「マルチメディア，分散，協調とモバイル
(DICOMO2013) シンポジウム」 平成25年7月

メタ学習に基づく加速度センサからの看護師行動識別

上田 修功[1]　田中 祐典[2]　中島 直樹[3]

概要：加速度センサを用いた行動識別手法はこれまで数多く提案されているが，それらの大半は，「歩く，走る，座る」などの識別が容易な少数の行動カテゴリを対象とするか，あるいは，多数の加速度センサを体に装着するなど，実用的な手法とは言い難い．本研究では，現実的な状況下 (少数の加速度センサ) で複雑な行動を高精度に識別する手法を確立すべく，問診などの認識困難な対象も含む，病院により既定された 41 種類の看護師行動識別をターゲットタスクとし，ベイズモデルの枠組みで，メタ学習法と呼ぶ高精度なパターン識別手法を提案する．メタ学習法では，異なる条件で個別に学習した複数の識別器の識別結果から真のクラスを推定する．実験により，個々の単一の分類器，および従来のアンサンブル分類手法の識別率を顕著に上回る性能を確認した．

Actual Nursing Recognition from Acceleration Data based on Meta-Learning

Naonori Ueda[1]　Yusukei Tanaka[2]　Naoki Nakashima[3]

1. まえがき

保健医療の分野では，より良い最終アウトカムを得るためには，どのような患者状況を経るべきかを知ることが重要であり [1]，その際，患者の容態のみならず，看護師が患者に対してどのような患者行動を取ったかについての情報も有用視されている．即ち，看護師行動の情報が保健医療において，より良いアウトカムを得るための重要情報と成り得る．一方，近年，加速度センサから人の行動を自動識別する手法が数多く提案されている [2]．しかし，それらの研究の大半は，研究者の視点で予め定めた比較的識別が容易な行動クラスであったり，あるいは，識別精度を上げるべく，体のいたるところにセンサを装着するなど非現実的な問題設定である．また，識別対象のクラス数も高々十数程度である．本論文で対象とする済生会熊本病院での看護師行動データでは，看護業務の制約から加速度センサの装着部位，個数も制限され，さらに，問診など加速度センサでは識別困難なクラスも識別対象となっている．本研究では，真に有用な行動識別手法を確立すべく，上記看護師データでも適用可能な新たな行動識別手法を提案する．従来，加速度センサからの行動識別手法では，slidingwindow と呼ばれるある時間幅で，加速度データ (時系列データ) から特徴を順次抽出し，抽出した特徴に基づいて，隠れマルコフモデルやサポートベクトルマシンなどの識別器を援用するというアプローチがとられている [3]．しかし，行動クラスによっては加速度の時間変化が大幅に異なる場合があり，最適な時間幅の slidingwindow を定めることが困難となる．また，特徴抽出においても，クラス毎に識別に有効な特徴が異なり，特に，クラス数が数十になるとその問題が顕著になる．即ち，あるクラスの識別精度を向上させると，別のクラスの識別精度が低下するなどの問題が生じる．このような問題に対し，機械学習の分野で開発されたアンサンブル識別手法 [4],[5],[6],[7] の適用が考えられる．アンサンブル識別手法とは，特徴や識別器を複数種類用意し，各々について各識別器を個別に学習し，クラスが未知のデータに対しては，学習済の複数の識別器の識別結果を何らかの方法で統合してクラス決定をする手法である．最

[1] 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
[2] NTT サービスエボリューション研究所
神奈川県横須賀市光の丘 1-1
[3] 九州大学病院メディカルインフォメーションセンター
福岡県福岡市東区馬出 3-1-1

も単純な統合手法として多数決決定がある [4]．しかし，従来のアンサンブル識別手法は，各識別器の性能がある程度良いことを前提としているため，識別性能の低い識別機が存在すると，アンサンブルの効果が期待できないという問題がある．この問題は，応用での特徴表現の質に大きく依存する．つまり，応用分野において識別に有効な特徴抽出が実現できていなければ，単一の識別器の性能には限界が生じる．また，前述した様に，識別器の性能はクラス毎に異なるので、全クラスに平均的に識別精度が高いもののみでは，結果的にアンサンブルによる精度向上が期待できない．

