

単語マルコフモデルを組み合わせた仮文節境界の補正法

2H-6

荒木 哲郎<sup>+</sup> 池原 悟<sup>++</sup> 笹島 伸一<sup>+</sup>

<sup>+</sup> 福井大学    <sup>++</sup> NTTコミュニケーション科学研究所

1. はじめに

べた書きの漢字かな混じり文の文節切りにおいては、これまでに高精度な技術が確立されている[1]。しかし、べた書きのかな文の場合には、あらゆる単語候補の組み合わせを考慮して解析しようとする、一般に探索木が爆発するという問題が生じ、これまでに有効な方法が得られていない。

このような問題を解決するために、音節の2重および3重マルコフ連鎖モデルを用いて仮文節境界を推定する方法が提案されている[2]。また、上の方法で推定された仮文節境界に対して、単語辞書引きや品詞接続テーブルを用いた接続検定を行うことによって、仮文節境界を補正する方法も提案され、有効性が示されている[3]。

しかしながら、単語辞書引きや品詞接続テーブルを用いた仮文節境界の補正での処理時間は、非常に長く、現在のマシンでは実用性に欠ける。

本文では、従来の仮文節境界の補正の結果を損わずに、処理時間を軽減するために、単語マルコフ連鎖モデルを用いて仮単語境界を推定することによって、仮文節境界を補正する方法を提案する。

2. 音節文における仮文節境界の補正方法

本文は、仮単語境界を推定することによって仮文節境界を補正するものである。仮単語境界の推定方法を次に述べる。

【定義】 音節2重単語マルコフ連鎖確率が、足切り値Tに対して、次の条件を満たすとき位置iが仮単語境界として推定される。

$$P(S_i | S_{i-2} S_{i-1}) < T$$

次に仮文節境界の補正方法を述べる。図1に補正の考え方を示す。

【方法1】品詞間の接続性による補正

仮文節境界を起点とし、前文節の末尾単語候補と、後文節の先頭単語候補を仮単語境界によって推定し、それぞれの単語候補について辞書引きを行う。それぞれ単語候補の品詞の相互接続性を、品詞接続テーブルによる接続検定を行うことによって、仮文節境界の補正をする。

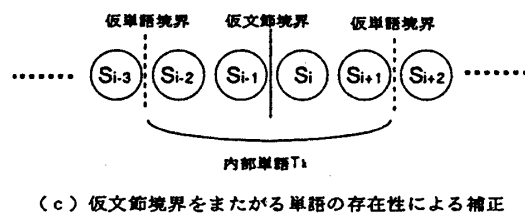
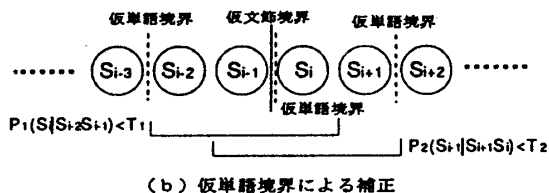
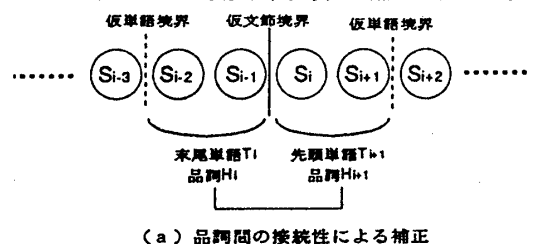


図1 仮文節境界補正の考え方

A Method of Correcting The Provisional Boundaries of "Bunsetsu" Using Word Markov Model

Tetsuo ARAKI<sup>+</sup> Satoru IKEHARA<sup>++</sup>  
Shinichi SASAJIMA<sup>+</sup>  
<sup>+</sup> Fukui University

