Vol.2013-HCI-155 No.13
Vol.2013-UBI-40 No.13
2013/11/6

半教師あり学習によるセンサデータからの日常行動分類

松重 龍之介[1]　伊藤 翔[1]　岡留 剛[1]

概要：加速度データを基に，日常行動を分類するためのいくつかの半教師あり学習法について議論する．具体的には，代表的な半教師あり学習法である Semi-Supervised Gaussian Mixture Model と Semi-Supervised Support Vector Machine，さらにカーネルロジスティック回帰を半教師あり学習に新たに拡張した手法である Semi-Supervised Kernel Logistic Regression のそれぞれを加速度データに適用して，分類性能の比較検討を行ない，それぞれの手法の特徴について議論する．

キーワード：半教師あり学習，センサデータ，日常行動分類

Semi-Supervised Learning based Activity Recognition from Sensor Data

Ryunosuke Matsushige[1]　Sho Ito[1]　Takeshi Okadome[1]

Abstract: The semi-supervised kernel logistic regression (SSKLR), proposed here, for probabilistic multi-calss classification takes the form of a linear combination of kernel functions associated with each of the labeled and unlabeled points from the training set. The EM algorithm determines the model parameters by maximizing the expectation of the joint distribution over the posterior for unlabeled data, where the joint distribution is represented by the softmax function. Several tests for SSKLR, together with those for the semi-supervised Gaussian mixture and semi-supervised support vector machine models, for sensor data obtained from human behaviors such as “ walk,” “ skip,” and “ run” reveal its high generalization ability.

Keywords: semi-supervised learning, sensor data, activity recognition

1. はじめに

人に装着したセンサーから得られるデータを解析し，人の行動を推定する技術は，周辺機器の制御や介護支援などの様々なコンテキストアウェアサービスへの応用が期待されている．行動推定の研究では，機械学習，とりわけ教師あり学習の手法が一般に用いられる．しかし，ラベルの付与は通常人手で行なわれるため，データが大量に得られる場合には，すべてのデータにラベルを付与するのにコストがかかる．本研究では，少数のラベルありデータと多数のラベルなしデータから分類を行なう半教師あり学習 [1] に着目する．半教師あり学習の手法を用いることによって，ラベルの付与にかかるコストを軽減し，かつ，教師あり学習と同等の汎化性能を得ることが期待できる．とりわけ，本研究では，センサデータからの日常行動分類に適した半教師あり学習法を検討することを目的とする．

[1] 関西学院大学
Kwansei Gakuin University

2. 半教師あり学習法

本研究では，代表的な半教師あり学習法である Semi-Supervised Gaussian Mixture Model (SSGMM) と Semi-Supervised Support Vector Machine (S3VM)，さらに，カーネルロジスティック回帰を半教師あり学習に新たに拡張した手法である Semi-Supervised Kernel Logistic Regression (SSKLR) を用いる．以下，$\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$ をラベルありデータとし，$\{\mathbf{x}_j\}_{j=l+1}^{l+u}$ をラベルなしデータとする．すなわち，入力 $\mathbf{x}_i$ に対するラベルが y_i であり，ラベルなし入力 $\mathbf{x}_j$ に対して $\{y_j\}_{j=l+1}^{l+u}$ は潜在変数である．

2.1 SSGMM

SSGMM は，データ集合に混合ガウス分布を当てはめ

るモデルである．このモデルの対数尤度は以下で書き表せる [1].

$$\log p(\mathcal{D}|\boldsymbol{\theta}) = \sum_{i=1}^{l} \log \pi_{y_i} \mathcal{N}(\mathbf{x}_i|\mu_{y_i}, \Sigma_{y_i}) + \sum_{i=l+1}^{l+u} \log \sum_{j=1}^{J} \pi_j \mathcal{N}(\mathbf{x}_i|\mu_j, \Sigma_j),$$

ここで，$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{l} \cup \{\mathbf{x}_j\}_{j=l+1}^{l+u}$ であり，$\boldsymbol{\theta} = \{\pi_{y_i}, \mu_{y_i}, \Sigma_{y_i}, \pi_j, \mu_j, \Sigma_j\}$ である．また，J はクラス数である．

SSGMM では，EM アルゴリズムを用いて，対数尤度を最大化するパラメータを推定する．まず，ラベルありデータの最尤推定値を求める．すなわち，$j = 1, ..., J$ に対して

$$\mu_j^{(0)} = \frac{1}{l_j} \sum_{i:y_i=j} \mathbf{x}_i,$$
$$\Sigma_l^{(0)} = \frac{1}{l_j} \sum_{i:y_i=j} (\mathbf{x}_i - \mu_j)(\mathbf{x}_i - \mu_j)^{\mathrm{T}},$$
$$\pi_j^{(0)} = \frac{l_j}{l},$$

ここで，l はラベルありデータの数，l_j はクラス j に属するラベルありデータの数である．次に，現在のパラメータ値を用いて，ラベルなしデータについて次の負担率 $\gamma_{ij}, i = l+1, ..., l+u, j = 1, ..., J$ を計算する (E ステップ).

$$\gamma_{ij} = \frac{\pi_j^{(t)} \mathcal{N}(\mathbf{x}_i|\mu_j^{(t)}, \Sigma_j^{(t)})}{\sum_{k=1}^{J} \pi_k^{(t)} \mathcal{N}(\mathbf{x}_i|\mu_k^{(t)}, \Sigma_k^{(t)})}.$$

