

# 複合1次元HMMを利用した手書き文字認識の研究

西村 広光・堤 正義

情報工学科

## Off-line Character Recognition Using Multiple 1-D Hidden Markov Model

Hirimitsu Nishimura, Masayoshi Tsutsumi

### Abstract

The purpose of our research is to improve the recognition rate of off-line character recognition systems using the HMM (Hidden Markov Model). Some 1-dimensional (1-D) HMM recognition systems have been proposed to increase the recognition rate of the conventional recognition systems. However, the 1-D HMM recognition systems have not achieved a sufficient recognition rate. On the other hand, 2-D HMM character recognition systems have been proposed to increase the representational power. However, the complex structure of 2-D HMM could not produce enough representational power with the actual small number of learning samples. To overcome these problems, we proposed a method using the multiple 1-dimensional HMM approach. In our experiment, the recognition rate of our multiple 1-D HMM method performed 97.1%, which was about 3% better than the conventional 1-D HMM systems, and as good as the 2-D HMM (HMM-MRF) system. Thus our method has better performance in practical use.

**Key Words:** Character Recognition, Off-line, Hidden Markov Model, HMM

## 1 序論

Hidden Markov Model (HMM) の認識系への適用は、音声認識の分野で確立し [1][2]、音声と同じく 1 次元の性質を有するオンライン文字認識においても有効性が示されている [3]。

音声認識、オンライン文字認識で有効性が確認され、高い表現能力が実証された HMM を利用した認識系を、オフライン文字認識にも適用しようという研究が近年進められてきているが [4]、[5] HMM を利用した文字認識系は、従来までの方向性パターンマッチングを利用した認識系 [6] と比較して、十分な性能を発揮するまでに至っていない。

2 次元のパターンを 1 次元的な系列に変換する 1 次元 HMM を利用したのでは、2 次元パターンであるオフライン文字の情報を十分にモデルに反映できないとする考えから、HMM 自体を拡張し 2 次元シンボルの入力を可能にして認識系に利用しようという研究が現在盛んに

行われている [4][7]。しかし、2 次元 HMM を利用した認識系は、モデルが複雑になり、学習に必要なサンプル数が膨大となる。また、認識系を構築するために膨大なサンプルが必要になるため、膨大なサンプル収集が困難であるオフライン文字認識において、実用的といえない。

本研究では、モデル学習の観点から実装が実用的な 1 次元の HMM を複合することで、従来の 1 次元 HMM 認識系で課題であった低認識性能の問題を解決し、高い認識性能を持つ実用的なモデル構築手法を提案し、その効果を検証した。

## 2 HMM による文字認識

### 2.1 HMM による認識モデルの基本原則

Hidden Markov Model (HMM) は、マルコフ過程の拡張であり、一つの状態遷移に対して一つの出力シンボルが一意に対応しているマルコフ過程に対し、同一の出力

シンボル系列を出力するモデルパラメータが多数存在する隠れマルコフ過程をモデル化したものである。それら HMM のパラメータ群の中には、ある学習サンプル集合が与えられたとき、これを出力するモデルの最適パラメータがあると考えられている。

このようなモデルにおいて、あるシンボル系列の集合を仮定し、そのシンボル系列を出力する出力確率が最大となるようにモデルのパラメータ推定を行うのが HMM の学習である。

また、学習によって決定したモデルを利用して、未知クラスのシンボル系列に対する状態遷移確率を求め、出力確率を比較し、最も確からしいクラス、すなわちデータ集合の分類ラベル (英文字認識では [A, B, ..., Z]) を特定することが HMM を利用した認識である。

## 2.2 1次元 HMM による文字認識

音声認識、オンライン文字認識では、1節でも示したように、音声、ペンの動き等の情報を1次元の系列として HMM の入力として扱う。本論文では、2.3節の2次元 HMM と区別するため、音声認識等で利用されているような1次元の系列を入力とする HMM を1次元 HMM と呼ぶこととする。しかし、2次元パターンである文字を認識するためには、2次元の文字パターンを1次元の系列に変換し、HMM の入力とする必要がある。

一般的な手法では、はじめに図1に示すように、パターンを短冊化し、各短冊独立に特徴ベクトルを決定し、ベクトル量子化することで、文字パターンを1次元のシンボル系列に変換する。

