

Semi-Supervised Sparse Bayesian Learning と そのセンサデータへの適用

Semi-Supervised Sparse Bayesian Learning and its Application to Sensor Data

伊藤 翔*¹ 角所 考*¹ 岡留 剛*¹
Sho ITO 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

Semi-supervised sparse Bayesian learning, proposed here, for multiclass classification takes the form of a linear combination of basis functions transformed by the softmax function. The basis functions are given by kernels, with one kernel in the model associated with each of the labeled and unlabeled points from the training set. Using the variational Bayesian method with the Laplace approximation enables us to determine the distributions of the posterior parameters in the model. Several tests for the learning model for artificial data and sensor data show its high generalization ability.

1. はじめに

教師あり学習は、明確な分類基準があるため精度の高い分類が可能であるが、量の増加にともないラベリングコストが大きくなる。そのため、少数のラベルありデータと、多数のラベルありデータから構成される訓練データで学習することにより、教師あり学習と同程度またはそれ以上の精度を持つ分類器を作成する技術である半教師あり学習が注目されている。

半教師あり学習の研究において、様々な学習手法が提案されている (Zhu 2009)。使用する半教師あり学習の手法における仮定が訓練データに合えば高い推定精度が見込めるが、「正しい」仮定をすることは未解決問題である。

本稿では、センサデータから人の行動推定を行なうことを目的に、半教師あり学習の枠組みを援用する。標準的なアプローチとして、データがある特定のモデルに従っていると仮定し、そのモデルを学習する方法がある。しかし、センサデータが従うモデルは自明ではない。別アプローチのグラフベースの学習手法では、その仮定としてデータに重なりがないことが挙げられる。人の行動には個人差が大きく出ることから、また識別する行動の数が多くなると、異なるクラスのデータ間に重なりがない特徴空間にデータを写像することは容易ではない。2クラス分類を複数個用意して多クラス分類を行なう方法も提案されているが、これは曖昧な分類領域が生じてしまったり、一意のクラスが定まらないなどの問題が生じる。

以上に鑑み、本研究ではより仮定の少ないモデルを作成する。本稿では、多クラスの拡張が容易かつ SVM と同等の汎化性能を持つ sparse Bayesian learning (Tipping 2001) に着目し、それを半教師あり学習に拡張した semi-supervised sparse Bayesian learning (S3BL) を提案する。画像の2クラス分類に対して同様のアプローチをとった研究として (Rin 2011) がある。

2. Semi-supervised sparse Bayesian learning

2.1 モデル

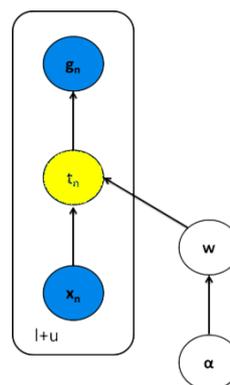


図 1: S3BL のグラフィカルモデル。青のノードは常に観測されるノードであり、黄色のノードは g_n が 0 のとき観測されるノードである。

訓練データは、ラベルありデータとラベルなしデータの両方からなり、それぞれを $\{\mathbf{x}_i, t_i\}_{i=1}^l, \{\mathbf{x}_j\}_{j=l+1}^{l+u}$ と表し、 $L = \{1, \dots, l\}$, $U = \{l+1, \dots, l+u\}$ とする。また、(Lawrence 2005) に従い、観測変数 g_n を導入する。(これについては後に詳述する)。

ここで、予測 $y(\mathbf{x}, \mathbf{w})$ が基底関数 $\phi_n(\mathbf{x})$ の線形和で表されるモデルを考える。以下の議論では、 K クラス分類問題を扱い、 K 値の目標変数 $t \in \{1, \dots, K\}$ を出力するものと仮定する。 K クラスの分類問題を解くために、ソフトマックス関数を用いる。

$$y_k(\mathbf{x}) = \frac{\exp(\mathbf{w}_k^T \phi(\mathbf{x}))}{\sum_{c=1}^K \exp(\mathbf{w}_c^T \phi(\mathbf{x}))}, \quad (1)$$