ただし，ラベルありデータについては，$y_i = j$ のとき $\gamma_{ij} = 1$，そうでないとき $\gamma_{ij} = 0$ とする．この負担率を用いて以下のようにパラメータを更新する (M ステップ) .

$$l_j = \sum_{i=1}^{l+u} \gamma_{ij},$$
$$\mu_j^{(t+1)} = \frac{1}{l_j} \sum_{i=1}^{l+u} \gamma_{ij} \mathbf{x}_i,$$
$$\Sigma_l^{(t+1)} = \frac{1}{l_j} \sum_{i=1}^{l+u} \gamma_{ij} (\mathbf{x}_i - \mu_j^{(t+1)})(\mathbf{x}_i - \mu_j^{(t+1)})^{\mathrm{T}},$$
$$\pi_j^{(t+1)} = \frac{l_j}{l+u}.$$

対数尤度が収束するまで E ステップと M ステップを繰り返す．

対数尤度を最大化するパラメータが求まると，新しい入力に対する出力の確率分布を予測できる．学習時のパラメータを用いて，テストデータの負担率を計算し，最も大きい値を返すクラスを予測ラベルとする．

2.2 S3VM

S3VM は，教師あり学習の SVM を半教師あり学習に拡張したもので，マージンを最大化する決定境界を求めるモデルである．マージンとは，分類境界と訓練データの間の最短距離を指す．S3VM の最適化にはいくつかの手法が存在する [2] が，本研究では SVM^{light} [3] を用いる．また，カーネル関数としてガウシアン RBF カーネルを使用し，その精度パラメータを 0.1 とした．

S3VM では，まず，ラベルありデータのみを用いて決定境界を求める．ラベルありデータに関する決定境界は，以下の 2 次計画問題を解くことで得られる．

$$\min_{\mathbf{w},\boldsymbol{\xi}} \quad \frac{1}{2}||\mathbf{w}^2|| + C\sum_{i=1}^{l} \xi_i,$$
$$\text{subject to} \quad \forall_{i=1}^{l} : y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1 - \xi_i.$$

ここで，$\mathbf{w}$ は重みパラメータ，b はバイアスパラメータ，$\boldsymbol{\xi}$ はスラック変数，C は正則化係数である．本研究では，$C = 1$ とした．この決定境界にしたがってラベルなしデータにラベルを付与する．正のクラスに割り当てられるラベルなしデータの数を num_+ としたとき，識別関数の値が大きい num_+ 個のラベルなしデータを正例とし，残りのラベルなしデータを負例とする．次に，目的関数の値が減少するような正例と負例の組を選択し，その正のラベルと負のラベルを入れ替える．すなわち，$\{y_j\}_{j=l+1}^{l+u}$ について

$$y_m * y_l < 0,$$
$$\xi_m > 0,$$
$$\xi_l > 0,$$
$$\xi_m + \xi_l > 2.$$

上記条件を満たす m と l に関して，m 番目のラベルなしデータのラベルと l 番目のラベルなしデータのラベルを入れ替え，再び決定境界を求める．ラベルなしデータを含めた場合の決定境界は，以下の 2 次計画問題を解くことで得られる．

$$\min_{\mathbf{w},\boldsymbol{\xi}} \quad \frac{1}{2}||\mathbf{w}^2|| + C\sum_{i=1}^{l} \xi_i + C_- \sum_{j:y_j=-1} \xi_j + C_+ \sum_{j:y_j=1} \xi_j,$$
$$\text{subject to} \quad \forall_{i=1}^{l} : y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1 - \xi_i,$$
$$\forall_{j=l+1}^{l+u} : y_j(\mathbf{w}^{\mathrm{T}}\mathbf{x}_j + b) \geq 1 - \xi_j,$$

ここで，C_-，C_+ はラベルなしデータに関する正則化係数である．上記の条件を満たす m，l が存在する間，ラベルの入れ替えと再訓練を繰り返す．

本研究では，多クラス分類を行なうため，1 対他方式を用いる．これは，K 個のクラスがあるときに，あるクラス C_k に属するデータを正例とし，それ以外のデータを負例として K 個の別々の SVM を学習する方法である．しかし，1 対他方式では，一つの入力に複数のクラスが割り当

てられる可能性がある．この問題を避けるために，新しい入力 $\mathbf{x}$ に対して

$$y(\mathbf{x}) = \max_k y_k(\mathbf{x}),$$

に従って，最も大きな識別関数の値を返すクラスを予測ラベルとする．また，訓練データの正例と負例のバランスを考慮して，正例に対する識別関数の値は $+1$，負例に対しては $-1/(K-1)$ となるように学習する [4]．

2.3 SSKLR

SSGMM および S3VM は，代表的な半教師あり学習法として広く用いられているが，いくつかの問題点を含んでいる．例えば，SSGMM は，各クラスのデータ集合がガウス分布から生成されるという仮定を置いている．また，S3VM は，出力が非確率的な識別結果であることや，本来は 2 クラス分類のためのモデルであるという問題がある．そこで，これらの問題点を解決するための新たな手法として SSKLR モデルを提案する．

