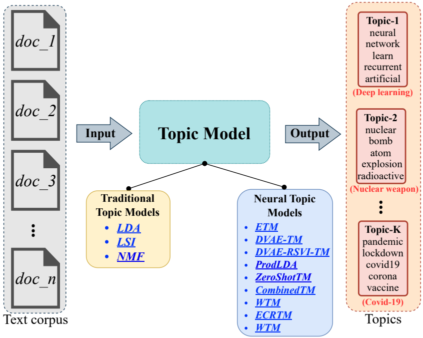
S2WTM: Spherical Sliced-Wasserstein Autoencoder for Topic Modeling

Suman Adhya, Debarshi Kumar Sanyal

School of Mathematical & Computational Sciences
Indian Association for the Cultivation of Science

ACL 2025
VIENNA
JULY 27 - AUGUST 1

What is a Topic Model?



- **Type**: Unsupervised learning;
- **Input**: Set of documents;
  Output: Set of topics;
- **Topic**: Distribution of the words;
- **Use cases**:
  1. Document clustering
  2. Corpus summarization
  3. Content recommendation
  4. Feature extraction
  5. Weakly supervised learning

Curse of Dimensionality - Gaussian Distribution

As the dimensionality $d \to \infty$, for $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, we have:

$$\|\mathbf{z}\| \xrightarrow{p} \sqrt{d} \quad \Rightarrow \quad \mathcal{N}(\mathbf{0}, \mathbf{I}_d) \approx \mathsf{Uniform}(\mathbb{S}^{d-1}(\sqrt{d}))$$

This geometric degeneration is a central aspect of the *curse of dimensionality*[1].

[1] Tim R. Davidson et al. "Hyperspherical Variational Auto-Encoders". In: *UAI*. 2018.

Limitation of KL-based Latent Alignment

$\mathcal{S}$-VAE replaces the Gaussian prior with a vMF prior to model hyperspherical latent spaces. However, it still uses KL divergence to align the posterior with the prior.

As shown in ELBO-surgery[2], this objective penalizes mutual information, often causing **posterior collapse**.

S2WTM addresses this by:

- Using a hyperspherical prior.
- Replacing KL with the Spherical Sliced-Wasserstein (SSW) distance.

This avoids posterior collapse while preserving meaningful latent representations.

[2] Matthew D Hoffman and Matthew J Johnson. "ELBO surgery: yet another way to carve up the variational evidence lower bound". In: *Workshop in advances in approximate Bayesian inference, NIPS*. 2016.

Hyperspherical Prior Distributions

von Mises–Fisher (vMF): Defined on the unit $(K-1)$-sphere $\mathcal{S}^{K-1}$:

$$\mathrm{vMF}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = c_K(\kappa) \exp(\kappa \boldsymbol{\mu}^\intercal \mathbf{x})$$

where $\boldsymbol{\mu}$ is the mean direction, κ controls concentration, and $c_K(\kappa)$ is a normalization constant.

Mixture of vMFs (MvMF): A weighted mixture of T vMF distributions:

$$\mathrm{MvMF}(\mathbf{x}) = \sum_{t=1}^{T} \alpha_t \, \mathrm{vMF}(\mathbf{x} \mid \boldsymbol{\mu}_t, \kappa_t)$$

with $\sum \alpha_t = 1$. Sampling involves selecting a component and drawing from its vMF.

Uniform on Sphere: Assigns equal density over $\mathcal{S}^{K-1}$. Sampling: draw $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, I)$ and normalize: $\mathbf{z} \leftarrow \mathbf{z}/\|\mathbf{z}\|$.

Spherical Sliced Wasserstein

The Spherical Sliced-Wasserstein (SSW) distance is a variant of the Sliced-Wasserstein (SW) distance, which itself is a computationally efficient approximation of the Wasserstein distance. SSW adapts SW to distributions supported on the unit hypersphere or other spherical domains.