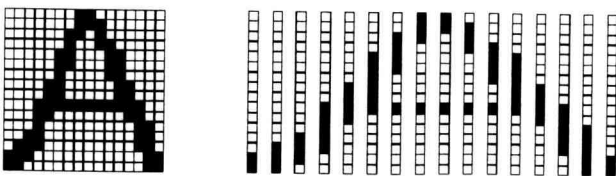


図1: パターンの短冊化

次に、学習サンプルの集合からクラスごとに作成したシンボル系列の集合を、文字クラス数の HMM に入力し、出力確率が最大となるようにパラメータ推定を行う。このとき、一般には、モデルの簡単化のため、図2に示すような、隣接する状態のみに遷移する Left to Right モデルを利用し HMM を構築する。

認識は、与えられた未知サンプルをベクトル系列に変換し、各文字クラスに対応する HMM の入力として与

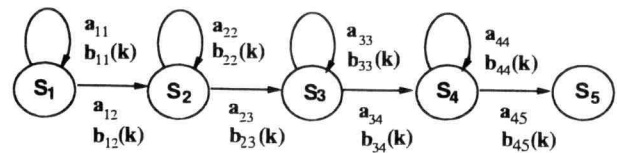


図2: 代表的な1次元 HMM (Left to Right モデル)

え、出力クラスが最大となる HMM に対応するクラスを認識結果とする。

このような1次元 HMM による文字認識は、従来までの方向性パターンマッチング、ニューラルネットワークを利用した文字認識等と比較して、十分に高い認識性能が得られていない。十分な認識性能が得られていない理由は、1次元 HMM 認識系ではパターンを単純に短冊化して特徴を求めており、従来の手法では2次元的なパターンの特徴がモデルに反映することが困難であるためであると考えられる。そこで、複数短冊にまたがる特徴も利用した、2次元的特徴採用1次元 HMM 認識系 [8] を考案し、認識性能の向上があったが、まだ十分に高い認識性能を得てはいない。

## 2.3 2次元 HMM による文字認識

文字パターンのような2次元のパターンを HMM 認識系で扱う場合、2.2節のような1次元 HMM を利用したのでは、パターンの2次元的特徴をモデルに反映することは難しい。そこで、文字パターンのような2次元の情報を扱うため、HMM で2次元のシンボル系列を扱えるように拡張したものを2次元 HMM と呼ぶ。

このような2次元の HMM 拡張モデルは、まだ一般化されたモデルはなく、現在も様々なモデルが研究されている。本論文では、既に報告されている2次元 HMM の一つで、高い認識性能が示された HMM-MRF (Mesh Random Fields) を利用した文字認識 [4] を追試し、提案手法と比較した。

次に、HMM-MRF の原理を解説する。画像空間を  $X = \{X_{ij} | (i, j) \in L\}$  を  $H \times W$  の格子  $L$  で定義されたランダム場とし、 $\Omega_{ij} = \{(k, l) \in L | l \leq k < i \text{ or } 1 \leq l < j\}$ 、 $\Theta_{ij} \subset \Omega_{ij}$  とする (図3)。 $X$  は以下の式を満たすとき MRF (Mesh Random Fields) といい、この状態を表現するように HMM を拡張したものを HMM-MRF と呼ぶ。

$$P(X_{ij} | X_{\Omega_{ij}}) = P(X_{ij} | X_{\Theta_{ij}}), \forall (i, j) \in L$$

このような2次元 HMM は、従来までの文字認識手法と比較して同等以上の性能を発揮することがわかってい

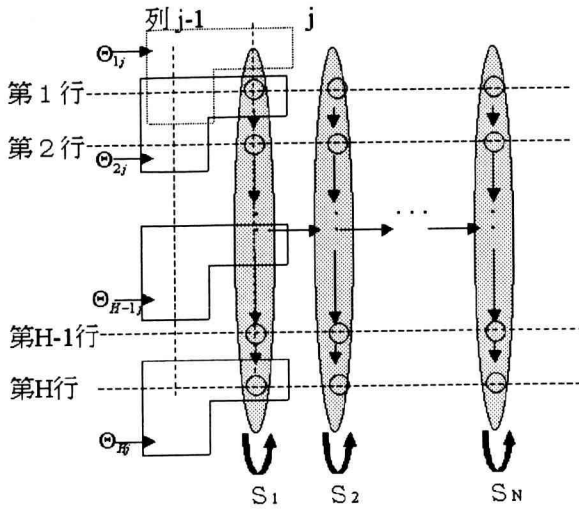


図 3: 代表的な 2 次元 HMM (HMM-MRF)

