

2重マルコフモデルの全域法と局所法による日本語の誤字訂正効果

荒木 哲郎⁺ 池原 悟⁺⁺ 土橋 潤也⁺ 堂元 一頼⁺
 +福井大学 ++NTT情報通信網研究所

Abstract

2重マルコフモデルを用いた日本語誤字の誤り訂正法を、マルコフモデルを適用する位置に着目して分類すると、検出された誤り音節または、漢字かな文字に限定してマルコフモデルを一回適用する方法(局所法)と、誤り音節または漢字かな文字を含む文字列全体に適用する方法(全域法)があり、前者は後者に比べて少ない処理時間で行える特徴がある。これまでに局所法については、文節並びに単語境界で検出された音節、漢字かな文字の誤りを訂正する問題に対して、誤り位置に応じて順方向、中間、逆方向タイプの2重マルコフモデルを適用する方法の有効性が報告されている[9]。

本論文では、全域法の誤り訂正能力を定量的に評価し、局所法との比較を行う。全域法においては、三つのマルコフ連鎖確率がそれぞれ中心的な役割を果たすことに着目して、これらのマルコフ連鎖確率を単独に用いて誤り訂正を行う三つのタイプの局所法について、文節境界における誤り文字の位置(文節の先頭誤り、2番目及び3番目以降)に関する誤り訂正の評価を行い、全域法と局所法の比較を行う。

新聞記事77日分の統計データを用いて、順方向、逆方向並びに中間タイプの2重マルコフ連鎖確率を求め、日本語音節文及び漢字かな交じり文の2000箇所誤りに対して、局所法と全域法による誤り訂正実験を行った。その結果、

- (1) 全域法では、順方向タイプの2重マルコフモデルが、常に他のタイプよりも優っていること、また
 - (2) 全域法は局所法よりも優っていること
- などがわかった。

Effect of Error Correction Using the Global and Local Application of 2nd-order Markov Models

Tetsuo Araki + Satoru Ikehara ++ Junya Tsuchihashi + Kazuyori Doumoto +
 +Fukui University ++NTT Network Information Systems Laboratories

This paper investigates two methods to correct erroneous syllables and kanji-kana characters located at the boundaries of "bunsetsu" using three types of 2nd-order Markov model, called as forward, backward and middle type respectively.

One is called the local method which is to correct erroneous characters using Markov model only once. The other is the global method which is to correct erroneous syllables and kanji-kana characters by applying Markov model to all the string of syllables and of kanji-kana characters. According to the experiment using 70 issues of a daily Japanese newspaper, the following results was obtained.

1. In the case of global method, the method to correct erroneous syllables and kanji-kana characters using Markov model of forward type is always superior to that of the other types.
2. The capability of error correction using the global method is always superior to that of the local method.

1. はじめに

マルコフモデルはニューラルネットやファジーモデルと同様に、機械学習モデルの一つと考えられ、自然言語の文字系列 [1],[2] や連続音声認識の音韻系列 [6],[7] に応用されている。

これまでに 2 重マルコフモデルが、漢字かな交じり文節の誤り訂正 [9]、べた書きかな文から漢字かな交じり文への変換における曖昧性の解消 [11]、及び日本語の音節文節候補から正しい音節を選択する問題 [10] に対して有効であることが示されている。

しかしこのモデルを、文の文節並びに単語境界にある誤り音節の検出や訂正に適用しようとすると、必ずしも有効に働かないと言う問題が生じるため、[9] では順方向、逆方向、中間の 3 タイプの 2 重マルコフモデルを用いて、文節並びに単語の境界にある誤り音節を訂正する方法が提案されている。

本論文では、検出された誤りの音節並びに漢字かな文字に対して、三つの 2 重マルコフ連鎖モデルを一回だけ適用して訂正する方法 (局所法) と、誤りの音節や漢字かな文字を含む連続した音節鎖及び漢字かな鎖全体に 2 重マルコフモデルを適用する方法 (全域法) を取り上げ、相互の誤り訂正能力を評価する。特に、局所法と全域法の関係を詳細に調べるために、誤りの音節または漢字かな文字の位置に応じて、三つの基本的な音節鎖または漢字かな鎖を定義し、それぞれの鎖に対する 2 重マルコフ連鎖確率を用いた誤り訂正能力を、文節境界における誤り位置に応じて評価する。

さらに全域法は、基本的な三つの鎖に対する 2 重マルコフ連鎖確率によって決定されるから、局所法に比べて 3 倍の時間かかり、特に漢字かな文字の誤り訂正の場合には、誤り訂正候補数が多く辞書アクセス時間に要する時間が増大する。このような問題に対処するための方法として、ここでは全域法により得られる誤り訂正候補を局所法によってどこまで得られるかを実験により調べる。

