

半教師あり擬 RVM による加速度データからの行動推定

Semi-Supervised Pseudo Relevance Vector Machine based
Activity Recognition from Acceleration Data

松重 龍之介*¹ 角所 考*¹ 岡留 剛*¹
 Ryunosuke MATSUSHIGE Koh KAKUSHO Takeshi OKADOME

*¹関西学院大学大学院理工学研究科
 Graduate School of Science and Engineering, Kwansei Gakuin University

The semi-supervised pseudo relevance vector machine (SSPRVM), proposed here, for probabilistic multiclass 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 SSPRVM, together with those for the semi-supervised Gaussian mixture and semi-supervised support vector machine models, for acceleration data obtained from human behaviors such as “walk,” “skip,” and “jog” reveal its high generalization ability.

1. はじめに

センサデータから行動を分類する場合には、通常、教師あり学習を用いて分類器を作成する [Bao 04]。教師あり学習は、入力データにクラスラベルを付与して分類器を作成する手法である。教師あり学習による行動推定の研究は数多く存在し、それぞれ高い正解率が示されている。しかし、ラベルの付与は、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、ラベルを付与するのにコストがかかる。多種類のセンサーを搭載したスマートフォンがインターネットに「常時接続」した状況では、多種多様な大量のデータが生成され、それゆえすべてのデータにラベルを付与するのは現実的ではない。

本研究では、半教師あり学習 [Zhu 09] に着目し、大量のセンサデータのもとで精度の高い行動分類器を効率的に作成する手法を構築することを目的とする。半教師あり学習は、少数のデータにラベルを付与し、残りのデータにはラベルを付与せずに分類器を作成する手法である。半教師あり学習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。

半教師あり学習では、モデルの仮定が重要な役割を果たす。半教師あり学習を用いて作成される分類器の性能は、モデルの仮定の正しさに依存する。例えば、半教師あり学習の既存手法である半教師あり混合ガウスモデル (SSGMM; Semi-Supervised Gaussian Mixture Model) では、各クラスのデータが単一のガウス分布に従って生成されるという仮定を置いている。また、半教師ありサポートベクトルマシン (S3VM; Semi-Supervised Support Vector Machine) では、各クラスのデータが十分に分離されているという仮定を置いている。学習データがモデルの仮定に従うとき、半教師あり学習を用いることによって、予測精度の高い分類器を作成することができる。

しかし、性別や身長・体重などの身体的特徴に依存して個人差が生じるため、行動中に得られるセンサデータは、既存手法の仮定に従わないと考えられる。すなわち、各クラスのセンサデータは、単一のガウス分布に従わず、さらに、クラス

ごとに峰の数が異なり、多峰的であるといった特徴を持つ可能性がある。センサデータから行動を分類する場合には、分布に対する強い仮定を置くことは避けるべきである。本研究では、分布に対する仮定を置かず、多クラス分類に適用することができ、予測に対する確率を計算することができる半教師あり学習法として、半教師ありカーネルロジスティック回帰 (SSKLR; Semi-Supervised Kernel Logistic Regression) を提案する。多クラス教師あり学習の一手法であるカーネルロジスティック回帰 (KLR; Kernel Logistic Regression) は、ロジスティック回帰の入力をカーネル関数により非線形化したモデルであり、SSKLR は、KLR を半教師あり学習に拡張したモデルである。

SSKLR において、すべての訓練データを使用して学習すると、汎化性能が低下してしまう現象が見られた。そのため、疎なモデルである S3VM より正解率が低い。したがって、ラベルなしデータをすべて用いるのではなく、分類に役立つラベルなしデータを選別することが重要であると考えられる。本研究では、ラベルなしデータの選別手法として 2 種類の手法を提案する。1 つはランダムセレクション法で、もう 1 つは半教師あり擬 RVM (SSPRVM; Semi-Supervised Pseudo Relevance Vector Machine) である。本稿では、これらの手法と、既存手法である SSGMM と S3VM のそれぞれについて、行動中に得られた加速度データを用い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。

2. 関連研究

半教師あり学習の枠組みで、センサデータから行動認識を行っている研究は筆者らが知る範囲では無い。ここでは、最近の教師あり学習による行動認識の研究について簡単に触れる。

