

固有名詞の認識を含むHMMによる英文形態素解析

揚石 亮平[†] 三浦 孝夫[†]

[†] 法政大学 工学部 情報電気電子工学科 〒184-8584 東京都小金井市梶野町 3-7-2

E-mail: †ryohei.ageishi.st@eng.hosei.ac.jp, ††miurat@k.hosei.ac.jp

あらまし 本稿では、英文形態素解析に広く用いられてきた隠れマルコフモデルの有用性を検証する．HMM を用いることで生じる未知語（学習データに出現していない語）問題により、固有名詞を正しく処理できそうに無い．このため、本稿では 2 グラム上で接続表を用いる手法を提案し、向上できることを示す．

キーワード 形態素解析、隠れマルコフモデル、2 グラム、固有名詞、未知語

Part-Of-Speech Tagging using by Hidden Markov Model and Named Entity Recognition

Ryohei AGEISHI[†] and Takao MIURA[†]

[†] Electronics,Electrical and Computer Engineering, Faculty of Engineering, HOSEI University 3-7-2,

KajinoCho, Koganei, Tokyo, 184-8584 Japan

E-mail: †ryohei.ageishi.st@eng.hosei.ac.jp, ††miurat@k.hosei.ac.jp

Abstract In this investigation we verify and confirm the availability of Hidden Markov Model that is widely used for English morphological analysis. Unfortunately it is impossible to identify Named Entity because of *unknown word* problem (that doesn't appear in training data), though it is an important issue in morphological analysis. Here to recognize Named Entity correctly, we introduce conjunction table using by 2 gram.

Key words Part-of-speech Tagging, Hidden Markov Model, 2 gram, Named Entity, unknown words

1. ま え が き

これまで英語文章の形態素解析のために、個々に生じる単語に品詞タグを識別する、タグ付け (tagging) 手法が論じられてきた．本研究では、確率過程に基づくタグ付け手法を提案し、とくに固有名詞を正しく処理することによって、認識制度を向上させる手法を論じる．多くの単語は、複数の品詞の働きを有し辞書を検査しただけでは特定できない．このため、前後の関連や共起語などから頻度解析や確率推定が効果的であるとされる．

形態素解析を高精度に行うことは、(英語に限らず) 自然言語処理の重要な基本技術の 1 つである．構文や意味を正しく解析し、その意図を正しく認識するためには不可欠であると言ってよい．特に、固有名詞を正しく認識することは、情報抽出や機械翻訳においてキーステップとなることがある．例えば、*I like White House* という表現からは特定の政治信条を読み取ることができる．固有名詞 *White House* を正しく認識できないと、全く意味の異なる文章となってしまう．

しかし固有名詞かどうかを判定することは容易ではない．例えば、確率推定を用いる場合、学習データに出現しない単語は

頻度 0 であり、確率値 0.0 と推定される．また、固有名詞は辞書に含まれず判定できない．実際には未知語がこれに相当することが多いため、固有名詞の認識は困難であると考えられる．

現在、固有名詞の認識には、外部辞書を用いたものが多く用いられている．Ji and Grishman [2] らでは、隠れマルコフモデルを用いてタグ付け後、4 つの Re-Rnaker を順次使用することによって、固有名詞を認識している．これらの Re-Rnaker はそれぞれ構造ベース、関係ベース、イベントベース、共参照ベースのものであり、それぞれ異なるデータを基に構成されている．この手法により、約 4 % ほどの精度が上昇している．しかし、このような手法は、外部に辞書やルールを持つため構造が複雑であり、またその構成には、大量の学習データが必要となっている．

本研究の目的は、2 つある．1 つは、英語形態素解析で広く用いられている隠れマルコフモデル (HMM) の有用性を確認することである．このため学習データから確率モデルを生成し、任意の文章に対して品詞推定を行い、有用性を確認する．もう 1 つは、未知語中の固有名詞を認識する手法を提案することである．本稿では、2 グラムを用いた発見的方法を提案し、効果的であることを実験で検証する．この手法を導入することによ

り，HMM によるタグ付けを改善し，未知語に対しても正しく解析できる能力を与えることができる．

第 2 章で，品詞を推定するために用いた隠れマルコフモデルの適用方法を示す．第 3 章では，固有名詞の認識手法を提案する．2 グラムモデルを用いた発見の手順を示して特徴や利点を述べる．第 4 章では，2 つの実験結果を述べ，本手法の有効性を示す．

