1E−1

日本語連続文音声認識における探索手法の検討

小島　英樹・岩見田　均・木村　晋太
（　株式会社　富士通研究所　）

1．はじめに

我々は、音響セグメントネットワークを用いた単語認識方式を提案し、１０万単語認識においてその有効性を示した[1]。また、日本語の格支配構造を柔軟に記述する方式としてCASE　FLAG伝播法を提案し、文節単位の音響セグメントネットワークを用いて日本語文音声認識を行ない有効性を確認した[2]。しかし、認識にかかる計算量が多いという問題点があった。

本報告では、コンテキスト依存型の音節単位の音響セグメントネットワークの導入と探索手法の検討により処理量の削減を行なったので、その結果について報告する。

2．処理の概要

図１は本方法の概念図である。深層格レベルの係り受け文法を用いて、文末からツリー状に文節の仮説を生成して入力音声とのマッチングを行う。本報告では、音節単位の音響セグメントネットワークの導入と探索アルゴリズムの検討を行ない実験を行なった。音節単位のネットワークは無声化などの音節間にまたがる音素変形も考慮したものである。探索手法としては従来のビームサーチの代わりにbest-firstサーチを導入した。

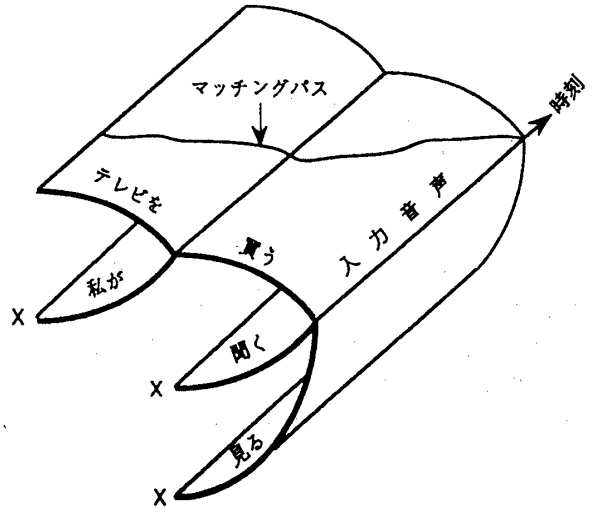

図１　方式概念図

3．コンテキスト依存型音節単位ネットワーク

図２は今回用いたコンテキスト依存型の音節単位の音響セグメントネットワークの例で、SHIという音節を表すものである。図中の各ノードは音響セグメントを示しており、スペクトルと時間長情報を持っている。図中のKU/U.などの記号はKUという音節のU.という音響セグメントからの遷移を示している。このように、入口と出口を複数持つネットワークに音節間の接続情報を付加することにより音節間の音素変形も考慮できる様にした。

音節単位の認識の方が文節単位の認識より有利な点は、文節単位の認識の場合候補に上った文節は必ず最後までマッチングをしてしまうのに対して、音節単位の認識では文節の途中でも距離が大きければ枝刈りできるため、計算量が削減できるということである。

4．ビームサーチとbest-firstサーチ

ビームサーチはbreadth-firstサーチの一種であり、仮説が生成された順序に従って、深さ一定になるように展開する。仮説数の爆発を防ぐために、スコアの低い仮説を除外し、仮説数を一定（ビーム幅）に抑さえる。最適解が局所的に低いスコアを取ると、最適解が枝刈りされる可能性があるため、ビーム幅

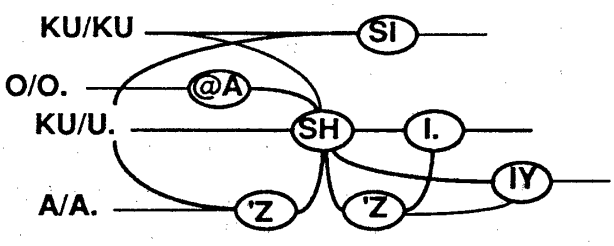

図２　音節単位ネットワーク

A Study of Search Algorithms for Japanese Sentence Recognition
Hideki KOJIMA,Hitoshi IWAMIDA, and Shinta KIMURA
Fujitsu Laboratories LTD

