Effective Neural Topic Modeling with Embedding Clustering Regularization

Xiaobao Wu ¹ Xinshuai Dong ² Thong Nguyen ³ Anh Tuan Luu ¹

Abstract
Topic models have been prevalent for decades with various applications. However, existing topic models commonly suffer from the notorious topic collapsing: discovered topics semantically collapse towards each other, leading to highly repetitive topics, insufficient topic discovery, and damaged model interpretability. In this paper, we propose a new neural topic model, Embedding Clustering Regularization Topic Model (ECRTM). Besides the existing reconstruction error, we propose a novel Embedding Clustering Regularization (ECR), which forces each topic embedding to be the center of a separately aggregated word embedding cluster in the semantic space. This enables each produced topic to contain distinct word semantics, which alleviates topic collapsing. Regularized by ECR, our ECRTM generates diverse and coherent topics together with high-quality topic distributions of documents. Extensive experiments on benchmark datasets demonstrate that ECRTM effectively addresses the topic collapsing issue and consistently surpasses state-of-the-art baselines in terms of topic quality, topic distributions of documents, and downstream classification tasks.

1. Introduction
Topic models have achieved great success in document analysis via discovering latent semantics. They have facilitated various downstream applications (Boyd-Graber et al., 2017), like content recommendation (McAuley & Leskovec, 2013), summarization (Ma et al., 2012), and information retrieval (Wang et al., 2007). Current topic models can be roughly classified as two lines: (1) conventional topic models with probabilistic graphical models (Blei et al., 2003) or matrix factorization (Kim et al., 2015; Shi et al., 2018) and (2) neural topic models (Miao et al., 2016; 2017; Srivastava & Sutton, 2017; Gupta et al., 2019).

However, despite the current achievements, existing topic models commonly suffer from topic collapsing: the discovered topics tend to semantically collapse towards each other (Srivastava & Sutton, 2017), as exemplified in Table 1. We see these collapsed topics include many uninformative and repetitive words. This brings about several issues: (1) Topic collapsing results in highly repetitive topics, which are less useful for downstream applications (Wallach et al., 2009; Nan et al., 2019). (2) Topic collapsing incurs insufficient topic discovery. Many latent topics are undisclosed, making the topic discovery insufficient to understand documents (Dieng et al., 2020). (3) Topic collapsing damages the interpretability of topic models. It becomes difficult to infer the real underlying topics that a document contains (Huynh et al., 2020). In consequence, topic collapsing impedes the utilization and extension of topic mod-

Table 1: Top related words of the discovered topics by NSTM (Zhao et al., 2021b) on IMDB. These topics semantically collapse towards each other with many uninformative and repetitive words. Repetitions are underlined.

| Topic#1: just show even come time one good really going know | Topic#2: just even really something come going like actually things get | Topic#3: just one even something come way really like always good | Topic#4: just get going come one know even really something way | Topic#5: just like inside get even look come one everything away |

Figure 1: t-SNE visualization of word embeddings (●) and topic embeddings (▲) under 50 topics. These show while the topic embeddings mostly collapse together in previous state-of-the-art models, our ECRTM successfully avoids the collapsing.
els; therefore it is crucial to overcome this challenge for building effective topic models.

To address the topic collapsing issue and achieve effective topic modeling, we in this paper propose a novel neural topic model, Embedding Clustering Regularization Topic Model (ECRTM). First, we illustrate the reason for topic collapsing: Figures 1a to 1c show topic embeddings mostly collapse together in the semantic space of previous state-of-the-art methods. This makes discovered topics contain similar word semantics and thus results in the topic collapsing. Then to avoid the collapsing of topic embeddings, we propose the novel Embedding Clustering Regularization (ECR) besides the reconstruction error of previous work. ECR regularizes topic embeddings as cluster centers and word embeddings as cluster samples. For effective regularization, ECR models the clustering soft-assignments between them by solving a specifically defined optimal transport problem on them. As such, ECR forces each topic embedding to be the center of a separately aggregated word embedding cluster. Instead of collapsing together, this makes topic embeddings away from each other and cover different semantics of word embeddings. Thus our ECR enables each produced topic to contain distinct word semantics, which alleviates topic collapsing. Regularized by ECR, our ECRTM achieves effective topic modeling performance by producing diverse and coherent topics together with high-quality topic distributions of documents. Figure 1d shows the effectiveness of ECRTM. We conclude the main contributions of our paper as follows:\footnote{Our code is available at \url{https://github.com/bobxwu/ECRTM}}:

- We propose a novel embedding clustering regularization that avoids the collapsing of topic embeddings by forcing each topic embedding to be the center of a separately aggregated word embedding cluster, which effectively mitigates topic collapsing.

- We further propose a new neural topic model that jointly optimizes the topic modeling objective and the embedding clustering regularization objective. Our model can produce diverse and coherent topics with high-quality topic distributions of documents at the same time.

- We conduct extensive experiments on benchmark datasets and demonstrate that our model effectively addresses the topic collapsing issue and surpasses state-of-the-art baseline methods with substantially improved topic modeling performance.

2. Related Work

Conventional Topic Models Conventional topic models (Hofmann, 1999; Blei et al., 2003; Yao et al., 2014; Das et al., 2015) mostly employ probabilistic graphical models to model the generation of documents with topics as latent variables. They infer model parameters with MCMC methods like Gibbs sampling (Steyvers & Griffiths, 2007) or Variational Inference (Blei et al., 2017). Some studies use matrix factorization to model topics (Yan et al., 2013b; Kim et al., 2015; Shi et al., 2018). Many various scenarios have been developed, like short texts (Yan et al., 2013a; Wu & Li, 2019), multilingual (Mimno et al., 2009), and dynamic topic models (Blei & Lafferty, 2006). These methods commonly need model-specific derivations for different modeling assumptions.

Neural Topic Models Due to the success of Variational AutoEncoder (VAE, Kingma & Welling, 2014; Rezende et al., 2014), several neural topic models have been proposed (Miao et al., 2016; Srivastava & Sutton, 2017; Ding et al., 2019; Meng et al., 2020; Nguyen & Luu, 2021; Wu et al., 2021; 2022; 2023). Different from conventional ones, neural topic models can directly apply gradient back-propagation, which enhances flexibility and scalability. Alternatively, some studies directly cluster pre-trained word or sentence embeddings to produce topics (Sia et al., 2020; Zhang et al., 2022), but they are not topic models since they cannot infer the topic distributions of documents as required. Recent state-of-the-art NSTM (Zhao et al., 2021b) and WeTe (Wang et al., 2022) measure the reconstruction error with optimal (conditional) transport distance. However, they still suffer from topic collapsing (see Sec. 4.2). Different from these, our proposed model aims to address the topic collapsing issue and achieve effective neural topic modeling. Besides the reconstruction error of these previous work, we propose a novel embedding clustering regularization that avoids the collapsing of topic embeddings by forcing each topic embedding to be the center of a separately aggregated word embedding cluster. Then our model learns topics under this effective regularization and particularly addresses the topic collapsing issue.

3. Methodology

3.1. Problem Setting and Notations

We recall the problem setting of topic modeling following LDA (Blei et al., 2003). Consider a document collection $X$ with $V$ unique words (vocabulary size), and each document is denoted as $x$. We require to discover $K$ latent topics from this document collection. The $k$-th topic is defined as a distribution over all words (topic-word distribution), denoted as $\beta_k \in \mathbb{R}^V$. We have $\beta = (\beta_1, \ldots, \beta_K) \in \mathbb{R}^{V \times K}$ as the topic-word distribution matrix of all topics. The topic distribution of a document (doc-topic distribution) refers to what topics it contains, denoted as $\theta \in \Delta_K$. Here $\Delta_K$ denotes a probability simplex $\Delta_K = \left\{ \theta \in \mathbb{R}^+_K \mid \sum_{k=1}^K \theta_k = 1 \right\}$.\footnote{Our code is available at \url{https://github.com/bobxwu/ECRTM}}
Effective Neural Topic Modeling with Embedding Clustering Regularization

3.2. What Causes Topic Collapsing?

Despite the current achievements, most topic models suffer from topic collapsing: topics semantically collapse towards each other (see Table 1). We illustrate what causes topic collapsing by analyzing a kind of recently proposed state-of-the-art neural topic models (Dieng et al., 2020; Zhao et al., 2021b). These models compute the topic-word distribution matrix with two parameters: \( \beta = W^\top T \). Here \( W = (w_1, \ldots, w_V) \in \mathbb{R}^{D \times V} \) are the embeddings of \( V \) words, and \( T = (t_1, \ldots, t_K) \in \mathbb{R}^{D \times K} \) are the embeddings of \( K \) topics, all in the same \( D \)-dimensional semantic space. They can facilitate learning by initializing \( W \) with pre-trained embeddings like GloVe (Pennington et al., 2014).

However, topic collapsing commonly happens in these state-of-the-art models. We believe the reason lies in that their reconstruction error minimization incurs the collapsing of topic embeddings. Specifically, these models learn topic and word embeddings by minimizing the reconstruction error between topic distribution \( \theta \) and word distribution \( x \) of a document. For example, to measure reconstruction error, ETM (Dieng et al., 2020) uses traditional expected log-likelihood, and recent NSTM (Zhao et al., 2021b) and WeTe (Wang et al., 2022) use optimal (conditional) transport distance. In fact, words in a document collection commonly are long-tail distributed following Zipf’s law (Reed, 2001; Piantadosi, 2014)—roughly speaking, few words are of high frequency and most are of low frequency. Therefore the reconstruction is biased as it mainly reconstructs high-frequency words regardless of the reconstruction error measurements. Since topic and word embeddings are learned by minimizing reconstruction error, this biased reconstruction pushes most topic embeddings close to the embeddings of some high-frequency words in the semantic space. As a result, topic embeddings collapse together in these state-of-the-art methods as shown in Figure 1. The topic-word distributions become similar to each other, leading to topic collapsing. We empirically demonstrate this argument by removing high-frequency words (See experiments in Sec. 4.6).

3.3. How to Design An Effective Regularization on Embeddings?

In this section, we discuss how to design an effective regularization on embeddings for the topic collapsing issue.

