

各クラスへの中間出力値を用いた 多クラス認識のための確信度処理

Confidence Analysis for Multi-Class Object Recognition using the Intermediate Values from Machine Learning Algorithms

正会員 山崎 俊彦[†], 学生会員 大島 辰之輔[†], 正会員 相澤 清晴[†]

Toshihiko Yamasaki[†], Shinnosuke Ohshima[†] and Kiyoharu Aizawa[†]

あらまし

我々は、一般物体認識において多クラス分類器が出力する各クラスへの帰属確率や各クラスの分離平面からの符号付き距離などの中間的な値を処理して認識結果に対する確信度を推定することで、正答の可能性が高い場合と誤答の可能性が高い場合を精度よく判別する手法について研究している。この確信度判定について従来の手法を拡張し、各クラスへの中間出力値を複数用いることで確信度判定の精度向上を図る。中間出力値を処理する手法として、数値計算によるものと機械学習を用いたものの二手法を提案する。これらにより、より高精度の確信度推定が実現でき、最終的な物体認識精度の向上が期待できる。また、これまで one-vs-all SVM に対してのみの実験であったが、最近注目を集めている Convolutional Neural Network (CNN) にも適用し、提案手法の一般性と有効性を確認した。

キーワード：一般物体認識, 機械学習, 確信度

1. ま え が き

一般物体認識は古くから研究がなされている分野であるが、近年の計算機能力の向上や種々のアルゴリズムが提案されてきたことにより、急速な進歩を遂げている^{1)~3)}。我々は、機械学習器にも人間と同じように出力ラベルに対する「確信度」のようなものがあり、確信度が低い場合には認識の正答率が極端に低下するというを示してきた⁴⁾⁵⁾。また、確信度が低い場合には別の相補的な画像特徴抽出・機械学習を用いるという多段階認識で行うことで認識精度の向上をさせることができることを示してきた。正解を知る前に正解である可能性が高いことがわかれば、それを最終的な認識結果とする。また、誤っている可能性が高い場合には別の手法を用いて認識しなおすことで正しい認識が行われることを期待するものである。たとえば、4) では exemplar-SVM⁶⁾ を二段目の認識器に利用することで認識精度が向上することを示している。また、5) では local Naive Bayes Nearest Neighbor (local NBNN) 法⁷⁾ を改良した inclined local NBNN 法という手法を提案し、確信度処理と組み合わせることで大幅な認識精度向上ができることを示している。

これらのアルゴリズムが正しく動作するには、機械学習器の確信度をいかに正確に推定するかが重要となる。また確信度判定は、物体認識以外にも半教師学習において反復的にラベル付きトレーニングデータを増やしていく際に、トレーニングデータとして有用なデータから優先的に選択する戦略として有効であることが示されている⁸⁾。この場合にも確信度推定の精度が重要となる。我々の先行研究においては、one-vs-all SVM による多クラス分類における分離超平面からの符号付距離に着目し、候補 1 位に対する距離の符号で判定する手法⁴⁾ と、候補 1 位と 2 位の距離の差分を取る手法⁵⁾ を提案している。

本論文では、より高い精度での確信度推定を目指して、候補 3 位以下のクラスに対する分離超平面からの符号付距離にも注目する。処理手法として、数値計算により求められる既存研究の改良手法と、機械学習によって判定する手法の 2 手法を示し、物体認識で一般的に用いられている 5 種類の画像データセットにより提案手法の有効性を確認する。さらに、これまで検証に用いてきた one-vs-all SVM だけでなく、最近特に注目を集めている CNN²⁾ が出力する帰属確率にも適用し、提案手法が CNN にもそのまま応用することができる一般性の高い手法であることを示す。本論文では確信度の判定精度のみについて議論し、最終的な物体認識精度については議論しないので注意されたい。なお、本論文は 9) の発表をもとに、大幅に実験規模を拡大して提

2014 年 12 月、冬季大会で発表

2015 年 3 月 3 日受付, 2015 年 5 月 11 日再受付, 2015 年 6 月 21 日採録

[†] 東京大学大学院情報理工学系研究科電子情報学専攻

(〒 113-8656 文京区本郷 7-3-1, TEL 03-5841-6761)

