

1. はじめに

知識転移グラフ [1] は、知識蒸留や相互学習を含む多様な学習手段を Gate 関数を用いたグラフとして表現し、複数モデルを用いた共同学習法である。Gate 関数はハイパーパラメータ探索により、最適な組合せとなるグラフを探索する。4 種類の Gate 関数により多様な知識転移を表現しているが、学習過程に伴う変化が少ないことから、時間軸方向に対する知識転移は限定的である。そこで本研究では、Gate 関数の時間軸方向に対する表現力向上を目的として、Positive-Gamma Gate, Negative-Gamma Gate を提案する。これらの Gate 関数は 0.01 から 100 までの連続値で表現される γ 値を持つことで時間軸方向への重み付け戦略を表現することが可能となる。これにより、従来の 4 種類であった Gate 関数が 2 種類に減少し、ハイパーパラメータ変数が削減され、ベイズ最適化による探索コストの削減が期待できる。

2. 知識転移グラフ

知識転移グラフは、モデルを表現するノードとモデル間の知識転移を表現するエッジで構成され、各エッジに適応する Gate 関数を「Through Gate, Cutoff Gate, Correct Gate, Linear Gate」の 4 種類から選択することでモデル間の知識転移を制御する。Gate 関数の組み合わせをハイパーパラメータとして自動探索することで、人手で設計した手法より効果的な知識転移グラフを発見することができる。これにより 1 つのモデルで学習する場合よりも高い精度を達成している。

3. 提案手法

Gate 関数の時間軸方向への重み付け戦略により、共同学習の精度向上を目的とし、 γ 値を導入した「Positive-Gamma Gate (PG Gate) と Negative-Gamma Gate (NG Gate)」を提案する。PG Gate を式 (1)、NG Gate を式 (2) に示す。

$$G_{s,t}^{PG}(a) = \left(\frac{k}{k_{end}} \right)^{\frac{1}{\gamma}} (a) \quad (1)$$

$$G_{s,t}^{NG}(a) = \left(\frac{k_{end} - k}{k_{end}} \right)^{\frac{1}{\gamma}} (a) \quad (2)$$

2 つの Gate 関数のグラフを図 1 に示す。式 (1) は、学習初期から学習後半にかけて損失に乗算する重みを増加させる。一方で式 (2) は学習初期から学習後半にかけて重みを減少させる。提案する Gate 関数は γ 値により、線形な関数では表現不可能な多様な重み付け戦略を表現する。これらの Gate 関数を導入した知識転移グラフを多変数ベイズ最適化で探索することで時間軸方向へ最適知識転移グラフを獲得することが期待できる。

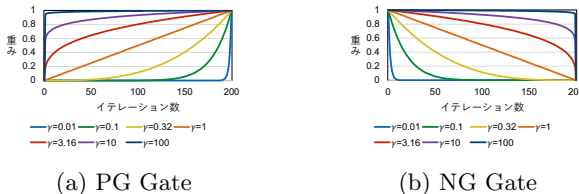


図 1: PG Gate と NG Gate のグラフ

4. 評価実験

PG Gate と NG Gate を知識転移グラフに導入し、その後ベイズ最適化を行って精度を評価する。ベイズ最適化による探索回数は 500 回とする。データセットには CIFAR-100、評価対象のノードに ResNet32、補助ノードには WRN-28 を使用する。

4.1 精度比較

学習結果を表 1 に示す。従来の知識転移グラフ (DCL) の精度 72.88% に対して、提案手法では 73.34% と 0.46pt 精度が向上した。

表 1: 教師あり学習における精度比較

手法	Vanilla	DCL	Ours
Acc[%]	70.71	72.88	73.34
Gate の種類	0	4	2
評価ノード	ResNet32	ResNet32	ResNet32
補助ノード	-	WRN-28	WRN-28
探索回数	1	1500	500

図 2 に、ベイズ最適化で獲得された知識転移グラフを示す。獲得された知識転移グラフの Gate 関数から、時間軸方向への重み付けが戦略が最適化されていることが分かる。最適化されたグラフから学習前半は教師ラベルから学習し、学習後半にかけて徐々に相互学習に切り替えていく学習方法が適していたことが分かる。また、探索回数は DCL の 1,500 回から 500 回に削減することができた。

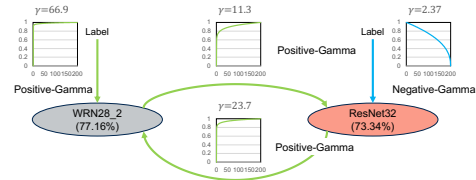


図 2: 最適化済みの知識転移グラフ

4.2 自己教師あり学習への適用

自己教師あり学習である SimCLR, SimSiam, MoCo, SwAV, BYOL, BarlowTwins, DINO の 7 種の学習手法をノード、提案した Gate 関数をエッジとし、これらの組合せをベイズ最適化で探索する。データセットは CIFAR-10、評価対象モデルは ResNet20、補助ノードは ResNet32 を使用する。知識転移には DoGo[2] で提案されたサンプル間の類似度と DisCO[3] で提案された同サンプル間の特徴量の差異を使用する。学習結果を表 2 に示す。従来の SimCLR の KNN 精度 64.16% と比較して、提案手法は 66.80% と 2.64pt 精度が向上した。

表 2: 自己教師あり学習における精度比較

手法	SSL1	SSL2	KNN Acc[%]
Vanilla	SimCLR	-	64.16
DML	SimCLR	SimCLR	64.24
Ours	SimCLR	BarlowTwins	66.80

5. おわりに

本研究では、時間軸方向への重み付け戦略が多様な深層共同学習を実現するために、PG Gate と NG Gate を提案した。提案手法では、学習経過に伴う損失に乘算する重みの変化により、従来手法に比べて、0.46pt 精度が向上した。また自己教師あり学習においても効果を確認した。今後は、ViT をベースとした、基盤モデルへの活用に取り組む予定である。

参考文献

- [1] S. Minami, *et al.*, “Knowledge Transfer Graph for Deep Collaborative Learning”, In ACCV, 2020.
- [2] P. Bhat, *et al.*, “Distill on the Go: Online knowledge distillation in self-supervised learning”, In CVPR, 2021.
- [3] Yuting Gao, *et al.*, “DisCo: Remedy Self-supervised Learning on Lightweight Models with Distilled Contrastive Learning”, In ECCV, 2022.