SSKLR は，カーネルロジスティック回帰（KLR）を半教師あり学習に拡張したモデルである．y_n が潜在変数である時 1，ラベルである時 0 を取る観測変数 g_n を導入する．さらに，以下の条件付き確率を定義する．

$$p(g_n = 1|y_n) = \begin{cases} \mu^{(1)} & \text{if} \quad y_n = 1 \\ \vdots & \\ \mu^{(K)} & \text{if} \quad y_n = K. \end{cases}$$

すなわち，$p(g_n = 1|y_n)$ は，データ中のあるクラスのラベルありデータの割合である．この変数 g_n と上記の条件付き確率を導入すると，ソフトマックス関数を用いて，$\mathbf{L} = \{y_i\}_{i=1}^{l}$ と $\mathbf{Z} = \{y_j\}_{j=l+1}^{l+u}$ との同時確率が以下で書き表せる [5]．

$$p(\mathbf{L}, \mathbf{Z}|\mathbf{X}, \mathbf{W}) = \prod_{i=1}^{l}\prod_{k=1}^{K} \frac{(\exp(\mathbf{w}_k^{\mathrm{T}}\phi(\mathbf{x}_i)))^{z_{ik}}}{\sum_{c=1}^{K}\exp(\mathbf{w}_c^{\mathrm{T}}\phi(\mathbf{x}_i))} \times \prod_{j=l+1}^{l+u}\prod_{k=1}^{K} \frac{(\exp(\mathbf{w}_k^{\mathrm{T}}\phi(\mathbf{x}_j)) + \mu_k)^{z_{jk}}}{\sum_{c=1}^{K}\exp(\mathbf{w}_c^{\mathrm{T}}\phi(\mathbf{x}_j))}$$

ここで，z_i は y_i の指示変数（1-of-K）で，z_j は y_j の指示変数（潜在変数）である．また，$\mathbf{X}$ は n 番目の行が $\mathbf{x}_n^{\mathrm{T}}$ で与えられるデータ行列，$\mathbf{W}$ は k 番目の行が $\mathbf{w}_k^{\mathrm{T}}$ で与えられる重みパラメータの行列，$\phi(\mathbf{x}_n) = k(\mathbf{x}_n, \mathbf{X})$，$\mu_k = \log\mu^{(k)}$，$K$ はクラス数である．また，カーネル関数 $k(\cdot,\cdot)$ としてガウシアン RBF カーネルを使用し，その精度パラメータを 0.1 とした．

この同時確率の z_{jk} は潜在変数であるため，尤度ではなく，最尤推定を行なうことができない．そこで，z_{jk} の事後分布に関する同時確率の期待値を最大化する．この期待値の最大化を EM アルゴリズムを用いて求める．まず，$\mathbf{W}$ の初期値を選ぶ．次に，現在のパラメータ値を用いて，z_{jk} の事後分布を求める (E ステップ)．すなわち，

$$p(\mathbf{Z}|\mathbf{X}, \mathbf{W}, \mathbf{L}) = \frac{(\exp(\mathbf{w}_k^{\mathrm{T}}\phi(\mathbf{x}_n) + \mu_k))^{z_{jk}}}{\sum_{c=1}^{K}\exp(\mathbf{w}_c^{\mathrm{T}}\phi(\mathbf{x}_n))}.$$

z_{jk} の事後分布に関する以下の同時確率の期待値を計算し，これを最大化するパラメータを求める （M ステップ）．

$$Q(\mathbf{W}, \mathbf{W}^{old}) = \sum_{i=1}^{l}\sum_{k=1}^{K} z_{ik}\log\left(\frac{\exp(\mathbf{w}_k^{\mathrm{T}}\phi(\mathbf{x}_i))}{\sum_{c=1}^{K}\exp(\mathbf{w}_c^{\mathrm{T}}\phi(\mathbf{x}_i))}\right) + \sum_{j=l+1}^{l+u}\sum_{k=1}^{K} b_{jk}\log\left(\frac{\exp(\mathbf{w}_k^{\mathrm{T}}\phi(\mathbf{x}_j) + \mu_k)}{\sum_{c=1}^{K}\exp(\mathbf{w}_c^{\mathrm{T}}\phi(\mathbf{x}_j))}\right),$$

ただし，

$$b_{jk} = \frac{\exp(\overline{\mathbf{w}}_k^{\mathrm{T}}\phi(\mathbf{x}_j) + \mu_k)}{\sum_{c=1}^{K}\exp(\overline{\mathbf{w}}_c^{\mathrm{T}}\phi(\mathbf{x}_j) + \mu_c)}, \quad \overline{\mathbf{w}}_k = \mathbf{w}_k^{old}$$

である．収束条件を満たすまで E ステップと M ステップを繰り返す．

3. 日常行動の加速度データ

3.1 HASC

本研究では，HASC （Human Activity Sensing Consortium） の 3 軸加速度データを使用して行動推定を行なう [6]．HASC は，人間行動理解のための装着型センサーによる大規模データベースの構築を目的とした研究団体であり，行動推定の研究およびセンサデータの収集を行なっている．

