Uncertainty-driven Embedding Convolution

Sungjun Lim[1], Kangjun Noh[1], Youngjun Choi[1], Heeyoung Lee[1], **Kyungwoo Song**[1]

[1] Yonsei University

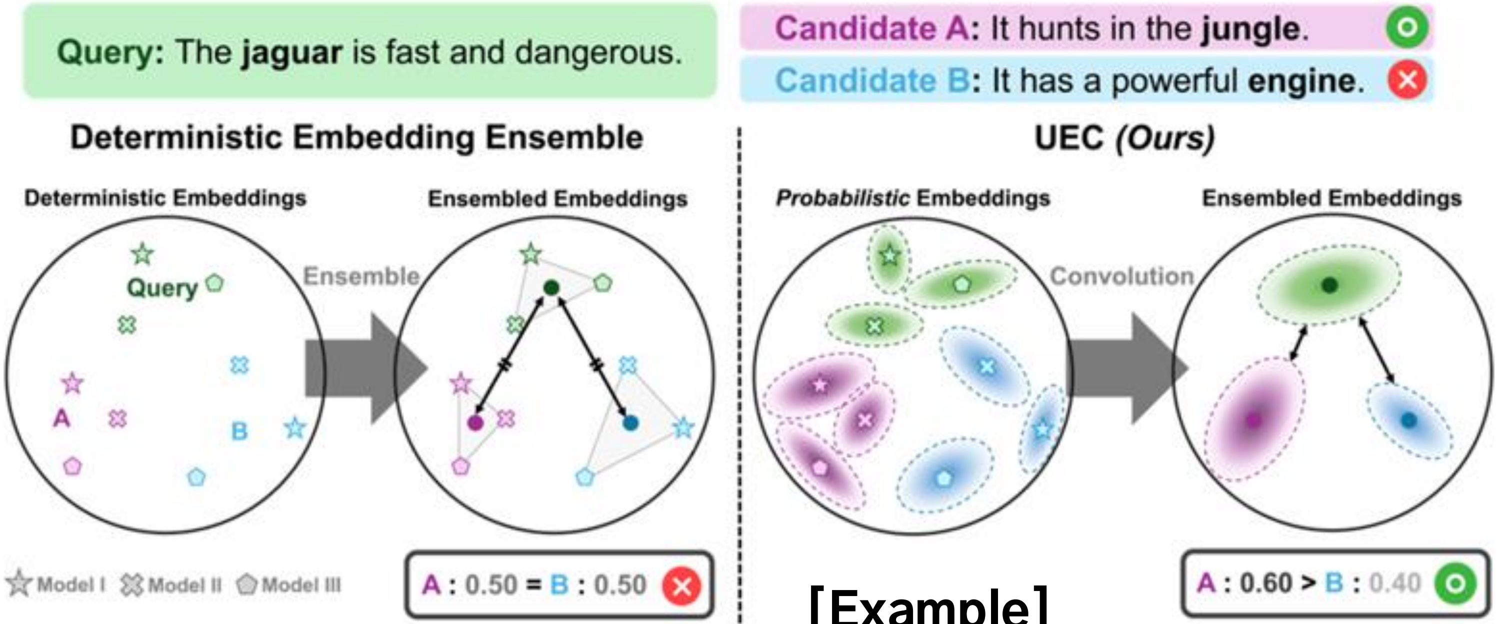
paper

ICLR2026, Poster Session 5
(In Room: Pavilion 3 Poster Location: P3–#220)

