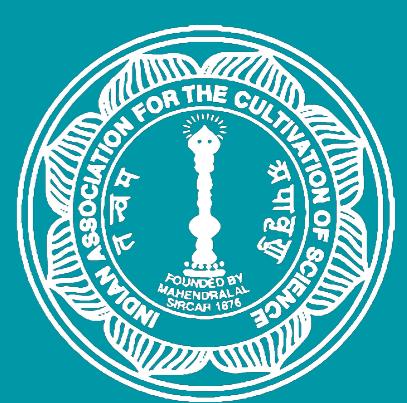


Studies on Topic Models of Document Corpora

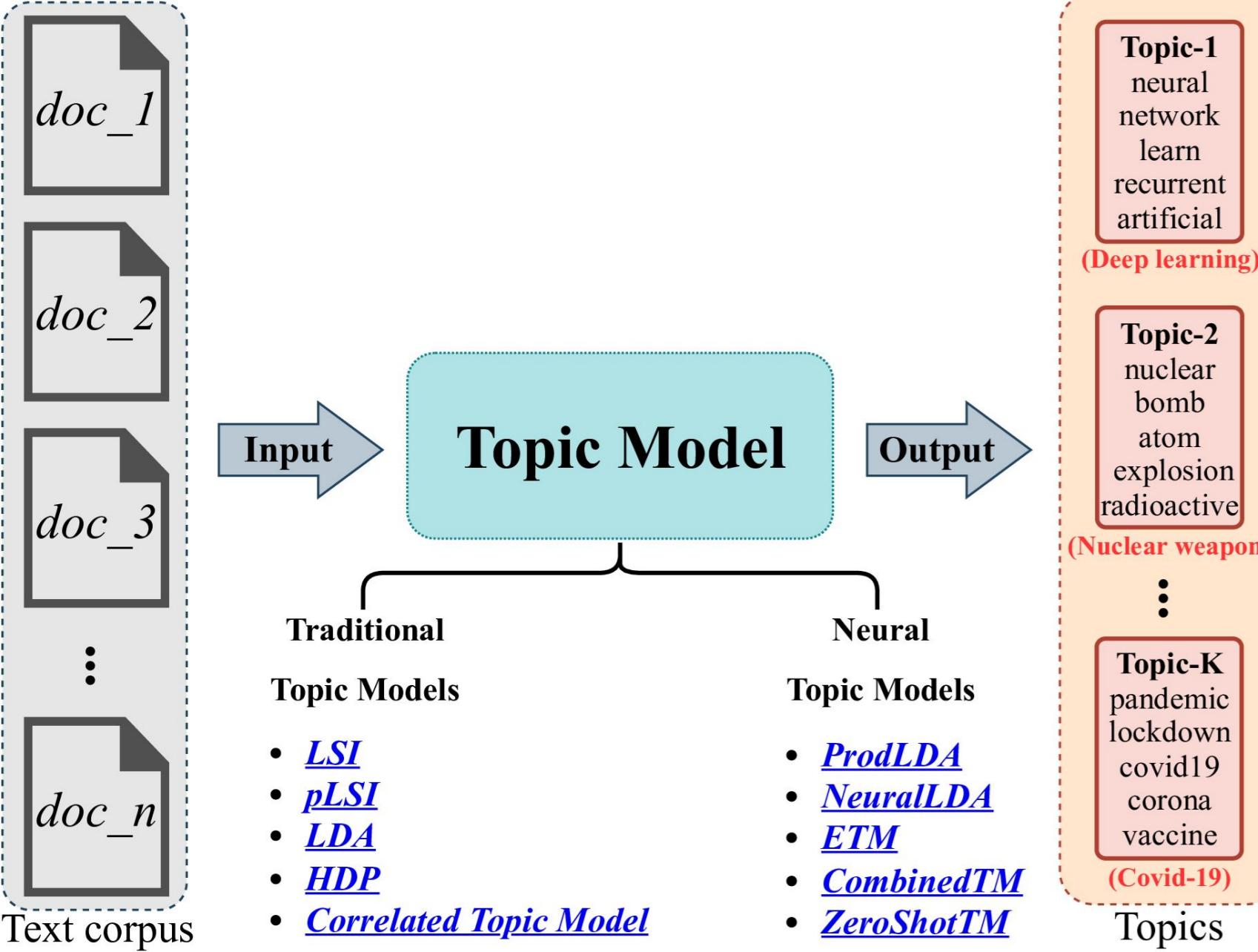


IACS, Kolkata

Suman Adhya

Supervised by: Dr. Debarshi Kumar Sanyal

Introduction to Topic Modeling



Key Objectives of this Doctoral Work:

- ★ Evaluate and improve negative sampling techniques to enhance topic quality, proposing a simple yet effective approach.
- ★ Incorporate document representations to refine topic modeling performance, leveraging graph-based approaches.