Definition (Spherical Sliced Wasserstein)

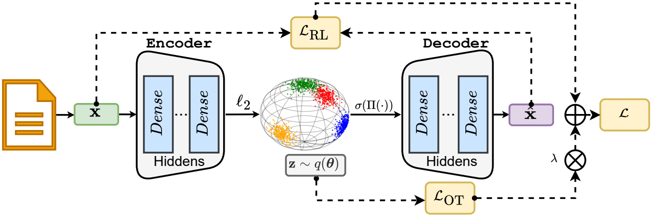
For $p \geq 1$, the p-SSW[a] between $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ is

$$\mathrm{SSW}_p^p(\mu, \nu) = \int_{S^{d-1}} W_p^p\Big((\tilde{\mathcal{R}}\mu)^\theta, (\tilde{\mathcal{R}}\nu)^\theta\Big) d\theta$$

where $(\tilde{\mathcal{R}}\eta)^\theta$ denotes the spherical Radon transform, which projects η onto a substructure of the sphere, such as a great circle.

[a] Clément Bonet et al. "Spherical Sliced-Wasserstein". In: *ICLR*. 2023.

Proposed Framework



Proposed framework for S2WTM.

$$\mathcal{L} = \mathcal{L}_{\mathrm{RL}} + \lambda \mathcal{L}_{\mathrm{OT}}$$

$$= \inf_{q(\boldsymbol{\theta}|\mathbf{x})} \mathbb{E}_{p(\mathbf{x})} \mathbb{E}_{q(\boldsymbol{\theta}|\mathbf{x})} [c(\mathbf{x}, \hat{\mathbf{x}})] + \lambda \, \mathrm{SSW}_2^2(q_{\boldsymbol{\theta}}, p_{\boldsymbol{\theta}})$$

where, $\mathrm{SSW}_2^2(q_{\boldsymbol{\theta}}, p_{\boldsymbol{\theta}}) \approx \frac{1}{M} \sum_{i=1}^{M} W_2^2\Big((\tilde{\mathcal{R}}_i q_{\boldsymbol{\theta}}), (\tilde{\mathcal{R}}_i p_{\boldsymbol{\theta}})\Big)$

Experimental Setup

All experiments are performed using **OCTIS**.

- **Datasets**:

Dataset	#Docs	#Labels	#Words
20NG	16309	20	1612
BBC	2225	5	2949
M10	8355	10	1696
SS	12295	8	2000
Pascal	4834	20	2630
Bio	19448	20	2000
DBLP	54595	4	1513

Statistics of the datasets used.

- **Baselines**:

1. **LDA**;
2. **LSI**;
3. **NMF**;
4. **ETM**;
5. **DVAE-TM**;
6. **DVAE-RSVI-TM**;
7. **ProdLDA**;
8. **ZeroshotTM**;
9. **CombinedTM**;
10. **WTM**;
11. **vONT**;
12. **ECRTM**;

Quantitative Analysis – Coherence Scores

Model Name	20NG		BBC		M10		SS		Pascal		Bio		DBLP	
	NPMI	CV	NPMI	CV	NPMI	CV	NPMI	CV	NPMI	CV	NPMI	CV	NPMI	CV
LDA	0.092	0.599	0.076	0.565	-0.047	0.369	-0.066	0.362	-0.072	0.356	0.019	0.444	0.015	0.348
LSI	0.006	0.457	0.064	0.539	0.001	0.390	-0.062	0.314	-0.045	0.293	-0.026	0.320	0.009	0.334
NMF	0.118	0.648	0.065	0.555	0.050	0.448	0.019	0.462	-0.042	0.378	0.100	0.537	0.016	0.354
ETM	0.066	0.564	0.070	0.581	0.018	0.374	0.000	0.420	-0.020	0.310	-0.062	0.182	-0.059	0.160
DVAE-TM	0.155	0.748	-0.032	0.530	-0.054	0.381	-0.175	0.357	0.000	0.421	0.113	0.540	-0.271	0.373
DVAE-RSVI-TM	0.146	**0.750**	-0.051,	0.523	-0.052	0.412	-0.192	0.411	-0.019	0.422	0.100	0.537	-0.269	0.356
ProdLDA	0.107	0.660	0.010	0.639	0.027	0.481	-0.009	0.560	-0.023	0.414	0.107	0.594	-0.065	0.472
ZeroshotTM	0.103	0.653	0.038	0.673	0.041	0.481	0.017	0.565	0.005	0.428	0.133	0.604	-0.062	0.474
CombinedTM	0.107	0.655	0.017	0.683	0.059	0.490	0.018	0.531	-0.002	0.421	0.133	0.608	-0.065	0.485
WTM	0.046	0.505	-0.006	0.454	-0.052	0.298	-0.013	0.405	-0.089	0.298	0.052	0.434	-0.044	0.202
vONT	0.045	0.505	-0.001	0.468	-0.053	0.301	-0.015	0.407	-0.090	0.302	0.052	0.442	-0.043	0.204
ECRTM	-0.089	0.416	0.170	0.804	-0.445	**0.516**	-0.333	0.423	-0.414	0.510	-0.421	0.510	-0.248	0.377
S2WTM	**0.167**	0.723	**0.252**	**0.863**	**0.101**	0.492	**0.146**	**0.683**	**0.045**	**0.572**	**0.191**	**0.663**	**0.133**	**0.558**