は充分に大きく取る必要がある。しかし、計算量はビーム幅に比例して大きくなる。

これに対しbest-firstサーチは、評価値の最も高い仮説を展開することにより探索を進める。この場合も、記憶できる仮説の数には限界があるため、仮説数を一定値（スタックの深さ）以内に抑さえることになるが、スタックの深さが増えても計算量はあまり増えないのが特長である。これは、best-firstサーチがスタックの深さに無関係に距離の小さい所だけを選んで展開を進めて行くためと考えられる。

5．距離のフレーム数による正規化

普通にDPマッチングを行ない距離を求めると長い文節ほど距離は大きくなるため、正解の文節が長い場合に枝刈りされやすくなる。そのため、文節単位のビームサーチを行なう場合には距離をフレーム数で割って正規化する必要が生ずる。

ところが、best-firstサーチの場合は、探索の枝を延ばすごとに距離が単調増加しないと最適解が得られないため正規化しない距離の方が望ましいとも考えられる。そこで、今回はbest-firstサーチに関しては正規化した距離と正規化しない距離の両方について実験を行ない比較した。

6．評価実験

6．1　音声データ

評価用の文として、テレビに関する問い合わせ文を選んだ。男性話者１名に地名184単語とテレビに関する問い合わせ文500文を発声してもらい、地名184単語を音響ﾃﾝﾌﾟﾚｰﾄの学習に用い、500文を評価に用いた。平均文節数は3.47、最大文節数は５である。

6．2　文法

約150単語から、２単語モデルにより約1300文節を生成し登録した。動詞の文節数は294個である。評価用の500文を受理できるような文法を書いた結果、５文節以内の文で、この文法によって生成できる文の数は約３０億になった。

6．3　実験結果

表１に実験結果を示す。ビーム（又はスタック）の幅は正解候補が枝刈りされないために必要充分な値を実験により求めた。意味理解率は意味的に問題のない誤りを除いた認識率である。意味的に問題のない誤りのほとんどは助詞の「は」と「が」の誤りである。計算量はSparc Station-2上で測ったcpu timeである。

文節単位のビームサーチと音節単位のビームサーチを比較すると、音節単位のネットワークの導入が認識率を落とさずに計算量を大きく削減していることが分かる。また、音節単位のビームサーチと正規化しない距離を用いた音節単位のbest-firstサーチを比較すると、best-firstサーチの導入も認識率を落とさずに計算量を大きく削減することが分かる。同じ音節単位のbest-firstサーチでも正規化距離を用いた方が、認識率はやや落ちるが、計算量は半分以下に減らせることが分かる。

表１　認識実験結果

認識　単位	文節	音節	音節	音節
探索　手法	ビーム	ビーム	ベスト	ベスト
距離正規化	あり	あり	なし	あり
ビーム　又は スタックの幅	１０	６０	６０	６０
認　識　率(%)	74.8	74.8	74.8	74.2
意味理解率(%)	99.2	99.2	99.2	98.8
計　算　量(秒)	117.15	27.73	3.58	1.25

7．考察

音節単位認識とbest-firstサーチの導入による計算量の削減は予想通りであった。正規化距離を用いたbest-firstサーチが正規化しない距離を用いたbest-firstサーチよりも認識率が悪いのは、距離を正規化することにより最適解を得る保証がなくなるためである。また、正規化距離を用いた方が速いのは、正規化しない距離を用いると、正しい結果ではないのに、まだマッチングした音節の数が少ないために距離が小さい枝を展開してしまう可能性が出てくるためと考えられる。

8．最後に

日本語文の認識において、コンテキスト依存型の音響セグメントネットワークの導入とbest-firstサーチの導入により、認識率を落とすことなく計算量を大幅に削減できることが分かった。今後はA*サーチなどの導入により更に計算量の削減を図る予定である。

参考文献

[1] 山崎，他：”音響ｾｸﾞﾒﾝﾄﾈｯﾄﾜｰｸを用いた１０万単語認識”，音響学会春季講論　1-3-24 (1990)
[2]小島，他：”深層格を用いた係り受け解析による日本語文音声の認識”，
音響学会秋季講論　2-Q-9 (1992)