1P2-2in

PSIの確率的表現とクラス分類への応用

Probabilistic expression of PSI and application for classification

箕浦 健太郎[*1] Kentaro Minoura
田村 哲嗣[*2] Satoshi Tamura
速水 悟[*2] Satoru Hayamizu

[*1] 岐阜大学工学研究科応用情報学専攻 Div. Information Science, Gifu Univ.
[*2] 岐阜大学工学部応用情報学科 Dep. Information Science, Gifu Univ.

We propose probabilistic Polynomial Semantic Indexing (pPSI) as a probabilistic expression of PSI. pPSI can stabilize output and classification precision, perform model-based training, promote the efficiency of semi-supervised learning, and use a parameters' approximation based on the amount of data. Making use of these peculiarities, we apply this method for classification. PSI is a model weighting pairs of needed input variables to approximate tensors as products of matrixes. The tensors are parameters to express weight between input variables' elements. For this reason, we can treat high-dimensional data without a process such as dimension reduction and figure extraction.

1. はじめに

機械学習において, クラス分類は音声処理, 音楽情報処理, 画像処理, テキスト情報処理など幅広い分野において利用されている重要なタスクである.

本稿では, 検索システムにおいてクエリに対して適切な Web ページを表示するためのページランク付け手法として提案された Polynomial Semantic Indexing(PSI;多項意味索引)[Bai 2009]をクラス分類へと応用する手法について提案する. PSI はパラメータであるテンソルを行列の積として近似することで次元数の大きいデータに対しても圧縮などの処理を行うことなく適応することが可能である. この近似は潜在的概念空間への写像とその空間上での内積及びその拡張(以降拡張内積と呼ぶ)によって定義される. この特徴により次元数の大きいデータに対しても次元圧縮(次元低減), 特徴抽出などの処理を行うことなく扱うことができる. 本稿では, この PSI について, 指数関数, 事後確率を用いて確率的表現 probabilistic PSI(pPSI;確率的多項意味索引)を提案する. これにより, 出力, 及び精度の安定化, モデルに基づく訓練, 半教師あり学習の効率化, データ数に応じた推定手法の利用が可能になるなどのメリットが挙げられる.

本稿では, 人工的に生成したデータ, 及びリポジトリのデータを用いて pPSI, PSI, 及び近年多くの分野で評価され多用されている SVM(多クラスに対応した m-SVM)のクラス分類性能を比較する. これによって pPSI の有効性を示す.

2. 関連研究

2.1 潜在的概念空間

PSI は訓練によってN次元潜在的概念空間を形成する. 同様にN次元概念空間を形成する手法に後述する Latent Semantic Indexing(LSI;潜在的意味索引)[Deerwester 1990], probabilistic LSI(pLSI; 確率的潜在的意味索引)[Hofmann 1999]などがある.

これらはいずれもテキスト情報処理のための手法であるが, ラベル付けされていないデータに基づく教師なしモデルであり, パラメータが学習データのみに依存する(このためデータの解析に向いているといえる). したがって限られたデータに基づいた結果となる. また処理後のデータの扱いには関与しないため, 目的とするタスクに直接的に作用しないという欠点がある.

(1) Latent Semantic Indexing(LSI)

LSI は, 処理の対象とするデータの特異値分解(SVD)によって得た特異値の大小関係を元にして優先的に使用する潜在的概念空間軸を定めることによって概念空間を生成し, その概念空間への写像によって元のデータをN次元へと次元圧縮する手法である. 多くはテキスト情報処理に利用されている. これは最小二乗法によって計算される.

(2) probabilistic LSI(pLSI)

pLSI は, 処理の対象とするデータに内在するトピックを潜在パラメータとして用いることによって, データの生成モデルを表現したものである. これは最尤法に基づいて EM アルゴリズムなどを利用して計算される. この場合, パラメータが SVD と同様の形式で表現されるため, それを用いてN次元概念空間を導き出すことができ, これによりデータに内包されたトピックを抽出することが可能である.

