

Studies on Topic Models of Document Corpora

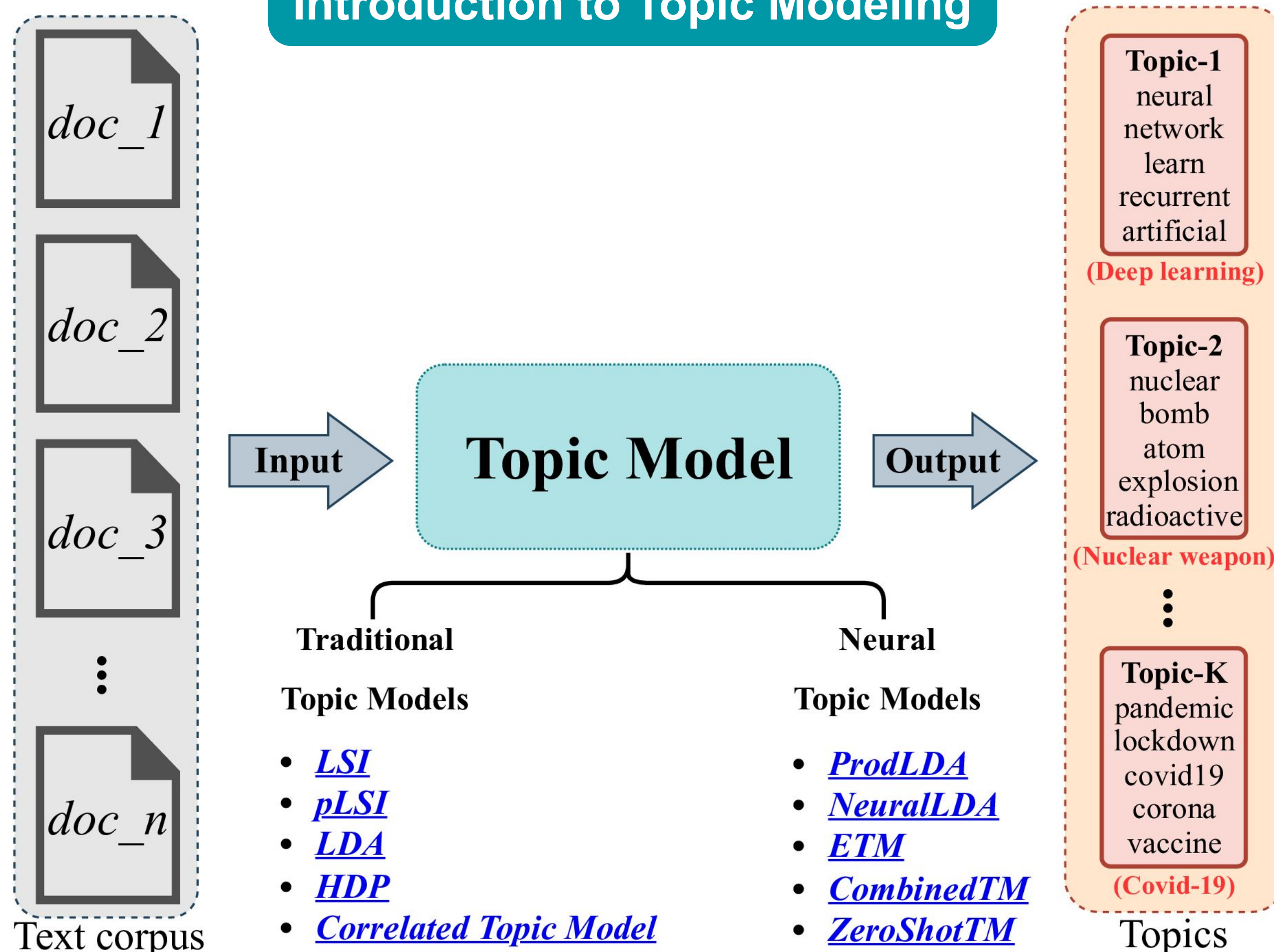
Suman Adhya

Supervised by: Dr. Debarshi Kumar Sanyal



IACS, Kolkata

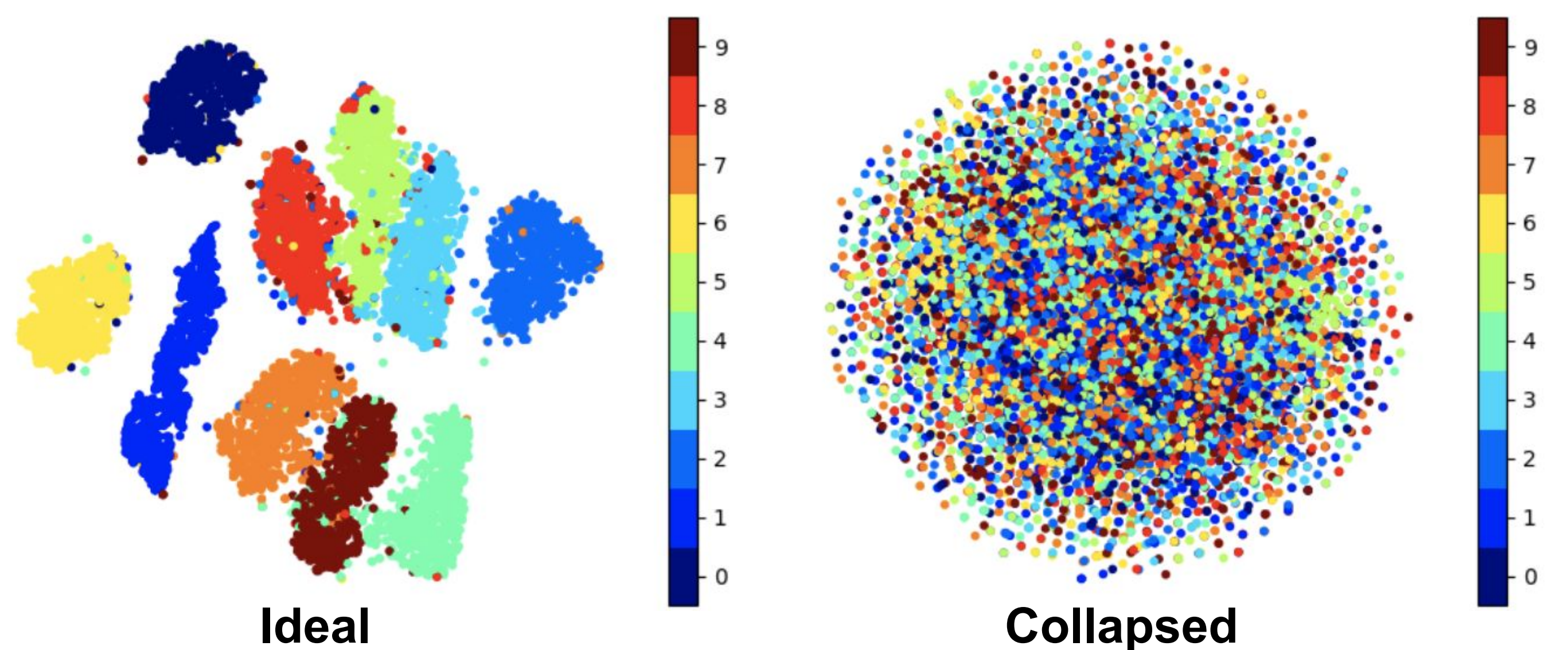
Introduction to Topic Modeling



Key Objectives of this Doctoral Work:

- ★ Analyze and address issues in existing topic models.
- ★ Build a student model by distilling knowledge from a complex teacher model, without compromising performance.
- ★ Introduce unsupervised negative sampling generation technique to enhance the topic representations.
- ★ Address posterior collapse issue in topic models.
- ★ Develop a neural dynamic topic model for sequential document corpus to capture the evolution of topics.