案手法の有効性を確認したものである。

2. 関 連 研 究

Support Vector Machine (SVM)¹⁰⁾ は原理的には2クラスを分類するための手法であるが、これを多クラス分類に拡張した one-vs-all SVM という手法がある。One-vs-all SVM では、識別したい K クラスそれぞれに対しそのクラス C_K に所属するかどうかという SVM を学習し、入力に対して k 個の SVM が個別に判別する。得られた K 個の結果のうち最も識別関数値 (分離超平面からの符号付距離) が大きい、つまりそのクラスに所属する可能性が最も高いものを識別結果とすることで、多クラス分類ができるようになっている。なお、類似の手法に one-vs-one SVM という方式もあるが、認識すべきクラス数が大きくなるにつれてメモリ効率の観点から one-vs-all SVM のほうが頻繁に用いられるようになっている。

この one-vs-all SVM による分類では、得られる識別関数値から認識精度を推定する研究¹¹⁾⁴⁾ が行われてきた。ここで、機械学習による認識結果がどれくらい自信を持って出されたものであるかを表すパラメータを、確信度と定義する。先行研究⁵⁾ では、確信度として候補第1位と候補第2位の識別関数値の差に注目した。候補第1位の識別関数値のみを用いる場合に対し、第1位と第2位の結果の比を計算する手法や、第1位と第2位の結果の差を取る手法を実験し比較している。この比較の結果、現在のところ第1位と第2位の結果の差を取る手法を用いた場合が最も確信度の精度が良いということが示されている。

3. 提 案 手 法

既存手法である、候補1位の値をそのまま用いる手法と候補1位と2位の差分を取る手法をそれぞれ $conf_{top}$, $conf_{diff}$ と呼ぶことにする。 k 位の候補に対する識別関数値を $DV(k)$ と定めると、 $conf_{top}$, $conf_{diff}$ はそれぞれ次の様に定式化できる。

$$conf_{top} = DV(1), \quad (1)$$

$$conf_{diff} = DV(1) - DV(2). \quad (2)$$

1つ目の提案手法は、候補3位以下にも目を向けるという視点から、第2位から第 n 位までの平均を算出し第1位との差分を取るという、 $conf_{diff}$ を拡張した方法である。この値を $conf_{topmean}(n)$ と名付け、次式で表す。

$$conf_{topmean}(n) = DV(1) - \frac{1}{n-1} \sum_{i=2}^n DV(i) \quad (3)$$

2つ目の提案手法は、機械学習により確信度算出を行うという手法である。入力された候補1位からの n 位までの値を順に並べたものを新たな特徴ベクトルとして捉え、その識別結果が正解か不正解かの2クラスを分類するタスクと考えて、SVMを用いて学習を行う。これを $conf_{SVM}(n)$

と定義し、式(4)で表される。なお、本実験では線形カーネルを用いる。

$$\begin{aligned} feature &= (DV(1), DV(2), \dots, DV(n))^T, \\ conf_{SVM}(n) &= SVM(feature) \end{aligned} \quad (4)$$

ここで、関数 SVM は認識結果が正しいと思われる場合に1を、誤りと思われる場合に-1として事前にトレーニングデータを用いて学習し、入力に対して分離超平面からの符号付き距離を返す関数である。

1つめの手法は文献5)の拡張を狙ったものである。候補第3位以下の識別関数値も考慮に入れることによって、候補クラス同士がどの程度拮抗しているかをより詳細に判断できると考えた。2つ目の手法は、1つめの手法が天下り的な式であるのに対し、機械学習によって確信度が高い場合と低い場合の識別関数値の分布の様子の違いを自動的に学習することによって確信度を精度よく判定することを狙ったものである。

式(1)~(4)による計算では「確信度」に相当する連続値が出力される。確信度は大きいほど認識結果が正解である可能性が高く、小さいほど可能性が低くなるような値である。ただし、この確信度の値をそのまま評価することは困難であるため、閾値を定め、その閾値よりも確信度が高い場合は「認識結果を正解と予測する」、低い場合は「認識結果を不正解と予測する」として、その判別性能によって評価する。なお、適切な n と閾値を定める必要があるが、これは5)で示されている通りトレーニングデータを使ってプレ・トレーニングとプレ・テストを行うことで事前にかつ自動的に定めることができるので本論文では議論の対象としない。

4. 実 験 結 果

本研究では、one-vs-all SVM による分類手法と深層学習の一手法である convolutional neural network (CNN) による分類手法について確信度の算出を行う。前述のように、one-vs-all SVM からはそれぞれのクラスに対する識別関数値が出力され、CNN からは各クラスに属する確率が出力される。これらの値を用いて式(1)~(4)により確信度を推定する。