- ★ Develop a knowledge distillation framework to compress NTMs while maintaining their performance.
- ★ Analyze the role of dropout in NTMs and its impact on stability and performance.
- ★ Design an LLM-based dynamic topic model to capture topic evolution.
- ★ Investigate and mitigate posterior collapse in VAE-based topic models using Spherical Sliced-Wasserstein distance.

Evaluating Negative Sampling Approaches for Neural Topic Models

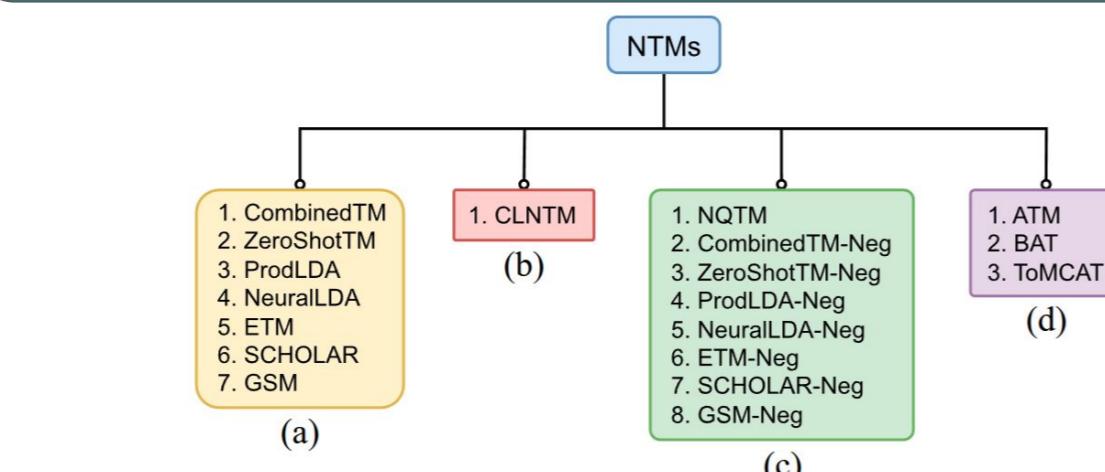
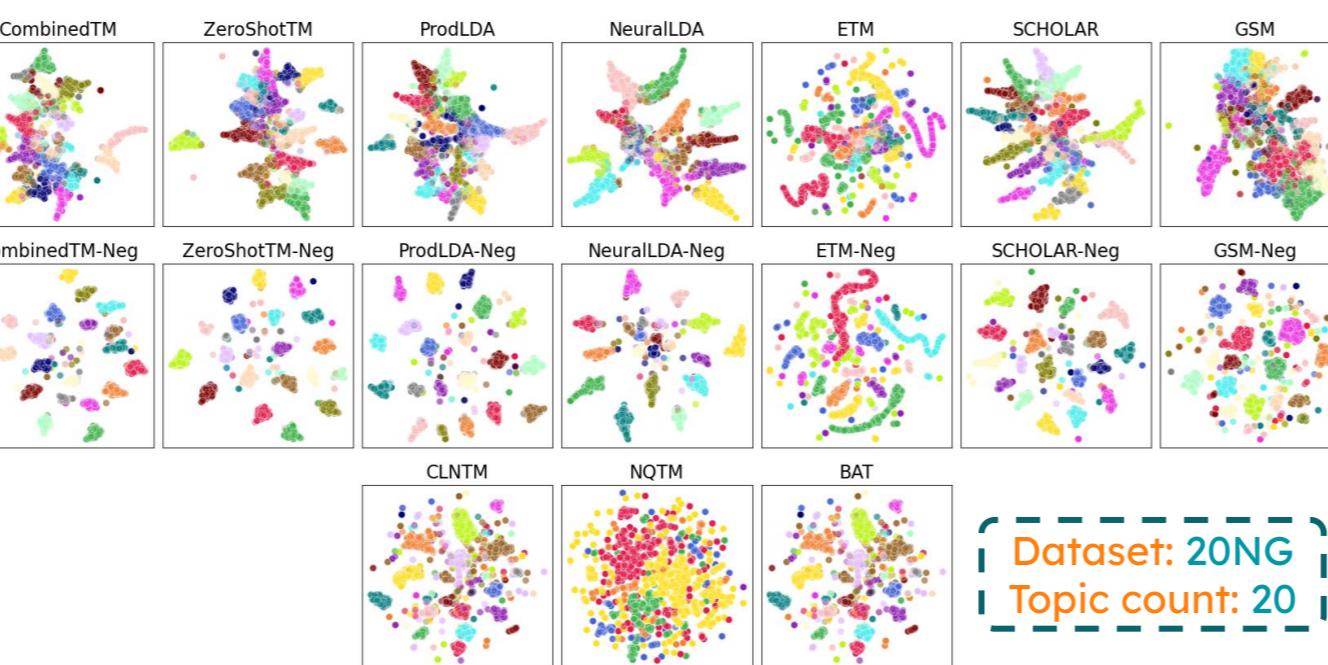