Ongoing Project: Mitigating Posterior Collapse

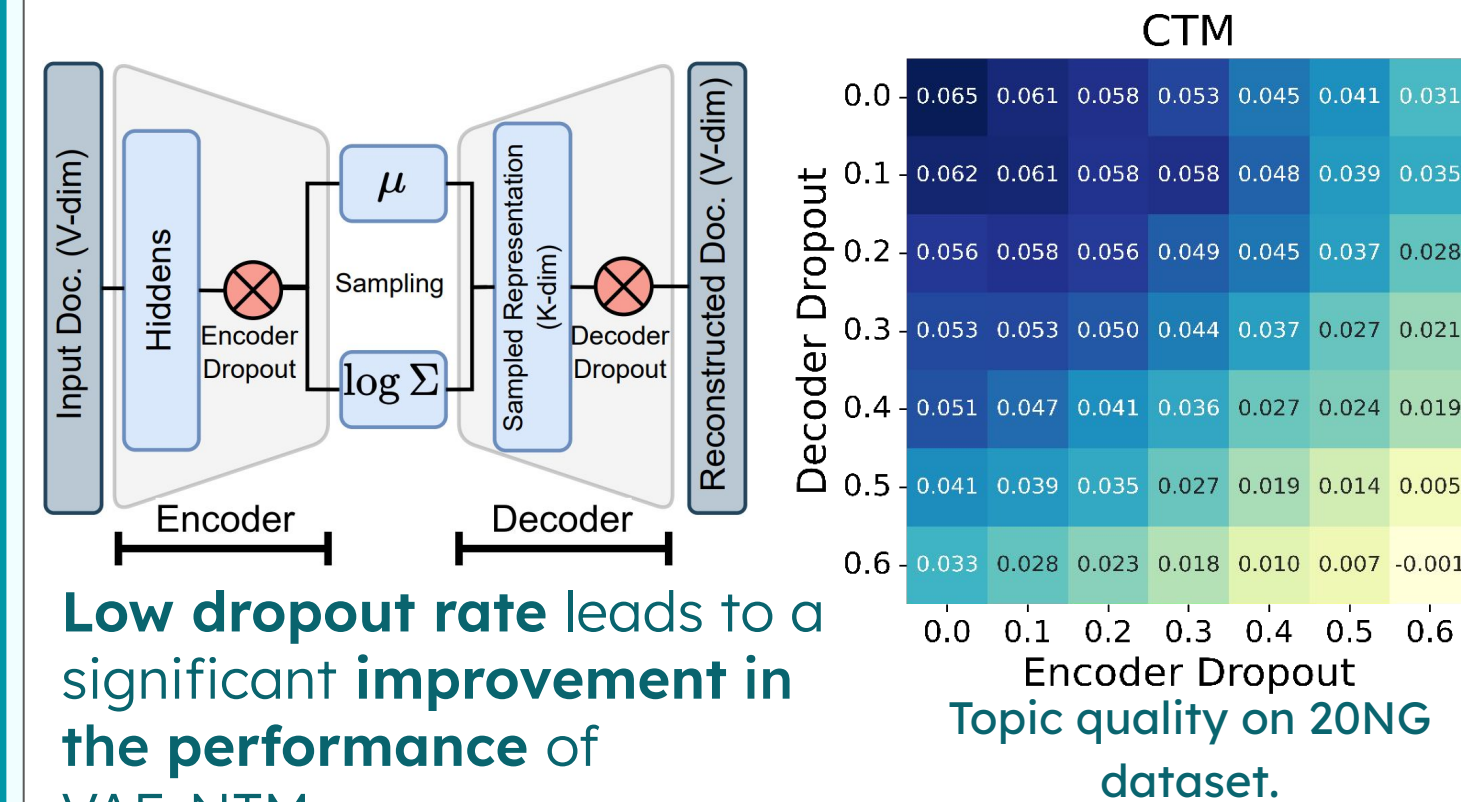


Jha et al. (CVPR, 2018)

“Posterior Collapse” occurs when the encoder-generated representation coincides with the prior, losing information from the input data latent structure.