3.2 特徴抽出

3 軸加速度センサーから得られるデータは，デバイスの保持姿勢によって値が変化するので，重力成分を推定し，除去する必要がある [7]．3 軸加速度センサーには常に 1G の重力がかかるので，一定の時間幅 w_G での XYZ 軸の平均ベクトルを重力ベクトルと推定する．すなわち，時刻 t での重力ベクトル $v_G(t)$ は，3 軸加速度ベクトル $v(t)$ を用いて以下で書き表せる．

$$v_G(t) = \frac{\sum_{i=t-w_G}^{t} v(i)}{w_G}.$$

3 軸加速度ベクトルから重力ベクトルを減算し，正規化された運動加速度ベクトルを得る．時刻 t での正規化された運動加速度ベクトル $v_n(t)$ は以下で書き表せる．

$$v_n(t) = v(t) - v_G(t).$$

加速度データからの教師あり学習に関する先行研究 [7] や [8] ・[9] を参考に，本研究では以下で述べる 11 の特徴量を用いる．まず，正規化された加速度ベクトルから特徴量として，以下のベクトル長 $F_1(t)$ と重力ベクトルとの内積値 $F_2(t)$ ・重力ベクトルとの外積値 $F_3(t)$ を算出する．

$$F_1(t) = ||v_n(t)||,$$
$$F_2(t) = v_n(t) \cdot v_G(t),$$
$$F_3(t) = v_n(t) \times v_G(t).$$

文献 [7] では，$F_1(t)$，$F_2(t)$，$F_3(t)$ それぞれについて平均値，最大値，最小値，分散値の 4 種類の統計量を計算しているが，これらの統計量は外れ値の影響を受ける可能性がある．そのため，本研究では，25%値，中央値，75%値の 3 種類の統計量を用いた．

また，HASC データにおける「歩く」や「走る」などの行動は，周期的に繰り返される動作であるので，周波数軸方向における特徴量を考慮する．本研究では，エネルギーと周波数領域エントロピーの 2 種類の特徴量を用いる．エネルギーは，各データに関して FFT を行ない，得られた周波数成分の絶対値の合計で表される [8]．すなわち，

$$E_n = \frac{1}{N}\sum_{i=2}^{n}|F_i|^2.$$

周波数領域エントロピーは，FFT の全成分の総和で各成分を正規化し，確率分布 p を求め，そのエントロピーで表される．すなわち，

$$p(i) = \frac{|F_i|^2}{\sum_{i=2}^{n}|F_i|^2},$$

$$FDE = -\sum_{i=2}^{n} p(i)\log p(i).$$

得られた 11 次元の特徴量に対して，白色化 (ホワイトニング) を行なう．ホワイトニングとは，データを正規化し，主成分分析を施す処理である [10]．すなわち，各データ点 $\mathbf{x}_n$ に対して

$$\mathbf{y}_n = \mathbf{L}^{-1/2}\mathbf{U}^{\mathrm{T}}(\mathbf{x}_n - \overline{\mathbf{x}})$$

で与えられる変換値を定義する．ここで，$\mathbf{L}$ はデータ集合の共分散行列の固有値 λ_i を持つ対角行列，$\mathbf{U}$ は i 番目の列が固有ベクトル $\mathbf{u}_i$ で与えられる直交行列，$\overline{\mathbf{x}}$ はサンプル平均である．これにより，特徴空間の各軸が無相関化される．

4. 評価

本研究では，HASC が提供している HASC2011corpus というデータ集合から，サンプリングレートが 100Hz のセンサーを腰に付けた被験者 32 名の 3 軸加速度データを用いて，「静止」，「歩く」，「走る」，「スキップ」，「階段を上る」，「階段を下りる」の 6 クラス分類を行なった．ここで，1 クラスのデータ数を 200 とし，合計 1200 のデータを用いた．

評価の方法として，半教師あり学習による評価と教師あり学習による評価，および後で述べるランダムリダクション法による評価を行なった．それぞれの評価において，全データの 1 割をテストデータ，残りの 9 割を学習データとする交差確認を行なった．

表 1　SSGMM の半教師あり学習による Confusion Matrix (%)

	stay	walk	jog	skip	stUp	stDown
stay	92.0	1.5	0.5	0.0	6.0	0.0
walk	2.0	41.0	4.0	7.5	25.0	20.5
jog	0.5	0.5	48.5	48.0	0.0	2.5
skip	0.0	0.0	24.5	73.0	0.0	2.5
stUp	0.0	12.0	1.0	2.0	73.5	11.5
stDown	0.0	13.0	0.0	5.0	30.0	52.0

4.1 半教師あり学習

半教師あり学習では，ラベルありデータとラベルなしデータを用いて学習する．HASC のデータにはすべてラベルが付与されているので，学習データの中でランダムに抽出したデータをラベルなしデータとみなして学習する．ここで，ラベルありデータの割合を 5%，10%，30%，50%とした．ただし，ラベルありデータの割合が 5% のときは 1 クラスのラベルありデータ数が特徴量の次元数より小さくなり，SSGMM において逆行列の計算ができないので，SSGMM の 5% の結果は省略した．半教師あり学習による各分類器の識別率を図 1 に示す．

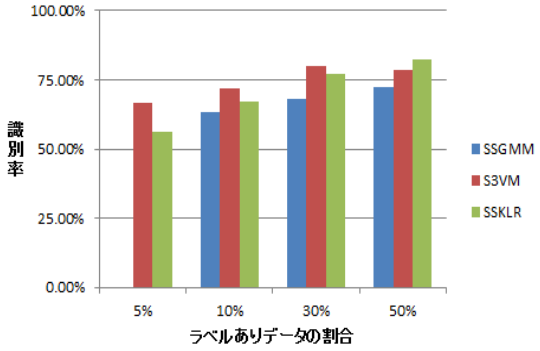

図 1　半教師あり学習による各分類器の識別率．

交差確認の結果，ラベルありデータの割合が 5%，10%，30% のときに S3VM が最も高い識別率を示した（ラベルありデータの割合が 10% のとき 71.8%）．SSKLR は，ラベルありデータの割合が 50% のときに S3VM よりも高い識別率を示した（ラベルありデータの割合が 10% のとき 67.3%）．SSGMM は 3 種類の分類器の中で識別率が最も低い（ラベルありデータの割合が 10% のとき 63.3%）．表 1 と表 2 ・表 3 は，ラベルありデータの割合を 10% としたときの各分類器の confusion matrix の平均である．

