An Efficient Memory Module for Graph Few-Shot Class-Incremental Learning

Dong Li, Zhang Aijia,
Junqi Gao, Biqing Qi

NEURAL INFORMATION PROCESSING SYSTEMS

Motivation

The motivation behind this paper is to address the challenges in graph learning, particularly in dynamic contexts where graphs grow over time with new nodes and edges. The paper highlights the issue of catastrophic forgetting in Graph Neural Networks (GNNs) when updating graph representation learning methods with new data, leading to a loss of previously acquired knowledge. Furthermore, it emphasizes the critical challenge of label scarcity for newly introduced nodes, which complicates the design of effective regularization strategies and hampers the ability of existing graph continual learning methods to generalize in graph few-shot learning scenarios. The main contributions of this paper are as follows:

- **Design of Mecoin:** The paper introduces Mecoin, a novel framework that effectively mitigates catastrophic forgetting in Graph Few-Shot Class-Incremental Learning (GFSCIL) by integrating the Structured Memory Unit (SMU) and Memory Representation Adaptive Module (MRaM).
- **Structured Memory Unit (SMU):** The SMU is designed to efficiently learn class prototypes by facilitating interaction between node features and existing class prototypes, while extracting local graph structures of input nodes.
- **Memory Representation Adaptive Module (MRaM):** The MRaM is proposed to reduce the loss of prior knowledge during parameter fine-tuning by decoupling the learning of class prototypes from node probability distributions.
- **Analysis of Separation Benefits:** The paper analyzes the benefits of separating class prototype learning from node probability distribution learning, considering generalization error bounds and VC dimensions. It also explores how different MRaM-GNN interaction patterns affect model performance.
- **Empirical Studies:** Through extensive empirical studies, the paper demonstrates Mecoin's significant advantages over current state-of-the-art methods.

