

1. はじめに

半教師あり学習は、ラベルありデータとラベルなしデータを用いた学習法である。半教師あり学習の分野で手法を大別すると、一致性正則化法、擬似ラベリング法、それらを組み合わせた手法がある。これらの学習手法は、人が設計した構成であり、最適な学習法であるとは限らない。本研究では、従来法を内包しつつ、新しく高精度な半教師あり学習法の自動設計を目的とする。従来の相互学習を内包する知識転移グラフに半教師あり学習法の主要な構成要素を導入し、グラフ構造を自動設計することで、従来法より高精度な半教師あり学習法を獲得する。

2. 半教師あり学習

半教師あり学習において一致性正則化法は、摂動とラベルなしデータを活用した正則化効果によりモデル性能を改善する。一致性正則化法の代表例として Π -model [1] がある。 Π -model の派生手法として、代表的な Mean Teacher [2] の他に、擬似ラベリング法を導入した FixMatch [3] がある。FixMatch は、ラベルなしデータに対して弱変換時に推測値をもとに生成した擬似ラベルと強変換時の推測値が一致するように学習する。そのため FixMatch は、従来法で個別に行うような、一致性正則化と擬似ラベリングを上手く合わせた正則化により、高い精度を発揮する。

3. 提案手法

FixMatch は一致性正則化と擬似ラベリングの組み合わせが効果的であることを示した。本研究では、半教師あり学習の従来法を内包しつつグラフで統一の形式に表現し、グラフの構成要素を組み合わせて新たな半教師あり学習法を自動設計する手法を提案する。半教師あり深層学習の従来法として、一致性正則化法、擬似ラベリング法、複合的な学習法をノードとエッジを用いてグラフで表現する。ノードはネットワークモデルを表し、エッジは誤差計算と誤差逆伝播による知識転移の方向を表す。

3.1. 一致性正則化法のグラフ表現

一致性正則化の代表的な学習法として Π -model と Mean Teacher の学習の流れをグラフに表現する。 Π -model のグラフ表現を図 1(a) に示す。

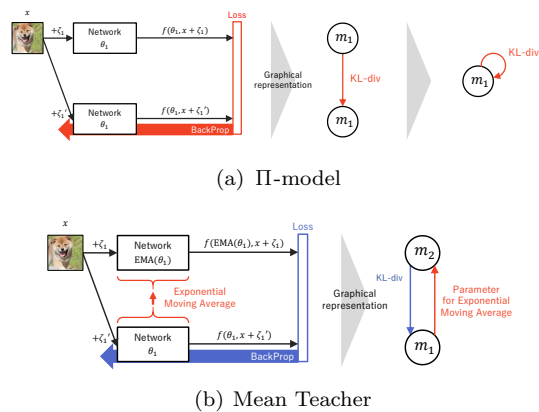


図 1：一致性正則化法のグラフ表現

$f(\theta, x)$ はパラメータ θ を持つネットワークに画像 x を入力した時のクラス確率を表す。1 枚の画像に対して異なる摂動を付与した 2 枚の画像 $x + \zeta_1, x + \zeta'_1$ をネットワークに入力し、2 つのクラス確率の差を誤差とする一致性損失を最小化するように学習する。一致性損失 $\mathcal{L}_{con}$ を式 (1) に示す。

$$\mathcal{L}_{con}(x) = \mathbb{E}_x \mathcal{R}(f(\theta_1, x + \zeta_1), f(\theta_1, x + \zeta'_1)). \quad (1)$$

ここで、 $\mathcal{R}(\cdot, \cdot)$ は 2 つの確率分布間の KL-divergence による誤差を表す。 Π -model は、誤差の算出に用いるネット

ワークと更新するネットワークが同一である。そのため、パラメータ θ_1 をもつネットワークモデルをノード m_1 として、 m_1 から m_1 への知識転移としてグラフに表現できる。一致性損失は始点と終点と同じノードにあるエッジで表現する。

Mean Teacher のグラフ表現を図 1(b) に示す。Mean Teacher は、学習を補助する役割を持つ指数移動平均 (EMA) モデルを使用し、EMA モデルのクラス確率 $f(\text{EMA}(\theta_1), x + \zeta_1)$ に $f(\theta_1, x + \zeta'_1)$ が類似するように学習する。EMA モデルのパラメータ $\text{EMA}(\theta_1)$ の更新方法を式 (2) に示す。

$$\text{EMA}(\theta_{1,t}) = \alpha \text{EMA}(\theta_{1,t-1}) + (1 - \alpha) \theta_{1,t} \quad (2)$$

ここで、 α はハイパーパラメータ、 t は学習ステップである。Mean Teacher は、指数移動平均モデルのパラメータ $\text{EMA}(\theta_1)$ は θ_1 の EMA で更新する。そのため、EMA モデルを表すノードを用いて、EMA モデルと m_1 による双方向の知識転移としてグラフに表現する。