2. 隠れマルコフモデル

2.1 隠れマルコフモデル

隠れマルコフモデル (Hidden Markov Model, HMM) とは，確率的な状態遷移と記号出力を備えたオートマトンである．次の 5 項組 $M = (X, Y, A, B, \pi)$ で定義される．

(1) 出力記号系列 $X = \{x_1, \dots, x_n\}$ であり，観測可能である．

(2) 状態遷移系列 $Y = \{y_1, \dots, y_n\}$ であり，観測不可能である．

(3) 状態遷移確率分布 $A = \{a_{ij}\}$ であり，状態 y_i から状態 y_j への遷移確率である．単純マルコフを仮定している．

(4) 記号出力確率分布 $B = \{b_i(x_t)\}$ であり，状態 y_i で記号 x_t を出力する確率である．出力は現在の状態にのみ依存すると仮定している．

(5) 初期状態確率分布 $\pi = \{\pi_i\}$ であり，状態 y_i が初期状態である確率である．

すなわち，図 1 のようなモデルが構成される．

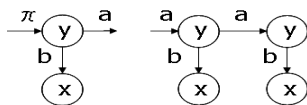


図 1 隠れマルコフモデル

本稿では，あらかじめ正解が与えられた学習データからモデルを生成し，これを基に，最適な状態遷移系列の推定を行う教師あり機械学習の手法を用いることとする．

また，その推定には，Viterbi アルゴリズムを用いる．Viterbi アルゴリズムは以下のようなステップを持つ．

(1) 各状態 $i=1, \dots, N$ に対して，変数の初期化を行う．

$$\begin{aligned}\delta_1(i) &= \pi_i b_i(o_1) \\ \psi_1(i) &= 0\end{aligned}$$

(2) 各状態の遷移ステップ $t=1, \dots, T-1$ ，各状態 $j=1, \dots, N$ について，再帰的に計算を行う．

$$\begin{aligned}\delta_{t+1}(j) &= \max_i [\delta_t(i) a_{ij}] b_j(o_{t+1}) \\ \psi_{t+1}(j) &= \operatorname{argmax}_i [\delta_t(i) a_{ij}]\end{aligned}$$

(3) 再起計算の終了

$$\begin{aligned}\hat{P} &= \max_i \delta_T(i) \\ \hat{q}_T &= \operatorname{argmax}_i \delta_T(i)\end{aligned}$$

(4) バックトラックによる最適状態遷移系列を復元する． $t=T-1, \dots, 1$ に対して行う．

$$\hat{q}_t = \psi_{t+1}(\hat{q}_{t+1})$$

Viterbi アルゴリズムにより，確率値を最大にする状態遷移系列を得ることができる．

2.2 HMM とタグ付けの対応

英語の文章は，観測できない状態遷移系列として，品詞を持っている．例えば，*The man saw a girl* という文章の場合，その観測できない (隠れ) 状態列として，図 2 のように，品詞を持っていると考えられる．

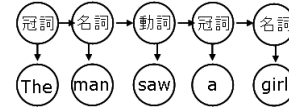


図 2 HMM の対応図

したがって，本稿で扱う品詞推定は，隠れマルコフモデルにおいて，状態遷移系列を品詞，出力記号系列を単語とし，学習データから， (A, B, π) を以下で示した式のように学習することにより，モデルを生成する．ただし， $P(y|x)$ は条件付き確率であり，条件 x の下で y が生じる確率を示したものである．また $C(x)$ は生起頻度であり，語 x が当該文書で出現する頻度を表す． BOS (Beginning Of Stream) は初期状態であり，文の先頭であることを表す．

$$\begin{aligned}\pi_i &= P(y_i | BOS) = \frac{C(y_i)}{C(BOS)} \\ a_{ij} &= P(y_j | y_i) = \frac{C(y_i, y_j)}{C(y_i)} \\ b_i(x_t) &= P(x_t | y_i) = \frac{C(y_i, x_t)}{C(y_i)}\end{aligned}$$

以下に，隠れマルコフモデルの学習例を示す．次の 3 つの文章を学習した場合を考える．

- *Dogs*_名 *like*_動 *my*_代 *red*_形 *arrows*_名
- *This*_代 *flies*_動 *an*_冠 *arrow*_名
- *No*_形 *time*_名 *like*_前 *the*_冠 *present*_名

まず，表 1 のように状態遷移頻度を計算する．表 1 中の値は状態 y_i から状態 y_j に遷移する頻度を示すものであり， $C(y_i, y_j)$ に対応する値である．例えば，状態 y_i を”名詞”とすると，”動詞”に遷移する頻度が 1 であることを示している．また合計の値は，状態 y_i から状態 y_j に遷移する頻度の合計で， $C(y_i)$ に対応する値である．