Motivation

Need for *Uncertainty-driven* Embedding Ensemble

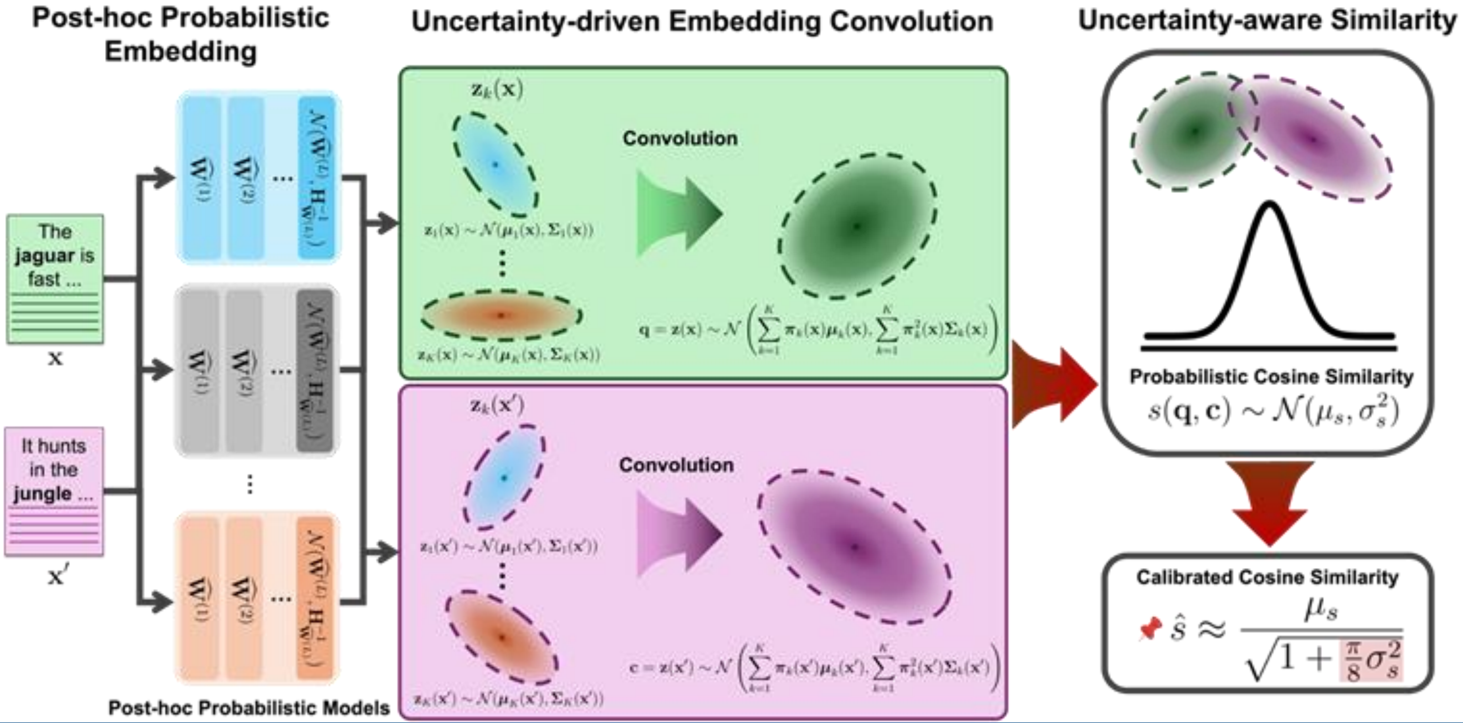
- No single embedding excels universally. ⇒ **Need for Ensemble**
- *Without uncertainty*, ensemble can lead to suboptimal performance.



[Example]