表 1 と表 2 ・表 3 より，「歩く」と「階段を上る」・「階段を下りる」の混同および「走る」と「スキップ」の混同がどの分類器でも多く見られる．各分類器の適合率と再現

表 2　S3VM の半教師あり学習による Confusion Matrix (%)

	stay	walk	jog	skip	stUp	stDown
stay	97.5	2.0	0.5	0.0	0.0	0.0
walk	3.0	61.5	6.0	2.5	11.0	16.0
jog	1.0	0.5	67.0	24.0	3.0	4.5
skip	0.0	1.0	25.0	69.0	1.5	3.5
stUp	3.0	14.0	0.0	1.5	67.5	14.0
stDown	0.5	9.5	3.5	3.0	15.0	68.5

表 3　SSKLR の半教師あり学習による Confusion Matrix (%)

	stay	walk	jog	skip	stUp	stDown
stay	97.5	2.0	0.5	0.0	0.0	0.0
walk	4.0	61.0	0.0	1.5	16.0	17.5
jog	5.0	4.0	53.0	25.5	0.0	12.5
skip	2.0	6.0	18.5	64.0	1.0	8.5
stUp	3.0	21.5	0.5	2.0	60.0	13.0
stDown	0.0	15.5	4.0	4.5	7.5	68.5

率を図 2 と図 3 に示す．いずれも識別率の結果と同様であることが分かる．

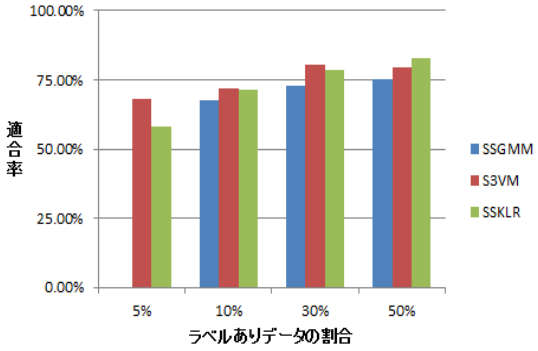

図 2　半教師あり学習による各分類器の適合率．

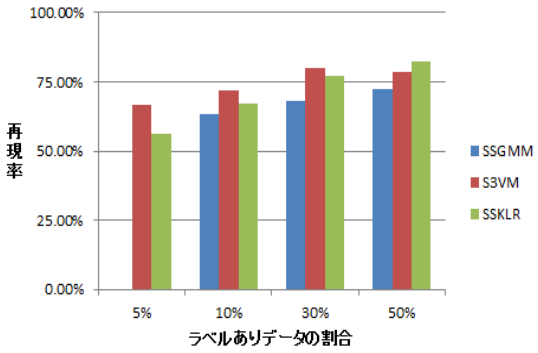

図 3　半教師あり学習による各分類器の再現率．

4.2　教師あり学習

ラベルなしデータの利用の有効性を検証するため，学習データからラベルなしデータを除いて学習する．すなわち，学習データをラベルありデータとラベルなしデータに分割し，ラベルありデータのみを用いて学習を行なう．ラベルありデータの割合はやはり 5%，10%，30%，50%とし，ラベルなしデータは用いずに，GMM，SVM，KLR の 3 手法で教師あり学習を行なう．各クラスの学習データの個数は，それぞれ 9，18，54，90 である．教師あり学習による各分類器の識別率を図 4 に示す．ここで，半教師あり学習による評価と同様に，SSGMM の 5% の結果は省略した．

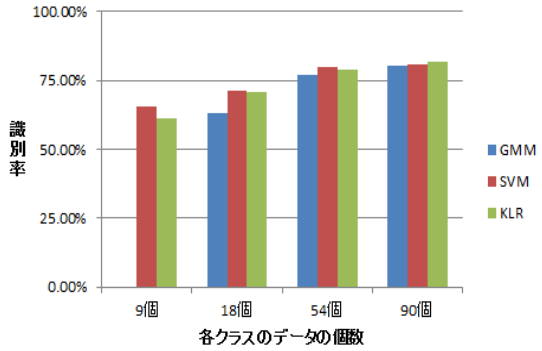

図 4　教師あり学習による各分類器の識別率．

交差確認の結果，ラベルありデータの割合が 5%，10%，30% のときに S3VM が最も高い識別率を示した（ラベルありデータの割合が 10% のとき 71.3%）．SSKLR は S3VM と同等の識別率を示し，ラベルありデータの割合が 50% のときには S3VM よりも高い識別率を示した（ラベルありデータの割合が 10% のとき 70.9%）．SSGMM は 3 種類の分類器の中で識別率が最も低い（ラベルありデータの割合が 10% のとき 62.9%）．半教師あり学習による識別率とともに，教師あり学習による各分類器の識別率を図 5 と図 6・図 7 に示す．

