论分层主题建模的亲和性、合理性和多样性

摘要

层次主题模型旨在从语料库中发现潜在的主题，并将其组织成层次结构，以便以合适的语义粒度理解文档。然而，现有的研究在构建主题层次时往往面临 Low Affinity, Rationality 和 Diversity 的问题 [5]，这使得文档理解受到限制。为了解决这些挑战，本文提出了一种新的方法——传输计划与上下文感知层次主题模型（TraCo）。与早期简单的主题依赖方法不同，我们提出了一种基于传输计划的依赖方法。该方法通过约束依赖关系，确保其稀疏性和平衡性，同时在构建主题层次时进行正则化，从而提高层次结构的关联性和多样性。此外，我们还提出了一种上下文感知的解耦解码器。与以往的耦合解码不同，该解码器通过解耦解码将不同语义粒度分配给不同层级的主题。这一改进促进了层次结构的理性化。在标准数据集上的实验结果表明，我们的方法超越了现有的最先进基准，显著提高了层次主题模型的关联性、理性和多样性，并在下游任务上取得了更好的表现。

关键词： Topic Model；文本；传输计划

1 引言

与传统的平面主题模型不同，层次主题模型旨在从文档中发现一个主题层次结构 [5]。每个主题都被解释为一组相关的词汇，以代表一个语义概念。该层次结构捕捉主题之间的关系，并按照语义粒度进行组织：较低层级的子主题通常较为具体，而较高层级的父主题则较为抽象。因此，层次主题模型能够以合适的粒度提供对复杂文档的更全面理解。正因为这种优势，它们已广泛应用于文档检索、情感分析、文本摘要以及文本生成等多个下游任务。

现有的层次主题模型主要分为两类。第一类是传统模型，如 hLDA [5] 及其变种 [7]。这些模型通过吉布斯采样或变分推断推导参数，但由于计算成本较高，难以处理大规模数据集。第二类是神经网络模型，包括 HNTM [3]、HyperMiner [12] 及其他模型 [4]。这些模型通常基于变分自编码器（VAE）框架，并利用反向传播加速参数推导 [11]。

然而，这些方法在生成高质量主题层次时面临三个问题：首先是 Low Affinity 问题，子主题与父主题的关联性较低，正如图1左侧所示，父主题与 “army” 相关，而其子主题却包含了与之无关的词汇，如 “game music” 和 “school”。这种 low affinity 的层次结构捕捉到的主题关系不准确。其次是 Low Rationality 问题：子主题与父主题过于相似，而非按预期表现出特定的区别。如图1右侧所示，父主题及其子主题都聚焦于 “image segmentation”，且语义粒度相同。low Rationality 的层次结构提供的主题粒度不够全面。最后是 Low Diversity 问题：兄弟主题重复而非多样。在图1右侧，两个兄弟主题相互重复，导致冗余，暗示可能存在其他未揭示的潜在主题。因此，Low Diversity 的层次结构生成的主题信息不足且不完整。

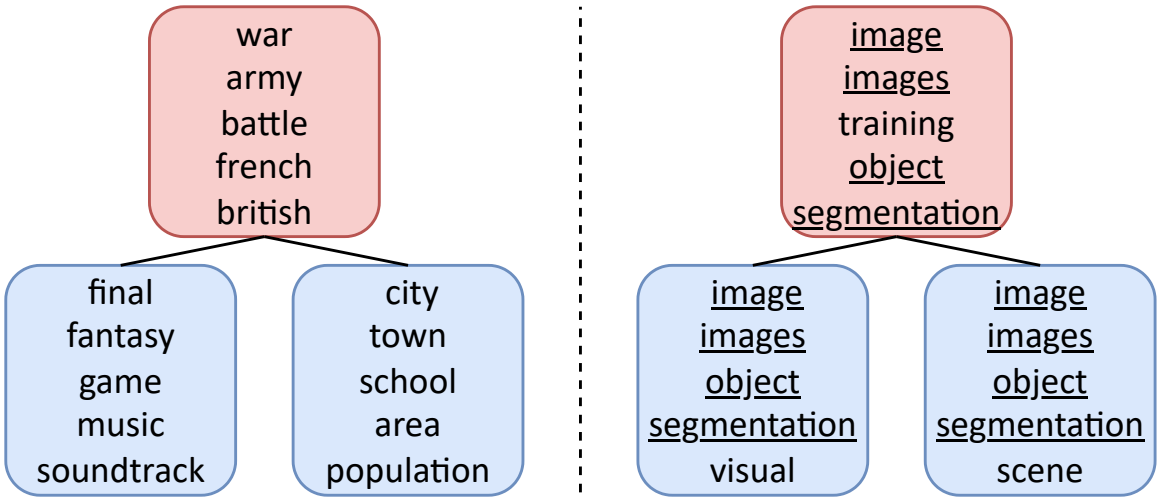


图 1. 展示了 Wikitext-103 和 NeurIPS 数据集中低关联性（左）以及低理性和多样性问题（右）的情况，每个矩形框内展示的是来自 HyperMiner 模型的主题中最相关的词汇，重复的词汇已被下划线标出。