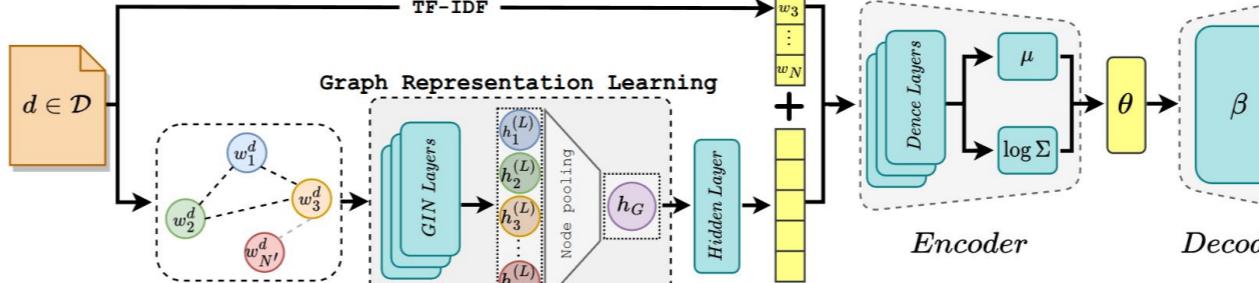
Fig. 1. Topic model categories based on negative sampling techniques.
(a) VAE-NTM. (b) VAE-ENS. (c) VAE-DNS. (d) Adv-Net.



