Context-guided Embedding Adaptation for Effective Topic Modeling in Low-Resourced Regimes

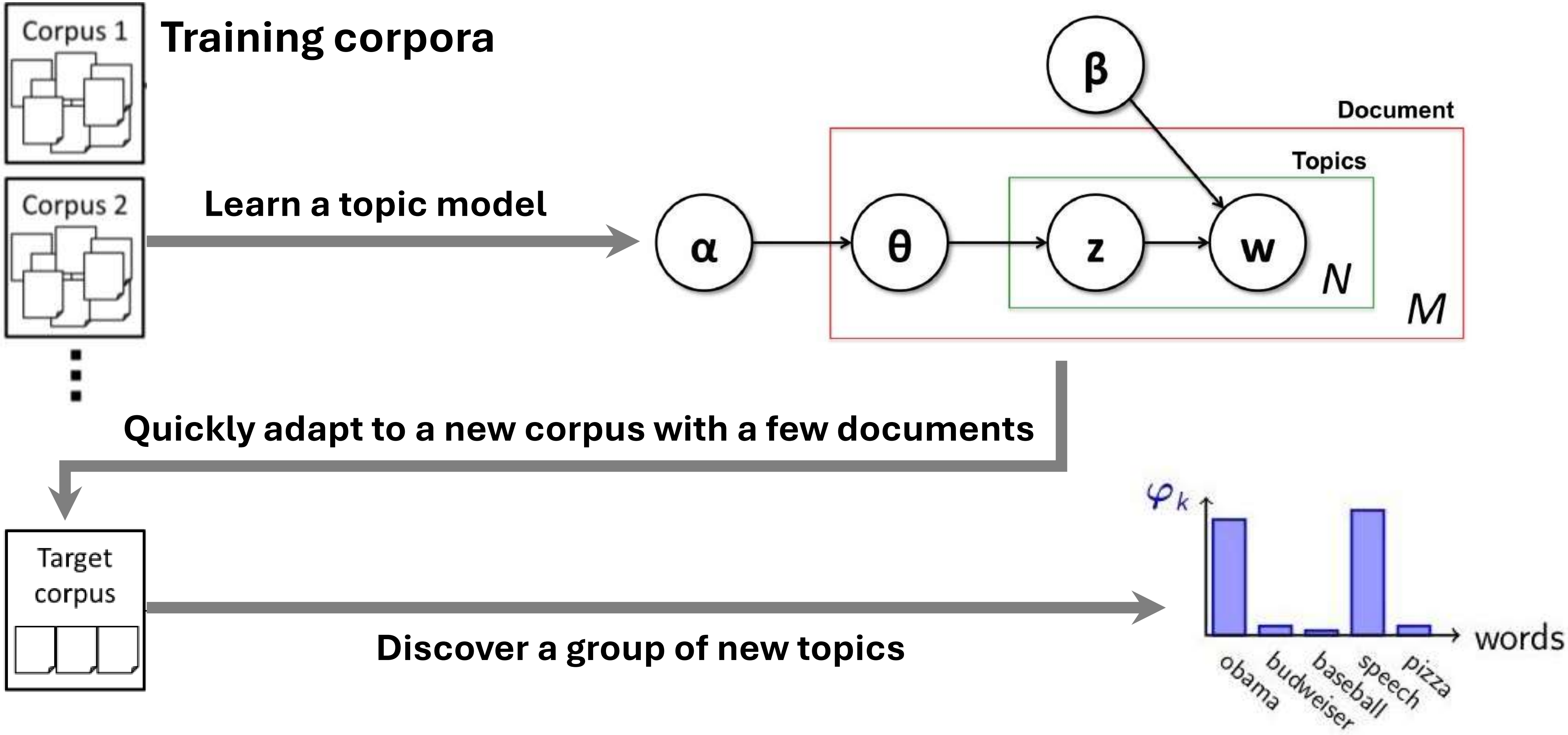
Yishi Xu[1,*] Jianqiao Sun[1,*] Yudi Su[1] Xinyang Liu[1] Zhibin Duan[1] Bo Chen[1] Mingyuan Zhou[2]

1 National Laboratory of Radar Signal Processing, Xidian University

2 McCombs School of Business, The University of Texas at Austin

Source code: https://github.com/NoviceStone/Meta-CETM

Topic modeling in low-data regimes



Motivation

Existing embedded topic models generally view the static word embeddings learned from source tasks as *general knowledge* that can be directly transferred to the target task with only a few documents.