由于这些问题，现有的层次主题模型往往生成低质量的层次结构，这阻碍了文档理解，从而影响了其可解释性和在下游应用中的表现。

为了解决这些挑战，本文提出了一种新颖的神经网络层次主题模型，称为**传输计划与上下文感知层次主题模型**（**TraCo**）。首先，针对 Low Affinity 和 Low Diversity 的问题，我们提出了一种新的**传输计划依赖**（**TPD**）方法。与以往研究中未加约束的依赖方法不同，TPD 将层次主题之间的依赖建模为最优的传输计划，通过约束依赖关系以确保其稀疏性和平衡性。在这些约束依赖的指导下，TPD 进一步对主题层次结构的构建进行正则化：它将子主题拉近其父主题并远离其他主题，避免过度聚集兄弟主题。因此，这种方法提升了子主题与父主题之间的关联性，并改善了兄弟主题的多样性。

其次，为了解决 Low Rationality 的问题，我们进一步提出了一种新颖的**上下文感知解耦解码器**（**CDD**）。与早期工作的耦合解码方法不同，CDD 通过分别使用每一层级的主题对输入文档进行解码，达到了解耦解码的效果。此外，每一层的解码过程中都会引入一个包含来自其上下文层级的主题语义的偏差。这种引入迫使每个层级的主题覆盖与其上下文层级不同的语义。因此，CDD 能够将不同的语义粒度分配给不同层级的主题，从而增强了层次结构的理性。

2 相关工作

2.1 传统的分层主题模型

与 LDA [1] 等平面主题模型不同，Griffiths 等人提出了 hLDA，利用嵌套的中国餐馆过程（nCRP）生成主题层次结构。为了缓解 nCRP 中的单路径结构，Paisley 等人 [8] 提出了嵌套的层次狄利克雷过程。另一种方法，Viegas 等人 [10] 使用了 NMF 与集群词嵌入；Shahid 等人 [9] 则通过引入超球面词嵌入扩展了这一方法，但这些方法并未涉及推断文档的主题分布。

2.2 神经层次主题模型

近年来，神经网络层次主题模型在变分自编码器（VAE）框架下得到了应用。一些模型遵循传统模型，Isonuma 等人（2020）[6] 首次提出了一种树形结构的主题模型，结合了两个简化的双重递归神经网络。Chen 等人 [4] 提出了 nTSNTM，采用了 stick-breaking 过程先验。最近，参数化设置受到了更多关注，即指定层次结构中每一层级的主题数量。Chen 等人 [3] 提出了对主题依赖关系的流形正则化方法。Li 等人使用跳跃连接进行解码，并通过策略梯度方法进行训练。Xu 等人 [12] 将主题和词嵌入建模为超球面空间中的向量。Chen 等人 [2] 使用高斯混合先验和非线性结构方程来建模依赖关系。

我们遵循了流行的参数化设置，但不同的是，我们关注层次主题建模中的低关联性、低理性和低多样性问题。为了解决这些问题，本文提出了传输计划依赖方法来正则化主题层次结构的构建，并提出了上下文感知解耦解码器以分离语义粒度。

3 本文方法

在本节中，我们回顾层次主题建模的基本问题设置和符号表示。随后，我们提出了传输计划依赖方法和上下文感知解耦解码器。最后，我们介绍了我们的传输计划与上下文感知层次主题模型（TraCo）。