Our analysis in Sec. 3.2 indicates topic collapsing happens because the reconstruction error minimization incurs the collapsing of topic embeddings in existing work. To address this issue, we propose to design a clustering regularization on embeddings in addition to the reconstruction error of existing work. We consider topic embeddings as cluster centers and word embeddings as cluster samples; then we require the regularization to force each topic embedding to be the center of a separately aggregated word embedding cluster. As such, instead of collapsing together, topic embeddings are away from each other and cover different semantics of word embeddings in the space. This will make each discovered topic contain distinct word semantics and thus alleviate topic collapsing. However, it is non-trivial to design such an effective regularization. We explore the requirements as follows.

Supporting Joint Optimization As we regularize on neural topic models, we require the clustering regularization to support joint optimization on topic and word embeddings along with a neural topic modeling objective. Some studies (Sia et al., 2020) apply classic clustering algorithms, e.g., KMeans and GMM, to produce topics by clustering pre-trained word embeddings. We clarify that they are not topic models as they only produce topics and cannot learn doc-topic distributions as required (but we...
Algorithm 1 Training algorithm for ECRTM.

Input: document collection \( \mathbf{X} \), number of epochs \( n_{\text{epoch}} \).
Output: model parameters \( \Theta \), \( \mathbf{W}, \mathbf{T} \);
1: for 1 to \( n_{\text{epoch}} \) do
2: for each mini-batch \( \{ \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(N)} \} \) from \( \mathbf{X} \) do
3: //Sinkhorn’s algorithm;
4: \( C_{jk} = ||w_j - t_k||^2 \) \( \forall j,k \);
5: \( M = \exp( -C/\epsilon ) \);
6: Initialize \( \mathbf{b} \leftarrow \mathbf{1}_K \);
7: while not converged and not reach max iterations do
8: \( \mathbf{a} \leftarrow \frac{1}{\mathbf{K}^\top \mathbf{a}} ; \mathbf{b} \leftarrow \frac{\mathbf{b}}{\mathbf{a}^\top \mathbf{1}} ; \)
9: end while
10: Compute \( \pi_j^k \leftarrow \text{diag} ( \mathbf{a}) \text{diag} ( \mathbf{b}) \);
11: Compute \( L_{\text{TM}} + \lambda_{\text{ECR}} L_{\text{ECR}} \) (Eq. (6));
12: Update \( \Theta \), \( \mathbf{W}, \mathbf{T} \) with a gradient step;
13: end for
14: end for

compare them in experiments). We do not adopt these classical clustering algorithms and some other work (Song et al., 2013; Huang et al., 2014; Xie et al., 2016; Hsu & Lin, 2017; Yang et al., 2017) as our clustering regularizer, because we cannot jointly optimize them along with a neural topic modeling objective.

Producing Sparse Soft-assignments We also require the clustering regularization to produce sparse soft-assignments. Even supporting joint optimization, existing clustering methods may still lead to topic collapsing. For example, we propose to use the state-of-the-art deep clustering method, Deep KMeans (DKM, Fard et al., 2020) that supports joint optimization. **Note we are the first to use DKM in topic modeling.** Its clustering objective is to minimize the total Euclidean distance between centers and samples weighted by soft-assignments. We use DKM as a clustering regularization on topic and word embeddings:

\[
\min_{\mathbf{W}, \mathbf{T}, \mathbf{p}} \sum_{j=1}^{V} \sum_{k=1}^{K} ||w_j - t_k||^2 p_{jk};
\]

where

\[
p_{jk} = \frac{e^{-||w_j - t_k||^2/\tau}}{\sum_{k'=1}^{K} e^{-||w_j - t_{k'}||^2/\tau}}.
\]

Here \( p_{jk} \) denotes the clustering soft-assignment of word embedding \( w_j \) assigned to topic embedding \( t_k \), which is modeled as a softmax function of the Euclidean distance between \( w_j \) and all topic embeddings \( (\tau \) is a temperature hyperparameter). Unfortunately, DKM still incurs topic collapsing (See quantitative results in Sec. 4.3). We see from Figure 2a that DKM cannot form separately aggregated clusters, so the topic embeddings (centers) cannot be separated but collapse together. To solve this issue, we require the clustering regularization to produce sparse soft-assignments—each word embedding is mainly assigned to only one topic embedding and rarely to others, which pushes each word embedding only close to one topic embedding and away from all others in the semantic space. This way can form separately aggregated word embedding clusters with topic embeddings as centers, which encourages topic embeddings to be away from each other. **Note that we do not directly model latent topics with these sparse soft-assignments (See Sec. 3.5).**

Fulfilling Preset Cluster Size Constraints We further require the clustering regularization to fulfill preset cluster size constraints. Only producing sparse soft-assignments may still result in topic collapsing. To make the soft-assignments sparse, we propose DKM+Entropy that jointly minimizes the entropy of soft-assignments,

\[
\sum_{j=1}^{V} \sum_{k=1}^{K} -p_{jk} \log p_{jk},
\]

with the clustering objective of DKM (Eq. (1)). However, this way still leads to topic collapsing (See quantitative results in Sec. 4.3). Figure 2b shows DKM+Entropy indeed forms separately aggregated clusters for some topic embeddings, but unfortunately the clustering solution is trivial—the clusters of most topic embeddings are empty, as quantitatively shown in Figure 2f. As a result, the topic embeddings of these empty clusters cannot be separated to cover distinct semantics but only collapse to others in the space. To avoid such trivial solutions of empty clusters, we propose to preset constraints on the size of each cluster (must not be empty) and require the clustering regularization to fulfill these constraints.

3.4. Embedding Clustering Regularization

To meet the above necessary requirements, we in this section introduce a novel method, **Embedding Clustering Regularization (ECR)**. Figure 3 illustrates ECR, and Figure 2e shows its effectiveness.
Presetting Cluster Size Constraints We first preset the cluster size constraints to be fulfilled to avoid trivial solutions of empty clusters. Recall that we have \( K \) topic embeddings as centers and \( V \) word embeddings as samples. We denote the cluster size of topic embedding \( t_k \) as \( n_k \) and the cluster size proportion as \( s_k = n_k / V \). We have \( s = (s_1, \ldots, s_K)^T \in \Delta_K \) as the vector of all cluster size proportions. Unfortunately, we usually lack prior knowledge about the cluster sizes of topic embeddings. Previous studies (Wallach et al., 2009) find that a symmetric Dirichlet prior over topic-word distributions achieves better performance in LDA. Inspired by this, we set all cluster sizes as uniform: \( n_k = V / K \) and \( s = (1/K, \ldots, 1/K)^T \). This setting can avoid the trivial solutions of empty clusters, and experiments show it works well across datasets (see Sec. 4.2). Note that \( s \) can be different values determined by prior knowledge from experts, and we leave this as future work.

Embedding Clustering Regularization (ECR) To meet the above requirements, we propose ECR that models clustering soft-assignments with the transport plan of a specifically defined optimal transport problem. Specifically, we define two discrete measures of topic \( (t_k) \) and word embeddings \( (w_j) \): 

\[ \gamma = \sum_{j=1}^V \delta_{w_j}, \quad \phi = \sum_{k=1}^K s_k \delta_{t_k}, \]

where \( \delta_x \) denotes the Dirac unit mass on \( x \). We formulate the entropic regularized optimal transport (Sinkhorn, 1964; Cuturi, 2013), a fast iterative scheme suited to the execution of GPU (Peyré et al., 2019). See Algorithm 1 for detailed algorithm steps. By doing so, \( \pi^*_\gamma = \text{sinkhorn}(\gamma, \phi, \varepsilon) \approx \arg \min_{\pi \in \mathbb{R}^{V \times K}_+} \mathcal{L}_{\text{OT}}(\gamma, \phi) = \epsilon \pi \mathbb{1}_V \) and \( \pi^T \mathbb{1}_V = s \).

Here the first term of \( \mathcal{L}_{\text{OT}} \) is the original optimal transport problem, and the second term with hyperparameter \( \varepsilon \) is the entropic regularization to make this problem tractable (Canas & Rosasco, 2012). Eq. (2) is to find the optimal transport plan \( \pi^*_\gamma \) that minimizes the total cost of transporting weight from word embeddings to topic embeddings. We measure the transport cost between word embedding \( w_j \) and topic embedding \( t_k \) by Euclidean distance: 

\[ C_{jk} = \| w_j - t_k \|^2, \]

and the transport cost matrix is \( C \in \mathbb{R}^{V \times K} \). The two conditions in Eq. (2) restrict the weight of each word embedding \( w_j \) as \( 1/V \) and each topic embedding \( t_k \) as \( s_k \), where \( \mathbb{1}_K(\mathbb{1}_V) \) is a \( K \) (\( V \)) dimensional column vector of ones. \( \pi_{jk} \) denotes the transport weight from \( w_j \) to \( t_k \); \( \pi \in \mathbb{R}^{V \times K}_+ \) is the transport plan that includes the transport weight of each word embedding to fulfill the weight of each topic embedding.

To meet the above requirements, we model clustering soft-assignments with the optimal transport plan \( \pi^*_\gamma \), i.e., the soft-assignment of \( w_j \) to \( t_k \) is the transport weight between them, \( \pi^*_{jk} \). Then we formulate our ECR objective by minimizing the total distance between word and topic embeddings weighted by these soft-assignments:

\[ \mathcal{L}_{\text{ECR}} = \sum_{j=1}^V \sum_{k=1}^K \| w_j - t_k \|^2 \pi^*_{jk}, \]

where \( \pi^*_{jk} = \text{sinkhorn}(\gamma, \phi, \varepsilon) \approx \arg \min_{\pi \in \mathbb{R}^{V \times K}_+} \mathcal{L}_{\text{OT}}(\gamma, \phi) \).

To solve this specifically defined optimal transport problem, here we compute \( \pi^*_\gamma \) through Sinkhorn’s algorithm (Sinkhorn, 1964; Cuturi, 2013), a fast iterative scheme suited to the execution of GPU (Peyré et al., 2019). See Algorithm 1 for detailed algorithm steps. By doing so, \( \pi^*_\gamma \) is a differentiable variable parameterized by transport cost matrix \( C \) (Salimans et al., 2018; Genevay et al., 2018). Due to this, minimizing \( \pi^*_{jk} \) increases transport cost \( C_{jk} \), i.e., the distance between \( w_j \) and \( t_k \); otherwise maximizing it decreases the distance (Genevay et al., 2019). Thus we can exactly model \( \pi^*_\gamma \) as differentiable clustering soft-assignments between topic and word embeddings.

