MaxEnt Loss: Calibrating Graph Neural Networks under Out-of-Distribution Shift (Student Abstract)

Dexter Neo

School of Computing, National University of Singapore
e0534450@u.nus.edu

Abstract

We present a new, simple and effective loss function for calibrating graph neural networks (GNNs). Miscalibration is the problem whereby a model's probabilities does not reflect it's correctness, making it difficult and possibly dangerous for real-world deployment. We compare our method against other baselines on a novel ID and OOD graph form of Celeb-A faces dataset. Our findings show that our method improves calibration for GNNs, which are not immune to miscalibration in-distribution (ID) and out-of-distribution (OOD). Our code is available for review at *https://github.com/dexterdley/CS6208/tree/main/Project*.

Introduction

Calibration seeks to align a model's predictions with its probabilities. This initial paper surveys general state-of-the-art objectives functions that can be used to calibrate GNNs. We develop a novel makeshift form of Celeb-A with the task of multi-label facial recognition and a novel loss function (MaxEnt loss) for calibrating GNNS. When trained with cross entropy (CE) loss, our findings agree with (Hsu et al. 2022) that GNN's have a tendency to be underconfident. However, we show that calibration can be greatly improved with MaxEnt loss.

Preliminaries

Expected Calibration Error (ECE) Calibration is the measure of how a model's predicted confidence is aligned with its correctness. Both over- and under-confident predictions lead to miscalibration and meaningless probabilities. Many different metrics have been proposed to measure calibration, with the most popular metric being ECE (Naeini, Cooper, and Hauskrecht 2015), which divides the model's probabilities into B bins. The number of samples, average accuracy and confidence for each bin is represented by n_b, *acc* and *conf*. The weighted differences between each *acc* and *conf* bins are measured: ECE $= \sum_{b=1}^{B} \frac{n_b}{N} |acc(B_b) - conf(B_b)|$.

Out-of-distribution Shifts Typically, a classifier is trained with the assumption that the test distribution is aligned with the training set. This assumption no longer holds true, when inputs deviate greatly from the train set. These shifted samples are considered OOD and can be caused by differences in resolution, lighting, blurs or noise (Hendrycks and Gimpel 2017). An illustration of OOD is shown in Figure 1, where for computer vision tasks images can be simply augmented to replicate an OOD shift, for graphs we propose a similar problem and use a resolution shift (e.g different number of nodes and edges). OOD samples generally cause a decrease in recognition and calibration performance.

Figure 1: Distribution shifts often occur in traditional computer vision tasks (Left). For graph ML tasks, we propose an OOD shift in node resolution (Right).

MaxEnt Loss

For classification tasks, models are typically trained using the ground-truth onehot labels y_k. The CE loss between the predictions p_k is minimized across all classes K and $\mathcal{Y} = \{1, ..., K\}$ is a fixed vector containing each random variable. The binary form of the CE loss is written as $\mathcal{L}_{CE} = -\sum_k y_k \log(p_k) - \sum_k (1 - y_k) \log(1 - p_k)$. As a simple and effective alternative to CE loss, Focal loss (Mukhoti et al. 2020) has also been show to improve model calibration. The hyperparameter γ controls the order of the polynomial focal term and the binary form of the Focal loss is given as:

$$\mathcal{L}_F = -\sum_k (1-p_k)^\gamma y_k \log(p_k) - \sum_k p_k^\gamma (1-y_k) \log(1-p_k) \tag{1}$$

Based on the Principle of Maximum Entropy (Jaynes 1957), we propose MaxEnt loss by adding the following constraint to the Focal loss. Where the scalar μ is a global constraint computed from the ratio for each class. Specifically, the prior distribution $\frac{N_k}{N}$ is used to compute the expectation where $\mathbb{E}[\mathcal{Y}] = \sum_k \mathcal{Y} \frac{N_k}{N} = \mu$. Subjected to the given constraint, the Lagrange multiplier λ_1 can thereafter be computed for each attribute numerically using Newton Raphson.

