

多クラス早期認識ブースティング法とその応用

石黒 勝彦[†] 澤田 宏[†] 坂野 鋭[†]

[†] NTT コミュニケーション科学基礎研究所

京都府相楽郡精華町光台 2-4

E-mail: †{ishiguro,sawada,keen}@cslab.kecl.ntt.co.jp

あらまし ジェスチャ認識のような時系列シーケンスの認識においては、応用によりシーケンスの終了を待たずにできるだけ早く認識結果を出すことが要求される。この問題は早期認識の問題と呼ばれ、内田らによって提案されたブースティング的解法が効果的であることが知られている。しかし、内田らの方法は2クラス識別手法であるため、例えば音声語彙認識などの多クラス識別問題への応用が困難であった。そこで我々は内田らの手法を多クラスへと拡張した多クラス早期認識ブースティング法を考案した。提案手法は簡潔に定式化可能であり、さらにその計算量はクラス数に対して線形で抑えることが可能である。自動車内の運転者行動認識および手書き数字認識をタスクとした実験を行い、提案手法の有効性を示した。

キーワード 早期認識, 時系列解析, ブースティング, 多クラス識別問題

Multiclass Boosting for Sequence Early Classification and its Applications

Katsuhiko ISHIGURO[†], Hiroshi SAWADA[†], and Hitoshi SAKANO[†]

[†] NTT Communication Science Laboratories

2-4 Hikari-dai, Seika-cho, Soraku-gun, Kyoto, Japan

E-mail: †{ishiguro,sawada,keen}@cslab.kecl.ntt.co.jp

Abstract We propose a novel boosting method for early classification of sequences. In many practical problems, we would like to classify a sequence into one of K classes as quickly as possible, without waiting for the end of the sequence. In this paper, we propose a multiclass early classification model based on Adaboost, which extends the previously proposed “Early-boost” for a binary classification to deal with many real-world problems. The formulation of the proposed multiclass early classification method is straightforward, and its computational cost is small. We show the effectiveness of our method through the classification experiments of driver behaviors and online hand writings.

Key words Early classification, time series data analysis, boosting, multiclass classification

1. はじめに

本稿では、時系列データをはじめとするシーケンスデータの多クラス識別問題、特に早期認識と呼ばれる問題について論ずる。

例として車載カメラを用いた運転者行動の認識問題 [1] を考える。カーナビゲーションシステムなどに代表される自動車の ICT 化は顕著であり、運転者の行動・動作・状態をカメラ等を通じて認識することで、運転者の状態に応じた、より快適で安全な運転環境を提供することが将来的に可能となると予想される。この際、運転者の役に立つ機能をシステムが発現するためには、運転者が行動を終了する前にその行動を予測・認識することが自然に要請される。たとえば運転中の携帯電話操作などの危険な行動に対しては、システム側はできるだけ早期に行動を認識してドライバーに警告を与えるなどの処理を行

う必要がある。また、運転者がサイドミラーを確認する動作から車線変更の可能性を早めに予知することができれば、複数車線の運転をスムーズに協調制御することで、例えば混雑した高速道路などでもより安全な車線変更の支援が実現できる。

このように、時系列シーケンスの終了をまたずにそのクラス等を認識する識別問題は早期認識 (早期識別, early classification) と呼ばれ、ジェスチャ認識 [2], オンライン手書き文字認識 [3] や音声発話認識システム [4] を始めとして多くの現実的なタスク・システムで必要となる課題である。

この課題に対して、一般に生成モデルあるいは識別モデルの利用による解決が考え得る。一つの解法として、HMM [5] のような時系列生成モデルの利用が容易に想像される。しかしこれらのモデルは一般に時系列シーケンス $x_{1:T}$ を最後まで入力しシーケンス全体の情報から最適

な識別結果 y を出力するという利用方法が一般である。モデルパラメータの学習においても上記の利用方法を想定しているため、そこで学習されたモデルパラメータがサブシーケンス $x_{1:t} (t < T)$ の認識問題に対して最も優れているという保証は無い。一方、識別モデルとしては SVM [6] や Adaboost [7] などが高い識別能力ゆえに良く知られている。これらのモデルを利用するためには一般にシーケンスサンプル x_i, x_j 間の距離等を良く表現する特徴量が必要となる。早期認識を行う場合、(部分)シーケンスの長さは一定ではないため、不確定なシーケンス長の識別に有効な特徴を定義する必要がある。これは識別モデルの開発と同程度かそれ以上に困難な課題である。