3.1 问题设置和符号

考虑一组包含 N 篇文档集合: $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\}$ 其中包含 V 个唯一词汇（词汇表大小）。我们的目标是从该文档集合中发现一个包含 L 个层级的主题层次结构，其中第 ℓ 层包含 $K^{(\ell)}$ 个潜在主题。我们通过依赖矩阵构建该层次结构，这些矩阵描述了两个层级之间主题的层次关系。例如，$\boldsymbol{\varphi}^{(\ell)} \in \mathbb{R}^{K^{(\ell+1)} \times K^{(\ell)}}$ 表示第 ℓ 层和第 $\ell+1$ 层主题之间的依赖矩阵，其中 $\varphi^{(\ell)}_{kk'}$ 表示第 $\ell+1$ 层第 k 个主题与第 ℓ 层第 k' 个主题之间的关系。子主题应当与父主题具有较高的依赖关系，并与其他主题具有较低的依赖关系。与 LDA 类似，我们将每个潜在主题定义为一个词分布（主题-词分布），例如第 ℓ 层第 k 个主题被定义为 $\boldsymbol{\beta}^{(\ell)}_k \in \mathbb{R}^V$。然后，$\boldsymbol{\beta}^{(\ell)} = (\boldsymbol{\beta}^{(\ell)}_1, \ldots, \boldsymbol{\beta}^{(\ell)}_{K^{(\ell)}}) \in \mathbb{R}^{V \times K^{(\ell)}}$ 表示第 ℓ 层的主题-词分布矩阵。

此外，我们在每个层次推断文档-主题分布，即文档中各主题的比例。例如，我们用 $\boldsymbol{\theta}^{(\ell)} \in \Delta_{K^{(\ell)}}$ 表示文档 $\mathbf{x}$ 在第 ℓ 层的文档-主题分布，其中 $\Delta_{K^{(\ell)}}$ 是一个概率 simplex。

3.2 参数化分层潜在主题

首先，我们对层次潜在主题进行参数化。我们将词汇表中的词和各层级的主题都投影到一个嵌入空间中。具体来说，我们有 V 个词嵌入：$\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_V) \in \mathbb{R}^{D \times V}$，其中 D 是维度。类似地，我们有第 ℓ 层的 $K^{(\ell)}$ 个主题嵌入：$\mathbf{T}^{(\ell)} = (\mathbf{t}^{(\ell)}_1, \ldots, \mathbf{t}^{(\ell)}_{K^{(\ell)}}) \in \mathbb{R}^{D \times K^{(\ell)}}$。每个主题（或词）嵌入代表其语义。

为了建模第 ℓ 层的潜在主题，我们计算其主题-词分布矩阵 $\boldsymbol{\beta}^{(\ell)}$，计算公式为：

$$\beta^{(\ell)}_{k,i} = \frac{\exp(-\|\mathbf{t}^{(\ell)}_k - \mathbf{w}_i\|^2/\tau)}{\sum_{k'=1}^{K} \exp(-\|\mathbf{t}^{(\ell)}_{k'} - \mathbf{w}_i\|^2/\tau)} \tag{1}$$

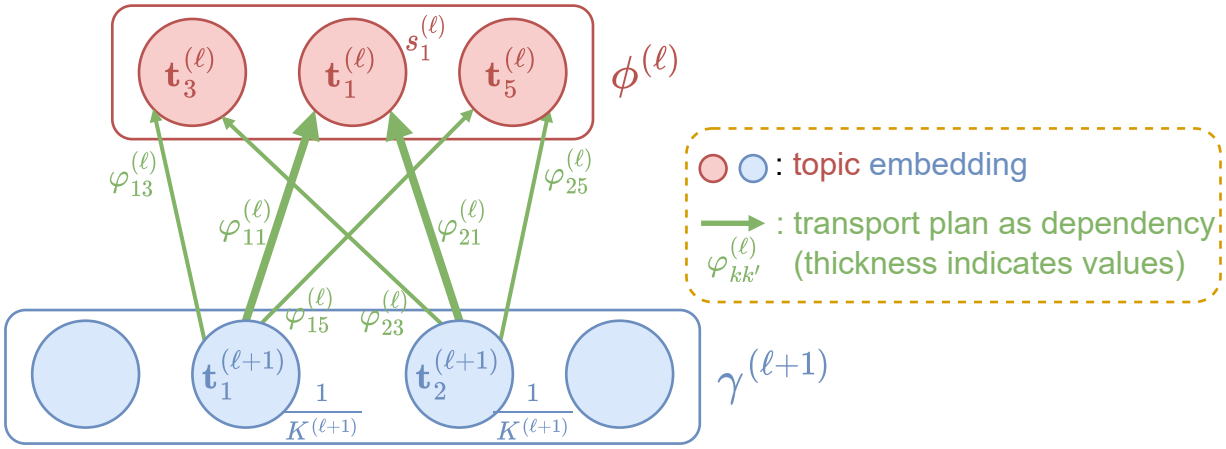
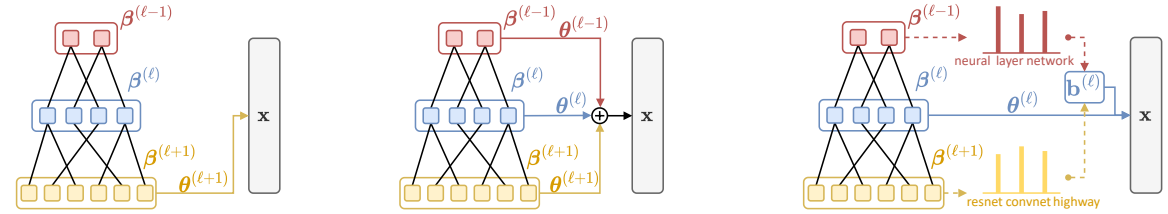


