

変形部品モデルの高速学習法と顔パーツ検出への応用

高岸 謙斗†

宇敷 卓哉†

加藤 毅†

†群馬大学大学院理工学専攻

1 はじめに

コンピュータビジョンの分野において、顔パーツ検出や姿勢推定などの構造予測を行うに当たり、変形部品モデル (Deformable Part Model, DPM) は広く用いられてきた。DPM は、ノードとエッジからなる無向グラフからなっており、顔パーツ検出の場合、各ノードは顔のパーツに対応する。各ノードとエッジはパラメトリックなスコア関数を保持し、従来、そのパラメータは SVM 学習によって獲得されてきた。SVM 学習は、2 クラス分類用の誤識別率を直接最小化する代わりに、ヒンジ損失と呼ばれるサロゲート損失の最小化を行う方法である。近年、Uricăřら [1] は、DPM のパラメータの学習に、構造学習 (structural learning) を導入して、検出のズレを減少させた。構造学習で用いられるサロゲート損失は、構造的出力の損失を評価できるように拡張されたヒンジ損失 (ヒンジ構造損失と呼ぶ) である。

一方、近年の機械学習では、確率的勾配法 (SGD 法) を使った学習が主流になっている。SGD 法は従来の最急勾配法よりも 1 反復あたりの計算量を小さくすることができたが、収束の遅さが問題になっていた。これを解決するために SGD 法に改良を施した、Stochastic Average Gradient (SAG) 法 [2] や Stochastic Variance Reduced Gradient (SVRG) 法 [3] などが出現した。これらは、強凸でスムーズなサロゲート損失関数を仮定していた。しかし、これまでの構造学習においては、筆者らの知る限り、スムーズなサロゲート損失関数は存在していなかったため、DPM を最適解まで収束させようと学習することが非現実的であった。

本論文では、DPM の学習方法として新たに開発したロジスティック構造学習を提案する。具体的には次の 3 つの貢献をする。(i) 利用者が定義した損失に対する新たなサロゲート損失 (ロジスティック構造損失と呼ぶ) を提案し、構造学習に、SAG 法や SVRG 法を適用可能にする。これによって、DPM の学習の精度を高め、結果として検出性能の向上をはかるものである。従来のヒンジ構造損失の場合、一つの訓練用画像あたりのサロゲート損失の勾配を効率的に計算する動的計画法が存

在した。今回新たに考案したロジスティック構造損失でも、一つの訓練用画像あたりのサロゲート損失の勾配を効率的に計算する動的計画法を組むことができると同時に、強凸性とスムーズ性を兼ね備えているため、SAG 法や SVRG 法によって、実用的な計算時間で、最適解に収束させることができる。(ii) 顔画像の実データセットに適用した数値実験を行い、提案するロジスティック構造損失と SAG 法や SVRG 法を組み合わせることで、最適解に実用的な計算量で到達できることを数値実験を通して示す。(iii) さらに、顔画像のデータセットにおいて顔パーツを検出する計算機実験により、提案するロジスティック構造損失は、従来のヒンジ構造損失や、Schmidtら [4] の損失関数より、検出性能が顕著に向上することを示す。

2 ロジスティック構造学習の提案

本節では、ロジスティック構造学習という新たな構造学習の枠組みを提案する。構造学習とは、 $h: X \rightarrow Y$ なる関数を学習する問題である。ただし、 X は入力空間、 Y は離散的な出力空間である。DPM を学習する問題の場合、 X は入力画像の空間であり、 Y は、各ノードの状態の組み合わせの空間である。以後、画像に焦点を絞って議論する。ここでは、 h が、 $h(I) := h(I; w) := \operatorname{argmax}_{y \in Y} \langle \Psi(I, y), w \rangle$ のような形式で与えられるとする。ただし、 $I \in X$ は入力画像、 $\Psi: X \times Y \rightarrow \mathbb{R}^d$ は状態の組み合わせ y に対する特徴ベクトルを返す関数、 $w \in \mathbb{R}^d$ はモデルパラメータとする。

構造学習の特徴は、予測値 $y \in Y$ が正解値 $y^* \in Y$ と異なったときに、どれほどの損失があるか、利用者が任意に決定できることである。すなわち損失 $\Delta: Y \times Y \rightarrow \mathbb{R}_+$ の定義は、 $\Delta(y^*, y^*) = 0$ を除けば任意である。所与の訓練用データ $\{(I_i, y_i)\}_{i=1}^{\ell}$ に対して、経験リスク $\frac{1}{\ell} \sum_{i=1}^{\ell} \Delta(y_i, h(I_i; w))$ を直接最小化することは難しいので、提案するロジスティック構造学習では、 $\Delta(\cdot, h(\cdot; \cdot))$ を次のサロゲート損失に置き換える。

$$l_{\log i}(w; I, y^*) := -\langle \Psi(I, y^*), w \rangle + \log \sum_{y \in Y} \exp(\Delta(y, h(I; w)) + \langle \Psi(I, y), w \rangle) \quad (1)$$

これをロジスティック構造損失と呼ぶ。ロジスティック構造損失 (1) は、 $\Delta(\mathbf{y}, h(I; \mathbf{w})) \leq l_{\text{logi}}(\mathbf{w}; I, \mathbf{y}^*)$ のように、 $\Delta(\cdot, \cdot)$ の上限になっている。 \mathbf{w} の値は、目的関数 $f_{\text{logi}}(\mathbf{w}) := \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\ell} \sum_{i=1}^{\ell} l_{\text{logi}}(\mathbf{w}; I_i, \mathbf{y}_i^*)$ を最小化することで決める。この関数を最小化することによってモデルパラメータ \mathbf{w} を決めることをロジスティック構造学習と呼ぶことにする。

