

文本样式和主题框架引导下的 大模型辅助儿童新闻生成

杜晓蒙¹, 于东^{1,*}, 刘鹏远^{1,2}

1.北京语言大学信息科学学院, 北京, 100083

2.国家语言资源监测与研究平面媒体中心, 北京, 100083

202221198701@stu.blcu.edu.cn, yudong@blcu.edu.cn, liupengyuan@blcu.edu.cn

摘要

主流新闻内容多针对成年人设计, 不易于儿童理解, 难以满足其阅读需求。对此, 我们提出了一种基于主题的儿童新闻篇章结构框架 (TNC-LLM)。该框架融合了文本样式定义 (TSD) 和主题类别定义 (TCD) 两大核心模块, TSD模块采用多种机器学习算法, 从不同粒度分析文本样式风格和段落布局等特点, TCD模块针对不同主题进行了内容分析, 以揭示儿童新闻的写作特点和内容的倾向性, 确保内容的教育性和适宜性。本文实验主要评估了ChatGPT3.5等四个模型在将成年人新闻转换为面向儿童的新闻的性能。实验结果表明, TNC-LLM在儿童新闻内容生成任务中对内容的准确性、文本的趣味性以及教育性等关键维度有显著提升。此外, 该框架具有普适性, 能够应用于不同类型的大型语言模型。

关键词: 文本生成; 大型语言模型; 儿童新闻

Text Styles and Thematic Framework Guided Large Modeling to Aid Children's News Generation

Xiaomeng Du¹, Dong Yu^{1,*}, Pengyuan Liu^{1,2}

1.Faculty of Computer Science, Beijing Language and Culture University, Beijing, 100083

2.National Language Resources Monitoring and Research Center for Print Media, Beijing, 100083

202221198701@stu.blcu.edu.cn, yudong@blcu.edu.cn, liupengyuan@blcu.edu.cn

Abstract

Mainstream news content is mostly designed for adults, which is not easy for children to understand and difficult to meet their reading needs. In this regard, we propose a topic-based children's news chapter structure framework (TNC-LLM). The framework integrates two core modules, Text Style Definition (TSD) and Theme Category Definition (TCD). The TSD module adopts various machine learning algorithms to analyze the characteristics of text style style and paragraph layout from different granularities, and the TCD module conducts content analysis for different topics to reveal the writing characteristics and content tendency of children's news and to ensure that the

*为通讯作者

基金项目:2023年国际中文教育研究课题重点项目 (23YH38B); 北京语言大学梧桐创新平台(中央高校基本科研业务费)(21PT04)

©2024 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

content is educational and appropriate. The experiments in this paper mainly evaluate the performance of four models, including ChatGPT3.5, in converting adult news into child-oriented news. The experimental results show that TNC-LLM significantly improves the key dimensions of content accuracy, text interest, and educational appropriateness in the children's news content generation task. In addition, the framework is universal and can be applied to different types of large-scale language models.

Keywords: Text Generation , Large Language Models , Children's News

1 引言

阅读新闻是儿童提高社会能力和阅读能力的重要途径。皮亚杰(1981)研究表明儿童读物应与儿童所处的认知发展阶段相匹配,儿童读物应根据儿童的年龄、认知水平和兴趣进行个性化设计(Pelletier, 2004; 饶芳芳, 2023)。现有的新闻报道大多以成年人为主要受众,其形式结构上不满足儿童阅读需求,难度上与儿童阅读能力不匹配。如图1所示,成年人版新闻和儿童版的新闻在语言风格、段落布局等有着明显的差异性。如何利用技术手段实现图1的成年人新闻到儿童新闻转写,成为具有应用价值的研究课题。此外,目前面向儿童的新闻文本数量有限,无法满足儿童阅读新闻的需求。

成年人版新闻:

太平洋深处的“第八大陆”,到底是什么来头?卫星图像显示,在北美洲西海岸至夏威夷群岛之间的北太平洋上,存在着一个地图上从未有过标注的巨型“大陆”,面积超过160万平方公里。但奇怪的是,这个“第八大陆”上并没有人居住,世界各国也无意争抢占领,反而唯恐避之而不及……这背后究竟隐藏着怎样的秘密?众所周知,地球七大洲、四大洋,三分陆地、七分海洋的格局以稳定存在的上亿年之久,而这块“第八大陆”就位于美国西海岸和夏威夷之间,面积相当于两个得克萨斯州,是太平洋上一片由400万吨塑料垃圾组成的漩涡,完全由海洋垃圾聚集而成!受到洋流的影响,整个太平洋内的垃圾塑料都会不断向这块“大陆”聚集,每天都在吸收着周围的海洋垃圾,因此面积也一天比一天变大。此外,海上风暴可能会进一步导致原本块状的塑料垃圾变得更加分散,从而产生更严重的生态灾难。一项调查显示,所有人类的生活用品都可以在这块“大陆”上找到,也让我们在内心意识到如今的海洋污染也已经达到非常可怕的地步……

(来源:百度TA说)

儿童版新闻:

“第八大陆”究竟什么来头?
很多同学都知道,意大利探险家哥伦布发现了新大陆
你知道吗?在广阔的太平洋里,也有一个“新大陆”,人们叫它
“第八大陆”
最近,科学家研究发现,“第八大陆”上竟然有一个“部落”
一、奇特的“部落”
清晨,第一缕阳光点亮北太平洋。……
(节选《快乐少年报》)

ChatGPT 3.5:

在太平洋深处,有一个神秘的地方被称为“第八大陆”。这个大陆面积超过160万平方公里,相当于两个得克萨斯州那么大!但你知道吗?这个“第八大陆”并不是由土地和山川构成的,而是由海洋垃圾组成的!
据卫星图像显示,这片“大陆”位于北美洲西海岸和夏威夷之间的北太平洋上,它完全由400万吨塑料垃圾聚集而成,就像一个巨大的垃圾漩涡,每天,整个太平洋内的塑料垃圾都会被洋流带到这里,使得这个“大陆”的面积越来越大……

(ChatGPT 3.5 生成)

图 1. 同一主题新闻的成年人版和对应儿童版示例

目前的研究领域中,针对将文本内容转化为适合儿童阅读理解的版本,主要采用了文本简化的方法。通过将原文中的复杂词汇替换为更简单、更易于儿童理解的词语,来降低文本的阅读难度(Xu et al., 2018)。在长文本简化任务上,采用将长篇文本内容压缩成简短、精炼的摘要的方法来降低文本难度(Zhong et al., 2020)。尽管这些方法在提高文本可读性方面取得了一定的成效,但它们在内容的敏感性、教育价值、语言表达方式、以及段落排版布局等方面,常常忽视了对儿童认知发展特性和兴趣偏好的系统性考虑和深入分析。

大型语言模型(LLM)通过在海量文本数据上的预训练,展现了卓越的语义理解和生成能力,这使得它们在文本生成和文本创作方面表现出色(Zhao et al., 2023)。无需特定任务训练,可以有效的解决小模型在文本生成所面临的问题如上下文连贯性等。在新闻文本生成方面,大型语言模型已经被应用于自动化地生成新闻报道。例如,一些新闻机构利用LLM来撰写财经报告、体育赛事结果等结构化新闻(Newman, 2024)。此外,LLM也被用于辅助记者进行研究和内容创作,通过生成新闻摘要或提出报道角度来提高工作效率。这些应用为如何利用LLM生成适合儿童阅读的新闻内容提供了有价值的启示。如图1所示,虽然可以通过简单的提示让ChatGPT 3.5基于成年人新闻内容生成适合儿童阅读的版本,但这种方法产生的文本往往只是对原文的简化,并没有充分考虑到儿童新闻在语言表达、风格以及段落结构等方面的特殊需求。

针对以上问题，我们提出了一种儿童新闻文本样式定义（TSD）模块。该模块专注于捕捉适合儿童阅读的新闻文本特征，这些特征覆盖了字词选择、句子结构、段落组织等多个层面，TSD模块能够协助LLM基于成年人新闻内容，生成难度适宜、适合儿童年龄的新闻文本。该模块还能优化文本的段落布局，使其更加符合儿童的阅读习惯。此外，本研究还针对不同类别的儿童新闻内容进行了深入分析，提出了主题类别定义(TCD)模块，旨在探讨各种类别下儿童新闻的写作特色。通过细致地剖析每个类别特点为儿童新闻创作提供更为精准的指导和参考。通过融合TSD和TCD两个模块的优势，我们借助LLM在语义理解和长文本生成方面的先进能力，成功地将成年人新闻文本转换成既有趣又具有教育意义，同时适合儿童阅读的新闻文本。我们基于此构建了儿童新闻文本的数据集，并构造了与儿童新闻主题相对应的成年人版的新闻文本。