この問題に対して、主にブースティング [7] の拡張モデルがいくつか提案されている。Sochman and Mates [8] は Sequential decision making problem のアイデアを組み合わせることによって、認識結果をユーザに報告するタイミング (できるだけ早く) と認識率 (できるだけ正確に) のトレードオフをモデル化した手法を提案している。一方、内田らは早期認識ブースティングと呼ばれる方法を提案した [9]。この手法では、複数シーケンスの時刻 t における観測量集合 $\{x_{i,t}\}_{i=1,\dots,N}$ だけを識別する弱識別器 h_t を学習する。弱識別器の学習では、重み伝播法によって前時刻での弱識別器で識別に失敗したシーケンスの観測量を優先して正しく識別できるように目的関数を設計する。この手法によれば、前時刻までの情報に基づく識別器の誤りを訂正するように各時刻の弱識別器 h_t を更新することによって、中途フレームまでの入力しか与えられない部分シーケンスに対してもできるだけ高い識別性能が得られるように識別器を構成することができる。

しかし、これらの既存モデルはいずれも 2 クラス識別モデルであり、より多クラスの早期認識問題への応用には課題がある。[9] においては、 $K > 2$ クラスの識別問題を $\binom{K}{2}$ 個の 2 クラス問題に分解することによって解決した。しかし、この方法は時間・空間計算量の観点で明らかに非効率的である上に、学習される識別モデルは $\binom{K}{2}$ 種の 2 クラス問題にそれぞれ最適化されたものにすぎず、 K クラスの識別問題として最適化されたモデルは得られない。

そこで我々は早期認識ブースティング法を多クラス識別問題へと適用可能とする新しい方法を提案する。我々の多クラス早期認識ブースティング法は K クラスの識別問題に対してその学習コストをクラス数 K に対して線形に抑えることができる。また、学習される識別モデルは個別最適化ではなく K クラス全体での識別性能を考慮して学習される。まず次節において背景となる既存手法について簡単に説明した後に、第 3 節で多クラス早期認識ブースティング法を提案する。その後、手書き文字認識データおよび動画データにおける運転者動作認識を例としてその性能を実験で確認する。

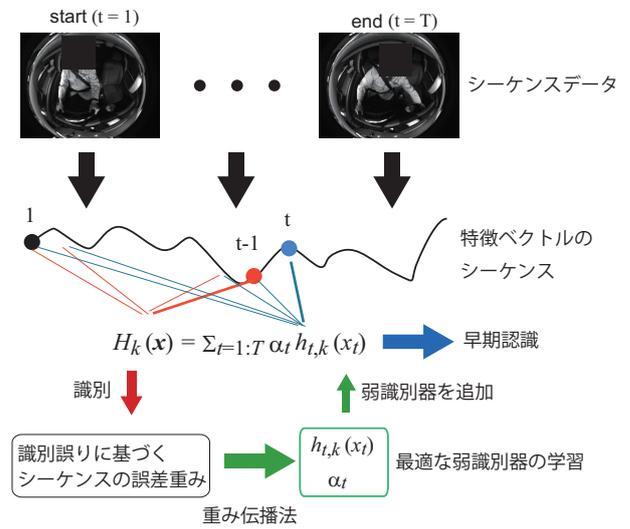


図 1 多クラス早期認識ブースティングの概念図。学習時には重み伝播法を利用した誤り訂正学習により弱識別器 h と識別重み α を決定する。学習された弱識別器の集合からなる強識別器 H はシーケンスの早期認識を可能とする。

2. 既存手法

2.1 Adaboost

まず、最も基本となる 2 クラスの Adaboost [7] について簡単に説明する。観測特徴ベクトルを $x_i \in \mathbb{R}^d$ 、そのクラスラベルを $y_i \in \{1, -1\}$ とする。訓練データセット $\mathcal{D} = \{x_i, y_i : i = 1, \dots, N\}$ が与えられた時、Adaboost の学習問題では以下の強識別器 $H: \mathbb{R}^d \rightarrow \{1, -1\}$ を決定する。

$$y_i = H(x_i) = \text{sign} \left(\sum_{m=1}^M \alpha_m h_m(x_i) \right). \quad (1)$$

ここで $h_m: \mathbb{R}^d \rightarrow \{1, -1\}$ は m 番目の弱識別器である。 $\alpha_m \in \mathbb{R}$ はその識別重みであり、 M は弱識別器の総数である。これらの変数は $m = 1$ より m を一つずつインクリメントしながら式 (2)-(4) に従って学習される。

$$h_m = \arg \min_h \epsilon_m, \quad (2)$$

$$\alpha_m = \frac{1}{2} \log \left(\frac{1 - \epsilon_m}{\epsilon_m} \right), \quad (3)$$

$$\epsilon_m = \sum_{i: y_i \neq h_m(x_i)} D_m(i). \quad (4)$$

式 (4) は弱識別器 h_m による識別が誤ったサンプル集合に対して、 $D_m(i) \in \mathbb{R}$ を足し込んでいる。 $D_m(i)$ は m 番目の弱識別器学習におけるサンプル i の誤差重みであり、以下のように計算される。

$$D_m(i) \propto D_{m-1}(i) \exp(-\alpha_{m-1} y_i h_{m-1}(x_i)). \quad (5)$$

式 (5) によれば、観測量 x_i の h_{m-1} による識別結果が真のクラスラベル y_i と異なると誤差重み $D_m(i)$ が増加し、逆に正しければ誤差重みは減少する。この学習過程により、