		状態 y_j						
		代名詞	名詞	冠詞	動詞	形容詞	前置詞	合計
状態 y_i	代名詞				1	1		2
	名詞				1		1	2
	冠詞		2					2
	動詞	1		1		1		3
	形容詞		2					2
				1				1

表 1 状態遷移頻度表

つぎに，この頻度表を基に，状態遷移確率を計算する (表 2)．例えば，状態 y_i を”名詞”，状態 y_j を”動詞”とすると，その状態遷移確率 $P(\text{”動詞”} | \text{”名詞”})$ は， $C(\text{”名詞”}, \text{”動詞”})$ が 1， $C(\text{”名詞”})$ が 2 であるので， $P(\text{”動詞”} | \text{”名詞”}) = 1/2$ となる．

) = $\frac{C(\text{"名詞"}, \text{"動詞"})}{C(\text{"名詞"})} = 0.5$ となる．このようにして，すべての確率値を計算する．その値を表 2 に示す．

		状態 y_j						
		代名詞	名詞	冠詞	動詞	形容詞	前置詞	
状態 y_i	代名詞	0	0	0	0.5	0.5	0	
	名詞	0	0	0	0.5	0	0.5	
	冠詞	0	1	0	0	0	0	
	動詞	0.33	0	0.33	0	0.33	0	
	形容詞	0	1	0	0	0	0	
	前置詞	0	0	1	0	0	0	

表 2 状態遷移確率表

同様にして，記号出力確率 (表 3)，初期状態確率 (表 4) も頻度を基にして計算する．たとえば，記号出力確率において，状態 y_i を”名詞”として，出力 x_t をそれぞれ”dogs”，”arrows”とした場合， $C(\text{"名詞"})$ は 5 であり， $C(\text{"名詞"}, \text{"dogs"})$ と $C(\text{"名詞"}, \text{"arrows"})$ はそれぞれ 1 と 2 であるので， $P(\text{"dogs"}|\text{"名詞"}) = \frac{C(\text{"名詞"}, \text{"dogs"})}{C(\text{"名詞"})} = 0.2$ となり， $P(\text{"arrows"}|\text{"名詞"}) = \frac{C(\text{"名詞"}, \text{"arrows"})}{C(\text{"名詞"})} = 0.4$ となる．

		出力 x_t						
		dogs	like	my	red	arrows	this	
状態 y_i	代名詞	0	0	0	0	0	1	
	名詞	0.2	0	0	0	0.4	0	
	冠詞	0	0	0	0	0	0	
	動詞	0	0.5	0	0	0	0	
	形容詞	0	0	0	0.5	0	0	
	前置詞	0	1	0	0	0	0	
		flies	an	no	time	the	present	
状態 y_i	代名詞	0	0	0	0	0	0	
	名詞	0	0	0	0.2	0	0.2	
	冠詞	0	0.5	0	0	0.5	0	
	動詞	0.5	0	0	0	0	0	
	形容詞	0	0	0.5	0	0	0	
	前置詞	0	0	0	0	0	0	

表 3 記号出力確率表

		状態 y_i					
		代名詞	名詞	冠詞	動詞	形容詞	前置詞
初期状態		0.33	0.33	0	0	0.33	0

表 4 初期状態確率表

そして表 2, 3, 4 により，図 3 のような状態遷移図を生成する．図 3 において，各品詞の下に数字は初期状態確率を表す．また，矢印は，それぞれの品詞が次の状態に遷移することを表し，その数字が状態遷移確率である．また，各品詞の [] はその品詞が出力する語を記号出力確率とともに表している．

このようにして，学習データから状態遷移図を生成し，それを基に，Vitebi アルゴリズムを用いて品詞の推定を行う．

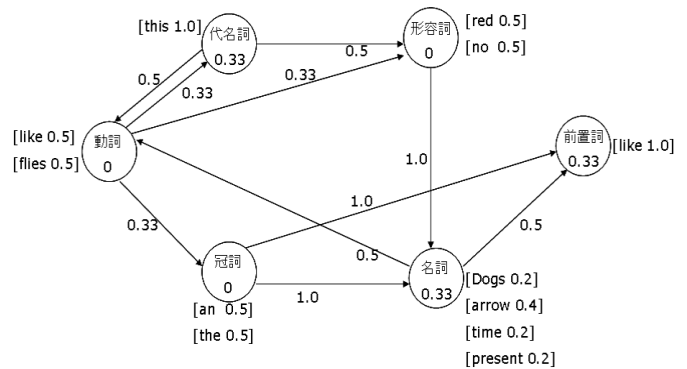


図 3 状態遷移図

2.3 ゼロ確率の回避

前項のモデルでは，学習データに出現しない状態遷移や記号出力のパターンは決して生じないとして，品詞の推定を行ってしまう．これでは，確率値がゼロとなるものが発生してしまい，品詞の推定が行われないおそれがある．