Overall

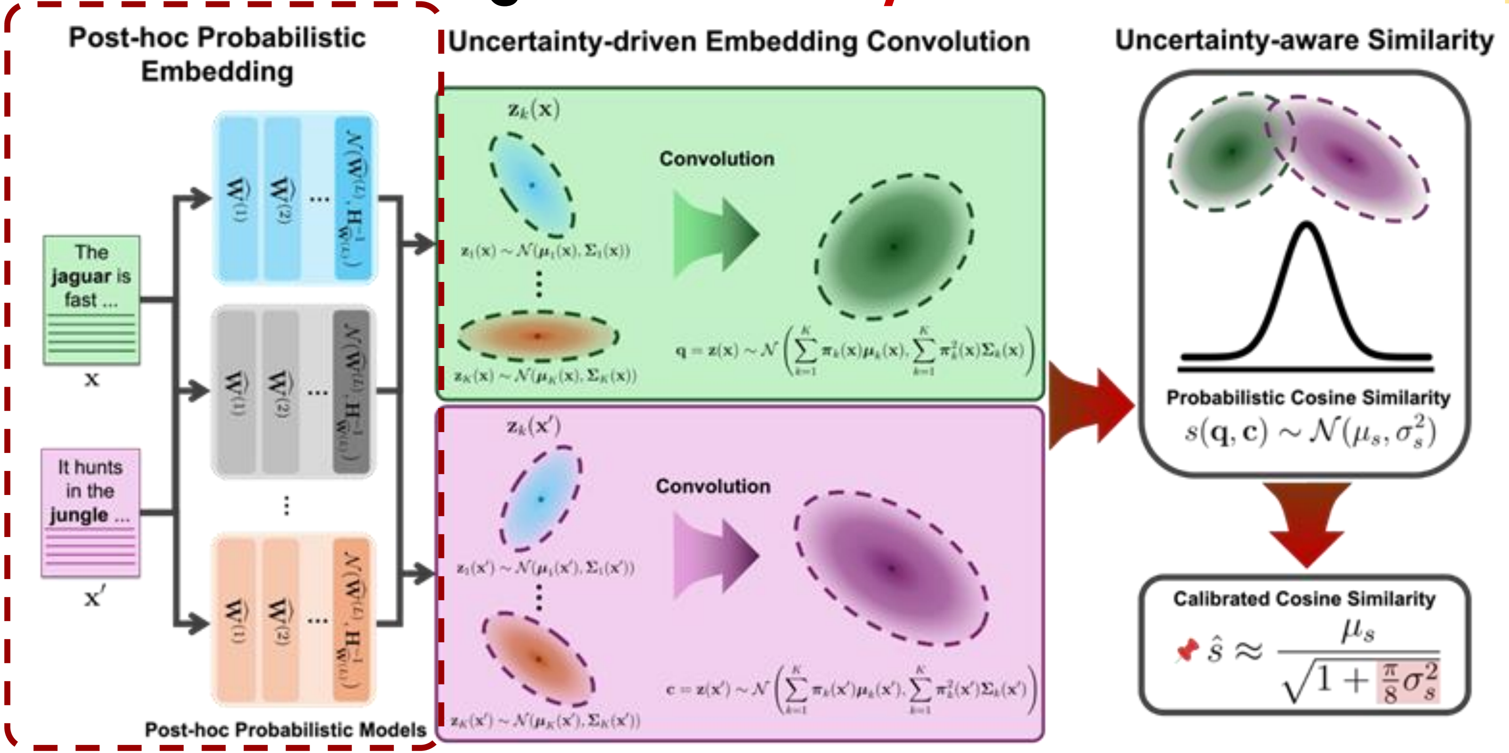
Propose Uncertainty-driven Embedding Convolution (UEC)!