Median coherence scores over five runs per metric, with the highest values in **bold** and the second-highest values underlined.

Quantitative Analysis – Diversity Scores

Model Name	**20NG**		**BBC**		**M10**		**SS**		**Pascal**		**Bio**		**DBLP**	
	IRBO	**wI-C**	**IRBO**	**wI-C**	**IRBO**	**wI-C**	**IRBO**	**wI-C**	**IRBO**	**wI-C**	**IRBO**	**wI-C**	**IRBO**	**wI-C**
LDA	0.970	0.845	0.968	0.844	0.949	0.838	0.960	0.842	0.902	0.824	0.976	0.844	0.856	0.833
LSI	0.911	0.840	0.899	0.841	0.820	0.829	0.789	0.833	0.792	0.820	0.814	0.833	0.510	0.804
NMF	0.970	0.844	0.963	0.845	0.956	0.840	0.944	0.842	0.931	0.828	0.972	0.843	0.892	0.831
ETM	0.828	0.829	0.969	0.844	0.451	0.797	0.957	0.842	0.207	0.750	0.127	0.761	0.021	0.734
DVAE-TM	0.986	0.850	0.979	0.849	**1.000**	0.840	**1.000**	0.842	0.976	0.833	0.993	0.845	0.669	0.813
DVAE-RSVI-TM	0.987	0.850	0.994	0.849	**1.000**	0.840	**1.000**	0.842	0.978	0.833	0.996	0.845	0.546	0.827
ProdLDA	0.991	0.850	**1.000**	0.848	0.997	0.842	**1.000**	0.845	0.987	0.833	0.997	0.846	**1.000**	0.845
ZeroshotTM	0.991	0.850	**1.000**	0.849	**1.000**	0.842	**1.000**	0.844	0.987	0.834	0.996	0.845	**1.000**	0.845
CombinedTM	0.992	0.850	**1.000**	0.848	0.999	0.842	**1.000**	0.845	0.987	0.833	0.993	0.846	**1.000**	0.845
WTM	0.787	0.831	0.938	0.843	0.960	0.839	0.995	0.844	0.898	0.824	0.976	0.844	0.891	0.842
vONT	0.887	0.835	0.983	0.844	0.847	0.830	0.933	0.840	0.873	0.819	0.819	0.829	0.313	0.779
ECRTM	**0.996**	0.843	**1.000**	0.850	**1.000**	0.836	**1.000**	0.840	**1.000**	0.814	**1.000**	0.839	**1.000**	0.845
S2WTM	0.994	**0.881**	**1.000**	**0.854**	0.999	**0.857**	**1.000**	**0.851**	0.994	**0.868**	0.998	**0.859**	**1.000**	**0.847**

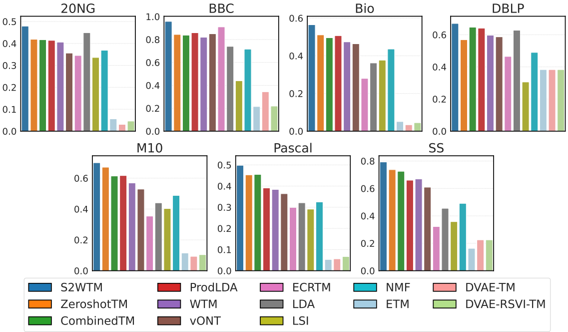
Median diversity scores over five runs per metric, with the highest values in **bold** and the second-highest values underlined.