本文的贡献：第一，提出了基于主题的儿童新闻篇章结构(TNC-LLM)框架，这是大型语言模型在儿童新闻文本生成领域的首次深入应用，经过大量的实验证明，TNC-LLM在文本内容以及趣味性、教育性等多方面得到了显著的提升。第二，设计了文本样式定义(TSD)模块，该模块集合了多种机器学习的算法，可以在字、词和句级别熟练的总结文本样式和难度特征，在段落级别可以准确地总结出文本的布局和格式特点。该模块不仅提升了LLM在文本内容的保留度，还可以将难度适配到小学生阅读水平。第三，设计了主题类别定义（TCD）模块，旨在深入剖析各类新闻内容的特点和撰写风格。该模块能够精准捕捉儿童新闻报道的倾向性、语言风格和内容展现形式，进而基于成年人文本进行动态结构转换，生成既符合儿童兴趣爱好又积极正面的儿童新闻文本，同时传递了有益的主题和价值观。

2 相关工作

可控文本生成（CTG）任务是指生成符合语法规则和语义需求，且满足给定约束的自然语言文本，具有重要应用价值。可控文本生成领域在情感控制、主题导向和写作风格调整等关键领域已经取得了显著进展。在风格转换领域(Xu et al., 2018)开创性地提出了一种循环强化学习模型，有效解决了文本风格转换中缺乏成对样本的情感转换难题。(Li et al., 2018)聚焦于风格转换过程中的文本属性，通过替换这些关键词汇可以直接改变文本的性质属性。尽管上述两项研究在风格转换方面取得了成就，但它们主要适用于简单句式结构的转换，且生成内容较为单一。随着需求的增加，研究者寻求更多样化和可控的文本生成技术。(Qian et al., 2022)研究提出了一种使用GPT-2模型的可控文本生成方法，通过属性向量引导生成过程，实现了细致的控制，提高了生成控制的精度。(Dathathri et al., 2020)中采用了基于预训练模型精调的可控文本生成方法，通过在预训练模型上添加属性判别模型，实现了在不重新训练整个模型的情况下对文本新属性的控制。然而，在新闻文本生成上仅仅依赖可控文本生成直接输出的内容，并不能确保其事实准确性。

故事生成任务是人工智能和自然语言处理领域的一个关键研究方向，旨在赋予计算机自动创作出既连贯又引人入胜的故事的能力。(Lebowitz, 1985)是故事生成领域的开创性工作之一，他提出了一种基于规划和学习的故事生成方法，将故事生成过程视为一个规划问题。随后，(Fan et al., 2018)提出了一种创新的故事生成框架，通过整合分层生成、模型融合和自注意力机制，显著提升了故事生成的质量。(Liu et al., 2020)研究中提出了一个以角色为中心的神经网络模型，通过明确捕捉人物信息以及情节与人物之间的关系，增强了故事生成的可解释性和一致性。尽管GPT-2等大型预训练语言模型在文本生成上表现出色，但它们在可控性上仍有不足。在需要知识驱动的生成任务，如故事创作时，这些模型可能会生成逻辑或语法错误的内容。为此(Xu et al., 2020)提出了一种新的框架，该框架允许将外部知识动态地融入语言模型，实现对文本生成过程的精细控制。为了应对现有语言模型在生成长篇故事时的挑战，(Park et al., 2023)提出了一个新的任务——长度可控的故事生成。该任务通过递归地生成故事片段，同时确保整体的连贯性和完整性。通过定量分析，Park证明了其方法LongStory在生成较长故事方面的优势，尤其是在提升连贯性、完整性、相关性以及减少重复性等方面。

经过大规模语料库的预训练，大型语言模型具备了处理各类任务的潜力(Wei et al., 2022)。然而，这种潜力在执行特定任务时可能并不总是明显。因此，设计恰当的任务指令或特定的上下文策略来激发这些潜能变得极为关键。(Chen et al., 2023)的研究表明，提示工程在释放LLM的功能方面发挥着核心作用。在(Kaddour et al., 2023)的研究进一步证实了提示在模型输入中的重要性，精心设计的提示能够显著改变输出结果。(Lu et al., 2022)的研究中揭示了提

示中包含的短语和示例序列对模型行为的重大影响。指令调优技术赋予了大型语言模型在缺乏显式样本的情况下执行新任务的能力，通过深化模型对指令的理解，该技术确保了LLMs能够更加高效地适应并完成各种任务。这种方法为我们在儿童文本生成等特定领域的应用提供了新的视角和解决方案。

3 数据集构建

本文选择了两份面向儿童读者的新闻出版物——《快乐少年报》和《阳光少年报》作为主要的数据来源。两份报刊风格相似，内容丰富多样，既适合儿童阅读，也符合我们构建儿童新闻数据集的需求。研究涵盖了《阳光少年报》2021年10月至2023年10月的“榜样”、“中国”、“世界”、“文化”和“科学”板块，以及《快乐少年报》2023年5月至12月的“世界大事”、“世界趣事”、“中国大事”、“中国趣事”和“文化”板块。鉴于两类报刊的新闻分类标准不一，我们重新分类了文本，建立了包括“榜样”、“政治”、“经济”、“文化”、“科学”、“社会”、“体育”和“环境保护”在内的八个类别。为确保分类的准确性，我们首先使用GPT-4模型进行新闻文本分类。我们给定了每个类别的定义和特征，如环境保护：着重呼吁儿童关注环境保护问题，如节约能源、减少垃圾、保护动植物等，以及简单易懂的环保知识和行动建议。随后，我们通过人工审核模型的分类结果进行复核和调整，以确定每篇新闻的类别。此外，根据新闻文本在报刊中的排版长度，我们将文本划分为“儿童长新闻”和“儿童短新闻”。如图4给出了详细的数据集统计结果。目前市面上针对儿童的新闻报刊资料仅限于纸质形式，并且是周期性出版。我们制定了长期的收集计划来扩展我们的儿童新闻语料库。如图3为儿童长新闻数量饼图。

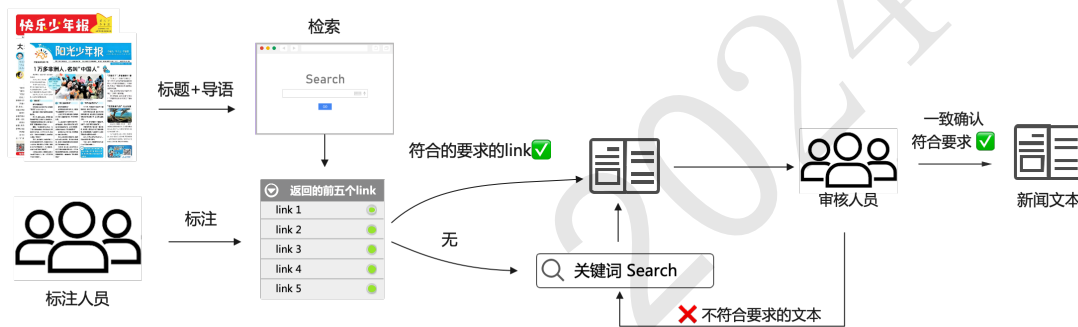


图 2. 成年人版新闻文本标注流程图

在选择成年人新闻文本的过程中，我们采取了自动化检索结合人工审核的方法。如图2为成年人新闻标注流程图。首先提取出儿童新闻标题和摘要作为百度搜索引擎的输入，自动检索并统一返回检索结果的前五个搜索结果，然后由五名硕士生组成的标注团队对返回的结果进行筛选，挑选出与儿童主题相关、内容适宜的新闻，并优先选择知名媒体的新闻。若返回的检索结果中没有符合要求的文本，标注人员根据关键词进行人工检索，找到符合要求的成年人新闻。在初步筛选后，两位语言学背景的硕士生对标注数据进行审核，剔除不合格的文本，并要求标注团队重新标注。通过多轮迭代审核，我们确保了最终结果的准确性和质量。最终，我们收集到了一批高质量、与儿童新闻内容相匹配的成年人新闻文本。所收集的数据数量如图4所示。

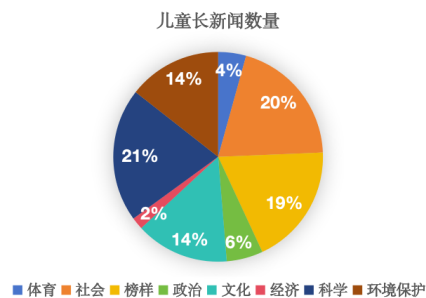


图 3. 儿童长新闻数量饼图

文本类别	儿童长新闻数量	儿童短新闻数量	成年人新闻数量
体育	23	29	46
社会	107	76	178
榜样	100	11	111
政治	30	9	39
文化	77	38	115
经济	10	4	14
科学	110	77	187
环境保护	77	46	123

图 4. 新闻数据集的统计结果

4 方法

4.1 任务定义和整体框架

假设现在有一组数据集 $\{D_i\}_{i=1}^K$ ，每个数据对中包含一篇儿童新闻文本和对应的成年人版新闻报道。我们将 D_i 语料库分为两个部分： D_{i1} 和 D_{i2} ，其中 D_{i1} 中的文本都共享特有的样式 $\{s_i\}_{i=1}^K$ ，以及在不同主题类别上的特征样式 $\{c_i\}_{i=1}^K$ 。任务是生成同时具有目标样式s和c的新文本 x' 。