图 2. TPD 的展示，它将依赖关系 $\boldsymbol{\varphi}_{kk'}^{(\ell)}$ 建模从主题嵌入 $\mathbf{t}_k^{(\ell+1)}$ 到 $\mathbf{t}_{k'}^{(\ell)}$ 的运输计划，其中度量 $\gamma^{(\ell+1)}$ 和 $\phi^{(\ell)}$ 受限于 $\mathbf{t}_k^{(\ell+1)}$ 的权重为 $1/K^{(\ell+1)}$，以及 $\mathbf{t}_{k'}^{(\ell)}$ 的权重为 $s_{k'}^{(\ell)}$，这使得 $\mathbf{t}_1^{(\ell+1)}$ 更接近 $\mathbf{t}_1^{(\ell)}$ 并远离其他的，类似地，$\mathbf{t}_2^{(\ell+1)}$ 也是如此。



(a) Lowest-Level Decoder (b) Aggregation Decoder (c) Context-aware Disentangled Decoder

图 3: 层次主题建模中解码器的比较。这里 $\boldsymbol{\beta}^{(\ell)}$ 和 $\boldsymbol{\theta}^{(\ell)}$ 分别是层次 ℓ 的主题-词分布矩阵和文档-主题分布，$\mathbf{x}$ 是待解码的输入文档。(a): 仅使用最低层进行解码。(b): 使用所有层次进行解码。(c): 分别使用每个层次进行解码。

其中 $\beta_{k,i}^{(\ell)}$ 表示第 ℓ 层第 k 个主题与第 i 个词之间的相关性，τ 是一个超参数。在这里，我们将相关性建模为词与主题嵌入之间的欧几里得距离，并在第 ℓ 层的所有主题上进行归一化。

3.3 运输计划依赖性

在这一节中，我们分析了为什么主题层次结构存在低亲和性和低多样性的问题，并提出了一种新的解决方案，称为运输计划依赖性。

为什么会导致 Low Affinity 和 Low Diversity? 之前的模型在解决低亲和性和低多样性的问题时遇到了困难，我们认为，问题的根源在于它们建模主题依赖关系的方式。具体来说，之前的方法通过计算主题嵌入之间的相似度来建模主题之间的依赖关系。例如，大多数研究通过计算主题嵌入的点积来衡量相似度，并通过 softmax 函数对其进行归一化。然而，这些依赖关系没有约束，无法规范化主题层次结构的构建。这种做法引发了低亲和性和低多样性的问题：首先，依赖关系可能缺乏稀疏性，即子主题的嵌入与父主题的嵌入没有足够的接近度。结果，子主题与父主题的关联不足，导致层次结构的亲和性受到损害。然后，依赖关系可能不平衡，即过多的子主题嵌入被聚集在少数几个父主题附近。因此，这些主题变成了同级主题，并且语义相似，这会削弱层次结构的多样性。

基于上述分析，为了解决低关联性和多样性的问题，我们提出了一种新的传输计划依赖

(Transport Plan Dependency) 方法，该方法通过稀疏和平衡的依赖关系来规范化主题层次的构建。图2展示了 TPD 方法的原理。

为了约束依赖关系，我们将其建模为一个特定定义的最优传输问题的传输计划。具体地，我们在 $\ell+1$ 和 ℓ 层的主题嵌入上分别定义离散度量为

$$\begin{aligned}
&\underset{\boldsymbol{\pi}^{(\ell)}\in\mathbb{R}_{+}^{K^{(\ell+1)}\times K^{(\ell)}}}{\arg\min}\quad \mathcal{L}_{\mathrm{OT}_\varepsilon}(\gamma^{(\ell+1)},\phi^{(\ell)}),\quad \text{where}\\
&\mathcal{L}_{\mathrm{OT}_\varepsilon}(\gamma^{(\ell+1)},\phi^{(\ell)})=\sum_{k=1}^{K^{(\ell+1)}}\sum_{k'=1}^{K^{(\ell)}} C_{kk'}^{(\ell)}\pi_{kk'}^{(\ell)}+\varepsilon\pi_{kk'}^{(\ell)}(\log\pi_{kk'}^{(\ell)}-1)\\
&\text{s.t. }\boldsymbol{\pi}^{(\ell)}\mathbb{1}_{K^{(\ell)}}=1/K^{(\ell+1)}\mathbb{1}_{K^{(\ell+1)}},(\boldsymbol{\pi}^{(\ell)})^{\top}\mathbb{1}_{K^{(\ell+1)}}=\mathbf{s}^{(\ell)}.
\end{aligned}\tag{2}$$

