3—4—8

音韻モデルの照合における制約の利用法*

◎伊藤克亘　△田中穂積　(東京工業大学)
速水　悟　(電子技術総合研究所)

連続音声認識で、さまざまな制約を利用して性能の向上をめざそうとする試みは数多くなされている。本稿では、連続音声認識システムにおいて、さまざまな制約を利用しながら音韻モデルを照合するときに、処理をまとめ、再計算を防ぐ枠組についてのべる。

1　音韻モデルの照合における制約の利用法

連続音声認識に何らかの制約を導入しようとした場合に、従来の自然言語処理で用いられている方法を利用しようとするには、考慮すべき点がある。ここでは、音声認識でもよく用いられる 一般化 LR 構文解析法 [1] (以下、LR 法とする) を例にとって、考慮すべき点についてのべる。

LR 法では、解析途中に再計算を防ぐために、マージという操作をおこなう。しかし、LR 法でマージをおこなうためには、途中で生じた複数の候補が同じ状態でシフト動作をするように同期をとらなければならない。この待ち合わせは、複数の候補を並列処理する場合に問題となることが知られている [2]。LR 法を音声認識に利用する場合は、テキストとして全く異なった複数の候補を並列してあつかうことになる。たとえば、ある発話を認識する途中に、/w a k a r a n a i/(分らない) と /k a m a w a n a i/(構わない) のふたつの候補があらわれた場合、その構文木は、下の図のようになる。

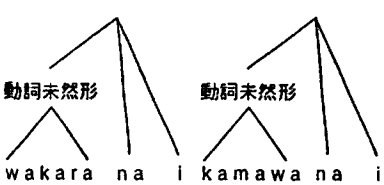

図 1 構文木の例

これらは、「動詞未然形」になってしまうと、それ以降は同じように、/n a i/の処理をおこなう。これらを共通してあつかうためには、異なるテキストを非同期で並列に処理できるスタックの構造が必要となる。(問題 1)。

このように、マージをそのまま導入するのも困難なのであるが、マージだけでは、完全に解析途中に生じる再計算を防ぐことはできない。例えば、以下のふたつの規則を考える。

```
A -> X Y Z
A -> W Y Z
```

これらの規則では、Y Z の部分の部分木の解析は同じであるが、LR 表ではことなる状態で処理される。LR 法を音声認識に利用する場合は、テキスト入力の場合と違い、先読み記号の処理のコストが大きいので、このように、同じ処理があちこちでおこなわれると無駄な処理が多くなる。また、解析中に生じる複数の候補は、同時に LR 表の様々な状態を参照する。同じ先読み記号の処理を複数の候補で別々に処理することもあるため、この場合も無駄な処理が多くなってしまう。(問題 2)

また、文法以外の制約も利用する場合には、マージをすることによる弊害もある。マージは、構文解析中の候補の状態をより上位の非終端記号に還元するという操作と対応して用いられる。しかし、構文以外の制約 (たとえば、単語と単語の間の意味的関係など) を利用して探索をおこなうためには、非終端記号に還元してしまうわけにはいかない。

たとえば、図 1 でマージをおこなってしまうと、「分らない」と「構わない」の意味の違いを利用した場合には、結局、まとめられている候補を展開しなければならなくなる。そこで、構文的な操作以外には、マージの影響を及ぼさないようにしなければならない (問題 3)。

問題 1 は、多くの音声認識システムで利用されている、解析の履歴を保持するという考え方で処理できる。LR 法の場合には、スタックのための仮説と、LR 表の先読み記号のための仮説を用意すればよい [3]。これらを用意することで、同じ構文解析を繰り返すことは避けられる。

問題 2 では、オートマトンの状態で、直接、処理をまとめていることから不十分さが生じている。そこで、オートマトンの状態で直接まとめるのをやめて、その時点で、考慮されている状態から出ている枝での処理が同じものはまとめるようにすればよい。このように対処すれば、プッシュ・ダウン・オートマトンを展開してしまわなければならない場合にも、状態数が多くなるのは防げないが、無駄な処理を避けることはできる。

問題 3 は、複数の制約を利用するときには、それぞれの制約を処理するオートマトンは全体として階層的に処理できない場合もあるということである。同様の例として、文脈自由文法と N-gram モデルを併用する場合があげられる。

しかし、その一方で、それぞれの制約の間には、階層関係がなりたつ場合もある。したがって、部分的な階層関係を利用してまとめることはできるが、上位の階層へ還元してしまうことはゆるされない。この原則は、オートマトンを利用して N-Best な探索をする場合にもあてはまり、途中で、候補をまとめておく以外には、結局、上位の階層へと還元できない。

本論文で提案する手法では、これらの問題点を解決する機構として、様々な知識を利用する場合に、それぞれのレベルにセルとよばれるものを用意する。

セルは、その時点で探索に利用する全てのスコアを参照できるようにする。また、それぞれの制約間の依存関係は、セルどうしの依存関係に反映させて、処理をまとめる。

次の節では、単語の音韻系列とその構文カテゴリを

*Utilization of Multiple Constraints on the Phone Model Matching
by ITOU Katunobu, TANAKA Hozumi (Tokyo Institute of Technology) and HAYAMIZU Satoru (Electrotechnical Laboratory)