ここで、重み \mathbf{w} はモデルのパラメータであり、

連絡先: 氏名: 伊藤 翔

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: s.ito@kwansei.ac.jp

$\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_K)^T$ とする。また、 ϕ_n は $\sum_{n=1}^{l+u} \phi_n(\mathbf{x}) = \sum_{n=1}^{l+u} k(\mathbf{x}, \mathbf{x}_n)$ であり、 $k(\cdot, \cdot)$ はカーネル関数である。

S3BL のグラフィカルモデルは図 1 で表現される。我々の目標は、高い汎化性能を持つ、確率変数 \mathbf{W} で重み付けされた、観測データ \mathbf{x} からラベル t に写像する分類器を求めることである。 t_j , $j \in U$ は欠損値であるが、観測値 g_j , $j \in U$ が共通の祖先にあるので、 \mathbf{W} と \mathbf{x}_j , $j \in U$ は条件付き依存となる。そのため、ラベルなしデータが分類の際に役に立つ。このとき、 $\mathcal{D}_l = \{(\mathbf{x}_i, t_i, g_i)\}_{i \in L}$, $\mathcal{D}_u = \{(\mathbf{x}_j, g_j)\}_{j \in U}$ と表すと、観測変数すべては $\mathcal{D} = \mathcal{D}_l \cup \mathcal{D}_u$ となる。そして潜在変数は、 $\mathcal{Z} = \{t_j, \mathbf{W}, \Lambda\}_{j \in U}$ と表すと、このモデルの同時分布は以下で書き表せる。

$$p(\mathcal{D}, \mathcal{Z}) = p(\mathbf{W}|\Lambda)p(\Lambda) \prod_{i=1}^l p(t_i|\mathbf{W}, \mathbf{x}_i) \times \prod_{j=l+1}^{l+u} p(g_j|t_j)p(t_j|\mathbf{W}, \mathbf{x}_j), \quad (2)$$

ここで、 \mathbf{W} の各要素はガウス分布に従うとし、 Λ はそれに対応する精度で、 $\Lambda = (\alpha_1, \dots, \alpha_K)^T$ とする。以下で、観測データ \mathcal{D} が与えられたときの潜在変数 \mathcal{Z} の事後分布 $p(\mathcal{Z}|\mathcal{D})$ を求める。さらに、この分布を用いた期待値を求める。それは、ベイズの規則により導出できるが、この場合解析的に解くことはできない。そのため、変分ベイズを用いて近似解を求める。

2.1.1 付加する観測変数

S3BL では、変数 t_n が観測されているとき 0 を、そうでないときは 1 を取る変数として、観測変数 g_n を導入する。さらに、以下の条件付き分布を定義する。

$$p(g_j = 1|t_j) = \begin{cases} \exp(\mu^1) / \sum_{c=1}^K \exp(\mu^c + \gamma_c) & \text{if } t_j = 1 \\ \vdots \\ \exp(\mu^K) / \sum_{c=1}^K \exp(\mu^c + \gamma_c) & \text{if } t_j = K \end{cases}$$

ここで γ_c は 0 以上の変数である。 g の条件つき確率は、あるクラスにおけるラベルなしデータである確率を、つまり、クラス事前確率の役割を果たす。

2.2 変分ベイズ法による semi-supervised sparse Bayesian learning の解法

パラメータベクトル \mathbf{W} の事前確率分布として、平均 0 のガウス事前分布を用いる。1 つの超パラメータの代わりに個々の重みパラメータ w_{kn} ごとに異なった超パラメータを用いる。すなわち、重みに対する事前分布は、

$$p(\mathbf{W}|\Lambda) = \prod_{k=1}^K p(\mathbf{w}_k|\alpha_k) \quad (4)$$

と表される。精度 Λ の事前分布として、共役事前分布のガンマ分布を用いる。すなわち、

$$p(\Lambda) = \prod_{k=1}^K p(\alpha_k) = \prod_{n=1}^{l+u} \prod_{k=1}^K \text{Gam}(\alpha_{nk}|a_{nk}, b_{nk}). \quad (5)$$