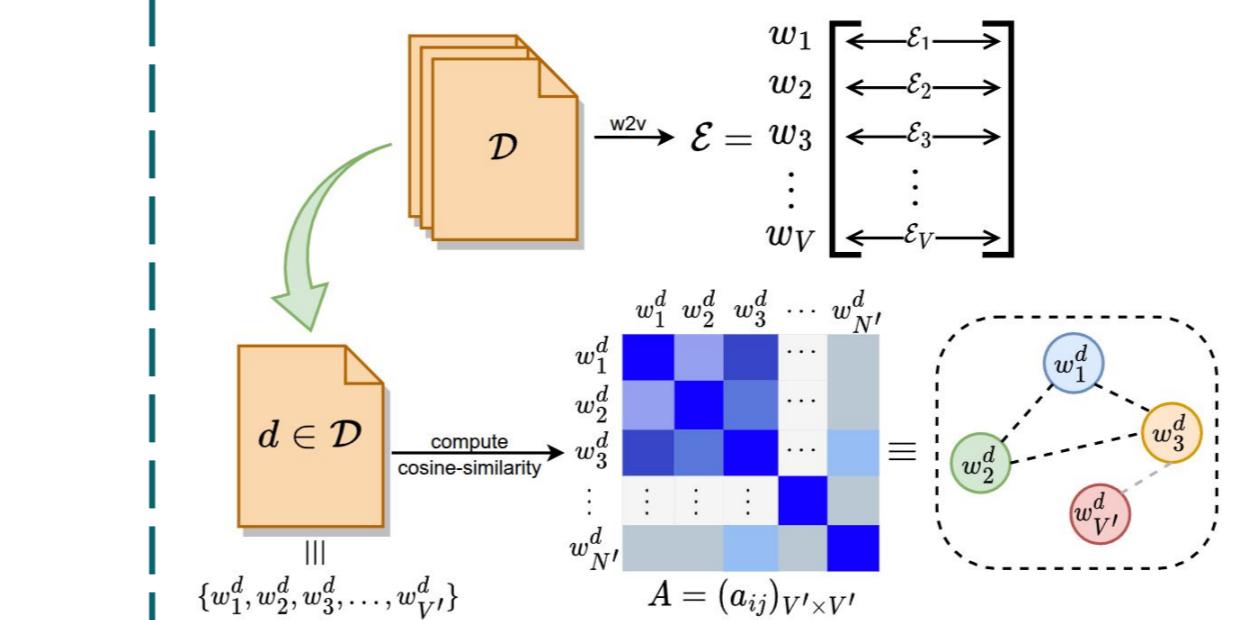
Suman Adhya
Avishek Lahiri
Debarshi Kumar Sanyal
Partha Pratim Das

Published Works

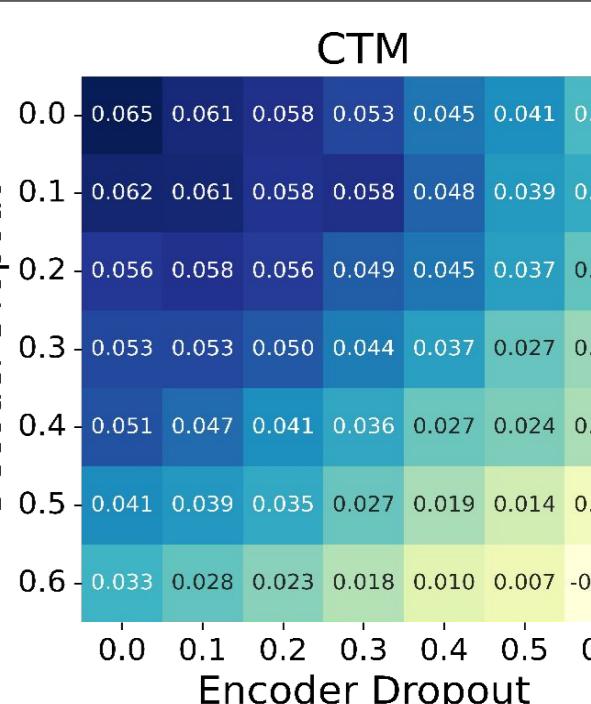
GINopic: Topic Modeling with Graph Isomorphism Network