Methodology

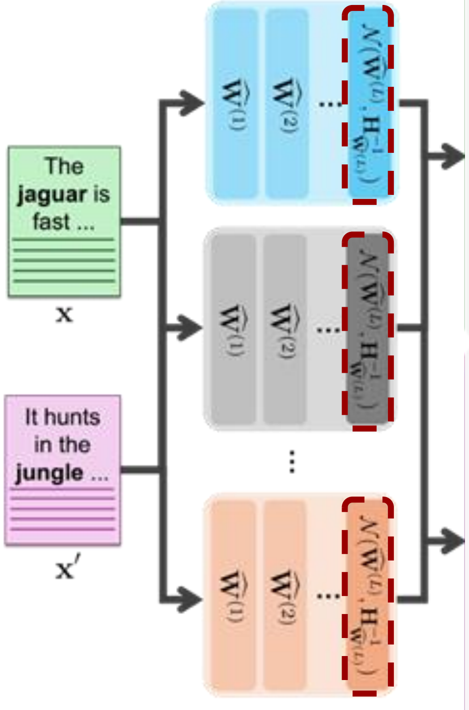
(1) Post-hoc Probabilistic Embedding

Convert embedding models into *probabilistic* ones *via* *LA*



(1) Post-hoc Probabilistic Embedding

Convert embedding models into *probabilistic* ones *via* LA



- Apply LA to **last layer only**
- Apply **diagonal covariance** for LA

Efficient

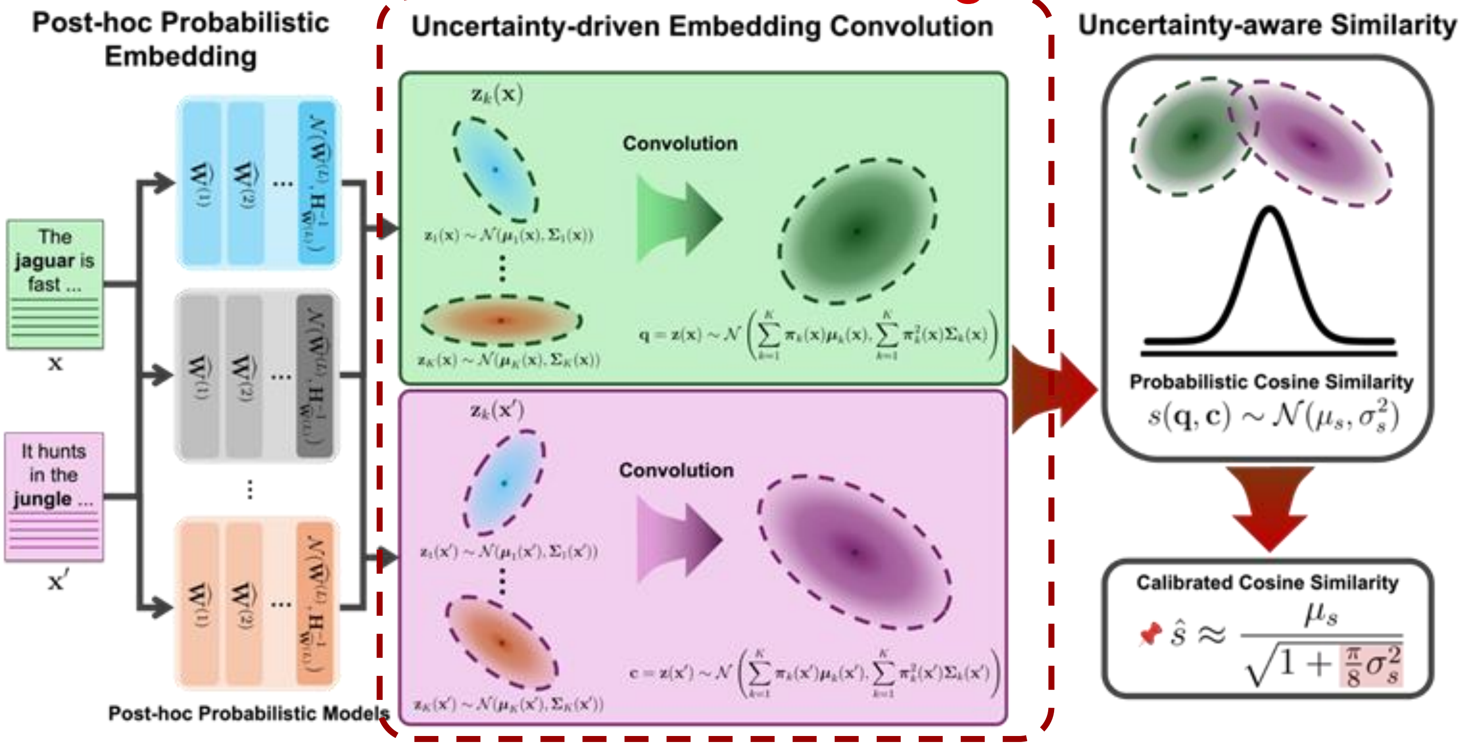
Generate Gaussian embeddings for k-th embedding model

$$\mathbf{z}_k(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}_k(\mathbf{x}), \boldsymbol{\Sigma}_k(\mathbf{x}))$$

