

品詞ならびに漢字かな文字のマルコフ連鎖確率情報を用いた 1Q-9 文節漢字かな列候補の絞り込み法

荒木哲郎⁺ 池原悟⁺⁺ 芳永寛史⁺⁺⁺ 真田陽一⁺

⁺福井大学工学部 ⁺⁺NTT情報通信網研究所 ⁺⁺⁺松下電器産業

1. はじめに

日本語を計算機に入力する方法としては、現在区切りを一切入れないべた書き文入力方式が主流となっている。べた書きかな文の場合に、総当たり法でかな漢字変換等の処理により生成される、あらゆる単語候補列の組み合わせを考慮して解析を行うと、一般に探索木が爆発する問題が生じる。

本論文では文節境界が正しく与えられた場合に対してべた書きかな文節より、かな漢字変換によって得られる膨大な漢字かな候補列を絞り込む問題について議論する。マルコフ連鎖モデルを用いた音節認識候補列の絞り込み法[1]及び漢字かな候補絞り込み法[2]を拡張し、文節内及び文節間の漢字かな文字及び品詞情報のマルコフ連鎖モデルを用いる方法を提案し、実験を行う。

2. マルコフ連鎖モデルによる漢字かな文節候補の絞り込み方法

文節内と文節間の漢字かな文字及び品詞についてのマルコフモデルの種類により、以下のような絞り込み方法を定義する。

(方法1) 漢字かな文節候補の集合を文節内漢字かなコスト値によって絞り込む方法をIK法とし、更にIK法に文節内の品詞コスト値による足切りを

加えた方法をIKH法と呼ぶ。

(方法2) 文節内及び前後文節間の漢字かなコスト値を組み合わせた方法によって絞り込む場合、両者のコストを加算した和が最小となるもので評価する方法をADD-IK-FBK法とし、更に文節間漢字かなコスト値で足切りを行い評価する方法をCHK-IK-FBK法と呼ぶ。

(方法3) 方法2の文節間漢字かなコスト値の代わりに文節間品詞コスト値を用いる場合をそれぞれADD-IK-FBH法、CHK-IK-FBH法と呼ぶ。

3. 実験条件

(1) 入力データ

①文の種類：新聞記事

②字種：べた書き音節文節

③総文章数(標本内データ)：218-1000文節

総文章数(標本外データ)：263-1000文節

(2) 使用辞書

①漢字かな候補の生成：40万語の単語辞書

②漢字かな候補の絞り込み：漢字かなマルコフ連鎖確率情報、品詞マルコフ連鎖確率情報(いずれも文節内と文節間の2重マルコフ連鎖確率)

4. 実験結果

2. で示した絞り込み方法による実験結果を示す。

(1) 方法1による効果

IK法及びIKH法の効果を図1に示す。特にIKH法はIK法より1%程度優れ、IK法の10位以内候補を約4割程度削減し、同じ正解率を6位以内で実現出来ることを意味している。

(2) 方法2及び方法3による効果

A Method of Deciding "Bunsetu" Candidates of "kan-jikana" Strings Using Markov Models

Tetsuo Araki⁺ Satoru Ikehara⁺⁺

Hiroshi Yoshinaga⁺⁺⁺ Youichi Sanada⁺

⁺Faculty of Engineering Fukui University

⁺⁺NTT Network Information Systems Laboratories

⁺⁺⁺Matsushita Electric Industrial Company Limited

文節間情報を用いる方法2のADD-IK-FBK法、CHK-IK-FBK法の絞り込みの結果を図2、3に示す。同図より、加算タイプの組み合わせ方法が、足切り方法よりも有効であることがわかり、第一位正解率がIK法よりも約2~3%向上し、83.7%（標本外データ）~98.2%（標本内データ）に達することがわかった。