2.2 クラス分類

本稿の目的であるクラス分類では, 一般的に Support Vector Machine(SVM), Naive Bayes Classifier(単純ベイズ分類器)などの分類器が用いられている.

(1) Support Vector Machine(SVM)

SVM は 2 クラス分類に用いられ, 学習データをクラス毎に各データ点との距離が最大となる平面(分離平面)によって判別する手法である. SVM の特徴的な点は, この分離平面に最も近いデータとの距離(マージン)を最大にするように分離平面を決定する点にある. これにより高精度な分類を実現している. また多クラスに対する適応についても様々な手法が編み出されている.

(2) Naïve Bayes Classifier

Naïve Bayes Classifier は, 入力ベクトルの各要素に対して強い独立性を仮定したモデルであり, 特定のクラスに該当する確率は入力ベクトルの要素毎の条件付き確率(入力ベクトルが既知のとき, 対象のクラスに該当する確率)の積で表現される. このモデルは学習に多くのデータを必要とせず学習が高速である点でメリットがある.

連絡先：箕浦建太郎，岐阜大学工学研究科応用情報学専攻速水・田村研究室，岐阜県岐阜市柳戸 1-1，058-293-2763，mino@asr.info.gifu-u.ac.jp

3. Polynomial Semantic Indexing(PSI)

3.1 拡張内積

PSI の定義を行うのに先立ち, 定義において利用する拡張内積について定義する. 要素数Iのベクトル$\boldsymbol{x_h}(h = 1 \ldots H)$があったとき, 拡張内積を(1)のように定義する.

$$<_{h=1}^{H} \boldsymbol{x_h} >= \sum_{i=1}^{I} \prod_{h=1}^{H} \boldsymbol{x_{h_i}} \tag{1}$$

ただし添字iはi番目の要素を示す. この拡張内積は, $H = 2$のときに内積となる.

3.2 定義

K次の Polynomial Semantic Indexing(PSI; 多項意味索引)の近似は次のような関数として定義される(元の定義については[Bai 2009]に詳しい). 任意個数$M(\geqq 2)$の入力変数ベクトル$\boldsymbol{x_{m \leq M}}$があるとき, 2 個以上$K$個以下の必要とする重複組み合わせの集合$V$について考える. このとき考えうる組み合わせの個数をG_{max}, うち必要とする組み合わせの個数$|V|$をGとする. Vのうち, g番目の組み合わせに含まれるK_g個の入力変数ベクトル$\boldsymbol{x_{gk_g}}$をそれぞれ行列$\boldsymbol{X_{gk_g}}$によってN次元の潜在的概念空間上へと写像する. それらの拡張内積を, G個全てについて計算し, その和をとったものが PSI の関数の値である. これを式で表すと(2)となる.

$$f^{[K]}(\boldsymbol{x_1}, \ldots, \boldsymbol{x_M}) = \sum_{g=1}^{G} <_{k_g=1}^{K_g} \boldsymbol{X_{gk_g}} \boldsymbol{x_{gk_g}} > \tag{2}$$

ここで, G_{max}: $\sum_{k=2}^{K} {}_{M}H_k$となる. また$\boldsymbol{X_{gk_g}}$はその行列によって写像されるベクトル$\boldsymbol{x_{gk_g}}$の次元数を$Dim.$とすると, $\boldsymbol{X_{gk_g}} \in R^{N \times Dim.}$である.

3.3 例

ここでは, $K = 3$の場合について具体的に述べる. なお, 入力変数を$\boldsymbol{x}$, $\boldsymbol{y}$とし, $\boldsymbol{y}$よりも$\boldsymbol{x}$を重視するものとして, $V = \{(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{y}), (\boldsymbol{x}, \boldsymbol{y})\}$とおく. これを式で表すと(3)となる.

$$f^{[3]}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{N} (\boldsymbol{Ax})_i (\boldsymbol{Bx})_i (\boldsymbol{Cy})_i + \sum_{i=1}^{N} (\boldsymbol{Dx})_i (\boldsymbol{Ey})_i \tag{3}$$

