

**General topic:** Ensemble methods for combining multiple text embedding models in NLP.

**Specific activity:** How to leverage model-specific uncertainty when combining embeddings, rather than treating all models as equally reliable.

**Research questions:** The paper's central question is whether "a principled framework for combining multiple embedding models by modeling predictive uncertainty" can "consistently improve both performance and robustness by leveraging principled uncertainty modeling."

**Problem:** No single embedding model excels across all tasks. Existing ensemble approaches use deterministic aggregation like uniform averaging, ignoring individual model reliability.

**Claims:** Post-hoc Laplace approximation can convert deterministic embeddings into probabilistic ones; uncertainty-driven adaptive weighting outperforms fixed-weight ensembles; their similarity estimator approximates the squared 2-Wasserstein distance with bounded error at negligible cost.

**Evidence:** Evaluated on MIRACL (4 languages), and MMTEB benchmarks spanning retrieval (5 datasets), classification (5 datasets), STS (10 datasets), plus bitext mining, clustering, and reranking. UEC achieves the highest average scores across nearly all tasks compared to individual models, uniform/weighted ensembles, task arithmetic, Borda count, and GroVE.

**Statistical analysis:** Uses nDCG@10, Recall, Accuracy, F1, Spearman correlation, AUROC. Introduces Var-ECE to evaluate variance calibration. Includes ablation studies and hyperparameter sensitivity analysis. No statistical significance tests are reported.