One-vs-all SVM による分類手法については、まず画像から SIFT 特徴量を抽出し、bag-of-features (BoF) の一手法である LLCSPM¹²⁾ を用いる。One-vs-all SVM にはフリーソフトウェアである Liblinear¹³⁾ を用いる。式(4)の算出にも Liblinear を利用する。CNN は Caffe²⁾ を用いる。

画像データセットとして Caltech-101, Caltech-256, Caltech-UCSD birds200, CIFAR-10, CIFAR-100 の5つを用いて実験を行う。Caltech-101, Caltech-256 は古くから物体認識に用いられている画像セット、Caltech-UCSD birds200 は fine-grained と呼ばれる細かな差異を認識するための画像セット、CIFAR-10, CIFAR-100 は CNN の

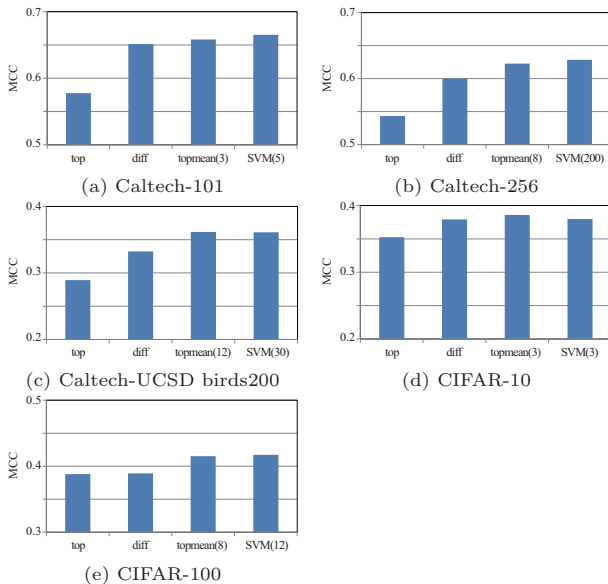


図 1 One-vs-all SVM を用いた場合の提案手法 $conf_{topmean}(n)$, $conf_{SVM}(n)$ の性能

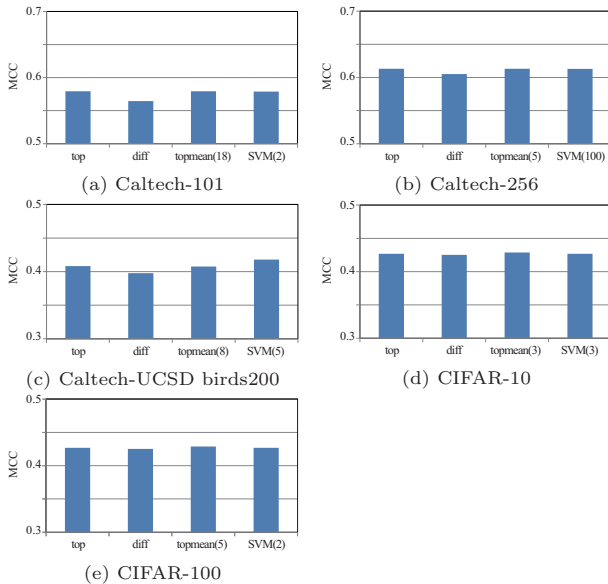


図 2 CNN を用いた場合の提案手法 $conf_{topmean}(n)$, $conf_{SVM}(n)$ の性能

性能を議論するのによく用いられる画像セットである。確信度の精度の指標には Matthews correlation coefficient (MCC)¹⁴⁾ を用いた。MCC は以下の式で定義されている。

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \quad (5)$$

なお, TP, TN, FP, FN はそれぞれ True Positive, True Negative, False Positive, False Negative の個数を表す。

One-vs-all SVM を用いた時の確信度推定の結果を図 1 に示す。ここで、式 (3), (4) における n は前述のとおり MCC 値を最大とするクラス数 n であり、事前に自動的に決定することができる。図 1 をみると、提案手法である $conf_{topmean}(n)$ と $conf_{SVM}(n)$ はともに既存の手法を上回るか、悪くとも同等以上の成績となっていることが確認で

きる。特に 5 つの中で難しいデータセットである Caltech-UCSD birds200 や CIFAR100 での結果においては確信度推定が大きく改善されている。候補 3 位以下まで着目するという新しい視点、機械学習を行うという異なったアプローチのどちらにおいても確信度推定において効果的な手法であったと言える。また、学習を用いる $conf_{SVM}$ が必ず最善とは言えない場合があるが、これはトレーニング時の過学習が原因と考えられる。今後データセットが大規模化し、充分多くのトレーニングデータが得られた場合にはこの問題は解消されることが期待される。