最後に、順方向、逆方向、中間タイプと呼ばれる三つの 2 重マルコフ連鎖確率の間に成り立つ関係式を導き、誤り訂正実験によりその関係を確認した。

7 7 日分の新聞記事の統計データを用いた実験により、局所法と全域法の評価、全域法における三つの 2 重マルコフモデルの評価を調べて、相互の関係を明らかにする。

2. 諸定義

日本語の文は、一つの自立語と一つ以上の付属語から成る文節と呼ばれる単位に分割される。すなわち、文節 = (自立語)(付属語)ⁿ で、自立語 = {名詞、動詞、形容詞、形容動詞、副詞、接続詞、連体詞、感動詞、形式名詞}、また付属語 = {助詞、助動詞、接辞} である。一般に日本語の文節として意味のある文節に構成されるためには、文節に現れる音節の列がランダムに構成されるのではなく、ある一つの拘束性に従って規定されると考えられる。(例えば、撥音「ン」は文節の先頭にはこないとか、「ア」の

次に「ル」がくる確率が高いなど)。このような音節列は m 重マルコフ連鎖によって表され、そのエントロピーはシャノンが与えたエントロピーを用いて同様に定義可能である。ここでは、べた書きの日本語の音節文並びに漢字かな交じり文に対して、音節並びに漢字かな文字の 2 重マルコフ連鎖確率を、次のように定義する。

[定義 1] 順方向、逆方向、中間タイプの 2 重マルコフ連鎖確率を次のように定める。

- (1) 順方向マルコフ確率の集合 $P_{2f} = \{P_{i-2,i-1}(i) \mid P_{i-2,i-1}(i) = \frac{P(x_{i-2}x_{i-1}x_i)}{P(x_{i-2}x_{i-1})}\}$
- (2) 逆方向マルコフ確率の集合 $P_{2b} = \{P_{i+2,i+1}(i) \mid P_{i+2,i+1}(i) = \frac{P(x_{i+1}x_{i+2})}{P(x_{i+1})}\}$
- (3) 中間マルコフ確率の集合 $P_{2m} = \{P_{i-1,i+1}(i) \mid P_{i-1,i+1}(i) = \frac{P(x_{i-1}x_ix_{i+1})}{P(x_{i-1}x_{i+1})}\}$

ここで $P_{i-2,i-1}(i)$ は条件付き確率を表す。順方向、逆方向、中間タイプの 2 重マルコフ連鎖確率モデルを図 1 に示す。

[定義 2] 順方向、逆方向、中間タイプの 2 重マルコフ連鎖確率のエントロピーが次のように定義される。

- (1) 順方向: $H_{2f} = \sum P(i, j, k) \log_2 P_{i,j}(k)$
- (2) 逆方向: $H_{2b} = \sum P(i, j, k) \log_2 P_{k,j}(i)$
- (3) 中間: $H_{2m} = \sum P(i, j, k) \log_2 P_{i,k}(j)$

新聞記事 7 7 日分を用いて、0 重、1 重、と三つの 2 重音節マルコフ連鎖確率のエントロピー (簡単に $H_0, H_1, H_{2f}, H_{2b}, H_{2m}$ と表す) が表 1 のように得られる。本実験では、清音 (ア、サなど総計 4 4)、濁音 (ガ、ザなど総計 1 8)、半濁音 (バ、ピなど総計 5)、拗音 (キャ、ジャなど総計 3 3)、撥音 (ン) 及び空白文字からなる 1 0 2 音節を用いる。

音節の全ての組み合わせの中で、マルコフ連鎖確率の値が零でない組み合わせの比率は、1 重マルコフでは全体 102^2 の約 5 7%、また 2 重マルコフでは 102^3 の約 5% であり、2 重マルコフの場合はほとんどの音節組み合わせが使用されていない。

2 重マルコフ連鎖確率のデータは、日本文音声出力の研究において実現された高精度な言語解析技術 [8] を用いて、新聞記事の漢字かな交じり文を音節文に変換する事によって得られたものを使用した。図 2 に音節文節の例を示す。

表 1 0 重、1 重並びに 2 重音節マルコフ連鎖確率のエントロピー

order	equation	entropy
H_0	$\sum P(i) \log_2 P(i)$	5.574
H_1	$\sum P(i, j) \log_2 aP_i(j)$	4.427
H_{2f}	$\sum P(i, j, k) \log_2 P_{i,j}(k)$	3.761
H_{2b}	$\sum P(i, j, k) \log_2 P_{k,j}(i)$	3.761
H_{2m}	$\sum P(i, j, k) \log_2 P_{i,k}(j)$	3.168

表1に従って、次の関係が認められる。

[基本式1] 三つのタイプのマルコフモデルのエントロピー間に次の関係が成り立つ。

$$H_{2f} = H_{2b} \quad \text{及び} \quad H_{2f} > H_{2m}$$

この関係は、マルコフ連鎖確率 P_{2f} , P_{2b} , 及び P_{2m} が音節 x_i に適用される位置によって異なることを意味している。またマルコフ連鎖確率のエントロピーが小さくなればなるほど、マルコフ連鎖確率による誤り訂正がより効果的になることが期待される。

