Vol.2016-CVIM-200 No.24
2016/1/22

Deformable Part Model のためのロジスティック構造学習

宇敷 卓哉[1]　高岸 謙斗[1]　加藤 毅[1]

概要：今日では，Deformable Part Model (DPM) は，顔パーツ検出や，姿勢推定などで，広く使われている．DPM を構造学習と呼ばれる学習法により高精度化するとの研究がある．学習のための最適化に勾配法を用いるとすると，各反復における勾配の計算が必要となるが，DPM では，パーツ間を木構造で繋ぐとことにより，動的計画法で効率的に勾配を計算できるようになっていた．しかし，その学習において最小化される損失関数がスムースではなかったために，適用できる確率的勾配法が限定され，それゆえに十分に最適解に到達できないことが多かった．本研究では，高速な確率的勾配法に適用できるよう，新たにスムースな損失関数を考案した．その損失関数は，従来の損失関数と同様，動的計画法で効率的に勾配を計算できるように設計した．これによって，高速かつ高精度な最適化手法が適用可能になり，結果として検出精度の向上が得られた．

キーワード：Deformable Part Model，確率的勾配法，ロジスティック構造学習

1. 導入

顔パーツ検出や姿勢推定などの構造予測を行うにあたり，**Deformable Part Model**（DPM） [1] は広く用いられてきた [2] [3] [4]．DPM は，モデルをノードとエッジで示される無向グラフとして捉える手法である．本研究で扱う顔パーツ検出であれば，ノードは顔の各パーツを意味する．すると，顔パーツを検出する問題は，各ノード間の位置関係を保持したまま，モデルが画像にもっともフィットされるような各パーツの位置を探す問題になる．各パーツは，離散化された位置からなり，位置がノードの状態を表す．DPM では，ノードごとに，あるノードの状態，すなわち位置が入力画像にどれほどマッチしているかを表すマッチ関数と，ノード間の位置関係がどれほど崩れていないかを表すエッジ関数からなっている．それらの総和をスコア関数と呼ぶことにする．顔パーツの検出問題は，スコア関数が最大となる各ノードの状態の組み合わせを探す問題になる．これによって，全てのノードの位置を，ある程度の変形を許しながら，同時に予測することができる．

DPM の大きな特色は，判別的学習 (discriminant learning) が可能な点である．DPM の出現まで，Active Shape Model (ASM) [5] [6] や Active Appearance Model (AAM) [7] [8] などが用いられていた．これらは，検出対象の画像から学習によってモデルパラメータを決定するものであったが，検出対象以外の画像は使っておらず，その学習は判別的ではなかった．DPM の学習は判別的である．DPM の各々のノードのマッチ関数と各々のエッジのエッジ関数はモデルパラメータに依存していて，このモデルパラメータを学習によって調整することで，DPM を顔パーツ検出や姿勢推定など所与の問題に特化したモデルにすることができる．各々のマッチ関数やエッジ関数は，すべて，その状態と入力画像に依存した特徴ベクトルと，モデルパラメータの内積で表されている．それらの特徴ベクトルをひとつに連結することで，スコア関数は，状態の組み合わせと入力画像に依存した特徴ベクトルとモデルパラメータとの内積の形式に帰着され，SVM などの多くの機械学習アルゴリズムに適用できるようになっている．本研究では，モデルパラメータの学習法を改良したロジスティック構造学習という枠組みを新たに開発し，DPM に適用することによって，DPM の検出性能の向上をはかる．

関連研究

構造学習 (structural learning) [9][10] と呼ばれる特殊な機械学習の枠組みがある．標準的な２クラス分類では，誤識別率を最小化するようにモデルパラメータを学習するものであった．誤識別率そのものの最小化は巨大な混合整数計画問題になってしまうので，標準的な SVM 学習は代わりにヒンジ損失 (hinge loss) の平均を最小化していた．ヒンジ損失の平均は，誤識別率の上限になっているため，標準的な SVM 学習は誤識別率を最小化する代わりに，誤識別率の上限を最小化することになる．誤識別率の上限を表す損失関

[1] 群馬大学理工学部，〒 376–8515 群馬県桐生市天神町 1–5–1
School of Science and Technology, Gunma University Tenjin-cho 1–5–1, Kiryu-shi, Gunma, 376–8515 Japan

数はサロゲート損失 (surrogate loss) [11] と呼ばれており，そのほかにも最小二乗損失 (least square loss)[12]，スムース化ヒンジ損失 (smoothed hinge loss)[13]，ロジスティック損失 (logistic loss)[12] などが開発されている．構造学習は，いわば，大規模なクラス数を持つ多クラス問題のために学習するような，一般化された SVM 学習の枠組みである．多クラス問題の場合，ある訓練用例題に対して，あるクラスに誤ったときの損失と，その他のクラスに誤ったときの損害の程度が異なることがある．その損害の程度を損失関数として，利用者が任意に定義することができ，構造学習では，学習用例題集合における損失の平均を最小化するように学習する．しかし，損失の平均を直接最小化するのは難しいので，損失の上限を表すヒンジ構造損失 (hinge structural loss) の平均を最小化する．

Uřičář ら [14] は，顔パーツ検出のため，構造学習に DPM を適用し，検出の精度に直結する独自の損失関数を直接学習することを可能にした．Uřičář ら [14] 以前の DPM は，従来の 2 クラス分類用の SVM 学習を用いていた．そのために，顔パーツ検出のような画像中の物体検出特有の損失の性質を学習出来ずにいた．顔パーツ検出の場合，目や鼻などが，正解より少しだけずれているのと，大きくずれていたのとでは，被る損失が異なる．たとえば，顔パーツ検出を個人認証の前処理として使うとするなら，パーツが少しだけずれて検出されるだけなら，後続の個人認証フェーズへの悪影響は少ない．一方，大きくずれて検出されてしまった場合，個人認証フェーズにおいて誤った情報を入力することになり，無視しがたい悪影響を及ぼす．よって，ずれの大きさに比例するように損失を設計するのが自然である．Uřičář ら [14] はそのような損失関数を設計して，構造学習の枠組みに当てはめることで，検出のズレを減少させることに成功した．