この問題を解決すべく，本研究では，学習データに対し，各識別器の識別結果と真のクラスラベルを用いて，複数の識別器の振る舞い (識別の癖) を学習する手法を検討し，アンサンブル識別器の中であるクラスに対し，誤識別する識別器が多数混在していたとしても，その情報も正しい識別に有効利用する「メタ学習法」を新たに提案する．提案手法は，ベイズモデルに基づき，クラス毎にどの識別器がそのクラス識別に有効か否かを自動判別する．重要な点は，その有効か否かの基準は，必ずしも正しく識別しているか否かではなく，識別結果に一貫性があるか否かである点である．直観的には，ある識別器がクラス 1 のサンプルを高頻度でクラス 2 と識別するのであれば，その識別器がクラス 2 と識別した場合，実はクラス 1 である確率が高いという考え方をモデル化している．

2. ベイズアンサンブルモデル

2.1 潜在変数の導入

提案モデルでは，学習データの各々に対して複数の識別器の識別結果が与えられているものとする．例えば，図 1 では，3 クラス問題における，10 個の識別器による各クラス 5 個の学習データでの識別結果を表している．つまり提案モデルでは，この識別結果行列が所与としてモデル化するため，学習の際にどのような特徴でどのような識別器が使われたかは問題にしていないことに注意．そして，この識別行列から，メタ識別器を構成し，クラスが未知のサンプル (テストデータ) に対しクラス識別を行う．

この時，各識別器は前述した様に，あるクラスのサンプルは比較的正しく識別できるが，別のクラスのサンプルは誤識別し易いという偏った識別結果となる場合がある．この傾向はクラス数が多くなるほど顕著である．それ故，全ての識別器の識別結果を対等に扱うのは適切ではない．そこで，提案モデルでは，ある識別器が各クラスの識別に有効か否かを示す潜在変数を導入する．

ただし，あるクラスにおいて多くの誤識別する識別器が必ずしもそのクラスの識別に有効でないとは限らない．つまり，その識別器があるクラスに対し，殆ど全て誤識別でも識別結果に一貫性があれば (図 1 の網掛けの部分)，その識別器はそのクラスの識別に有効と言える．例えば，図 1 の識別器 2 は全クラスの識別には有効であるが，識別器 10 はクラス 1，2 の識別には有効であるが、クラス 3 のサンプルに対しては，一貫性のない誤識別をしているのでクラス 3 の識別には有効でないと考える．

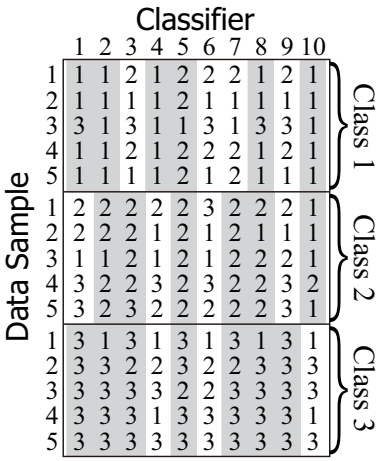

図 1 複数の識別器による識別結果の例

では，この潜在変数を用いてどのようにモデル化するかについて次節で詳しく説明する．ただし，本論文は行動識別への応用に焦点をあてており，式や計算の詳細については論文の本質ではないため省略する．

2.2 提案モデル: Bayesian Ensemble Model (BEM)l

今，K クラス問題を対象に J 個の識別器を考える．J 個の識別器は，個別に学習済とする．J 個の識別器は，同一の識別器を異なる特徴表現で学習して得ることもでき，あるいは異なる識別器で学習して得ることも可能である．$\mathbf{C}^{(k)}$ を識別行列とする．明らかに，行列のサイズはクラス ω_k に対し，$N^{(k)} \times J$ となり，第 (i,j) 要素はクラス ω_k の第 i 学習データに対する第 j 識別器の識別結果を表す．明らかに，$c_{i,j}^{(k)} \in \{1,\ldots,J\}$．ここで，$N^{(k)}$ はクラス ω_k の学習データ数を表す．また，$\mathbf{C} = \cup_{k=1}^{K} \mathbf{C}^{(k)}$ とする．