る [4]。しかし、HMM を 2 次元に拡張することはパラメータ数が膨大になるので、膨大な学習サンプルの収集が困難であるオフライン文字認識においては実用的とは言えない。

### 3 複合 1 次元 HMM による文字認識

前節までで示したように、1 次元 HMM 認識系は性能が不十分、2 次元 HMM は実装が困難であるため、本研究では、新たな HMM 認識系構築法を提案した。提案した認識系の構築方法は、認識系の構築が容易であるが、認識性能が不十分である 1 次元 HMM 認識法を利用して、それらを効果的に組み合わせ高い認識性能の実現を目指すものである。

本節では、はじめに利用した 1 次元 HMM 認識系を解説し、本研究で提案した 1 次元 HMM を利用して高い性能を実現する組み合わせ方法について説明する。

#### 3.1 採用した 1 次元 HMM 文字認識

本研究では、3 つの独立した 1 次元 HMM 文字認識系を構築し、それら 3 つの効果的な組み合わせ方法を提案した。

それぞれの 1 次元 HMM 認識系は、特徴抽出したあるベクトル系列を、基本的な 1 次元 HMM 認識系と同様の手法で、ベクトル量子化によりシンボル系列とし、HMM の入力として扱った。

以下に、3 つの HMM 認識系で採用した、特徴抽出法

と、ベクトル系列を説明する。なお、本手法では 1 次元 HMM を利用するため、一般的なベクトル量子化を用いて、ベクトル系列からシンボル系列を作成し、HMM で利用した。

**短冊独立特徴抽出型 HMM** 一つ目の特徴抽出法は、文字パターンを左から右に、縦長の短冊状に分割してそれぞれの短冊内で求めた特徴をベクトル値とし、左から右のベクトル系列を作成した。

具体的には、図 4 に示すように、短冊状に分割した各領域で、矢印で示したような、黒画素の位置、大きさ、数などを数値化し、ベクトル系列を作成した。

本論文では、このように短冊化して各短冊独立してベクトル系列を求める HMM 認識系 (構築手法) を、短冊独立特徴抽出型 HMM と呼ぶことにする。

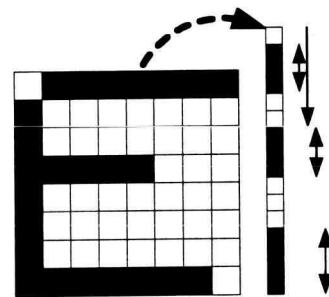


図 4: 短冊独立特徴抽出型 HMM

**隣接短冊特徴抽出型 HMM** 二つ目の特徴抽出法は、短冊独立特徴抽出型 HMM と同様にしてパターンを短冊化し、短冊内の画素と隣接する短冊の画素との関係の特徴として、ベクトル系列を作成した。

具体的には、図 5 に示すように、短冊状に分割した際の右側隣接短冊との画素の一致を、(黒・黒)、(黒・白)、(白・黒)、(白・白) の 4 種類に分類して数値化し、ベクトル系列を作成した。