Uřičář ら [14] など，近年の機械学習は，確率的勾配法 (stochastic gradient method) を使って，学習を行うことが主流となっている [15] [16]．従来の最急降下法では，訓練用のすべての画像に対して，サロゲート損失の勾配を計算し，その平均で最急降下方向を計算していた．しかし，画像数が多い場合や，勾配の計算に時間がかかる場合は，学習に用いる目的関数の勾配を 1 回計算するのに大きな計算量がかかってしまっていた．これに対して，確率的勾配法では，各反復で，訓練用画像の中から，無作為に一つだけ画像を選択してその損失の勾配からモデルパラメータを動かすので，1 反復あたりの計算時間を激減できる．確率的勾配法は，更新方向の期待値をとると，たしかに，最急降下方向になるが，その分散は大きく，選択される画像によっては最急降下方向と逆向きになることさえあるため，収束の遅さが問題になっていた．理論的には，凸問題であれば最適解に収束することが保証されているが，実際には，最適解に収束するまで計算し続けるのは非現実的であるため，十分に最適解に到達しないまま，学習を停止せざるを得なかった．

DPM の出現や構造学習による改良とは独立して，機械学習分野では，近年，最適化アルゴリズムが大きな発展を遂げている．それは，確率的勾配法 (stochastic gradient method) の改良である．たとえば，Stochastic Average Gradient (SAG) 法 [17] や Stochastic Variance Reduced Gradient（SVRG）法 [18] などがあげられ，これらを使うと，現実的な計算時間で，最適解に収束させることができる．しかし，これらは，2 クラス分類を想定していて，強凸でスムースなサロゲート損失を仮定していた．一方，構造学習においては，筆者の知る限り，強凸でスムースなサロゲート損失が存在していなかったため，これまでの構造学習は，従来の確率的勾配法を使わざるを得なかった．Schmidt ら [19] は，構造学習にスムースな損失関数を導入して，SAG 法を適用しようとしたが，彼らの定式化は利用者が定義した損失を取り込める形式にはなっていなかった．

本研究の貢献

本研究の貢献をまとめると次のようになる．

- 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，強凸性とスムース性を兼ね備えているため，SAG 法や SVRG 法によって，現実的な計算時間で，最適解に収束させることができる．
- 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 法を組み合わせると，最適解に現実的な計算量で到達できることを数値実験を通して示す．
- 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 [19] の損失関数より，検出性能が顕著に向上することを示す．

本論文の構成

本論文の構成は次のようである．2 節で，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．3 節で，本研究に使用した DPM の詳細を述べ，その DPM がロジスティック構造学習の枠組みで効率的に計算できることを示す．4 節で，スムース性を仮定することによって高速化された確率的勾配法である SAG 法や SVRG 法をロジスティック構造損失に適用したところ，最適解に現実的な計算量で収束することを示し，結果として，検出性能が向上す

ることを，顔パーツ検出の実データを使って示す．5 節で結論を述べる．

記法

- $\mathbb{R}^n$ ， $\mathbb{N}^n$ はそれぞれ n 次元の実数集合，n 次元の自然数集合を表し，$\mathbb{R}^{m\times n}$ は $m\times n$ の実数行列の集合を表す．
- $\mathbb{R}^n_+$ はすべての要素が 0 以上である n 次元実数ベクトルの集合を表す．$\mathbb{R}^n_{++}$ はすべての要素が 0 より大きい n 次元実数ベクトルの集合を表す．
- ある $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$ があったとき，$\langle \boldsymbol{x},\boldsymbol{y}\rangle := \sum_{i=1}^n x_i y_i$ のように定義する．ただし，x_i と y_i はそれぞれ $\boldsymbol{x}$ と $\boldsymbol{y}$ の第 i 成分とする．
- 行列 $\boldsymbol{A}$ の転置行列は $\boldsymbol{A}^\top$ ，逆行列は $\boldsymbol{A}^{-1}$ のように表す．

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，$h:\mathcal{X}\to\mathcal{Y}$ なる関数を学習する問題である．ただし，$\mathcal{X}$ は入力空間，$\mathcal{Y}$ は離散的な出力空間である．DPM を学習する問題の場合，$\mathcal{X}$ は入力画像の空間であり，$\mathcal{Y}$ 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，

$$h(I) := h(I;\boldsymbol{w}) := \operatorname*{argmax}_{\boldsymbol{y}\in\mathcal{Y}} \langle \boldsymbol{\Psi}(I,\boldsymbol{y}),\boldsymbol{w}\rangle$$

のような形式で与えられるとする．ただし，$I\in\mathcal{X}$ は入力画像，$\boldsymbol{\Psi}:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}^d$ は状態の組み合わせ $\boldsymbol{y}$ に対する特徴ベクトルを返す関数，$\boldsymbol{w}\in\mathbb{R}^d$ はモデルパラメータとする．

構造学習の特徴は，予測値 $\boldsymbol{y}\in\mathcal{Y}$ が正解値 $\boldsymbol{y}^*\in\mathcal{Y}$ と異なったときに，どれほどの損失があるか，利用者が決定できることである．すなわち損失 $\Delta:\mathcal{Y}\times\mathcal{Y}\to\mathbb{R}_+$ の定義は，$\Delta(\boldsymbol{y}^*,\boldsymbol{y}^*)=0$ を除けば任意である．