Qualitative Analysis

Model	Topics
WTM	game, team, player, play, year, good, win, season, make, time drive, work, price, good, sell, make, buy, monitor, card, system key, government, system, encryption, make, program, time, space, chip, number
vONT	game, year, good, time, make, team, car, win, play, player drive, problem, card, run, system, disk, scsi, driver, bus, window key, chip, encryption, government, bit, system, make, clipper, time, phone
ECRTM	baseball, fan, pitcher, player, expansion, pitch, league, draft, suck, apple scsi, card, ide, pin, ranger, modem, port, ram, mouse, disk illegal, transmit, patient, warrant, taxis, fund, restriction, secret, budget, crack
S2WTM	game, team, win, score, player, playoff, goal, play, stat, season drive, card, scsi, ide, bus, controller, driver, disk, system, ram encryption, secure, chip, encrypt, phone, secret, communication, clipper, agency, security

The three topics broadly represent sports, hardware, and encryption/security from the 20NG dataset, with unrelated words highlighted in blue

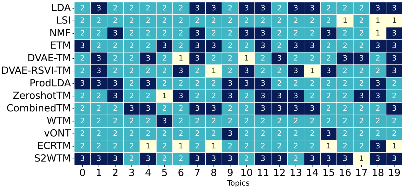
Extrinsic Evaluation – Document Classification



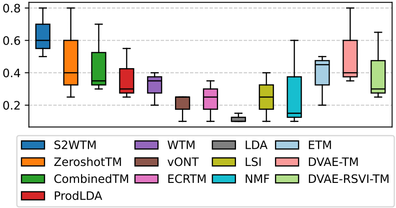
Median document classification accuracy of the models for all datasets.

S2WTM learns more discriminative *document-topic* embeddings, leading to superior classification accuracy on all datasets.

LLM-based Topic Quality Assessment



(a) LLM-assigned coherence ratings (1–3) for topics from all models.



(b) Accuracy in detecting intruder words across three test runs.

LLM-based evaluation approximates human judgment on **topic coherence** and **intrusiveness**[3], using `GPT-4` on the **20NG** dataset.

- **Rating:** S2WTM scores highest (mean: 2.7, median: 3.0).
- **Intrusion:** S2WTM achieves top accuracy across 3 runs.

[3] Dominik Stammbach et al. "Revisiting Automated Topic Model Evaluation with Large Language Models". In: *EMNLP*. 2023.

Beyond Euclidean Assumptions: Empirical Insights

Dataset	Coherence		Diversity	
	NPMI	CV	IRBO	wI-C
20NG	0.108 (0.167)	0.623 (0.723)	0.984 (0.994)	0.821 (0.881)
BBC	0.081 (0.252)	0.592 (0.863)	0.941 (1.000)	0.835 (0.854)
M10	0.038 (0.101)	0.494 (0.492)	0.964 (0.999)	0.826 (0.857)
SS	0.024 (0.146)	0.549 (0.683)	0.940 (1.000)	0.809 (0.851)
Pascal	0.007 (0.045)	0.425 (0.572)	0.887 (0.994)	0.826 (0.868)
Bio	0.094 (0.191)	0.572 (0.663)	0.972 (0.998)	0.818 (0.859)
DBLP	0.004 (0.133)	0.369 (0.558)	0.658 (1.000)	0.838 (0.847)

Comparison of *Coherence* and *Diversity* scores with and without hyperspherical latent space

Setup: Replace S2WTM's hyperspherical prior with a Dirichlet prior and use standard SW instead of SSW.

Result: Spherical modeling improves both coherence and diversity across datasets.

Conclusion

Introducing S2WTM, WAE-based topic model that models the hyperspherical latent space while effectively mitigating the posterior collapse problem. In most experiments, S2WTM consistently achieved superior performance compared to competitive topic models from the literature.

Thank you for listening...