[基本式2] 三つの2重マルコフ連鎖確率の間に次の関係が成り立つ。

(1) 中間タイプと順方向タイプの関係

$$P_{i-2,i}(i-1) = \frac{P_{i-2,i-1}(i)P_{i-2}(i-1)}{P_{i-2}(i)}$$

(2) 逆方向タイプと順方向タイプの関係

$$\prod_{i=1}^{n+2} P_{i,i-1}(i-2) = \prod_{i=1}^{n+2} P_{i-2,i-1}(i)$$

但し、 $i < 0$ 及び $i > n$ のときは、 x_i は空白文字とする。

(略証) 簡単のために $x_1x_2x_3$ に対する順方向、逆方向、中間タイプの2重マルコフ連鎖確率を、それぞれ次のように表す。

$$P(x_3|x_1x_2) = \frac{P(x_1x_2x_3)}{P(x_1x_2)}$$

$$P(x_1|x_3x_2) = \frac{P(x_1x_2x_3)}{P(x_3x_2)}$$

$$P(x_2|x_1x_3) = \frac{P(x_1x_2x_3)}{P(x_1x_3)}$$

これより、 $P(x_1x_2x_3)$ が順方向タイプの確率 $P(x_3|x_1x_2)$ を用いて次のように表される。

$$P(x_1x_2x_3) = P(x_3|x_1x_2)P(x_1x_2)$$

これを用いて、中間タイプの確率 $P(x_2|x_1x_3)$ は

$$\text{次のように表される。} \quad P(x_2|x_1x_3) = \frac{P(x_1x_2x_3)}{P(x_1x_3)}$$

$$= \frac{P(x_3|x_1x_2)P(x_1x_2)}{P(x_1x_3)}$$

$$= \frac{P(x_3|x_1x_2)P(x_2|x_1)P(x_1)}{P(x_3|x_1)P(x_1)} = \frac{P(x_3|x_1x_2)P(x_2|x_1)}{P(x_3|x_1)}$$

3. 局所法並びに全域法による誤り訂正法

ここでは音節文及び漢字かな交じり文において、誤りの音節及び漢字かな文字が検出されたとき、2章で述べた三つのタイプの2重マルコフモデルを用いて誤り訂正を行う局所法と全域法について述べる。

最初に、上記関係で示された三つのエントロピーの間の関係に従って、誤り訂正能力を評価する為に三つの誤りケースを定義する。

[定義3] 音節列で検出された誤り音節 x_i の位置に従って、三つの誤りのケースを定義する。

1. ケース S_1 : 誤り音節並びに漢字かな文字 x_i が文節の先頭位置にあるケース。
2. ケース S_2 : 誤り音節並びに漢字かな文字 x_i が文節の先頭から2番目に位置するケース。

3. ケース S_3 : 誤り音節 x_i が上記の(1)及び(2)以外に位置するケース。

[定義4] x_i を誤りの音節または誤りの漢字かな文字とすると、 x_i を含む次のような三つの音節連鎖または漢字かな連鎖を定義し、誤り x_i が音節連鎖または漢字かな連鎖のどこに位置するかによって、それぞれ前方連鎖、中間連鎖、及び後方連鎖と呼ぶ。

- (1) 前方連鎖: $x_{i-2}x_{i-1}x_i$
- (2) 中間連鎖: $x_{i-1}x_i x_{i+1}$
- (3) 後方連鎖: $x_i x_{i+1} x_{i+2}$

各々の連鎖に対して、図3に示すような順方向、逆方向、中間タイプの2重マルコフ連鎖確率が定義される。

[局所法による誤り訂正方法] C を音節または漢字かな文字の集合とすると、次の手順により、 P_{2f} , P_{2b} 及び P_{2m} に対して、集合 C の要素 \bar{x} を決定する。

求めた \bar{x} を、誤り音節または漢字かな文字と置き換える。

1. $\exists x_{i-2}, x_{i-1} \in C, \exists P_{i-2,i-1}(i) \in P_{2f}$ に対して、 $P_{i-2,i-1}(\bar{x})$ が最大となるような \bar{x} を求める。
2. $\exists x_{i+2}, x_{i+1} \in C, \exists P_{i+2,i+1}(i) \in P_{2b}$ に対して、 $P_{i+2,i+1}(\bar{x})$ が最大となるような \bar{x} を求める。
3. $\exists x_{i-1}, x_{i+1} \in C, \exists P_{i-1,i+1}(i) \in P_{2m}$ に対して、 $P_{i-1,i+1}(\bar{x})$ が最大となるような \bar{x} を求める。

[全域法による誤り訂正方法] C を音節または漢字かな文字の集合とすると、次の手順により、 P_{2f} , P_{2b} 及び P_{2m} に対して、集合 C の要素 \bar{x} を決定する。