それぞれの弱識別器は、大半の識別容易なサンプルを識別するもの、大半のサンプルには弱い識別困難な少数のサンプルの識別が得意なものなどに特化し、結果的に強力な強識別器を得ることが期待できる。

2.2 “Earlyboost”

続いて、早期認識ブースティング法 [9] を説明する。なお、本稿ではこの手法を早期認識ブースティング法あるいは“Earlyboost”と呼ぶこととする。

i 番目の観測時系列シーケンスデータを $x_i = \{x_{i,t} \in \mathbb{R}^d\}$ と表現する。シーケンス数は N であり、 $i \in \{1, 2, \dots, N\}$ でインデックスする。シーケンス長は T で一定と仮定し、 $t \in \{1, 2, \dots, T\}$ でインデックスする。 $y_i \in \{1, -1\}$ はこのシーケンスサンプル x_i に対するラベルを表す。

Earlyboost の学習では、以下の強識別器 H を訓練データ $\mathcal{D} = \{x_i, y_i\}$ から学習する。

$$y_i = H(x_i) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x_{i,t}) \right). \quad (6)$$

ここで $h_t: \mathbb{R}^d \rightarrow \{1, -1\}$ は各シーケンスの時刻 t の観測量 $\{x_{i,t}\}_{i=1, \dots, N}$ だけを入力とする、**frame-wise** の弱識別器である。学習時には、 $t = 1$ より t を 1 ずつインクリメントしながら弱識別器 h_t と識別重み $\alpha_t \in \mathbb{R}$ を以下のように決定する。

$$h_t = \arg \min_h \epsilon_t, \quad (7)$$

$$\alpha_t = \frac{1}{2} \log \left(\frac{1 - \epsilon_t}{\epsilon_t} \right), \quad (8)$$

$$\epsilon_t = \sum_{i: y_i \neq h(x_{i,t})} D_t(i). \quad (9)$$

弱識別器 h_{t-1} と h_t は、入力とする観測特徴ベクトルの集合が異なる。従って、これらを接続するためには何らかの工夫が必要である。[9] では、誤差重みを個別のサンプル $x_{i,t}$ ではなく各時系列シーケンス x_i 全体にかかるものと捉え、以下のように時刻 $t-1$ から時刻 t へ識別情報を伝播させることを提案した。

$$D_t(i) \propto D_{t-1}(i) \exp(-\alpha_{t-1} y_i h_{t-1}(x_{i,t-1})). \quad (10)$$

この重み伝播法によって、時刻 $t-1$ において誤識別されたシーケンスを優先して正しく識別するように h_t を学習してゆく。

識別時は、上記で決定された強識別器 H を用いて、テスト (部分) シーケンス $x_j = \{x_{j,t}, t = 1, 2, \dots, L\}$ のクラスラベル y_j を以下のように決定する。

$$y_j = H(x_j) = \text{sign} \left(\sum_{t=1}^L \alpha_t h_t(x_{j,t}) \right) \quad (11)$$

式 (10) で示される重み伝播法は前時刻での誤りを最小化するように弱識別器を構成している。従って、学習され

た H はシーケンス長が短い ($L \leq T$) 部分シーケンスに対してもできる限り正確な識別結果を与える早期認識器になっていることが期待される。

3. 提案法: Earlyboost.MH

3.1 アルゴリズム

このように、Earlyboost は時系列シーケンスの早期認識のために Adaboost を自然に再解釈したモデルとなっている。しかし、現実には早期認識を必要とする課題は多クラス問題であることが多い。特に大規模語彙音声認識 [4] や手書き漢字認識などの場合には認識すべきクラス数は数百から数千と多数にのぼるため、[9] で採用されたように $\binom{K}{2}$ 個の 2 クラス問題に分割して対処することは識別時・認識時の双方において非効率である。また、多数の 2 クラス識別器は独立に学習されるため、 K クラス識別器として最適化されていない。

そこで、本節では我々が考案した多クラス早期認識ブースティング法 Earlyboost.MH (図 1) を説明する。我々のモデルは Adaboost.MH [10] で提案された Adaboost の多クラス化の試みを Earlyboost に直接適用したものである。

与えられる学習データは $\mathcal{D} = \{x_i, y_i\}, i = 1, 2, \dots, N$ と表現する。 x_i は d 次元特徴ベクトルのシーケンスであり、 $x_i = \{x_{i,t} \in \mathbb{R}^d\}, t = 1, 2, \dots, T$ である。また、 y_i は i 番目のシーケンスのクラスラベルである。ただし、その値域はこれまでと大きく異なり、 K クラス識別問題に応用するため $y_i \in \{1, 2, \dots, K\}$ と定義する。