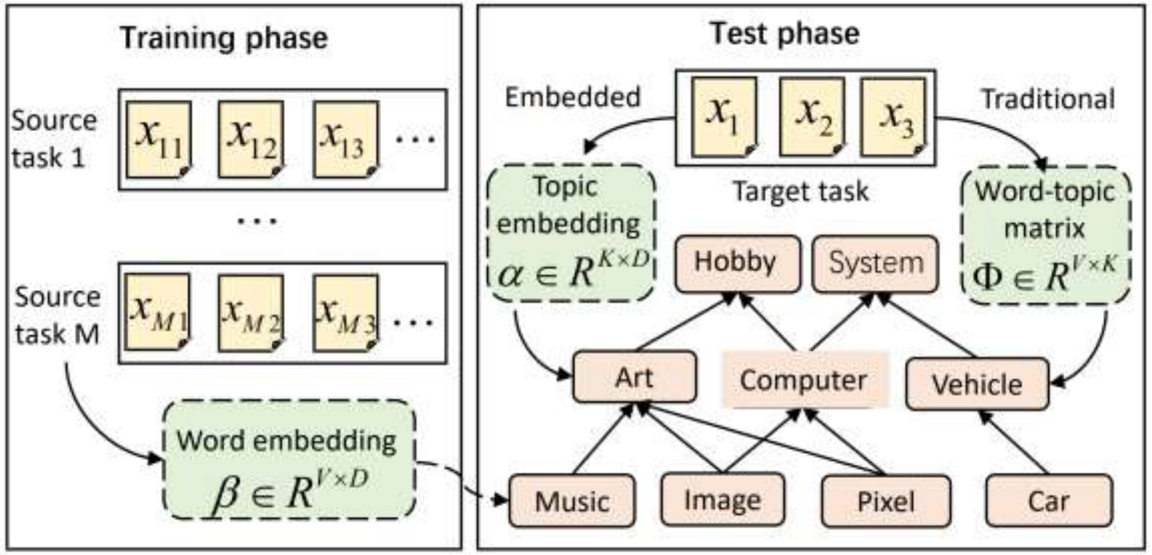


Figure 1: Illustration of the advantage of embedded topic models over traditional topic models in low-resourced regimes.