**ECR is an Effective Regularization on Embeddings** First, ECR supports joint optimization since \( \pi^*_\gamma \) is differentiable as aforementioned. Second, ECR produces sparse soft-assignments. It is proven that \( \pi^*_\gamma \) converges to the optimal solution of the original optimal transport problem when \( \varepsilon \to 0 \), which leads to a sparse transport plan (Peyré et al., 2019). Hence ECR (Eq. (3)) produces sparse soft-assignments under a small \( \varepsilon \). With sparse soft-assignments, ECR pushes each word embedding only close to one topic embedding and away from all others, which forms separately aggregated clusters. We illustrate this property in Figures 2c to 2e. Last, ECR fulfills preset cluster size constraints. In Eq. (2), the transport plan is restricted by two conditions indicating the weight of each word embedding \( w_j \) is \( 1/V \) and each topic embedding \( t_k \) is \( s_k \). These ensure the sparse optimal transport plan \( \pi^*_\gamma \) needs to transport \( n_k \) word embeddings to topic embedding \( t_k \) to balance the weight, such that \( n_k \times \frac{1}{V} = s_k \). Accordingly, ECR fulfills the preset cluster size constraints with \( \pi^*_\gamma \) as clustering soft-assignments. This avoids trivial clustering solutions of empty clusters as shown in Figure 2f.

To sum up, our ECR effectively forces each topic embedding to be the center of a separately aggregated word embedding cluster. This makes topic embeddings away from each other and cover different semantics of word embeddings, which alleviates topic collapsing.

3.5 Embedding Clustering Regularization Topic Model

In this section, we propose a novel topic model, Embedding Clustering Regularization Topic Model (ECRTM)
that jointly optimizes the topic modeling objective and the ECR objective. Algorithm 1 shows its training algorithm.

**Inferring Doc-Topic Distributions** We devise the prior and variational distribution following VAE (Kingma & Welling, 2014) to infer doc-topic distributions. In detail, we draw a latent variable, \( r \), from a logistic normal distribution: \( p(r) = \mathcal{LN}(\mu_0, \Sigma_0) \), where \( \mu_0 \) and \( \Sigma_0 \) are the mean and diagonal covariance matrix (Srivastava & Sutton, 2014). Then we use an encoder network that outputs parameters of the variational distribution, the mean vector \( \mu = f_\mu(x) \) and covariance matrix \( \Sigma = \text{diag}(f_\Sigma(x)) \). So the variational distribution is \( q_\theta(r|x) = \mathcal{N}(\mu, \Sigma) \) where \( \Theta \) denotes the parameters of \( f_\mu \) and \( f_\Sigma \). By applying the reparameterization trick (Kingma & Welling, 2014), we sample \( r = \mu + \Sigma^{1/2} \varepsilon \) where \( \varepsilon \sim \mathcal{N}(0, I) \). We obtain the doc-topic distribution \( \Theta \) with a softmax function as \( \theta = \text{softmax}(r) \).

**Reconstructing Documents** We then reconstruct the input documents with topic-word distribution matrix \( \beta \in \mathbb{R}^{V \times K} \). Recall that \( \beta \) indicates the weights between all topics and words. Previous methods commonly model \( \beta \) as the product of topic and word embeddings (Sec. 3.2). Differently, our model uses the proposed ECR as a clustering regularization on topic and word embeddings, so \( \beta \) also needs to reflect the learned clustering assignments between them. We do not directly model \( \beta \) with the soft-assignments \( \pi^\star \) of our ECR. This is because it makes one word only belong to one topic as \( \pi^\star \) is very sparse (most values are close to zero as aforementioned); but in reality, one word should be able to belong to different topics (Blei et al., 2003). To this end, we propose to model \( \beta \) as

\[
\beta_{jk} = \frac{e^{-\|w_j - r_k\|^2 / \tau}}{\sum_{k'=1}^{K} e^{-\|w_j - r_{k'}\|^2 / \tau}} \quad (4)
\]

where \( \tau \) is a temperature hyperparameter. This formulation is similar to the less sparse soft-assignments of DKM (Fard et al., 2020). It is less sparse and can reflect the learned clustering assignments between topic and word embeddings. Thus one word can belong to different topics, which agrees with reality. With the doc-topic distribution \( \theta \) and the topic-word distribution matrix \( \beta \), we routinely sample the reconstructed document from a Multinomial distribution \( \text{Multi}(\text{softmax}(\beta \theta)) \).

**Overall Objective Function of ECRTM** Given a batch of \( N \) documents \( (x^{(1)}, \ldots, x^{(N)}) \), we write the topic modeling objective function following VAE as

\[
\mathcal{L}_{\text{TM}} = \frac{1}{N} \sum_{i=1}^{N} - (x^{(i)})^\top \log(\text{softmax}(\beta \theta^{(i)})) + \text{KL} \left[ q_\theta(r^{(i)}|x^{(i)}) \left| p(r^{(i)}) \right. \right]. \quad (5)
\]

The first term is the reconstruction error, and the second term is the KL divergence between the prior and variational distribution. ECRTM learns topics regularized by our ECR. We define the overall objective function of ECRTM as a combination of \( \mathcal{L}_{\text{TM}} \) (Eq. (5)) and \( \mathcal{L}_{\text{ECR}} \) (Eq. (3)):

\[
\min_{\theta, \omega, T} \mathcal{L}_{\text{TM}} + \lambda_{\text{ECR}} \mathcal{L}_{\text{ECR}} \quad (6)
\]

where \( \lambda_{\text{ECR}} \) is a weight hyperparameter. This overall objective enables ECRTM to aggregate the embeddings of related words to form separate clusters with topic embeddings as centers and avoids the collapsing of topic embeddings. Thus our ECRTM can alleviate the topic collapsing issue and learn coherent and diverse topics together with high-quality doc-topic distributions at the same time.

4. Experiment

4.1. Experiment Setup

**Datasets** We adopt the following benchmark document datasets for experiments: (i) 20 News Groups (20NG, Lang, 1995) is one of the most popular datasets for evaluating topic models, including news articles with 20 labels; (ii) IMDB (Maas et al., 2011) is the movie reviews containing two labels (positive and negative); (iii) Yahoo Answer (Zhang et al., 2015) is the question titles, contents, and the best answers from the Yahoo website with 10 labels, such as Society, Culture, and Family & Relationships; (iv) AG News (Zhang et al., 2015) contains news titles and descriptions, divided into 4 categories like Sports and Business. Note that Yahoo Answer and AG News belong to short texts. See Appendix A for pre-processing details.

**Evaluation Metrics** Following previous mainstream work, we evaluate topic models concerning topic quality and doc-topic distribution quality. Topic quality includes: (i) **Topic Coherence** measures the coherence between the top words of discovered topics (Newman et al., 2010; Wang & Blei, 2011). We employ the widely-used metric, Coherence Value (\(CV\)) which has been empirically shown to outperform the traditional metrics, NPMI, UCI, and UMass (Röder et al., 2015). We also exemplify this in Appendix E. We use the public Wikipedia article collection \(^2\) as the external reference corpus. This removes the bias of using relatively small datasets (e.g., training sets) as the reference corpus, so we can reach fair comparisons and good reproducibility. (ii) **Topic Diversity** measures the differences between discovered topics to verify if topic collapsing happens. We use the Topic Diversity metric (TD, Dieng et al., 2020) to evaluate this performance, which computes the proportion of unique words in the discovered topics. We select the top 15 words of discovered

\(^2https://github.com/dice-group/Palmetto
Effective Neural Topic Modeling with Embedding Clustering Regularization

<table>
<thead>
<tr>
<th>Model</th>
<th>20NG K=50</th>
<th>20NG K=100</th>
<th>IMDB K=50</th>
<th>IMDB K=100</th>
<th>Yahoo Answer K=50</th>
<th>Yahoo Answer K=100</th>
<th>AG News K=50</th>
<th>AG News K=100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CV</td>
<td>TD</td>
<td>CV</td>
<td>TD</td>
<td>CV</td>
<td>TD</td>
<td>CV</td>
<td>TD</td>
</tr>
<tr>
<td>LDA</td>
<td>0.385</td>
<td>0.655</td>
<td>0.387</td>
<td>0.622</td>
<td>0.347</td>
<td>0.342</td>
<td>0.691</td>
<td>0.224</td>
</tr>
<tr>
<td>KM</td>
<td>0.251</td>
<td>0.204</td>
<td>0.294</td>
<td>0.317</td>
<td>0.213</td>
<td>0.244</td>
<td>0.302</td>
<td>0.271</td>
</tr>
<tr>
<td>WLDAP</td>
<td>0.378</td>
<td>0.375</td>
<td>0.369</td>
<td>0.273</td>
<td>0.311</td>
<td>0.320</td>
<td>0.069</td>
<td>0.333</td>
</tr>
<tr>
<td>DVAE</td>
<td>0.331</td>
<td>0.598</td>
<td>0.372</td>
<td>0.658</td>
<td>0.294</td>
<td>0.290</td>
<td>0.229</td>
<td>0.338</td>
</tr>
<tr>
<td>ETM</td>
<td>0.375</td>
<td>0.704</td>
<td>0.369</td>
<td>0.573</td>
<td>0.346</td>
<td>0.341</td>
<td>0.371</td>
<td>0.354</td>
</tr>
<tr>
<td>HyperMiner</td>
<td>0.371</td>
<td>0.613</td>
<td>0.368</td>
<td>0.446</td>
<td>0.347</td>
<td>0.343</td>
<td>0.258</td>
<td>0.344</td>
</tr>
<tr>
<td>NSTM</td>
<td>0.395</td>
<td>0.427</td>
<td>0.391</td>
<td>0.473</td>
<td>0.334</td>
<td>0.340</td>
<td>0.255</td>
<td>0.390</td>
</tr>
<tr>
<td>WeTe</td>
<td>0.383</td>
<td>0.949</td>
<td>0.352</td>
<td>0.742</td>
<td>0.368</td>
<td>0.931</td>
<td>0.293</td>
<td>0.367</td>
</tr>
</tbody>
</table>

ECRTM 0.431 0.964 0.405 0.904 0.393 0.974 0.373 0.887 0.405 0.985 0.389 0.903 0.466 0.961 0.416 0.981

Table 2: Topic quality of topic coherence (C_V) and topic diversity (TD) under 50 and 100 topics (K=50 and K=100). The best scores are in **bold**. ‡ means the gain of ECRTM is statistically significant at 0.05 level.