本論文では、このように、隣接する短冊の関係を利用してベクトル系列を求める HMM 認識系 (構築手法) を、隣接短冊特徴抽出型 HMM と呼ぶことにする。

**黒画素分布特徴抽出型 HMM** 三つ目の特徴抽出法は、格子空間に分割したパターンに対して、各格子内の黒画素数を特徴とし、パターンの縦方向の格子の並びからベクトル化、さらに横方向の並からベクトル系列化して求めるものである。

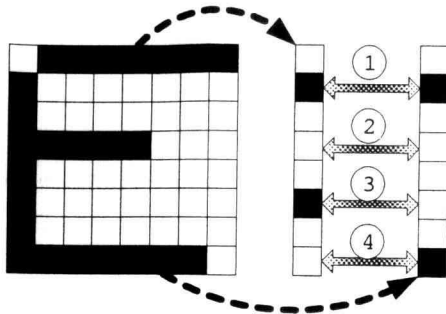


図 5: 隣接短冊特徴抽出型 HMM

具体的には、図 6 に示すように、一定領域ごとに黒画素数を調べ、縦方向の並びをベクトルとして並び、それらベクトルをまとめてベクトル系列を作成した。

本研究では、このように、格子空間内の黒画素数を利用してベクトル系列を求める HMM 認識系 (構築手法) を、黒画素分布特徴抽出型 HMM と呼ぶこととする。

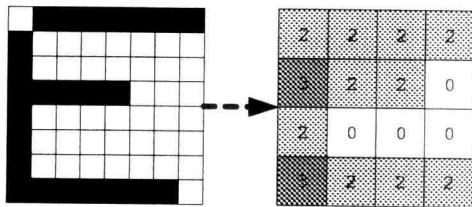


図 6: 黒画素分布特徴抽出型 HMM

### 3.2 複数 HMM 認識系の複合法

本研究では、前節に示した 1 次元 HMM を利用して、高い認識性能を実現する複合 1 次元 HMM 認識系構築法を提案した。本節では、提案した複合法を説明する。

クラス  $m$  の学習により構築した  $n$  個の 1 次元 HMM の出力確率を  $HMM_n^m$  であらわすと、1 次元 HMM 認識系  $HMMRec_n$  は、次式であらわされる。

$$HMMRec_n = \operatorname{argmax} \{ HMM_n^0, \dots, HMM_n^m \}$$

ここで、HMM の出力値は、各モデルに対しての適合確率であるので、図 7 に示すように、複合 1 次元 HMM 認識系  $Multiple - HMMRec(MHMMRec)$  を次式に示すような、複数の HMM の出力確率の和によって定義した。

$$M - HMMRec = \operatorname{argmax} \{ \sum_{n=1}^n HMM_n^0, \dots, \sum_{n=1}^n HMM_n^m \}$$

このようにして複合 1 次元 HMM として認識系を構築したのは、次の理由からである。1 次元 HMM を利用した文字認識において、誤認識が多く見られるパターンを解析したところ、特徴ベクトルの値では微小な変動が、HMM の出力値で大きな差を生む場合があり、微小変動したパターン群を同一なクラスとして認識できないことがあるのがわかった。その原因を検討したところ、変動パターンに対して、ある HMM の出力確率が不安定になることがわかった。そこで、不安定な HMM の出力確率を安定化するため、複数の方式で作成した 1 次元 HMM の出力確率の和によって入力に対する出力値を定義することで、モデルからの出力値の安定化を試みたのである。

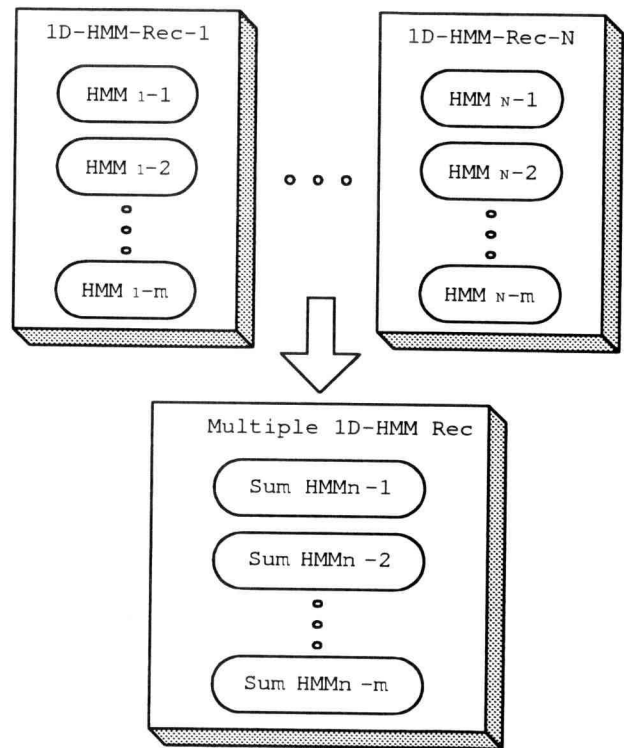


図 7: 複合 1 次元 HMM の認識系

## 4 評価実験

本研究では、3 節で示した、各 HMM 認識系を構築し、共通データベースを利用して評価実験を行った。

利用したデータベースは、ETL6[9] の常用手書き英文字データベース、英文字 [A-Z] を利用した。前処理として、2 値化、ノイズ除去、大きさの正規化を行い、その段階で一般に行われるように、人間が読み取れないほどノイズの多いものは目視により実験データから除外した。