求めた \bar{x} を、誤り音節または漢字かな文字と置き換える。

1. 前方連鎖 $x_{i-2}x_{i-1}x_i$ 、中間連鎖 $x_{i-1}x_i x_{i+1}$ 、後方連鎖 $x_i x_{i+1} x_{i+2}$ において、それぞれ $\exists x_{i-2}, x_{i-1} \in C, \exists P_{i-2,i-1}(\bar{x}) \in P_{2f}$, $\exists x_{i-1}, x_{i+1} \in C, \exists P_{i-1,i+1}(\bar{x}) \in P_{2f}$, $\exists x_{i+1}, x_{i+2} \in C, \exists P_{i,i+1}(\bar{x}) \in P_{2f}$ で、 $P_{i-2,i-1}(\bar{x}) \cdot P_{i-1,i+1}(\bar{x}) \cdot P_{i,i+1}(\bar{x})$ が最大となるような \bar{x} を求める。
2. 前方連鎖 $x_{i-2}x_{i-1}x_i$ 、中間連鎖 $x_{i-1}x_i x_{i+1}$ 、後方連鎖 $x_i x_{i+1} x_{i+2}$ において、それぞれ $\exists x_{i-2}, x_{i-1} \in C, \exists P_{i-2,\bar{x}}(i-1) \in P_{2b}$, $\exists x_{i-1}, x_{i+1} \in C, \exists P_{i-1,i+1}(\bar{x}) \in P_{2b}$, $\exists x_{i+1}, x_{i+2} \in C, \exists P_{\bar{x},i+2}(i+1) \in P_{2b}$ で、 $P_{i-2,\bar{x}}(i-1) \cdot P_{i-1,i+1}(\bar{x}) \cdot P_{\bar{x},i+2}(i+1)$ が最大となるような \bar{x} を求める。

3. 前方連鎖 $x_{i-2}x_{i-1}x_i$ 、中間連鎖 $x_{i-1}x_ix_{i+1}$ 、後方連鎖 $x_ix_{i+1}x_{i+2}$ において、それぞれ $\exists x_{i-2}, x_{i-1} \in C, \exists P_{\bar{x}, i-1}(i+2) \in P_{2m}, \exists x_{i-1}, x_{i+1} \in C, \exists P_{i+1, \bar{x}}(i-1) \in P_{2m}, \exists x_{i+1}, x_{i+2} \in C, \exists P_{i+2, i+1}(\bar{x}) \in P_{2m}$ で、 $P_{\bar{x}, i-1}(i+2) \cdot P_{i+1, \bar{x}}(i-1) \cdot P_{i+2, i+1}(\bar{x})$ が最大となるような \bar{x} を求める。 ■

図3に各連鎖に対する全域法と局所法の違いを示す。

4. 実験結果

4.1 実験条件

1. マルコフモデルの統計データに用いた日本語文のタイプ: 新聞記事
2. 誤りの音節数並びに漢字かな文字の総数: それぞれ2000箇所
3. 誤りの音節文と漢字かな交じり文の種類: 標本外データ(マルコフの統計データに用いた日本語文とは異なった文として、ミンスキーの本「心の社会」の中の日本語文から選ぶ)
4. 2重マルコフタイプの種類: べた書き文の音節並びに漢字かな文字の順方向、逆方向、中間タイプの2重マルコフ連鎖確率
5. 2重マルコフモデルによる訂正法: 全域法(順方向、中間、逆方向タイプ)と局所法(順方向、中間、逆方向タイプ)

4.2. 実験結果

[1] 全域法と局所法による誤り訂正結果

音節文における誤りの音節の位置が、文節先頭から1番目、2番目及び3番目であるとき、前方鎖、中間鎖及び後方鎖の三つに対する2重マルコフモデルを各々独立に用いた局所法と、それらを組み合わせた全域法による誤り訂正の結果をそれぞれ図4～図6に示す。また漢字かな交じり文における誤りの漢字かな文字の位置が、文節先頭から1番目、2番目及び3番目であるとき、局所法と全域法による誤り訂正の結果をそれぞれ図7～図9に示す。これらの実験結果から以下のような結果を得る。

1. 誤り位置が文節先頭から1番目の場合:
 - (A) 音節文の場合:
 - (a) 第一位正解率: 全域法 = 33-40% 及び局所法 = 9-32%
 - (b) 10位内累積正解率: 全域法 = 76-82% 及び局所法 = 47-79%
 - (B) 漢字かな交じり文の場合:

- (a) 第一位正解率: 全域法 = 42-44% 及び局所法 = 9-42%
- (b) 10位内累積正解率: 全域法 = 71-75% 及び局所法 = 29-73%

2. (2) 誤り位置が文節先頭から2番目の場合:
 - (A) 音節文の場合:

- (a) 第一位正解率: 全域法 = 50-56% 及び局所法 = 24-47%
- (b) 10位内累積正解率: 全域法 = 89-90% 及び局所法 = 75-86%

- (B) 漢字かな交じり文の場合:

- (a) 第一位正解率: 全域法 = 61-63% 及び局所法 = 15-56%
- (b) 10位内累積正解率: 全域法 = 83-85% 及び局所法 = 50-81%

3. (3) 誤り位置が文節先頭から3番目の場合:
 - (A) 音節文の場合:

- (a) 第一位正解率: 全域法 = 45-54% 及び局所法 = 21-34%
- (b) 10位内累積正解率: 全域法 = 86-88% 及び局所法 = 67-80%

- (B) 漢字かな交じり文の場合:

- (a) 第一位正解率: 全域法 = 52-57% 及び局所法 = 33-38%
- (b) 10位内累積正解率: 全域法 = 82-83% 及び局所法 = 66-76%

以上より、次のことがわかる。

1. 順方向、中間、逆方向タイプによる全域法では、誤り位置に無関係に順方向タイプによる誤り訂正が、他のタイプのものよりも優れている。
2. 全域法と局所法を比べると、全般的に全域法が局所法よりも優れていることがわかる。その差は、誤りの位置が文節先頭から1番目及び2番目の所では小さく、3番目以降では7%から16%とやや大きくなっている。
3. 全域法と局所法の差は、第一位正解率では5-20%と大きい、10位内累積正解率では10%前後と比較的小さくなっている。

更に、標本内データ(マルコフ統計データに用いた新聞記事より選ぶ)のときは、10-20%程度正解率が向上する。

[2] 前方鎖、中間鎖及び後方鎖における局所法の比較

図4～図6の標本外の音節文における誤り訂正結果から、前方鎖、中間鎖及び後方鎖における局所法の優劣を比較すると、以下ようになる。

1. 誤り位置が文節先頭から1番目のとき: 後方連鎖の逆方向タイプ > 順方向、中間タイプ
2. 誤り位置が文節先頭から2番目のとき: 中間連鎖の中間タイプ > 順方向、逆方向タイプ
3. 誤り位置が文節先頭から3番目のとき: 前方連鎖の順方向 > 中間、逆方向タイプ

図7~図9の標本外の漢字かな交じり文でも、音節文の局所法の関係と同じになることがわかる。

[3] 局所法による全域法の10位内訂正候補の選出能力

全域法は局所法に比べると、誤り訂正能力は高い反面、漢字かな交じり文のように誤り訂正候補の漢字が多くなると、辞書アクセス時間が大きくなるため実用性の面で問題となる。これを解決する一つの方法として、局所法を誤り訂正候補の選出に用いる方法が考えられる。ここではその目安を与えるために、全域法によって得られる10位内誤り訂正候補を、局所法によって選出しようとする際、どのくらいに範囲まで求めればよいかを調べた。図10~図11に実験結果を示す。これらの結果より、20位から30位間で、ほぼ十分な候補が得られることがわかった。

[4] 中間タイプと順方向タイプの相互変換による等価性

2章で示した順方向、中間、及び逆方向タイプの2重マルコフ連鎖確率の相互変換式の妥当性を調べるために、相互の変換式による結果2章で示した順方向、中間、及び逆方向タイプの2重マルコフ連鎖確率の相互変換式の妥当性を調べるために、中間タイプと順方向による誤り訂正実験を行い、図12に示すように変換式による誤り訂正の精度が相互に一致することを確認した。

5. おわりに

本論文では、順方向、逆方向及び中間タイプの2重マルコフモデルを用いた、全域法と局所法による誤り訂正方法の効果を定量的に評価し、その比較を行った。その結果、(1) 全域法では順方向タイプの2重マルコフモデルが、常に他のタイプよりも優れていること、(2) 全域法が局所法よりも優れていること、(3) 局所法で、10位から30位内の訂正候補を取ると、その中に全域法の10位内候補がほとんど現れることがわかった。このことは漢字かな交じり文の場合に、辞書アクセスに時間のかかる全域法の誤り訂正の代わりに、局所法を用いることが可能であることを示している。

[参考文献]

1. C.E.Shannon: "Mathematical Theory of Communication", Bell System Technical Journal, Vol.27, pp379-423, 623-656, October (1948)