これを防ぐため，スムース化係数 s を導入し，スムージングを行うことにする．すなわち， (A, B, π) を以下のように再定義することである．ただし， N_T は品詞数であり， N_W は単語数である．

- $\pi_i = P(y_i | BOS) = s \times \frac{1}{N_T} + (1 - s) \times \frac{C(y_i)}{C(BOS)}$
- $a_{ij} = P(y_j | y_i) = s \times \frac{1}{N_T} + (1 - s) \times \frac{C(y_i, y_j)}{C(y_i)}$
- $b_i(x_t) = P(x_t | y_i) = s \times \frac{1}{N_W} + (1 - s) \times \frac{C(y_i, x_t)}{C(y_i)}$

このスムージングされた確率値を用いて，品詞の推定を行う．ただし，未知語に対する記号出力確率は一定で与えられているものとする．

3. 未知語と固有名詞判定

3.1 HMM の問題点

前章で，隠れマルコフモデルを英文形態素解析に適用する手法を述べた．しかしながら，このモデルを用いても未知語中の固有名詞を認識することは難しい．ここで，未知語とは，語自体が未知であり辞書には含まれない場合と，語は既知であるが未知の意味を含む場合に分けられる．前者の例としては *Yasuo Fukuda* のような人名が，後者の場合は *White House* や *Red Socks* などがある．後者の場合，さらに語の分かち書きに応じて異なる意味を表すことがある．例えば *New York* は地名であるが *New York Times* は新聞名である．これら固有名詞を正しく認識することは，学習データに依存して確率を計算するアプローチでは正しく判断しにくい面がある^(注1)

この問題を解決するためには，*New York* のような語を連続する二語として扱うのではなく単一語として扱う必要がある．これは分かち書き (segmentation) の問題と呼ばれる．*New York* や *White House* といった語は，それぞれの単語が独立して意味を持っているのではないため独立した単語とみなす機構が必

(注1): 実際，以下の実験では，未知語の中の半数が固有名詞であったが，正しく判読できなかった．

要である。

3.2 固有名詞の認識

未知語中の固有名詞の認識の発見的手法として、本稿では、固有名詞に限定した接続表を用いる。これは、学習データから、固有名詞の前の1単語とその品詞と、後ろ2単語とその品詞を取り出し、固有名詞を出現させる特定のパターンを抽出したものを表にしたものである。実験では、未知語に対して、その前後の単語や、隠れマルコフモデルを用いてタグ付けした品詞を用い、接続表との比較を行い、接続表と一致したものの未知語の品詞を、“固有名詞”であるとする。

また、前項で示した、連続する二語を単一語とみなす機構を実現する手法として、本稿では、“2グラム語”を提案する。例えば、*New York* を *New-York* のようにして単一語として扱うということである。

ここで、Nグラムとは、自然言語処理において最も広く用いられている言語モデルの1つである。その構造は単純で、単語の生起が直前のN-1単語にのみ依存すると考えたモデルである。ここでは、N=2とした2グラム、つまり、単語の生起が直前の単語にのみ依存すると考えたモデルを扱うこととする。その条件付き確率は $P(x_t|x_{t-1}) = \frac{C(x_{t-1}, x_t)}{C(x_{t-1})}$ で定義されている。

このようにして2グラムを用いることには、様々な利点が考えられる。まず初めに、2グラム語により、*New-York* を単一語として扱うことで、*New* が単独で出現することがなくなる。したがって、確率的に形容詞として品詞が与えられることがなくなり、HMM タグ付けにおいて、固有名詞と見なしやすくなると考えられる。

さらに、条件付き確率Pが高確率である場合、その2つの語は結びつきが強いと考えることができる。結びつきが強い語は固有名詞として振る舞うことが多い。以下に、閾値が0.8の場合の2グラム語の例を示す(表5)。