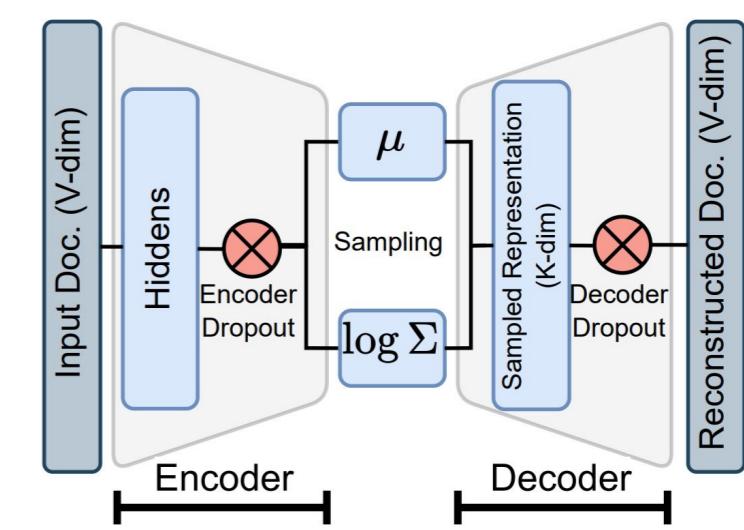
Graph Construction Methodology



Suman Adhya
Debarshi Kumar Sanyal

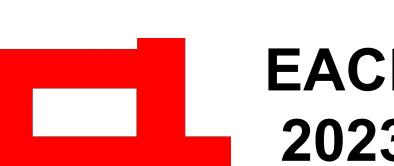


Do Neural Topic Models Really Need Dropout? Analysis of the Effect of Dropout in Topic Modeling