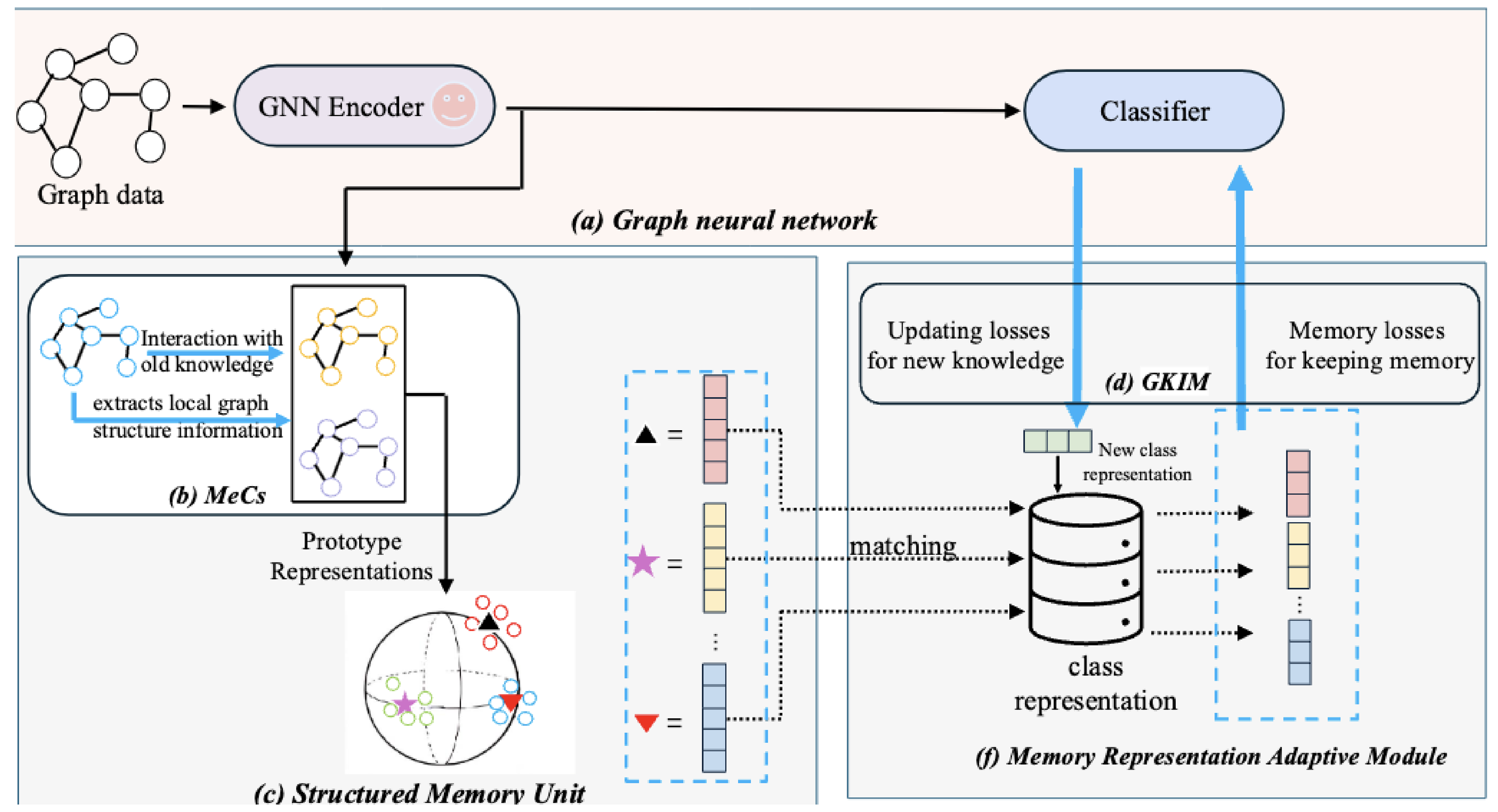
Theory

$$\mathcal{R} \leq \mathcal{R}_\epsilon + \mathcal{B}_{\hat{f}}\mathbb{I}\{f = \hat{f}\} + c\sqrt{\frac{2\ln(e/\delta)}{N}},$$

$$VCD = \begin{cases} \mathcal{O}(n+1) & \text{voting classifier} \\ \mathcal{O}(p^2H^2) & \text{MLP} \\ \mathcal{O}(p^2n^2H^2) & \text{GKIM} \end{cases}$$

Mecoin has a lower upper bound on the generalization error, and a larger VC dimension compared with other GNN models.