これは, (2)において$M = 2$, $G = 2$, $K_1 = 3$, $\boldsymbol{X_{11}} = \boldsymbol{A}$, $\boldsymbol{x_{11}} = \boldsymbol{x}$, $\boldsymbol{X_{12}} = \boldsymbol{B}$, $\boldsymbol{x_{12}} = \boldsymbol{x}$, $\boldsymbol{X_{13}} = \boldsymbol{C}$, $\boldsymbol{x_{13}} = \boldsymbol{y}$, $K_2 = 2$, $\boldsymbol{X_{21}} = \boldsymbol{D}$, $\boldsymbol{x_{21}} = \boldsymbol{x}$, $\boldsymbol{X_{22}} = \boldsymbol{E}$, $\boldsymbol{x_{22}} = \boldsymbol{y}$と対応する.

3.4 訓練

このモデルの訓練は, 損失関数を margin rank loss[Herbrich 2000]によって定め, これを最小化することによって行う. 入力変数を$\boldsymbol{x}$, $\boldsymbol{y}$としたとき, ある$\boldsymbol{x}$について, $\boldsymbol{y}^-$より$\boldsymbol{y}^+$の方が高く評価されるべきであるとする. このとき, 損失関数$E(\theta)$は(4)のようになる.

$$E(\theta) = \sum_S \max(0, 1 - f^{[K]}(\boldsymbol{x}, \boldsymbol{y}^+; \theta) + f^{[K]}(\boldsymbol{x}, \boldsymbol{y}^-; \theta)) \tag{4}$$

ここでθはパラメータ(潜在的概念空間上へ写像する行列), Sはデータ集合である. この関数の最小化は確率的勾配法を用いてデータのペア$s = (\boldsymbol{x}, \boldsymbol{y}^+, \boldsymbol{y}^-) \in S$について(5)のように逐次的に行う.

$$if(1 - f^{[K]}(\boldsymbol{x}, \boldsymbol{y}^+; \theta^{(t)}) + f^{[K]}(\boldsymbol{x}, \boldsymbol{y}^-; \theta^{(t)}) > 0)$$

$$\theta^{(t+1)} \leftarrow \theta^{(t)} - \lambda^{(t)} \frac{\partial}{\partial \theta} E_{s \in S}(\theta^{(t)}) \tag{5}$$

ここでtはステップ数, $\lambda^{(t)} \leqq 1$は学習率パラメータである.

4. probabilistic PSI(pPSI)

4.1 定義

PSI を確率の枠組みで構成するため, その確率的表現として probabilistic PSI(pPSI; 確率的多項意味索引)を, 指数関数, 及び事後確率を用いて(6)のように定義する.

$$\Pr^{[K]}(\boldsymbol{y}|\boldsymbol{x}; \theta) = \frac{\exp(f^{[K]}(\boldsymbol{x}, \boldsymbol{y}; \theta))}{\sum_{\boldsymbol{y} \in Y} \exp(f^{[K]}(\boldsymbol{x}, \boldsymbol{y}; \theta))} \tag{6}$$

ここでYは有限集合であり, θはパラメータ(潜在概念空間上へ写像する行列)である. これは$\boldsymbol{x}$が既知のとき考え得るYにおける対象としている$\boldsymbol{y}$のスコアの割合を表現したものであり, $\boldsymbol{x}$が既知のときの$\boldsymbol{y}$の事後確率となっている.

4.2 訓練

このモデルの訓練は最尤推定法を用いて行う. ある$\boldsymbol{x}$に対して$\boldsymbol{y}$が$\boldsymbol{y}' \in Y'(= Y - \{\boldsymbol{y}\})$より高く評価されるべきであるとき, 更新は(7)によって逐次的に行う.

$$if(Pr^{[K]}(\boldsymbol{y}|\boldsymbol{x}; \theta) < e * \max_{\boldsymbol{y}' \in Y'} Pr^{[K]}(\boldsymbol{y}'|\boldsymbol{x}; \theta))$$

$$\theta^{(t+1)} \leftarrow \theta^{(t)} + \lambda^{(t)} \frac{\partial}{\partial \theta} l(\theta^{(t)}) \tag{7}$$