画像 I とその正解 $\boldsymbol{y}$ がある同時分布 $P(I,\boldsymbol{y})$ に従っているとする．構造学習の目標は，リスク

$$R_P(\boldsymbol{w}) := \int_{\mathcal{X}\times\mathcal{Y}} \Delta(\boldsymbol{y},h(I;\boldsymbol{w}))dP(I,\boldsymbol{y})$$

を最小にするモデルパラメータ $\boldsymbol{w}\in\mathbb{R}^d$ を見つけることである．しかし，実際には $P(\cdot,\cdot)$ は未知である．よって，有限の訓練用データ集合 $(I_1,\boldsymbol{y}_1^*),\ldots,(I_\ell,\boldsymbol{y}_\ell^*)$ を収集して，正則化経験リスク

$$R_S(\boldsymbol{w}) := \frac{\lambda}{2}\|\boldsymbol{w}\|^2 + \frac{1}{\ell}\sum_{i=1}^{\ell}\Delta(\boldsymbol{y}_i^*,h(I_i;\boldsymbol{w}))$$

を最小化しようとする．$\lambda\in\mathbb{R}_{++}$ は正則化定数である．しかし，これでも $\boldsymbol{w}$ の最小化は難しい．なぜならば，$R_S(\cdot)$ が凸関数ではないからである．従来の構造学習 [9][10] [14] では，計算可能な最適化問題に変換するため，Δ を次の関数に置き換えていた：

$$l_{\text{hinge}}(\boldsymbol{w};I,\boldsymbol{y}^*) := \max_{\boldsymbol{y}\in\mathcal{Y}} \left(\Delta(\boldsymbol{y},h(I;\boldsymbol{w})) + \langle\boldsymbol{\Psi}(I,\boldsymbol{y}),\boldsymbol{w}\rangle\right) - \langle\boldsymbol{\Psi}(I,\boldsymbol{y}^*),\boldsymbol{w}\rangle\,. \tag{1}$$

本稿では (1) をヒンジ構造損失と呼ぶ．ヒンジ構造損失は，

$$\Delta(\boldsymbol{y},h(I;\boldsymbol{w})) \le l_{\text{hinge}}(\boldsymbol{w};I,\boldsymbol{y}^*)$$

のように，$\Delta(\cdot,\cdot)$ の上限になっている．モデルパラメータ $\boldsymbol{w}$ の値を決定するための目的関数は

$$f_{\text{hinge}}(\boldsymbol{w}) = \frac{\lambda}{2}\|\boldsymbol{w}\|^2 + \frac{1}{\ell}\sum_{i=1}^{\ell} l_{\text{hinge}}(\boldsymbol{w};I_i,\boldsymbol{y}_i^*)$$

と表される．この目的関数 f_{hinge} を最小化することは，正則化経験リスクの上限を最小化していることに他ならない．目的関数 f_{hinge} の最小化には確率的勾配法などを利用できる．ここで問題となるのは，この目的関数 f_{hinge} は強凸ではあるがスムースではないため，確率的勾配法を高速化した SAG 法 [17] や SVRG 法 [18] の仮定には違反してしまう．

一方，Schmidt ら [19] は，SAG 法を構造学習に適用するために，ヒンジ構造損失に代わるサロゲート損失として，

$$l_{\text{sch}}(\boldsymbol{w};I,\boldsymbol{y}^*) := \log\sum_{\boldsymbol{y}\in\mathcal{Y}}\exp\left(\langle\boldsymbol{\Psi}(I,\boldsymbol{y}),\boldsymbol{w}\rangle\right) - \langle\boldsymbol{\Psi}(I,\boldsymbol{y}^*),\boldsymbol{w}\rangle \tag{2}$$

を使った．しかし，このサロゲート損失は損失 Δ の上限にはなっておらず，損失 Δ の値の違いを反映できない．

そこで，筆者らは，ヒンジ構造損失や Schmidt ら [19] のサロゲート損失に代わって，次のサロゲート損失を考案した：

$$l_{\text{logi}}(\boldsymbol{w};I,\boldsymbol{y}^*) := \log\sum_{\boldsymbol{y}\in\mathcal{Y}}\exp\left(\Delta(\boldsymbol{y},h(I;\boldsymbol{w})) + \langle\boldsymbol{\Psi}(I,\boldsymbol{y}),\boldsymbol{w}\rangle\right) - \langle\boldsymbol{\Psi}(I,\boldsymbol{y}^*),\boldsymbol{w}\rangle\,. \tag{3}$$

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 (3) を $\log(2)$ で割ったものは，

$$\Delta(\boldsymbol{y},h(I;\boldsymbol{w})) \le \frac{l_{\text{logi}}(\boldsymbol{w};I,\boldsymbol{y}^*)}{\log(2)}$$

のように，$\Delta(\cdot,\cdot)$ の上限になっている．$\boldsymbol{w}$ の値の決定は，目的関数

$$\frac{\lambda}{2}\|\boldsymbol{w}\|^2 + \frac{1}{\ell}\sum_{i=1}^{\ell}\frac{l_{\text{logi}}(\boldsymbol{w};I_i,\boldsymbol{y}_i^*)}{\log(2)}$$

の最小化によって行う．$\log(2)$ を正則化定数に含めてしまえば，目的関数を

$$f_{\text{logi}}(\boldsymbol{w}) := \frac{\lambda}{2}\|\boldsymbol{w}\|^2 + \frac{1}{\ell}\sum_{i=1}^{\ell} l_{\text{logi}}(\boldsymbol{w};I_i,\boldsymbol{y}_i^*)$$