2. C.E.Shannon: "Prediction and Entropy of Printed English", Bell System Technical Journal, Vol.30, pp50-64, January (1951)
3. F.Jelinek: "Continuous Speech Recognition by Statistical Methods", Proc. of the IEEE, Vol.64, No.4, pp532-556 (1976)
4. L.R.Rabiner, S.E.Levinson and M.M. Sondai: "On the Application of Vector Quantization and Hidden Markov Models to Speaker-independent, Isolated Word Recognition", Bell System Technical Journal, Vol.62, No.4, pp1075-1105 (1983)
5. 宮崎、大山: "日本文音声出力のための言語処理方式", 情報処理学会論文誌, Vol.27, No.11, pp1053-1061 (1986)
6. 池原、白井: "単語解析プログラムによる日本文誤字の自動検出と二次マルコフモデルによる訂正候補の抽出", 情報処理学会論文誌, Vol.25, No.2, pp298-305 (1984)
7. 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", 情報処理学会論文誌, Vol.30, No.4, pp467-477 (1989)
8. 村上、荒木、池原: "日本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仮名交じり文節候補の抽出精度", 電子情報通信学会論文誌, Vol.J75-D-II, No.1, pp11-20 (1992)
9. 荒木、池原、土橋、堂元: "順方向、逆方向及び中間タイプの2重マルコフ連鎖モデルによる日本語の誤字訂正候補選択の効果", 人工知能学会全国大会, 19-4 (1993)

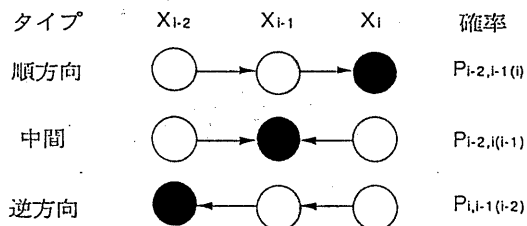


図1 3タイプの2重マルコフモデル

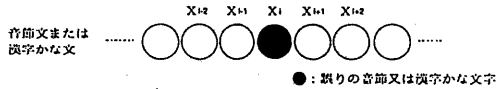
文節切りされた漢字かな交じり文：

わが国 経済の 国際化に 伴い、日本に 進出する
外国銀行は 急増している

文節切りされた音節文：

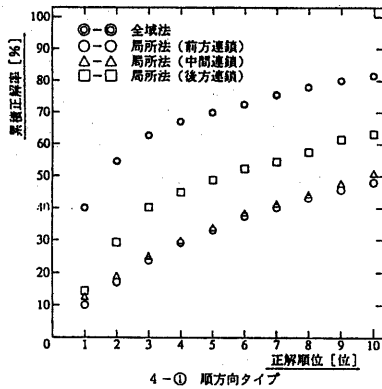
ワガクニ ケイザイノ コクサイカニ トモナイ
ニホンニ シンシュツスル ガイコクギンコウウ
キュウゾウシテイル

図2 文節切りされた漢字かな交じり文及び音節文の例

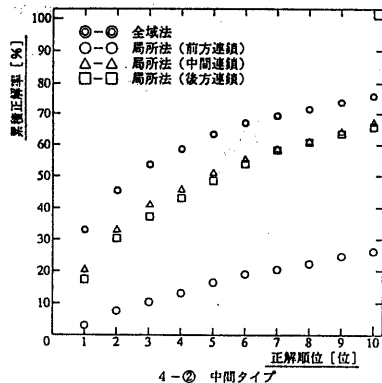


2 取マルコフタイプ		順方向	中間	逆方向
局所法	前方鎖 (X ₁₊₂ X ₁₊₁ X ₁)	(1) P _{1+2,1} (0) 順方向による前方鎖 における局所法	(4) P ₁₊₂ (1-1)	(7) P ₁₊₁ (1-2)
	中間鎖 (X ₁₊₁ X ₁ X ₁₊₁)	(2) P _{1+1,1} (+1)	(5) P _{1+1,1} (0)	(8) P _{1+1,1} (1-1)
	後方鎖 (X ₁ X ₁₊₁ X ₁₊₂)	(3) P _{1+1,1} (+2)	(6) P ₁₊₂ (1+1)	(9) P _{1+2,1} (0)
全域法		(1)×(2)×(3) 順方向による全域法	(4)×(5)×(6)	(7)×(8)×(9)