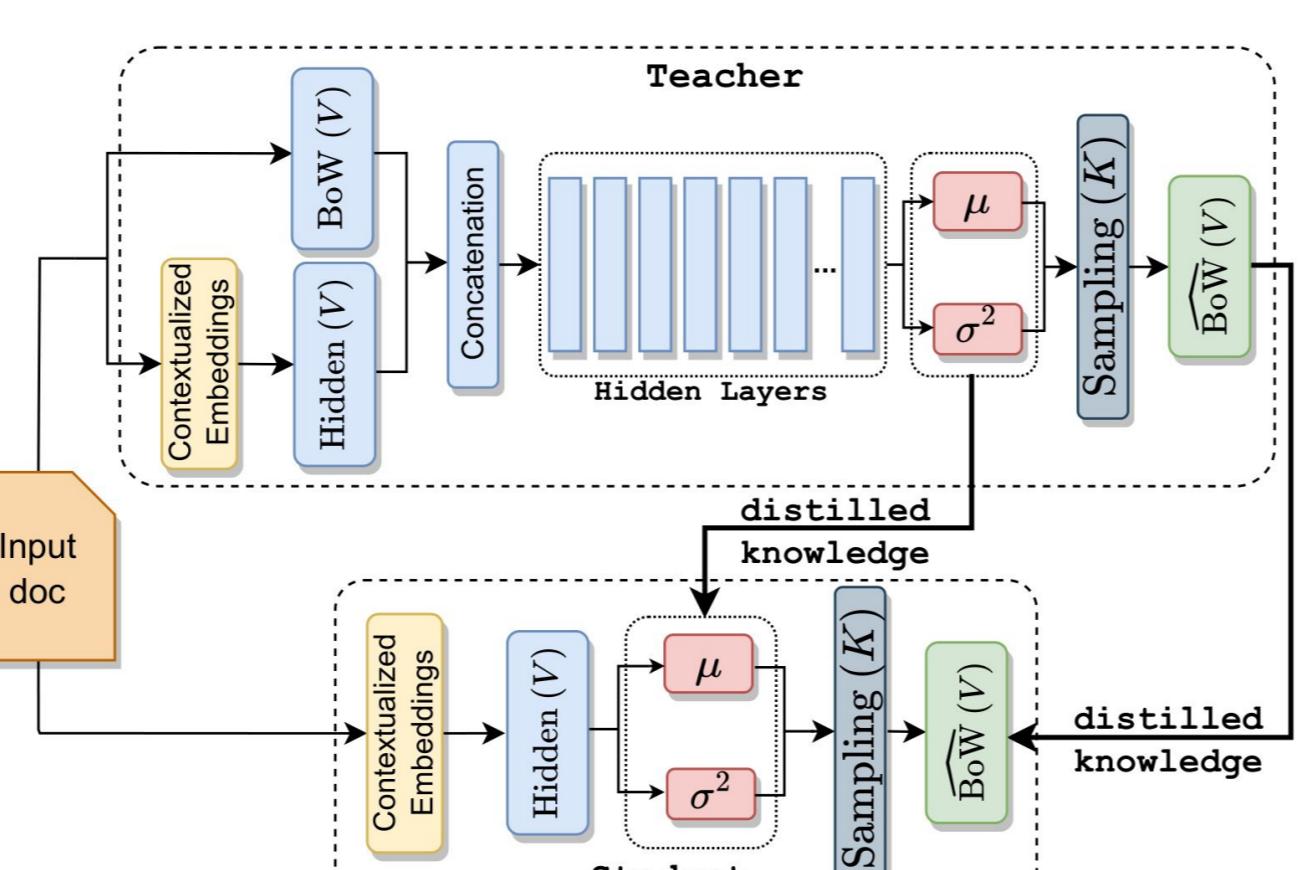


Low dropout rate leads to a significant improvement in the performance of VAE-NTMs.

Suman Adhya
Avishek Lahiri
Debarshi Kumar Sanyal



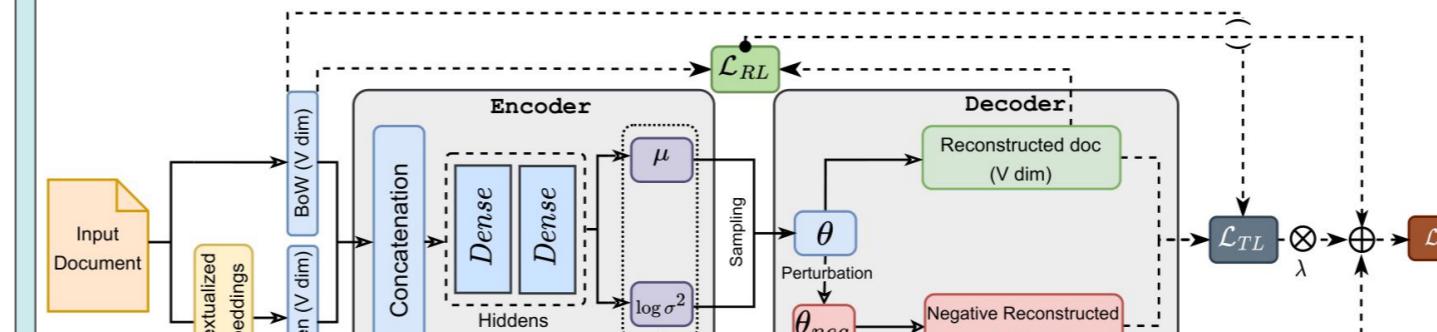
Improving Neural Topic Models with Wasserstein Knowledge Distillation