と書くことができる．この関数を最小化することによって

モデルパラメータ $\boldsymbol{w}$ 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 $(I, \boldsymbol{y})$ のサロゲート損失 $l(\cdot; I, \boldsymbol{y})$ に対して，モデルパラメータに関する勾配

$$\nabla_{\boldsymbol{w}} l(\boldsymbol{w}; I, \boldsymbol{y})$$

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. Deformable Part Model (DPM)

本節では，本研究で用いた DPM について述べる．ここでは，顔のパーツ検出の問題に特化して記述するが，ほかの DPM の問題にも適用可能である．本研究では，Ûřîčář らの先行研究 [14] と検出性能を比較しやすいように，Ûřîčář らとほぼ同じモデルを用いた．ただし，Ûřîčář らは損失関数 Δ を定義するとき，画像中の顔の大きさに依存しないよう，目の中心と鼻との距離で正規化していたが，本研究では，より一般的な問題にもそのまま適用できるよう，標準偏差で正規化した Δ を用いた（詳細は 3.3 節を参照）．

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，顔パーツを検出するので，ノード集合を

$$\begin{aligned} V :=& \{ \text{顔中心}, \text{鼻}, \text{左目頭}, \\ & \quad \text{右目頭}, \text{左口角}, \text{右口角}, \text{左目尻}, \text{右目尻} \} \\ =& \{v_1, v_2, \ldots, v_M\} \end{aligned}$$

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 とし，$v_m = m$ とも表現する．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

$$\mathcal{Y} := \Sigma_1 \times \cdots \times \Sigma_m$$

で与えられる．よって，モデル全体の形態 $\boldsymbol{y} \in \mathcal{Y}$ は $\boldsymbol{y} = [y_1, \ldots, y_m]^\top$ のように表現できる．したがって，DPM の目的は，出力空間 $\mathcal{Y}$ から適切な形態 $\boldsymbol{y}$ 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である．

本研究における DPM に基づくスコア関数は，

$$J(\boldsymbol{y}; I, \boldsymbol{w}) := \sum_{v_m \in V} q_{v_m}(y_m) + \sum_{e_m \in E} g_{e_m}(y_{\mathrm{pa}(m)}, y_m) \quad (4)$$

である．ここで，E はエッジ集合を表す．DPM では木のモデルを用いるので $|E| = |V| - 1$ を満たし，$\mathrm{pa}(m)$ は $\forall v_m \in V$ における，隣接関係上の親ノードを表す．v_1 を根ノードとして約束しておく．また，$e_m = (\mathrm{pa}(m), m)$ はノード v_m とその親ノードとの間のエッジを意味する．マッチ関数 $q_{v_m} : \Sigma_m \to \mathbb{R}$ はノード v_m の状態がどれほど入力画像にマッチしているかを表す．エッジ関数 $g_{e_m} : \Sigma_{\mathrm{pa}(m)} \times \Sigma_m \to \mathbb{R}$ は，エッジ e_m が繋ぐ 2 つのノード $\mathrm{pa}(m)$ と v_m の関係が崩れるほど小さな値になるよう学習される．各項 q_{v_m} および g_{e_m} は，入力画像 I やモデルパラメータ $\boldsymbol{w}$ にも依存しているが，表記を簡略化するため，依存していることを明示的に表現するのを避けた．それぞれの関数の定義については，3.1 節および 3.2 節で与える．構造学習を行うための損失関数 $\Delta(\cdot, \cdot)$ は 3.3 節で述べ，この損失関数がヒンジ構造学習にもロジスティック構造学習にも適用可能なことを 3.4 節で示す．

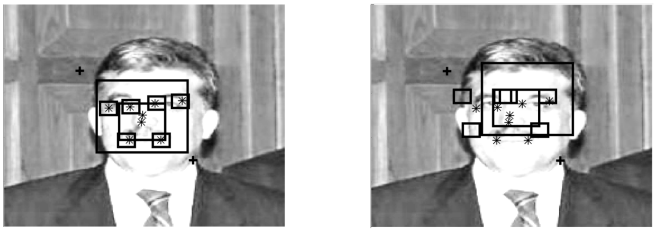

図 1　良好な形態（左）と不適切な形態（右）における損失 Δ の比較．この画像においては，状態集合より最適な形状を選ぶと損失関数値は 0.381 となり，不適切な形状を選ぶと損失関数値は 5.436 となった．

3.1 マッチ関数

マッチ関数は，ノード v_m の状態が入力画像にマッチしているほど大きな値になるよう学習される関数である．本研究においてはマッチ関数をノード v_m のある状態 y_m に対し，

$$q_{v_m}(y_m) := \langle \boldsymbol{w}_{q,m}, \boldsymbol{\Psi}^q_{v_m}(y_m) \rangle \quad (5)$$

として定義する．ただし，$\boldsymbol{w}_{q,m}$ はノード v_m のマッチ関数のパラメータである．本研究においては，各 $\boldsymbol{\Psi}^q$ には，Histogram of Oriented Gradients（HOG）特徴 [20] を用いた．

3.2 エッジ関数

エッジ関数は，ノード間の隣接関係の類似度を計る項であり，本研究においては，ノード v_m とその親ノード $\mathrm{pa}(m)$ を繋ぐエッジ e_m に対し，子ノードのある状態を y_m と親ノードのある状態を $y_{\mathrm{pa}(m)}$ としたときのエッジ関数を，

$$g_{e_m}(y_{\mathrm{pa}(m)}, y_m) := \langle \boldsymbol{w}_{g,m}, \boldsymbol{\Psi}^g_{e_m}(y_{\mathrm{pa}(m)}, y_m) \rangle \quad (6)$$