本文中，我们提出的基于主题的儿童新闻篇章结构框架（TNC-LLM）主要分为三个阶段。我们的语料库 D_i 包含两个部分： D_{i1} 和 D_{i2} ，其中 D_{i1} 是儿童新闻文本语料， D_{i2} 是对应成年人版新闻文本语料。第一阶段我们提出了文本样式定义(TSD)模块，该模块在语料 D_{i1} 的基础上，分别从字、词、句和段落四个层面综合考虑了文本样式的风格特征。由于文本样式定义部分的文本统计是比较随机和广泛的，所以我们将总结出来的文体特征作为文章的文本样式的风格特征。第二阶段我们提出了主题类别定义(TCD)模块，该模块在语料 D_{i1} 的基础上，细致分析不同主题类别在写作板块特征及其内容侧重点的差异性，通过这种方法，我们能够对每个主题类别进行精准的特征描述，并将其归纳整理成一个结构化的主题类别特征库，以便在给定成年人文本 x 进行主题类别分类后，准确地匹配到与之相对应的主题特征。在第三阶段，我们构建儿童新闻文本生成的提示，以指导LLM依据成年人新闻文本的核心内容，生成适合儿童阅读的新闻文本。最后我们对新生成的文本 D'_{i2} 与儿童新闻文本语料 D_{i1} 从多维度进行对比分析，以验证我们方法的有效性。TNC-LLM的整体框架图如图5所示。

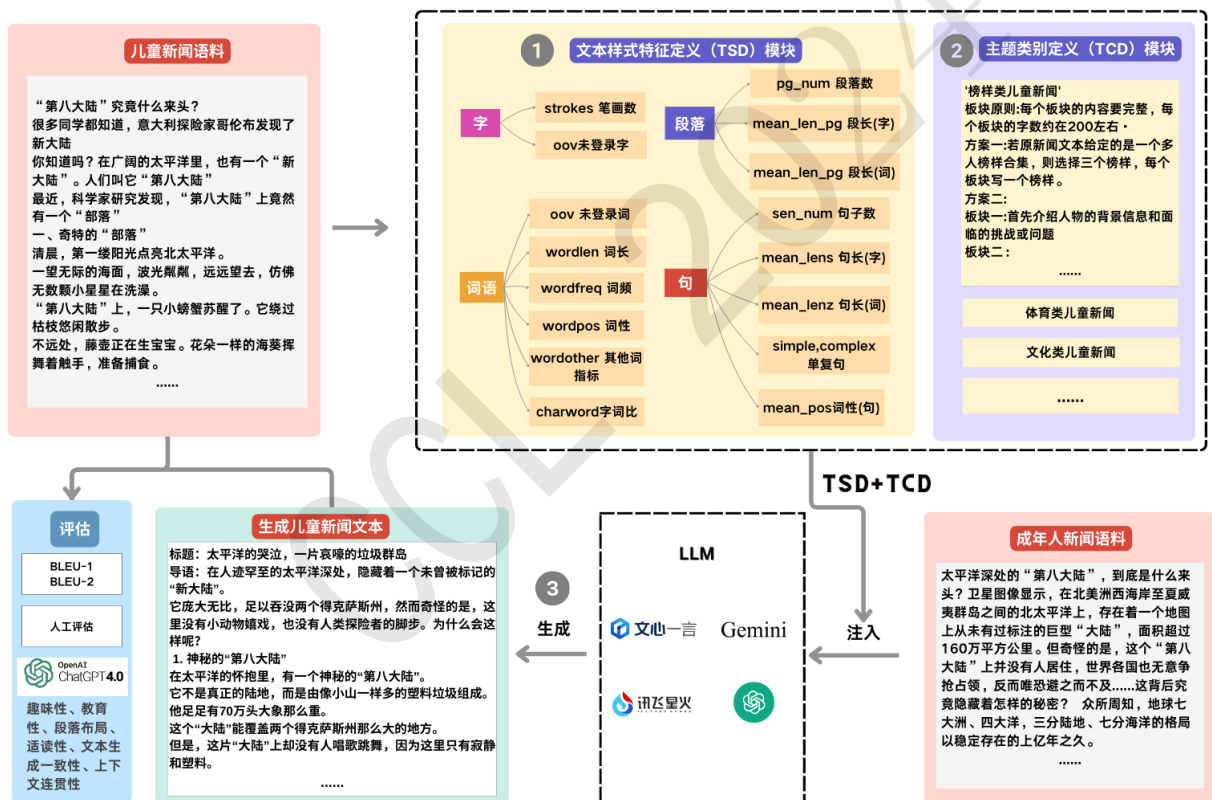


图 5. 基于主题的儿童新闻篇章结构框架（TNC-LLM）图

4.2 文本样式定义(TSD)模块

如图6我们设计了一个专为儿童新闻文本定制的文本样式定义（TSD）模块。该模块从字、词、句及段落四个级别统计了 D_{i1} 的文本样式，并给出了具体的样式定义。TSD模块能够为大型语言模型提供准确的文本样式提示，有利于在成年人新闻内容转化处理上，降低文本难度从而适配儿童的阅读理解能力，还可以优化文章版式布局。

字 在研究汉字结构的复杂性时，笔画数是一个重要的量化指标，通常笔画越多，记忆和认

知的难度越大。本研究统计分析了儿童新闻语料库中汉字的笔画分布，以了解其复杂度特征。同时，考虑到儿童对常见汉字的熟悉度，我们分析了汉字的熟悉度和未登录字的比例，这两个维度对理解儿童汉字认知具有重要意义。

笔画数：综合了(Sung et al., 2015)和(孙汉银, 1992)分类方法，本研究将笔画数小于或等于8的汉字划分为低笔画字，将笔画数大于16的汉字划分为高笔画字，其余的汉字则被归类为中等笔画字(吴思远et al., 2020)。 **汉字熟悉度：**为了准确评估未登录字的比例，本研究参考了六个不同的字表进行量化分析，这些字表包括《现代汉语语料库字频表》、《常用字表》以及该表中前500、前1000、前1500和前2000位的高频字集合。

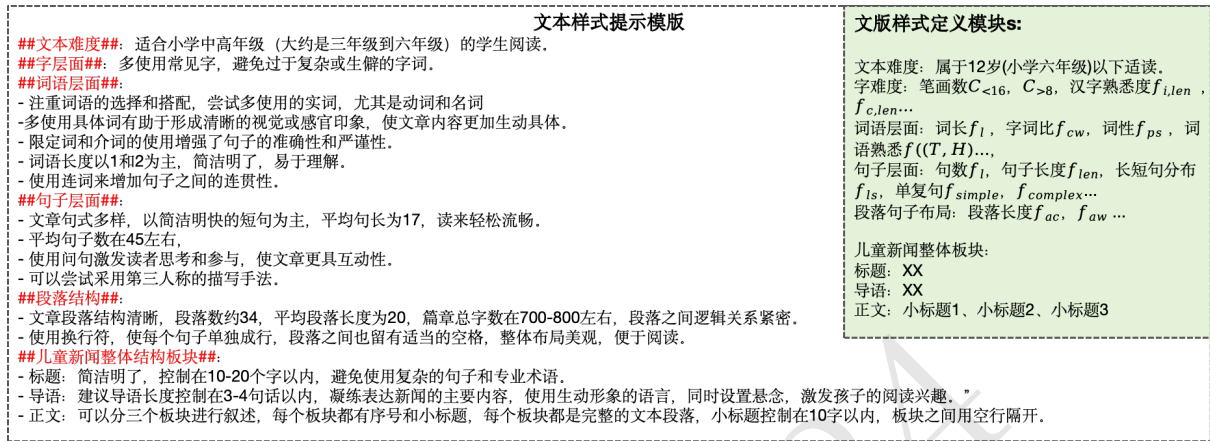


图 6. 文本样式定义模块示意图

词语 词汇是构建文本的基础，对于塑造语言的特色具有极其重要的作用，(Sung et al., 2015)从字词复杂度、词语长度、词语频率、词语计数和词汇多样性等多个维度评估词汇复杂度。通过对文本的词汇进行定量分析可以帮助我们更好地理解文本的特征和含义，本文中从词长、字词比、词频、词性等多个维度来分析。

词长：在本研究中，我们专注于从汉字的角度来量化词长，即通过计算词语中所含汉字的数量来评估其长度。为了量化词长，我们采用公式1，其中 f_l 表示平均词长； $argTop$ 是对加权长度进行求和以得到平均值； U 表示集合的并集操作，在这里是将所有不同长度单词的加权长度合并起来； K_l 是权重因子，它可能与长度 l 的单词在文本中的相对重要性或频率有关， I_l 是长度为 l 的单词的计数， nw 表示文本中不同单词长度的最大值(Tao et al., 2024)。

$$f_l = argTop \left(\bigcup_{i \in \{1, 2, \dots, nw\}} K_l \cdot I_l(len(w_i)) \right) \quad (1)$$