すべての観測変数と潜在変数の同時分布は以下で表される。

$$p(\mathcal{D}, \mathcal{Z}) \propto p(\mathbf{W}|\Lambda)p(\Lambda) \prod_{i=1}^l p(t_i|\mathbf{W}, \mathbf{x}_i) \prod_{j=l+1}^{l+u} p(g_j|t_j)p(t_j|\mathbf{W}, \mathbf{x}_j) = \prod_{k=1}^K \mathcal{N}(\mathbf{w}_k|\mathbf{0}, \alpha_k) \text{Gam}(\alpha_k|a, b) \times \prod_{i=1}^l \left(\frac{\exp(\mathbf{w}_k^T \phi(\mathbf{x}_i))}{\exp(\sum_c \mathbf{w}_c^T \phi(\mathbf{x}_i))} \right)^{z_{ik}} \times \prod_{j=l+1}^{l+u} \left(\frac{\exp(\mathbf{w}_k^T \phi(\mathbf{x}_j) + z_{jk}\mu_k)}{\exp(\sum_c \mathbf{w}_c^T \phi(\mathbf{x}_j) + \mu_c + \gamma_c)} \right)^{z_{jk}}, \quad (6)$$

ここで、 z_{nk} は n 番目のデータがクラス k に属するときに 1、それ以外は 0 をとる指示変数である。

さて、事後分布 $p(\mathbf{W}, \Lambda, \{y_j\}_{j \in U} | X, \{y_i\}_{i \in L}, \mathbf{g})$ を $q(\mathbf{W}, \Lambda, \{y_j\}_{j \in U}) = q(\mathbf{W})q(\Lambda) \prod_{j=l+1}^{l+u} q(y_j)$ のように分解した変分事後分布を考える。変分ベイズ法により、この分布の各因子を再推定する方程式を求めることができ、以下の変分事後近似が求まる。

$$q^*(\Lambda) = \prod_{n=1}^{l+u} \prod_{k=1}^K \text{Gam}(\alpha_{nk}|\tilde{a}_{nk}, \tilde{b}_{nk}), \quad (7)$$

$$q^*(t_j) \propto \prod_{k=1}^K \exp\left\{z_{jk}(\mathbb{E}_{\mathbf{w}_k}[\mathbf{w}_k]^T \phi(\mathbf{x}_j) + \mu_k)\right\}, \quad (8)$$

$$\ln q^*(\mathbf{w}_k) = -\frac{1}{2} \mathbf{w}_k^T \mathbf{A}_k \mathbf{w}_k + \ln \sum_{i=1}^l z_{ik} \frac{\exp(\mathbf{w}_k^T \phi(\mathbf{x}_i))}{\sum_{c=1}^K \exp(\mathbf{w}_c^T \phi(\mathbf{x}_i))} + \ln \sum_{j=l+1}^{l+u} \mathbb{E}_{t_{jk}}[z_{jk}] \frac{\exp(\mathbf{w}_k^T \phi(\mathbf{x}_j) + \mathbb{E}_{t_{jk}}[z_{jk}]\mu_k)}{\sum_{c=1}^K \exp(\mathbf{w}_c^T \phi(\mathbf{x}_j) + \mathbb{E}_{t_{jc}}[z_{jc}](\mu_c + \gamma_c))}, \quad (9)$$

ここで、 $\tilde{a}_{nk} = a - \frac{1}{2}$, $\tilde{b}_{nk} = b + \frac{1}{2} \mathbb{E}_{w_{nk}}[w_{nk}^2]$, $\mathbf{A}_k = \text{diag}(\mathbb{E}_{\alpha_{1k}}[\alpha_{1k}], \dots, \mathbb{E}_{\alpha_{nk}}[\alpha_{nk}])$ とした。