として定義した．$\boldsymbol{\Psi}^g$ には，Ûřîčář ら [14] と同じ 4 次元のベクトルを使用した．

3.3 損失関数

構造学習では，予測値 $\boldsymbol{y} \in \mathcal{Y}$ が正解値 $\boldsymbol{y}^* \in \mathcal{Y}$ と異なったときに，どれほどの損失があるか，任意に決めることができるので，応用に適合した損失を設定できる．本研究で用

いる損失関数 $\Delta(\cdot,\cdot)$ は Uřiĉář らのと基本的には同じであるが，より一般的な問題にもそのまま適用できるよう，標準偏差による正規化を行うよう変更を加えている．すなわち，次式で与えるものを用いる：

$$\Delta(\boldsymbol{y}^*,\boldsymbol{y}) := \frac{1}{M}\sum_{m=1}^{M}\sqrt{(\tilde{\boldsymbol{y}}_m-\tilde{\boldsymbol{y}}^*_m)D^{-2}_{\sigma^*_m}(\tilde{\boldsymbol{y}}_m-\tilde{\boldsymbol{y}}^*_m)} \quad (7)$$
$$= \frac{1}{M}\sum_{m=1}^{M}\sqrt{\sum_{h=1}^{2}\frac{(\tilde{y}_{h,m}-\tilde{y}^*_{h,m})^2}{\sigma^2_{h,m}(\boldsymbol{y}^*)}},$$

ただし，

$$\sigma_{h,m} := \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\tilde{y}^i_{h,m}-\mu_{h,m})^2}, \quad \mu_{h,m} := \frac{1}{N}\sum_{i=1}^{N}\tilde{y}^i_{h,m},$$

$$\tilde{y}^i_{h,m} := \frac{1}{S(\boldsymbol{y}^*)}(y^i_{h,m}-\bar{y}^*_h),$$
$$\tilde{\boldsymbol{y}}_m := \begin{bmatrix}\tilde{y}_{1,m}\\ \tilde{y}_{2,m}\end{bmatrix} = \frac{1}{S(\boldsymbol{y}^*)}\begin{bmatrix}(y_{1,m}-\bar{y}^*_1)\\ (y_{2,m}-\bar{y}^*_2)\end{bmatrix},$$
$$S(\boldsymbol{y}^*) := \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left\|\begin{matrix}y^*_{1,m}-\bar{y}^*_1\\ y^*_{2,m}-\bar{y}^*_2\end{matrix}\right\|^2},$$

$$\bar{y}^*_1 := \frac{1}{M}\sum_{m=1}^{M}y^*_{1,m} \qquad \bar{y}^*_2 := \frac{1}{M}\sum_{m=1}^{M}y^*_{2,m}.$$

$y_m=(y_{1,m},y_{2,m})$ はノード m の予測した状態の水平座標，および垂直座標，$y^*_m=(y^*_{1,m},y^*_{2,m})$ は正解の水平座標，および垂直座標を表す．$S(\cdot)$ はスケールを表し，各画像における形状の大きさによる違いを緩和する．また，$(y^i_{1,m},y^i_{2,m})$ は i 枚目の画像のノード m における，アノテーション情報を示す．

図 1 に，形態 $\boldsymbol{y}$ と損失の値との関係を例示する．図 1 において，$*$ 印は各ノード v_m の正解位置 y^*_m を表す．矩形は，与えた形態 $\boldsymbol{y}$ に対応する検出窓を表している．すべてのノードの状態が，真の位置と一致するならば，損失関数 Δ の値は 0 になる．図 1 左では，与えた形態 $\boldsymbol{y}$ が正解位置 $\boldsymbol{y}^*$ に近いため，損失の値が 0.381 と小さくなっているが，図 1 右のように，与えた形態 $\boldsymbol{y}$ が正解位置 $\boldsymbol{y}^*$ から離れるほど，損失の値は大きくなる．しかし，後述の実験では，現在用いることができるアノテーションの情報 [21] の精度が悪いため，どんなに精度よく顔のパーツを検出したとしても，アノテーション情報に基づく正解位置 $\boldsymbol{y}^*$ にぴったりあわせることは難しいのが現状である．

3.4 サロゲート損失とその勾配

関数 (5) , (6) の定義より，スコア関数 (4) は，

$$J(\boldsymbol{y};I,\boldsymbol{w}) := \langle \boldsymbol{w},\boldsymbol{\Psi}(I,\boldsymbol{y})\rangle \quad (8)$$

のようにも表せる．ただし，

$$\boldsymbol{w} := [\boldsymbol{w}^\top_{q,1},\ldots,\boldsymbol{w}^\top_{q,m},\boldsymbol{w}^\top_{g,2},\ldots,\boldsymbol{w}^\top_{g,m}]^\top,$$

及び，

$$\boldsymbol{\Psi}(I,\boldsymbol{y}) := [\boldsymbol{\Psi}^q_{v_1}(y_1)^\top,\ldots,\boldsymbol{\Psi}^q_{v_m}(y_m)^\top,$$
$$\boldsymbol{\Psi}^g_{e_2}(y_{\mathrm{pa}(2)},y_2)^\top,\ldots,\boldsymbol{\Psi}^g_{e_m}(y_{\mathrm{pa}(m)},y_m)^\top]^\top$$

である．よって，

$$\max_{\boldsymbol{y}\in\mathcal{Y}}\left(J(\boldsymbol{y};I,\boldsymbol{w})-J(\boldsymbol{y}^*;I,\boldsymbol{w})+\Delta\left(\boldsymbol{y}^*,\boldsymbol{y}\right)\right) = l_{\mathrm{hinge}}(\boldsymbol{w};I,\boldsymbol{y}^*)$$