実験に利用した 26 クラスのデータ数は約 1,300 で、そのうち各 600 を学習用サンプルセット、残りを検査用サンプルセットとした。検査用サンプルセットの総数は、18,636 である。

#### 4.1 単独 1 次元 HMM を利用した文字認識実験結果

表 1 に、3.1 で示した 3 つの 1 次元 HMM についての検査用サンプルの認識結果を表 2 に学習用サンプルの認識結果を示す。

表 1,2 の認識結果の中で、 $n$  位累積正解率という評価基準を利用している。これは、正解が認識系の出力の  $n$  番目以内に含まれている確率を表しており、文字認識性能の評価に多く利用されている。文字認識結果を知識処理を利用して 2 段階判定で確定すれば、その多くは正解にできるためである。

表 1 からわかるように、短冊独立特徴抽出型・黒画素分布特徴抽出型 1 次元 HMM 認識系の認識結果は、1 位正解率はやや高いといえるが、3 位累積正解率の 96% 程度で一般的評価に 99 高い認識性能とはいえない。また、隣接短冊特徴抽出型 1 次元 HMM 認識結果は、1 位正解率は上記 2 つの手法より認識性能は劣るものの、3 位累積正解率ではやや高い性能を持っている。

このような 3 つの 1 次元 HMM 文字認識系を利用して次節で示す複合 1 次元 HMM 認識系の評価実験を行った。

表 1: 単独 1 次元 HMM 文字認識結果 (検査サンプル)

	n 位累積正解率		
	1	2	3
短冊独立特徴抽出型	95.1%	96.5%	96.9%
隣接短冊特徴抽出型	89.6%	96.0%	97.2%
黒画素分布特徴抽出型	93.2%	95.9%	96.5%

表 2: 単独 1 次元 HMM 文字認識結果 (学習サンプル)

	n 位累積正解率		
	1	2	3
短冊独立特徴抽出型	98.1%	99.5%	99.7%
隣接短冊特徴抽出型	93.2%	98.2%	99.5%
黒画素分布特徴抽出型	96.0%	98.9%	99.5%

#### 4.2 複合 1 次元 HMM を利用した文字認識実験結果

表 3,4 に、3.2 節で示した複数組み合わせ法を利用して、表 1 の 3 つの 1 次元 HMM のうち 2 つを組み合わせた場合の認識結果を検査用サンプル、学習用サンプル別に示す。

さらに、表 1 の 1 次元 HMM の 3 つを組み合わせた場合の認識結果と、既に報告されている 1 次元 HMM 文字認識結果 [8] を表 5 示す。

本節で示した表 3, 4, 5 では、3.1 節の 3 つの HMM 認識系をそれぞれ、下記のように記述している。

- T1 短冊独立特徴抽出型 1 次元 HMM
- T2 隣接短冊特徴抽出型 1 次元 HMM
- T3 黒画素分布特徴抽出型 1 次元 HMM

表 3: (2) 複合 1 次元 HMM 認識結果 (検査サンプル)

	n 位累積正解率		
	1	2	3
T1+T2	96.9%	98.4%	98.9%
T1+T3	97.0%	98.4%	98.7%
T2+T3	95.5%	98.0%	98.7%

表 4: (2) 複合 1 次元 HMM 認識結果 (学習サンプル)

	n 位累積正解率		
	1	2	3
T1+T2	99.0%	99.6%	99.8%
T1+T3	98.9%	99.6%	99.7%
T2+T3	97.9%	99.5%	99.7%