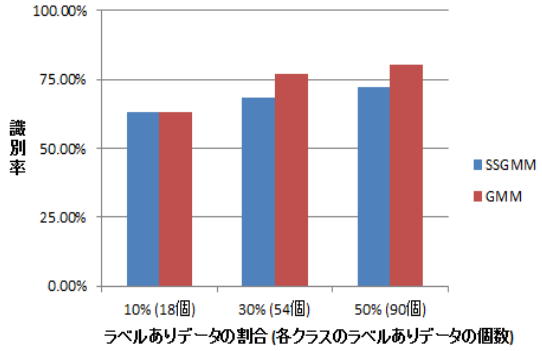

図 5　SSGMM による識別率と，GMM による識別率．() 内は，各クラスのラベルありデータ数．

半教師あり学習による評価と比較すると，図 5 と図 7 より，SSGMM および SSKLR は，ラベルありデータのみを用いた方が識別率が高いことが分かる．一方，図 6 より，S3VM は，ラベルなしデータを用いた方が識別率が高くなった．

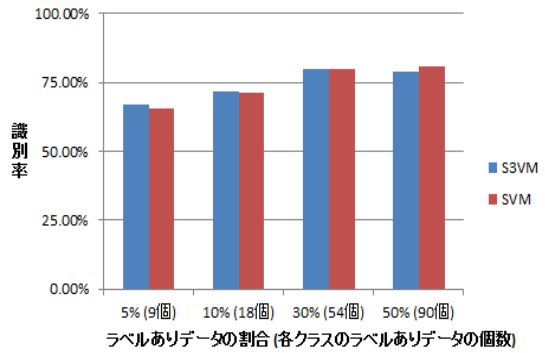

図 6 S3VM による識別率と，SVM による識別率．() 内は，各クラスのラベルありデータ数．

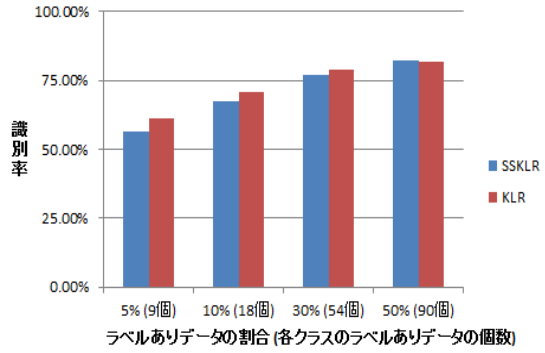

図 7 SSKLR による識別率と，KLR による識別率．() 内は，各クラスのラベルありデータ数．

4.3 ランダムリダクション法

第 4.2 節で述べたように，SSGMM ならびに SSKLR では，ラベルなしデータをすべて用いた場合，識別率の劣化をもたらしている．本研究では，ランダムリダクション法を新たに導入し，その評価を行なった．これは，ラベルなしデータの全てを用いるのではなく，数に制限を置いてランダムにラベルなしデータを選択して学習する方法である．本研究では，ラベルありデータと同数とした．SSGMM および SSKLR について，ラベルなしデータをランダムに 50 回選んで学習し，ラベルありデータに対する尤度が最大になる結果を使用してテストを行なった．ここで，ラベルありデータの割合が 50% のときは，通常の半教師あり学習と同じであるので，ラベルありデータの割合を 5%，10%，30% とした．また，S3VM には尤度がないので，ランダムリダクション法による評価は省略した．半教師あり学習ならびに教師あり学習による識別率とともに，ランダムリダクション法による SSGMM と SSKLR の識別率を図 8 と図 9 に示す．

交差確認の結果，SSGMM では，通常の半教師あり学習よりも高い識別率を示した（ラベルありデータの割合が 10% のとき 65.0%）．SSKLR では，教師あり学習よりも高い識別率を示した（ラベルありデータの割合が 10% のとき 72.6%）．