Methodologies



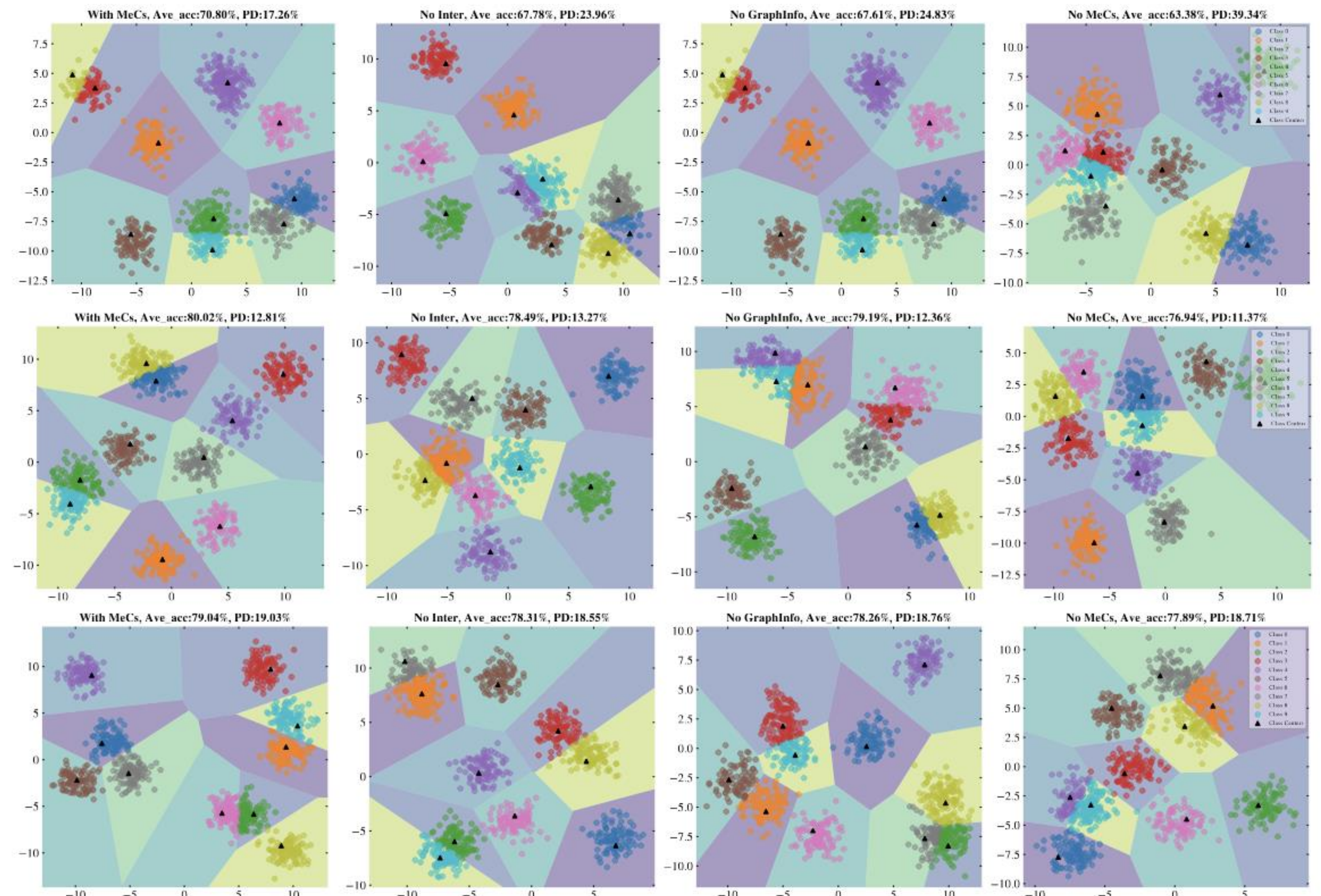
Overview of the Mecoin framework for GFSCIL. (a)Graph neural network: Consists of a GNN encoder and a classifier(MLP) pre-trained by GNN. In GFSCIL tasks, the encoder parameters are frozen. (b)Structured Memory Unit: Constructs class prototypes through MeCs and stores them in SMU. (c)Memory Representation Adaptive Module: Facilitates adaptive knowledge interaction with the GNN model.