<table>
<thead>
<tr>
<th>Model</th>
<th>20NG K=50</th>
<th>20NG K=100</th>
<th>IMDB K=50</th>
<th>IMDB K=100</th>
<th>Yahoo Answer K=50</th>
<th>Yahoo Answer K=100</th>
<th>AG News K=50</th>
<th>AG News K=100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Purity</td>
<td>NMI</td>
<td>Purity</td>
<td>NMI</td>
<td>Purity</td>
<td>NMI</td>
<td>Purity</td>
<td>NMI</td>
</tr>
<tr>
<td>LDA</td>
<td>0.367</td>
<td>0.364</td>
<td>0.364</td>
<td>0.346</td>
<td>0.614</td>
<td>0.041</td>
<td>0.600</td>
<td>0.037</td>
</tr>
<tr>
<td>WLDAP</td>
<td>0.233</td>
<td>0.157</td>
<td>0.292</td>
<td>0.207</td>
<td>0.589</td>
<td>0.011</td>
<td>0.602</td>
<td>0.013</td>
</tr>
<tr>
<td>DVAE</td>
<td>0.087</td>
<td>0.018</td>
<td>0.104</td>
<td>0.035</td>
<td>0.517</td>
<td>0.000</td>
<td>0.525</td>
<td>0.001</td>
</tr>
<tr>
<td>ETM</td>
<td>0.347</td>
<td>0.319</td>
<td>0.394</td>
<td>0.339</td>
<td>0.660</td>
<td>0.038</td>
<td>0.648</td>
<td>0.037</td>
</tr>
<tr>
<td>HyperMiner</td>
<td>0.433</td>
<td>0.405</td>
<td>0.454</td>
<td>0.386</td>
<td>0.655</td>
<td>0.046</td>
<td>0.641</td>
<td>0.032</td>
</tr>
<tr>
<td>NSTM</td>
<td>0.354</td>
<td>0.356</td>
<td>0.383</td>
<td>0.363</td>
<td>0.658</td>
<td>0.040</td>
<td>0.659</td>
<td>0.039</td>
</tr>
<tr>
<td>WeTe</td>
<td>0.268</td>
<td>0.304</td>
<td>0.338</td>
<td>0.348</td>
<td>0.587</td>
<td>0.031</td>
<td>0.589</td>
<td>0.025</td>
</tr>
</tbody>
</table>

ECRTM 0.560 0.524 0.555 0.494 0.694 0.058 0.694 0.049 0.550 0.295 0.563 0.311 0.802 0.367 0.812 0.428

Table 3: Document clustering of Purity and NMI under 50 and 100 topics (K=50 and K=100). The best scores are in **bold**. ‡ means the gain of ECRTM is statistically significant at 0.05 level.

topics for the above topic quality evaluation. We furthermore conduct document clustering experiments to evaluate doc-topic distribution quality with Purity and NMI following Zhao et al. (2021b); Wang et al. (2022).

**Baseline Models** We consider the following state-of-the-art models for comparison: (i) **LDA** (Blei et al., 2003), one of the most widely-used probabilistic topic models; (ii) **KM** (Sia et al., 2020), directly clustering word embeddings to produce topics. Note that we cannot use it for document clustering since it cannot infer the doc-topic distributions to produce topics. Note that we cannot use it for doc-topic modeling. (iii) **DVAE** (Burkhardt & Kramer, 2019), Dirichlet VAE that approximates Dirichlet priors with rejection sampling; (iv) **WLDAP** (Nan et al., 2019), a WAE-based topic model; (v) **ETM** (Dieng et al., 2020), a neural topic model which models the topic-word distribution matrix with word and topic embeddings; (vi) **HyperMiner** (Xu et al., 2022), using embeddings in the hyperbolic space to model topics. (vii) **NSTM** (Zhao et al., 2021b), using optimal transport distance between doc-topic distributions and documents to measure reconstruction error. (viii) **WeTe** (Wang et al., 2022), following NSTM and using conditional transport distance as reconstruction error.

4.2. **Topic and Doc-Topic Distribution Quality**

Table 2 reports the topic quality results concerning C_V and TD, and Table 3 summarizes doc-topic distribution quality results concerning Purity and NMI of document clustering. We have the following observations: (i) **ECRTM** effectively addresses the topic collapsing issue and outperforms baselines in topic quality. In Table 2, the much lower TD scores imply baselines generate repetitive topics and thus suffer from topic collapsing. As aforementioned, these repetitive topics are less useful for downstream tasks and damage the interpretability of topic models. In contrast, we see our ECRTM achieves significantly higher TD scores across all datasets and mostly the best C_V scores. We emphasize although the C_V of ECRTM is slightly higher than NSTM (0.389 vs. 0.387) on Yahoo Answer, ECRTM completely outperforms on TD (0.903 vs. 0.659). These results demonstrate that ECRTM produces more coherent and diverse topics than state-of-the-art baselines. These improvements are because our ECRTM makes topic embeddings away from each other and cover different semantics of word embeddings in the space instead of collapsing together as some baselines. (ii) **ECRTM** surpasses baselines in inferring high-quality doc-topic dis-
**Table 4: Ablation study.** The terrible TD scores of DKM, DKM+Entropy and w/o ECR indicate topic collapsing still exists, making their high $C_V$ less meaningful. In contrast, ECRTM achieves much higher TD with the best Purity and NMI. The $C_V$ scores of ECRTM also outperforms state-of-the-art baselines (See Table 2). ‡ means the gain of ECRTM is statistically significant at 0.05 level.

<table>
<thead>
<tr>
<th>Model</th>
<th>20NG</th>
<th>Yahoo Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Purity</td>
<td>NMI</td>
</tr>
<tr>
<td>DKM</td>
<td>0.510</td>
<td>0.471</td>
</tr>
<tr>
<td>DKM+Entropy</td>
<td>0.222</td>
<td>0.148</td>
</tr>
<tr>
<td>w/o ECR</td>
<td>0.504</td>
<td>0.446</td>
</tr>
<tr>
<td>ECRTM</td>
<td>0.560</td>
<td>0.524</td>
</tr>
</tbody>
</table>

Table 5: Case study: each row is the top 10 related words of a discovered topic. Repetitive words are underlined.

<table>
<thead>
<tr>
<th>Model</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETM</td>
<td>like better good especially end look much done way just like just one way made times really even feel one like around sort looking kind good main look just</td>
</tr>
<tr>
<td>HyperMiner</td>
<td>one even end way little part character make never plot even even little why one plot character enough part make even seems fact enough plot end least character audience make</td>
</tr>
<tr>
<td>NSTM</td>
<td>just show even come time one good really going know just even really something come going like actually things get just one even something come way really like always good</td>
</tr>
<tr>
<td>WeTe</td>
<td>just like really bad good get one think see even man back gets goes two get takes house around away jokes jackson lisa predictable recycled wasted murphy writers williams rock</td>
</tr>
<tr>
<td>DKM</td>
<td>christmas disney musical songs bill timeless prince art rock holiday christmas santa childrens holiday betty age ann adult children toy fantasy christmas magic effects magical santa special holiday childrens child</td>
</tr>
<tr>
<td>DKM+Entropy</td>
<td>funny day physical semi ever way old due seen zone funny ever day old seen way physical due semi relationship funny ever day semi seen physical way old due psychological</td>
</tr>
<tr>
<td>ECRTM</td>
<td>jackie martial chan kung arts kong hong stunts bruce fight nominated nancy academy award awards oscar oscars jake nomination dracula vampires vampire freddy zombies zombie nightmare serial halloween killer slasher</td>
</tr>
</tbody>
</table>

Table 4 shows our ECRTM consistently outperforms the baseline models by a large margin in terms of Purity and NMI. For example, ECRTM reaches 0.560 and 0.524 for Purity and NMI on 20NG, while the runner-up only has 0.367 and 0.364. These manifest that ECRTM not only achieves higher-quality topics but also better document-level topic distributions. See Appendices C to F for more experiments like robustness to the number of topics and visualization results.

### 4.3. Ablation Study

We conduct ablation studies and show the necessity of our proposed Embedding Clustering Regularization (ECR). Specifically, we remove the ECR from our ECRTM, denoted as w/o ECR. We also compare with the state-of-the-art deep clustering method, DKM (Fard et al., 2020) and DKM with minimizing entropy (DKM+Entropy, see Sec. 3.3). Note that we are the first to use DKM in topic modeling. Table 4 shows DKM, DKM+Entropy, and w/o ECR all suffer from topic collapsing as indicated by their much lower TD scores. Although they have high $C_V$, their terrible TD scores mean most topics are repetitive and less useful for downstream tasks, making their high $C_V$ scores less meaningful (see examples in Sec. 4.5 for illustrations). Conversely, our ECRTM improves TD scores by a large margin and achieves the best document clustering performance with much higher Purity and NMI. This is because our ECR, as an effective regularization, can avoid the collapsing of topic embeddings while DKM, DKM+Entropy, and w/o ECR cannot. These results demonstrate our ECR is necessary to address the topic collapsing issue and achieve effective topic modeling performance.
processing. A smaller max-df removes more high-frequency words. The best scores are in Table 6: Influence of high-frequency words. Here max-df denotes the maximum document frequency of dataset pre-processing? Driven by this, we alter the max-df to remove the high-frequency words using reliable pre-processing. This inspires us to ask: what if we carefully remove high-frequency words using reliable pre-processing? As aforementioned in Sec. 3.2, we argue that topic collapsing results from the reconstruction error minimization on high-frequency words. This inspires us to ask: what if we carefully remove high-frequency words using reliable dataset pre-processing? Driven by this, we alter the maximum document frequency (max-df) to remove the high-frequency words when pre-processing the 20NG dataset. A smaller max-df removes more high-frequency words. From Table 6 we see that most baselines, such as LDA, ETM, and NSTM, reach higher TD scores under small max-df, indicating that topic collapsing is alleviated to some extent. But our model consistently outperforms all baselines on the topic quality and document clustering. These results empirically confirm our argument that the topic collapsing issue arises from reconstructing high-frequency words. Besides, these results verify one of our advantages: our model requires no reliable pre-processing to achieve state-of-the-art performance. This advantage is vital since the definition of reliable pre-processing is inconclusive: A large max-df may not remove any high-frequency words while a small one may remove most of the important words. More critically, brutally searching for reliable pre-processing is time-consuming and laborious. This advantage becomes more significant when meeting many large-scale datasets from various domains.

