (d7) であり、生起確率は $1 \times 0.2 \times 0.2 \times 1 = 0.04$ となる。

$w_1 \dots w_i$ が入力されたときの、項 $\tau \in T_i$ の生起確率 $p_i(\tau)$ は次のように求められる。なお、 τ の生起確率を $p(\tau)$ と書く。

$$p_i(\tau) = \frac{p(\tau)}{\sum_{\tau' \in T_i} p(\tau')}$$

例えば、項 (3) ~ (5) の生起確率はそれぞれ 0.04, 0.08, 0.08 であり、その和は 0.2 である。このとき、英語文 (1) の “saw” が入力された段階での項 (3) の生起確率は $0.04/0.2 = 0.2$ となる。

項 σ が適切となる可能性について考える。 σ が適切となるのは、文に対する項で、 σ に包含されるものが生起するときである。したがって、そのような項が生起する確率が高いほど、 σ が適切となる可能性は高い。ところで、 σ の生起確率は、まさにこの確率を表している。したがって、生起確率の高い項ほど、適切である可能性が高い。例えば、項 (3)、及び (4) の “saw” が入力された段階での生起確率はそれぞれ 0.2, 0.4 であるが、これは、(3) に比べて、(4) が適切となる可能性が高いことを表している。

次に、未決定範疇列の一致する項をまとめたブロック $T_{i,k}$ が適切な項の集合となる可能性について考える。 $T_{i,k}$ 中のいずれかの項が適切となるとき、 $T_{i,k}$ が適切な項の集合にな

表 1: 文法規則と確率値

文法規則	確率値
(g1), (g4), (d2) ~ (d7)	1
(g2), (g3), (g6), (g7)	0.4
(g5), (d1)	0.2

る。よって、 $T_{i,k}$ 中の項の生起確率の和により $T_{i,k}$ が適切な項の集合となる可能性を評価する。

以上から、 i 番目の語 w_i が入力されたときの項 σ が適切である可能性 $score_i(\sigma)$ を次のように定義する。なお、 σ に包含される項を要素にもつブロックを $T_{i,1}, \dots, T_{i,l}$ とする。

$$score_i(\sigma) = \sum_{s=1}^l \sum_{\tau \in T_{i,s}} p_i(\tau)$$

$score_i(\sigma)$ は σ に包含される項を要素にもつブロックの評価値の和である。 $0 \leq score_i(\sigma) \leq 1$ であり、 $score_i(\sigma) = 1$ のとき、 σ が適切であることが保証される。 $score_i(\sigma) = 0$ となるのは、 σ が適切でないときである。これにより、語が入力されるごとに随時、項が適切である可能性を評価する。

例えば、英語文 (1) に対して、“saw” が入力されたとき項 (3) ~ (5) が作成され、その生起確率 p_2 はそれぞれ 0.2, 0.4, 0.4 である。これらの項は、 $\{(3)\}$, $\{(4)\}$, $\{(5)\}$ の 3 つのブロックに分割され、その評価値はそれぞれ 0.2, 0.4, 0.4 となる。結局、(3) ~ (5) の評価値 $score_2$ はそれぞれ 0.2, 0.4, 0.4 となる。実際、英語文 (1) に対して、評価値の低い (3) は適切でない項であり、評価値の高い (4), (5) は適切な項である。次の語 “the” が入力されたとき、項 (6) ~ (9) が作成され、その生起確率はすべて 0.032 である。これらの項は $\{(6)\}$, $\{(7), (8)\}$, $\{(9)\}$ の 3 つのブロックに分割され、その評価値はそれぞれ、0.25, 0.5, 0.25 となる。(3) はいずれの項も包含せず、(4) は (6), (7) を、(5) は (8), (9) を包含するので、(3) ~ (5) の評価値 $score_3$ はそれぞれ 0, 0.75, 0.75 となる。

3.2 評価値に基づく構文構造の出力

本手法では、 i 番目の語 w_i が入力されたとき、 $score_i$ があらかじめ与えた閾値を超えた項を随時出力する。例として、閾値を 0.7 と定めた場合を考える。英語文 (1) に対して、“saw” が入力されたとき、項 (3) ~ (5) はいずれも閾値を超えないため出力されない。“the” が入力されたとき、(3) は閾値を超えないため出力されないが、(4), (5) は閾値を超えるため出力される。実際に、(4), (5) は適切な項である。

4 おわりに

本稿では、漸進的な構文解析において入力途中の段階で随時、構文構造が適切となる可能性を評価し、評価値が閾値を超えた構文構造を出力する手法を提案した。実験により、本手法の有効性を確認することは今後の課題である。

参考文献

- [1] Matsubara, S. et al.: Chart-based Parsing and Transfer in Incremental Spoken Language Translation, *Proc. of NLPRS-97*, pp.521-524 (1997).
- [2] 加藤他: 漸進的構文解析における構文構造の動的決定手法, 言語処理学会第 5 年次大会, pp.84-87(1999).