字词比：字词比是指文本中汉字总数与词语总数之间的比率，这一概念由(张卫国, 2006)提出。通过这个量化指标，我们可以更准确地评估和比较不同文本的词汇复杂度。**词频：**我们应用了jieba2中的Cut函数库，对文本 D_{i1} 进行标记，以获得每个单词的词性注释。词类主要分为实词、虚词和否定词等。通过对文章中的词汇进行分类和统计分析，可以推断出词性的分布。**词语熟悉度：**首先通过jieba对 D_{i1} 中文本进行分词，分别统计词出现在《现代汉语语料库词频表》的前500、前1000、前1500 与前2000 高频词的数量。

句 在中文文本中，句子是由一系列词构成的表达单元，它表达了完整的思想或信息，句子的长度和难度等信息可以帮助我们了解文本的结构和组织方式，可以评估文本的阅读难度等级和适用人群，本文从句子数以及句子的长度、长短句、单复句等来分析文本的特征和结构特点。

句子数：我们以句号、问号或感叹号等标点符号结束来统计文本中的句子数。**长短句：**我们评估每个句子的长度，并根据20个字的阈值将其分类为短句或长句，其中 $I_{ls}()$ 是长句和短句计数的函数。如果文章主要包含长句，则被归类为长句较多；相反，它被归类为有更多的短

句。如果长短句数量大致相等，则认为文章是长句和短句的组合(Tao et al., 2024)。

$$f_{ls} = i \in \{1, 2, \dots, n_s\} \operatorname{argmax} \{I_{ls}(s_i)\} \quad (2)$$

单复句：复句通常包含更多的信息和复杂的句子结构，统计复句的数量可以帮助我们评估文本的复杂性。较高的复句比例可能意味着文本内容较为复杂，需要读者有较高的理解能力。为了计算文本中的单复句，用句法分析器构建句子的句法树，这些树表示句子的结构和成分。通过分析这些树，我们可以识别主句和从句，从而判断一个句子是单句还是复句。

段落 段落长度对文本的可读性和理解性有重要影响。较短的段落通常更容易阅读和理解，而较长的段落可能需要读者更多的注意力和努力(Peter, 1996)。分析段落长度可以帮助改进文本的布局和格式，使其更适应目标读者。通过计算文本段落数和平均段落字数和平均段落词数，可以了解文本的结构和组织方式。设 T 为给定的文本，它由一系列段落 $P = (p_1, p_2, \dots, p_m)$ 组成，其中 m 是段落的总数。每个段落 p_i 包含 n_i 个字。段落的平均字数计算公式如下，其中， $|p_i|_{chars}$ 表示段落 p_i 中的字符数（包括空格和标点符号）。统计结果在附录B.3。

$$f_{ac} = \frac{\sum_{i=1}^m |p_i|_{chars}}{m} \quad (3)$$

4.3 主题类别定义(TCD)模块

成年人新闻遵循标准的报道模式，包括标题、导语、正文、背景和结尾。这些报道使用正式语言和专业术语，以及复杂句法，确保信息权威准确。它们常用被动语态以显客观中立，并深入分析事件原因、影响和后果，涵盖政治、经济、社会和文化等关键领域，旨在激发读者批判性思考，提供多元视角和深度讨论。报道采用“倒金字塔”结构，优先展示核心信息，并通过逻辑和过渡词保持文章连贯性，同时融入背景资料和专家意见。

相比之下，儿童新闻的结构更为简化，主要由标题、导语和正文三个部分组成。内容上侧重于儿童日常生活相关主题，如学校、社会和自然等。目的在于简化信息，使其易于理解，重点在于传授基础知识和道德教育，而不是深入分析。儿童新闻在文本表现上追求形象、声音和色彩的丰富性，强调故事性和生动性，运用各种修辞手法来提升语言的吸引力和文本的可读性(Alon-Tirosh, 2014)。在段落布局上，儿童新闻采用分节式结构，每篇新闻都分为三个板块，每个板块讨论一个特定主题或情景，通过具体情境的描述吸引儿童的注意力，传递积极正面的主题和价值观，助力儿童树立正确的人生观和价值观，增强社会责任感。

在深入分析儿童新闻的撰写特点后，我们发现不同主题类别的儿童新闻在内容呈现上各有侧重。因此在本文中我们将儿童新闻划分为不同的主题类别，以便更精确地满足儿童受众的需求和兴趣。该模块针对特定的主题，采用适合该主题的写作风格和内容策略，确保新闻报道既有趣又具有教育意义，同时能够吸引儿童的注意力并促进他们的理解。通过这种方式，儿童新闻能够更有效地传递信息，激发儿童的好奇心和学习兴趣。如图7是主题类别定义模块中环境保护类儿童新闻的提示模版，附录C.1中为完整的主题类别定义(TCD)模块。

环境保护类儿童新闻：

板块原则：每个板块的内容要完整，每个板块的字数约在200左右，每个板块有小标题

板块一：首先介绍背景和问题的存在，让读者了解目前面临的挑战或问题的严重性。

板块二：深入探讨问题带来的具体影响或潜在危害，这一部分旨在增强读者对问题紧迫性的认识。

板块三：提出解决问题的方法、已经采取的措施或呼吁采取的行动。这部分旨在激励读者或社会采取行动，对问题进行缓解或解决。

环境类儿童新闻的撰写特点：

互动性和参与感：通过提问或呼吁读者思考，激发儿童对环境问题的兴趣和参与感，如通过展示积极的改变，激发他们对保护环境的兴趣和责任感。

科普性：在故事化的叙述中融入科学知识和环保理念，既娱乐又教育，帮助儿童读者在轻松愉快的阅读过程中学习到有关环境保护的知识。

生动的比喻和描绘：通过将复杂的环境问题比作人类熟悉的情景(如“地球之肺”生病)，使儿童更容易理解和感受到问题的紧迫性。简化的数据和事实：使用容易理解的数据(如“每过8秒，就有足球场大小的森林消失”)来说明问题，避免了复杂和抽象的统计数据，使信息更加亲近儿童的认知水平。

图 7. 主题类别定义模块中环境保护类儿童新闻的提示模版

5 实验与结果

5.1 实验设计

为了评估我们框架的全面性和适用性，我们选取了四种当前领先的大型语言模型（LLM）作为实验模型。这些模型包括：GPT-3.5-turbo-16k、讯飞星火-v3.5、ERNIE-3.5、Gemini-1.0-pro。通过它们，我们可以全面地评估我们的框架在不同场景下的表现和适用性。我们研究了在Zero-Shot和One-Shot学习场景下，利用文本样式定义模块和主题类别定义模块的组合提示，四个不同的模型生成儿童新闻内容的能力。其中所有模型的温度均为0.5。在Zero-Shot场景中，模型没有接受到任何具体的示例或指导，需要直接根据给定的模块和成年人版新闻文本内容生成儿童新闻文本。Zero-Shot结合文本样式定义模块s和主题类别定义模块c的提示模版在附录A.3。而在One-Shot场景中，模型首先提供一个成年人版新闻文本和对应的儿童新闻文本示例，然后基于这个示例结合给定模块组合生成新的儿童新闻内容。通过这种设置，我们可以评估模型在不同学习条件下的文本生成性能，以及它们理解和应用给定模块提示的能力。此外，我们将采用不提供任何模块提示的情况作为实验的基准线。One-Shot结合文本样式定义模块s和主题类别定义模块c的提示模版在附录A.2中。

5.2 实验数据与评估指标

我们使用的数据集包括两个部分： D_{i1} 中包含534篇长篇儿童新闻文本，以及 D_{i1} 中与长篇儿童新闻文本主题内容相对应的534篇成年人新闻文本。我们的评估主要分为两个方面，一方面是在验证内容和主题的保留，另一方面是评估文本的趣味性、教育性、段落布局、适读性、文本生成一致性、上下文连贯性。其中在内容相似性上的评估使用BLEU-n($n=1,2$)，在趣味性、教育性、段落布局、适读性、文本生成一致性、上下文连贯性方面使用GPT-4来进行评估(Chang et al., 2024)。为确保GPT-4评估结果的准确性和可信度，我们从每个评估模型中随机抽取了50篇文本，总计200篇，邀请四位语言学硕士来进行背靠背人工评审。这一步骤旨在检验GPT-4的评分是否与人工评估人员的专业判断保持一致。通过这种方法，我们能够对GPT-4的评估效果进行严格的验证，并据此调整和优化我们的评估流程。完整的GPT-4评估提示在附录A.1中，考虑LLM生成文本的随机性，我们将每个类型的文本传输3次，并计算最终的平均结果，其中实验结果统计为百分制。