文献 [Maekawa 13] では、ユーザから得られたラベルあり/ラベルなし加速度センサデータを使用せずに、エンドユーザの行動をモデル化する行動認識手法を提案している。まず、あらかじめ用意した多数の他のユーザー（ソースユーザー）のセンサデータから、性別や身長などの身体的特徴情報を用いて、エンドユーザー（ターゲットユーザー）と類似しているセンサデータを、機械学習のアプローチを用いて発見する。そして、そのソースユーザーのラベルありセンサデータを用いて、ターゲットユーザーの行動モデルを学習する。すなわち、エンド

連絡先: 氏名: 松重 龍之介

所属: 関西学院大学大学院理工学研究科

住所: 〒 669-1337 兵庫県三田市学園 2-1

メールアドレス: ryunosuke@kwansei.ac.jp

ユーザーは、自身のセンサデータを集める必要がなく、センサデータにラベルを付与する必要もない。分類する行動は、「歩く」や「歯を磨く」・「掃除機をかける」などの14種類である。使用したセンサーは、サンプリングレートが30Hzの3軸加速度センサーである。センサーの取付位置は、被験者の両手首、腰、右腿の4箇所である。被験者数は40名である。特徴量には、平均やエネルギー・周波数領域エントロピーが用いられている。分類器には、隠れマルコフモデル(HMM; Hidden Markov Model)が用いられている。ソースユーザーの人数を10人としたときの識別率は50%とかなり低い結果となっている。

文献[上田13]では、異なる条件で個別に学習した複数の分類器の分類結果から真のクラスを推定するメタ学習法を提案している。まず、ある時間幅のsliding windowを設定し、50%のオーバーラップですりながら特徴抽出を行なう。特徴量には、平均値や標準偏差・エネルギー・周波数領域エントロピーが用いられている。次に、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を学習する。そして、これら50個のHMMで得られた識別結果からメタ学習器を学習する。ここで、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して、メタ学習器を構成する。分類する行動は、「ベッド運搬」や「採血」・「歩行介助」などの22種類である。使用したセンサーは3軸加速度センサーである。センサーの取付位置は、看護師の両手首・胸ポケット・腰の4箇所である。データ数は合計1097である。メタ学習器の識別率は分類クラス数が多いためか、62.8%と低い。

3. 半教師ありカーネルロジスティック回帰

$\{(\mathbf{x}_i, y_i)\}_{i=1}^l$ をラベルありデータ、 $\{\mathbf{x}_j\}_{j=l+1}^{l+u}$ 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。 K をクラス数とすると、ソフトマックス関数を用いて、 $\mathbf{T} = \{\mathbf{t}_i\}_{i=1}^l$ と $\mathbf{Z} = \{\mathbf{z}_j\}_{j=l+1}^{l+u}$ との完全データ対数尤度が以下で書き表せる。

$$\ln p(\mathbf{T}, \mathbf{Z} | \Phi, \mathbf{W}) = \sum_{i=1}^l \sum_{k=1}^K t_{ik} \ln \left(\frac{\exp(\mathbf{w}_k^T \phi(\mathbf{x}_i))}{\sum_{c=1}^K \exp(\mathbf{w}_c^T \phi(\mathbf{x}_i))} \right) + \sum_{j=l+1}^{l+u} \sum_{k=1}^K z_{jk} \ln \left(\frac{\exp(\mathbf{w}_k^T \phi(\mathbf{x}_j))}{\sum_{c=1}^K \exp(\mathbf{w}_c^T \phi(\mathbf{x}_j))} \right).$$

ここで、 $\mathbf{t}_i = (t_{i1}, \dots, t_{iK})^T$ は y_i の指示変数(1-of- K)で、 $\mathbf{z}_j = (z_{j1}, \dots, z_{jK})^T$ は y_j の指示変数(潜在変数)である。また、 Φ は n 番目の行が $\phi(\mathbf{x}_n)^T$ で与えられる計画行列、 \mathbf{W} は k 番目の行が \mathbf{w}_k^T で与えられる重みパラメータの行列、 $\phi(\mathbf{x}_n) = (\phi_{n1}, \dots, \phi_{n(l+u)})^T$ 、 $\phi_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$ 、 $K(\cdot, \cdot)$ はカーネル関数である。本研究では、予備的検討により正解率が高くなる傾向が見られたため、カーネル関数としてガウスカーネルを使用し、その精度パラメータを0.1とした。