Methodology

(2) Uncertainty-driven Embedding Convolution

Convolution Probabilistic Embedding



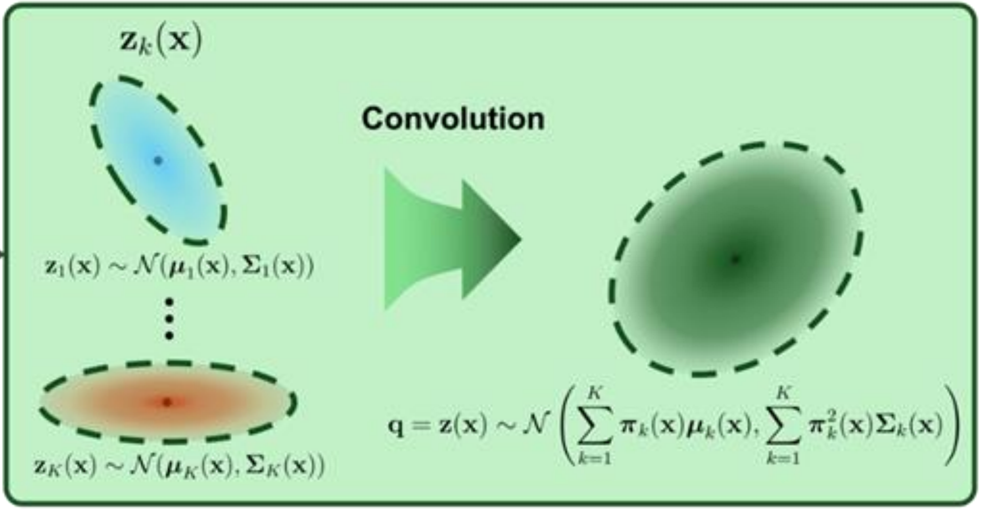
Convolution Probabilistic Embedding

- k-th probabilistic embedding

$$\mathbf{z}_k(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}_k(\mathbf{x}), \boldsymbol{\Sigma}_k(\mathbf{x}))$$