および

$$\log\sum_{\boldsymbol{y}\in\mathcal{Y}}\exp\left(J(\boldsymbol{y};I,\boldsymbol{w})-J(\boldsymbol{y}^*;I,\boldsymbol{w})+\Delta\left(\boldsymbol{y}^*,\boldsymbol{y}\right)\right)$$
$$= l_{\mathrm{logi}}(\boldsymbol{w};I,\boldsymbol{y}^*)$$

を得，ヒンジ構造損失やロジスティック構造損失の形式に帰着できることが導かれた．

ヒンジ構造損失の勾配

ヒンジ構造損失やロジスティック構造損失を用いて学習を行うには，勾配もしくは劣勾配が必要になる．ヒンジ構造損失の劣勾配は次式で与えられる：

$$\nabla l_{\mathrm{hinge}}(\boldsymbol{w};I,\boldsymbol{y}^*) = \boldsymbol{\Psi}(\hat{\boldsymbol{y}})-\boldsymbol{\Psi}(\boldsymbol{y}^*)$$
$$\text{where } \hat{\boldsymbol{y}} := \operatorname*{argmax}_{\boldsymbol{y}\in\mathcal{Y}}\left(J(\boldsymbol{y};I,\boldsymbol{w})+\Delta\left(\boldsymbol{y}^*,\boldsymbol{y}\right)\right).$$

よって，勾配を高速に計算するには，このスコアと損失の和 $\left(J(\boldsymbol{y};I,\boldsymbol{w})+\Delta\left(\boldsymbol{y}^*,\boldsymbol{y}\right)\right)$ が最大になる形態 $\hat{\boldsymbol{y}}$ を効率的に計算しなくてはならない．幸い，DPM の場合，木構造で与えられているため，動的計画法を使って効率的に $\hat{\boldsymbol{y}}$ を計算できる [14][22].

ロジスティック構造損失の勾配

ロジスティック構造損失は，$\forall\boldsymbol{w}\in\mathbb{R}^d$ で微分可能である．ロジスティック構造損失の勾配は

$$\nabla l_{\mathrm{logi}}(\boldsymbol{w};I,\boldsymbol{y}^*) = -\boldsymbol{\Psi}(\boldsymbol{y}^*)$$
$$+\sum_{\boldsymbol{y}\in\mathcal{Y}}\boldsymbol{\Psi}(\hat{\boldsymbol{y}})\exp\left(J(\boldsymbol{y};I,\boldsymbol{w})+\Delta\left(\boldsymbol{y}^*,\boldsymbol{y}\right)-S(\boldsymbol{w},I,\boldsymbol{y}^*)\right) \quad (9)$$

と表される．ただし，

$$S(\boldsymbol{w},I,\boldsymbol{y}^*) := \log\sum_{\boldsymbol{y}\in\mathcal{Y}}\exp\left(J(\boldsymbol{y};I,\boldsymbol{w})+\Delta\left(\boldsymbol{y}^*,\boldsymbol{y}\right)\right)$$

とおいた．この計算は，一見膨大な計算量がかかるように見えるかもしれない．しかし，DPM は木構造で表現されているため，次の命題が成り立つ [23].

Proposition 3.1. ロジスティック構造損失の勾配 (9) は，$O(MH^2)$ で計算できる．ただし，$H := \max_m|\Sigma_m|$ である．

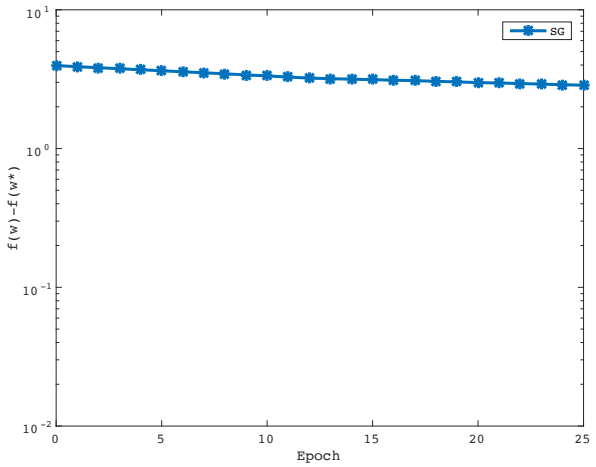

図 2　ヒンジ構造損失による目的関数 $f_{\mathrm{hinge}}(\boldsymbol{w})$ の値の変化．ヒンジ構造損失による目的関数では収束が遅いことが観測される．

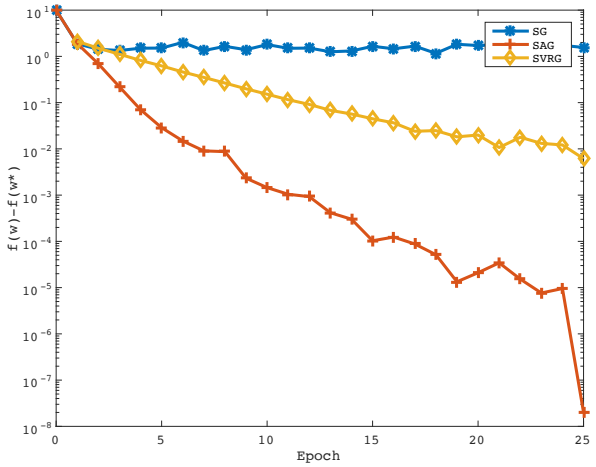

図 3　ロジスティック構造損失による目的関数 $f_{\mathrm{logi}}(\boldsymbol{w})$ の値の変化．ロジスティック構造損失による目的関数では，確率的勾配法の改良版 SAG 法や SVRG 法の仮定を満たしており，現実的なエポック数で最適解にほぼ収束していることが観測できる．