この完全データ対数尤度の z_{jk} は潜在変数であるため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完全データ対数尤度の期待値を最大化するパラメータを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、

$$p(\mathbf{Z} | \Phi, \mathbf{W}, \mathbf{T}) = \prod_{j=l+1}^{l+u} \prod_{k=1}^K \frac{(\exp(\mathbf{w}_k^T \phi(\mathbf{x}_j)))^{z_{jk}}}{\sum_{c=1}^K \exp(\mathbf{w}_c^T \phi(\mathbf{x}_j))}.$$

さらに、 z_{jk} の事後分布に関する以下の完全データ対数尤度の期待値を計算し、これを最大化するパラメータを求める(Mステップ)。

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

ただし、

$$b_{jk} \equiv \frac{\exp(\bar{\mathbf{w}}_k^T \phi(\mathbf{x}_j))}{\sum_{c=1}^K \exp(\bar{\mathbf{w}}_c^T \phi(\mathbf{x}_j))}, \quad \bar{\mathbf{w}}_k \equiv \mathbf{w}_k^{old},$$

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。

4. ラベルなしデータの選別手法

4.1 ランダムセレクション法

ランダムセレクション法は、ラベルなしデータのすべてを用いるのではなく、数に制限を置いてランダムに複数回選別し、ラベルありデータに対する尤度が最大になるラベルなしデータを使用して評価を行なう手法である。予備実験の結果より、選別するラベルなしデータの数は、ラベルありデータの数と同数とした。また、学習の回数は予備実験により適切と判断した50回とした。

4.2 半教師あり擬RVM

半教師あり擬RVMは、カーネル空間の基底ベクトル φ と観測値ベクトル \mathbf{t} とのコサイン距離を計算し、値の小さい基底ベクトルを計画行列から取り除いて学習を行なう手法である。コサイン距離が1に近ければ2つのベクトルは相関が高く、-1に近ければ相関が低い。半教師あり擬RVMでは、観測値ベクトルとの相関が低い基底ベクトルを計画行列から取り除く。コサイン距離は以下で書き表せる。

$$\cos(\varphi, \mathbf{t}) = \frac{(\varphi, \mathbf{t})}{|\varphi| |\mathbf{t}|},$$

ここで、 $\varphi_n = (\phi_{n1}, \dots, \phi_{nl})^T$ 、 $n \in \{l+1, \dots, l+u\}$ 、 $\mathbf{t}_k = (t_{1k}, \dots, t_{lk})^T$ 、 $k \in \{1, \dots, K\}$ である。 $k=1, \dots, K$ について、 φ_n と \mathbf{t}_k のコサイン距離を計算し、 K 個の値の中で最も大きい値を d_n とする。 $n=l+1, \dots, l+u$ について、 d_n を計算し、値の小さい基底ベクトルを計画行列から取り除く。予備実験の結果より、選別するラベルなしデータの数は、ラベルありデータの数と同数とした。

5. 評価実験

5.1 実験条件

本研究では、HASC (Human Activity Sensing Consortium) [Kawaguchi 12]の3軸加速度データを使用して、「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。データは、HASC2011corpusから、サンプリングレートが100Hzの3軸加速度センサーを腰に付けた被験者32名分を用いた。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、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: k-NN 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

先行研究 [Bao 04] や [池谷 08] を参考に、本研究では、以下で述べる 11 種類の特徴量を用いる。まず、重力成分を除去した運動加速度ベクトルから、ベクトル長・重力ベクトルとの内積値・重力ベクトルとの外積値を算出する。文献 [池谷 08] では、上記 3 種類の特徴量について、平均値や最大値・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。したがって、本研究では、25%値、中央値、75%値の 3 種類の統計量を用いる。また、「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して、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。

5.2 実験結果

5.2.1 k-nearest neighbor

k-nearest neighbor (k-NN) 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。 $k \in \{1, \dots, 10\}$ について検討した結果、 $k = 3$ のときに正解率が最も高くなり、このときの正解率の平均値は 89.8% である。これが学習による分類器の正解率の上限であると思われる。k-NN による confusion matrix の平均を表 1 に示す。分類するのが難しい「歩く」と「階段を上る」・「階段を下りる」に対しても正解率が高いことが分かる。

5.2.2 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% とした。各分類器の正解率を図 1 に示す。交差確認の結果、半教師あり擬 RVM が最も高い正解率を示した（ラベルありデータの割合が 5% のとき 65.7%、10% のとき 74.7%、15% のとき 79.4%）。