### 4.4. Text Classification

To evaluate extrinsically, we further conduct text classification experiments as downstream tasks. Specifically, we use the doc-topic distributions learned by topic models as document features and train SVMs to predict the class of each document. As reported in Figure 4, ECRMT significantly outperforms baseline models on all datasets. These results demonstrate that our ECRMT can be better utilized in the downstream classification tasks.

### 4.5. Case Study: Examples of Discovered Topics

For case study, Table 5 shows examples of discovered topics by different models from IMDB. We observe that ETM and NSTM both have highly uninformative and similar topics including common words like “just”, “like”, or “something”. HyperMiner generates repetitive topics with the words “one”, “even”, and “end” WeTe produces some less informative topics like “just like really bad good...”. DKM and DKM+Entropy also have repetitive topics with the words “christmas”, “holiday”, and “funny”. Accordingly, we observe that the topic collapsing issue commonly exists in these methods. These collapsed topics are uninformative and redundant, which are less useful for downstream applications and damage the interpretability of topic models. In contrast, the topics discovered by ECRMT are more distinct instead of repeating each other. Besides, they are more coherent, such as the first topic with relevant words like “christmas”, “holiday”, and “funny”. Appendix G shows the full topics lists of models.

### 4.6. Influence of Dataset Pre-processing

As aforementioned in Sec. 3.2, we argue that topic collapsing results from the reconstruction error minimization on high-frequency words. A smaller max-df may not remove any high-frequency words while a small one may remove most of the important words. More critically, brutally searching for reliable pre-processing is time-consuming and laborious. This advantage becomes more significant when meeting many large-scale datasets from various domains.

### 5. Conclusion

In this paper, we propose the novel Embedding Clustering Regularization Topic Model (ECRMT) to address the topic collapsing issue. ECRMT learns topics under the new Embedding Clustering Regularization that forces each topic embedding to be the center of a separately aggregated word embedding cluster. Extensive experiments demonstrate that ECRMT achieves effective neural topic modeling, successfully alleviates topic collapsing, and consistently achieves state-of-the-art performance in terms of producing high-quality topics and topic distributions of documents.

### Acknowledgements

We thank all anonymous reviewers for their helpful comments. This research/project is supported by the National Research Foundation, Singapore under its AI Singapore Programme, AISG Award No: AISG2-TC-2022-005 and AISG Award No: AISG-100E-2019-046.
References


Yan, X., Guo, J., Liu, S., Cheng, X., and Wang, Y. Learning topics in short texts by non-negative matrix factorization on term correlation matrix. In *proceedings of the*


Effective Neural Topic Modeling with Embedding Clustering Regularization

### Table 7: Topic quality of coherence ($C_V$) and diversity (TD) under topic number $K = 10, 20, 30, 40, 60, 70, 80, 90$. The best scores are in **bold**.

<table>
<thead>
<tr>
<th>Model</th>
<th>Purity</th>
<th>NMI</th>
<th>Purity</th>
<th>NMI</th>
<th>Purity</th>
<th>NMI</th>
<th>Purity</th>
<th>NMI</th>
<th>Purity</th>
<th>NMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>0.295</td>
<td>0.408</td>
<td>0.340</td>
<td>0.396</td>
<td>0.347</td>
<td>0.375</td>
<td>0.368</td>
<td>0.356</td>
<td>0.354</td>
<td>0.352</td>
</tr>
<tr>
<td>WLDA</td>
<td>0.174</td>
<td>0.119</td>
<td>0.194</td>
<td>0.124</td>
<td>0.223</td>
<td>0.152</td>
<td>0.238</td>
<td>0.161</td>
<td>0.260</td>
<td>0.176</td>
</tr>
<tr>
<td>ETM</td>
<td>0.183</td>
<td>0.274</td>
<td>0.275</td>
<td>0.307</td>
<td>0.307</td>
<td>0.288</td>
<td>0.331</td>
<td>0.281</td>
<td>0.351</td>
<td>0.291</td>
</tr>
<tr>
<td>HyperMiner</td>
<td>0.240</td>
<td>0.299</td>
<td>0.338</td>
<td>0.390</td>
<td>0.416</td>
<td>0.421</td>
<td>0.407</td>
<td>0.389</td>
<td>0.478</td>
<td>0.422</td>
</tr>
<tr>
<td>NSTM</td>
<td>0.228</td>
<td>0.284</td>
<td>0.295</td>
<td>0.327</td>
<td>0.355</td>
<td>0.373</td>
<td>0.349</td>
<td>0.349</td>
<td>0.362</td>
<td>0.353</td>
</tr>
<tr>
<td>WeTe</td>
<td>0.055</td>
<td>0.004</td>
<td>0.119</td>
<td>0.150</td>
<td>0.197</td>
<td>0.244</td>
<td>0.252</td>
<td>0.317</td>
<td>0.281</td>
<td>0.332</td>
</tr>
</tbody>
</table>

ECRTM 0.390 0.485 0.373 0.420 0.463 0.435 0.462 0.426 0.554 0.522 0.559 0.498 0.581 0.506 0.564 0.497

### Table 8: Document clustering of Purity and NMI under topic number $K = 10, 20, 30, 40, 60, 70, 80, 90$. The best scores are in **bold**.

<table>
<thead>
<tr>
<th>Model</th>
<th>$K=10$</th>
<th>$K=20$</th>
<th>$K=30$</th>
<th>$K=40$</th>
<th>$K=60$</th>
<th>$K=70$</th>
<th>$K=80$</th>
<th>$K=90$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>0.376</td>
<td>0.627</td>
<td>0.384</td>
<td>0.626</td>
<td>0.377</td>
<td>0.678</td>
<td>0.390</td>
<td>0.697</td>
</tr>
<tr>
<td>KM</td>
<td>0.208</td>
<td>0.207</td>
<td>0.230</td>
<td>0.180</td>
<td>0.247</td>
<td>0.151</td>
<td>0.255</td>
<td>0.217</td>
</tr>
<tr>
<td>WLDA</td>
<td>0.354</td>
<td>0.533</td>
<td>0.354</td>
<td>0.443</td>
<td>0.360</td>
<td>0.389</td>
<td>0.357</td>
<td>0.430</td>
</tr>
<tr>
<td>ETM</td>
<td>0.380</td>
<td>0.820</td>
<td>0.372</td>
<td>0.763</td>
<td>0.373</td>
<td>0.744</td>
<td>0.383</td>
<td>0.698</td>
</tr>
<tr>
<td>HyperMiner</td>
<td>0.361</td>
<td>0.800</td>
<td>0.377</td>
<td>0.750</td>
<td>0.373</td>
<td>0.671</td>
<td>0.378</td>
<td>0.665</td>
</tr>
<tr>
<td>NSTM</td>
<td>0.397</td>
<td>0.606</td>
<td>0.381</td>
<td>0.487</td>
<td>0.389</td>
<td>0.418</td>
<td>0.392</td>
<td>0.444</td>
</tr>
<tr>
<td>WeTe</td>
<td>0.422</td>
<td>1.000</td>
<td>0.380</td>
<td>0.980</td>
<td>0.387</td>
<td>0.980</td>
<td>0.388</td>
<td>0.978</td>
</tr>
</tbody>
</table>

ECRTM 0.487 1.000 0.454 1.000 0.437 1.000 0.435 0.993 0.413 0.993 0.405 0.910 0.410 0.957 0.410 0.957 0.402 0.906

### Table 9: Statistics of datasets after pre-processing.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#docs</th>
<th>Vocabulary Size</th>
<th>Average Length</th>
<th>#labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>20NG</td>
<td>18,846</td>
<td>5,000</td>
<td>110.5</td>
<td>20</td>
</tr>
<tr>
<td>IMDB</td>
<td>50,000</td>
<td>5,000</td>
<td>95.0</td>
<td>2</td>
</tr>
<tr>
<td>Yahoo Answer</td>
<td>12,500</td>
<td>5,000</td>
<td>35.4</td>
<td>10</td>
</tr>
<tr>
<td>AG News</td>
<td>12,500</td>
<td>5,000</td>
<td>20.1</td>
<td>4</td>
</tr>
</tbody>
</table>

### A. Dataset

We follow the dataset pre-processing steps of (Card et al., 2018): (1) tokenize documents and convert to lowercase; (2) remove punctuation; (3) remove tokens that include numbers; (4) remove tokens less than 3 characters; (5) remove stop words. The statistics of pre-processed datasets are reported in Table 9.

### B. Model Implementation

For pre-trained word embeddings, we employ 200-dimensional GloVe (Pennington et al., 2014) \(^3\). For the Sinkhorn’s algorithm of ECRTM, we set the maximum number of iterations as 1,000, the stop tolerance 0.005, and $\varepsilon$ 0.05 following Cuturi (2013). For our ECRTM, the prior distribution is specified with Laplace approximation (Henig et al., 2012) to approximate a symmetric Dirichlet prior as $\mu_{0,k} = 0$ and $\Sigma_{0,kk} = (K - 1)/(\alpha K)$ with hyperparameter $\alpha$. We set $\alpha$ as 1.0 following Card et al. (2018). Our encoder network is the same as Srivastava & Sutton (2017); Wu et al. (2020a;b): a MLP that has two linear layers with softplus activation function, concatenated with two single layers each for the mean and covariance matrix. We use Adam (Kingma & Ba, 2014) to optimize model parameters. See other implementation details in our code.

\(^3\)https://nlp.stanford.edu/projects/glove/
Effective Neural Topic Modeling with Embedding Clustering Regularization

Figure 5: Annotations of top words of discovered topics in the semantic space.

Figure 6: t-SNE (van der Maaten & Hinton, 2008) visualization of word embeddings (●) and topic embeddings (▲) under 100 topics. Topic embeddings commonly collapse together in state-of-the-art models (ETM (Dieng et al., 2020), NSTM (Zhao et al., 2021b), and WeTe (Wang et al., 2022)). In contrast, ECRTM can avoid the collapsing by forcing each topic embedding to be the center of a separately aggregated word embedding cluster.