また方法3は、漢字かな文字のマルコフを用いる方法2に比べて効果が小さいことがわかった。

5. おわりに

本論文では文節内及び文節間の漢字かな文字及び品詞情報のマルコフ連鎖モデルを用いる方法を提案しその有効性を定量的に把握した。その結果、文節内ではIKH法が有効であり、標本外データの場合10位内候補の内約4割を削減する効果があった。また文節間ではADD-IK-FBK法により最終的に第一位正解率が、83.7%~98.2%になることがわかった。

今後の課題としては構文解析情報及び意味解析情報を用いた文候補絞り込みの研究があげられる。

参考文献

- [1] 荒木、村上、池原：“2重音節マルコフモデルによる日本語の文節音節認識候補の曖昧さの解消効果”，情報処理, Vol. 30, 4, pp467-477(1989)
- [2] 村上、荒木、池原：“日本語文音節入力に対して2重マルコフ連鎖モデルを用いた漢字かな交じり候補の抽出精度”，信学論, Vol. J75-DII, pp 11-20(1992)

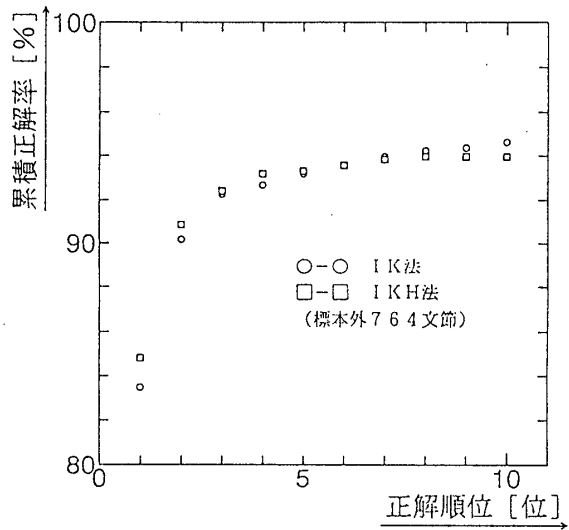


図1 IKH法による効果

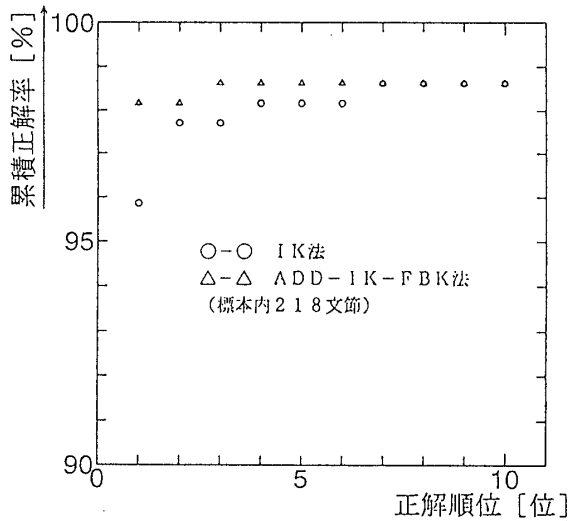


図2 ADD-IK-FBK法による効果

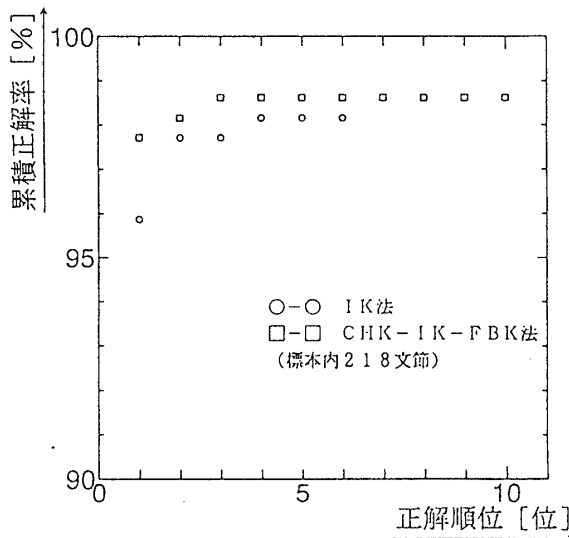


図3 CHK-IK-FBK法による効果