其中，$\mathcal{L}_{\mathrm{OT}_\varepsilon}$ 的第一项是原始的最优传输问题，第二項是具有超参数 ε 的熵正则化，用于使得這個问题变得可处理。公式2旨在找到一个传输计划 $\boldsymbol{\pi}^{(\ell)}$，使得其最小化将 $\ell+1$ 层主题嵌入的权重传输到 ℓ 层主题嵌入的总成本，并满足两个约束条件。这里，$\pi_{kk'}^{(\ell)}$ 表示从 $\mathbf{t}_k^{(\ell+1)}$ 到 $\mathbf{t}_{k'}^{(\ell)}$ 的传输权重，我们计算它们之间的传输成本为欧几里得距离：$C_{kk'}^{(\ell)}=\|\mathbf{t}_k^{(\ell+1)}-\mathbf{t}_{k'}^{(\ell)}\|^2$。我们将 $\mathbf{C}^{(\ell)}$ 表示为传输成本矩阵。公式2对 $\boldsymbol{\pi}^{(\ell)}$ 有两个约束条件，以平衡传输权重，其中 $\mathbb{1}_K$ 是一个 K 维的列向量，其所有元素为 1。

为了规范化主题层次结构的构建，我们制定了 TPD 的目标函数，考虑了依赖关系：

$$\mathcal{L}_{\mathrm{TPD}}^{(\ell)}=\sum_{k=1}^{K^{(\ell+1)}}\sum_{k'=1}^{K^{(\ell)}} C_{kk'}\varphi_{kk'}^{(\ell)}\tag{3}$$

其中，我们最小化两个层次之间主题嵌入的总距离，该距离由依赖关系加权。如图2所示，由于依赖关系 $\boldsymbol{\varphi}^{(\ell)}$ 是稀疏的，公式3仅将子主题嵌入推向其父主题，并远离其他主题。这有助于增强学习到的层次结构的亲和性。由于依赖关系也是平衡的，它适当地聚合了子主题嵌入，避免了过多地聚集在一起。这提高了学习到的层次结构的多样性。

3.4 上下文感知的解耦解码器

在本节中，我们探讨了低合理性问题发生的原因。然后，我们提出了一种新颖的上下文感知解耦解码器（CDD）来解决这一问题。

为什么会导致 low rationality？如图1所示，早期方法存在低合理性问题，即子主题与父主题具有相同的粒度，而不是特定于它们。我们认为根本原因在于它们的解码器。如图3.3所示，之前的解码器可以分为两类。第一类是最低层解码器 [?, 12]。它们的解码仅涉及最低层的主题。高层主题通过依赖矩阵是这些最低层主题的线性组合。因此，这使得所有层次的主题纠缠在一起，以覆盖相同的语义粒度，导致合理性低。第二类是聚合解码器 [?, 2–4]。它们的解码涉及所有层次，仍然使所有层次的主题纠缠在一起。这使得这些主题具有相同的语义，因此它们变得相关，但具有相似的粒度。因此，即使具有高亲和性，学习到的层次结构也倾向于具有低合理性。

受到上述启发，我们的目标是为每个层次分离语义粒度，以解决低合理性问题。不幸的是，这并不简单，因为语义粒度是未知的，并且在每个领域中都有所不同。一些研究借用外部知识图谱 [?, ?]，但这种辅助信息无法适应各种领域，且大多不可用。为了克服这一挑战，我们提出了一种新的上下文感知解耦解码器（CDD），图3c展示了 CDD。

为了分离语义粒度，我们提出在每个层次的解码中引入一个上下文主题偏差。我们用可学习变量 $\mathbf{b}^{(\ell)} \in \mathbb{R}^V$ 来表示层次 ℓ 的这种偏差。我们希望它包含层次 ℓ 在层次结构中的上下文层次的主题语义，从而使层次 ℓ 涵盖其他不同的语义。设 $\mathbf{p}^{(\ell)}$ 表示层次 ℓ 的这种主题语义，我们将其建模为：

$$\mathbf{p}^{(\ell)} = \sum_{\ell' \in \{\ell-1, \ell+1\}} \sum_{k=1}^{K^{(\ell')}} \mathrm{topK}(\boldsymbol{\beta}_k^{(\ell')}, N_{\mathrm{top}}). \tag{4}$$