各分類器の再現率と適合率のマクロ平均を図 2 と図 3 に示す。いずれも正解率の結果と同様であることが分かる。

ラベルありデータの割合を 5% としたときの半教師あり擬 RVM の confusion matrix の平均を表 2 に示す。表 2 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」の混同が見られる。

5.2.3 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% とし、ラベルなしデータは用いずに、KLR 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 個、18 個、54 個である。SSKLR およびランダムセレクション法・半教師あり擬 RVM の正解率と

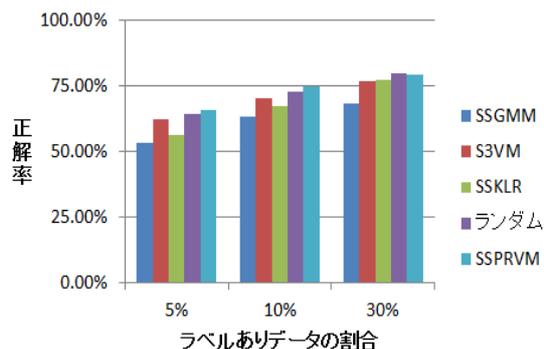


図 1: 各分類器の正解率。

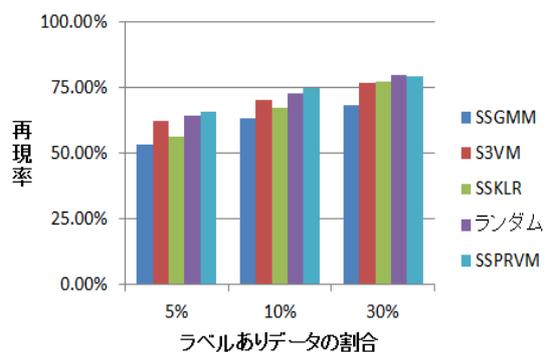


図 2: 各分類器の再現率。

もに、KLR の識別率を図 4 に示す。交差確認の結果、SSKLR は、KLR よりも低い正解率を示した。一方、ランダムセレクション法と半教師あり擬 RVM は、KLR よりも高い識別率を示した。

6. 議論

SSGMM は、全分類器の中で正解率が最も低い。これは、各クラスのデータが単一のガウス分布に従って生成されるという SSGMM の仮定を満たしていないことが原因であると考えられる。性別や身長・体重などの身体的特徴に依存して個人差が生じるため、行動中に得られるセンサデータは多峰的であり、さらに、クラスごとに峰の数が異なる。したがって、SSGMM の仮定を満たさず、正解率が低いと考えられる。

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

表 2: ラベルありデータの割合を 5% としたときの半教師あり擬 RVM の 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	0.5	0.5	14.5	19.5
jog	1.5	1.5	52.5	22.0	4.0	18.0
skip	1.0	3.5	14.0	65.0	3.0	13.5
stUp	13.0	19.0	0.0	0.0	52.0	16.0
stDown	2.5	19.0	0.0	0.5	12.0	66.0

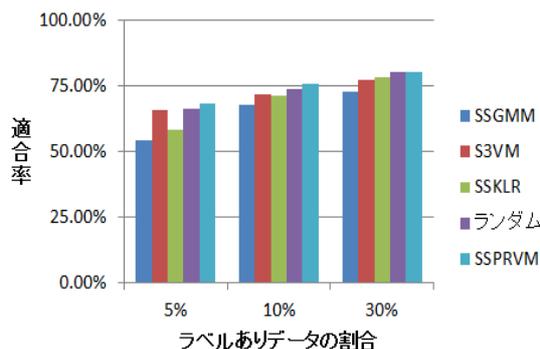


図 3: 各分類器の適合率.

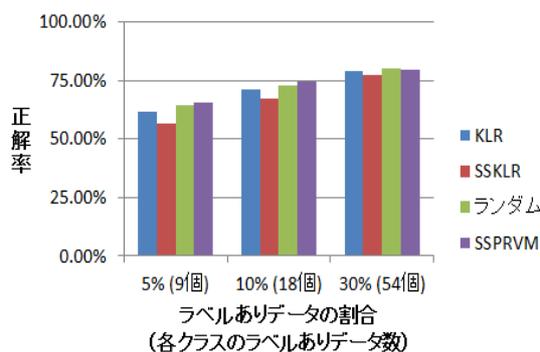


図 4: 各分類器の正解率.