$q^*(\mathbf{w}_k)$ が解析的に積分できないことから、 \mathbf{w}_k に関する期待値を計算できない。ここではラプラス近似によりガウス分布で近似することで、各変数の期待値を求める。 \mathbf{w}_k に関するヘシアンを求めると、

$$\nabla_{\mathbf{w}_k} \nabla_{\mathbf{w}_k} \ln q^*(\mathbf{w}_k) = -\mathbf{A}_k - \sum_{i=1}^l \frac{\exp(\mathbf{w}_k^T \phi(\mathbf{x}_i))}{\sum_{c=1}^K \exp(\mathbf{w}_c^T \phi(\mathbf{x}_i))} \times \left(1 - \frac{\exp(\mathbf{w}_k^T \phi(\mathbf{x}_i))}{\sum_{c=1}^K \exp(\mathbf{w}_c^T \phi(\mathbf{x}_i))} \right) \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_i) - \sum_{j=l+1}^{l+u} \frac{\exp(\mathbf{w}_k^T \phi(\mathbf{x}_j) + \mathbb{E}_{t_{jk}}[z_{jk}]\mu_k)}{\sum_{c=1}^K \exp(\mathbf{w}_c^T \phi(\mathbf{x}_j) + \mathbb{E}_{t_{jc}}[z_{jc}](\mu_c + \gamma_c))} \times \left(1 - \frac{\exp(\mathbf{w}_k^T \phi(\mathbf{x}_j) + \mathbb{E}_{t_{jk}}[z_{jk}]\mu_k)}{\sum_{c=1}^K \exp(\mathbf{w}_c^T \phi(\mathbf{x}_j) + \mathbb{E}_{t_{jc}}[z_{jc}](\mu_c + \gamma_c))} \right) \phi(\mathbf{x}_j)^T \phi(\mathbf{x}_j). \quad (10)$$

となる。これにより、近似分布は以下となる。

$$q^*(\mathbf{w}_k) = N(\mathbf{w}_k | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (11)$$

ここで、 $\boldsymbol{\mu}_k = \mathbf{w}_k^{MAP}$, $\boldsymbol{\Sigma}_k = (\nabla_{\mathbf{w}_k} \nabla_{\mathbf{w}_k} \ln q^*(\mathbf{w}_k))^{-1}$ である。 \mathbf{w}_k の MAP 解については Nelder-Mead 法で求める。変分事後分布を求めるには、まず $q(\mathbf{W})$ もしくは $q(t_j), q(\boldsymbol{\alpha})$ のどちらかの分布のパラメータを初期化し、これらを適当な収束基準が満たされるまで交互に再推定していく。

3. 人工データにおけるクラス推定

3.1 実験

人工的に生成したデータセットを用いて、既存の半教師あり学習手法と S3BL の精度比較を行なう。今回の実験で用いた既存の半教師あり学習手法は、semi-supervised Gauss mixture model (SSGMM) と Laplacian regularized least square (LapRLS) とした。

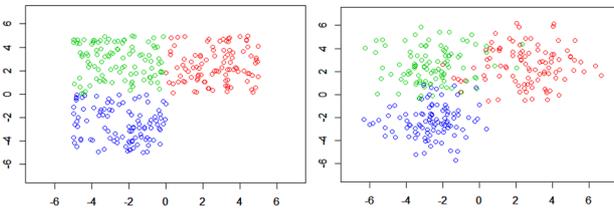


図 2: 一様分布から生成したデータ。決定境界はそれぞれのデータ。少しの重なりのあるものを用意した。

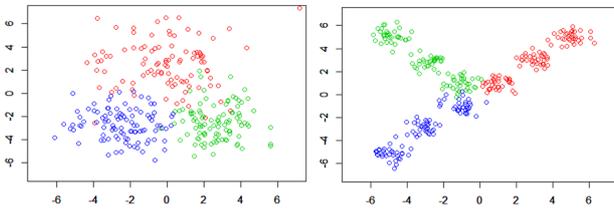


図 4: ガウス分布から生成したデータ。赤のクラスの範囲を他のクラスより多くした。

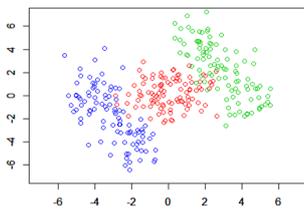


図 6: ガウス分布と混合ガウス分布から生成したデータ。赤のクラスはガウス分布から生成したデータ、その他のクラスは 2 要素の混合ガウス分布とした。

図 2 から図 6 に示す人工データを用いて半教師あり学習を行なった。訓練データ数は各クラス 30 個の総計 90 個、テストデータ数とともに各クラス 100 個の総計 300 個とした。ラベルありデータの割合を 2 個, 3 個, 4 個とした。また精度については、学習を 10 回行ない、その精度の平均とした。