Motivation

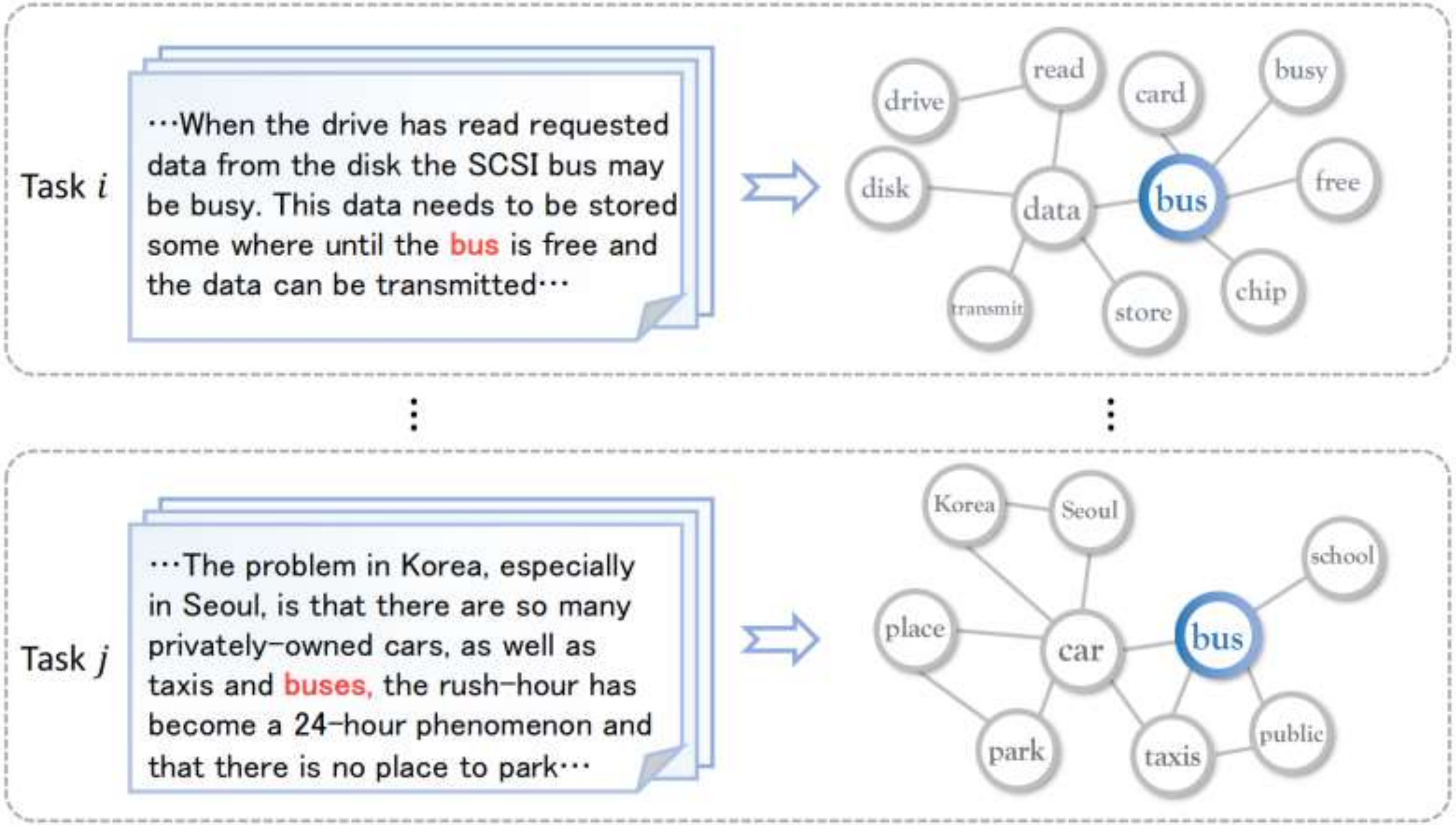


Figure 2: An example of word sense variation caused by different contexts. The task i is sampled from a corpus about "hardware", and the task j is sampled from a corpus related to "autos". By means of established dependency parsing tools, we build a semantic graph for each task to capture syntactic dependencies between words in the context.

Word meanings inevitably change with different contexts.

How to effectively adapt the word embedding to a new task?