表 5: 複合 1 次元 HMM 認識結果 (3 つの HMM)

	n 位累積正解率		
	1	2	3
T1+T2+T3	97.1%	98.7%	99.1%
基本的 1 次元 HMM	86.2%	94.3%	96.2%

### 4.3 評価

表 2、表 4 からわかるように、本提案組み合わせ手法を利用して 2 つの 1 次元 HMM 認識系を統合することでより高い学習サンプルに対する認識率が得られた。この結果から、本提案組み合わせ手法により統合した認識系は、単独の HMM 認識系よりも高い表現能力を持つことがわかる。

また、表 1、表 3 からわかるように、本提案組み合わせ手法を利用して 2 つの 1 次元 HMM 認識系を統合することでより高い認識率が得られた。この結果から、本提案組み合わせ手法により統合した認識系は、単独の HMM 認識系よりも高い認識性能を発揮することがわかる。特に、隣接短冊特徴抽出型 HMM は他 2 つの手法と比べ、単独では 1 位正解率が低いが、高認識性能の他 2 つのモデルと統合しても、高認識性能モデル単独よりも高い認識性能を発揮した。

また、表 5 からわかるように、3 つの 1 次元 HMM 認識系を統合した場合、表 3 の認識結果よりも高い性能を発揮していることがわかる。具体的には、従来示されていた 1 次元 HMM 認識性能にくらべ、1 位正解率で約 11%、3 位累積正解率でも約 3% の認識性能の向上が得られた。

## 5 結論

本研究は、オフライン文字認識で高い認識性能が期待されている HMM 認識系において、1 次元 HMM は認識性能が不十分であり、2 次元 HMM は膨大な学習サンプルを必要とするため実用化が困難であるという課題を解決する手法として、複数の 1 次元 HMM を複合し認識系を構築することで高い認識性能の実現を試み、検証を行った。

実装した認識系を利用した評価実験から、提案手法により統合した認識モデルの表現能力の高さ、認識性能の高さを実証した。

具体的には、1 位正解率で約 95%、3 位累積正解率でも約 97% の 1 次元 HMM 認識系 3 つを提案手法で統合した認識系を構築し、1 位正解率 97%、3 位累積正解率 99% という高い認識性能を実現した。

以上から、本研究で提案した手法の導入により、HMM を利用した、実用的かつ高い認識性能をもつ認識系の実現が可能となるものと考えられる。

**謝辞** 本研究をまとめるにあたり、ご意見を戴いた信州大学教授中野康明博士に深く感謝する。

## 参考文献

- [1] L. R. Rabiner, "A tutorial on hidden Markov model and selected applications in speech recognition," Proc. IEEE Vol.77[2], pp. 257-286, 1989
- [2] 中川 聖一, "確率モデルによる音声認識," 電子情報通信学会, 1998
- [3] J. J. Lee, J. H. Kim and M. Nakajima, "A hierarchical HMM network-based approach for on-line recognition of multiple cursive handwritings," IEICE Trans. INF&SYSTEM Vol. E81-D[8], pp. 881-888, 1998
- [4] G. Saon and A. Belaid, "Off-line handwritten word recognition using a mixed HMM-MRF approach", Proc. ICDAR'97 Vol. 1, pp. 118-122, 1997
- [5] A. Kundu and P. Bahl, "Recognition of handwritten script: a hidden Markov model based approach," Proc. Int. Conf. on Acoustics, Speech and Signal Processing, New York City, USA, pp. 928-931, 1988
- [6] 丸山, 小林, 山田, 中野, "複数の文字認識を用いた連続手書き英単語認識," 電子情報通信学会論文誌 Vol. J82-DII[9], pp. 1435-1443, 1999
- [7] H. S. Park and S. W. Lee, "Off-line recognition of large-set handwritten character recognition," Pattern Recognition Vol. 31[12], pp. 1849-1864, 1996
- [8] H. Nishimura, M. Kobayashi, M. Maruyama and Y. Nakano, "Off-line Character Recognition Using HMM by Multiple Directional Feature Extraction and Voting with Bagging Algorithm," Proc. IC-DAR'99, pp. 49-52, 1999
- [9] 斎藤 泰一, 山田 博三, 森 俊二, "手書き文字データベースの解析 (III)," 電総研彙報 42[5], pp. 385-434, 1978