弱識別器 $h_{t,k}(x): \mathbb{R}^d \rightarrow \{1, -1\}$ は時刻 t の観測サンプル $x_{1:N,t}$ のみを入力として受け取る。出力は、入力された x がクラス k に属すると識別した場合は 1、それ以外のクラスに属すると識別される場合には -1 を出力する。また、新たに $g_k(y): \{1, 2, \dots, K\} \rightarrow \{1, -1\}$ という関数を定義する。この関数はクラスラベル y を入力として受け取り、 $y = k$ ならば 1 を、そうでなければ -1 を出力する関数である。

Earlyboost.MH では、下記の強識別器を学習する。

$$y_i = \arg \max_{k=1,2,\dots,K} H_k(x_i) = \arg \max_{k=1,2,\dots,K} \sum_{t=1}^T \alpha_t h_{t,k}(x_{i,t}). \quad (12)$$

すなわち、弱識別器とその識別重み $\{h_{t,k}, \alpha_t\}$ を学習データ \mathcal{D} を用いて学習し、 K 個の (2 クラス) 強識別器 H_k を生成する。そして、識別フェーズではそれらのスコアを比較することによってテストシーケンス x_j を $y_j \in \{1, 2, \dots, K\}$ へと識別する。

具体的な学習アルゴリズムの説明を行う。Adaboost.MH [10] での多クラス定式化に従って、まず、式 (4) 及び式 (9) における誤識別スコア ϵ の最小化ではなく、識別スコア $r_{t,k}$ の最大化によって $h_{t,k}$ を決定する。以上を数式で表現すると式 (13)、式 (14) のようになる。

$$h_{t,k} = \arg \max_h r_{t,k}, \quad (13)$$

$$r_{t,k} = \sum_{i=1}^N g_k(y_i) h_{t,k}(x_{i,t}) D_t(i, k). \quad (14)$$

上式より明らかなように， $r_{t,k}$ は $g_k(y_i)$ と $h_{t,k}(x_{i,t})$ の符号が一致した場合，すなわち与えられた観測量 $x_{i,t}$ がクラス k に属するか否かの識別結果が正しい場合に増加する．各シーケンス i に対する識別結果は誤差重み $D_t(i, k) \in \mathbb{R}$ によって重み付けされている．

この重みは Earlyboost に従って時刻 $t-1$ より重み伝播に従って定義される．

$$D_t(i, k) \propto D_{t-1}(i, k) \exp(-\alpha_{t-1} g_k(y_i) h_{t-1,k}(x_{i,t-1})). \quad (15)$$

式 (10) と異なり，クラス k に依存して重みを伝播させる．これは，クラス k' とクラス k'' の識別時において，各シーケンスの重要度は自然に異なる事が想定されるからである．上記の誤差重み $D_t(i, k)$ が与えられたもとの，全ての $k \in \{1, 2, \dots, K\}$ について $h_{t,k}$ を $r_{t,k}$ に関して最適化する．

その後，識別重みは以下のように $h_{t,k}$ について最大化された $r_{t,k}$ を用いて計算する．

$$\alpha_t = \frac{1}{2} \log \left(\frac{1 + \sum_k r_{t,k}}{1 - \sum_k r_{t,k}} \right). \quad (16)$$

以上の計算を $t=1$ より t を 1 ずつインクリメントしながら $t=T$ に到達するまで繰り返すことで学習が完了する．

識別フェーズでは，上記で決定された強識別器 H_k を用いて，テスト (部分) シーケンス $x_j = \{x_{j,t}, t=1, 2, \dots, L\}$ のクラスラベル y_j を以下のように決定する．

$$y_j = \arg \max_k H_k(x_j) = \arg \max_k \sum_{t=1}^L \alpha_t h_{t,k}(x_{j,t}) \quad (17)$$

強識別器 H_k は式 (15) による学習の効果により，早期認識のために最適化されている．従って，入力されるテストシーケンス x_j が $t=1 \sim L (\leq T)$ の範囲の観測量のみから構成されていても，できる限り正確な多クラス識別が行われると期待できる．

3.2 提案法の特徴

提案法では，シーケンス全体に対する”良い”特徴量を開発する代わりに，各時刻固有の弱識別器を学習し，その集合を利用することでシーケンス全体の強識別器を構成した．各弱識別器の識別能力は高くないが，それらを組み合わせることによって識別を行うアプローチは，シーケンスに対するブースティングの設計理念として自然な発想と言える．また，式 (13), 式 (14) にあるように，弱識別器 $h_{t,k}$ を各時刻ごとに定義しているのだから，時系列シーケンスの特徴ベクトルの分布が時刻発展に応じてダイナミックに変化する場合にも自然と追従することが可

能となる．弱識別器の構成方法として，前後数フレームから各弱識別器を学習するよう拡張することも可能であり，またその拡張によってモデル全体の学習方法等は大きく変化しない事を注意しておく．