テストフェーズでは，まず，テストデータの各々に対し，学習済の J 個の識別器で識別する．$c_m^* = (c_{m,1}^*,\ldots,c_{m,J}^*)$ を第 m テストデータに対する J 個の識別器による識別結果を表すものとする．すなわち，ここでの目的は，$\mathbf{C}$ から最良のメタ識別器を構築することとなる．以下では，i, j, k および m は各々，学習データ，識別器，クラス，そしてテストデータのインデックスを表すものとする．

図 1 に示す様に，$\mathbf{c}_i^{(k)} = (c_{i,1}^{(k)},\ldots,c_{i,J}^{(k)})$ は，クラス ω_k におけるサンプル i に対する J 次元の離散特徴ベクトルと見なすことができる．J 個の識別器が統計的に独立とすると，クラス ω_k の確率分布はナイーブベイズモデル：

$$P(\mathbf{c}_i^{(k)}|\omega_k) = \prod_{j=1}^{J} P(c_{i,j}^{(k)}|\omega_k) \qquad (1)$$

と書ける．そして，$c_{i,j}^{(k)}$ は $1,...,K$ のいずれかの値をとるので，$P(c_{i,j}^{(k)}|\omega_k)$ は K-離散 (Discrete) 分布となる．

しかし，前述した様に，識別結果は必ずしも正しくないので，このような単純なモデル化では実験で示す様に不十分である．そこで，前節で説明した，第 j 識別器がクラス ω_k に有効か否かを示す 2 値の潜在変数 $r_j^{(k)} \in \{0,1\}$ を導入する．$r_j^{(k)}=1$ のとき，$c_{i,j}$ はクラス ω_k に関連するとし，クラス ω_k に属すサンプルの第 j 次元の特徴はクラス k に固有の分布に従って生成されたと仮定する．一方，$r_j^{(k)}=0$ のとき，$c_{i,j}$ はクラス ω_k に関連しないとし，クラス ω_k に属すサンプルの第 j 次元の特徴はクラスに依らない分布に従って生成されたと仮定する．$r_j^{(k)}$ の事前分布としてベータ・ベルヌーイ分布を仮定する．

$$\lambda_j|a,b \sim \mathrm{Beta}(\lambda_j;a,b)$$
$$r_j^{(k)}|\lambda_j \sim \mathrm{Bernoulli}(r_j^{(k)};\lambda_j)$$

$\mathrm{R}=\{r_j^{(k)} \mid j=1,\ldots,M;\ k=1,\ldots,K\}$ と表記する．

$r_j^{(k)}=1$ の時，$c_{i,j}^{(k)}, i=1,...,N^{(k)}$ は k に依存する確率 (K 項の離散分布):

$$\boldsymbol{\phi}_j=(\theta_{j,1}^{(k)},\ldots,\theta_{j,K}^{(k)}),\quad \theta_{j,l}^{(k)}\geq 0\quad \sum_{l=1}^{K}\theta_{j,l}^{(k)}=1 \qquad (2)$$

から生成されると仮定する．また，確率ベクトル $\boldsymbol{\theta}_j^{(k)}, \boldsymbol{\phi}_j$ の事前分布として共役事前分布であるディリクレ (Dirichlet) 分布を仮定する．以上より，観測データ $c_{i,j}^{(k)}$ の生成モデルは以下となる．

$$\begin{aligned}
&\lambda \sim \mathrm{Beta}(a,b)\\
&\boldsymbol{\phi} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha})\\
&r_j^{(k)}|\lambda \sim \mathrm{Bernoulli}(\lambda), \forall k,j\\
&\boldsymbol{\theta}_j^{(k)} \sim \mathrm{Dirichlet}(\boldsymbol{\beta}_j^{(k)}), \forall k,j,\\
&c_{i,j}^{(k)}|r_j^{(k)},\boldsymbol{\theta}_j^{(k)},\boldsymbol{\phi} \sim \mathrm{Discrete}((\boldsymbol{\theta}_j^{(k)})^{r_j^{(k)}}(\boldsymbol{\phi})^{1-r_j^{(k)}}), \forall k,j,i
\end{aligned} \qquad (3)$$