各学習におけるパラメータの設定を以下に述べる。SSGMM のクラス事前確率は等確率とした。LapRLS ではカーネル関数として RBF カーネルを使用し、その精度パラメータを 1 とした、また、グラフ作成には、k-近傍法を使用し、k を 3 とした。LapRLS のパラメータ数が多く、最も汎化性能が高くなるパラメータ設定をすることは容易ではない。そのため、いくつかの組み合わせを試し、相対的に精度のよかったものを使用した。S3BL について、カーネル関数として RBF カーネルを用いて、その精度パラメータを 1 とした。クラス事前確率は等確率とした。

3.2 結果

各手法で訓練データから学習し、テストデータで精度を求めた。その結果を表 1 から表 5 に示す。

手法	数	精度	手法	数	精度
SSGMM	2 個	×	SSGMM	2 個	×
	3 個	77.8%		3 個	73.1%
	4 個	81.1%		4 個	85.6%
LapSVM	2 個	68.7%	LapSVM	2 個	56.5%
	3 個	79.4%		3 個	68.8%
	4 個	83.7%		4 個	83.7%
S3BL	2 個	93.7%	S3BL	2 個	99.4%
	3 個	94.1%		3 個	99.0%
	4 個	89.0%		4 個	99.4%

表 1: 図 2 の一様分布から生成したデータで学習したときの、各手法におけるテストデータの予測精度。

手法	数	精度	手法	数	精度
SSGMM	2 個	×	SSGMM	2 個	×
	3 個	76.5%		3 個	84.3%
	4 個	80.4%		4 個	72.2%
LapSVM	2 個	56.9%	LapSVM	2 個	44.6%
	3 個	69.9%		3 個	61.2%
	4 個	72.1%		4 個	67.8%
S3BL	2 個	87.2%	S3BL	2 個	88.7%
	3 個	87.9%		3 個	90.3%
	4 個	90.1%		4 個	88.4%

表 3: 図 4 のガウス分布から生成したデータで学習したときの、各手法におけるテストデータの予測精度。

手法	数	精度
SSGMM	2 個	×
	3 個	72.3%
	4 個	88.6%
LapSVM	2 個	46.2%
	3 個	51.1%
	4 個	58.6%
S3BL	2 個	81.7%
	3 個	83.8%
	4 個	84.2%

表 5: 図 6 のガウス分布と混合ガウス分布から生成したデータで学習したときの、各手法におけるテストデータの予測精度。

この表から、S3BL が平均的にかなり高い推定精度を示していることが分かる。S3BL はあらゆるデータセットに対して良

い汎化性能であった。表 2 の SSGMM の結果、学習の仮定とデータがあっても、精度は思わしくなかった。これは、ラベルありデータがかなり少ないために、ラベルありデータの位置により、予測精度にかなり影響したことが原因である。S3BL では少数のラベルありデータに対しても安定した汎化性能を誇ることが分かった。S3BL はカーネル関数やパラメータの設定により、より汎化性能の高い分類器が作成できることが期待できる。

4. センサデータによる行動推定

4.1 実験

実験データとして、HASC Challenge2010 (Kawaguchi 2011) によって得られたデータを用いる。行動推定として、「静止」・「歩行」・「走行」の 3 種類を分類する 3 クラス分類を行なう。学習データ数はそれぞれ 30 個の総計 90 個とし、テストデータはそれぞれ 100 個の総計 300 個とする。用いる既存の半教師あり手法並びにそのパラメータ設定は上の実験と同様とした。S3BL のカーネル関数として上と同様に RBF カーネルを用い、その精度パラメータを 2 とした。HASC データの特徴量として、加速度のノルムの平均値とエネルギー、エントロピーエネルギーの 3 つを使用した。HASC データのセンサ取付位置については指定しなかった。