すると、ラベルなしデータをすべて用いて学習を行なう SSKLR よりも、ラベルありデータのみを用いて学習を行なう KLR の方が正解率が高いことが分かる。また、疎なモデルである S3VM と比較すると、S3VM の方が正解率が高い。したがって、ラベルなしデータをすべて用いるのではなく、数に制限を置いて選別することによって、正解率が向上すると考えられる。

本研究では、ラベルなしデータを選別する手法として、ランダムセレクション法と半教師あり擬 RVM を提案した。これらの手法は、ラベルありデータのみを用いて学習を行なう KLR およびラベルなしデータをすべて用いて学習を行なう SSKLR よりも高い正解率を示した。ラベルなしデータをすべて用いるのではなく、数に制限を置いて選別することによって、ラベルありデータのみを用いて学習する場合よりも正解率が向上すると考えられる。ランダムセレクション法は、ラベルなしデータをランダムに選別しているが、半教師あり擬 RVM は、ランダムセレクション法より高い正解率を示していることから、分類に役立つラベルなしデータを効率的に選別していることが分かる。また、ランダムセレクション法は、複数回学習を行なうので時間がかかる。

関連ベクトルマシン (RVM; Relevance Vector Machine) は、多クラス教師あり学習の一手法であり、疎なカーネルベースのベイズ流学習手法である。RVM では、観測値ベクトルとの相関が低い基底ベクトルが計画行列から取り除かれる。RVM は、ラプラス近似の共分散行列を求めるときに、逆行列を計算しなければならないので学習に時間がかかる。しかし、半教師あり擬 RVM は、逆行列を計算する必要がないので、計算量

を軽減したモデルである。

7. おわりに

本研究では、分布に対する仮定を置かず、多クラス分類に適用することができ、予測に対する確率を計算することができる半教師あり学習法として、SSKLR を提案した。また、ラベルなしデータの選別手法として、ランダムセレクション法と半教師あり擬 RVM を提案した。ランダムセレクション法は、ラベルなしデータのすべてを用いるのではなく、数に制限を置いてランダムに複数回選別し、ラベルありデータに対する尤度が最大になるラベルなしデータを使用して評価を行なう手法である。半教師あり擬 RVM は、カーネル空間の基底ベクトルと観測値ベクトルとのコサイン距離を計算し、値の小さい基底ベクトルを計画行列から取り除いて学習を行なう手法である。これらの手法と、半教師あり学習の既存手法である SSGMM と S3VM のそれぞれについて、行動中に得られた加速度データを用いて分類性能の比較検討を行なった。交差確認の結果、半教師あり擬 RVM は、既存手法およびランダムセレクション法より高い正解率を示し、センサデータからの行動推定における有用性を示した。

RVM は、ラプラス近似の共分散行列を求めるときに、逆行列を計算しなければならないので学習に時間がかかる。しかし、半教師あり擬 RVM は、逆行列の計算をする必要がないので、RVM よりも計算量が少ないモデルである。今後は、半教師あり擬 RVM と RVM の理論的な関係について研究する予定である。

参考文献

- [Bao 04] Bao, L. and S. S. Intille (2004). Activity recognition from user-annotated acceleration data. *Proceedings of the Second International Conference on Pervasive Computing*, 3001, 1-17.
- [Zhu 09] Zhu, X. and A. B. Goldberg (2009). *Introduction to Semi-Supervised Learning*. Morgan & Claypool Publishers.
- [Maekawa 13] Maekawa, T. and S. Watanabe (2013). Training data selection with user's physical characteristics data for acceleration-based activity modeling. *Journal of Personal and Ubiquitous Computing*, 17, 3, 451-463.
- [上田 13] 上田修功・田中祐典・中島直樹 (2013). メタ学習に基づく加速度センサからの看護師行動識別. マルチメディア, 分散, 協調とモバイル (DICOMO 2013) シンポジウム論文集, 663-667.
- [Kawaguchi 12] Kawaguchi, N. et al. (2012). HASC 2012 corpus: Large scale human activity corpus and its application. *Proceedings of the Second International Workshop of Mobile Sensing: From Smartphones and Wearables to Big Data*, 10-14.
- [池谷 08] 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. 電子情報通信学会研究報告, ユビキタス・センサネットワーク (USN), 108, 138, 75-80.