社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2008－CVIM－165 (35)
2008／11／28

多クラスSupport Vector Machineを用いた一般物体認識での複数候補提示下における分類性能の傾向

栗田哲平† 近山隆†

現在，一般物体認識問題に対する手法への評価は，主にテストデータに対応するクラスを一意に推定する際の精度で性能を決定している．しかし一般には複数候補を提示するようなシステムを必要とする場合も多く，その場合の性能は正解を上位候補と出来るか否かの問題になる．そこで本稿では，一般物体認識のデータセットに対し多クラスSupport Vector Machine(SVM)を適用した場合の複数候補提示下での性能の傾向について報告する．多クラスSVMは複数の1対1あるいは1対他の2値分類SVMを組み合わせて構成するが，その際に生成された各分類器でテストデータ及び学習データのみを用いてソフトマージンのペナルティパラメータ，カーネル及びそのハイパーパラメータを調整する．そして1対1・1対他各々の拡張，また誤り訂正出力符号での複数候補提示出力について一般化を行い，それらを用いた場合の，候補提示数に対する性能について比較し，候補提示数によって異なる精度傾向を示すことについて述べる．

Clasification precision of several candidate using multiclass support vector machine in generic object recognition

Teppei KURITA† and Takashi CHIKAYAMA†

Generic object recognition is one of the most important research topics in Image Recognition. For these several years, research on generic object recognition has progressed greatly. But sometimes, we need a system which presents several candidate. In this paper, we present a unifying framework for studying the solution of multiclass classification probrems with several candidate. And we apply One-versus-One, One-versus-All and Coding Outputs of SVMs to recognize database images.

1. はじめに

近年，画像情報はインターネット上のコンテンツなどを通してその利用が急激に増加している．その画像情報の検索技術についても，web上でのテキストと関連付けた検索手法のみならず，画像情報を頼りにした検索手法も多くなり，画像認識の研究が活躍している．そのように画像認識技術の需要が高まっている中で，画像認識でも一般の物体をその対象名で認識する事をGeneric object recognition（一般物体認識）と言い，数十年の間，研究がなされている．この研究が長期間続けられ，今再度盛んになっている理由として，一般の物体を計算機に認識させる事は非常に困難であるが，計算機の進化と共に大規模なデータを扱えるようになり様々な手法を試す事が可能になってきたことが挙げられる．また，画像は人々の生活に深く密接しているということも大きな原動力となっていると考えられる．

計算機による認識とは，結局は様々なセンサから入力される信号を抽象的な概念に写像をする処理になる．パターンと概念の間には，大きなセマンティックギャップが存在し，計算機による処理は困難を極める．これが一般物体認識を困難にさせている大きな原因である．また，実用性を考慮すると，たとえ計算機がパターンを概念に写像できたとしても，その認識結果をどのようにして使用するのか，といういわゆるアクションの部分が存在しなければ人間の役には立たなく，そのような有効なアクションが取れる分野は限られてくる．画像認識の各段階において，どのような手法を採用するのかは分野や処理の目的により変化してくる．実時間性が必要になってくる場合には，処理の精度が多少悪くても高速なアルゴリズムを採用しなければならなく，逆に間違った例を提示する事が致命傷になる場合は，提示を無理にしないという判断も必要になってくる．人間の直感に近い認識結果が必要な時は，認識率の精度が多少悪くなっても，認知科学的な分野の手法を用いたアルゴリズムを選択する場合も出てくる．

また，システムによっては必ずしも一つの候補に絞る必要もない場合が考えられる．現在の一般物体認識問題に対する手法への評価は，主にテストデータに対応するクラスを一意に推定する際の精度によって性能を比べ，手法の優劣を決定している事が多い．しかし一般には複数候補を提示するようなシステムを必要とする場合も多く，その場合の性能は正解を上位候補と出来るか否かの問題になってくる．

今回，我々はその複数候補提示下での性能について着目する．そこで複数のクラスを分類を行うSupport Vector Machine（以下SVM）における順位出力手法について記し，一般物体認識のデータセットに対しBag-of-Keypointに判別モデルとして多クラスSVMを適用した場合の複数候補提示下での性能の傾向について報告する．

† 東京大学大学院工学系研究科
Graduate School of Engineering, The University of Tokyo

2. 関連研究

一般物体認識の研究として，近年特に成果を上げているものを挙げる．Varmaらは，特徴としての記述子が複数種類あった場合，それを分類器に組み込む時の各記述子の識別力と汎化性のトレードオフを学習し，それを使用して特徴を学習して最終的な一般物体認識を行っている．そして様々な物体認識のデータベースセットにおいて近年で最高（条件はあるがCaltech101で80%程度）の精度を出している[1]．Boschらは，あらかじめ学習画像群からその物体がある推定される関心領域(Region of interest)を自動的に得て，その中でSpatial pyramid表現[2]を用いてSVMおよびrandom forestsで判別モデルを構成し分類を行い，高い精度の認識に成功している[3]．

だがどれも複数候補提示下での精度については着目していなく，多くの種類の特徴量を組み合わせて使用している．本報は複数候補提示下での性能の傾向を調査する事を目的としている．シンプルな特徴量でどの程度までデータセットの2値問題を分類でき，多値的な分類を成功させられるかを示す．そこで，根本的に分離できない2値問題を探索・発見する．また，得られたパラメータを知識として使用した場合にどこまで有効であるのかを示し，既存研究と比較をする．2値分類問題についてはSupport Vector Machineを用いる．次章からSVMについて，その多クラスへの拡張と順位付け問題まで述べる．