Context-guided embedding adaption

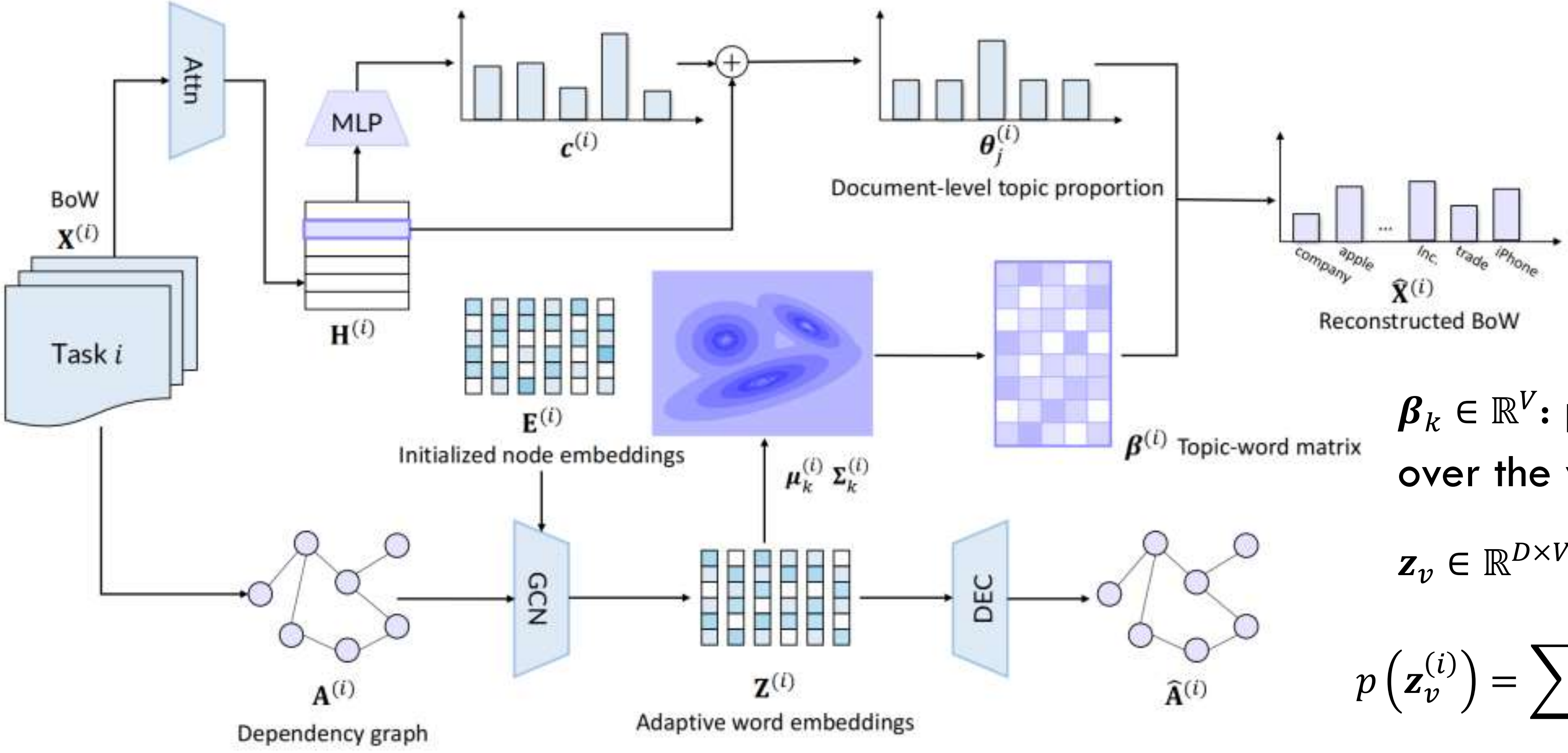


Figure 3: Overview of the proposed framework. The top branch establishes a standard neural topic modeling pipeline, with the topic-word matrix derived according to the word embeddings' probability densities. The bottom branch creates a graph VAE to learn contextualized word embeddings, with a Gaussian mixture prior imposed on the latent space to yield topic representations.

$\boldsymbol{\beta}_k \in \mathbb{R}^V$: per-topic distribution over the vocabulary terms

$\boldsymbol{z}_v \in \mathbb{R}^{D \times V}$: word embeddings

$$p\left(\boldsymbol{z}_v^{(i)}\right) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(\boldsymbol{z}_v^{(i)} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$$

$$\boldsymbol{\beta}_k^{(i)} = \mathrm{Softmax}\left(p\left(\boldsymbol{Z}^{(i)} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)\right)$$

Per-holdout-word perplexity results

- Our model yields the best predictive performances.

Methods	20NG		Yahoo		DB14		WOS	
	5	10	5	10	5	10	5	10
LDA[42]	4021±1528	3502±1277	4476±1544	4028±1097	4410±1918	3697±1747	3439±671	3246±461
PFA[12]	3463±1452	3150±1119	3257±1328	3122±1040	3443±1937	3170±1562	3113±819	3431±830
ProdLDA[43]	4853±1034	4523±817	5765±1104	5378±826	5477±846	5297±740	4311±469	4220±392
ETM[18]	3192±895	3107±671	2868±909	2817±620	3217±1960	3054±1539	3135±704	3310±455
MAML-ProdLDA*	4292±1123	4355±997	4354±1369	4250±919	4844±1337	4678±1119	4117±462	4068±332
MAML-ETM*	3849±1064	3725±841	3653±1081	3642±776	4448±2737	4279±2301	3483±4044	3277±644
Meta-SawETM[30]	2872±869	2984±740	2365±934	2487±756	2047±1374	1914±1009	2031±445	2253±315
CombinedTM[21]	2660±659	2595±625	2700±590	2674±575	1851±767	1774±731	2562±633	2648±658
ZeroShotTM[22]	2904±851	2569±663	2822±732	2795±721	1938±758	1835±739	2863±704	2775±558
Meta-CETM	**954**±543	**1170**±606	**1074**±442	**1219**±455	**802**±571	**1084**±643	**1293**±542	**1528**±218

Table 1: Performance comparison of different topic models on the per-holdout-word perplexity (5 and 10 documents in each task are considered).