本稿で提案した Earlyboost.MH は Earlyboost に対し 2 つの点で優れている．まず第一は識別重み α_t の計算方法である．式 (15) にあるように，Earlyboost.MH では全 K クラスの識別スコア $r_{t,k}$ の総和を用いて α_t を決定している．従って，学習される強識別器 H_k は，各クラス k のための one vs. all の識別性能ではなく K クラス識別問題の性能を向上するように設計されていることになる．一方通常の Earlyboost [9] では「クラス k とクラス k' 」の 2 クラス識別に最適化された強識別器を多数用意することになる．従って， K クラス識別問題としての最適化は図られていない．

第二に，Earlyboost.MH では必要となる学習コストがクラス数 K に対して線形に抑えられる．具体的には，長さ T の時系列シーケンスに対して K クラスの識別を行う場合，学習が必要となる弱識別器 $H_{t,k}$ は TK 個であり，また計算が必要な識別重み α_t の個数は T 個である．一方 [9] に示された方法では，弱識別器は $T \binom{K}{2}$ 回の学習が必要であり，識別重みの個数も $T \binom{K}{2}$ 個必要となる．これは時間計算量，空間計算量ともに Earlyboost.MH より大幅に増加してしまうことを意味しており，特にクラス数の多い漢字認識や大規模語彙音声認識などの問題では致命的になりうる．この点でも Earlyboost.MH は多クラス識別に適した識別モデルであると言える．

4. 実験

実験では，2 種類のタスクにおいて Earlyboost.MH の性能を確認する．一つはオンライン手書き数字認識，もう一つは動画像を用いた運転者動作認識である．

なお，運転者動作認識実験のセットアップは [11] を参考にした．

4.1 実験データ

オンライン手書き数字認識実験では，Earlyboost の原論文 [9] で利用されたデータと同じく Ethem Alpaydin Digit データベースを利用した．このデータベースでは孤立数字 (“0”-“9”) の手書き軌跡データが 800 サンプル提供されている．

このデータはシーケンスの伸縮が非常に大きいため [9] にならって単純な線形補間によって全てのシーケンスの長さを $T=50$ に揃える前処理を行った．特徴量はペン先位置の 2 次元座標，および前フレームからの差分による速度特徴を用いた．これらの特徴を白色化した $d=4$ 次元の実数ベクトル系列が入力される特徴シーケンスである．

識別する文字は数字 (“0” - “9”)，すなわち $K=10$ クラスの識別問題である．一つの孤立文字の軌跡を一つの

シーケンスとみなす．学習では弱識別器として $D_{i,k}$ による重み付きに従ってリサンプリングされた kNN($k=5$) を利用した．学習および識別は 6-fold cross validation を用いて行った．

運転者動作認識実験では，実寸大のドライビングシミュレータの運転席にカメラを設置し，運転者の運転中の動作を録画，認識する実験を行う．本実験では 7 名のドライバに，それぞれ 30 回の運転シミュレーションを実施させ，その様子をバックミラー付近に設置した一般の単眼ビデオカメラで撮影した．フレームレートは 60FPS である．

画像より抽出する特徴量は [1] に従った．すなわち，オプティカルフローを用いて運転者の両手首，両肘，両肩の位置を追跡し，その軌跡を観測量とする．本実験では，そのうち車内の動作に大きな影響を及ぼすと思われる右手首，左手首および左肘の画像座標 (w, h) を特徴量として採用した．従って，入力される観測ベクトル x_{it} は $d=6(=2 \times 3)$ 次元の正数ベクトルである．

識別する動作のクラス数は $K=12$ である．いずれもステアリングハンドルを握った状態から，エアコンやミラーなど K 種の車内設備を操作する，という動作を認識する．収集された時系列シーケンスは，まず 1 回の動作ごとのシーケンス x_i に分割したのち，全てのフレームについて $y_i = 1, 2, \dots, K$ のうち一つの値を付与する．以上の操作は全て人手で行った．得られたシーケンス数は運転者ごとに異なり，またシーケンス長の正規化も行っていない．実験では弱識別器 h として decision stump 識別器 [12] を用いた．学習および識別は同じドライバーの時系列シーケンスデータを 6-fold cross validation を用いて行った．

双方の実験において，識別率はシーケンス単位で計算する．全 K 種の動作それぞれについてシーケンス識別率を算出し， K クラスの平均識別率をもって学習された識別器を評価する．

4.2 比較手法

提案法の比較対象としては，オリジナルの Earlyboost [9] を用いた．Earlyboost は 2 クラス用識別器のため，本実験では多クラスへの対応が必要となる．本実験では計算量の観点で公平を期すため [9] とは異なる単純な方法による多クラス対応を行った．すなわち K 種のクラスそれぞれに対して one vs. all タイプの 2 クラス強識別器 (式 (6)) を生成する方法である．