3. 2クラス分類問題におけるSVM

本章では画像認識を行うために用いる分類器としての2値クラスSVMについて述べる．SVMは最適分離超平面の考え方を元にして考案され，近年になって非線形への問題への拡張としてカーネル学習法などと組み合わされてきた．入力ベクトルを非線形に写像した高次元の特徴空間上で，1つの線形識別関数を構成する．そもそもSVMでのクラス分類の目的は，高次元特徴空間において，2値問題について，「正しく」分類するような，つまり汎化限界を最適にするような超平面を，計算量的に「効率良く」（10万事例程のオーダーの問題も扱えるように）学習するような方法を提供する事である．高次元特徴空間において写像された分離超平面は，元の入力空間では局面になって，最終的に非線形な識別関数を構成する．

以上で述べたSVMは数多くある現在のパターン認識手法の中でも，最も性能が優秀な学習モデルの一つとして知られている．だが，SVMは2クラス識別問題において基本的には定式化されているので，画像認識など，多クラスの問題を扱うような識別器を構成するには更に工夫が必要となる．具体的な手法については後に述べるが，多数の2クラスSVMを組み合わせる事で，多クラスの識別を可能としている．そのアプローチも多数あるが，それらを統一化した手法も提案されている．

3.1 線形SVM

本節では，線形のSVMについて解説する．

SVMの最も単純なモデルとして，最大マージンクラス分類器がある．これは，線形分離可能なデータにのみ適用化であり，実世界のデータに用いる事は現実的ではない．しかし，構造が非常にシンプルで理解もしやすいアルゴリズムであり，これを基本として現在の実用的なSVMは構成されている．

ラベル付けがされたN個の特徴ベクトルとしてのデータがあり，それを$\{\mathbf{x}_i, y_i\}_{i=1}^{N}$とする．ここで，$\mathbf{x}_i$は$i$番目のデータの特徴ベクトル，$y_i$はそのデータについているラベルである．

学習データの集合Rが線形分離可能であるならば，関数のマージンが1である以下の式を満たす識別関数のパラメータ$\{\mathbf{w}, b\}$が存在する事になる．

$$\mathbf{w}\cdot\mathbf{x}_i + b \geq 1 \quad (if\ y_i = +1) \tag{1}$$

$$\mathbf{w}\cdot\mathbf{x}_i + b \leq 1 \quad (if\ y_i = -1) \tag{2}$$

上記の境界を関数とした2枚の超平面で学習データは完全に分離されており，この間にはデータが存在しないという事を示している．

そのような超平面の場合，識別する平面とこれらの超平面との距離をマージンと呼ぶ．この時のマージンが大きい方が，汎化性能が高い事が知られている．実際に上記の平面間のマージン距離γを求めると以下のようになる．

$$\gamma = \frac{|\mathbf{w}^T\mathbf{x}' + b - 1|}{\|\mathbf{w}\|} + \frac{|\mathbf{w}^T\mathbf{x}' + b + 1|}{\|\mathbf{w}\|} \tag{3}$$

$$= \frac{|-1|}{\|\mathbf{w}\|} + \frac{|1|}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \tag{4}$$

SVMの最大の特徴はこのマージンγが最大となる分離平面を求める「マージン最大化」にあり，学習データの近くをぎりぎり通るのではなく，出来るだけ余裕があるように分離するような識別平面が求められる．これによって，SVMは高い汎化能力（未学習データに対しても高い識別性能）を発揮できる．

このマージンを最大化するためには$\|\mathbf{w}\|$を最大化すればいい事がわかる．しかし一般には，訓練サンプル集合を誤りなく分けるパラメータは一意には決まらない．

ここから，マージンγを最大化するようなパラメータ$\mathbf{w}$とbを求める事は，結局以下の制約が付いた最適化問題を解く事と等価になる．

$$\text{minimise}_{\mathbf{w},b}\ \ L(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 \tag{5}$$

$$\text{subject to}\ \ y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1\ \ i = 1, ...l \tag{6}$$

超平面$(\mathbf{w}, b)$はこの最適化問題の最適解となる．

この最適化問題は二次計画問題であり，対応する双対問題へと変換する方法などによって解かれている．この主問題のラグランジアンを考えると，目的関数は以下のようになる．

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l}\alpha[y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1] \tag{7}$$

$\alpha_i \geq 0$はラグランジュ乗数である．ここで対応する双対問題を得るために$\mathbf{w}$とbに関し偏微分を行い定常性を仮定する．その時，停留店では以下の関係が成り立つ．

$$\frac{\partial L(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{l} y_i\alpha_i\mathbf{x}_i = \mathbf{0} \tag{8}$$

$$\frac{\partial L(\mathbf{w}, b, \alpha)}{\partial b} = \sum_{i=1}^{l} y_i\alpha_i = 0 \tag{9}$$

この関係を上の主問題としての目的関数，式(7)に代入すると，以下のような関係がわかる．

$$L(\mathbf{w},b,\alpha)=\frac{1}{2}\sum_{i,j=1}^{l}y_iy_j\alpha_i\alpha_j\mathbf{x}_i^T\mathbf{x}_j$$
$$-\sum_{i=1}^{l}y_iy_j\alpha_i\alpha_j\mathbf{x}_i^T\mathbf{x}_j+\sum_{i=1}^{l}\alpha_i \quad (10)$$
$$=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{l}y_iy_j\alpha_i\alpha_j\mathbf{x}_i^T\mathbf{x}_j \quad (11)$$

ここから以下のような命題が得られる．

$$\text{maximise W}(\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{l}y_iy_j\alpha_i\alpha_j\mathbf{x}_i^T\mathbf{x}_j \quad (12)$$
$$\text{subject to}\quad y_i(\mathbf{w}^{\mathbf{T}}\mathbf{x}_i+b)\geq 1\ \ i=1,...l \quad (13)$$

パラメータ α^* がこの二次最適化問題の最適解の時，重みベクトル $\mathbf{w}^*=\sum_{i=1}^{l}y_i\alpha_i^*\mathbf{x}_i$ は，マージンが $\gamma=\frac{2}{\|\mathbf{w}\|}$ である最大マージンをもつ超平面を実現する．