CNN を用いた時の確信度推定の結果を図 2 に示す。CNN はもともと認識精度が高いため、確信度推定は One-vs-all SVM に対するそれほど大きな意味を持たない。しかし、提案手法により微妙ではあるがすべての場合において確信度推定の MCC 値が向上しており、これは CNN と別の認識アルゴリズムを組み合わせることによって物体認識の精度がさらに向上することを示唆している。

5. む す び

本稿では、多クラス分類のための機械学習アルゴリズムに対して、従来手法より精度よく認識結果に対する確信度を推定する手法を提案した。候補第 3 位以下についても着目する $conf_{topmean}$ と、確信度推定に機械学習を用いる $conf_{SVM}$ とを提案した。これら 2 つの提案手法について、複数の難易度の異なる 5 種類のデータセットと 2 種類の認識手法を用いて実験を行い、既存手法との比較検討を行った。その上で、提案手法の $conf_{topmean}$ と $conf_{SVM}$ はともに既存手法よりも精度が高く、認識条件の変化に強いことを示した。本論文では物体認識を対象として議論したが、多クラス分類を用いるほかの研究分野にも適用可能であると思われる。

謝 辞

本研究の一部は科学研究費補助金 (26540078)、および一般財団法人テレコム先端技術研究支援センター研究助成 (SCAT) の支援プログラムを受けて行われた。

【文 献】

- 1) A. Andreopoulos and J. K. Tsotsos. 50 years of object recognition: Directions forward. *Computer Vision and Image Understanding*, 117(8), pp. 827 – 891 (2013)
- 2) Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093* (2014)
- 3) S.S. Bucak, J. Rong, and A.K. Jain. Multiple kernel learning for visual object recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), pp. 1354–1369 (July 2014)
- 4) T. Yamasaki and T. Chen. Confidence-assisted classification result refinement for object recognition featuring topn-exemplar-svm. In *ICPR*, pp. 1783–1786 (2012)
- 5) T. Maeda, T. Yamasaki, and K. Aizawa. Multi-stage object classification featuring confidence analysis of classifier and inclined local naive bayes nearest neighbor. In *ICIP* (2014)

- 6) A. A. Efros T. Malisiewicz, A. Gupta. Ensemble of exemplar-svms for object detection and beyond. In *ICCV* (2012)
- 7) S. McCann and D. G. Lowe. Local naive bayes nearest neighbor for image classification. In *CVPR* (2012)
- 8) 貴志一樹, 山崎俊彦, and 相澤清晴. 机上行動のライフログのための行動認識. In *HCG シンポジウム*, pp. 582–586 (2014)
- 9) 大島辰之輔, 山崎俊彦, and 相澤清晴. マルチクラス物体認識のための学習を用いた確信度処理. In *映像情報メディア学会冬季大会* (2014)
- 10) C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3), pp. 273–297 (1995)
- 11) J. Fu and S. Lee. A multi-class svm classification system based on learning methods from indistinguishable chinese official documents. *Expert Systems with Applications*, 39(3), pp. 3127–3134 (2012)
- 12) J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *CVPR*, pp. 3360–3367 (2010)
- 13) R. Fan, K. Chang, C. Hsieh, X. Wang, and C. Lin. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9, pp. 1871–1874 (2008)
- 14) B.W. Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405(2), pp. 442–451 (1975)



やまさき としひこ
山崎 俊彦 東京大学工学部電子工学科卒業。東京大学工学系研究科電子工学専攻修了。博士(工学)。現在、東京大学大学院情報理工学系研究科電子情報学専攻 准教授。大規模マルチメディアデータ処理、3次元映像処理、物体認識・機械学習などの研究に従事。正会員。



おおしましんのすけ
大島辰之輔 東京大学工学部電子情報工学科卒業。現在、東京大学大学院情報理工学系研究科電子情報学専攻修正課程在籍。学生会員。



あいざわ きよ はる
相澤 清晴 1983年 東京大学工学部電子工学科卒業。1988年 同大学院博士課程修了。工博。東大工学部助手、講師、助教授を経て、2001年より、教授。現在、大学院情報理工学系研究科教授。この間、画像・マルチメディア処理、ライフログ、3次元映像、漫画画像処理等に関する研究に従事。学術会議連携会員。本学会フェロー。