ここで，$\boldsymbol{\alpha}=\{\alpha_l\}_{l=1}^K$ および $\boldsymbol{\beta}_j^{(k)}=\{\beta_{j,l}^{(k)}\}_{l=1}^K$ はディリクレ分布のハイパーパラメータであり，実験では交差検定法でデータから決定している．

$\{\mathbf{c}_i^{(k)}\}_{i=1}^K$ が統計的に独立とすると，結合分布は以下と書ける．

$$\begin{aligned}
P(\mathrm{C}|\mathrm{R},\Theta,\boldsymbol{\phi}) &= \prod_{k=1}^{K}\prod_{i=1}^{N^{(k)}}\prod_{j=1}^{J}\prod_{l=1}^{K}\{(\theta_{j,l}^{(k)})^{r_j^{(k)}}(\phi_l)^{1-r_j^{(k)}}\}^{\delta(c_{i,j}^{(k)},l)}\\
&= \prod_{k=1}^{K}\prod_{j=1}^{J}\prod_{l=1}^{K}\{(\theta_{j,l}^{(k)})^{r_j^{(k)}}(\phi_l)^{1-r_j^{(k)}}\}^{n_{j,l}^{(k)}}. \qquad (4)
\end{aligned}$$

ここで，${n_{j,l}}^{(k)}$ は第 j 識別器により，クラス ω_k のサンプルがクラス ω_l に識別されたサンプル数を表す．すなわち，${n_{j,l}}^{(k)}=\sum_{i=1}^{N^{(k)}}\delta(c_{i,j}^{(k)},l)$．ただし，$\delta(x,y)$ はデルタ関数で $x=y$ の時のみ 1 でそれ以外は 0 の値をとる．$\Theta=\{\boldsymbol{\theta}_j^{(k)}\}$．さらに，共役性より Θ, $\boldsymbol{\phi}$ および λ は積分消去できる．

$$\begin{aligned}
P(\mathrm{C}|\mathrm{R};\boldsymbol{\alpha},\boldsymbol{\beta}) &= \int P(\mathrm{C}|\mathrm{R},\boldsymbol{\phi},\Theta)p(\boldsymbol{\phi};\boldsymbol{\alpha})p(\Theta;\boldsymbol{\beta})d\boldsymbol{\phi}d\Theta\\
&= \left(\frac{\Gamma(\alpha_\bullet)}{\prod_l\Gamma(\alpha_l)}\frac{\prod_l\Gamma(\sum_{k,j}\delta(r_j^{(k)},0)n_{j,l}^{(k)}+\alpha_l)}{\Gamma(\sum_{k,j}\delta(r_j^{(k)},0)N^{(k)}+\alpha_\bullet)}\right)\\
&\quad\left(\prod_{k=1}^{K}\prod_{j=1}^{J}\frac{\Gamma(\beta_{j,\bullet}^{(k)})}{\prod_l\Gamma(\beta_{j,l}^{(k)})}\frac{\prod_l\Gamma(r_j^{(k)}n_{j,l}^{(k)}+\beta_{j,l}^{(k)})}{\Gamma(r_j^{(k)}N^{(k)}+\beta_{j,\bullet})}\right) \qquad (5)
\end{aligned}$$

ここで，$\mathrm{R}=\{r_j^{(k)}\}$，$\alpha_\bullet=\sum_{l=1}^K\alpha_l$ および $\beta_{j,\bullet}^{(k)}=\sum_{l=1}^K\beta_{j,l}^{(k)}$．同様に，

$$P(\mathrm{R};a,b)=\frac{\Gamma(\sum_{k,j}r_j^{(k)}+a)\Gamma(\sum_k\sum_j(1-r_j^{(k)})+b)\Gamma(a+b)}{\Gamma(KJ+a+b)\Gamma(a)\Gamma(b)}. \qquad (6)$$

を得る．ただし，$\Gamma(x)$ はガンマ関数を表す．式 (5),(6) より全てのパラメータが積分消去された確率分布 $P(\mathrm{R}|\mathrm{C};\boldsymbol{\alpha},\boldsymbol{\beta},a,b)\propto P(\mathrm{C}|\mathrm{R};\boldsymbol{\alpha},\boldsymbol{\beta})P(\mathrm{R};a,b)$ を解析的に計算でき，これを用いて R をギブスサンプリングアルゴリズムを用いて推定できる (詳細は省略).