ここで b の値は双対問題には存在せず，解 b^* は制約から以下のように解かれる．

$$b^*=-\frac{\max\limits_{y_i=-1}(\mathbf{w}^T\mathbf{x}_i)+\min\limits_{y_i=1}(\mathbf{w}^*\mathbf{x}_i)}{2} \quad (14)$$

3.2 ソフトマージン最適化

前節で解説した最大マージンクラス分類器としての線形SVMは，SVMの概念として非常に重要であるが，前述したように実世界の問題に対して適用をするのは，線形分離の可能性から言って困難である．最大マージンクラス分類器は完全に無矛盾な仮説，つまり訓練誤差が無いような識別関数を常に生成する．これはマージンに依存した汎化誤差の限界を考慮しているためである．

マージンに依存するような設計にすると，全体での少数のノイズの影響が大きくなり，ノイズが散在する実世界のデータにその分類器をそのまま適用することは，信頼性の面から言って限りなく利用価値が低い．そして線形分離不可能の問題に対して，主問題の実行領域が空であり双対問題は限界の無い目的関数を持つことになり，最適化問題を主問題として解く事が出来ない．

そこで多少の誤識別について外れ値を設けることで許容するような尺度を用いて分類器を設計し直す方法がある．これはソフトマージンの手法と呼ばれている．

3.3 カーネルトリック

線形分類器を用いて，データの形が非線形であるような事象を学習するには．非線形な特徴集合を選択して，データを異なった表現に書き換える必要がある．

線形分類器の特性の一つとして，双対問題として表現できるという事が挙げられる．つまり仮説が学習データの線形結合で表現が可能であるということである．ここから，決定規則がテストデータと学習データの内積を用いて表現できることがわかる．ここで，もし特徴空間内の内積 $(\Phi^T(\mathbf{x}_i)\Phi(\mathbf{x})$ が元の入力データでの関数として直接計算が可能ならば，非線形学習マシンを構築する固定した非線形写像によりデータを特徴空間に変換し，特徴空間内での識別に線形マシンを使用するという段階を併合することが出来るようになる．このような直接計算を行うために用いる以下のようなカーネル関数を用いて非線形データを計算量を削減しつつように扱えるようにする手法はカーネルトリックと呼ばれている．

$$K(\mathbf{x},\mathbf{z})=\Phi(\mathbf{x})\cdot\Phi(\mathbf{z}) \quad (15)$$

このような性質を持ったカーネル関数を用いれば，非線形のデータをSVMに適用するための特徴ベクトルの写像および内積によって起こる膨大な演算が1回カーネルを計算するだけですんでしまう．

4. 複数のクラスを対象としたSVM

前章で扱ったSVMは基本的には2値の識別問題を対象として定式化されている．しかし実世界の問題では2クラスより多いようなクラスを識別するような分類問題に直面する場合も多い．そのため2値問題の複数クラス問題へ拡張する手法が必要になる．

多クラス分類とは，各データが予め用意されたクラスのどれかに分類をすることを言う．データ $\mathbf{x}$ が条件としてあったときのクラス分類の事後確率 $P(i|\mathbf{x})$ が最大となるようなクラスを選択するように通常設計する．事後確率を近似するために，正規混合分布・カーネルk-nearest neighborやカーネル密度推定などのパラメトリック・ノンパラメトリックな手法が用いられる．

しかし2値分類問題に限ると，SVMのようなマージン制御に従った高い汎化性能を持つような分類器を用いる事が出来る．前述したように，基本的に多クラスの識別問題を2値分類器で扱うためには，2クラスの判別モデルを組み合わせる事になる．これは複数クラス問題を2値問題のセットへと縮小させる考えを前提として提案されている．SVMの基本的な構成のまま，多値の判別関数を直接構成する手法（Multiclass-SVM）も提案されているが，今回は組み合わせで構成する手法を対象とする．これは，各クラス間の分類問題が同じ仮説の上に成り立っているとは限らないこと，個々の問題においての詳細なデータの解析が難しい，ことが主な理由である．

4.1 One-versus-All

全クラス数が k で，$\mathbf{x}$ を入力としたとき，k 個の各SVMの識別関数 $g_j(\mathbf{x})$ は，

$$g_j(\mathbf{x})=\mathbf{w}_j\mathbf{x}+b_j \quad (16)$$

と定義されるとする．ここでは単純化のため対象のクラス j とその他のクラスの間に分離が可能であるとすると仮定している．その時，もし入力 $\mathbf{x}$ がクラス j のラベルを保持していれば，式(17)が成立する事がわかる．

$$\mathbf{w}_j\mathbf{x}+b_j\geq 0 \quad (17)$$

各識別境界は上式の境界で与えられる．

以上の事を利用し，この識別関数において各データが正しくその保持クラスに分離される状況においては，マルチクラスへの適用も出来ている．この手法は従来のSVMをいわばそのまま用いている最も単純な手法であり，計算量も少なく実装も簡単である．しかし，クラス数が増加すれば確実に分離出来る状況はなくなり，上で述べた仮定が成り立たなくなる．すると，誤って他のクラスに属してしまう場合だけでなく，どこかのクラスに確実に属している前提があっても，そのデータがどこのクラスにも属していないという結果が出てしまうこともある．これは全体で分離性を仮定しているため，データ量が多くなる