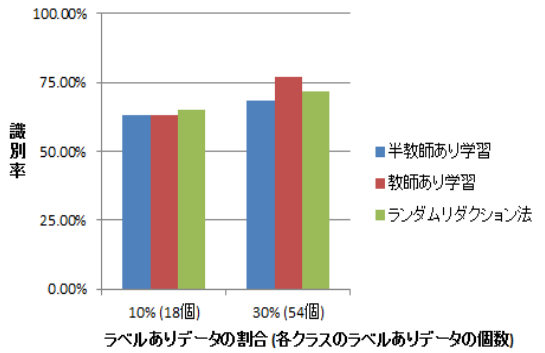

図 8 各方法による SSGMM の識別率．() 内は，各クラスのラベルありデータ数．

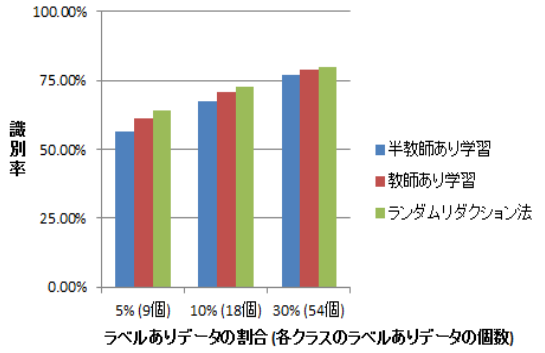

図 9 各方法による SSKLR の識別率．() 内は，各クラスのラベルありデータ数．

5. 議論

SSGMM は，3 種類の分類器の中で識別率が最も低い．これは，各クラスのデータがガウス分布に従って生成されるという強い仮定を置いていることが原因であると考えられる．すなわち，SSGMM では，各クラスのデータがそれぞれ単一のガウス分布に従って生成されるとし，全体として混合数が 6 の混合ガウス分布を当てはめているが，センサデータの場合，各クラスのデータがそれぞれ単一のガウス分布に従うという根拠は無い．各クラスのデータがそれぞれ混合ガウス分布に従い，全体として混合数がクラス数よりも大きい混合ガウス分布となる可能性がある．また，各ガウス分布の平均値や分散値が近い場合，誤識別が起こりやすくなる．図 10 は，ホワイトニングしたデータの固有値の大きい 2 軸を選択し，プロットしたものであり，図 11 は，6 クラスの中で「歩く」と「階段を上る」の 2 クラスを取り出したものである．

図 10 と図 11 より，「歩く」と「階段を上る」，「階段を下りる」の 3 クラス，あるいは「走る」と「スキップ」の 2 クラスが互いに重なっており，誤識別が起こりやすいことが分かる．従って，データに応じて混合数を決定するモデルを検討する必要がある．

S3VM は，3 種類の分類器の中で最も高い識別率を示し

Vol.2013-HCI-155 No.13
Vol.2013-UBI-40 No.13
2013/11/6

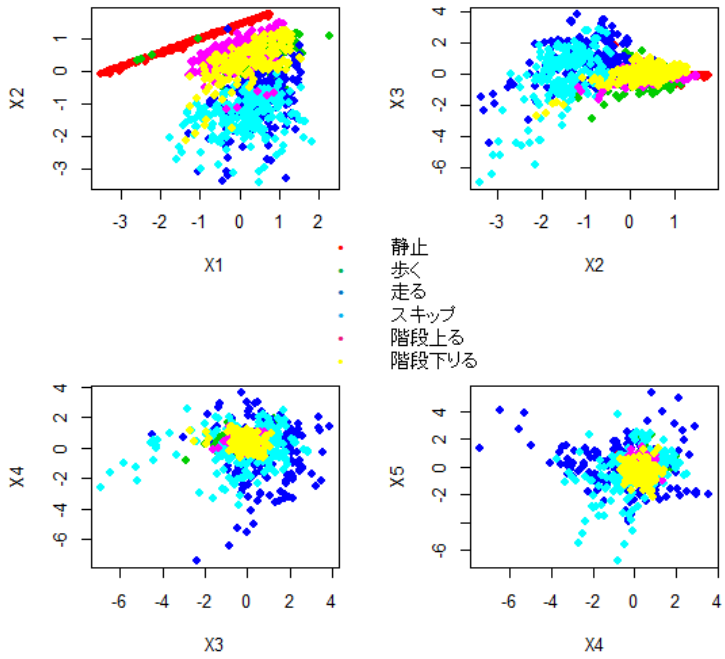

図 10 ホワイトニングしたデータの 2 次元プロット．左上の図は，1 番目と 2 番目の固有値の軸であり，右上の図は，2 番目と 3 番目の固有値の軸である．左下の図は，3 番目と 4 番目の固有値の軸であり，右下の図は，4 番目と 5 番目の固有値の軸である．

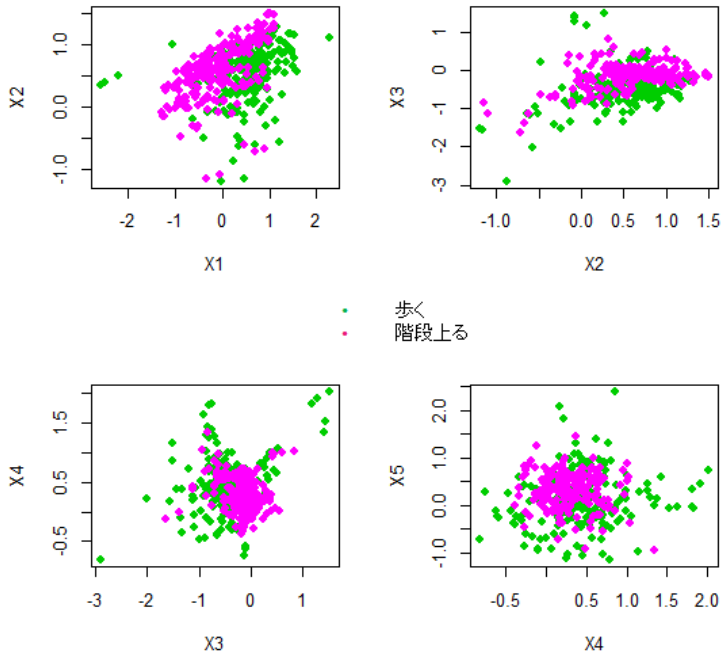

図 11 ホワイトニングしたデータから，「歩く」と「階段を上る」の 2 クラスを取り出した 2 次元プロット．左上の図は，1 番目と 2 番目の固有値の軸であり，右上の図は，2 番目と 3 番目の固有値の軸である．左下の図は，3 番目と 4 番目の固有値の軸であり，右下の図は，4 番目と 5 番目の固有値の軸である．