各 2 クラス識別器は時系列シーケンス x が「クラス k に属する ($y=1$)」か「他のクラスに属する ($y=-1$)」を学習する．ただし，全 K 個の強識別器はそれぞれ独立に学習する点が提案法である Earlyboost.MH との相違点となる．識別フェーズではテストシーケンス x を全ての強識別器へと入力する．ただし今回は式 (11) のように sign 関数を用いず式 (17) のように弱識別器出力の重み付きス

コアのみを計算し，その値が最も大きいクラスへと識別する．このスコアの大きさは各強識別器の識別の確信度を表している [10] ので，Earlyboost が正しく K クラス識別問題に対しても適用可能であるならば，真のクラスの強識別器が最も高いスコアを出力するはずである．この方法によれば，Earlyboost と Earlyboost.MH の計算量はほぼ同等となるため公正な性能比較が可能である．

なお [9] の論文では明確には示されていないが，より正確に 2 クラスペア間識別器による K クラス問題の識別モデルを構成する場合には，誤り訂正符号にのっとったクラス分けを行う方法がより適切であると考えられる [13]．ただし，この場合も元のクラス数 K 以上の強識別器が必要となる．

4.3 実験結果

最初に，オンライン手書き数字認識データの実験結果を示す．図 2 は $T=50$ フレームに正規化されたシーケンスデータ集合に対する $K=10$ クラスの平均識別率である．強識別器 H_k に対して，テスト時系列シーケンスを $L(\leq T)$ フレーム目まで入力したときに，どれだけ正確にクラスを識別できるかを測ることで，早期認識モデルとしての性能を評価する．縦軸は平均識別率，横軸が入力されるシーケンス長，すなわち時刻フレーム数 L を表す．太青線は提案法である Earlyboost.MH による識別率の時間発展を示し，赤細線は既存手法である Earlyboost による識別率の推移である．

最初の 10 フレームほどでは提案法がわずかに劣る結果になっているが，その後提案法が既存手法を上回る認識率を示している．特に 40 ~ 45 フレーム付近では，10% 以上の識別率向上を記録している．これは式 (16) にあるように多クラス間の識別スコアを利用した学習則を設計しているからであると考えられる．この結果より， K クラスの間の関係性を無視しない提案法のモデルの有効性が良く証明されたと考えられる．

次に，運転動作識別実験の結果を示す．図 3 は 7 人の運転者の平均識別率の時間発展のグラフである．太青線は提案法である Earlyboost.MH による識別率の時間発展を示し，赤細線は既存手法である Earlyboost による識別率の推移である．図より明らかなように，Earlyboost.MH による認識率は既存法を大幅に上回っており，手書き数字認識実験の場合に比べてもその差は歴然としている．

これはおそらくデータの特性によるものと思われる．オンライン手書き数字認識の場合，書かれた数字 (クラス) によってシーケンスの始点がある程度分離されていると思われる．また，その後のペン先の移動シーケンスに関しても，例えば“1”と“3”などはまったく相関を持たないであろう．すなわち，手書き文字認識データにおいては各クラスのシーケンスの時間発展にある程度の独立性が予想できる．従って，提案法のように K クラスの識別能力の関係をモデルにとりこまなくてもある程度良

識別率の時間発展グラフ(手書き文字)

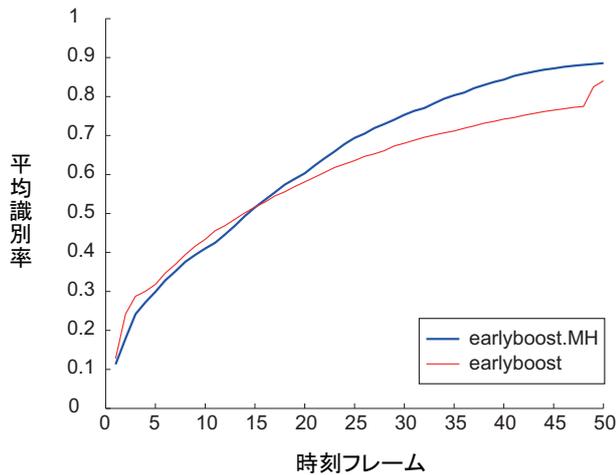


図2 手書き文字データの平均識別率の時間発展グラフ．太線は Earlyboost.MH, 細線は Earlyboost による結果．横軸は識別を行う時刻フレーム，縦軸は平均識別率を表す．

識別率の時間発展グラフ(運転動作)