Published Works

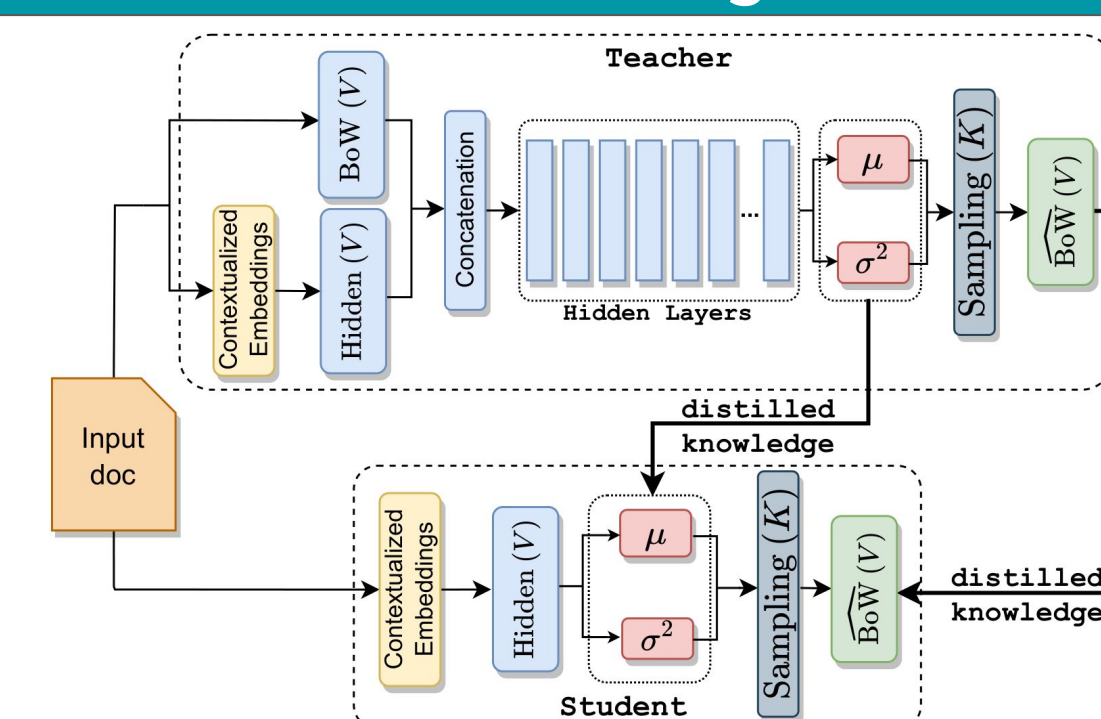
Do Neural Topic Models Really Need Dropout? Analysis of the Effect of Dropout in Topic Modeling



Suman Adhya
Avishek Lahiri
Debarshi Kumar Sanyal

EACL
2023

Improving Neural Topic Models with Wasserstein Knowledge Distillation

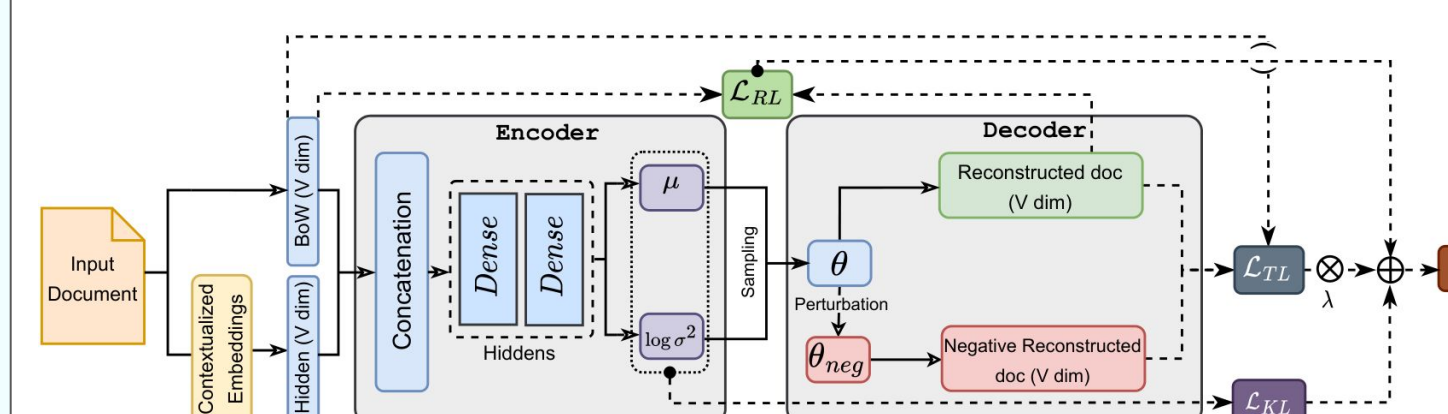


Introduced a knowledge distillation framework using **2-Wasserstein loss** to compress neural topic models without performance compromise.