2.3 未知クラスサンプルの識別

$\boldsymbol{c}_m^*=(c_{m,1}^*,\ldots,c_{m,J}^*)$ を第 m サンプルに対する J 個の識別結果とする．J 個の識別器が学習できれば，c_m は第 m テストデータに J 個の識別器を適用することにより得られる．そこで，$\boldsymbol{c}_m^*$ の最適クラスは事後確率: $P(\omega_k|\boldsymbol{c}_m^*,\mathrm{C})$ を最大ならしめるクラスとすればベイズ最適な識別が実現できる．ベイズの定理より次式を得る．

$$k^*=\arg\max_k P(\omega_k|\boldsymbol{c}_m^*,\mathrm{C})=\arg\max_k P(\boldsymbol{c}_m^*|\omega_k,\mathrm{C})P(\omega_k), \qquad (7)$$

ここで，$P(\omega_k)$ はクラスの事前確率で通常は実験では一様分布とした．ただし，$P(\boldsymbol{c}_m^*|\omega_k,\mathrm{C})$ は次式で計算される．

$$P(\boldsymbol{c}_m^*|\omega_k,\mathrm{C})=\prod_{j=1}^{J}\sum_{r_j^{(k)}}P(c_{m,j}^*|r_j^{(k)},\mathrm{C})P(r_j^{(k)}|\mathrm{C}) \qquad (8)$$

式 (8) は解析的に計算困難であるが，ギブスサンプリングにより容易に計算可能（詳細略）.

3. 実験

表 2 単一識別器の看護師行動識別率 (%)

	HMM	C4.5	RF
22 クラス	42.4 (1.7)	24.3 (2.9)	32.2 (3.4)
14 クラス	54.2 (2.5)	34.1 (1.8)	45.2 (1.0)

表 1 に示す 22 種類の看護師行動データを用いて識別実験を行った．看護師の両手首，胸ポケット，腰に装着され

表 1 Nursing activities

アナムネ（看護師立位）	○ ギャッジアップ	○ ベッド運搬
ポータブル（胸部臥位正面）	○ ポータブルでの排泄介助	○ 看護記録（PC 入力）
○ 看護記録（手書き）	○ 血圧測定（仰臥時）	血糖測定（デキスター）
○ 採血（仰臥時）	持続静注 (輸液ポンプ)	○ 車いす介助
○ 手洗い	心電図モニター装着	○ 心電図モニター除去
○ 心電図検査	体位交換	○ 体重測定 (仰臥時)
○ 動脈触知	浮腫診察（仰臥時）	歩行介助
○ 時刻合わせ		

表 3 各手法での看護師行動の識別率 (%)

	BEM	MV	NB	LR+Lasso	SVM(L)	SVM(P)
22 クラス	62.8 (3.3)	42.3 (1.9)	56.8 (2.9)	51.0 (2.9)	46.6 (3.8)	43.0 (3.2)
14 クラス	75.6 (1.6)	57.7 (1.2)	71.5 (1.4)	71.8 (1.5)	70.2 (1.5)	69.8 (1.5)

た 4 つの 3 軸加速度センサから模擬患者に対する 22 種類およびそのサブセットである 14 種類の看護行動に対して，従来法と提案法を適用し識別率を比較した．データ数はトータル 1,097 であった．ただし，各看護行動は，事前に前処理?でセグメンテーションされている．学習データとテストデータから成る 5 つのデータセットを作成し，テストデータでの識別率の平均値と標準偏差を比較した．データセット当たりの平均学習データ数，平均テストデータ数は各々 39.9, 9.9 であった．

ある時間幅の sliding window を設定し，50%のオーバーラップでずらしながら特徴抽出を行った．特徴として，平均値，標準偏差，周波数領域でのエネルギーとエントロピーを採用した．これらの特徴は文献 [3] を踏襲している．これらの特徴を連結することにより時系列特徴を得る．そして識別器として隠れマルコフモデル (HMM) を用いた．