<sup>++</sup> NTT Communication Science Laboratories

〔方法2〕仮単語境界による補正

仮文節境界において、仮単語境界の条件を満たすか否かによって仮文節境界の補正をする。

〔方法3〕仮文節境界をまたがる単語の存在性による補正

仮文節境界を起点とし、その仮文節境界をまたがる単語候補を、仮単語境界によって推定し、その単語候補について辞書引きを行う。このとき漢字かな候補が存在するか否かを検定することによって、仮文節境界の補正をする。

3. 実験条件

(1) 入力データ

文の種類：新聞記事

字種：べた書き音節文

総文章数：200文(1458文節, 7272文字)

(2) 使用辞書

単語辞書(15万語)

音節の2重単語マルコフ連鎖確率辞書

品詞接続テーブル

4. 仮文節境界補正の実験結果

(1) 処理時間軽減の効果

図2に従来の方法と今回の方法による処理時間軽減の効果を示す。従来の方法に要する処理時間を1としたとき、最高で60倍高速に処理できることが分かった。

(2) 仮文節境界補正の効果

図3、4に仮文節境界を補正した結果を示す。図より〔方法3〕では補正前より適合率が最高15%向上する結果が得られた。

5. 終わりに

単語マルコフモデルを用いて仮単語境界を推定したのち仮文節境界を補正する方法を提案した。実験により有効性を評価した結果、従来の方法に比べて処理速度が約60倍向上するとともに、補正効果が再現率をあまり低下させずに〔方法2〕で適合率を9%向上させ、〔方法3〕で適合率を15%向上させる結果が得られた。

今後の課題としては、より正確に仮文節境界を補正する方法の研究、仮単語境界を推定する単語マルコフモデルの研究などが挙げられる。

参考文献

[1]宮崎、大山 “日本語音声入力のための言語処理” 情処, Vol. 27, No. 11, pp1053-1061(1986)  
 [2]荒木、池原、土橋、笹島 “3重マルコフモデルによるべた書きかな文の仮文節境界の推定法” 情処NL研究会, Vol. 102-14, pp105-112(1994)  
 [3]荒木、池原、土橋 “べた書きかな文の仮文節境界の補正方法” 情処NL研究会, Vol. 98-1, pp1-7(1993)

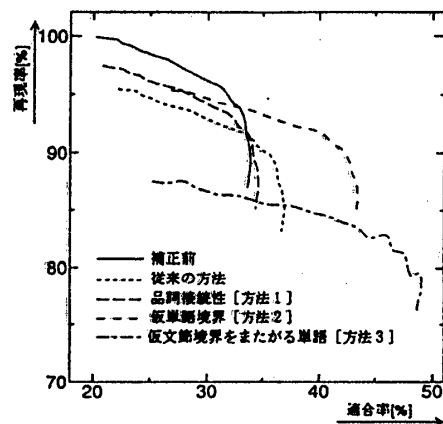
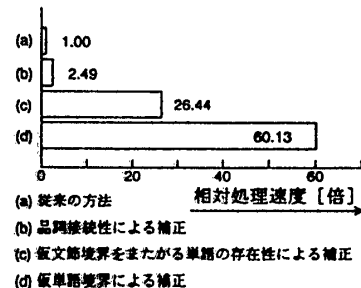


図3 仮文節境界の補正結果 (LM法を適用した場合)

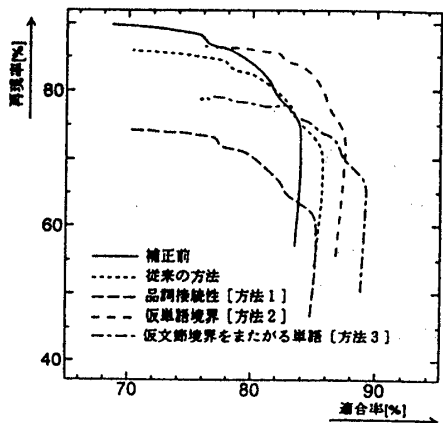


図4 仮文節境界の補正結果 (2BLM法を適用した場合)