4. 実験

提案法に対して，以下の目的で数値実験を行った．

- 提案するロジスティック構造損失と SAG 法や SVRG 法を組み合わせると，最適解に現実的な計算量で収束するか検証する．
- ロジスティック構造学習により，顔パーツの検出の性能が，従来法より向上するか検証する．

本節では，それらの検証実験の結果を報告する．

データセットには，LFW [24] を用い，それに対するアノテーションには Uřičář ら [14] も用いている [21] を用いた．LFW からランダムに選択した 1000 枚の顔画像を実験データとして用いている．訓練データとテストデータは 7:3 として分割した．また，各ノードにおける状態集合は，各ノード 11 × 11 箇所から構成し，各ノードの検出窓の大きさは，顔中心を 80 × 80 画素，鼻を 40 × 40 画素，左右目尻・目頭・口角を 14 × 14 画素として設定した．確率的勾配法におけるステップサイズは候補 $10^{-6}, 10^{-5}, \ldots, 10^{1}$ の中より hindsight で選んだ．

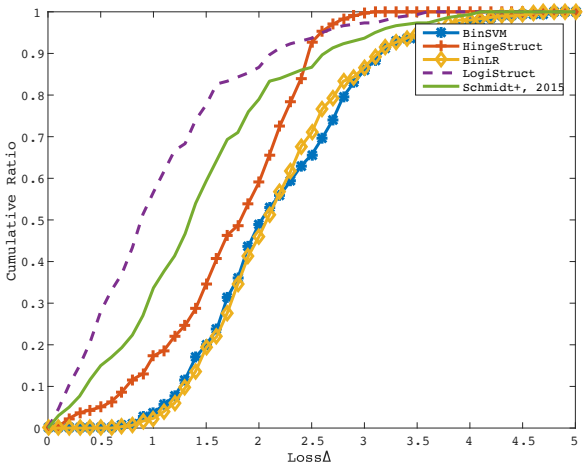

図 4　損失関数値の出現頻度．横軸は損失関数 Δ の値を表す．縦軸は損失がその値以下になった評価用画像の割合を表す．ロジスティック構造損失の導入によって検出性能が顕著に向上していることがわかる．

4.1 目的関数値の収束性の検証

ヒンジ構造学習に SGD 法を適用した場合と，ロジスティック構造学習に SGD 法，SAG 法，SVRG 法をそれぞれ適用した場合の，目的関数の収束性を検証した．$\lambda = 10^{-1}$ とした．目的関数 $f(\boldsymbol{w})$ の最適解 $\boldsymbol{w}^{\star}$ を LBFGS 法で，十分な反復数を回して収束した点とし，エポックごとに $f(\boldsymbol{w}) - f(\boldsymbol{w}^{\star})$ を観測した．SVRG 法は，SGD 法を 1 反復実行したときの解を初期値 $\boldsymbol{w}^{0}$ とするため，その他の勾配法とは初期値が異なるので，プロットの開始が 1 から始まっている．

図 2 にヒンジ構造学習の結果を示す．25 エポック経っても最適解に到達していないことがわかる．ロジスティック構造学習の結果は図 3 にプロットした．SGD 法と比べ，SVRG 法のほうが収束が早く，SAG 法は急速に最適解に収束していることがわかる．これにより，ロジスティック構造損失の導入によって，最適解に到達可能な DPM 学習ができるようになったことが実証された．

4.2 検出性能の検証

検出精度の検証を行うため，次の方法を比較した．

- **BinSVM**: 構造学習ではなく，検出結果が正解位置と誤差なく等しくなったときを陽性，それ以外を陰性とする 2 クラス分類問題とし，ヒンジ損失による SVM と比較する．学習の際に必要な最適化には，SGD 法を用いる．訓練用例題の負例には各ノードの状態を無作為

に選んだものから作成した．

- **BinLR**: BinSVM と同様，２クラス分類問題として学習するが，損失関数にはロジスティック損失を用いる．すなわち，ロジスティック回帰を行う．学習の際に必要な最適化には，SAG 法を用いる．
- **HingeStruct**: Uřičář ら [14] の方法．ヒンジ構造損失を用いて，SGD 法で学習を行う．
- **Schmidt+, 2015**: Schmidt ら [19] の方法．サロゲート損失に (2)　を用い，SAG 法で学習を行う．
- **LogiStruct**: 提案するロジスティック構造損失 (3) を用い，SAG 法で学習を行う．

検出性能は，評価用画像を使って検証した．Uřičář ら [25] に倣い，損失関数 Δ がある値以下になった評価用画像の割合を算出した．図 4 にその結果をプロットした．提案するロジスティック構造学習が損失関数値 2.5 までの損失関数値の出現率が最も高く，適切に顔パーツの学習がなされたデータが多かったことが確認できた．

5. 結論

本論文では，ロジスティック構造学習という新しい構造学習の枠組みを提案し，その枠組みが DPM に適用可能なことを示した．数値実験の結果，提案するロジスティック構造損失は最適解への収束が早いことが確認できた．また，ロジスティック構造学習によって得られた DPM のモデルパラメータは，顔パーツの検出において，従来の方法よりも顕著に高い検出精度を得ることができた．今後は顔パーツ検出だけではなく，姿勢推定や物体検出に対しても適用し，提案法の有効性を積極的に検証していきたい．

参考文献

[1] Crandall, D. J., Felzenszwalb, P. F. and Huttenlocher, D. P.: Spatial Priors for Part-Based Recognition Using Statistical Models, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA*, pp. 10–17 (2005).