- Convolution K embeddings

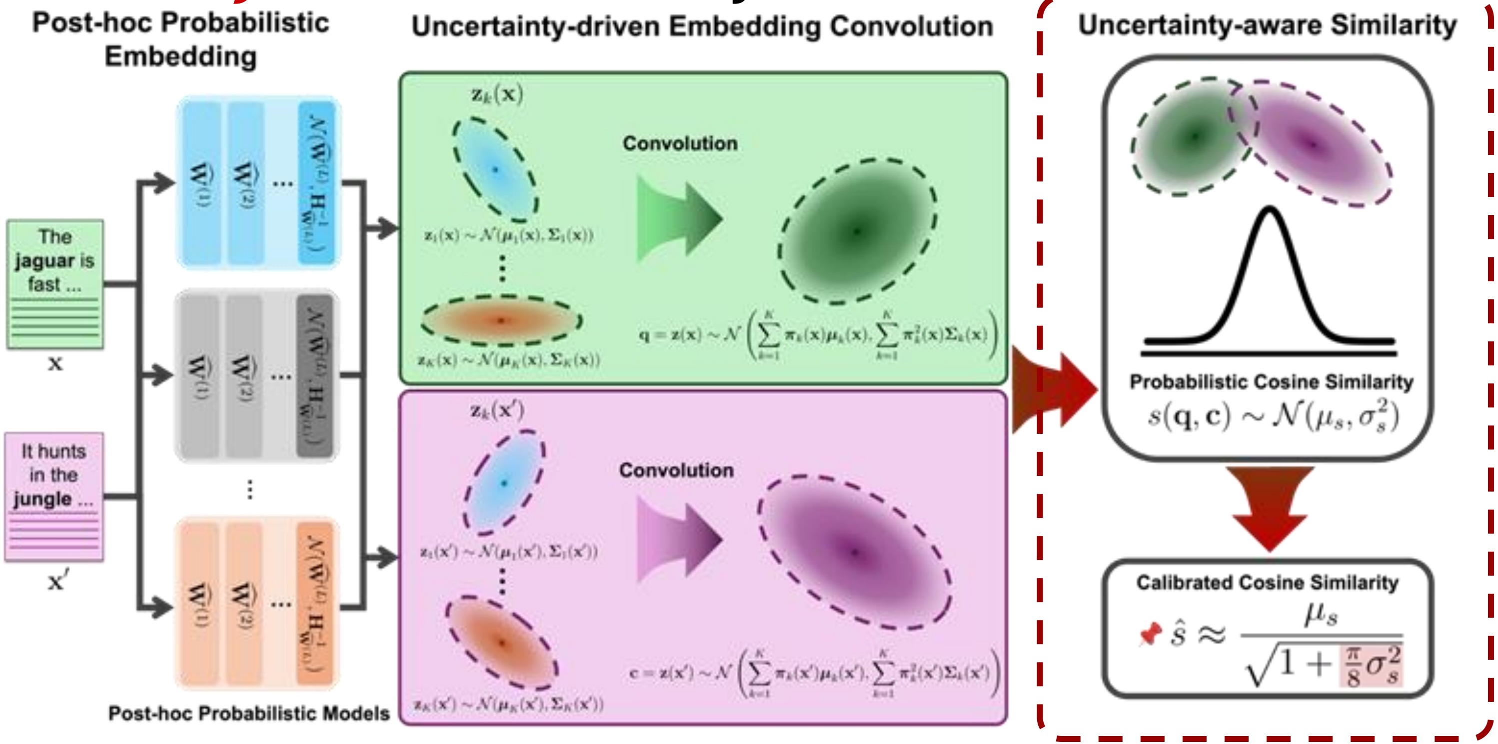
$$\mathbf{z}(\mathbf{x}) = \sum_{k=1}^{K} \pi_k(\mathbf{x}) \cdot \mathbf{z}_k(\mathbf{x})$$



How can we determine the convolution coefficients?

Uncertainty-driven coefficients $\pi_k^\star(\mathbf{x}) \approx \dfrac{1/\operatorname{tr}(\boldsymbol{\Sigma}_k(\mathbf{x}))}{\sum_{j=1}^{K} 1/\operatorname{tr}(\boldsymbol{\Sigma}_j(\mathbf{x}))}$

Uncertainty-aware Similarity



Uncertainty-aware Similarity

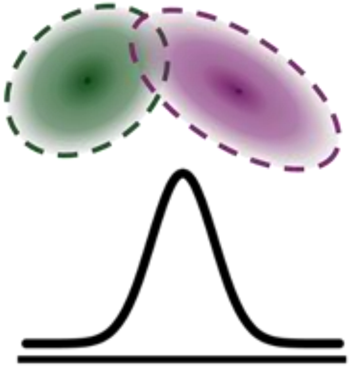
- Estimate the similarity $s = \mathbf{q}^\top \mathbf{c}$ between two ensembled embeddings $\mathbf{q}, \mathbf{c}$

$$\mu_s = \boldsymbol{\mu}_\mathbf{q}^\top \boldsymbol{\mu}_\mathbf{c}, \quad \sigma_s^2 = \boldsymbol{\mu}_\mathbf{q}^\top \boldsymbol{\Sigma}_\mathbf{c} \boldsymbol{\mu}_\mathbf{q} + \boldsymbol{\mu}_\mathbf{c}^\top \boldsymbol{\Sigma}_\mathbf{q} \boldsymbol{\mu}_\mathbf{c} + \mathrm{tr}(\boldsymbol{\Sigma}_\mathbf{q} \boldsymbol{\Sigma}_\mathbf{c})$$