ただし, 更新条件は(6)を PSI について解いたものを$\boldsymbol{y}^+ \rightarrow \boldsymbol{y}$, $\boldsymbol{y}^- \rightarrow \boldsymbol{y}'$と置き換えた (5)に代入したものであり, (7)中の対数尤度の微分値$\frac{\partial}{\partial \theta} l(\theta)$は(8)のようになる.

$$\frac{\partial}{\partial \theta} l(\theta) = \frac{\partial}{\partial \theta} f^{[K]}(\boldsymbol{x}, \boldsymbol{y}; \theta) - \sum_{\boldsymbol{y} \in Y} Pr(\boldsymbol{y}|\boldsymbol{x}; \theta) \frac{\partial}{\partial \theta} f^{[K]}(\boldsymbol{x}, \boldsymbol{y}; \theta) \tag{8}$$

ここでtはステップ数, $\lambda^{(t)} \leqq 1$は学習率パラメータである.

なお訓練には条件に応じて MAP 推定などを用いることも可能である.

4.3 確率モデルであるメリット

PSI を確率モデルとしたことで, 次の利点がある.

- 出力が[0,1]と正規化され安定する.
- 異なる入力変数$\boldsymbol{x}$についてどちらがより$\boldsymbol{y}$とのスコアが高いか, 他の$\boldsymbol{y}$を考慮した上で比較することができる.
- パラメータの更新式に確率が存在することで, 更新が現在のパラメータの分布に則ったものとなる. これは訓練後半のパラメータの微調整や外れ値による過学習の対策に有効であると考えられる. またこの特性は半教師あり学習のときにより強く働く.
- 半教師あり学習を行う際に, 求めた数値をそのまま利用できるため効率が良い.
- 最尤推定の他, MAP 推定などデータ数に応じた確率に基づく推定法を用いることができる. また他の確率モデルとの統合を図ることも可能である.

4.4 CRFs との関連

Conditional Random Fields(CRFs; 条件付き確率場)は, 入力変数$\boldsymbol{x}$について素性$\boldsymbol{y}$の成立する箇所数をベクトル$\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})$で表現したものと, 各素性の重要度を示す重み$\boldsymbol{w}$の内積を条件付き確率の形で表現したものである. これは(9)のように表される.

$$f(\boldsymbol{y}|\boldsymbol{x}) = \frac{\exp<\boldsymbol{w}, \Phi(\boldsymbol{x}, \boldsymbol{y})>}{\sum_{\boldsymbol{y} \in Y} \exp<\boldsymbol{w}, \Phi(\boldsymbol{x}, \boldsymbol{y})>} \tag{9}$$

CRFs は条件付き確率を用いて表現されている点, 内積(及び拡張内積)を用いるという点で pPSI と類似の表現であるといえる. CRFs はテキスト情報処理に特化したモデルであるが, pPSI はどのような情報に対しても用いることができる.

5. pPSI のクラス分類への利用

5.1 目的変数パラメータ

pPSI をクラス分類に適用するに先立ち, (10)のように目的変数パラメータ$\boldsymbol{t}^{(\boldsymbol{u})} \in R^U$を定める. ここで$t_i^{(u)}$は$\boldsymbol{t}^{(\boldsymbol{u})}$の$i$次元目の要素を示す.

$$t_i^{(u)} = \begin{cases} i = u & 1 \\ otherwise & 0 \end{cases} \ (i = 1 \dots U) \tag{10}$$

ただし, Uはクラス数, uはクラス番号である. このuは分類するクラスに対して一意に与えられるものとする. またすべての$\boldsymbol{t}^{(\boldsymbol{u})}$の集合を$T$とおく.