这里 $\mathrm{topK}(\cdot,\cdot)$ 返回一个向量，保留 $\boldsymbol{\beta}_k^{(\ell')}$ 的前 N_{top} 个元素，并将其他元素填充为 0。因此，$\mathbf{p}^{(\ell)}$ 代表上下文主题语义，因为它包含了层次 $\ell-1$ 和 $\ell+1$ 中所有主题的前相关词（如果层次 ℓ 是顶层（最低层），则只涉及层次 $\ell+1$（$\ell-1$））。

然后我们将这些上下文主题语义分配给偏差 $\mathbf{b}^{(\ell)}$：

$$b_i^{(\ell)} = p_i^{(\ell)} \quad \text{where} \quad p_i^{(\ell)} \neq 0. \tag{5}$$

因此，$\mathbf{b}^{(\ell)}$ 包含了上下文层次的主题语义，并且还允许在这些层次未涵盖的语义上进行灵活的偏差学习。

3.5 运输计划和上下文感知的层次主题模型

最后，我们为我们的运输计划和上下文感知的层次主题模型（TraCo）制定了目标函数。主题建模的目标函数：遵循 VAE 的 ELBO ，我们用公式**??**写出主题建模的目标为

$$\begin{aligned}\mathcal{L}_{\mathrm{TM}}(\mathbf{x}) =& \frac{1}{L}\sum_{\ell=1}^{L} -\mathbf{x}^\top \log\big(\mathrm{softmax}(\boldsymbol{\beta}^{(\ell)}\boldsymbol{\theta}^{(\ell)} + \lambda_{\mathrm{b}}\mathbf{b}^{(\ell)})\big) \\ &+ \mathrm{KL}\Big[q(\mathbf{r}|\mathbf{x}) \| p(\mathbf{r}))\Big]\end{aligned} \tag{6}$$

第一项衡量所有层次的平均重构误差；第二项是先验分布和变分分布之间的 KL 散度。

TraCo 的目标函数：基于上述内容，我们通过结合公式**??**写出 TraCo 的总体目标：

$$\min_{\Theta, \mathbf{W}, \{\mathbf{T}^{(\ell)}\}_{\ell=1}^{L}} \lambda_{\mathrm{TPD}} \frac{1}{L-1}\sum_{\ell=1}^{L-1} \mathcal{L}_{\mathrm{TPD}}^{(\ell)} + \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_{\mathrm{TM}}(\mathbf{x}^{(i)}) \tag{7}$$

其中 λ_{TPD} 是一个权重超参数。这里 $\mathcal{L}_{\mathrm{TPD}}^{(\ell)}$ 通过稀疏和平衡的依赖关系规范化主题层次结构的构建；$\mathcal{L}_{\mathrm{TM}}$ 在每个层次分配具有不同语义粒度的主题，并推断文档-主题分布。

4 复现细节

4.1 与已有开源代码对比

原论文采用了传输计划和上下文感知的层次主题模型，在距离度量方面仍然使用的是欧几里得距离。然而，欧几里得空间在处理复杂层次结构时可能存在一定的局限性。因此，为了提升模型在单词和主题表示上的能力，本次复现工作在其基础上引入了双曲空间。双曲空间由于其独特的几何性质，可以更有效地捕捉和表示数据中的层次关系，从而使单词和主题的表示更加准确和灵活。

4.2 实验环境搭建

所有复现实验均在 Windows 系统中的 Pycharm 软件上执行，本文的实验使用了 pytorch1.7 版本和 python3.8 的环境，为了搭建主题模型，还安装了 4.3 版本的 gensim 和 1.5 版本的 scipy 库。

4.3 创新点

原论文的距离度量使用的是欧几里得距离，我把它改为了双曲空间作为度量距离，通过在双曲嵌入空间中测量词和主题之间的距离，模型被鼓励更好地捕捉词之间的潜在语义层次结构。

Lorentz model：洛伦兹模型（也称为双曲面模型）是一个具有曲率 C（$C<0$）的 n 维双曲空间的模型，定义为黎曼流形 $\mathcal{L}^n=(\mathcal{H}^n,g_l)$, 其中 $\mathcal{H}^n=\{\boldsymbol{x}\in\mathbb{R}^{n+1}:\langle\boldsymbol{x},\boldsymbol{x}\rangle_{\mathcal{L}}=1/C\}$ 和 $g_l=\operatorname{diag}\left([-1,\mathbf{1}_n^\top]\right)$. $\langle\cdot,\cdot\rangle_{\mathcal{L}}$ 表示 洛伦兹内积。设 $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{n+1}$, 由 g_l 诱导的洛伦兹内积计算为

$$\langle\boldsymbol{x},\boldsymbol{y}\rangle_{\mathcal{L}}=\boldsymbol{x}^\top g_l\boldsymbol{y}=-x_0y_0+\sum_{i=1}^{n}x_iy_i \tag{8}$$