各クラスに最適な window サイズの決定は困難故，時間幅を 2,4,6...,100 として 50 種類の特徴表現を構成し，各特徴表現で HMM を学習した．その内，テストデータで最良の識別率を有する HMM(window サイズが 48 の特徴表現で学習した HMM) の識別率は表 2 に示す様に約 42%であった．そして，これら 50 個の HMM で得られた識別結果からメタ識別器を学習した．

比較のため，代表的な識別器 C4.5，および文献 [3] の行動識別で用いられ，高い識別率が得られているアンサンブル識別器 (Random Forest: RF)[5] も比較した．C4.5 と RF の最良の結果を表 2 に示す．両クラス (22 クラスおよび 14 クラス), C4.5 と RF の識別率は最良の HMM よりも低かった．

表 3 に各手法での識別率を比較する．従来法は，ナイーブベイズ (NB)，ロジスティック回帰 (L1 正則化付）[10]，サポートベクトルマシン（線形カーネル:SVM(L)), 多項式カーネル:SVM(P)) と比較した．表から明らかなように，提案法 (BEM) が顕著に優位な結果を得られていることが確認できる．提案法は NB モデルの潜在変数拡張と見なせるが，両手法の識別結果から，潜在変数の導入の有効性が実験的にも検証できていると言える．尚，提案法はギブスサンプリングに基づいてメタ識別器を学習しているが，高々 100 回程度の反復で収束し，計算時間的にも問題はなく，また，テストデータに対する識別は一サンプル辺り実時間で計算可能である．

4. まとめと今後の課題

本研究では，ベイズモデルに基づくメタ学習法を提案し，加速度センサからの看護師行動識別タスクにおいてその有用性を検証した．現時点では，識別カテゴリ数に対して，学習データ量がそれほど十分でない状況ではあるが，少なくとも従来手法に比べ識別率において顕著な優位性を確認し，提案アプローチの有効性を検証した．今後の課題として，学習データ量の拡充による絶対精度の向上，および，予測行列の作成部分において，より良い特徴表現と識別器の選択などの最適化を行い，実システムとしての完成度を高める予定である．

謝辞 看護師行動データ収集にご協力頂いた済生会熊本病院の福島秀久先生に深謝します．本研究の一部は，内閣府最先端研究支援プログラム (FIRST) の援助による．

参考文献

[1] 日野原 重明，福井次矢: 監訳臨床決断分析 –医療における意思決定理論，医歯薬出版 (1992)
[2] 寺田 努 : ウェアラブルセンサを用いた行動認識技術の現状と課題, コンピュータソフトウェア, vol.28, no.2, pp.43–54 (2011).
[3] L. Bao and S. Intille : Activity recognition from user-annotated acceleration data, Proceedings of International Conference on Pervasive Computing, Pervasive2004, Springer-Verlag, pp.1–17 (2004)
[4] L. Breiman : Bagging predictors, Machine Learning, vol.24, pp.123–140 (1996).
[5] L. Breiman : Random forests, Machine Learning, vol. 45, pp.5–32 (2001).
[6] T. G. Dietterich : Ensemble methods in machine learn-

ing, Proceedings of the First International Workshop on Multiple Classifier Systems, Springer-Verlag, London UK, pp.1–15 (2000).
[7] Y. Freund, and R. E. Schapire : Experiments with a new boosting algorithm, Proceedings of International Conference on Machine Learning ICML96, pp.148–156 (1996).
[8] P. D. Hoff: Subset clustering of binary sequences, with an application to genomic abnormality data, Biometrics, vol.61, no.4, pp.1027–1036 (2005).
[9] C. Hsu, C. Chang, and C. Lin : A practical guide to support vector classification, http://www.csie.ntu.edu.tw/ cjlin (2010).
[10] R. Tibshirani : Regression shrinkage and selection via the Lasso, Journal of Royal Statistical Society. vol.58, no.1, pp.267–288 (1996).
[11] 井上 創造, 林田 興祐, 中村 優斗, 野原 康伸, 中島 直樹, "Capturing Nursing Interactions from Mobile Sensor Data and In-room Sensors", International Conference on Human-Computer Interaction (HCI International), to appear, July 21, 2013, Las Vegas, USA.