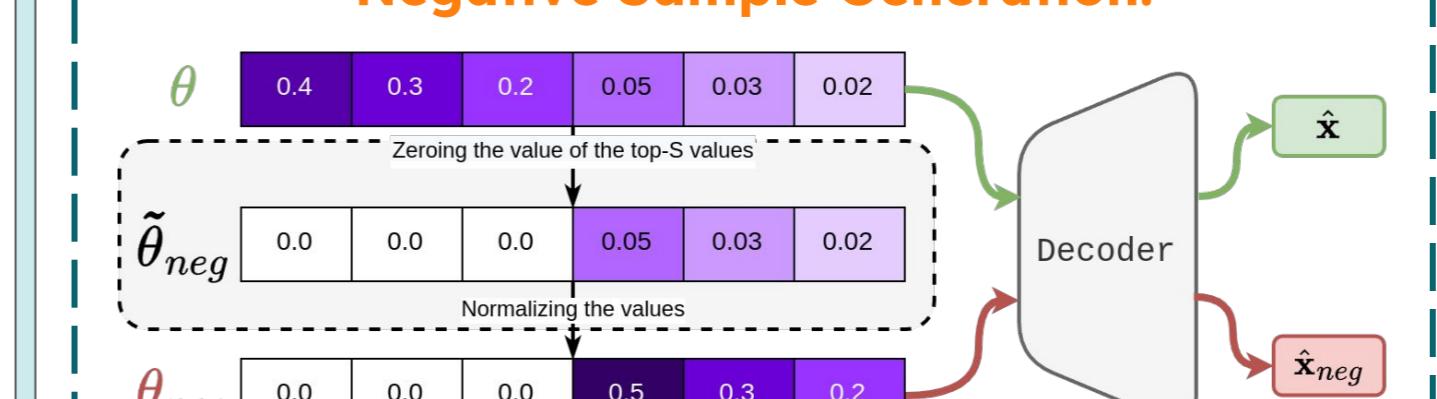
Introduced a knowledge distillation framework using 2-Wasserstein loss to compress neural topic models without performance compromise.

Suman Adhya
Avishek Lahiri
Debarshi Kumar Sanyal

Improving Contextualized Topic Models with Negative Sampling



Negative Sample Generation:

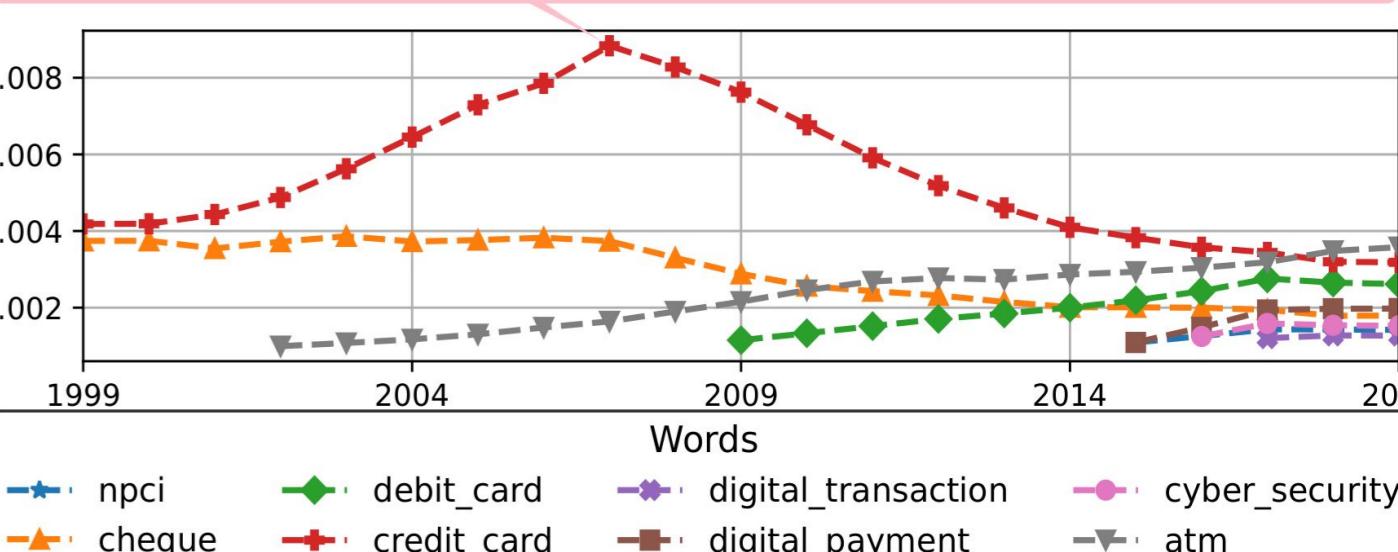


Proposed a simple yet effective method for generating unsupervised negative samples to enhance the quality of topic representations.

Suman Adhya
Avishek Lahiri
Debarshi Kumar Sanyal
Partha Pratim Das

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