Topic quality

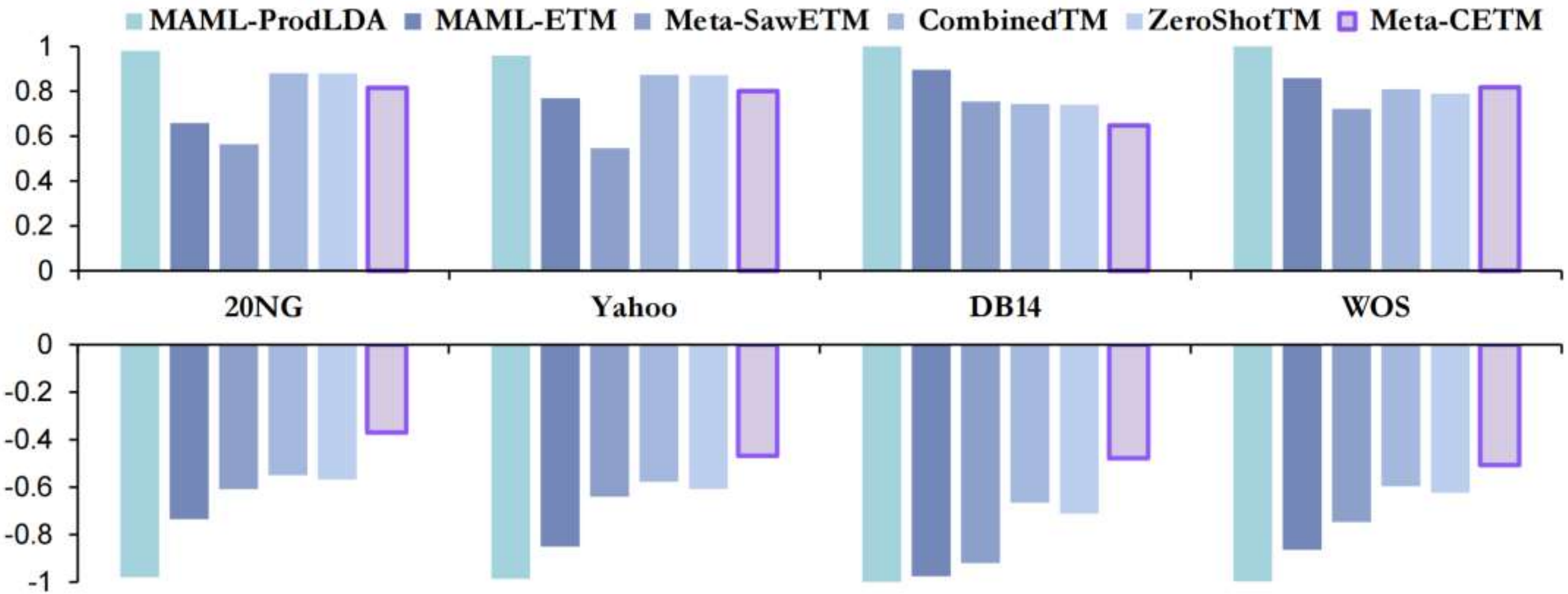


Figure 4: Performance comparison of six selected methods for **topic diversity** (top row) and **topic coherence** (bottom row) on four datasets. The topics are adapted from each task with 10 documents.

Topic visualizations

- Our model can adapt to the target task effectively.