4.2 結果

各手法で訓練データから学習し、テストデータで精度を求めた。その結果を表 6 に示す。

手法	数	精度
SSGMM	2 個	×
	3 個	69.8%
	4 個	78.2%
LapSVM	2 個	62.7%
	3 個	70.9%
	4 個	72.2%
S3BL	2 個	74.3%
	3 個	77.0%
	4 個	80.1%

表 6: HASC データで学習したときの、各手法におけるテストデータの予測精度。

人工データと同様に HASC データでも、S3BL は他の半教師あり手法と比べてよい汎化性能を誇った。このことから、より仮定の小さいモデルを用いることが、半教師あり学習における行動推定でもかなり有効な手法であると分かった。

5. 議論

既存の半教師あり学習の手法では、データに対する仮定が正しい場合でも予測精度がよくない結果もいくらか観測された。例えば、表 2 の SSGMM において、仮定は正しいが、汎化性能は S3BL より低い結果となっている。

SSGMM の学習の結果、ラベルありデータが密になっているクラスでは分散は狭く、ラベルありデータが広い範囲に存在しているクラスでは分散は広くなる。そのため、本来ラベルありデータが密になっているクラスのデータのうち、ラベルありデータと離れているデータが別のクラスとして予測されるとい現象が起きる。

LapRLS もグラフベースな学習であることから、予測精度が良くないときに同様の現象が観測された。既存の学習法ではラベルありデータの位置が予測精度にかなり依存する。

しかし S3BL では、基底関数にカーネル関数と、クラス事前確率の導入により、ラベルありデータの位置が予測精度にあまり影響を与えず安定した挙動を示す。

S3BL の利点はまず、強い仮定のモデルを仮定しないことにある。そのため、データについて知見が得られないような場合でも、高い汎化性能を示す。さらに、新しいデータ点について確率的な出力を行なうことも長所である。

一方、欠点は、計算量の大きさが挙げられる。変分ベイズにおける再推定により、データ数を N とすると、 N 次元の重みベクトルの最適化や (N, N) 行列の逆行列の計算を何度も行なう必要がある。また、パラメータの最適計算において、その初期値の与え方によってはローカルミニマムに落ち、解が求まらない場合がある。さらに、解析解が直接得られないことも欠点である。本研究では変分ベイズに加えさらにラプラス近似により解を求めた。この解の「真の解」に対する近似の程度が分からない。

6. 終わりに

本稿では、新たな半教師あり学習手法について提案し、人工データとセンサデータによるクラス推定により性能の高さを示した。具体的には、SBL を新たな観測変数を追加し、半教師あり学習に拡張した。解析的に解けないことから、変分ベイズ法とラプラス近似を援用した解法を用いた。そして人工データとセンサデータを用いてクラス推定・行動推定を行ない、他の半教師あり学習との比較によりその有用性を確認した。

今後は、大規模での実データでの検証、さらには、変分ベイズ法とラプラス近似の援用による解法だけでなく、MCMC での定式化も行なう予定である。

参考文献

- [Kawaguchi 2011] Kawaguchi, N., N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, S. Inoue, Y. Kawahara, Y. Sumi, and N. Nishio.: HASC Challenge: Gathering large scale human activity corpus for the real-world activity understandings, *proceedings of ACM AH 2011*, pp 27:1-27:5 (2011)
- [Lawrence 2005] Lawrence, N. D. and M. I. Jordan.: Semi-supervised learning via Gaussian processes, *Advances in neural information processing systems*, 17 Cambridge, MA: MIT Press (2005)
- [Rin 2011] Rin, D., K. hassan, and C. Pal.: Semi supervised learning in wild faces and videos, *Proceedings of the British Machine Vision Conference*, 3 pp 1-12 (2011)
- [Tipping 2001] Tipping, M. E.: Sparse bayesian learning and the relevance vector machine, *Journal of Machine Learning Research*, 1 pp 211-244 (2001)
- [Zhu 2009] Zhu, X. and A. B. Goldberg.: *Introduction to Semi-Supervised Learning*, Morgan & Claypool Publishers (2009)