C. Perplexity Results

We report the perplexity results in Table 10. Here we do not include some neural topic models (WLDA, NSTM, WeTe) as they are inapplicable to the perplexity approximation with ELBO (See explanations in Miao et al. (2016); Srivastava & Sutton (2017); Nan et al. (2019); Zhao et al. (2021a)). Table 10 shows our ECRTM also achieves the best perplexity results (lower is better).

D. Robustness to the Number of Topics

Besides the aforementioned results under $K = 50, 100$ (Tables 2 and 3), we also experiment under $K = 10, 20, 30, 40, 60, 70, 80, 90$ on 20NG to verify the robustness of our method. As shown in Tables 7 and 8, we see that ECRTM consistently outperforms baseline models in terms of both topic quality and clustering. These show that the performance improvements of our ECRTM are robust to the number of topics.

E. Comparison of Coherence Metrics

Röder et al. (2015) have empirically shown that $C_V$ is a better coherence metric which has better consistency with human judgment than traditional metrics like NPMI, UCI, and UMass (Bouma, 2009; Chang et al., 2009; Newman et al., 2010; Mimno et al., 2011). We also confirm this argument in our experiments: we find NPMI, UCI, and UMass tend to
Effective Neural Topic Modeling with Embedding Clustering Regularization

<table>
<thead>
<tr>
<th>NPMI</th>
<th>UCI</th>
<th>UMass</th>
<th>(C_V)</th>
<th>Top words of topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.069</td>
<td>0.746</td>
<td>-0.900</td>
<td>0.272</td>
<td><strong>Topic#1:</strong> way come actually make yet example even fact though rather</td>
</tr>
<tr>
<td>0.065</td>
<td>0.615</td>
<td>-0.967</td>
<td>0.272</td>
<td><strong>Topic#2:</strong> fact even indeed though kind something way always actually things</td>
</tr>
<tr>
<td>0.065</td>
<td>0.628</td>
<td>-3.367</td>
<td>0.255</td>
<td><strong>Topic#3:</strong> really pretty something seem seems quite things nothing thing ridiculous</td>
</tr>
<tr>
<td>-0.076</td>
<td>-2.660</td>
<td>-8.196</td>
<td>0.497</td>
<td><strong>Topic#4:</strong> vampires vampire freddy zombies zombie nightmare serial halloween killer slasher</td>
</tr>
<tr>
<td>-0.075</td>
<td>-2.678</td>
<td>-9.032</td>
<td>0.476</td>
<td><strong>Topic#5:</strong> disney lion cartoons animation cartoon animated anime animals toy fox</td>
</tr>
<tr>
<td>-0.114</td>
<td>-3.367</td>
<td>-6.120</td>
<td>0.452</td>
<td><strong>Topic#6:</strong> jackie martial chan kung arts kong hong stunts bruce fight</td>
</tr>
</tbody>
</table>

Table 11: Comparison of coherence metrics. \(C_V\) gives high scores to coherent topics (Topic#4-6) while traditional metrics (NPMI, UCI, UMass) assigns high scores to less informative topics (Topic#1-3).

give higher scores to trivial and less informative topics. For example, we see in Table 11, Topic#4-6 are more coherent and informative than Topic#1-3. But unfortunately, NPMI, UCI, and UMass give much higher scores to Topic#1-3 instead of Topic#4-6. In contrast, \(C_V\) is more reasonable as it gives higher scores to Topic#4-6. Note that it is incorrect to directly compare a NPMI/UCI/UMass score with a \(C_V\) score because they are in different scales.

F. Visualization of Embedding Space

We visualize the learned topic and word embeddings with t-SNE (van der Maaten & Hinton, 2008) under 100 topics (Figure 1 is under 50 topics). Figure 6 shows while the topic embeddings mostly collapse together in the state-of-the-art baselines, our ECRTM avoids the collapsing of topic embeddings by forcing each topic embedding to be the center of a separately aggregated word embedding cluster. This illustrates that our ECR also works effectively under a larger number of topics.

We furthermore annotate the semantic embedding space with the top related words of discovered topics by ECRTM under \(K=50\) as shown in Figure 5. We see that each word embedding cluster represents a diverse and coherent topic. This shows that our ECRTM effectively clusters the embeddings of coherent words.
G. Full Lists of Discovered Topics

Below are the discovered topics of different models from IMDB under 50 topics ($K=50$).

**ETM**

**Topic#1**: director etc dialog cult ben joe steve mans todays bruce
**Topic#2**: killer police kill killed murder car head shot killing body
**Topic#3**: instead rather put mind together despite come beyond without alone
**Topic#4**: series first later one time also since made set years
**Topic#5**: art rather use special little like sci king bit style
**Topic#6**: dark fantasy earth full begin world open end setting set
**Topic#7**: family father young wife son mother child life children daughter
**Topic#8**: show shows episode watching see watch episodes television star dvd
**Topic#9**: boring entertaining pointless silly dull humour pretentious sub unfunny entertained
**Topic#10**: like better good especially end look much done way just
**Topic#11**: war world history people documentary american life society german soldiers
**Topic#12**: film films cinema directors filmed hollywood filmmakers cinematography welles noir
**Topic#13**: new one way time world different place two live day
**Topic#14**: like just one way made much times really even feel
**Topic#15**: music musical song songs dance stage oscar singing voice career
**Topic#16**: even one though also far way still made yet better
**Topic#17**: part like end also along two short lines now line
**Topic#18**: self somewhat sense seems narrative character sexual becomes often person
**Topic#19**: story book great perfect gives comes work novel excellent brilliant
**Topic#20**: supposed keep get wanted got trying never kept going turn
**Topic#21**: people just really know think say maybe something things understand
**Topic#22**: john michael james robert richard paul george jack cast played
**Topic#23**: scene scenes sex plot slow violence blood boring twist nudity
**Topic#24**: action man death gets fight hero back another comes face
**Topic#25**: plot original poor pretty unfortunately decent terrible nothing budget completely
**Topic#26**: story also makes interesting many stories however quite make style
**Topic#27**: great really good see recommend fan liked definitely thought watch
**Topic#28**: seen ever one worst years saw see beginning time remember
**Topic#29**: one time make give anyone made take get making another
**Topic#30**: time minutes just long hour half back night couple looking
**Topic#31**: house goes starts gets looks red night tries takes strange
**Topic#32**: love life heart beautiful human loved lives wonderful amazing world
**Topic#33**: first one ending time two end second just made last
**Topic#34**: best role actor performance played also good performances actors play
**Topic#35**: characters character plot director dialogue film real drama development acted
**Topic#36**: actors cast good acting throughout script moments story film written
**Topic#37**: really thing good guy watch movie whole yes want waste
**Topic#38**: course much seems audience quite less picture overall bit opinion
**Topic#39**: man young men women two woman town small female local
**Topic#40**: like high school game video one quality camera sound shot
**Topic#41**: bad even acting like just awful made stupid worse crap
**Topic#42**: old little girl get kids girls boy kid cool year
**Topic#43**: movie movies watch watching actors people theater theaters disaster entertainment
**Topic#44**: make see sure say think need really looks know good
**Topic#45**: version american truly classic japanese british era french cartoon tale
**Topic#46**: funny comedy fun laugh humor jokes watch hilarious comic imagine
**Topic#47**: one like around sort looking kind good main look just
**Topic#48**: actually real black just white believe mean see either know
**Topic#49**: horror effects low dvd english movies films zombie dead genre
**Topic#50**: fact reason actually even never simply none however obvious nothing
HyperMiner

Topic#1: jack musical oscar song songs george dance jane mary scott
Topic#2: just people like think know something see get say want
Topic#3: effort attempts lacks poor merely intended lacking whilst grace appeal
Topic#4: one even end way little part character make never plot
Topic#5: also quite interesting however rather many much although bit though
Topic#6: one great seen ever best see watch time every enjoy
Topic#7: even one end way little part character make plot never
Topic#8: just like people know something think say see want get
Topic#9: life love beautiful real heart TRUE romantic lives romance dream
Topic#10: even one end part little character way make never point
Topic#11: film films made cinema making silent festival director makers filmmakers
Topic#12: scene get gets guy girl around back away head getting
Topic#13: movie movies bad made watch make acting watching even plot
Topic#14: see saw thought watched dvd watching got video felt went
Topic#15: good like really just better lot much look pretty nice
Topic#16: bad horror worst awful terrible acting waste budget worse low
Topic#17: police crime cop gun soldiers gang prison church agent charlie
Topic#18: even end way character seems make never little part fact
Topic#19: funny comedy fun kids school laugh humor jokes hilarious christmas
Topic#20: just people know like think something see say get want
Topic#21: even seems fact enough plot end least character audience make
Topic#22: show series shows episode episodes television season pilot trek writers
Topic#23: car game red dog cat camp van steve chase eye
Topic#24: one first two new time another second star also world
Topic#25: effects earth monster sci space battle island match computer adventure
Topic#26: one great seen ever best see watch time every fan
Topic#27: man takes find woman help place small comes becomes along
Topic#28: also many however quite much interesting rather although though bit
Topic#29: scene get gets around guy girl sex getting back away
Topic#30: cast role performance actor john excellent play played performances actors
Topic#31: just like people know think see get something say going
Topic#32: film films made cinema making silent director independent makers festival
Topic#33: scenes action dark slow genre fight opening scene sequence violence
Topic#34: scene get gets guy girl around back sex guys getting
Topic#35: one first two time new another second star also world
Topic#36: even end way one little character make plot enough part
Topic#37: director work camera production music script sound writer shot direction
Topic#38: human art french world nature reality powerful images experience deep
Topic#39: war american black white history world country english british documentary
Topic#40: young family wife father boy son mother children child daughter
Topic#41: one great seen ever see best watch time every fan
Topic#42: one great seen ever best see watch time fan every
Topic#43: old years still now last time year three long back
Topic#44: death dead evil blood killer house kill night murder killed
Topic#45: even end little way one plot character enough part make
Topic#46: lack attempt self premise dull flat fails failed unfortunately pretentious
Topic#47: story characters original book version read based stories character king
Topic#48: good like really just better lot much look pretty acting
Topic#49: cast role performance actor john excellent played play actors performances
Topic#50: saw see thought dvd watched watching felt got went video
Effective Neural Topic Modeling with Embedding Clustering Regularization