とどうしても無理が出てきてしまうからであると考えられる．また，各クラスで別々に識別問題を解いているので，そこで出力された結果全体に対する最適解の保証は無い．

そこで，同じ1対他の分類器を構成する考え方であるが，部分的な分離可能性を仮定する手法について述べる．ある任意のクラス j とそれ以外のクラスの間に分離可能性を仮定したとき，上で述べた通り識別関数は式 (16) となる．もし入力 $\mathbf{x}$ がクラス j のラベルを保持していれば，他の全てのクラス $j'(j'=1...k, j'\neq j)$ の識別関数と比較して，式 (18) が成立する事がわかる．

$$\mathbf{w}_j\mathbf{x}+b_j \geq \mathbf{w}_{j'}\mathbf{x}+b_{j'} \tag{18}$$

このとき，クラス j と j' の間の識別境界は $(\mathbf{w}_j-\mathbf{w}_{j'})\mathbf{x}+(b_j-b_{j'})=0$ となる．以上のような識別境界を構成すれば，部分的な分離性を仮定している事になる．これは，各識別関数における入力データに対する出力マージンの値を比較し，クラスを決定している事となる．結果的に誤っていても，全てのデータはどこかのラベルに属することになる．k 個の識別関数を構成し，その各々に全体のデータ数 N を入力して学習する事になるので，計算量的には，クラスが多い問題にも対処できるようになっている．各2値 SVM の識別関数をデータ数 n で構築するオーダーは，大体どのような実装においても $O(n^3)\sim O(n^4)$ 程度であると知られている．

このような任意のクラスとその他全てのクラスで構成された識別関数を利用し，2値クラス SVM を多クラス問題対応させるような手法を One-versus-All(または One-versus-Rest) のアプローチと呼ぶ．

4.2 One-versus-One

前節で述べた One-versus-All の他に，Hastie らによって提案された，One-versus-One（または All-Pair）という2値問題の多クラスへの拡張手法がある[6]．これは任意のクラスのペアどうしでの学習データを用いて識別関数を構成し，それを多クラス分類問題に利用している．

任意の2つのクラス $j_1, j_2(j_1, j_2=1...k, j_1\neq j_2)$ の間に分離可能性があると仮定する．その時，全ての2クラスの組み合わせ $\frac{k(k-1)}{2}$ 個の識別関数が以下の式で構成される．

$$g_{j_1,j_2}(\mathbf{x})=\mathbf{w}_{j_1,j_2}\mathbf{x}+b_{j_1,j_2} \tag{19}$$

この時，もし入力 $\mathbf{x}$ がクラス j のラベルを保持していれば以下が成り立つ．

$$\mathbf{w}_{j_1,j_2}\mathbf{x}+b_{j_1,j_2}\geq 0 \tag{20}$$

各識別境界は上式の境界 $\mathbf{w}_{j_1,j_2}\mathbf{x}+b_{j_1,j_2}=0$ で与えられる．この手法では，ペアというそれぞれの仮説において分離性があるとして，学習器を構築，その結果を統合（多数決など）して扱う，といったものである．各2クラスの学習データ同士で部分的に識別関数を構築するため，非常にデータ量が小さくなり，完全分離可能性が高まる．また，2値分類器の元々の意味として持つ，2つの仮説を正しく分離するというものとも一致していると考えられる．$\frac{k(k-1)}{2}$ の識別関数を構成し，その各々にその学習器に用いるクラスのデータ $s(s=\frac{N}{k})$ を2つ入力して学習するので，計算量の観点でみると，多くのクラスがある場合には大きくなってしまうが，学習データの数が少ない場合には，各分類器を高速に作成できる．

しかし，最終的なそのクラスのラベルを結果として出力するために，その識別関数を構築する2ペア以外のクラスを持つデータもそこに当てはめるので，その他の識別関数においてどのような分離が行われるかの予測が困難である．また，各ペアで仮説を立て，そこで識別関数を構築するので，最終的な結果が決定不能な領域に属するデータというものがしばしば出てしまう．複数の学習器での出力を多数決で取るという処置は単純で実装も容易であるが，最適性の観点からは裏づけがない．

4.3 Error Correcting Output Codes

複数クラスの問題を扱うためのより一般的な手法が，Dietterich らによって与えられている[7]．それは，各クラスのラベルと分類器によっての出力が関連付けされた符号行列 $M\in\{-1,+1\}^{k\times l}$ を元に考えられている．これはラベルが y であるクラスが，s 番目の分類器によって出力される値 f_s を $M(y,s)$ にマッピングしている事を示している．ここで入力 $\mathbf{x}$ が与えられたとき，行列 M における y の仮説の各値が $f_1(\mathbf{x})...f_l(\mathbf{x})$ に最も近いような（例えば Hamming 距離などを取って）ラベル y を求める．この処理をデコードと呼ぶ．その y が入力 $\mathbf{x}$ のラベルとして予測される．

以上のような考え方を誤り訂正符号行列，Error Correcting Output Codes(ECOC) と呼ぶ．この手法の利点として，構築する識別平面，つまり分類器の数に制約がないことである．もちろん One-versus-All や One-versus-One の手法のように分類器の数を多くすれば頑強性は高まる．しかし同時に計算量も増える．そのような精度と計算量のトレードオフを図るための手法といって良い．

4.4 多クラス SVM の一般化

以上のようなアルゴリズムを一般化統合する[10]．行列の各分類器によって得られる値 f_s を拡張し，符号行列 $M\in\{-1,0,+1\}^{k\times l}$ を定める．ここで $M(y,s)$ が 0 の場合は，分類器によってそのデータがどのように分類されているかを考慮しなくて良いという事である．分類器 $s=1..l$ について，その中の分類アルゴリズムを使用して，全学習データデータ $\mathbf{x}_i$ によって，$M(y_i,s)$ にラベル付けを行う．

例えば，One-versus-All での多クラス SVM は，M は全対角要素が+1 であり，他の要素が 0 であるような $k\times k$ 行列となる．また，One-versus-One の手法では，M は各列が別個のペアクラス (j_1,j_2) の分類器である $k\times\frac{k(k-1)}{2}$ 行列となる．ペア (j_1,j_2) での分類器を $s(j_1,j_2)$ とすると，$M(j_1,s(j_1,j_2))=+1, M(j_2,s(j_1,j_2))=-1$ となり，また j_1,j_2 以外の全てのクラス j_{other} に対して，$M(j_{other},s(j_1,j_2))=0$ となる．