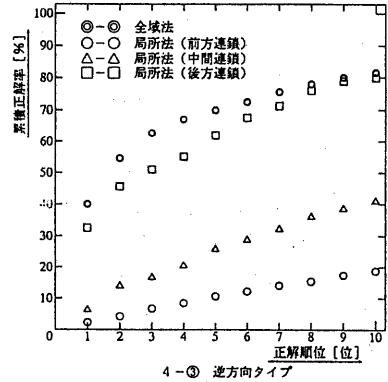
図3 局所法と全域法



4-① 順方向タイプ

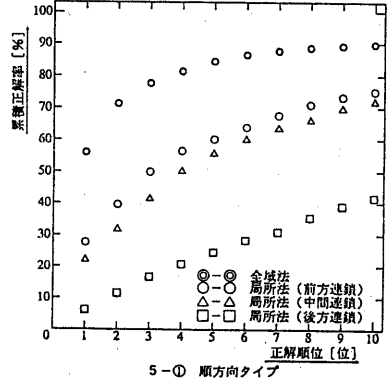


4-② 中間タイプ

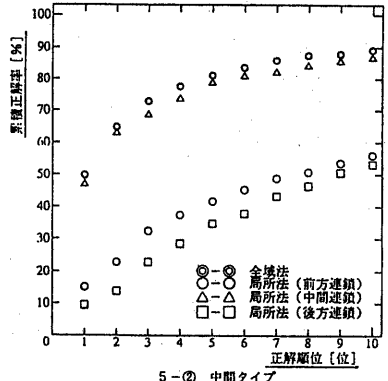


4-③ 逆方向タイプ

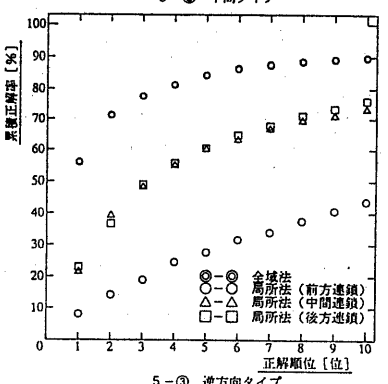
図4 音節文の文節先頭1文字目における誤り訂正結果(標本外)



5-① 順方向タイプ

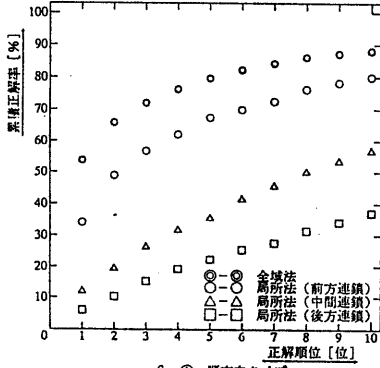


5-② 中間タイプ

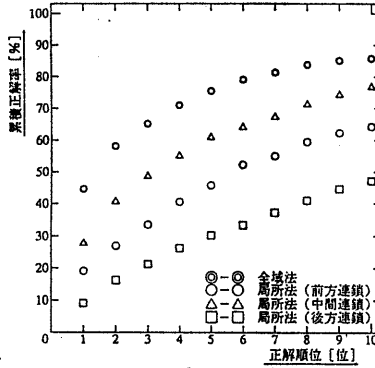


5-③ 逆方向タイプ

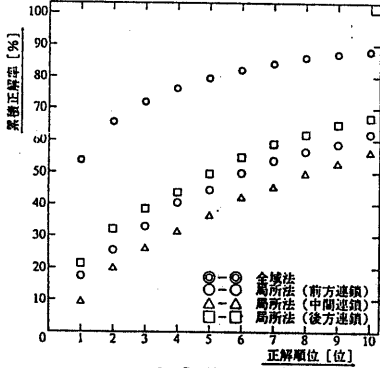
図5 音節文の文節先頭2文字目における誤り訂正結果(標本外)



6-① 順方向タイプ

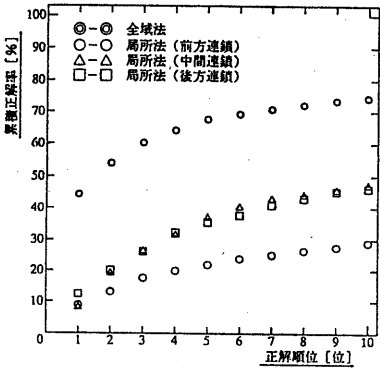


6-② 中間タイプ

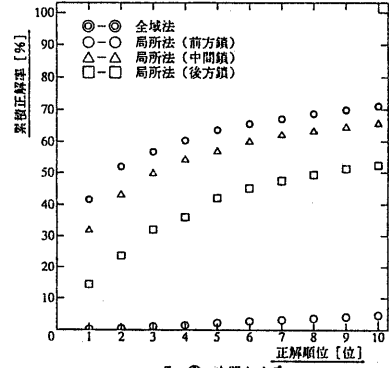


6-③ 逆方向タイプ

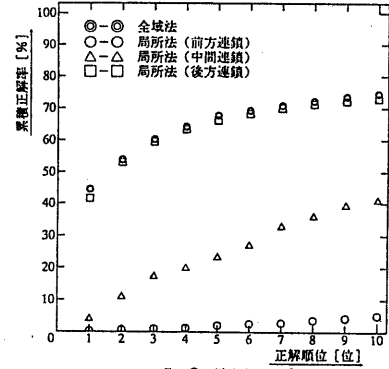
図6 音節文の文節先頭3文字目における誤り訂正結果(標本外)