5.3 实验结果及分析

在多个自然语言处理任务中，例如事件抽取、对话生成和文本摘要，大型语言模型在只有少量样本（Few-shot）的情况下表现出色。在本研究中，将TNC模块分别与Zero-Shot和One-Shot学习方法相结合，进行了相关实验。如图表1是在BLEU-n下的评估结果，其中TNC表示使用文本样式定义（TSD）模块和主题类别定义（TCD）模块的组合。在Zero-Shot实验中，Gemini-1.0-pro模型在BLEU-1值在40.9%，显示出生成文本和原文本内容上中等程度的一致性。与GPT-3.5-turbo-16k的得分相近，在加入我们的基于主题的儿童新闻篇章结构框架（TNC-LLM）结果显示，这两个模型的分数有显著提升，尤其在GPT-3.5-turbo-16k的BLEU-1上达到了50.35%，这表明TNC框架对于提升模型性能是非常有效的。讯飞星火-v3.5模型即使没有TNC框架其BLEU-1也达到了44.38%，可能因其在生成儿童文本方面有优势。这四个模型在加入TNC后BLEU值都有所提升。在One-Shot实验中，One-Shot情况下的分数普遍高于Zero-Shot情况，在多个模型中应用了TNC框架的版本普遍比未应用的版本表现更好，其中讯飞星火-v3.5在加入TNC后，BLEU-n值达到最高值，再次表明TNC框架是一种有效的性能提升手段。也同时反映出不同的模型在处理儿童新闻文本方面有不同的内在能力。

为了验证GPT-4评估结果的准确性，我们采取了随机抽样的方法进行人工复审。在对200篇随机选取的评估结果进行人工审核后，我们发现大约有168篇的评估与人工判断相符，其余32篇的人工评审结果与GPT-4评审结果差值统计详情在附录B.2，这一结果表明GPT-4的评估具有较高的准确度。基于这一发现，我们决定将GPT-4的评估结果作为我们分析的重要参考依据。如图表2是使用GPT-4在六个维度上的评估结果，结果统计为等比百分制。如图9为各个模型的实验结果在趣味性等六个维度的雷达图。在Zero-Shot条件下，结果显示，TNC框架的引入在多数维度上提升了模型的表现。这一现象在结合GPT-3.5-turbo-16k模型时尤为明显，其中TNC+GPT-3.5-turbo-16k在趣味性、教育性和段落布局上的得分相较于未加入TNC框架的原始模型提升较为明显，分别提高了4.22、10.66和9.39。此外，TNC+讯飞星火-v3.5在趣味

性、教育性、段落布局、适读性上表现最为出色，再次表明该模型在生成儿童新闻文本方面具有优势。在One-Shot条件下，大多数模型在实验设置下都取得了更高的分数，尤其是在趣味性、教育性和段落布局方面。TNC+GPT-3.5-turbo-16k在多个维度上取得了最高分，其中趣味性得到显著的提升，TNC+讯飞星火-v3.5在教育性和段落布局上均取得最高分，这一结果表明TNC框架不仅能够提升模型的文本生成质量，还能够在儿童教育中发挥出更大的潜力。TNC框架通过引入特定的教育性内容，优化了文本结构，使得生成的文本更加符合教育目的，并提高了儿童的阅读兴趣和理解能力。在表2所展示的数据中，对于一致性和连贯性这两个指标，不同模型的表现并没有显著差异。然而，当我们将TNC模块注入系统后，我们发现这两个指标有轻微的提升。这表明TNC模块对于提高文本的一致性和连贯性具有积极作用。然而，在ERNIE-3.5模型中，我们的方法在一致性和连贯性上的表现出现了轻微的下降。经过深入分析，我们推测这一现象的原因可能是由于将两个模块组合的提示注入到ERNIE-3.5模型中，这可能对ERNIE-3.5模型的生成过程施加了不必要的限制和约束，从而影响了其性能。

模型名称	Zero-Shot		One-Shot	
	BLEU-1	BLEU-2	BLEU-1	BLEU-2
Gemini-1.0-pro	40.9	33.74	42.15	35.15
TNC+Gemini-1.0-pro	43.03	35.8	45.95	37.63
GPT-3.5-turbo-16k	41.19	33.89	43.98	35.92
TNC+GPT-3.5-turbo-16k	50.35	41.11	51.48	41.71
讯飞星火-v3.5	44.38	36.5	44.89	37.05
TNC+讯飞星火-v3.5	46.06	38.04	52.07	45.41
ERNIE-3.5	41.02	33.33	42.53	34.27
TNC+ERNIE-3.5	44.73	35.38	48.41	40.32

表 1. 生成文本在BLEU-n指标下的评估结果(%)

模型名称	趣味性	教育性	段落布局	适读性	一致性	连贯性
Gemini-1.0-pro	67.84	75.36	72.06	80.14	81.48	85.64
TNC+Gemini-1.0-pro	70.06	79.34	79.16	81.32	81.68	85.58
GPT-3.5-turbo-16k	65.82	70.32	70.88	79.52	81.26	85.49
TNC+GPT-3.5-turbo-16k	69.84	80.98	80.27	81.52	81.49	85.88
讯飞星火-v3.5	72.76	79.3	78.35	81.39	81.35	85.27
TNC+讯飞星火-v3.5	75.77	81.31	81.72	81.56	81.2	85.46
ERNIE-3.5	63.74	72.01	69.83	79.26	81.66	85.43
TNC+ERNIE-3.5	67.34	73.82	77.22	81.49	81.31	85.35
Gemini-1.0-pro	68.14	79.43	78.85	81.3	81.31	85.52
TNC+Gemini-1.0-pro	71.74	80.45	81.94	82.09	81.34	85.39
GPT-3.5-turbo-16k	67.33	75.05	77.6	80.08	82.05	85.47
TNC+GPT-3.5-turbo-16k	75.56	82.91	82.04	82.54	82.79	86.36
讯飞星火-v3.5	75.04	80.22	81.33	81.17	81.65	85.64
TNC+讯飞星火-v3.5	75.53	83.28	82.78	82.27	82.77	85.56
ERNIE-3.5	65.67	73.21	78.23	81.13	81.94	85.91
TNC+ERNIE-3.5	68.78	75.29	79.46	81.87	81.09	85.35

表 2. GPT-4在六个维度上对生成文本的评估结果

综合来看，TNC框架的引入在提高大型语言模型在儿童新闻文本生成任务上的表现方面起到了关键作用。无论是在零样本还是一样本条件下，引入TNC框架的模型普遍优于未加入该框架的模型。这可能是由于TNC框架通过提供更细粒度的控制和优化，使得模型能够更好地理解和适应儿童新闻文本生成任务的需求，从而在生成文本的内容一致、趣味性、教育性、段落布局、适读性和连贯性等多个维度上实现了显著提升。

5.4 消融研究

为了探究不同模块对儿童新闻生成的影响，我们进行了消融实验，分别评估了文本样式定义(TSD)模块和主题类别定义(TCD)模块的独立作用对儿童新闻文本生成的影响。消融实验的评估结果如表3和4。根据实验结果以及附录B.1中图9，可以看出在Zero-Shot情况下，我们观察到单独使用TSD模块与单独使用TCD模块在结合Gemini-1.0-pro、GPT-3.5-turbo-16k和ERNIE-3.5模型时，在多数评估指标表现相似。但对于讯飞星火-v3.5模型，单独

使用TSD模块在BLEU评分上取得了较高的分数，表明其在词汇重叠度和流畅性方面具有优势。在趣味性、教育性和段落布局这些维度上，单独使用TCD模块的表现更佳，这可能是由于TCD模块在内容生成方面更加注重创意和教育价值的融入。在One-Shot情况下，独立使用TCD模块时，ERNIE-3.5和讯飞星火-v3.5模型在BLEU值、趣味性、教育性上的表现显著优于单独使用TSD模块。这说明TCD模块借助于示例，能更有效地增强文本的适用性和教育价值。

模型名称	Zero-Shot		One-Shot	
	BLEU-1	BLEU-2	BLEU-1	BLEU-2
TCD+Gemini-1.0-pro	41.59	32.41	42.78	35.47
TSD+Gemini-1.0-pro	42.73	35.72	42.64	35.42
TNC+Gemini-1.0-pro	43.03	35.8	45.95	37.63
TCD+GPT-3.5-turbo-16k	43.73	35.32	44.17	35.97
TSD+GPT-3.5-turbo-16k	43	35.37	43.44	35.56
TNC+GPT-3.5-turbo-16k	50.35	41.11	51.48	41.71
TCD+ERNIE-3.5	42.97	34.96	45.59	36.37
TSD+ERNIE-3.5	42.07	34.57	42.45	34.67
TNC+ERNIE-3.5	44.73	35.38	48.41	40.32
TCD+讯飞星火-v3.5	43.19	35.4	46.29	38.57
TSD+讯飞星火-v3.5	45.47	37.38	45.17	37.1
TNC+讯飞星火-v3.5	46.06	38.04	52.07	45.41