4.5 多クラス SVM の順位出力への拡張

前節まで，SVM の多クラス化について議論してきた．今回，我々は一般物体認識において，複数候補を提示した場合の精度を確認する．その為に，多クラス SVM での出力を最もそれらしいラベルのみならず，もっともらしさの基準を決定し，それを元に順位推定をしなければならない．先ほどの符号行列と各々の分類器によって生成された個々のデータに対する出力値が小さい程，そのデータのクラスであるというように順位付けを行う．

さて，One-versus-All の元の形を符号行列とデコードで記述するには，前述したように全対角要素が +1 であり，他の要素が 0 であるような $k\times k$ 行列である M を用い，個々のデータに

対する各分類器での出力値 ±1 の以下に定められる Humming 距離 d_H を求める．

$$d_H(M(r), f(\mathbf{x})) = \sum_{s=1}^{l} \frac{1 - \mathrm{sign}(M(r,s) f_s(\mathbf{x}))}{2} \quad (21)$$

通常のクラス分類では各データについて以下に示される y にクラスが決定される．

$$y = \arg\min_r d_H(M(r), f(\mathbf{x})) \quad (22)$$

また順位出力の形では，i 番目にそれらしいクラスは，$i-1$ 番目にて出力されたクラスを除いた集合を用いて上述した式で決定される．しかし前述したように One-versus-All は全体で分離性を仮定しているため，どこかのクラスに確実に属している前提があっても，そのデータがどこのクラスにも属していない結果が出てしまう．また，符号行列が対角成分にしか値を持たないので，分類器ラベルの出力が ±1 しか持たないと順位付けがほぼ不可能になる．

部分的な分離可能性を仮定した One-versus-All を符号行列とデコードで記述するには，先ほどと同じ符号行列と Humming 距離を用いた場合，個々のデータに対する各分類器の識別関数値の中で最も大きいものを+1 とし，その他を 0 とした出力を使用する．この操作を行うとどこかのクラスには必ず属することになるが，そのクラス以外の優位性が全く出なくなるので順位出力は不可能になる．そこで識別関数の値をそのまま，個々のデータに対する各分類器の出力として用いる．識別関数値を原形で用いると前述した符号行列との Humming 距離を取ることが出来なくなるので，Humming 距離の代わりに以下に示す損失関数に基づいた距離 d_L を用い，その値を元にクラスを決定する．関数は以下の式のような指数ベースの損失関数などが代表として挙げられる．

$$d_L(M(r), f(\mathbf{x})) = \sum_{s=1}^{l} \frac{1}{1 + e^{2M(r,s) f_s(\mathbf{x})}} \quad (23)$$

同じように，各データについて以下に示される y にクラスが決定される．

$$y = \arg\min_r d_L(M(r), f(\mathbf{x})) \quad (24)$$

このようなデコーディング手法を loss-based decoding と呼ぶ．

One-versus-One を符号行列で一般化するには，前述したように，M は各列が別個のペアクラス (j_1, j_2) の分類器である $k \times \frac{k(k-1)}{2}$ 行列となる．その符号行列と各データの出力の間で式 (21) で表される Humming 距離を取る．One-versus-One では，部分的な分離性を仮定しているために，このアプローチでも順位出力が可能である．また，One-versus-All で行ったように各データでの出力に識別関数値をそのまま用いて，符号行列との間で損失ベースの距離を取って順位出力に変換する事も出来る．

この各処理手法での実験結果について次章でそれぞれ述べる．

5. Caltech データセットの解析

今回，我々は Caltech101 のデータベースセットに対して，多クラス SVM を適用し，候補数にしたがった精度がどの程度になるかを検証する．Caltech101 は 2004 年にカルフォルニア工科大学の Fei-Fei らによって収集され，作られた画像データベースである．101 の物体カテゴリと追加背景からなる 102 のクラスで構成され，31 から 800 の画像をカテゴリ毎に含んでいる．今日利用できるデータベースの中で最も多種多様なオブジェクトを含んでいるとされている．殆どの画像は 300 × 300 程度の中解像度のものである．色，姿勢そして照明が画像ごとに変化しているので，分類は困難とされ，研究者達のなかでは非常にチャレンジングなデータベースとして，現在，評価画像データの業界標準となっている．

このデータセットにおける認識率とは訓練画像を 30 枚とし，テスト画像を残りとした結果の平均値となっている．2006 年の時点で最も高い認識率を出しているのは Zhang らの手法[5]で，66.23%となっていた．また 2007 年には，関連研究で触れたようなカーネルの重みを推定して組み合わせる手法や，学習領域を絞る手法などを用いて，80%以上の精度を出す研究も発表されてきている[1)3)]．

だが，序章でも書いたように，複数候補を提示した場合での精度が必要となる場合も多い．更に今回問題としているのは，特徴量や学習領域を上手く選択して結果を良くするというよりは，シンプルな特徴でどの程度まで分類できるかを確かめ，その特徴量ではどのようにしても分離が出来ないクラス間の問題を発見し，根本的な問題を提示する事にある．従って今回の実験では用いる特徴量やその組み合わせについてはあまり議論をしない．

5.1 局所特徴量

認識を行うための局所特徴量として，今回は SIFT を用いる．SIFT(Scale-Invariant Feature Transform) とは，特徴点の検出・局所特徴量の記述を行うアルゴリズムである．この局所特徴量は，画像の回転・スケール変化・照明変化に対してロバストに働く事が知られており，近年の物体認識において，最も用いられている特徴記述となっている．本報ではその特徴量の導出手法などの詳細については省略する．文献を参考して欲しい[4]．今回，SIFT で抽出された局所特徴量は 128 次元で表現されるものを用いる．SIFT は前述の通り，画像のスケール・回転変化に対してロバストに働くため，特定物体の同定には非常に有効な特徴量と言っていい．しかし，一般物体認識問題などに関するクラス分類に対して，SIFT 特徴量をそのまま利用する事は困難である．そこで Bag-of-Keypoints のアプローチを用いる．