7-① 順方向タイプ

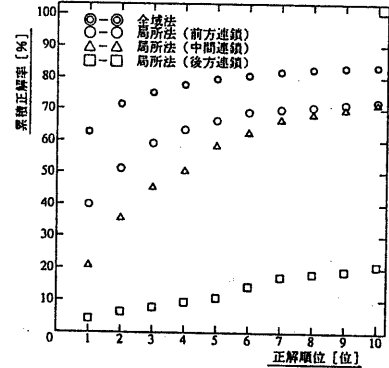


7-② 中間タイプ

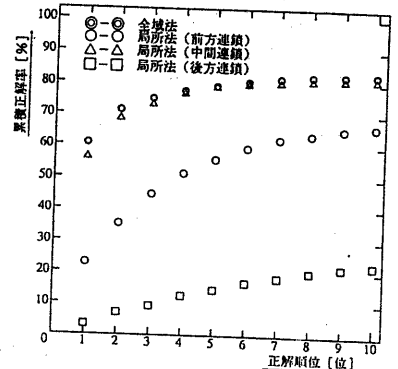


7-③ 逆方向タイプ

図7 漢字かな文の文節先頭1文字目における誤り訂正結果(標本外)



8-① 順方向タイプ



8-② 中間タイプ

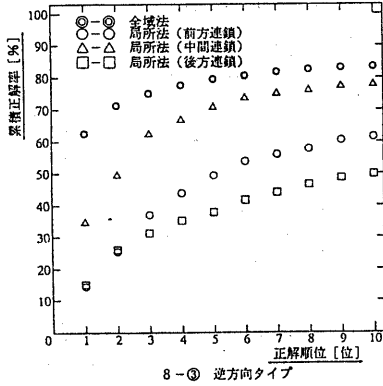


図8 漢字かな文の文節先頭2文字目における誤り訂正結果 (標本外)

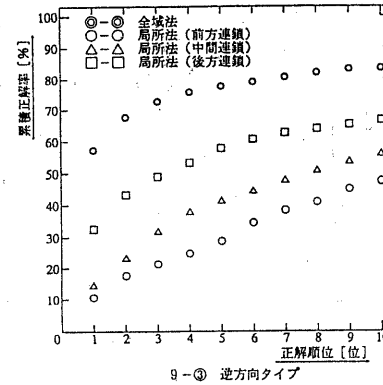
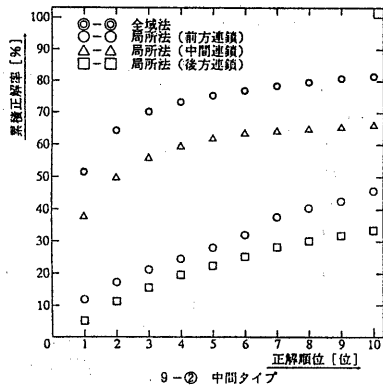
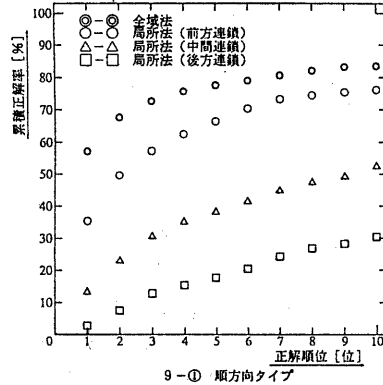


図9 漢字かな文の文節先頭3文字目における誤り訂正結果 (標本外)

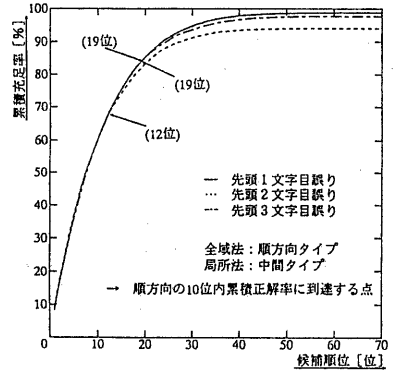


図10 局所法による全域法の10位内候補の選出能力 (音節文) 標本外データ

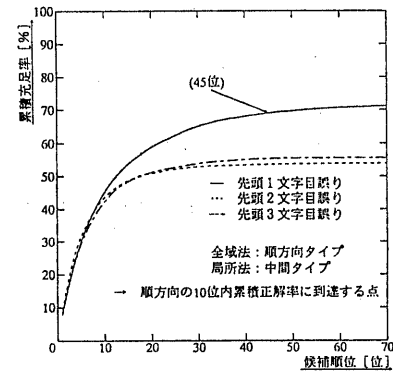


図11 局所法による全域法の10位内候補の選出能力 (漢字かな文) 標本外データ

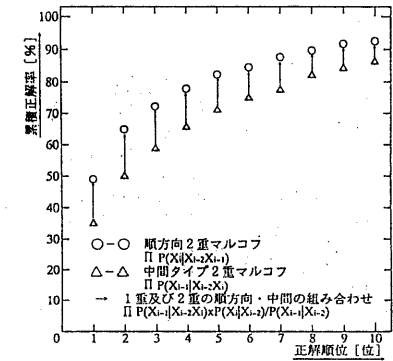


図12 2重マルコフ連鎖確率の順方向タイプと中間タイプの相互交換の関係 (全域法による誤り訂正の場合)