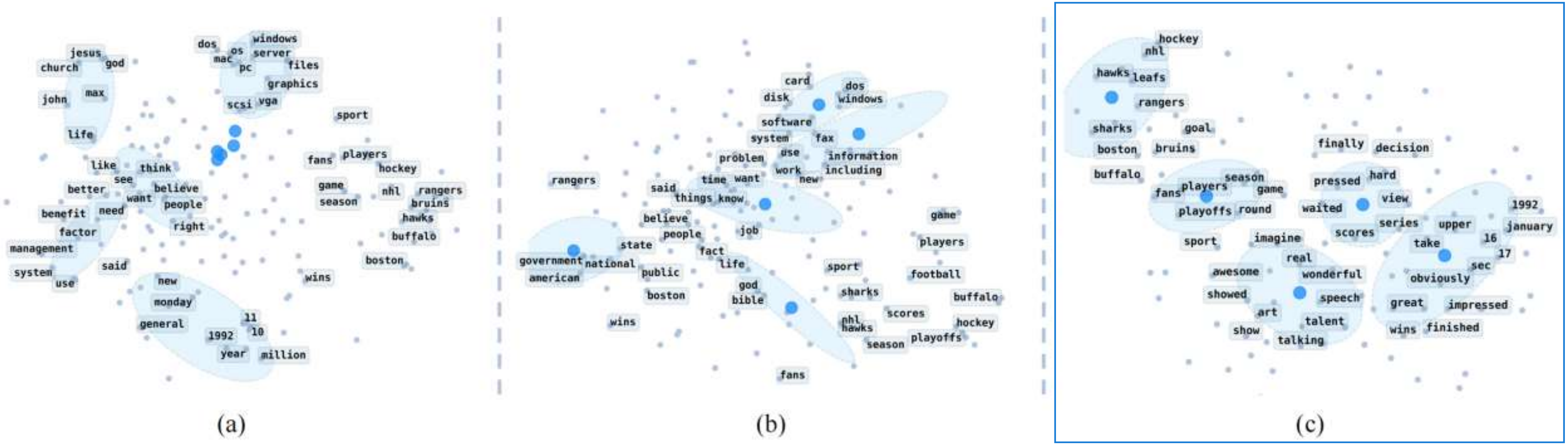


Figure 5: Visualization of the **adapted embedding space** for (a) MAML-ETM, (b) Meta-SawETM, and (c) Meta-CETM (ours). The small grey points represent word embeddings, and the large blue points denote topic embeddings for MAML-ETM, topic embedding means for Meta-SawETM and Meta-CETM. The ellipse coverages display topic embedding covariances (note that MAML-ETM has not modeled topics as distributions so the ellipse coverages are plotted approximately based on the top words. The example task is sampled from the corpus of sub-topic "rec.sport.hockey" in *20Newsgroups* dataset.

Few-shot text classification results

Methods		20NG		DB14		Yahoo		WOS	
Rep.	Alg.	5 shot	10 shot	5 shot	10 shot	5 shot	10 shot	5 shot	10 shot
MLP	MAML[44]	32.01	36.20	50.20	60.30	45.42	51.00	37.77	40.43
	PROTO[52]	35.20	38.30	54.13	57.16	50.01	56.16	39.61	41.46
	FT[53]	29.70	33.04	51.11	53.83	48.59	53.06	36.52	37.22
	FT*	38.87	48.52	71.12	77.94	50.73	56.74	45.02	51.20
CNN	MAML[44]	34.08	45.40	66.28	75.96	48.81	56.50	47.28	57.32
	PROTO[52]	39.86	49.71	**78.58**	**81.01**	53.16	63.66	59.05	**67.75**
	FT[53]	45.70	53.63	74.68	80.75	56.78	66.04	54.68	63.39
	FT*	44.53	51.92	72.49	80.07	53.28	52.56	51.42	61.98
HNS-SawETM[30]		39.37	43.78	65.93	71.08	52.35	57.86	42.09	56.91
Meta-SawETM[30]		39.19	45.83	67.20	72.31	52.45	60.58	43.39	57.44
CombinedTM[21]		46.17	52.73	68.42	73.26	57.94	64.75	56.16	65.97
ZeroShotTM[22]		46.65	52.08	71.93	76.09	58.12	66.21	58.50	66.10
Meta-CETM		**50.57**	**58.47**	76.85	79.34	**63.84**	**72.67**	**61.47**	67.62

Thank you.

Please feel free to contact us by e-mail

xuyishi@stu.xidian.edu.cn jianqiaosun@stu.xidian.edu.cn

bchen@mail.xidian.edu.cn mingyuan.zhou@mccombs.utexas.edu

Paper can be downloaded from