4.4 数据集描述

我们使用以下基准数据集进行实验：(i) **NeurIPS** 包含 1987 年至 2017 年在 NeurIPS 会议上发表的论文。(ii) **ACL** 是 1970 年至 2015 年从 ACL 文献库中收集的论文集。(iii) **NYT** 包含《纽约时报》的新闻文章，共 12 个类别。(iv) **Wikitext-103** 包括维基百科的文章。(v) **20NG** 包含带有 20 个标签的新闻文章。

4.5 评估指标

我们考虑以下指标来评估主题层次结构：

(i) **P**arent and **C**hild Topic **C**oherence (**PCC**) 表示父主题和子主题之间的一致性。我们使用 CLNPMI 来衡量它。CLNPMI 计算父主题及其子主题中每两个词的 NPMI [**?**]。(ii) **P**arent and **C**hild Topic **D**iversity (**PCD**) 衡量父主题及其子主题之间的多样性。PCC 和 PCD 共同验证父主题和子主题是否相关且覆盖不同的语义粒度。这评估了主题层次结构的合理性。(iii) **P**arent and **n**on-**C**hild Topic **D**iversity (**PnCD**) 衡量父主题及其非子主题之间的多样性。它验证子主题是否仅与其父主题具有高亲和性。(iv) **S**ibling Topic **D**iversity (**SD**) 衡量兄弟主题之间的多样性。注意，PCD 不能替代 SD，因为父主题可能有重复的子主题。

5 实验结果分析

在本节中，我们进行实验以展示方法的有效性。

我们复现的结果与论文结果并无太大差异，表1 描述了主题层次结构的质量结果，我们有以下观察：(**i**) **TraCo 模型显示出更高的亲和性**。可以看到，TraCo 在 PCC 和 PnCD 方面显著超过了所有基线模型。这表明 TraCo 层次结构中的父主题与其子主题更相关，并且与非子主题不同，显示出其增强的亲和性。(**ii**) **TraCo 模型获得了更好的合理性**。除了最佳的

Model	NeurIPS				ACL				NYT				Wikitext-103				20NG			
	PCC	PCD	SD	PnCD	PCC	PCD	SD	PnCD	PCC	PCD	SD	PnCD	PCC	PCD	SD	PnCD	PCC	PCD	SD	PnCD
nTSNTM	-0.348‡	0.603‡	0.195‡	0.566‡	-0.214‡	0.674‡	0.268‡	0.653‡	-0.450‡	0.501‡	0.193‡	0.479‡	-0.026‡	0.816‡	0.500‡	0.777‡	-0.089‡	0.745‡	0.323‡	0.765‡
HNTM	-0.214‡	0.719‡	0.410‡	0.775‡	-0.095‡	0.867‡	0.568‡	0.887‡	-0.137‡	0.757‡	0.380‡	0.723‡	-0.190‡	0.903‡	0.637‡	0.941‡	-0.332‡	0.832‡	0.425‡	0.796‡
NGHTM	0.014‡	0.905‡	0.635‡	0.954‡	0.055‡	0.902‡	0.633‡	0.947‡	-0.026‡	0.816‡	0.351‡	0.887‡	0.054‡	0.933‡	0.548‡	0.956‡	-0.011‡	0.831‡	0.446‡	0.863‡
SawETM	-0.093‡	0.785‡	0.816‡	0.986‡	-0.095‡	0.772‡	0.782‡	0.977‡	-0.234‡	0.641‡	0.680‡	0.970‡	-0.190‡	0.709‡	0.683‡	0.931‡	-0.332‡	0.563‡	0.543‡	0.945‡
DCETM	-0.361‡	0.605‡	0.485‡	0.858‡	-0.353‡	0.584‡	0.387‡	0.804‡	-0.041‡	0.802‡	0.756‡	0.978‡	-0.522‡	0.471‡	0.344‡	0.506‡	-0.085‡	0.742‡	0.644‡	0.900‡
ProGBN	-0.119‡	0.746‡	0.576‡	0.976‡	-0.058‡	0.781‡	0.611‡	0.976‡	-0.049‡	0.753‡	0.614‡	0.983‡	0.068‡	0.885‡	0.707‡	0.983‡	-0.009‡	0.780‡	0.626‡	0.981‡
HyperMiner	-0.084‡	0.771‡	0.808‡	0.991‡	-0.063‡	0.757‡	0.824‡	0.990‡	-0.229‡	0.638‡	0.713‡	0.984‡	-0.207‡	0.703‡	0.685‡	0.949‡	-0.256‡	0.604‡	0.584‡	0.959‡
TraCo	**0.077**	**0.958**	**0.972**	**0.999**	**0.081**	**0.932**	**0.967**	**0.999**	**-0.021**	**0.946**	**0.946**	**0.998**	**0.167**	**0.947**	**0.960**	**0.999**	**0.037**	**0.895**	**0.894**	**0.997**