3.2. 擬似ラベリング法のグラフ表現

擬似ラベリング法の代表的な手法として Pseudo-Label [4] のグラフ表現を図 2 に示す。Pseudo-Label は、 $f(\theta_1, x + \zeta_1)$

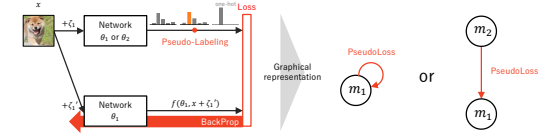


図 2：Pseudo-Label のグラフ表現

の推測クラスを 1 とする one-hot ベクトルを生成し、生成した one-hot ベクトルを入力画像の擬似的なラベル y' として、 y' と $f(\theta_1, x + \zeta'_1)$ が類似するように学習する。損失関数 $\mathcal{L}_{pse}$ を式 (3) に示す。

$$\mathcal{L}_{pse}(x) = -\mathbb{E}_x y' \log f(\theta_1, x + \zeta'_1) \quad (3)$$

Pseudo-Label の損失関数を PseudoLoss としてエッジに表現し、 m_1 から m_1 への知識転移としてグラフに表現する。また、疑似ラベルを生成するネットワークと推測を行うネットワークが異なる場合、始点と終点のノードが異なるエッジを用いて容易に拡張できる。

3.3. 複合的な学習法のグラフ表現

複合的な学習法として FixMatch のグラフ表現を図 3 に示す。擬似ラベルは、ネットワーク性能が不十分な学習序

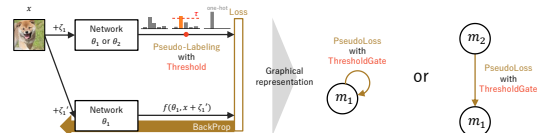


図 3：FixMatch のグラフ表現

盤において誤識別により学習を阻害する可能性がある。そのため FixMatch は、擬似ラベルとクラス確率が一致するように学習する際、クラス確率が閾値 τ を越える場合のみ擬似ラベルを生成することで損失値を制御する。グラフでは入力する損失値を a として、擬似ラベリングに対する損失値の制御をゲート関数 $G^{Threshold}(a)$ で表し、計算方法を式 (4) に示す。

$$G^{Threshold}(a) = \mathbb{1}(\max f(x) > \tau) \cdot a \quad (4)$$

FixMatch は、PseudoLoss で得た損失を Threshold Gate で制御するエッジを用いてグラフに表現する。また、摂動 ζ'_1 を RandAugment などの強い摂動とすることで、さらに正則化効果を高めることができる。

3.4. グラフ最適化による半教師あり学習法の自動設計

従来の学習法を内包しつつ、図4のようにノードとエッジを組み合わせて学習法を自動設計する。知識転移を制御するために複数種のゲート関数を導入することで、多様な共同学習を実現することができる。ゲート関数は、誤差値を制御して対象ノードを効率的に学習させる。

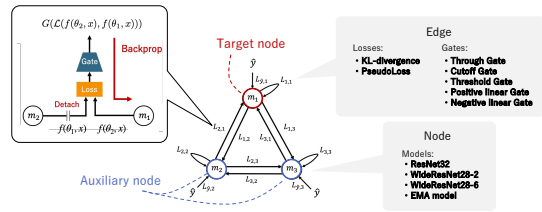


図4：各エッジの処理の流れと設計空間の例

グラフ構造の自動設計には、知識転移グラフ [5] の最適化法を用いる。最適化したい評価対象ノードを除くノードを補助ノードとして、評価対象ノードの精度が最大化するように補助ノードとエッジをハイパーパラメータサーチにより設計する。グラフ要素の組み合わせ数は膨大になることが多く、並列かつ非同期的に実行可能な最適化法 ASHA を用いる。ASHA は、ランダムサーチによって提案されたパラメータを用いて試行し、学習途中の評価値が過去に試行したパラメータと比較した時に既定の割合を下回る順位であれば枝刈りを行う。

4. 評価実験

自動設計した新しい半教師あり学習法を評価する。データセットには CIFAR-100 を用い、ラベルありデータ数を 2,000~10,000、試行回数を 15,000、ノード数は 2 と 3、評価対象ノードは ResNet32 とする。設計空間は、モデルが ResNet18, WideResNet28-2 と 28-6, EMA モデルの 4 種、損失計算が KL-divergence, PseudoLoss の 2 種、ゲート関数が Through, Cutoff, Threshold, Positive/Negative linear の 5 種である。Through はノード間の接続, Cutoff は切断, Positive linear と Negative linear は動的な接続を表す。Positive linear は学習が進むにつれて重みが大きくなり学習序盤は Cutoff と同様に切断する役割をもつ。それに対して、Negative linear は学習終盤に切断する役割をもつ。

