Bayesian based Long-Tailed Continual Learning

Hao Dai
University of Southampton
h.dai@soton.ac.uk

Chong Tang
University of Southampton
chong.tang@soton.ac.uk

Jagmohan Chuahan
University College London
jagmohan.chauhan@ucl.ac.uk

I. Introduction

Continual learning enables machine learning systems to adapt to evolving data distributions, with class-incremental learning addressing the challenge of sequentially acquiring new classes while preserving performance on previous ones; however, real-world deployment is hindered by non-stationary data streams and long-tailed distributions, where tail classes suffer from catastrophic forgetting and bias amplification, particularly when relying on traditional methods that struggle to balance representation learning across imbalanced classes. Recent advances in pre-trained models and generative classifiers show promise for mitigating these issues, though challenges remain with extremely sparse tail-class samples limiting the accuracy of distribution estimation.

II. Variational Inference-based Continual Learning Classification

To address inter-task bias caused by class imbalance, we use a generative classifier modeling class-conditional distributions to avoid discriminative bias, employing Gaussian distributions with Mahalanobis similarity for improved high-dimensional discrimination. To stabilize estimation for sparse tail classes, we apply variational inference and optimize the ELBO with a conjugate prior, enabling robust distribution approximation even with limited samples.

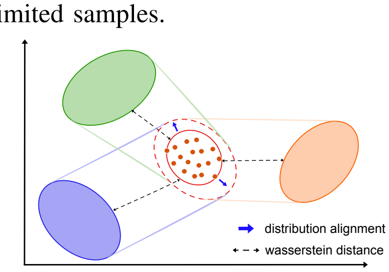

Fig. 1. The trilateration method for reconstructing distributions.

Variational inference provides robust posterior estimates but struggles with distributional shifts caused by long-tailed data, leading to biased or overconfident results for tail-class samples. To address this, we propose a redistribution mechanism that leverages geometric relationships between classes in the embedding space. We compute the 2-Wasserstein (W_2) distance between pairs of Gaussian distributions to identify semantically similar classes and use K-nearest neighbors based on minimal distances to reconstruct tail-class distributions via geometric interpolation. Neighbor weights, derived from W_2 distances and an adaptive weighting factor α, adjust the initial VI estimates by balancing fusion between the VI estimate and reconstructed distribution. Our method, ViRN, combines variational inference with trilateration-based redistribution to enhance robustness in long-tailed distribution estimation. By aligning distributions with geometric centroids of head classes, ViRN preserves manifold structure while improving classification performance in limited-sample scenarios.

III. Experiments

Our experiments evaluate ViRN on six long-tailed datasets across acoustic and visual domains with an imbalance setting, including Speechcommands, AudioMNIST, ESC50, UrbanSound8k, CIFAR100, and TinyImageNet. We compare ViRN against state-of-the-art class-incremental methods: iCaRL [1], DGR [2], SLCA [3], and LAE [4].

We evaluated average Top-1 accuracy to measure robustness against catastrophic forgetting and long-tailed bias. ViRN achieves state-of-the-art performance, surpassing prior methods by an average of 10.24%, particularly excelling on acoustic datasets with small sample sizes and extreme class imbalance. Unlike SLCA, which collapses on tasks like ESC50 due to covariance estimation failures, ViRN demonstrates strong handling of long-tailed scenarios through distributional trilateration.

Ablation studies reveal the critical roles of variational inference (VI) and distribution redistribution (RD). Without VI, tail class accuracy drops significantly (44.96% on ESC50-LT), while omitting RD leads to overfitting and increased forgetting (7.94%). ViRN's fusion of these components achieves a balanced performance trade-off, demonstrating robustness in long-tailed CIL scenarios.

IV. Conclusion and Discussion

We present ViRN, a framework for LCIL that unifies variational inference with distributional geometry. ViRN reduces bias toward head classes through class-conditional modeling with VI, while trilateration reconstructs tail-class representations by leveraging semantically similar neighbors. Extensive experiments across six benchmarks demonstrate ViRN's effectiveness, highlighting its potential for real-world applications.

References

[1] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and et al., "icarl: Incremental classifier and representation learning," *CVPR*, pp. 5533–5542, 2016.
[2] J. He and F. M. Zhu, "Gradient reweighting: Towards imbalanced class-incremental learning," *CVPR*, pp. 16 668–16 677, 2024.
[3] G. Zhang, L. Wang, and et al., "Slca: Slow learner with classifier alignment for continual learning on a pre-trained model," *ICCV*, 2023.
[4] Q. Gao, C. Zhao, Y. Sun, and et al., "A unified continual learning framework with general parameter-efficient tuning," *ICCV*, 2023.