

1. はじめに

深層学習において複数タスクを同時に学習する手法として、Task-wise early stopping が提案されている [1]。この手法は、1つのタスクの識別精度を向上させるために、他のタスクの学習を途中で停止するアプローチである。このとき、特定タスク以外の識別精度は向上しない。そこで、本研究では複数タスクの精度向上を目的とする逐次更新型重み付き Heterogeneous Learning を提案する。評価実験では、顔属性認識を対象として提案手法の有効性を示す。

2. Task-wise early stopping

複数のタスクを同時に学習する際、1つのタスクを基準タスク、他のタスクをサブタスクとする。Task-wise early stopping は式 (1) を満たしたサブタスクの学習を停止する。

$$\frac{k \cdot \text{med}_{j=t-k}^t E_{tr}^a(j)}{\sum_{j=t-k}^t E_{tr}^a(j) - k \cdot \text{med}_{j=t-k}^t E_{tr}^a(j)} \cdot \frac{E_{val}^a(t) - \min_{j=1..t} E_{tr}^a(j)}{\lambda^a \cdot \min_{j=1..t} E_{tr}^a(j)} > \epsilon \quad (1)$$

これにより、複数のサブタスクの過学習による基準タスクの性能の低下を抑えている。ここで、med は中央値を算出する関数、 E_{val}^a 、 E_{tr}^a はそれぞれ検証誤差と学習誤差の値、 ϵ はしきい値を示す。この手法は、基準タスクの識別精度が向上する一方、サブタスクは図 1 のように学習を途中で停止するため、識別精度が向上しないという問題がある。

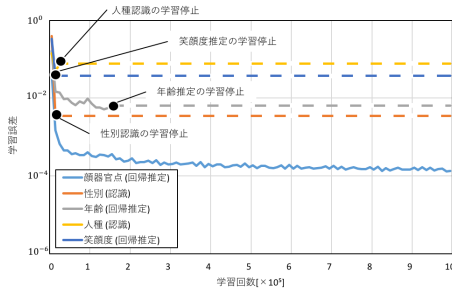


図 1：Task-wise early stopping の学習誤差の推移

3. 提案手法

Heterogeneous Learning は、複数の回帰推定タスクと認識タスクを 1つの DCNN で同時に学習する。これにより、学習時間を削減し、複数のタスクを識別することが可能となる。しかし、複数タスクを同時に学習することから各タスクの学習誤差にばらつきが発生するため、単一の DCNN に比べ識別精度は低下する。本研究では逐次更新型 Heterogeneous Learning を提案する。提案手法では、各サブタスク t に重み $w_{t,e}$ を与える。重み $w_{t,e}$ は、学習過程で各タスクの学習誤差から算出し、1 epoch 学習したときに更新する。重み $w_{t,e}$ の算出過程を以下に示す。まず、0 epoch から n epoch までの学習誤差から安定して学習しているかを基準に、式 (2) のようにタスク毎に安定度 $N_{t,e}$ を算出する。ここで、 e は更新回数、 $\mu_{t,e}$ 、 $\sigma_{t,e}$ はそれぞれタスク t における学習誤差の平均と標準偏差を示す。

$$N_{t,e} = \mu_{t,e} + 3\sigma_{t,e} \quad (2)$$

算出された各タスクの安定度 $N_{t,e}$ から、式 (3) を満たすタスク t を基準タスク $T_{main,e}$ とし、他のタスクをサブタスクとする。

$$T_{main,e} = \arg \min_t N_{t,e} \quad (3)$$

基準タスク $T_{main,e}$ に与える重みと安定度をそれぞれ $w_{main,e}$ 、 $N_{main,e}$ とし、式 (4) により各サブタスクに与える重み $w_{t,e}$ を算出する。ここで、 $w_{t,0} = 1.0$ とする。

$$w_{t,e} = w_{main,e} \cdot \frac{N_{main,e}}{N_{t,e}} \quad (4)$$

算出された重み $w_{t,e}$ は、式 (5) のように学習誤差関数 E に付与する。 M はバッチサイズ、 T はサブタスクの数、 L は各タスクの教師信号、 O は各タスクの出力値である。

$$E = \frac{1}{M} \sum_{m=1}^M \left(w_{main,e} \|L_{main,m} - O_{main,m}\|^2 + \sum_{t \neq main} w_{t,e} \|L_{t,m} - O_{t,m}\|^2 \right) \quad (5)$$

各サブタスクに重み $w_{t,e}$ を与えることで、基準タスク $T_{main,e}$ への影響を軽減しながらサブタスクの学習を同時に行うことが可能となる。

4. 評価実験

評価実験では、Task-wise early stopping と提案手法を比較する。データセットには、Web から収集した顔画像を使用する。学習サンプルは 42,663 枚、評価サンプルは 11,000 枚とし、 100×100 画素のグレースケール画像を入力する。本研究では、回帰タスクに顔器官点検出と年齢推定、笑顔度推定、認識タスクに性別認識と人種認識を行う。本研究に用いるネットワーク構造は図 2 のような畳み込み層が 3 層と全結合層が 1 層とする。

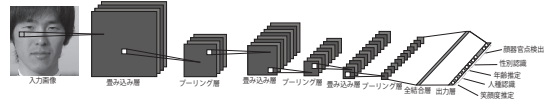


図 2：本研究に用いるネットワーク構造

4.1. 学習誤差の比較

提案手法の学習誤差の推移を図 3 に示す。図 3 から、誤差関数に重み付けをすることで、サブタスクの学習を停止することなく学習可能となった。

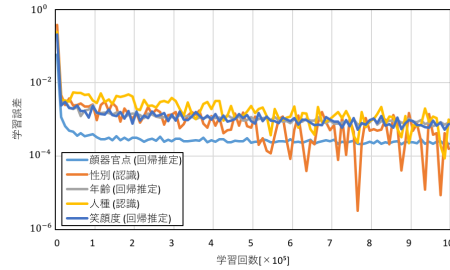


図 3：提案手法の学習誤差の推移

4.2. 識別精度の比較

従来法と提案手法の識別精度の比較を図 4 に示す。図 4 より、従来法と比較すると、提案手法は全体精度が約 10% 向上した。タスク毎に比較すると、顔器官点検出において識別率が約 4% 低下した。しかし、他のタスクにおいては識別精度が大幅に向上した。特に、笑顔度推定は識別率が約 30% 向上している。

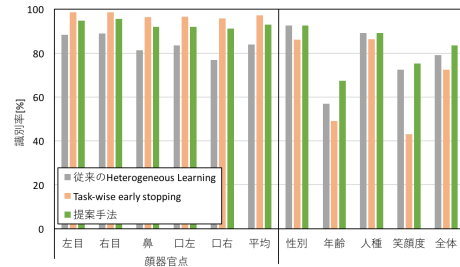


図 4：従来法と提案手法の識別精度の比較

5. おわりに

本研究では、逐次更新型重み付き Heterogeneous Learning を提案した。提案手法を用いることで、複数タスクを高精度に識別することが可能となった。今後は、顔属性認識のみならず歩行者の属性認識等への応用を検討する。

参考文献

[1] Z.Zhang, et al. "Facial landmark detection by deep multi-task learning." ECCV, 2014.