Suman Adhya
Debarshi Kumar Sanyal

ECIR
2023

Improving Contextualized Topic Models with Negative Sampling



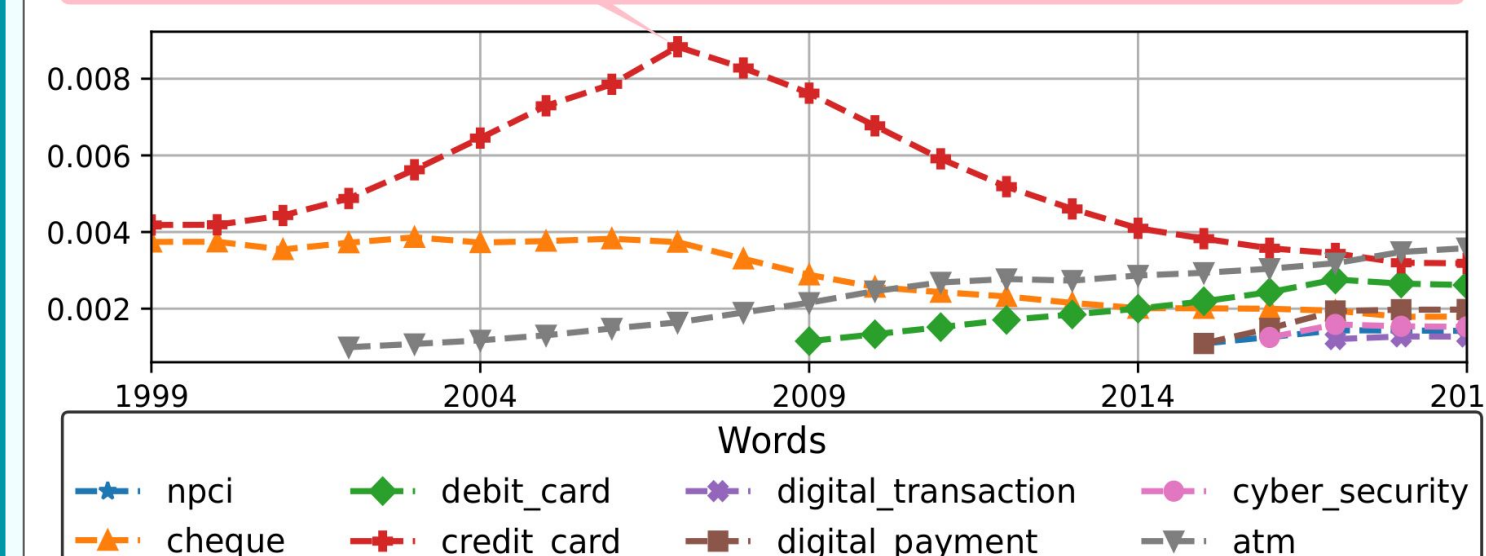
Suggested a **simple yet effective method** for generating unsupervised negative samples to enhance the quality of topic representations.

Suman Adhya
Avishek Lahiri
Debarshi Kumar Sanyal
Partha Pratim Das

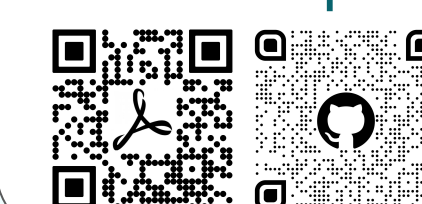
ICON

What Does the Indian Parliament Discuss? An Exploratory Analysis of the Question Hour in the Lok Sabha

Doc_1 : whether it has come to the notice of the Government that credit card number and CW number have been misused for fraudulent buying of e-tickets and other economic crimes;
Doc_2 : the details of Banks which are issuing Kisan Credit Cards in the State of Andhra Pradesh;
Doc_3 : whether the Government is aware of the credit card frauds being committed by well known internet service providers;



Illustrated the **temporal evolution of topics** for three important ministries.



Suman Adhya
Debarshi Kumar Sanyal

EUROPEAN
LANGUAGE
RESOURCES