表 3. 消融实验生成文本在BLEU-n指标下的评估结果

模型名称	趣味性	教育性	段落布局	适读性	一致性	连贯性
TCD+Gemini-1.0-pro	67.79	75.36	74.69	81.47	81.48	85.83
TSD+Gemini-1.0-pro	68.62	80.52	78.32	81.27	81.35	85.69
TNC+Gemini-1.0-pro	70.06	79.34	79.16	81.32	81.68	85.58
TCD+GPT-3.5-turbo-16k	68.62	80.52	78.32	81.27	81.35	85.69
TSD+GPT-3.5-turbo-16k	67.79	75.36	74.69	81.47	81.48	85.83
TNC+GPT-3.5-turbo-16k	69.84	80.98	80.27	81.52	81.49	85.88
TCD+ERNIE-3.5	65.28	73.3	75.1	81.34	81.84	85.83
TSD+ERNIE-3.5	64.42	73.27	73.47	81.26	81.61	85.85
TNC+ERNIE-3.5	67.34	73.82	77.22	81.49	81.31	85.35
TCD+讯飞星火-v3.5	75.35	81.37	80.59	81.37	81.73	85.52
TSD+讯飞星火-v3.5	72.54	80.33	78.77	81.52	81.37	85.67
TNC+讯飞星火-v3.5	75.77	81.31	81.72	81.56	81.2	85.46
<hr/>						
TCD+Gemini-1.0-pro	69.65	79.97	79.52	81.45	81.08	85.39
TSD+Gemini-1.0-pro	68.69	79.38	78.35	80.98	81.01	85.16
TNC+Gemini-1.0-pro	71.74	80.45	81.94	82.09	81.34	85.39
TCD+GPT-3.5-turbo-16k	69.6	81.55	79.59	81.51	81.5	85.43
TSD+GPT-3.5-turbo-16k	67.73	75.24	76.83	81.43	81.26	85.48
TNC+GPT-3.5-turbo-16k	75.56	82.91	82.04	82.54	82.79	86.36
TCD+ERNIE-3.5	67.83	74.36	78.91	81.25	81.41	85.67
TSD+ERNIE-3.5	65.63	73.34	78.31	81.37	81.5	85.38
TNC+ERNIE-3.5	68.78	75.29	79.46	81.87	81.09	85.35
TCD+讯飞星火-v3.5	75.41	82.28	82.01	81.33	81.54	85.63
TSD+讯飞星火-v3.5	74.86	81.18	81.8	81.11	81.29	85.04
TNC+讯飞星火-v3.5	75.53	83.28	82.78	82.27	82.77	85.56

表 4. 消融实验在GPT-4在六个维度上对生成文本的评估结果

TCD和TSD模块各有所长，它们结合不同模型在实验条件下展现出各自的优势。TCD模块在内容创意和教育价值方面表现突出，而TSD模块在提升文本的流畅性和BLEU评分方面更具优势。综合实验的结果，将TCD模块与TSD模块结合使用，能够显著提升儿童新闻生成框架的整体性能。这种组合不仅在文本的流畅性和一致性方面表现出色，同时也在趣味性、教育性、段落布局和适读性等关键维度上达到了更高的标准。通过这种协同作用，我们能够生成更符合儿童阅读习惯和认知水平的新闻内容，从而更好地满足青少年儿童的新闻阅读需求。因此，TCD与TSD的结合不仅是一种有效的策略，更是实现高质量儿童新闻生成的关键。

5.5 案例分析

如图8右侧文本是将我们的TSD模块和TCD模块注入到讯飞星火v-3.5根据左侧的成年人新

闻内容生成的儿童版新闻文本。观察生成的儿童新闻，我们可以发现其排版布局整洁有序，段落短小精悍，且广泛使用简短的句子。相比之下，成年人版新闻在语言风格上更为严肃和直接，频繁使用如“垃圾塑料”、“生态灾难”和“可怕的生态污染”等词汇，以此突出问题的紧迫性。与此相对照，儿童新闻倾向于运用生动形象的比喻，例如“像小山一样多的塑料垃圾”和“像贪吃的小怪兽”，为了更好地抓住孩子们的注意力，文本中巧妙地融入了互动性强的语言元素，比如使用拟声词“滴答滴答”和“轰隆轰隆”等，以吸引儿童的注意力。成年人版新闻在文中高亮部分提供了详尽的数据和事实依据，如“面积超过160万平方公里”、“400万吨塑料垃圾”和“1.8万亿块塑料垃圾”，而儿童新闻则通过类比——如“足足有70万头大象那么重”，来帮助儿童把握问题的规模。在文章的左侧画线文字部分，成年人版深入探讨了问题的严重性，这部分内容在制作儿童版新闻时被刻意省略。在文章结尾（浅色部分），成年人版客观地阐述了问题的严重性，目的在于提升公众对环境问题的认识，同时隐舍地对政策制定者和行业实践提出批评。相较之下，儿童版新闻更注重培养环保意识，通过一系列生动的比喻和形象化的描述，激发孩子们对保护地球的责任感和行动力。

成年人版新闻：

太平洋深处的“第八大陆”，到底是什么来头？卫星图像显示，在北美洲西海岸至夏威夷群岛之间的北太平洋上，存在着一个地图上从未有过标注的巨型“大陆”，面积超过160万平方公里。但奇怪的是，这个“第八大陆”上并没有人居住，世界各国也无意争夺占领，反而唯恐避之而不及……这背后究竟隐藏着怎样的秘密？众所周知，地球七大洲、四大洋，三分陆地、七分海洋的格局以稳定存在的上亿年之久，而这块“第八大陆”就位于美国西海岸和夏威夷之间，面积相当于两个得克萨斯州，是太平洋上一片由400万吨塑料垃圾组成的漩涡，完全由海洋垃圾聚集而成！

受到洋流的影响，整个太平洋内的垃圾塑料都会不断向这块“大陆”聚集，每天都在吸收着周围的海洋垃圾，因此面积也一天比一天变大。此外，海上风暴可能会进一步导致原本块状的塑料垃圾变得更加分散，从而产生更严重的生态灾难。一项调查显示，所有人类的生活用品都可以在这块“大陆”上找到，也让我们在内心意识到如今的海洋污染也已经达到非常可怕的地步。

在卫星传过来的图像中就能够看得出来，这一片垃圾大陆像是北太平洋上的一个移动板块，根据粗略估计，这片“世外”垃圾源由1.8万亿块、400万吨的塑料垃圾组成，虽然这些垃圾并没有堆积在任何一个国家的国土上，但最终造成的后果却需要全人类共同来承担，人类每年都会生产超过3.2亿吨的塑料垃圾，但其中仅有一小部分能够得到回收，乃至焚烧、填埋，大部分都进入了海洋。随着大家生活节奏越来越快，我们所居住的地球竟然成为一个大型垃圾站，海、陆、空的全方位立体污染不断加剧。可悲的是，面对第八大陆，各个国家唯恐避之不及，难道我们真的应该用这样的态度来对待环境日益恶化的地球吗？如何处治“第八大陆”？全球绝大部分科学家都认为，“防”大于“治”。此前，“第八大陆”发现者查尔斯·摩尔曾悲观地表示，“就算耗尽所有国家的财力，并杀死大量野生生物，也不一定能清除北太平洋环流系统的所有塑料垃圾。”

海洋是世界上最接近自然的地方，来自海洋的水如同地球的血液，值得每个地球人去关注并保护它。

生成的儿童新闻：

太平洋的哭泣，一片哀嚎的垃圾群岛

在人迹罕至的太平洋深处，隐藏着一个未曾被标记的“新大陆”。它庞大无比，足以吞没两个得克萨斯州，然而奇怪的是，这里没有小动物嬉戏，也没有人类探险者的脚步，为什么会这样呢？

1.神秘的“第八大陆”

在太平洋的怀抱里，有一个神秘的“第八大陆”。它不是真正的陆地，而是由像小山一样多的塑料垃圾组成，他足足有70万头大象那么重。

这个“大陆”能覆盖两个得克萨斯州那么大的地方，但是，这片“大陆”上却没有人唱歌跳舞，因为这里只有寂静和塑料，世界各国的船只也不会来这里探险，而是像躲开怪物一样绕着它走。

2.塑料垃圾的危害

洋流就像是海洋的呼吸，不断地“呼啦呼啦”地把太平洋里的塑料垃圾吹向这个“大陆”。这个“大陆”就像贪吃的小怪兽一样，

一口一口地吃掉周围的海洋垃圾，让自己变得越来越庞大。当海上风暴袭来时，它们就像一群愤怒的巨人，“轰隆轰隆”地把原本紧紧抱在一起的塑料垃圾撕成碎片，给海洋里的小鱼小虾带来了灾难。

3.共同保护我们的地球

全球的科学家们都在呼吁，预防比治疗更重要，我们需要从源头上减少塑料垃圾的产生，就像关上水龙头一样，不让塑料垃圾再“滴答滴答”流入海洋。

此外，我们还可以参加海滩清洁活动，就像给海滩做一次大扫除，支持环保组织，为保护海洋环境出一份力。地球是我们共同的家园，海洋的每一滴水都如同地球体内流淌的生命之血，一旦这血液受到污染，地球也将陷入疾病之中。