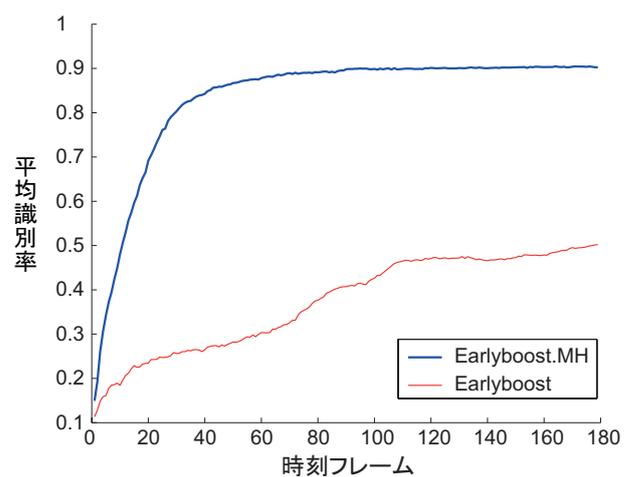


図3 Earlyboost.MH(太実線) とオリジナルの Earlyboost(細破線) による7 運転者の平均動作識別率の時間発展グラフ．横軸は時刻フレームを，縦軸は平均識別率をそれぞれ表す．

好な識別率が得られるものと思われる．これは，とくに認識初期において Earlyboost が提案法を若干上回っていることから予想される．

一方，動画像からの運転者動作認識実験データでは，全てのシーケンスがステアリングを握っている状態から始まっている．これは車内機器操作の早期認識(予測)というタスクからは自然な要請であるが，このために認識初期では全てのクラスのサンプル分布が分離不可能なほど重複する．また，例えば”ナビゲーションの操作”と”エアコンの操作”では，機器へと向かう手先位置の移動経路もかなり似通ってくる．すなわち，多クラス間の識別において，クラスごとの分布の相関が非常に強いデータである．従って，各クラスの強識別器を独立に学習する Earlyboost 法では高い識別率を得られない一方，全クラスの相関を考慮して学習を行う Earlyboost.MH 法は良好な認識結果を得ることができたものと考えられる．

すなわち，提案法は各クラスの間に関係が強く疑われるようなデータにおいては非常に良好に性能を発揮できると結論できる．

次に，図4では重み伝播の効果の大きさを観察する．太青線は提案した Earlyboost.MH による識別率の推移である．細赤線は式(15)に示す重み伝播を行わず，全ての $D_t(i, k)$ を等しい値 $(\frac{1}{NK})$ に固定した場合の識別率推移である．図より明らかなように，最終的な識別率，さらに少ないフレーム数における早期認識の識別率の向上率の両点において重み伝播の効果が顕著であることが見て取れる．

最後に，ある運転者のデータに対して学習された識別重み α_t の時間発展を図5に示す．興味深いことに，識別重み α_t は時間経過とともに単調に減少している． α_t は式(16)によって計算されるため， $\sum r_{t,k}$ が減少すると α_t も減少する．このことは， $h_{t,k}$ によって得られる識別ス

コア $r_{t,k}$ が単調に減少していくことを示唆する．これはある意味で妥当である．なぜなら Earlyboost.MH は早期認識を実現するために前時刻の弱識別器の誤差を修正するように学習を進めるからである．よって，時刻が進むごとに新しい弱識別器によって修正可能な誤りは減っていく．これが識別スコア $r_{t,k}$ の減少を予想させることは明らかであろう．

以上の結果より，提案する Earlyboost.MH は，時系列シーケンスの早期認識，とくに多クラスの識別に有効であることが示された．

識別率の時間発展(重み伝播の有無, 運転動作)

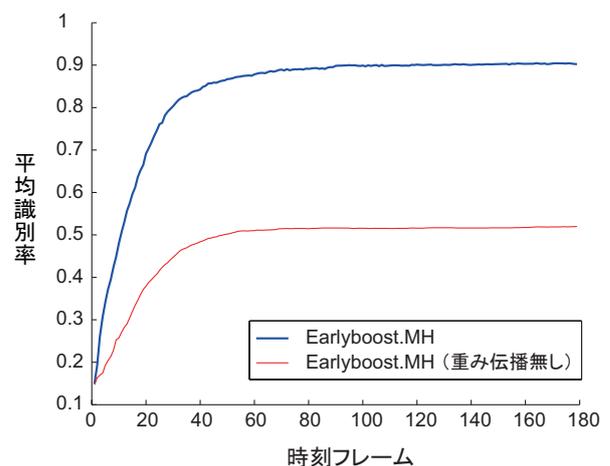


図4 7人の運転者による車内動作の平均動作識別率の時間発展グラフ．太実線は Earlyboost.MH, 細破線は重み伝播を行わない Earlyboost.MH による結果．横軸は識別を行う時刻フレーム，縦軸は平均識別率を表す．