5.2 Bag-of-Keypoints Approach

Bag-of-Keypoints は局所領域の特徴量（keypoints）のみで認識を行う手法であり，Constellation model（局所領域の相対的位置の情報も確率モデル化する手法）と同等の認識結果を出している．統計的な言語処理での Bag-of-word model と類似しており，対象画像を，位置を無視して局所特徴の集合として考える (図 1)．具体的な処理としては，クラスタリング手法（今回は k-means を利用した）を用い，局所特徴で構成される特徴ベクトルをベクトル量子化させ，局所特徴を word として扱えるようにする． 最終的な分類を行う際には画像中におけるその word の数をカウントして，ヒストグラム化する．そして各クラスの学習画像を用いて判別モデルを構成して，未

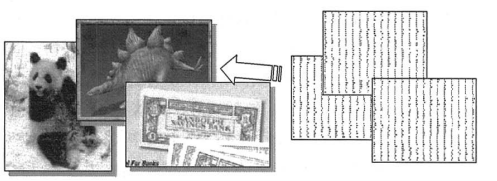

図 1 Bag-of-Keypoints

図 2 SIFT の抽出と，属するクラスタの Feature Weight

知のテストデータに対して分類を行う．ここでの分類には，前章で説明した多クラスの SVM を使用する．

ここで各ヒストグラムの各ビンの値が最終的に使用される特徴量となるが，その値を 2 通りの形にする．1 つは単純に各 word クラスタ中心に近いものをカウントして，その合計で正規化した値を利用する．もう片方は，Feature Weight という考え方を導入する．これは，言語処理で導入されている特徴量であるが，各 word がどれだけ重要なものであるかを考慮し，言わば重み付けを行っている．自然言語処理における tf-idf の考え方と同じであり，背景などで頻繁に登場するような特徴クラスタの影響を少なくするような手法であり，画像においても意味が出てくると考えられる．画像の総数を N とし，特徴クラスタ x が出現する画像の個数を n_x とした時，特徴クラスタ x の逆画像頻度 idf，は式 (25) となる．

$$\mathrm{idf}_x = \log \frac{N}{n_x} \tag{25}$$

また，特徴クラスタ x が画像 d 中に出現する数を oc_{xd} とし，出現特徴の集合を W としたとき，特徴クラスタ x の画像 d 内での出現頻度 tf は式 (26) となる．

$$\mathrm{tf}_{xd} = \frac{\mathrm{ptf}_{xd}}{\sqrt{\sum_{i \in W} \mathrm{ptf}_{id}^2}} \tag{26}$$

$$\left(\mathrm{ptf}_{xd} = 0.5 + 0.5 \times \frac{oc_{xd}}{\max_{i \in W} oc_{id}} \right) \tag{27}$$

上の tf 値，idf 値を用いて，重要度 Feature Weight は以下のようにして求まる．

$$\text{Feature Weight}_{xd} = \mathrm{tf}_{xd} \times \mathrm{idf}_x \tag{28}$$

図 2 では Feature Weight 値で重要であると判断されたクラスタに量子化されるような SIFT の局所特徴を濃い円で，重要でないと判断されたクラスタに入るような SIFT を薄い円で描写している．背景などでも多く出てくるような場所では SIFT が抽出されていてもその Feature Weight は小さくなる．

以上の Feature Weight の値と通常の正規化の 2 通りの値をヒストグラムの特徴量として用いる．

5.3 SVM に用いるカーネル関数

今回サポートベクターマシーンはソフトマージンのパラメータが組み込まれているものを使用する．更に，カーネルは以下に示される多項式カーネル K_{poly}，シグモイドカーネル K_{sig}，RBF カーネル K_{RBF} を使用する．

$$K_{poly}(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z} + c)^d \tag{29}$$

$$K_{sig}(\mathbf{x}, \mathbf{z}) = \tanh(\mathbf{x}^T \mathbf{z} + c) \tag{30}$$

$$K_{RBF}(\mathbf{x}, \mathbf{z}) = \exp(-\frac{1}{\sigma}^2 |\mathbf{z} - \mathbf{x}|^2) \tag{31}$$

5.4 各識別関数でのパラメータの調整

前述した Caltech のデータベースに対して，Bag-of-Keypoints で得られたヒストグラムを特徴として，判別モデルを構成した SVM で順位出力を行った結果について考察を行う．また，学習データ及びテストデータを用いて，SVM におけるソフトマージン制御パラメータ，カーネル及びそのハイパーパラメータを調整をして，今回用いる特徴を用いた場合にどの程度分類できるかの可能性を知る．

多クラス分類では，複数の識別平面を構成し，それを用いて最終的なクラスの決定・順位付けを行っている．ここで各分類器が同じ仮説の上に成り立っているとは限らないので，各々において最適なカーネルおよびパラメータを調整する必要がある．調整用のデータを用いてソフトマージンのパラメータ，カーネルのハイパーパラメータを調整する場合，グリッド的な探索が現在最も妥当だと言われている．つまり指数関数的にパラメータを変化させて，良い結果のパラメータをその中で採用し，徐々に最適なパラメータに近づけるようなヒューリスティックな手法である．今回は各 2 値分類器上でそのグリッド的な探索を用いて最も良い精度が出るようにパラメータの調整を行った．また，カーネルの選択もその探索に含め，最適なカーネルおよびそのパラメータの選択を行った．