従来の確率的勾配法 (SGD 法) でも、SAG 法でも SVRG 法でもある入出力ペア (I, \mathbf{y}) のサロゲート損失 $l(\cdot; I, \mathbf{y})$ に対して、モデルパラメータに関する勾配 $\nabla_{\mathbf{w}} l(\mathbf{w}; I, \mathbf{y})$ が各反復で必要になる。この勾配を効率的に計算できないとどの方法を使ったとしても現実的な時間で最適解に到達できない。

3 変形部品モデル (DPM)

DPM は、各パーツをグラフのノードで表し、それをループができないようにエッジで連結したモデルである。本研究では、Uřičář ら [1] のモデルを採用した。すなわち、顔パーツを検出するので、ノード集合を $V := \{\text{顔中心, 鼻, 左目頭, 右目頭, 左口角, 右口角, 左目尻, 右目尻}\} = \{v_1, v_2, \dots, v_M\}$ とした。ノード数は $M = 8$ である。各ノード $v_m \in V$ の取り得る状態集合 Σ_m とする。 v_m の状態を $y_m \in \Sigma_m$ で表す。出力空間は、各ノードの状態空間の直積空間 $\mathcal{Y} := \Sigma_1 \times \dots \times \Sigma_m$ で与えられる。よって、モデル全体の形態 $\mathbf{y} \in \mathcal{Y}$ は $\mathbf{y} = [y_1, \dots, y_m]^T$ のように表現できる。DPM で予測すると、出力空間 \mathcal{Y} から適切な形態 \mathbf{y} を選択することによって、各顔パーツを同時に検出することになる。

ロジスティック構造損失を用いて学習するには、勾配の計算が必要になる。ロジスティック構造損失の勾配は

$$\nabla l_{\text{logi}}(\mathbf{w}; I, \mathbf{y}^*) = -\Psi(I, \mathbf{y}^*) + \sum_{\mathbf{y} \in \mathcal{Y}} \Psi(I, \hat{\mathbf{y}}) \exp(\langle \Psi(I, \mathbf{y}), \mathbf{w} \rangle + \Delta(\mathbf{y}^*, \mathbf{y}) - S(\mathbf{w}, I, \mathbf{y}^*)) \quad (2)$$

と表される。ただし、

$$S(\mathbf{w}, I, \mathbf{y}^*) := \log \sum_{\mathbf{y} \in \mathcal{Y}} \exp(\langle \Psi(I, \mathbf{y}), \mathbf{w} \rangle + \Delta(\mathbf{y}^*, \mathbf{y}))$$

とおいた。この計算は、ナイーブに計算すると $O(H^M d)$ の計算時間がかかる。ただし、 $H := \max_m |\Sigma_m|$ である。しかし、Uřičář ら [1] のように分解可能な損失 $\Delta(\cdot, \cdot)$ を使い、エッジ特徴 [1] の数が $O(1)$ の場合、次の定理を示すことができる。

Theorem 3.1. ロジスティック構造損失の勾配 (2) は、 $O(MH^2 + Hd)$ で計算できる。

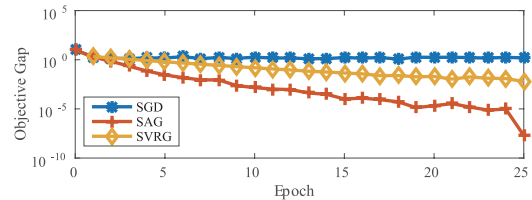


図1 ロジスティック構造損失による目的関数 $f_{\text{logi}}(\mathbf{w})$ と最適解との差。

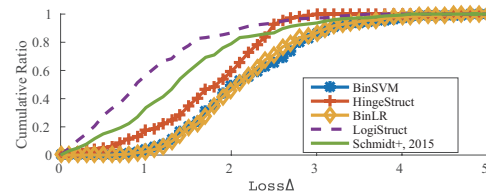


図2 累積エラー曲線。

4 実験

提案した学習アルゴリズムの最適解への収束性と汎化能力を検証するため、顔画像のデータセット Labeled Faces in the Wild (LFW) のうち、無作為に選んだ 700 枚を学習用に、300 枚を評価用に用いた。

図 1 に、ロジスティック構造損失による目的関数 $f_{\text{logi}}(\mathbf{w})$ と最適解との差をプロットした。SAG 法や SVRG 法では、実用的なエポック数で最適解にほぼ収束していることが観測できる。

次に、汎化性能を次の手法と比較した：BinSVM(2 クラス SVM)、BinLR(2 クラス分類用ロジスティック回帰)、HingeStruct[1]、Schmidt+ 2015[4]、LogiStruct(提案法)。累積エラー曲線を図 2 にプロットした。提案法の汎化性能がどの方法よりもっとも良いことが確認できる。

謝辞: 本研究は JSPS 科研費 26249075, 40401236 の助成を受けたものである。

参考文献

[1] M. Uřičář et al. Real-time multi-view facial landmark detector learned by the structured output SVM. In *BWILD'15*, New York, US, May 2015. IEEE Computer Society.
 [2] N. L. Roux et al. A stochastic gradient method with an exponential convergence rate for finite training sets. In *NIPS 25*, pp. 2663–2671. Curran Associates, Inc., 2012.
 [3] R. Johnson and T. Zhang. Accelerating stochastic gradient descent using predictive variance reduction. In *NIPS 26*, pp. 315–323, 2013.
 [4] M. Schmidt et al. Non-uniform stochastic average gradient method for training conditional random fields. In *AISTATS 2015*, 2015.