NSTM

Topic#1: miller got smith moore just johnson really know davis think
Topic#2: even one though time way just come fact much make
Topic#3: know just really going think something come maybe get even
Topic#4: one another man murder even wanted others death just taken
Topic#5: movie like monster something actually just movies come thing kind
Topic#6: one just along now part way another come time around
Topic#7: just get coming one know even really something way
Topic#8: one first time also though best another even although came
Topic#9: sense kind something love really feel thing feeling nothing sort
Topic#10: movie movies film films just really best like something thing
Topic#11: just like really come going get even maybe something good
Topic#12: really movie thing something just things maybe good stuff kind
Topic#13: interesting movie something funny kind really quite wonderful things fun
Topic#14: just going really get something maybe come know even thing
Topic#15: one time first movie best though even just like also
Topic#16: best one smith moore miller time first just davis james
Topic#17: movie film just even best though films actually fact really
Topic#18: just show even come time one good really going know
Topic#19: wondervul amazing terrific good really fantastic best something thing pretty
Topic#20: wearing wore dress look dressed wears clothes worn shirt
Topic#21: even just come way going get make time though one
Topic#22: really something think know thing maybe just things going good
Topic#23: movie film best fact one though example even story life
Topic#24: music best musical songs movie like featured one song playing
Topic#25: really something seems pretty quite things seem thing nothing think
Topic#26: goes takes tells gets comes finds makes knows everyone happens
Topic#27: just even really something come going like actually things get
Topic#28: just one even something come way really like always good
Topic#29: funny silly movie stuff amusing scary fun hilarious cheesy boring
Topic#30: even though much make way just come actually fact one
Topic#31: just come one way good really though going something
Topic#32: even fact though way come actually make yet indeed something
Topic#33: just one even time way going coming another get
Topic#34: something really even just things good actually always way kind
Topic#35: one even time now though come came last another also
Topic#36: best movie film one good films actor like movies time
Topic#37: movie film movies film comedy drama best hollywood starring story
Topic#38: mother daughter wife sister friend husband married couple actress love
Topic#39: fact even example though way one rather much indeed life
Topic#40: nose eyes hand just mouth legs fingers neck teeth like
Topic#41: really just something think going know things maybe even thing
Topic#42: even come know just fact think something way really actually
Topic#43: one even just come now time life mother know though
Topic#44: just like inside get even look come one everything away
Topic#45: really maybe thing know something think things just going everybody
Topic#46: one just another came time back now come went got
Topic#47: fact even something really though always way actually things kind
Topic#48: time first one just last second next came coming play
Topic#49: horrible awful terrible horrific thing horrifying frightening kind shocking really
Topic#50: film movie films movies best directed drama feature picture though
WeTe

Topic#1: tonight fifth terrific kings loser wow fourth lucky bang grabs
Topic#2: australia progress reached secondly environment suspects interests scores continued press
Topic#3: profanity striking refuse stress sue complain survived fatal contact cracking
Topic#4: located population neighborhood nation distance owns cox traffic centered bell
Topic#5: graphics wholly puzzle map chapter attraction medium reader composed edition
Topic#6: acclaimed exquisite gothic vivid splendid stark sublime lively photographer literary
Topic#7: theaters cinemas preview trailers ratings extras studios rental par nyc
Topic#8: journalist calls debate complaint behaviour flawed defense civil charge innocence
Topic#9: discovering pursuit rid junk disappears cheating pretending petty discovers saving
Topic#10: john big new star match city james game george stars
Topic#11: predator heartbreaking terrifying bravo suspenseful intrigue paranoia menace unsettling mafia
Topic#12: story characters life people way real sense love much
Topic#13: pal moody greek trend global corporate depression combat transformed sin
Topic#14: shaky hung gray shine rough heights staring screens clad casts
Topic#15: blacks teams merit races thru selection junior earned ranks groups
Topic#16: interestingly damned awfully astounding unreal incidentally screwed rendered alas instantly
Topic#17: original version book classic read novel adaptation written king sequel
Topic#18: adding sticking thread added process easier hang repeating suspend pieces
Topic#19: recommendation months raped policeman restored retired morning month weeks six
Topic#20: sorry understands regret lately teenager fond disliked worry unhappy troubles
Topic#21: show first years series time see saw since still now
Topic#22: records bollywood pulp futuristic romp bars circus punk hardcore gems
Topic#23: huh gotta cried swear guessed shouting shake fooled dude yelling
Topic#24: builds adds reaches agrees teaches introduces marries explains resembles threatens
Topic#25: rat tank burn duck trees dirt rabbit tree snake burning
Topic#26: film one films scenes director also time story plot even
Topic#27: brad ron betty matt dan ryan flynn glover anderson ann
Topic#28: war american world black white documentary history america people political
Topic#29: family young kids school father old girl children child mother
Topic#30: music musical songs song dance voice dancing singing rock stage
Topic#31: familys anyones hitchcocks wouldve shouldve everyones couldve wifes itll expect
Topic#32: great best good role cast comedy actor love character actors
Topic#33: active musician venture learning overcome teaching taught remained accomplished concepts
Topic#34: sport gross caliber definition mere waves design games rip aforementioned
Topic#35: women sex girls woman female men violence sexual gay scenes
Topic#36: killer death murder police thriller cop crime kill mystery michael
Topic#37: abc holiday aired eve midnight remake introduce mtv broadcast began
Topic#38: movie movies watch watching acting plot story scenes wifes familys
Topic#39: just like really bad good get one think see even
Topic#40: man back gets two get takes house around away
Topic#41: alcoholic wine cruise vegas serving con businessman beverly bent california
Topic#42: tech korean victory capital temple wwii riot seconds empire dragon
Topic#43: recorded guitar album tracks reviewer recording blues noted singers sung
Topic#44: horror action effects special budget evil gore blood low fight
Topic#45: hoot tarantino bergman dracula buster creator mario olds hack buff
Topic#46: origin cultural versus dracula buster creator mario olds hack buff
Topic#47: confident keen tad flair shy lyrics smart impressed nod calm
Topic#48: shoddy tasteless formulaic imaginable inane unfunny tripe overacting yawn incompetent
Topic#49: doom beware compassion foul mayhem rides subtlety cue awe wicked
Topic#50: simultaneously readers mute dimension complexity critical derivative essence perspective account
<table>
<thead>
<tr>
<th>Topic</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic#1</td>
<td>cooper india family victor town indian dorothy uncle alice jake</td>
</tr>
<tr>
<td>Topic#2</td>
<td>animated animation disney voiced lion bugs batman cartoon cartoons anime</td>
</tr>
<tr>
<td>Topic#3</td>
<td>christmas disney musical songs bill timeless prince art rock loved</td>
</tr>
<tr>
<td>Topic#4</td>
<td>gags gag footage channel jokes television pilot nostalgic smoking comedy</td>
</tr>
<tr>
<td>Topic#5</td>
<td>rock magazine christian chris daddy roger access page jesus school</td>
</tr>
<tr>
<td>Topic#6</td>
<td>jennifer eva comedy jokes genre plot humor sex nicole pie</td>
</tr>
<tr>
<td>Topic#7</td>
<td>christ games jesus religion bible game freeman christian theory vampires</td>
</tr>
<tr>
<td>Topic#8</td>
<td>eddie woody allen murphy comedians comedian keaton funnier sandra williams</td>
</tr>
<tr>
<td>Topic#9</td>
<td>noises father fear heroine summer toys humor adolescent premise trick</td>
</tr>
<tr>
<td>Topic#10</td>
<td>wedding prince grace betty affair sally marry opera secretary marriage</td>
</tr>
<tr>
<td>Topic#11</td>
<td>winner neil star oscar debut hit screenplay jeff ben stars</td>
</tr>
<tr>
<td>Topic#12</td>
<td>politically political racist garbage black wing jokes free foul stereotype</td>
</tr>
<tr>
<td>Topic#13</td>
<td>children gay baby garbage andy virgin parents fox offensive abuse</td>
</tr>
<tr>
<td>Topic#14</td>
<td>jokes jackson lisa predictable recycled wasted murphy writers williams rock</td>
</tr>
<tr>
<td>Topic#15</td>
<td>waste crap worst costs garbage horrible wasting ashamed sucks pile</td>
</tr>
<tr>
<td>Topic#16</td>
<td>erotic sexuality nudity explicit nude porn lesbian sexual sex photos</td>
</tr>
<tr>
<td>Topic#17</td>
<td>seasons season episodes episode abc trek show aired sitcom series</td>
</tr>
<tr>
<td>Topic#18</td>
<td>delightful parker grant gentle comedies henry witty grim delight arthur</td>
</tr>
<tr>
<td>Topic#19</td>
<td>white black dated english clothing costumes costume queen period heroine</td>
</tr>
<tr>
<td>Topic#20</td>
<td>halloween horror Freddy scares slasher carpenter eerie haunted gory scary</td>
</tr>
<tr>
<td>Topic#21</td>
<td>cried ned meryl streep charlie cry touched heartbreaking dan warming</td>
</tr>
<tr>
<td>Topic#22</td>
<td>seagal abc jet tripe hardy stan martial heist arts drivel</td>
</tr>
<tr>
<td>Topic#23</td>
<td>andrews noir detective eastwood harry hopper investigation fbi cop criminals</td>
</tr>
<tr>
<td>Topic#24</td>
<td>Lynch art imagery noir poetry images waters landscapes poetic symbolic</td>
</tr>
<tr>
<td>Topic#25</td>
<td>moore female soap arthur daughter island butler copy nurse pregnant</td>
</tr>
<tr>
<td>Topic#26</td>
<td>laugh sam winner comedy loser cop funniest golden hilarious simon</td>
</tr>
<tr>
<td>Topic#27</td>
<td>bollywood indian india hudson khan soap dialogs indians dialogues opera</td>
</tr>
<tr>
<td>Topic#28</td>
<td>santa homeless andrews truck airport car husband security mom wakes</td>
</tr>
<tr>
<td>Topic#29</td>
<td>spike match angle baseball tag indians hudson stewart ring flynn</td>
</tr>
<tr>
<td>Topic#30</td>
<td>sinatra musicals kelly powell mgm broadway musical dance numbers gene</td>
</tr>
<tr>
<td>Topic#31</td>
<td>christmas santa childrens holiday betty age ann adult children toy</td>
</tr>
<tr>
<td>Topic#32</td>
<td>robin don batman hood villains hurt pacino dick delivery adam</td>
</tr>
<tr>
<td>Topic#33</td>
<td>adaptation emma jane adaptations timothy bbc novel kenneth book bates</td>
</tr>
<tr>
<td>Topic#34</td>
<td>sons philip son daniel father mother jewish hoffman norman davis</td>
</tr>
<tr>
<td>Topic#35</td>
<td>cars rent scared channels vhs lol expecting cage renting cried</td>
</tr>
<tr>
<td>Topic#36</td>
<td>woody disjointed pacing distracting uninteresting development leonard tension subplots hour</td>
</tr>
<tr>
<td>Topic#37</td>
<td>trek alien planet superman science scientists invisible aliens outer scientist</td>
</tr>
<tr>
<td>Topic#38</td>
<td>fantasy christmas magic effects magical santa special holiday childrens child</td>
</tr>
<tr>
<td>Topic#39</td>
<td>gay blah gray spike indulgent lesbian jerk insulting rape racist</td>
</tr>
<tr>
<td>Topic#40</td>
<td>martial arts kung jet hong chan kong ninja jackie sword</td>
</tr>
<tr>
<td>Topic#41</td>
<td>zombies zombie werewolf vampires monster creature dinosaur rubber shark topless</td>
</tr>
<tr>
<td>Topic#42</td>
<td>indie chaplin festival budget pleasantly independent jake distribution spike victor</td>
</tr>
<tr>
<td>Topic#43</td>
<td>computer plot imaginative tech sub slasher budget holes video jennifer</td>
</tr>
<tr>
<td>Topic#44</td>
<td>low code monkey rental store computer scientist laughs budget executive</td>
</tr>
<tr>
<td>Topic#45</td>
<td>drew daughter adult father adults sports parents younger teen kids</td>
</tr>
<tr>
<td>Topic#46</td>
<td>propaganda jews hitler germans nazis soviet nazi civil americans jewish</td>
</tr>
<tr>
<td>Topic#47</td>
<td>baby jokes daughter humor amusing unfunny silly cute comedy funny</td>
</tr>
<tr>
<td>Topic#48</td>
<td>bands metal album punk kids tap band santa nerd school</td>
</tr>
<tr>
<td>Topic#49</td>
<td>bollywood manager eddie rap van store dance tap singer choreography</td>
</tr>
<tr>
<td>Topic#50</td>
<td>eddie album nominated bands concert vhs murphy jackie awards oscars</td>
</tr>
</tbody>
</table>
Effective Neural Topic Modeling with Embedding Clustering Regularization