5.5 実験結果

One-versus-All，One-versus-One 各々の多クラス拡張において，テストデータをチューニングデータとしてパラメータを調整する．以下の結果は全て学習データを３０とし，テストデータは５０を限度とし，１０回ランダムに各データを各クラスで収集した際の結果の平均である．学習時間は CPU が Xeon 3.06GHz dual の計算機を用いて，クラスタ数 200・Feature Weight で RBF カーネルを使用した場合，One-versus-One で 621 秒，One-versus-All で 246 秒であった．One-versus-One は識別関数の分類結果から，One-versus-All は識別関数値から，それぞれの符号行列との前述した別種の距離を取り，精度を求める．学習データのみで各識別関数を構成して，最終的なこの特徴量でカーネル及びパラメータを調整して分類できる限界についての複数候補提示化における結果を，元のチューニングしていないもの（正確には RBF カーネルを用い，全体で最適な結果が出る一意のパラメータに設定しているもの）と比較したものが図 3 である（特徴量は Feature Weight を用いており，クラスタ数は 200 である）．パラメータ等を調整していない場合には，少数の候補提示化における精度は One-

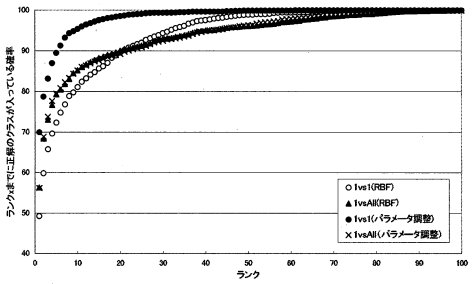

図 3 各分類器でパラメータを調整した場合との比較（One-versus-All と One-versus-One）

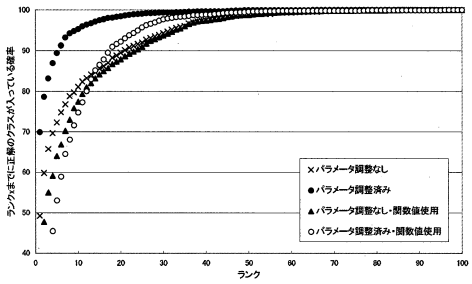

図 4 識別関数値を用いた場合と識別結果を用いた場合の比較

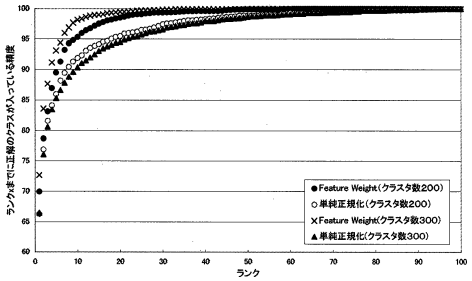

図 5 特徴量での精度比較（単純正規化と Feature Weight）

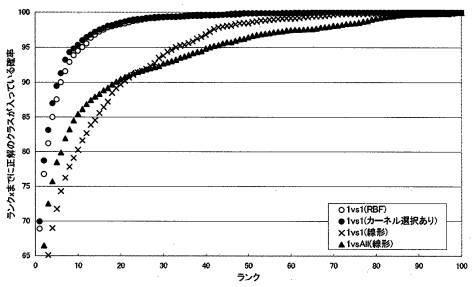

図 6 カーネル選択の有無による精度の違い

versus-All の方が One-versus-One よりもかなり高い値が出る．しかし，候補提示数を増やしていくと提示数２０辺りから One-versus-One の方が高い精度が出てくるようになる．また，パラメータを調整した場合では，One-versu-All は殆ど変化が見られないのに対して，One-versus-One では，著しく精度が上昇している．これは各分類器でパラメータ類を調整する事が，One-versus-All の識別関数値を順位推定に利用する事の正当性を失わせているのに対して，One-versus-One では出力は分類結果となり，その Humming 距離で候補数に従った精度を求めているため，各２値分類器の分類精度の上昇が直接結果に影響するためであると考えられる． 図４は，その One-versus-One に関して，前述のアプローチ（識別結果 (±1) をそのまま用いる）と識別関数値と loss decoding を用いた手法での比較を，パラメータ調整有り・無しのそれぞれについて行った結果である．識別関数値を用いると，今回の損失関数では精度が下がってしまうが，やはりパラメータ調整を行うとその評価の正当性が無くなってしまうため，精度が下がることがわかる．

特徴量間（単純な正規化と Feature Weight）の精度比較を行う．One-versus-One にて各分類器でカーネル及びパラメータを調整した状態での単純な正規化と Feature Weight の精度を比べたものが図 5 である（クラスタ数は 200 と 300）．単純な正規化よりも Feature Weight の方が候補提示数によらず精度が高くなっている．これはクラスタ数を増減させても同じような結果が出た（クラスタ数 100 以下では実験を行っていない）．

One-versus-One において RBF カーネルのみを用いてパラメータのみを調整したものと，カーネル関数も選択して調整したもの，および線形 SVM を用いたものでの精度の違いを示したものが図 6 である（Feature Weight，クラスタ数 200）．カーネルを選択していった方が多少精度の高い結果になっていることがわかる．仮説に適した分離平面を構成したのか，偶然に分離できたのかは不明であるが，３つのカーネルのうち最も精度の高い結果の出た RBF カーネル単体を用いるよりも可能性は上がることがわかる．精度が同程度の時は RBF、シグモイド、多項式の順でカーネルを選択している．その識別平面全体の内のカーネル選択の割合（200・300 のクラスタ数における、多項式 (Poly)，シグモイド (Sig)，RBF，多項式かシグモイド (PorS)，シグモイドか RBF(SorR)，RBF か多項式 (RorP)，どのカーネルを選択しても最高精度が同じ (どれでも)，の選択割合）を示したものが表 1 である．

表 1 カーネルの選択割合 [%]

数	Poly	Sig	RBF	PorS	SorR	RorP	どれでも
200	2.65	4.16	17.11	0.98	4.16	2.02	68.93
300	2.35	5.94	17.56	1.21	4.46	2.39	66.08

一般に Caltech101 に対する Bag-of-Keypoints のクラスタ数は 200～1000 程度が望ましい結果が得られるという事が既存研究によって分かっており，今回の事前実験としてより単純なデータセットで Feature Weight でのクラスタ数に従った精度傾向を調べたところ，200～400 程度が良いという結果が出た．今回は，その妥当なクラスタ数での候補提示数による傾向の違いを観察した．その結果が図 7 である．クラスタ数 200 での実験結果よりも 300・400 での結果の方が提示数全般に渡ってよい精度を出しているが，提示数が 30 を越えるとほ