保护海洋，便是维护我们自己的家园。

让我们手拉手，一起守护我们美丽的地球家园！每当我们拒绝使用一个塑料袋，就像是给地球妈妈一个温暖的拥抱。

图 8. 成年人版新闻与讯飞星火v-3.5模型生成的儿童新闻示例

6 结论与未来工作

在本研究中，我们提出了一种创新性的基于主题的框架儿童新闻篇章结构框架（TNC-LLM），这是大型语言模型在儿童新闻领域的首次系统性应用。本框架的核心由两个关键模块构成：文本样式定义（TSD）模块和主题类别定义（TCD）模块。经过实验我们发现TCD模块在内容创意和教育价值方面表现突出，而TSD模块在提升文本的流畅性和BLEU评分方面更具优势。综合实验的结果，TNC框架通过提供更细粒度的控制和优化，使得模型能够更好地理解和适应儿童新闻生成任务，从而在生成文本的趣味性等多个维度上实现了显著提升。同时我们发现讯飞星火-v3.5在趣味性、教育性表现最为出色，表明该模型在生成面向儿童教育文本方面具有优势。尽管我们的框架已经取得了相对成功，但我们认识到，未来的工作仍有进一步完善和提升的空间。我们计划在未来的研究中更加关注知识补充和整合方面的工作。由于儿童新闻往往需要综合多篇成年人新闻的信息，如何有效地提炼和融合这些知识以生成既准确又富有教育意义的儿童新闻，将成为我们未来研究的重点。我们期望通过这些努力，能够进一步提升儿童新闻的教育价值，使其成为儿童学习和成长过程中的有益资源。

参考文献

- Alon-Tirosh. 2014. If i was making the news”: What do children want from news. *Participations*, 11(1):108–129.
- Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. 2024. A survey on evaluation of large language models. *ACM Trans. Intell. Syst. Technol.*, 15(3), mar.
- Banghao Chen, Zhaofeng Zhang, Nicolas Langrené, and Shengxin Zhu. 2023. Unleashing the potential of prompt engineering in large language models: a comprehensive review.
- Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. 2020. Plug and play language models: A simple approach to controlled text generation.
- Angela Fan, Mike Lewis, and Yann Dauphin. 2018. Hierarchical neural story generation.
- Jean Kaddour, Joshua Harris, Maximilian Mozes, Herbie Bradley, Roberta Raileanu, and Robert McHardy. 2023. Challenges and applications of large language models.
- Michael Lebowitz. 1985. Story-telling as planning and learning. *Poetics*, 14:483–502.
- Juncen Li, Robin Jia, He He, and Percy Liang. 2018. Delete, retrieve, generate: A simple approach to sentiment and style transfer.
- Danyang Liu, Juntao Li, Meng-Hsuan Yu, Ziming Huang, Gongshen Liu, Dongyan Zhao, and Rui Yan. 2020. A character-centric neural model for automated story generation. In *AAAI Conference on Artificial Intelligence*.
- Yao Lu, Max Bartolo, Alastair Moore, Sebastian Riedel, and Pontus Stenetorp. 2022. Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity.
- N. Newman. 2024. Journalism, media, and technology trends and predictions 2024. Retrieved from <https://policycommons.net/artifacts/11300794/journalism-media-and-technology-trends-and-predictions-2024/12186166/>. Accessed on 17 Apr 2024, CID: 20.500.12592/fbg7g5r.
- Kyeongman Park, Nakyeong Yang, and Kyomin Jung. 2023. Longstory: Coherent, complete and length controlled long story generation.
- Pelletier. 2004. Action, consciousness and theory of mind: Children’s ability to coordinate story characters’ actions and thoughts. *Early Education Development*, 15(1):5–22.
- Peter. 1996. Working memory, intelligence and reading ability in children. *Personality and Individual Differences*, 21(6):1007–1020.
- Jing Qian, Li Dong, Yelong Shen, Furu Wei, and Weizhu Chen. 2022. Controllable natural language generation with contrastive prefixes.
- Yao Ting Sung, Ju Ling Chen, Ji Her Cha, Hou Chiang Tseng, Tao Hsing Chang, and Kuo En Chang. 2015. Constructing and validating readability models: the method of integrating multilevel linguistic features with machine learning. *Behavior Research Methods*.
- Zhen Tao, Dinghao Xi, Zhiyu Li, Liumin Tang, and Wei Xu. 2024. Cat-llm: Prompting large language models with text style definition for chinese article-style transfer. *ArXiv*, abs/2401.05707.
- Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. 2022. Emergent abilities of large language models. *Transactions on Machine Learning Research*. Survey Certification.
- Jingjing Xu, Xu Sun, Qi Zeng, Xiaodong Zhang, Xuancheng Ren, Houfeng Wang, and Wenjie Li. 2018. Unpaired sentiment-to-sentiment translation: A cycled reinforcement learning approach. In Iryna Gurevych and Yusuke Miyao, editors, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 979–988, Melbourne, Australia, July. Association for Computational Linguistics.

- Peng Xu, Mostofa Patwary, Mohammad Shoeybi, Raul Puri, Pascale Fung, Anima Anandkumar, and Bryan Catanzaro. 2020. MEGATRON-CNTRL: Controllable story generation with external knowledge using large-scale language models. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2831–2845, Online, November. Association for Computational Linguistics.
- Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A survey of large language models.
- Yang Zhong, Chao Jiang, Wei Xu, and Junyi Jessy Li. 2020. Discourse level factors for sentence deletion in text simplification.
- 吴思远, 于东, and 江新. 2020. 汉语文本可读性特征体系构建和效度验证. *世界汉语教学*, 34(1):81–97, 1.
- 孙汉银. 1992. 中文易懂性公式. 硕士学位论文, 北京师范大学.
- 张卫国. 2006. 阅读:覆盖率,识读率和字词比. *语言文字应用*, (3):8.
- 皮亚杰. 1981. 儿童心理学. 商务印书馆, 北京.
- 饶芳芳. 2023. 具身认知视角下儿童科普读物设计研究——以《小小牛顿幼儿馆steam》系列为例. *上海包装*, (11):137–139.

A

A.1 GPT-4评估生成文本的提示模版

你是个儿童新闻文本编辑专家，请客观的将模型生成文本的质量参照儿童新闻语料，从以下四个维度：趣味性、教育性、段落布局、适读性来进行参考评估，文本风格越接近原文本分数越高，从1（最差）到10（最好）进行评分。其中每个维度的具体信息如下：

趣味性

故事元素：文本是否包含吸引人的故事，如有趣的角色、激动人心的冒险或富有教育意义的情节。

修辞手法：文本是否使用了比喻、拟人等修辞手法来增强描述的生动性和形象性。

想象力和创造力：文本是否鼓励儿童进行想象，提供创造性思考的机会，或者激发他们的创造性表达。

教育性

知识传授：文本是否介绍了新的知识点，如科学发现、历史事件或文化习俗。

生活技能：文本是否提供了实用的信息或建议，帮助儿童学习生活中的技能，如健康饮食、安全意识等。

道德价值：文本是否讨论了道德价值观念，如诚实、勇敢、同情心等，并以儿童易于理解的方式呈现。

段落布局

结构清晰：是否一篇文章分三个小标题板块进行叙述，每个段落是否有明确的开始和结束，以及是否围绕单一主题或想法组织内容。

视觉呈现：段落是否通过标题、子标题等视觉元素来增强可读性和吸引力。

适读性

语言难度：文本使用的词汇和句子结构是否适合12岁（小学六年级）以下儿童的年龄和阅读水平。

信息量：文本是否提供了适量的信息，避免过于简单或过于复杂，确保儿童能够理解和吸收。

文本生成一致性

主题一致性：文本是否在整篇文章中保持对主题的一致关注，避免偏离主题。

风格一致性：文本的语气、风格和语言特点是否全文保持一致，没有突兀的变化。

上下文连贯性

内部逻辑：文本中的事件、描述和论点是否逻辑上连贯，形成有意义的整体。

连接词使用：是否恰当使用连接词和过渡短语来连接句子和段落，确保文本的流畅性。

上下文关联：文本中的信息是否与上下文相关联，确保读者能够理解每个部分在整个故事或讨论中的位置和作用。

原文本={儿童新闻语料}

需要评估的文本={模型生成的文本}

A.2 One-Shot下的文版样式定义模块s和主题类别定义模块提示模版

示例example:

成年人版新闻文本: XXX

儿童版新闻文本: XXX

成年人版新闻文本 D_{i2} : “央视网消息：北京时间今天（6月15日）13时30分，我国在太原卫星发射中心使用长征二号丁运载火箭，成功将吉林一号高分06A星等41颗卫星发射升空，卫星顺利进入预定轨道，发射任务获得圆满成功。该批卫星主要用于提供商业遥感服务及相关技术验证。此次任务是长征系列运载火箭的第476次飞行。41位“乘客”“一箭多星”再创纪录。“1箭41星”再次刷新了中国航天“一箭多星”纪录，41颗卫星就像是41位“乘客”，发射前它们要被合理地安排在火箭的整流罩里；发射后，它们是分组“下车”、依次分离，这就给发射任务带来了不小的挑战。太原卫星发射中心的科技人员从转运、吊装、卫星安装、星罩合体等各个

环节，精心准备、精心组织、精心实施，确保了发射任务圆满成功……”

提示：你是一位资深的儿童文学专家，现在有一个根据成年人版新闻文本 D_{i2} 内容来生成一篇儿童版的新闻，请参考以下两个特征模块：文本样式定义模块s和主题类别定义模块c。请参考以上示例example。（注：确保内容真实准确，避免误导儿童，选择儿童感兴趣的话题和事件。）