5.2 クラス分類への適用

あるクラスC_{j}において，そのクラスに属するi番目のデータベクトルを$\boldsymbol{d}_i^{(\mathbf{j})}$，対応する目的変数パラメータを$\boldsymbol{t}^{(j)}$とする．ここですべてのクラスについて$(\boldsymbol{d}_i^{(\boldsymbol{u})}, \boldsymbol{t}^{(\boldsymbol{u})})(u = 1 \dots U)$のペアを考える．この二変数のペアについてデータ$\boldsymbol{d}_i^{(\boldsymbol{u})}$が既知のときに正しい組み合わせの目的変数パラメータ$\boldsymbol{t}^{(\boldsymbol{u})}$の事後確率が最も高くなることを目標として pPSI を用いてクラス分類を行う．このとき，あるデータ$\boldsymbol{d}$のクラスC_uに属する確率$p(\boldsymbol{d}, C_u)$は，(3)において$\boldsymbol{x} \to \boldsymbol{d}$, $\boldsymbol{y} \to \boldsymbol{t}^{(\mathbf{u})}$とすることで(11)と表される．

$$p(\boldsymbol{d}, C_u) = Pr^{[3]}\left(\boldsymbol{t}^{(\mathbf{u})} \middle| \boldsymbol{d}\right) = \frac{\exp\left(f^{[3]}(\mathbf{d}, \boldsymbol{t}^{(\mathbf{u})}; \theta)\right)}{\sum_{\mathbf{t} \in \mathrm{T}} \exp\left(f^{[3]}(\boldsymbol{d}, \boldsymbol{t}; \theta)\right)}$$

$$f^{[3]}\left(\mathbf{d}, \boldsymbol{t}^{(\mathbf{u})}\right) = \sum_{i=1}^{N} (\boldsymbol{Ax})_i (\boldsymbol{Bx})_i \boldsymbol{C}_{\mathrm{u}i} + \sum_{i=1}^{N} (\boldsymbol{Dx})_i \boldsymbol{E}_{\mathrm{u}i} \tag{11}$$

ただし, 行列に対する添字 ui はu行i列の要素を示すものとする.

5.3 クラスの推定

あるデータベクトルdについて, その属するクラスのクラス番号をuとすると, その推定値$\hat{u}$は(12)のようになる.

$$\hat{u} = argmax_u \Pr^{[3]}(\boldsymbol{t}^{(\boldsymbol{u})} | \boldsymbol{d}) = argmax_u f^{[3]}(\mathbf{d}, \boldsymbol{t}^{(\mathbf{u})}; \theta) \tag{12}$$

また, 推定されたクラスは$C_{\hat{u}}$である.

5.4 分類器としての pPSI のメリット

pPSI を分類器として利用することは, 他の分類器に対して次のような利点をもつ.

- 入力変数の分布を仮定しない. したがってデータ及び目的変数パラメータがそれぞれどのような分布を示すものであっても訓練によってそれぞれに適合することとなる. これにより入力変数の分布を考慮する必要がない.
- 予め入力変数全体の素性を知る必要がない. これは前述の分布に加え, 平均, 分散, 共分散などが未知であったとしても, モデルの形成に影響しないということである. したがって逐次更新を行うことが可能であり, 更新時にすべての学習データを同時にメモリ上に配置する必要がない.
- 訓練, 推定において目的変数パラメータを保存しておく必要はない. 実際には(11)で示されるようにパラメータ$\boldsymbol{C}$, $\boldsymbol{E}$から直接要素を参照しさえすればよい.
- (11)に示されるように, あるデータに対して式中のデータに関わる積$\boldsymbol{Ax}$, $\boldsymbol{Bx}$, $\boldsymbol{Dx}$を一度計算しておくと, クラスの推定についてはただ内積と和の計算のみを行えばよい. これは訓練時にも推定時にも寄与し, 計算コストが軽減されることとなる. 特にデータの次元数が大きいとき, また対象とするクラスが多いときに, この特徴が強く活かされる.
- 高次元のデータであっても, 近似されたパラメータを用いるため, 一更新あたりに必要なメモリ, 計算時間ともに次元数に比例して増加するのみでそれ以上を必要としない. したがって, 文書分類のようなタスクを次元圧縮などの手法を用いることなく行うことができる.