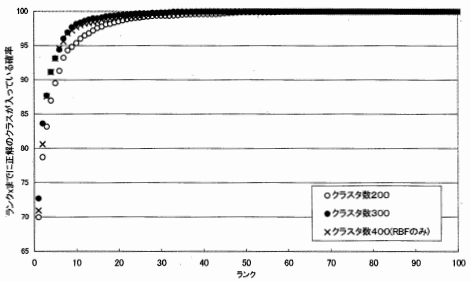

図 7 クラスタ数の違いによる比較

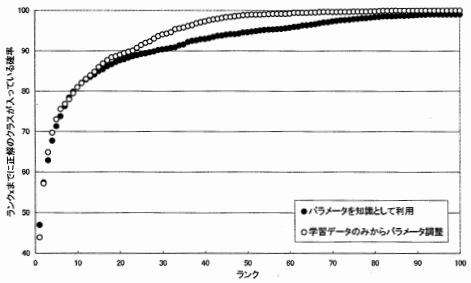

図 8 パラメータが知識としてある場合・ない場合の比較

表 2 分類で多くの誤りを出した2値問題例

2値問題	正例正解	正例不正解	負例正解	負例不正解
scorpion-pigeon	0	50	15	0
dolphin-flamingo	11	24	27	10
octopus-platypus	0	5	4	0
ewer-flamingohead	2	48	15	0
ewer-minaret	50	0	0	46
Faces-camera	18	32	19	1

ぼ100%に近くなり，あまり変わらなくなる．

識別においてどのような2値分類問題が難しく，上手く分離できなかったか，またどれだけ調整を行った2値SVMで分離が出来ている分類器が存在しているかを調査した．テストデータに対する分類において多くの誤りを出している2値問題例を示したものが表2である．これはどれだけ識別関数のカーネルやパラメータの値を調整しても上手く分類が出来なかった問題例である．多くが片方のクラスに因ってしまい，分離が出来なくなっているものが多い．これは今回用いた特徴量では根本的に判別が無理であるという事を示している．

一般的に未知のデータに対してパラメータ調整を行うには，学習データ・テストデータの他にチューニング用のデータがあるのが望ましい．しかし，この一般物体認識の性能の標準的な測定では学習データを30程度用いて残りをテストデータとしているので，合計データ数が少ないクラスもある事を考慮すると，学習データのみで行う場合はcross-validationを用いるのが妥当だと言える．10分割交差で学習データのみでパラメータを調整した場合と，あらかじめ実験したデータのパラメータ（データをランダムにシャッフルを行い5通り）を知識として得ている場合での精度の平均との比較をした図が8である．各分類器のパラメータの調整は，異なった学習データおよびテストデータから予め得られている知識を用いて行ってもあまり意味を成さない事がわかる．学習データのみで交差確認で調整を行っても，提示数50まで見るとほぼ精度100%となり，フィルタなどの活用においては十分実用的な精度であると言える．

6. おわりに

本報ではCaltechのデータセットを対象として，SVMを用いて一般物体認識における複数候補提示下の分類性能傾向について述べた．単純な特徴量だけを用いてBag-of-Keypointsのアプローチを取っても，パラメータの調整が上手くいけば実用的な精度が十分出る可能性を持っている．特に30程度の複数候補提示した場合はその条件下でのパラメータの完全調整知識があればほぼ100%の精度を出す事が出来る．しかし，基本的に実験で得られた各分類器のパラメータを，完全に条件同じである場合以外に適用して精度を出す事は難しく，また根本的に分類が不可能な2値問題も存在し，それに対処するには適した特徴量の考案をしたり，学習領域を上手く認識するような工夫が必要となってくる．

参考文献

1) M. Varma, D. Ray. *Learning The Discriminative Power-Inveriance Trade-off*, Proc. of IEEE International Conference on Computer Vision,2007.
2) S. Lazebnik, C. Schmid and J. Ponce. *Spatial Pyramid Matching for Recognizing Natural Scene Categories*, Proc. of IEEE Computer Vision and Pattern Recognition,pp.2169-2178, 2006.
3) A. Bosch, A. Zisserman, X. Munoz. *Image Classification using Random Forests and Ferns*, Proc. of IEEE International Conference on Computer Vision,2007.
4) 藤吉弘亘「Gradientベースの特徴抽出 - SIFTとHOG -」, 情報処理学会 Computer Vision・Image Media 研究会,160,pp.211-224, 2007.
5) H. Zhang, A.C. Berg, M. Maire, J. Malik. *SVM-KNN : Discriminative Nearest Neighbor Classification for Visual Category Recognition*, Proc. of IEEE Computer Vision and Pattern Recognition,pp.2126-2136, 2006.
6) T. Hastie, R. Tibshirani *Classification by pairwise coupling*, The annals of Statics,26,pp.451-471, 1998.
7) T.G. Ditterich, G. Bakiri. *Solving multiclass learning problems via error-correcting output codes*, Journal of Artifical Intelligence Research,2,pp.263-286, 1995.
8) E.L. Allwein, R.E. Schapire and Y. Singer. *Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers*, Journal of Machine Learning Research,2000.
9) F. Aiolli, A. Sperduti *Multiclass Classification with Multi-Prototype Support Vector Machines*, Journal of Machine Learning Research,2005 pp.817-850.
10) J. Weston, C. Watkins. *Multi-class Support Vector Machines*, Technical Report CSD-TR-98-04,1998.
11) M. Gonen, A. Gonul and E. alpaydm. *Multiclass Posterior Probability Support Vector Machines*, Proc. of IEEE Transaction on Neural Networks,2008.
12) 柳井啓司「一般物体認識の現状と今後」, 情報処理学会 Computer Vision・Image Media 研究会, 2006.