**DKM+Entropy**

- **Topic#1**: zone european even choppy showing shows evening shy china side
- **Topic#2**: sincerely workers warn thank yes ever seen old way semi
- **Topic#3**: finally morgan drug tape tap tank drugs remember remind reminded
- **Topic#4**: laughter distracting discussion social disbelief disappoint moronic motivation dinosaurs dimension
- **Topic#5**: bin camera hollywoods couldve amazingly ambition pacing paced titled awe
- **Topic#6**: reviewers spend spoke got split gore gordon motion gone spike
- **Topic#7**: expected opened entry sea canada june north installment ticket april
- **Topic#8**: zone period overcome overlooked pace pacino discovering painful painfully paper
- **Topic#9**: naked cousin raped fat neo crazed fans twins indian indians
- **Topic#10**: abandoned tons discovers spoiled died die diane rat near nearby
- **Topic#11**: clothing fitting trained amazingly robot borrowed masterpiece primarily heavily comic
- **Topic#12**: overwhelming disjointed heartfelt flair understandable captures sincere angst spirited dreck
- **Topic#13**: buff dude like saw survivor genuine start estate saving ever
- **Topic#14**: physical day semi ever way old due funny seen harm
- **Topic#15**: zone insipid interestingly interesting interest single intentions intensity intelligence insulting
- **Topic#16**: roof practically underneath tame colour albeit colors shorter color badly
- **Topic#17**: episode identify corruption costs project ignored promise ignore couple couples
- **Topic#18**: getting spring dollars stevens steven message stephen step doctors mickey
- **Topic#19**: rip dogs images discovered chinese illegal objects baseball possibly display
- **Topic#20**: picked patients sandra hour hours howard boat human continually sarah
- **Topic#21**: zone brilliance principal prior brief describing prize description process produced
- **Topic#22**: urge decision rushed agree navy agreed agrees ahead aid defense
- **Topic#23**: won japan spain union april italy south bank north entry
- **Topic#24**: homeless womens disturbed helps cares compassion souls incompetent caring ashamed
- **Topic#25**: episode identify corruption costs project ignored promise ignore couple couples
- **Topic#26**: improved lacked relative reflection reflect stress corporate respect lighting supporting
- **Topic#27**: zone insipid interestingly interesting interest single intentions intensity intelligence insulting
- **Topic#28**: checked growing cole come comes hal gun guessed study comments
- **Topic#29**: rocket rescue slice mates burns strong stronger strongest bus respective
- **Topic#30**: semi seen due old way physical day ever funny sentimental
- **Topic#31**: benefit buying paid guarantee suits patricia wealthy pay paying pays
- **Topic#32**: cult japanese killers sympathy kill kevin justin josh joins jimmy
- **Topic#33**: zone playing physical picking piece battles credited place places plane
- **Topic#34**: canada lower april june entry north installment sea ticket opened
- **Topic#35**: struggle americans note notch barry barrel nostalgic task screen tape
- **Topic#36**: funny day physical semi ever way old due seen zone
- **Topic#37**: zone place phil philosophical philosophy class claire phrase physical claims
- **Topic#38**: invasion army wwi opened june bank join third next fourth
- **Topic#39**: day ever old due seen funny physical semi person
- **Topic#40**: positive crisis elements painful serial television effects seen two causes
- **Topic#41**: gas plenty receives dislike average regarding treatment household safety concern
- **Topic#42**: zone honesty holy hollywood holiday holds shared hoffman hitting hits
- **Topic#43**: system college supported identical skip trial mentioned furthermore actual half
- **Topic#44**: funny ever day old seen way physical due semi relationship
- **Topic#45**: funny ever day semi seen physical way old due psychological
- **Topic#46**: stuff cried intentionally technical innocence anyway tender anything anyone incident
- **Topic#47**: compared conviction howard ted technology sci human officers office hundred
- **Topic#48**: tight understand positive came television seen standard win like semi
- **Topic#49**: ever semi due old funny physical day way seen solution
- **Topic#50**: nominated pass nod nomination suspect willing joined warner jean stanley
Effective Neural Topic Modeling with Embedding Clustering Regularization

ECRTM

Topic#1: students school teacher student sean high specially class texas shelf
Topic#2: disney lion cartoons animation cartoon animated anime animals toy fox
Topic#3: dimensional merely clumsy unpleasant draw potentially consistent handed develop wholly
Topic#4: santa christmas children kids adults adult child parents relax age
Topic#5: stories york season match lifetime davis tony currently respected episodes
Topic#6: wars alien burton dinosaurs outer graphics trek futuristic aliens sci
Topic#7: terrible costs horrible renting sucks awful avoid rented sounded rent
Topic#8: budget low values stinks producing violence gratuitous blast spare lighting
Topic#9: seasons abc episode aired season show episodes network program television
Topic#10: eastwood andrews hopper fbi westerns investigation policeman crime investigating showdown
Topic#11: worst superman hats ever les shorter guitar banned maniac mates
Topic#12: funniest comedies comedy laugh laughing dan black mario white jokes
Topic#13: japanese japan russian wrestling reynolds kim biased marketing industry vincent
Topic#14: album bands concert broadway musicians dancers rap lyrics sung musical
Topic#15: bates adams purchased ann library lasted listed quinn map native
Topic#16: jackie martial chan kung arts kong hong stunts bruce fight
Topic#17: cars swedish car thumbs files cop airport sappy jet tomorrow
Topic#18: bat stuart angela shouldve flynn hoot plague trailer abysmal bath
Topic#19: sucked crappy stupid twins monkey stupidity darn horrid idiotic puppet
Topic#20: vhs bought copy tape rental dvd store dvds video cable
Topic#21: dad son jackson jack father hotel god saved king thinks
Topic#22: games game bond victor chris germany reunion mexico french marie
Topic#23: martha familys dysfunctional cowboy cope illness daniel bergman financial property
Topic#24: titanic waste spike flop dreck advise someday blockbuster junk taylor
Topic#25: grim page gentle timothy understated magnificent captures poignant passionate debut
Topic#26: development unlikeable drew roy main pitt character implausible descent believable
Topic#27: sequel remake original beginning credits improved van scene missed spoilers
Topic#28: ted lou chaplin bridges mel warren russell sandra butler elizabeth
Topic#29: football teams randy kicked jeff ensues airplane thugs roof bus
Topic#30: woody walken jerry kelly hanks allen sinatra tom dances musicals
Topic#31: lynch ireland bang unusual worthwhile rabbit tops twist quotes temple
Topic#32: excited waters comments reading expectations yelling book urge disturbed sticking
Topic#33: excellent fantastic recommend performance highly job brilliant amazing enjoyed definitely
Topic#34: vampires vampire freddy zombies zombie nightmare serial halloween killer slasher
Topic#35: christ christian religious faith theory interviews media holy intellectual studying
Topic#36: jane emma novel adaptation novels book version versions faithful books
Topic#37: freeman justin nicole holly carrie annie austin glover btw matthew
Topic#38: artistic medium landscape breathtaking painting imaginative poetry technique contemporary movements
Topic#39: demons carpenter spooky eerie sleaze myers gothic karen hammer patients
Topic#40: eddie murphy unfunny funnier tacky clown humour yawn gags tripe
Topic#41: lynch ireland bang unusual worthwhile rabbit tops twist quotes temple
Topic#42: rubber cabin ninja rat barrel cave corpses splatter lake predator
Topic#43: festival indie harry welles films buffs film stan lucy gay
Topic#44: mean maybe sorry hate honestly understand saying else noises someone
 Topic#45: half improvement stretch ford hour limit respectable swimming covers portion
Topic#46: heaven cage brave segment earth robot vegas science planet drove
Topic#47: festival indie harry welles films buffs film stan lucy gay
Topic#48: hitler germans soviet civil war wwi russia fought union holocaust
Topic#49: virgin women andy sex sarah male diane kate woman boyfriend
Topic#50: touched bollywood sadness cried joy feelings emotion inspiring relationships warm