- Incorporate **the estimated uncertainty** into the similarity ***without any sampling via probit approximation***

$$\hat{s} \approx \frac{\mu_s}{\sqrt{1 + \frac{\pi}{8}\boxed{\sigma_s^2}}}$$

$\boldsymbol{\mu}_\mathbf{q}, \boldsymbol{\Sigma}_\mathbf{q}$	Mean / Covariance of query
$\boldsymbol{\mu}_\mathbf{c}, \boldsymbol{\Sigma}_\mathbf{c}$	Mean / Covariance of document (corpus)



Probabilistic Cosine Similarity

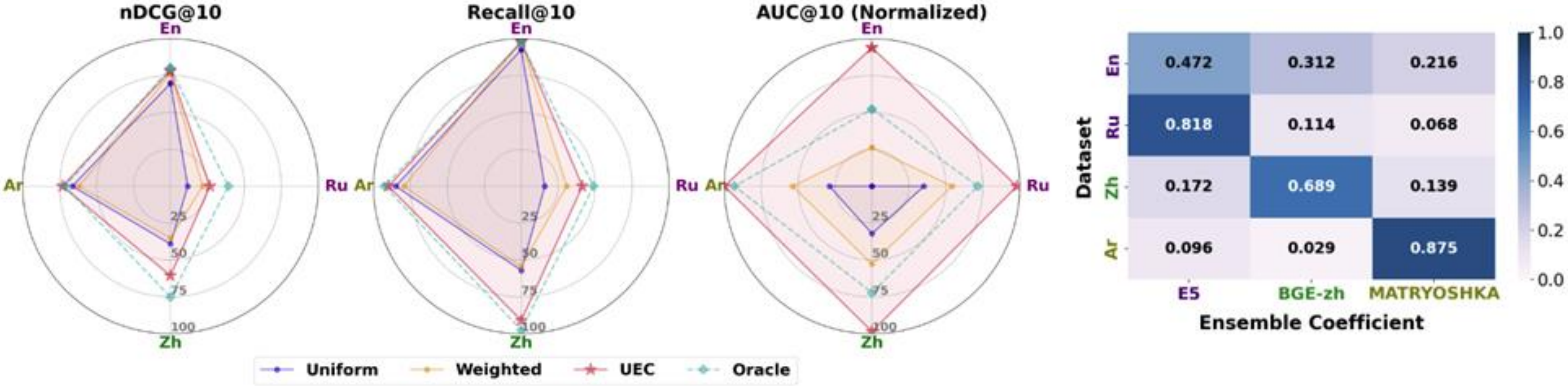
$s(\mathbf{q}, \mathbf{c}) \sim \mathcal{N}(\mu_s, \sigma_s^2)$

Calibrated Cosine Similarity

$$\hat{s} \approx \frac{\mu_s}{\sqrt{1 + \frac{\pi}{8}\sigma_s^2}}$$

Results

MIRACL Subset (Toy)