[2] Felzenszwalb, P. F. and Huttenlocher, D. P.: Pictorial Structures for Object Recognition, *Int. J. Comput. Vision*, Vol. 61, No. 1, pp. 55–79 (2005).

[3] Felzenszwalb, P. F., Girshick, R. B. and Mcallester, D.: D.M.: Cascade Object Detection with Deformable Part Models, *In: Proc. CVPR.* (2010).

[4] Fischler, M. A. and Elschlager, R. A.: The Representation and Matching of Pictorial Structures, *IEEE Trans. Comput.*, Vol. 22, No. 1, pp. 67–92 (1973).

[5] Cootes, T. F. and Taylor, C. J.: Active Shape Models - Smart Snakes, *in Proceedings of the British Machine Vision Conference* (1992).

[6] Cootes, T. F., Taylor, C. J., Cooper, D. H. and Graham, J.: Active Shape Models, *Comput. Vis. Image Underst.*, Vol. 61, No. 1, pp. 38–59 (1995).

[7] Cootes, T. F., Edwards, G. J. and Taylor, C. J.: Active Appearance Models, *ECCV Transactions on Pattern Analysis and Machine Intelligence*, Springer, pp. 484–498 (1998).

[8] Cootes, T. F., Edwards, G. J. and Taylor, C. J.: Active Appearance Models, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 23, No. 6, pp. 681–685 (2001).

[9] Tsochantaridis, I., Joachims, T., Hofmann, T. and Altun, Y.: Large Margin Methods for Structured and Interdependent Output Variables, *Journal of Machine Learning Research*, Vol. 6, pp. 1453–1484 (2005).

[10] Joachims, T., Finley, T. and Yu, C. J.: Cutting-plane training of structural SVMs, *Machine Learning*, Vol. 77, No. 1, pp. 27–59 (2009).

[11] Bartlett, P. L., Jordan, M. I. and McAuliffe, J. D.: Convexity, classification, and risk bounds, *Journal of the American Statistical Association*, Vol. 101, No. 473, pp. 138–156 (2006).

[12] Hastie, T. J., Tibshirani, R. J. and Friedman, J. H.: *The elements of statistical learning : data mining, inference, and prediction*, Springer series in statistics, Springer (2009).

[13] Zhou, T., Tao, D. and Wu, X.: NESVM: A Fast Gradient Method for Support Vector Machines, *ICDM 2010, The 10th IEEE International Conference on Data Mining, Sydney, Australia, 14-17 December 2010*, pp. 679–688 (2010).

[14] Uřičář, M., Franc, V. and Hlavác, V.: Detector of Facial Landmarks Learned by the Structured Output SVM, *VISAPP 2012 - Proceedings of the International Conference on Computer Vision Theory and Applications, Volume 1, Rome, Italy, 24-26 February, 2012.*, pp. 547–556 (2012).

[15] Shalev-Shwartz, S., Singer, Y., Srebro, N. and Cotter, A.: Pegasos: primal estimated sub-gradient solver for SVM, *Math. Program.*, Vol. 127, No. 1, pp. 3–30 (2011).

[16] Bottou, L.: Large-Scale Machine Learning with Stochastic Gradient Descent, *Proceedings of the 19th International Conference on Computational Statistics (COMPSTAT'2010)* (Lechevallier, Y. and Saporta, G., eds.), Paris, France, Springer, pp. 177–187 (2010).

[17] Roux, N. L., Schmidt, M. and Bach, F. R.: A Stochastic Gradient Method with an Exponential Convergence _Rate for Finite Training Sets, *Advances in Neural Information Processing Systems 25* (Pereira, F., Burges, C., Bottou, L. and Weinberger, K., eds.), Curran Associates, Inc., pp. 2663–2671 (2012).

[18] Johnson, R. and Zhang, T.: Accelerating Stochastic Gradient Descent using Predictive Variance Reduction, *Advances in Neural Information Processing Systems 26: Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pp. 315–323 (2013).

[19] Schmidt, M., Babanezhad, R., Ahmed, M. O., Defazio, A., Clifton, A. and Sarkar, A.: Non-Uniform Stochastic Average Gradient Method for Training Conditional Random Fields, *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2015, San Diego, California, USA, May 9-12, 2015* (2015).

[20] Dalal, N. and Triggs, B.: Histograms of Oriented Gradients for Human Detection, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA*, pp. 886–893 (2005).

[21] : flandmark, `http://cmp.felk.cvut.cz/~uricamic/flandmark/`.

[22] Felzenszwalb, P. F. and Huttenlocher, D. P.: Efficient Matching of Pictorial Structures, *PROC. IEEE COM-*

PUTER VISION AND PATTERN RECOGNITION CONF., pp. 66–73 (2000).

[23] Kschischang, F. R., Frey, B. J. and Loeliger, H.: Factor graphs and the sum-product algorithm, *IEEE Transactions on Information Theory*, Vol. 47, No. 2, pp. 498–519 (2001).

[24] : Labeled Faces in the Wild, `http://vis-www.cs.umass.edu/lfw/`.

[25] Uřičář, M., Franc, V., Thomas, D., Akihiro, S. and Hlaváč, V.: Real-time Multi-view Facial Landmark Detector Learned by the Structured Output SVM, *BWILD'15: 11th IEEE International Conference on Automatic Face and Gestu re Recognition Workshops, Biometrics in the Wild* (Bir, B., Abdenour, H., Qiang, J., Mark, N. and Vitomir, Š., eds.), New York, US, IEEE Computer Society (2015).