Armed-Forces	Part-J
Brooklyn-College	Peace-Corps
Civil-War	Puerto-Rico
Gross-Income	Rhode-Island
Hong-Kong	Saxon-Shore
Internal-Revenue	Simms-Purdew
Jesus-Christ	Small-Business
Kansas-City	Supreme-Court
Kennedy-Administration	U.-S.
Linda-Kay	United-States
Lo-Shu	Vice-President
Los-Angeles	Wall-Street

表5 2グラム語 (閾値 0.8)

表5にあるように、*Hong-Kong* や *Wall-Street* など、単一語で固有名詞と扱われている語が多い。よって、2グラムは固有名詞の認識に適していると考えられる。

未知語中の固有名詞の認識の具体的な手法としては、以下のようなアルゴリズムを持つ。

(1) 学習データから2グラムを計算する。

(2) 計算した確率値がある閾値より高い場合に限り、2グラム語とし、これに見合うテストデータ中の語を2グラム語に変換する。

(3) 隠れマルコフモデルを用いて、テストデータの品詞の推定を行う。

(4) さらに、未知語に対しては、接続表を用いて、固有名詞の認識を行う。

4. 実験

以下に、隠れマルコフモデルを用いた品詞の推定(実験1)と2グラムを用いた未知語中の固有名詞の認識(実験2)の2つの実験結果を示す。

4.1 準備

本研究では、学習データとしてBrownコーパスを用いる。Brownコーパスとは、1960年代にアメリカ人によって書かれた英語を、約100万語集めて編集されたコーパスであり、アメリカで出版された本、新聞などから15のカテゴリー、500のテキストで構成されている。また、その品詞は、“New-JJ York-NP”, “White-JJ House-NN”のように与えられている(JJ:形容詞, NP:固有名詞, NN:名詞)。学習には、Brownコーパス全57,082行のうち55,082行(総単語数約110万語)を用いる。品詞(状態)数はBrownコーパスに従い、その数は311語である。また、テストデータには、Brownコーパスの残りの2,000行を品詞を隠した状態で使用する。

4.2 評価方法

実験1では、隠れマルコフモデルによりテストデータ(Brownコーパス2,000行)の品詞の推定を行う。そしてその結果を、隠しておいた品詞を復元したものとの一致率で評価する。一致率は次のように定義する。

$$\text{一致率} = \frac{\text{品詞が一致した単語数}}{\text{全単語数}}$$

実験は学習量を変化した場合と、スムース化係数*s*を変化した場合の2つを行い、その一致率で評価する。

実験2では、テストデータを2グラム語に変換してしまっているため、Brownコーパスとの品詞の比較は行えない。そこで、評価には、GoTaggerという、BrillTagger^(注2)をWindowsOS上で実行できるようにしたタグ付け機を用いて品詞を与え、その品詞を正解とすることにする。その結果と、提案手法を用いて未知語中の固有名詞の判定を行ったものとを比較し、その適合率、再現率を計算することにより評価する。ただし、*a*:提案手法が固有名詞とみなした語、*b*:GoTaggerが固有名詞とみなした語、*c*:提案手法とGoTaggerの両方が固有名詞とみなした語、である。

$$\text{適合率 (Precision)} = \frac{c}{a}$$

$$\text{再現率 (Recall)} = \frac{c}{b}$$

実験は、2グラムの閾値を変化させて行い、適合率と再現率で評価する。

(注2): Eric Brill が開発したタグ付け機であり、Brownコーパスを基に製作した品詞辞書、文法辞書などによりタグ付けを行う。

4.3 実験結果

実験1の、学習量を変化した場合と、スムース化係数 s を変化した場合の一致率の変化を表6、表7に示す。

学習量 (行)	正解率 (%)
1000	61.79
5000	83.12
10000	87.11
20000	89.19
30000	90.02
40000	90.46
45000	90.61
50000	90.67
52000	90.69
55082	90.81

表6 実験1:学習量変化

スムース化係数 s	正解率 (%)
0.1	92.31
0.05	92.88
0.01	93.27
0.005	93.34
0.001	93.40
0.0005	93.40
0.0001	93.41
0.00005	93.41
0.00001	93.40
0.000001	93.37

表7 実験1:スムース化係数 s 変化

学習量を変化させた場合、全てを学習した場合が最も一致率が高く一致率は90.81%であった。また、学習量を55,082行としてスムース化係数 s を変化した場合、 $s=0.0001$ で最も一致率が高く一致率は93.41%であった。