6. 実験

6.1 概要

pPSI の性能評価のため, 人工データ及びリポジトリのデータを用いてクラス分類を試行する. 実験 1 では pPSI, PSI, 及び m-SVM[†]の比較を行う. これによって pPSI の他の分類器との比較を示す. 実験 2 では$N = 10{,}100$の pPSI,PSI の比較を行う. これによって潜在的概念空間の次元数Nを考慮した確率モデル pPSI と元のモデル PSI との性能の比較を示す.

6.2 データセット

実験にあたって, 以下に示す 5 種のデータセットを用いた.

1. クラス数 5, 次元数 5, 訓練データ数 50000, テストデータ数 5000. 正規乱数による人工データ. ただし平均は指数乱数に基づき, 各要素はそれぞれ独立である.
2. クラス数 5, 次元数 100, 訓練データ数 50000, テストデータ数 5000. 単純マルコフ情報源をソースとする人工データ. ただし遷移行列は正規乱数による.
3. 同上. ただし遷移行列は異なる.
4. クラス数 5, 次元数 100, 訓練データ数 50000, テストデータ数 5000. 2 において, 遷移行列を$\lambda = 7.734$の指数乱数によるものに変更した人工データ. このデータは文書が単純マルコフ情報源に基づき, 指数乱数に基づく bigram データが遷移行列に相当すると仮定したものである.
5. クラス数 3, 次元数 4, 訓練データ数 120, テストデータ数 30. Iris[††](ICS Machine Learning Repository)

6.3 実験 1: 他の分類器との比較

実験 1 では, 他の分類器との比較を目的として, pPSI, PSI, m-SVM で精度を求めた. なお, pPSI, PSI は$N = 100$で(11)で示したモデルを使用し, 初期値依存性を考慮して 4-cross validation での最良値について, 100 回の試行の平均値/最大値を求めた. m-SVM は線形カーネルを用いて regulation rate $Rr = \{1{,}10{,}100{,}1000\}$の場合について求めた. その結果を表 1 に示す.

表 1　分類器による精度の比較(%)

data set	pPSI	PSI	m-SVM			
			Rr = 1	10	100	1000
1	91.99/92.60	92.65/92.94	66.24	20.04	61.98	66.06
2	99.78/99.94	98.87/99.58	99.52	100.0	100.0	100.0
3	99.59/99.78	99.24/99.40	99.84	99.24	99.98	99.98
4	99.76/99.92	98.82/99.76	98.44	100.0	100.0	100.0
5	99.20/100.0	91.43/100.0*	66.67	66.67	66.67	100.0

* データセット 5 の PSI の結果において, 再現率を計算できない場合が 5 ケースあったことを特記する. これはテストデータセッ

† m-SVM / SVM-Light Support Vector Machine http://svmlight.joachims.org/svm_multiclass.html

†† Iris Data Set / Frank, A. & Asuncion, A. (2010). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

トのクラスが推定されたときに一度も推定されたクラスとして該当しなかったクラスが存在していたことを示す.

6.4 実験 2:pPSI と PSI の比較

実験 2 では, pPSI と PSI の比較, 潜在的概念空間の次元数Nを目的として, $N = 10{,}100$と条件を変えて pPSI, PSI で精度を求めた. なお, pPSI, PSI は実験 1 と同様, (11)で示したモデルを使用し, 初期値依存性を考慮して 4-cross validation での最良値について, 100 回の試行の平均値/最大値を求めた. その結果を表 2 に示す.

表 2　pPSI と PSI の精度の比較(%)

data	pPSI		PSI	
set	N=10	100	10	100
1	91.50/93.28	91.99/92.60	91.35/92.58	92.65/92.94
2	99.88/99.96	99.78/99.94	99.74/99.96	98.87/99.58
3	99.56/99.76	99.59/99.78	99.16/99.62	98.24/99.40
4	99.85/99.96	99.76/99.92	99.73/99.92	98.82/99.76
5	98.33/100.0	99.20/100.0	93.03/100.0*	91.43/100.0*

* データセット 5 の PSI の結果において, 再現率を計算できない場合が N=10 のとき 1 ケース, N=100 のとき 5 ケースあったことを特記する. これはテストデータセットのクラスが推定されたときに一度も推定されたクラスとして該当しなかったクラスが存在していたことを示す.