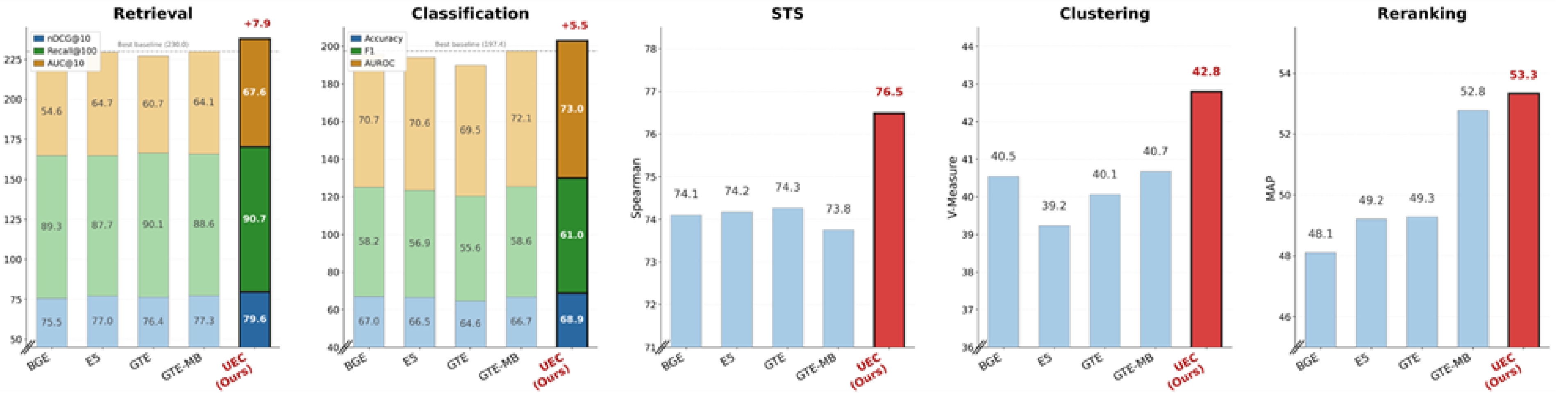


Multi-lingual Setting: Dataset with 4 languages / 3 language-specific models

- UEC achieves performance comparable to the oracle and even surpasses in some cases, with particularly strong gains in AUC@10.
- UEC computes ensemble coefficients adaptively modulated by the uncertainty of each embedding.

Results

MMTEB

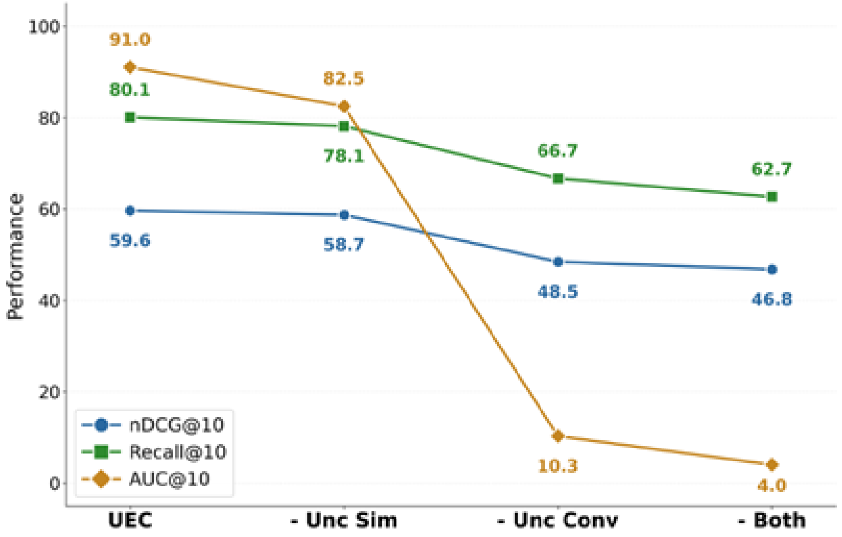


MMTEB (Benchmark) with 5 tasks

UEC achieves the highest average performance across all diverse metrics and tasks.

Comparison with other ensemble and ablation study

Property	Uniform	Weighted	UEC (Ours)
Auto Coefficient Selection	-	✗	✓
Data-wise Coefficient	✗	✗	✓
Uncertainty-aware Similarity	✗	✗	✓
Memory Complex.	$\mathcal{O}(KD)$	$\mathcal{O}(KD)$	$\mathcal{O}(KD)$
Time (Rel.)	1.000	1.000	1.006
Time (Complex.)	$\mathcal{O}(KD)$	$\mathcal{O}(KD)$	$\mathcal{O}(KD)$



- UEC shows practical advantages for real-world deployment.
- Both uncertainty convolution and similarity are essential.

Thank you!

Looking forward to connect with you :)