あらわす辞書と構文カテゴリに関する文法の制約を利用する場合の方法を例にとって説明する。

2 セルを利用した再計算の防止

連続音声認識システム niNja では、辞書は構文カテゴリごとに駆動され、音韻レベルの仮説を生成する。例えば、「ここに本が一冊あります」という文を認識しているときには、下のような仮説が生成される。

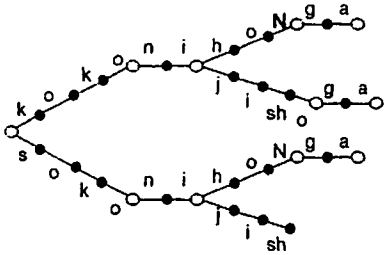

図 2 音韻レベルの仮説の例

この音韻仮説の枝ごとに、音韻モデルを照合する。niNja では、音韻の照合の部分には、音韻セルを導入して、辞書の制約を用いて処理をまとめている [4]。たとえば、/k o k o/ と /s o k o/ の /k/ や /o/は、照合をひとつにまとめておこなう。しかし、照合はひとつにまとめられても、音韻セルは、図 2 の節点の数だけ生成されるため、認識がすすんで候補が多くなると数がふえる。

音韻仮説に、構文の情報を利用する単語セルを導入する。単語セルを導入すると以下のようになる。

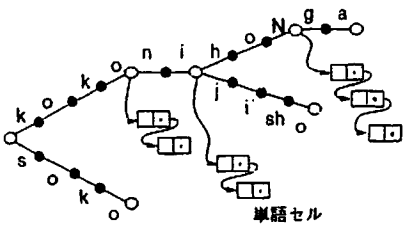

図 3 単語セルの例

この仮説では、/s o k o n i/ と /k o k o n i/ のあとに、/h o N/ などを含む辞書が駆動されている。この /h o N/ などを含む辞書名を "Z" とすると、構文レベルのオートマトンで "Z" が先読み記号となっているアークのうち、その時点で考慮されている全てのアークでは、/h o N/, /j i sh o/ などを照合することになる。つまり、"Z" の辞書引きが終了するまでの動作は、その時点でのスコアによって、多少枝刈りされるものがあるだけで、同じである。

音韻セルは図 3 の場合でも、節点に対応して生成されるので、単語セルを導入しない場合と比べると、単語セルでまとめられる分だけ減る。

セルは探索に必要なスコアを全て含むので、フレーム同期のシステムであれば、各フレームで、音韻同期のシステムであれば、各音韻で更新される。

ある時点でのセルの数は、セルどうしの依存関係の最上位に位置するレベルに関しては、更新する時点で考慮している全候補数となる。それ以外の、上位に位置するレベルをもつセルの数は、最悪の場合で、そのレベルでの終端記号の数となる。

ここで説明した例では、構文レベルの制約が最上位に位置しているので、単語セルは、その時点での候補数だけ必要となる。しかし、音韻セルの数は、単語セルを導入しなかったときには、候補数の数だけ必要だったのが、単語セルを導入すると総単語数ですむ。タスクが難しくなったり、枝刈りの制限をゆるめると、候補数は、組合せ的に増大する。単語セルを導入すると、音韻セルの数は減るが、単語セルが新たにふえる。ある時点での、単語セルの数は候補数だが、更新の時機が、音韻セルの場合は音韻が完成したときで、単語セルは単語が完成したときなので、結局、生成される単語セルの総数は、単語セルを導入しない場合の音韻セルの総数よりも減ることになる。

単語セルは音韻セル自体をまとめる操作しかしないので、探索の精度が単語セル導入によってかわることはない。

3 実験結果

単語セルの導入によって、照合のスコアは変化しないので、ここでは、生成されるセルの数と、処理時間を比較した結果について報告する。

文節の平均分岐数 650、平均文節数 3.0 の文法を用いて、枝刈りの条件を 3 段階に変化させて、実験をおこなった。表には、11 文を認識したときのセルの総数の平均値を示す。「単語セルあり」のセルの数は、音韻セルと単語セルの和である。

枝刈り条件			→ ゆるい
単語セルあり	163751	570212	2010510
単語セルなし	327860	1811130	11890900
比率	2.00	3.18	5.9

枝刈りの条件をゆるくすれば、ゆるくするほど、セルの数の比率が広がっていくのがわかる。

4 おわりに

本稿で述べた手法は、オートマトンを利用するものなら導入することができる。ただし、LR 法のように、入力の順序にしたがって、状態を遷移するものや、一般化弁別ネットワークなどの漸進的処理とよばれる手法を用いて、入力がはいるごとに状態を遷移させるものの方が、音韻モデルの照合のための制約としては適しているだろう。

謝辞

日頃御支援と御討論を頂く東工大田中研の皆様、並びに電総研音声研究室の皆様に感謝致します。

参考文献

[1] M. Tomita. An efficient augmented-context-free parsing algorithm. *Computational Linguistics*, Vol. 13, No. 1-2, pp. 31–46, 1987.

[2] 峯他. 文脈自由文法の並列構文解析. 情処 NL 研, Vol. NL73-1, 1989.

[3] 伊藤他. 拡張 LR 構文解析法を用いた連続音声認識. 信学技報, Vol. SP90–74, pp. 49–56, 1990.

[4] 伊藤他. 音素文脈依存モデルと高速な探索手法を用いた連続音声認識. 信学論, Vol. J75-D-II, No. 6, pp. 1023–1030, 1992.