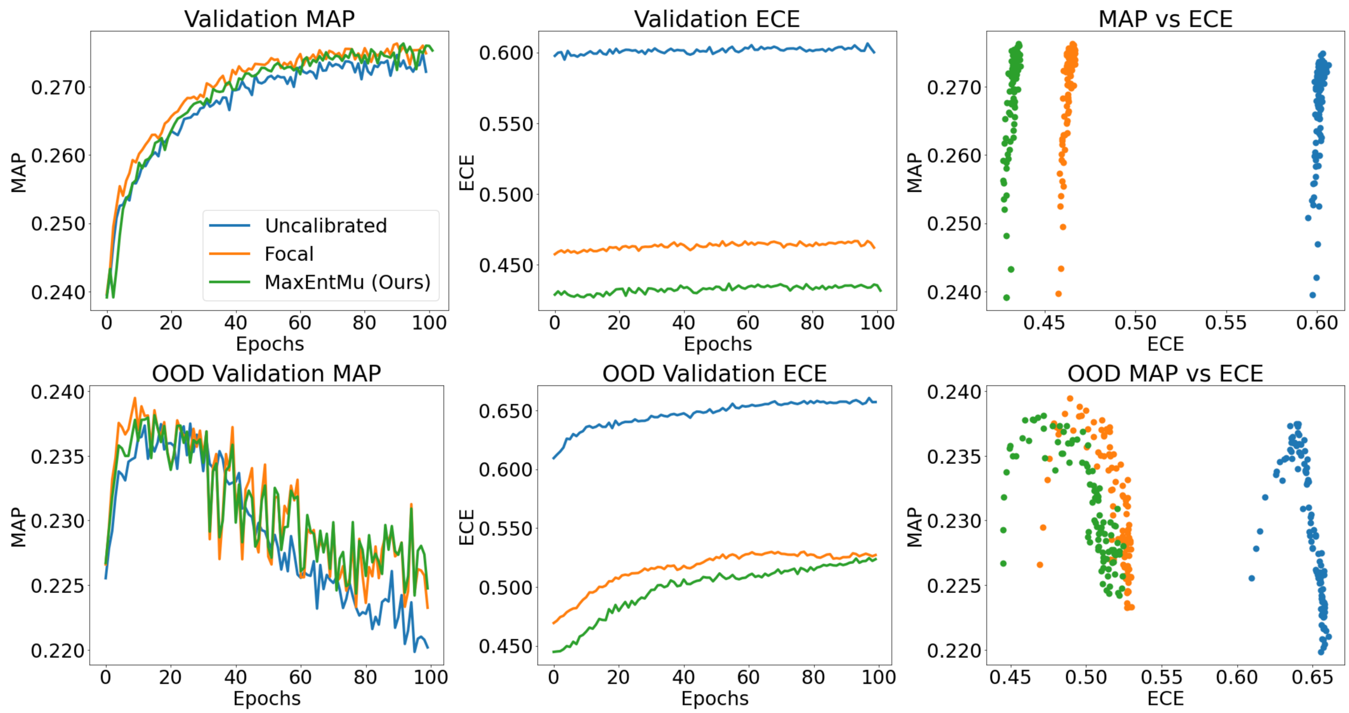

Figure 2: Learning curves comparing the performances of different loss functions on ID (top) and OOD (bottom) inputs. When evaluated OOD all methods see a drop in performance, however MaxEnt loss remains robustly calibrated.

The MaxEnt loss (Neo, Winkler, and Chen 2023) is given by:

$$\mathcal{L}_{ME}^{M} = \underbrace{\mathcal{L}_F}_{\text{Focal Loss}} + \underbrace{\lambda_1 \left(\sum_k \mathcal{Y} p_k - \mu \right)}_{\text{Mean constraint}} \tag{2}$$

Experiments

For our experiments, we use Pytorch's GCNConv and train uncalibrated baselines using CE loss and compare it with Focal and MaxEnt loss. We generate graphs from the facial images of the CelebA dataset (Liu et al. 2015). Specifically, the nodes are facial landmarks and the edges are derived via triangulation. In Figure 2, we can see that all loss functions converge to similar mean average precision (MAP), however CE loss results in the worst calibration performance. Similar to standard deep learning tasks, Focal loss improves the calibration performance greatly with an overall improvement of roughly 10%. Our method achieves the best performance, with an additional improvement of 5%. As expected, we observe a drop in MAP and ECE for all methods when evaluated OOD, but MaxEnt Loss is still able to deliver the best calibration performance OOD.

Conclusion and Future Work

We presented a novel loss function for calibrating GNNs. For future work, we plan to evaluate more GNN algorithms trained with MaxEnt loss compared against additional calibration loss functions on graph OOD benchmarks (Gui et al. 2022).

References

Gui, S.; Li, X.; Wang, L.; and Ji, S. 2022. GOOD: A Graph Out-of-Distribution Benchmark. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Hendrycks, D.; and Gimpel, K. 2017. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. *arXiv*, abs/1610.02136.

Hsu, H. H.-H.; Shen, Y.; Tomani, C.; and Cremers, D. 2022. What Makes Graph Neural Networks Miscalibrated? In Oh, A. H.; Agarwal, A.; Belgrave, D.; and Cho, K., eds., *Advances in Neural Information Processing Systems*.

Jaynes, E. T. 1957. Information Theory and Statistical Mechanics. *Physical Review*, 106: 620–630.

Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep Learning Face Attributes in the Wild. In *Proc. International Conference on Computer Vision (ICCV)*.

Mukhoti, J.; Kulharia, V.; Sanyal, A.; Golodetz, S.; Torr, P. H.; and Dokania, P. K. 2020. Calibrating Deep Neural Networks using Focal Loss. In *Advances in Neural Information Processing Systems*.

Naeini, M. P.; Cooper, G. F.; and Hauskrecht, M. 2015. Obtaining Well Calibrated Probabilities Using Bayesian Binning. In *Proc. 29th AAAI Conference on Artificial Intelligence*, 2901–2907.

Neo, D.; Winkler, S.; and Chen, T. 2023. MaxEnt Loss: Constrained Maximum Entropy for Calibration under Out-of-Distribution Shift. *arXiv*, 2310.17159.