表 1. 主题层次结构质量结果。PCC 和 PCD 分别表示父主题和子主题之间的一致性和多样性；PnCD 是父主题和非子主题之间的多样性；SD 是兄弟主题之间的多样性。最佳结果以粗体显示。上标 ‡ 表示 TraCo 的增益在 0.05 水平上具有统计显著性。

PCC 外，TraCo 模型在 PCD 方面也达到了最佳，与所有基线模型相比。例如，在 NeurIPS 上，TraCo 的 PCC 为 0.077，PCD 为 0.958，而亚军分别为 0.014 和 0.905。这表明父主题和子主题不仅包含相关语义，还具有不同的粒度，这显示出我们方法的更高合理性。(**iii**) **TraCo 模型实现了更高的多样性**。Table 1 显示 TraCo 模型在 SD 方面优于基线模型。例如，NGHTM 在 NYT 上的 PCC 得分接近，但 TraCo 达到了更高的 SD（0.946 对 0.351）。这表明 TraCo 模型产生了更多多样化的兄弟主题，而不是重复的主题。

6 总结与展望

在本文中，我们提出了用于层次主题建模的 TraCo。TraCo 使用传输计划依赖方法来解决低亲和性和多样性问题，并利用上下文感知的解耦解码器来缓解低合理性问题。实验表明，TraCo 能够持续优于基线模型，生成质量更高的主题层次结构，在亲和性、多样性和合理性方面显著改善。特别是，TraCo 在下游任务中表现更佳，能生成更准确的文档主题分布。

参考文献

[1] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[2] HeGang Chen, Pengbo Mao, Yuyin Lu, and Yanghui Rao. Nonlinear structural equation model guided gaussian mixture hierarchical topic modeling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10377–10390, 2023.

[3] Ziye Chen, Cheng Ding, Yanghui Rao, Haoran Xie, Xiaohui Tao, Gary Cheng, and Fu Lee Wang. Hierarchical neural topic modeling with manifold regularization. *World Wide Web*, 24:2139–2160, 2021.

[4] Ziye Chen, Cheng Ding, Zusheng Zhang, Yanghui Rao, and Haoran Xie. Tree-structured topic modeling with nonparametric neural variational inference. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International*

Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 2343–2353, 2021.

[5] Thomas Griffiths, Michael Jordan, Joshua Tenenbaum, and David Blei. Hierarchical topic models and the nested chinese restaurant process. *Advances in neural information processing systems*, 16, 2003.

[6] Masaru Isonuma, Junichiro Mori, Danushka Bollegala, and Ichiro Sakata. Tree-structured neural topic model. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 800–806, 2020.

[7] Joon Hee Kim, Dongwoo Kim, Suin Kim, and Alice Oh. Modeling topic hierarchies with the recursive chinese restaurant process. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 783–792, 2012.

[8] John Paisley, Chong Wang, David Blei, and Michael I Jordan. A nested hdp for hierarchical topic models. *arXiv preprint arXiv:1301.3570*, 2013.

[9] Simra Shahid, Tanay Anand, Nikitha Srikanth, Sumit Bhatia, Balaji Krishnamurthy, and Nikaash Puri. Hyhtm: Hyperbolic geometry based hierarchical topic models. *arXiv preprint arXiv:2305.09258*, 2023.

[10] Felipe Viegas, Washington Cunha, Christian Gomes, Antônio Pereira, Leonardo Rocha, and Marcos Goncalves. Cluhtm-semantic hierarchical topic modeling based on cluwords. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 8138–8150, 2020.

[11] Xiaobao Wu, Thong Nguyen, and Anh Tuan Luu. A survey on neural topic models: methods, applications, and challenges. *Artificial Intelligence Review*, 57(2):18, 2024.

[12] Yi Xu, Dongsheng Wang, Bo Chen, Ruiying Lu, Zhibin Duan, Mingyuan Zhou, et al. Hyperminer: Topic taxonomy mining with hyperbolic embedding. *Advances in Neural Information Processing Systems*, 35:31557–31570, 2022.