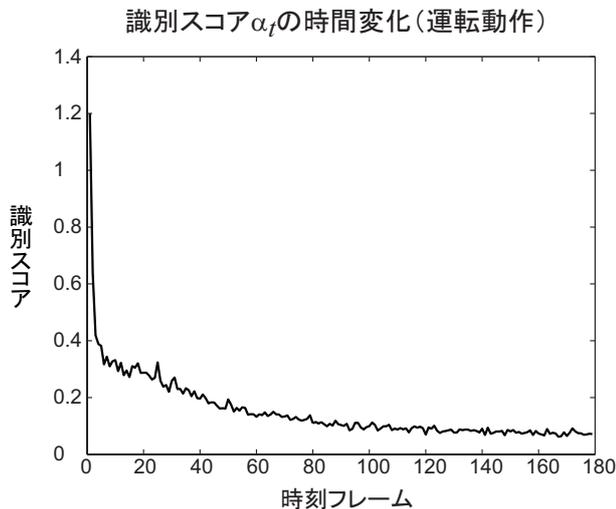


図5 ある運転者における識別重み α_t の学習結果. 横軸は時刻フレーム t を, 縦軸は α_t の値を示す.

5. 結 論

本稿では, 多クラスの時系列シーケンスデータを早期認識するための新しいブースティング法を提案した. 我々は [9] によって提案された重み伝播法と [10] で提案された Adaboost の多クラス拡張手法を組み合わせた Earlyboost.MH モデルを考案した. Earlyboost.MH はその定義は非常に簡潔で, また学習および識別に必要な計算量も少なくなっている. 実験では, オンライン手書き数字認識の問題, 及び動画像を利用した自動車の運転者動作の認識問題に対し, 既存の Earlyboost モデルよりも高い認識性能を示すことを確認した.

提案手法は時系列シーケンスの種類を問わず, 様々なデータの早期認識問題に応用可能であると考えられる. 従って, 今後は手書き漢字認識データやジェスチャー認識, 音声認識など様々な時系列データに提案法を適用していきたいと考えている. また, Adaboost は確率・統計モデルとしても研究が進んでいる [14], [15]. こういった先行研究を参考にして, 多クラス早期認識モデルとしての理論的な最適性の検証も行いたいと考えている.

運転者動作データベースを提供していただいた株式会社デンソー様に感謝いたします.

文 献

[1] Y. A. Sheikh, A. Datta and T. Kanade: "On the sustained tracking of human motion", Proceedings of 8th IEEE International Conference on Automatic Face and Gesture Recognition (FG) (2008).

[2] 森, 内田, 倉爪, 谷口, 長谷川, 迫江: "ジェスチャーの早期認識・予測ならびにそれらの高精度化のためのネットワークモデルに関する検討", 画像の認識理解シンポジウム予稿集 (MIRU2005), pp. IS3-106 (2005).

[3] C. Bahlmann and H. Burkhardt: "The writer independent online handwriting recognition system frog on hand and cluster generative statistical dynamic time warping", IEEE Transactions on Pattern Analysis and Machine Intelligence,

26, 3, pp. 299-310 (2004).

[4] T. Hori, C. Hori and Y. Minami: "Fast on-the-fly composition for weighted finite-state transducers in 1.8 million-word vocabulary continuous speech recognition", Proceedings of the 8th International Conference on Spoken Language Processing (Interspeech), Vol. 1, pp. 289-292 (2004).

[5] L. R. Rabiner: "A tutorial on hidden Markov models and selected applications in speech recognition", Proceedings of the IEEE, 77, 2, pp. 257-286 (1989).

[6] V. N. Vapnik: "The Nature of Statistical Learning Theory", Springer (1995).

[7] Y. Freund and R. E. Schapire: "A decision-theoretic generalization of on-line learning and an application to boosting", Journal of Computing Systems and Science, 55, 1, pp. 119-139 (1997).

[8] J. Sochman and J. Matas: "Waldboost learning for time constrained sequential detection", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 第 2 巻, pp. 150-156 (2005).

[9] S. Uchida and K. Amamoto: "Early recognition of sequential patterns by classifier combination", Proceedings of the 19th International Conference on Pattern Recognition (ICPR) (2008).

[10] R. Schapire and Y. Singer: "Improved boosting algorithms using confidence-rated predictions", Machine Learning, 37, 3, pp. 297-336 (1999).

[11] 坂野, 山田, 石黒: "誤り訂正学習による動作認識の試み", 情報処理学会論文誌: 数理モデル化と応用 (TOM), 2, 3, pp. 153-161 (2009).

[12] C. M. Bishop: "Pattern Recognition and Machine Learning", Springer-Verlag New York (2006).

[13] T. G. Dietterich and G. Bakiri: "Solving multiclass learning problems via error-correcting output codes", Journal of Artificial Intelligence Research, 2, pp. 262-286 (1995).

[14] J. Friedman, T. Hastie and R. Tibshirani: "Additive logistic regression: A statistical view of boosting (with discussion)", Annals of Statistics, 28, 2, pp. 337-407 (2000).

[15] H. Zou, J. Zhu and T. Hastie: "New multicategory boosting algorithms based on multicategory Fisher-consistent losses", The Annals of Applied Statistics, 2, 4, pp. 1290-1306 (2008).