7. 考察

7.1 実験 1:他の分類器との比較

実験 1 の結果を総合的に判断すると, m-SVM の方が pPSI, PSI より精度がよい(ただしRrに対する依存はある). しかし数%の差であり, PSI, pPSI は m-SVM に迫る精度で推定できることが分かる.

中でも注目すべきはデータセット 1 の結果であり, pPSI, PSI の入力変数を仮定しないという特徴から, 正規分布によって構成されたデータでも m-SVM に比べ高精度で分類することができている. m-SVM がガウシアンカーネルによって構成された場合, 線形カーネルよりも高精度で推定することが可能であると考えられるが, ガウシアンカーネルによる計算は線形カーネルに比べ大幅に時間を要し, またそのようにデータに合わせたカーネルを選択しなければならないというデメリットがある. pPSI, PSI の場合はそういった選択を行うことなく高精度な推定を実現できている.

なお, 参考としてデータセット 1 をデータの生成元に準じる GMM(Gaussian Mixture Model)によって推定した場合, 96.48%となった. この精度を上限としてみると, pPSI, PSI でもまだ改善の余地があると考えられる.

7.2 実験 2:pPSI と PSI の比較

実験 2 の結果を見ると, PSI と pPSI の差はあまりなく, 確率モデルとしたことによって精度に対して悪い影響を与えていないということが分かる.

またデータセット 5 では, pPSI, PSI とも精度の最大値は 100.0%であるものの, 平均値は pPSI の方が遥かに高い. これは*で述べたように再現率の計算できないような結果が含まれていたからであり, この点から PSI に比べ pPSI の方が安定的にモデルを形成できているといえる.

一方, 潜在的概念空間の次元数であるNについては, 実験の結果から必ずしも大きければよいというものではないと考えられる. これは入力データの次元数に依存せず, 概念空間上で圧縮された場合, 拡張された場合に関わらない. この値はデータに応じて適切な数値に調整することが必要であると考えられるが, どの程度左右されるのかについては検討が必要である.

8. 結論

pPSI, PSI では, SVM とは異なり入力変数の分布を考慮することなくモデルを構築できることを示した. 精度としては m-SVM に及ばないものの, いずれの分布においても高精度なクラスの推定を実現できる点において強みを発揮する. また本稿では具体的には示していないが, 訓練の特性上半教師あり学習が可能であり, 個人適応や推定の修正を行うことができる.

pPSI では, PSI とは異なり更新時に現在のモデルを考慮することができるため, 半教師あり学習などで元のモデルをあまり損なうことなく学習することができる. また MAP 推定などによって潜在的特徴空間への写像に方向性を持たせることが可能であり, 恣意的な概念空間の生成も可能であると考えられる.

9. 補足

本稿で示したパラメータの推定手法では, PSI に比べ pPSI は数倍程度の時間を要する. 期待できるメリットに対してこの計算時間の増加は見合わないため, より高速に訓練を行えることが求められる. 具体的には, ある段階まで PSI の更新式を用い, それ以降 pPSI の更新式に基づいてパラメータを推定するなどといった方法が考えられる. 例えば, パラメータの推定時の精度が目標とするクローズドの精度の 80%となるまで PSI の更新式を用いた場合, pPSI の更新式のみを用いた場合に比べ, 23%の時間短縮となった.

参考文献

[Bai 2009] B. Bai, J. Westion, D. Grangier, R. Collobert, K. Sadamasa, Y. Qi, C. Cortes, and M. Mohri.: Polynomial Semantic Indexing, Neural Information Processing Systems (NIPS) Conferences 2009.

[Deerwester 1990] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman.: Indexing by latent semantic analysis, JASIS, 41 (6) pp. 129-136, 1990.

[Hofmann 1999] T. Hafmann.: Probabilistic latent semantic indexing, SIGIR, pp. 50-57, 1999.

[Herbrich 2000] R. Herbrich, T. Graepel, and K. Obermayer.: Advances in Large Margin Classifiers, chapter Large margin rank boundaries for ordinal regression, MIT Press, Cambridge, MA, 2000.