次に、実験2の、閾値を変化させた場合の未知語中の固有名詞に対する適合率と再現率の結果を表8に示す。

閾値	適合率	再現率
0.1	94.87	36.54
0.2	94.78	37.05
0.3	94.59	37.12
0.4	94.52	37.31
0.5	94.52	37.44
0.6	94.49	37.68
0.7	94.65	37.79
0.8	94.64	37.99
0.9	94.59	39.27

表8 実験2:固有名詞の認識

適合率は0.1で最も高く、再現率は0.9で最も高い値を示している。

4.4 考察

実験1の学習量の変化においては、すべてを学習した場合が最も一致率が高くなっている。これは、学習量を増やすことにより未知語の数を減少させることができたからであることが考えられる。未知語が多く存在するとそれに比例してゼロ確率が発生してしまう。それにより推定が行われな可能性が高くなってしまふからであると考えられる。一般に、過学習(特定の入出力の組み合わせに対して系が収束しすぎたため、汎用性がなくなってしまうこと)も誤りを生じる原因とされるが、未知語の生じやすいタグ付けでは、過学習よりも未知語の出現が、より多くの誤りを生じる原因となると推測できる。

また、実験1のスムース化係数 s の変化においては、スムース化係数 s を変化させることにより2.6%の利得を得ることができている。未知語の出力確率を一定で与えているため、その

推定は状態遷移確率に依存している。つまり、未知語に対しては、最も頻繁に現れるパターンの品詞を選択することができたため、一致率の上昇につながったと考えられる。

学習量を55,082行、スムース化係数 s を0.0001とした場合の不正解総数は2863語であり、その中に未知語が半分(1451語)含まれている。また、未知語総数は2188語であったので、約66%が誤りとなっている。このことから、隠れマルコフモデルを用いて品詞の推定を行う場合、未知語が多く存在することがその結果に悪影響を与えられられる。未知語を生じないように学習させた場合の一致率が97.35%だったことからこの推測が正しいといえよう。したがって、隠れマルコフモデルをタグ付けに用いる場合、その精度は学習データの語分布に依存するということになる。

実験2の固有名詞の認識において、適合率はほぼ一定(94%付近)である。適合率はその手法の精度を示すものであるので、この値が高いということは、提案手法の有用性を示していると考えられる。また、適合率は0.1で最も高くなっているが、これは、閾値が0.1~0.2の間で、2グラム語の数が大幅に増加することによって、接続表で正解と見なせる語が増加したため、適合率の増加につながったと考えられる。

閾値が0.7~0.9の間で、適合率、再現率ともに高い値を示している。閾値が高いということは、それだけ2グラム語の結び付きの強い語、ということであるので、2グラムを用いることにより、固有名詞の認識精度が向上できると考えられる。

閾値を0.8とした場合の結果を、隠れマルコフモデルのみの場合と比較すると、以下の表9のようになる。

	適合率	再現率
HMMのみ	94.47	34.02
HMM+2グラム	+0.17	+3.97

表9 適合率, 再現率の変化

この結果、2グラムを用いることは、特に再現率の点において効果があると考えられる。

しかしながら、中には固有名詞でないものを固有名詞としてしまうケースも存在する。例えば、*High-School*のような語である。こういった語は隠れマルコフモデルのみの場合では正しく推定できていたので、2グラムを用いることで、精度が低下してしまう可能性がある。

また、再現率は閾値が低くなるほど低下している。これは適合率の場合とは反対に、2グラムを用いることにより、接続表が正解と見なせる語の数よりも、GoTaggerが固有名詞とする語の数が増大するためであると考えられる。よって、接続表を改善することによりさらなる精度の向上を期待できるかもしれない。

5. 結論

本研究では、あらかじめ品詞が与えられた文章から、未知の文章に対して品詞の推定を行う方法として、隠れマルコフモデルを用いた方法の有用性を確認した。実際に実験を行い、その

結果から学習データの語分布に依存することも確認した。また，未知語中の固有名詞を認識する方法として，2 グラムを用いた発見的方法を提案した。その結果から，この手法が特に再現率の点で有効な手法であることを確認した。

文 献

- [1] Bernard Merialdo: Tagging English Text with a Probabilistic Model. *Association for Computational Linguistics* pp. 155-171, 1994
- [2] Heng Ji. and Ralph Grishman: Analysis and Repair of Name Tagger Errors. *Proceedings of the COLING/ACL* pp. 420-427, 2006
- [3] 浅原 正幸: 系列ラベリング問題に関するメモ，奈良先端科学技術大学院大学, 2006
- [4] 北 研二: 確率的言語モデル，東京大学出版会, 1999