

1. はじめに

IoT の普及により、オンラインで蓄積した膨大なデータを深層学習モデルの学習に利用することが期待されている。一方、全てのデータを保持することはハードウェアなどの問題から困難である。この問題に対して、学習に利用したデータを破棄しながら継続的に学習を行う継続学習がある。ラベル無しデータを利用する継続学習法である Lifelong Unsupervised Mixup (LUMP) [1] は、破滅的忘却を抑制するためにデータ拡張法の mixup と過去の学習に利用したデータの一部を保存するリプレイバッファを活用する。リプレイバッファに保存するデータは、ランダムで選択されるため、データに偏りが生じ、性能が劣化することがある。そこで本研究では、k-means 法を用いてデータの分布を考慮したデータ選択を行うことで破滅的忘却を抑制する。

2. Lifelong Unsupervised Mixup

継続学習は、一定の間隔で学習に利用したデータを破棄し、新しいデータで学習を行う。学習過程において学習用データが変化するため、新しいタスクのデータに適合し、以前に学習したタスクを解くことができなくなる破滅的忘却が生じる。教師なし継続学習法の LUMP は、現在のタスクのデータとリプレイバッファに保存されている過去に学習したタスクのデータを mixup により合成して学習に利用する。これにより、破滅的忘却を抑制する。

3. 提案手法

LUMP は、リプレイバッファに保存するデータをランダムに選択しているため、データ分布を考慮した学習が難しい。そこで本研究では、LUMP にクラスタリング手法の k-means 法を導入して、データ分布を考慮したデータ選択を行う。提案手法の流れを図 1 に示す。

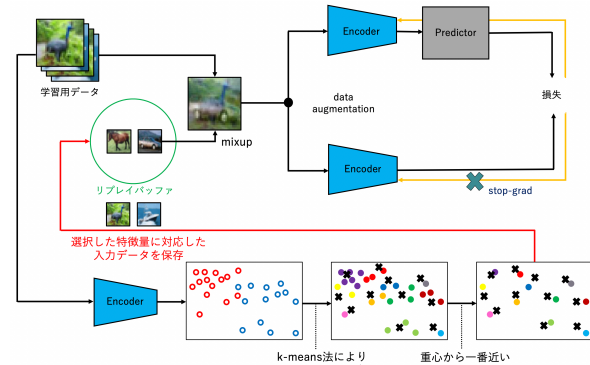


図 1：提案手法の流れ

提案手法では、k-means 法により学習用データの分布を複数のクラスタに分割し、各クラスタから 1 つずつデータを選択することになるため、元のデータ分布を保持したデータ選択を行う。以下に提案手法の流れを示す。

Step1. 学習

1 タスク目は学習用データ、2 タスク目以降は学習用データとリプレイバッファのデータを mixup により合成したデータを入力データとし、自己教師あり学習を行う。ここでは、入力データに異なるデータ拡張を適用してそれぞれをエンコーダに入力する。そして、片方のエンコーダの出力とプレディクタの出力のコサイン類似度を損失として求め、片方のエンコーダにのみ勾配を逆伝播し、エンコーダを更新する。

Step2. k-means 法によるデータ選択

学習用データをエンコーダに入力して特徴量を獲得し、k-means 法によりクラスタリングする。k-means 法のクラスタ数は、リプレイバッファに保存するデータ数とする。各クラスタの重心から一番近いデータを選択し、リプレイ

バッファに保存する。2 タスク目以降でリプレイバッファにデータを保存する時、これまで学習したタスクのデータ数が均等になるように、リプレイバッファのデータを選択したデータで置き換える。

4. 評価実験

提案手法と従来手法 LUMP において、クラス分類問題の分類精度と忘却率を比較する。

4.1. 実験概要

1 タスクずつ継続学習を行い、タスクごとの分類精度を評価する。データセットは 2 クラスずつ 5 個のタスクに分割した Split CIFAR-10、ネットワークは ResNet-18、自己教師あり学習法は SimSiam、学習回数は 200 エポック、バッチサイズは 256、リプレイバッファのサイズは 256 である。評価の際には、k-NN 法を用いてクラス分類を行う。k-NN 法は、テストデータから近い k 個の学習用データに付与された正解ラベルの中で最も多いクラスを推論結果とする。

4.2. 評価指標

評価指標として、平均精度と平均忘却率を用いる。平均精度は、すべてのタスクを学習し終えたモデルを利用した時の各タスクの精度の平均値である。また、平均忘却率は、各タスクの最高精度となった精度とすべてのタスクを学習し終えたモデルの精度の差の平均値である。

4.3. 評価結果

LUMP の評価結果を表 1、提案手法を表 2 に示す。表 1、表 2 から、提案手法は LUMP と比較して最終タスクまで学習した時の平均精度を 0.98pt 向上し、平均忘却率を 5.67pt 抑えることができた。以上より、k-means 法を利用してデータ選択したリプレイバッファを用いて学習することは、破滅的忘却の抑制に有効であるといえる。

表 1：LUMP の精度 [%] と忘却率 [pt]

		評価タスク					平均
		1	2	3	4	5	
学 習 タ ス ク	1	96.90	-	-	-	-	96.90
	2	88.50	89.80	-	-	-	89.15
	3	88.25	83.99	95.20	-	-	89.14
	4	87.95	80.25	89.14	98.80	-	89.04
	5	92.85	79.00	86.90	92.65	98.15	89.91
忘却率		4.05	10.80	8.30	6.15	-	7.33

表 2：提案手法の精度 [%] と忘却率 [pt]

		評価タスク					平均
		1	2	3	4	5	
学 習 タ ス ク	1	96.65	-	-	-	-	96.65
	2	95.50	83.20	-	-	-	89.35
	3	95.15	80.25	86.15	-	-	87.18
	4	92.20	80.50	88.45	96.90	-	89.51
	5	94.90	81.70	87.60	94.35	95.90	90.89
忘却率		1.75	1.50	0.85	2.55	-	1.66

5. おわりに

本研究では、k-means 法を用いてデータ選択したリプレイバッファによる継続学習法を提案した。評価実験により、提案手法が LUMP と比べ高精度かつ低い忘却率となり、破滅的忘却をより抑制することを示した。今後は、勾配空間に k-means 法を適用したデータ選択を導入した継続学習を行う予定である。

参考文献

- [1] D. Madaan, *et al.*, “Representational Continuity for Unsupervised Continual Learning”, ICLR, 2022.