Experimental results

Method	backbone	0	1	2	3	4	5	6	7	8	9	10	PD ↓	Average ACC ↑
		Acc. in each session (%) ↑												
ERGNN	GCN	73.43	77.64	29.38	32.14	20.93	21.84	16.13	15.28	15.2	11.77	11.30	62.13	29.55
	GAT	73.43	77.64	29.38	32.14	20.93	21.84	16.13	15.28	15.20	11.77	11.30	62.13	29.55
Mecoin	GCN	73.43	45.92	25.20	18.40	19.44	13.69	11.70	11.47	9.86	8.83	8.27	65.16	22.38
	GAT	69.06	49.49	25.30	19.85	19.44	14.78	12.79	11.91	9.90	9.15	7.52	61.54	22.65
MAS	GCN	73.43	46.40	25.20	18.40	19.44	13.69	11.70	11.47	9.86	8.83	8.27	65.16	22.38
	GAT	69.06	66.64	44.45	48.13	37.08	47.04	48.58	49.16	44.26	49.13	46.39	22.67	49.99
LWF	GCN	73.43	45.92	25.75	17.82	19.71	14.16	10.95	11.38	9.90	8.08	7.73	65.7	22.26
	GAT	73.60	48.83	24.59	24.03	19.95	13.91	13.76	13.24	10.83	14.02	7.46	66.14	24.02
EWC	GCN	73.43	45.92	26.48	18.52	20.27	14.59	11.41	11.59	10.04	7.79	7.75	65.68	22.53
	GAT	69.06	48.5	25.58	25.34	20.39	18.59	14.86	22.71	19.7	21.31	12.55	56.51	27.14
TWP	GCN	70.54	44.90	26.71	20.65	24.16	14.56	18.77	14.37	19.37	12.48	14.48	56.06	25.54
	GAT	69.06	50.27	26.53	28.27	21.4	23.29	15.04	26.74	14.86	17.03	13.77	55.29	27.84
HAG-Meta	/	**87.62**	**82.78**	**79.01**	**74.5**	67.65	63.38	60.00	57.36	55.5	52.98	51.47	36.15	66.57
Geometer	/	72.23	61.73	43.34	37.61	36.24	32.79	29.97	22.11	21.01	17.34	16.32	55.91	35.52
OURS	GCN	82.18	81.56	77.47	71.53	**70.43**	**69.70**	**68.78**	**68.07**	**66.70**	**62.10**	**61.36**	20.82	**70.90**
	GAT	75.53	73.04	70.74	67.52	66.05	64.97	64.18	63.68	62.02	60.62	60.10	**15.43**	66.22

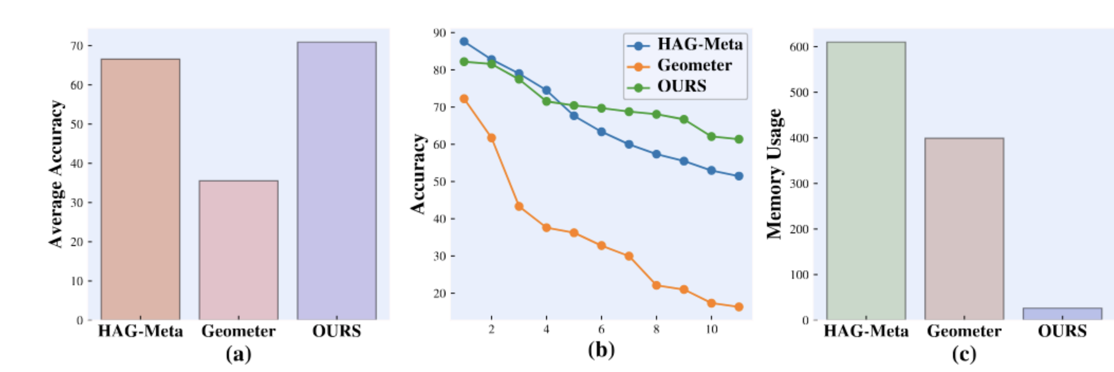
Method	backbone	0	1	2	3	4	5	6	7	8	9	10	PD ↓	Average ACC ↑
		Acc. in each session (%) ↑												
ERGNN	GCN	100	50.00	33.33	25.00	20.00	16.67	14.29	12.5	11.11	10.00	14.02	85.98	27.90
	GAT	100	50.00	33.33	46.17	33.95	36.93	24.64	23.34	18.60	25.44	30.12	69.88	38.41
Mecoin	GCN	100	50.00	33.33	25.00	20.00	16.67	14.29	12.50	11.11	10.00	12.12	87.88	27.73
	GAT	100	50.00	38.71	25.00	20.00	17.03	14.85	13.98	14.17	21.52	18.09	81.91	30.30
MAS	GCN	100	50.00	33.33	25.27	20.00	16.67	14.29	12.36	11.11	18.75	9.70	90.30	28.32
	GAT	100	50.00	34.05	48.23	56.16	59.54	65.57	64.28	61.41	64.36	63.92	36.08	60.68
LWF	GCN	100	50.00	33.33	25.00	20.00	16.73	14.29	12.36	11.11	16.36	15.44	84.56	28.60
	GAT	100	50.00	33.33	24.71	36.28	36.40	28.92	28.84	21.23	29.12	32.51	67.49	38.30
EWC	GCN	100	50.00	33.33	25.00	20.00	16.67	14.29	12.5	11.11	10.00	12.12	87.88	27.73
	GAT	100	50.00	33.33	31.65	38.60	36.62	25.64	29.04	16.19	26.44	38.63	61.37	38.74
TWP	GCN	100	50.00	33.33	25.00	20.00	16.67	14.29	12.43	11.04	16.81	15.03	84.97	28.6
	GAT	100	50.35	33.33	25.18	38.14	47.14	39.34	27.67	30.52	45.11	52.02	47.98	44.44
HAG-Meta	/	20.00	16.67	14.29	12.50	11.11	10.00	9.09	8.33	7.69	7.14	6.66	**13.34**	11.24
Geometer	/	60.60	33.85	24.05	19.03	24.87	28.86	24.46	18.18	19.30	26.39	29.63	30.97	28.11
OURS	GCN	98.07	94.09	**91.01**	**73.75**	**79.95**	**82.21**	74.13	72.17	69.48	63.01	59.66	38.41	**77.96**
	GAT	97.83	**95.13**	90.75	68.02	71.79	77.88	**75.35**	**75.06**	**72.52**	**65.92**	**62.21**	35.62	77.50