表 4　k-nearest-neighbor による Confusion Matrix (%)

	stay	walk	jog	skip	stUp	stDown
stay	99.0	0.5	0.5	0.0	0.0	0.0
walk	3.0	80.0	0.0	0.0	11.5	5.5
jog	0.5	1.0	90.0	6.0	1.0	1.5
skip	0.0	0.0	5.0	93.5	0.5	1.0
stUp	0.0	7.5	0.0	0.0	88.5	4.0
stDown	0.5	4.0	0.0	0.0	8.0	87.5

た．これは，SSGMM のような強い仮定が無いことが原因であると考えられる．しかし，SVM の多クラス分類のための 1 対他方式では，個々の SVM は独立した分類問題を解くように学習しているため，識別関数の値を比較することが意味を持つかどうかは保障されていない．また，尤度や誤差関数が定義されないため，SSGMM や SSKLR で有効であったランダムリダクション法は S3VM では適用できない．

これに対して，新たに提案した手法である SSKLR は，多クラス分類に適用することができ，かつ，予測に対する確率を得ることができるので，既存手法の問題点を克服したモデルであると考えられる．また，S3VM と同等の識別率を示すため，センサデータによる行動推定において有用であると考えられる．しかし，計算量が多いことが問題点として挙げられる．

半教師あり学習による評価と教師あり学習による評価を比較すると，SSGMM および SSKLR では，ラベルなしデータを用いて学習するよりも，ラベルありデータのみを用いて学習する方が識別率が高くなることが示された．この 2 つのモデルは，すべてのデータを用いて学習を行なうため，外れ値の影響を受ける．そのため，ラベルなしデータを用いることによって，識別率が低下したと考えられる．

新たに提案した評価方法であるランダムリダクション法を用いると，SSKLR では，通常の半教師あり学習および教師あり学習よりも高い識別率を示した．ラベルなしデータをすべて用いるのではなく，識別率の向上に貢献するデータを選別することによって，通常の半教師あり学習よりも識別率が高くなると考えられる．

なお，参考のため，ラベルありデータをすべて保持・利用する k-nearest-neighbor を用いて同様の分類を行なった．k-nearest-neighbor は，新しい入力に対して，学習データの中から k 近傍のデータを選び，その多数派のクラスラベルを割り当てるモデルである．$k = 3$ として，交差確認を行なった結果，識別率の平均値は 89.8% となった．これが学習による分類器の識別率の上限であると思われる．表 4 は，k-nearest-neighbor による confusion matrix の平均である．分類するのが難しい「歩く」と「階段を上る」，「階段を下りる」に対しても，識別率が高いことが分かる．

6. おわりに

本研究では，半教師あり学習の既存の手法である SSGMM と S3VM，および新たに提案した手法である SSKLR のそれぞれを，HASC データに適用して，分類性能の比較検討を行なった．SSGMM は，3 種類の分類器の中で最も低い識別率を示した．S3VM は，3 種類の分類器の中で最も高い識別率を示した．SSKLR は，S3VM と同等の識別率を示し，ラベルなしデータの全てを用いるのではなく，数に制限を置いてランダムにラベルなしデータを選択して学習するランダムリダクション法では S3VM よりも高い識別率を示した．このモデルは，多クラス分類へ適用することができ，かつ，予測に対する確率を計算することができるため，既存手法の問題点を克服したモデルである．

今後は，Relevance Vector Machine を半教師あり学習に拡張したモデルである Semi-Supervised Relevance Vector Machine (SSRVM) を実装する予定である．SSRVM は，SSKLR を疎にしたモデルであるので，SSKLR よりも高い識別率を示すことが期待できる．また，複数のモデルを結合させて，識別率の向上に取り組む予定である．

参考文献

[1] Zhu, X. and A. B. Goldberg (2009). *Introduction to Semi-Supervised Learning*. Morgan & Claypool Publishers.

[2] Chapelle, O. et al. (2008). Optimization techniques for semi-supervised support vector machine. *Journal of Machine Learning Research*, 9, 203-233.

[3] Joachims, T. (1999). Transductive inference for text classification using Support Vector Machines. *Proceedings of the Sixteenth International Conference on Machine Learning*, 200-209.

[4] Lee, Y., Y. Lin, and G. Wahba (2001). Multicategory support vector machine. *Technical Report 1040*, Department of Statistics, University of Madison, Wisconsin.

[5] 伊藤 翔, 角所 考, 岡留 剛 (2013). Semi-Supervised Sparse Bayesian Learning とそのセンサデータへの適用. **人工知能学会**, 476.

[6] Kawaguchi, N. et al. (2012). HASC2012corpus: Large scale human activity corpus and its application. *Proceedings of the 2nd International Workshop of Mobile Sensing: From Smartphones and Wearables to Big Data*, 10-14.

[7] 池谷直紀, 菊池匡晃, 長 健太, 服部正典 (2008). 3 軸加速度センサを用いた移動状況推定方式. **電子情報通信学会 (USN)**, 75-80.

[8] 小川兼人, 伊藤雄一, 安部登樹, 岸野文郎 (2009). 実物体によるモーションクエリを用いた 3 次元形状モデル検索. **情報処理学会**, 9-16.

[9] Bao, L. and S. Intille (2004). Activity recognition from user-annotated acceleration data. *Proceedings of Second International Conference on Pervasive Computing*, 1-17.

[10] Bishop, C. M. (2012). **パターン認識と機械学習**, 丸善出版.