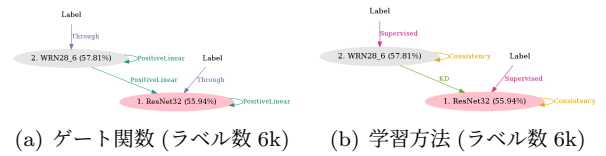
4.1. 評価結果

ノード数とラベルありデータ数のそれぞれ学習条件に対して個別に学習法を自動設計し、自動設計した学習法と従来法の認識精度の比較を表1に示す。

表1：自動設計した学習法と従来法の精度比較 [%]

	Number of labeled data				
	2,000	4,000	6,000	8,000	10,000
Supervised	20.85	32.35	41.56	50.38	53.61
Pseudo-Label	29.01	40.59	47.79	54.51	56.67
II-model	29.54	41.98	50.69	53.44	55.98
Mean Teacher	31.28	43.10	48.85	49.98	54.52
FixMatch	29.31	41.87	47.33	51.28	54.03
Ours(2 nodes)	42.04	51.92	55.94	60.65	63.08
Ours(3 nodes)	42.55	54.60	57.12	62.49	63.81

Supervised の行はラベルありデータのみで ResNet32 で教師あり学習した場合の認識精度である。表1より、ラベルありデータ数がいずれの場合においても自動設計した学習法が従来法に比べて高精度であり、ノード数が2から3になると更に精度が向上する。これは、自動設計した学習法は新しく高精度な学習構造を持ち、ノード数が増えて更に多様なグラフ構造から自動設計することの有効性を表す。また、従来法に注目すると、ラベルありデータ数が2,000~6,000 と少ない場合は一貫性正則化法の II-model や Mean Teacher の精度が高く、ラベルありデータ数が8,000, 10,000 と多い場合は擬似ラベリング法の Pseudo-Label の精度が高い。そのため、ラベルありデータ数によって適切な学習法が異なることがわかる。



(a) ゲート関数 (ラベル数 6k)

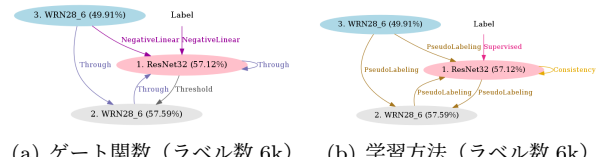
(b) 学習方法 (ラベル数 6k)



(c) ゲート関数 (ラベル数 10k)

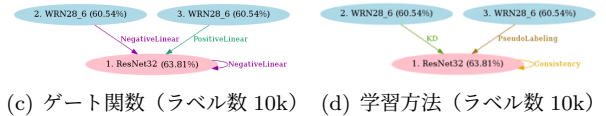
(d) 学習方法 (ラベル数 10k)

図5：ノード数2における自動設計した学習法のグラフ構造



(a) ゲート関数 (ラベル数 6k)

(b) 学習方法 (ラベル数 6k)



(c) ゲート関数 (ラベル数 10k)

(d) 学習方法 (ラベル数 10k)

図6：ノード数3における自動設計した学習法のグラフ構造

4.2. 自動設計した半教師あり学習のグラフ構造

自動設計したグラフの代表例として、ラベルありデータ数が少ない6,000 と多い10,000 の場合に、CIFAR-100 を用いて自動設計した学習法のグラフ構造を図5, 6に示す。エッジがゲート関数と損失関数という2つの要素を表すことから、ゲート関数と学習方法の2つの観点で表記する。赤のノードは評価対象、青と灰色のノードは補助ノードであり、青は教師あり学習による学習済みモデルを表す。

学習経過に注目すると、図5よりノード数2の場合、ラベルありデータ数が少ない場合は学習序盤に教師あり学習、多い場合は学習終盤にラベル情報を持つ学習済みモデルから擬似ラベリングをすることから、半教師あり学習においては学習序盤と終盤で異なる学習法が効果的である。また、図6よりノード数3の場合においても、ラベルありデータ数が少ない場合は Negative linear Gate により学習序盤のラベル情報活用が強調され、ラベルありデータ数が多い場合は Positive linear Gate と擬似ラベリングにより学習終盤のラベル情報活用が強調される。したがって、ラベルありデータ数が少ない場合は学習序盤、多い場合は終盤にラベル情報を活用する学習法が効果的であることがわかる。

5. おわりに

本論文では、従来の半教師あり学習法をグラフで統一的に表現し、自動設計により新しく高精度な半教師あり学習法を獲得した。今後は、ベイズ最適化の導入、自己教師あり学習への拡張について検討する。

参考文献

- [1] S. Laine, *et al.*, “Temporal Ensembling for Semi-Supervised Learning”, ICLR, 2017.
- [2] A. Tarvainen, *et al.*, “Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results”, NeurIPS, 2017.
- [3] K. Sohn, *et al.*, “FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence”, NeurIPS, 2020.
- [4] D. Lee, *et al.*, “Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks”, ICML, 2013.
- [5] S. Minami, *et al.*, “Knowledge Transfer Graph For Deep Collaborative Learning”, ACCV, 2020.

研究業績

- [1] 村本 等, “Refined Consistency による知識蒸留を用いた半教師あり学習”, 人工知能学会全国大会, 2021.
(他1件)