SMU for class prototypes

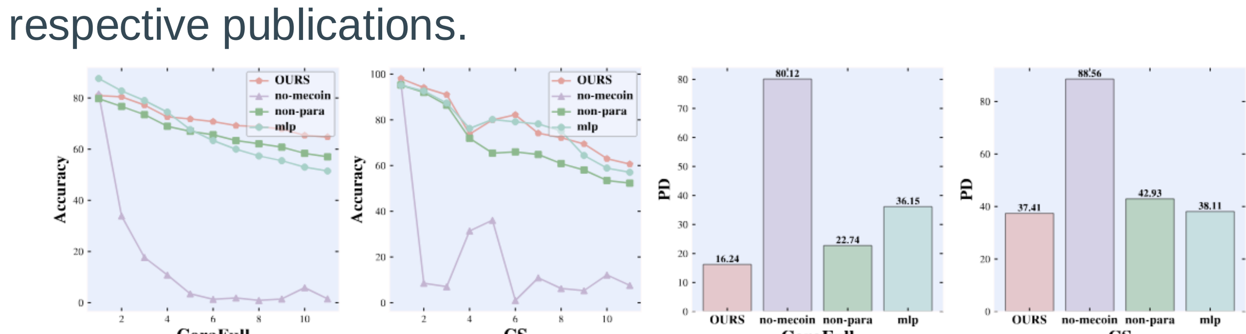


The outcomes of GKIM when conducting the few-shot continuous learning task on the CoraFull, Computers and CS datasets. The results are presented sequentially from left to right: GKIM with full capabilities, GKIM where node features do not interact with class prototypes in the SMU, GKIM without GraphInfo and GKIM without MeCs . The experimental results for CoraFull are shown in the above figure, the results for Computers are in the middle and the results for CS are in the figure below.

SMU for class prototypes



The comparative analysis of the mean performance, accuracy curves and memory utilization of HAG-Meta, Geometer and Mecoin across 10 sessions on CoraFull, conducted under the experimental conditions delineated in their respective publications.



Left 2 columns: Line charts depict the performance of models across various sessions on the CoraFull and CS datasets when using different distillation methods; Right 2 columns: Histograms illustrate the forgetting rates of different distillation methods on these two datasets.