文版样式定义模块s:

文本难度：属于12岁（小学六年级）以下适读。

字难度：笔画数 $C_{<16}$ ， $C_{>8}$ ，汉字熟悉度 $f_{i,len}$ ， $f_{c,len}$

词语层面：词长 f_l ，字词比 f_{cw} ，词性 f_{ps} ，词语熟悉度 $f(T, H)$

段落句子布局：句数 f_{sl} ，句子长度 f_{len}, f_{lw} ，长短句分布 f_{ls} ，单复句 f_{simple} ， $f_{complex}$ ，段落长度 f_{ac} ， f_{aw}

儿童新闻整体板块：标题：XX、导语：XX、正文：小标题1、小标题2、小标题3。

主题类别定义模块c:

该文本可以写成XX类别的新闻，该类别的儿童新闻主题特征如下： c_i

A.3 Zero-Shot下的文版样式定义模块s和主题类别定义模块提示模版

成年人版新闻文本 D_{i2} ：“央视网消息：北京时间今天（6月15日）13时30分，我国在太原卫星发射中心使用长征二号丁运载火箭，成功将吉林一号高分06A星等41颗卫星发射升空，卫星顺利进入预定轨道，发射任务获得圆满成功。该批卫星主要用于提供商业遥感服务及相关技术验证。此次任务是长征系列运载火箭的第476次飞行。41位“乘客”“一箭多星”再创纪录。“1箭41星”再次刷新了中国航天“一箭多星”纪录，41颗卫星就像是41位“乘客”，发射前它们要被合理地安排在火箭的整流罩里；发射后，它们是分组“下车”、依次分离，这就给发射任务带来了不小的挑战。太原卫星发射中心的科技人员从转运、吊装、卫星安装、星罩合体等各个环节，精心准备、精心组织、精心实施，确保了发射任务圆满成功……”

提示：你是一位资深的儿童文学专家，现在有一个根据成年人版新闻文本 D_{i2} 内容来生成一篇儿童版的新闻，请参考以下两个特征模块：文本样式定义模块s和主题类别定义模块c。（注：确保内容真实准确，避免误导儿童，选择儿童感兴趣的话题和事件。）

文版样式定义模块s:

文本难度：属于12岁（小学六年级）以下适读。

字难度：笔画数 $C_{<16}$ ， $C_{>8}$ ，汉字熟悉度 $f_{i,len}$ ， $f_{c,len}$

词语层面：词长 f_l ，字词比 f_{cw} ，词性 f_{ps} ，词语熟悉度 $f(T, H)$

段落句子布局：句数 f_{sl} ，句子长度 f_{len}, f_{lw} ，长短句分布 f_{ls} ，单复句 f_{simple} ， $f_{complex}$ ，段落长度 f_{ac} ， f_{aw}

儿童新闻整体板块：标题：XX、导语：XX、正文：小标题1、小标题2、小标题3。

主题类别定义模块c:

该文本可以写成XX类别的新闻，该类别的儿童新闻主题特征如下： c_i

B

B.1 多维度实验评估结果雷达图

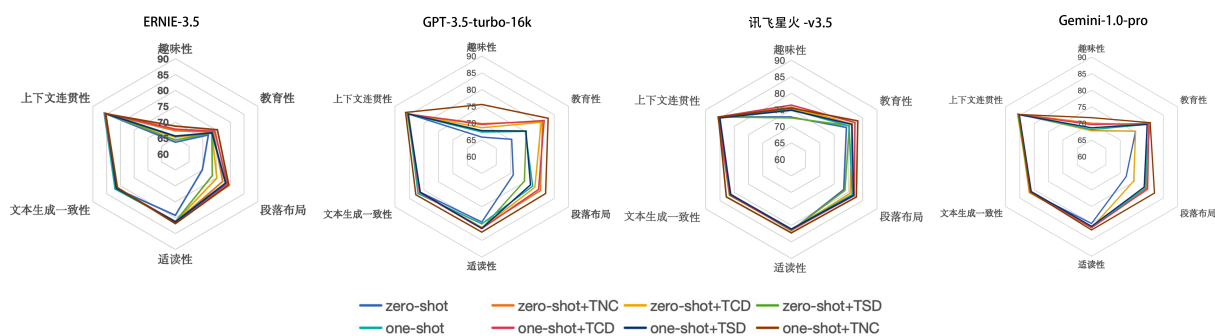


图 9. 多维度实验评估结果雷达图

B.2 32篇人工评估与GPT-4评估不符统计详情

段落布局	趣味性	教育性	可读性
-2	+1	+1	+1
-1	+1	+1	+1
-1	+1	+2	-1
-2	-1	+1	-1
-2	+1	-1	-1
+1	-1	-1	+1
-2	+2	-	-2
-2	+1	-	-
+2	+2	-	-
-	+2	-	-

表 5. 人工评估与GPT-4评估对比 (+2代表人工评估比GPT-4评估在该维度多2分)

B.3 TSD实验结果

平均笔画数 (argv_bh)	7.42	字词比 (charword_token)	1.57
最大笔画数(maxi_bh)	19.38	词长为1的字数(len1)	212.07
最小笔画数(mini_bh)	1.00	词长为2的字数(len2)	169.11
笔画数小于8的字数(low8_bh)	423.03	词长为3的字数(len3)	20.32
笔画数大于16的字数(hi16_bh)	9.50	词长为4的字数(len4)	6.75
平均字数(var_docfreq_word)	787.65	词长大于4的字数(len4_)	1.05
不在字频表高频500的字数(oov_corf500_char)	200.64	否定词数(neg)	14.01
不在字频表高频1000的字数(oov_corf1000_char)	99.18	成语数(idiom)	2.18
不在字频表高频1500的字数(oov_corf1500_char)	53.30	实词数(cont)	318.72
不在字频表高频2000的字数(oov_corf2000_char)	30.03	虚词数(func)	95.79
不在常见字表的字数(oov_commonchar)	6.58	形容词 (a)	23.68
汉字难度等级均值(argv_dl_tokenchar_newhsk)	2.53	其他名词修饰语 (b)	1.25
出现在常见字表1000的字数(common_10)	405.26	连词 (c)	5.46
出现在常见字表2000的字数(common_20)	574.78	副词 (d)	37.16
出现在常见字表3500的字数(common_35)	637.32	感叹词 (e)	0.87
不在词频表高频500的词数(oov_corf500_word)	207.18	成语 (i)	4.33
不在词频表高频1000的词数(oov_corf1000_word)	175.71	数字 (m)	20.07
不在词频表高频1500的词数(oov_corf1500_word)	154.33	普通名词 (n)	93.13
不在词频表高频2000的词数(oov_corf2000_word)	140.83	指示名词 (nd)	8.29
疑问代词数 (PronWhat)	2.17	人名 (nh)	1.42
并列连词 (binlieConj)	3.13	地理名词 (ns)	9.28
递进连词 (dijinConj)	0.64	时间名词 (nt)	5.96
假设连词 (jissheConj)	2.61	其他专有名词 (nz)	0.29
让步连词 (rangbuConj)	1.51	拟声词 (o)	0.93
顺承连词 (shunchengConj)	0.16	介词 (p)	12.43
条件连词 (tiaojianConj)	1.01	数量词 (q)	13.26
因果连词 (yinguoConj)	1.46	代词 (r)	24.28
转折连词 (zhuanzheConj)	1.64	助动词 (u)	38.93
连词总数 (ConjAll)	12.17	动词 (v)	110.32
动词短语数 (VP_num)	62.39	标点符号 (wp)	117.78
名词短语数 (NP_num)	103.01	描写性词 (z)	1.72
副词短语数 (ADVP_num)	49.64	代词数 (PronAll)	10.39
介词短语数 (PP_num)	10.37	指示代词数 (PronDemo)	4.51
形容词短语数 (ADJP_num)	9.68	人称代词数 (PronPers)	3.71
句子数 (sent)	45.24	第一人称代词数 (PronPers_Firs)	0.88
单句数 (easy_sent)	13.07	第二人称代词数 (PronPers_Secd)	0.63
复句数 (complex_sent)	32.17	第三人称代词数 (PronPers_Thir)	2.21
平均句子长度(字) (mean_lenz)	14.38	段落数 (pg_num)	34.99
最大句子长度(字) (max_lenz)	35.89	平均段落长度(字) (mean_lenz_pg)	18.58
最小句子长度(字) (min_lenz)	2.46	平均段落长度(词) (mean_lenw_pg)	15.12
平均句子长度(词) (mean_lenw)	11.94		
最大句子长度(词) (max_lenw)	29.17		
最小句子长度(词) (min_lenw)	2.54		

表 6. TSD模块的统计结果

C

C.1 主题类别定义(TCD)板块

榜样类儿童新闻:

板块原则:每个板块的内容要完整,每个板块的字数约在200左右,每个板块有小标题。

方案一:若原新闻文本给定的是一个多人榜样合集,则选择三个榜样,每个板块写一个榜样。

方案二:

板块一:首先介绍人物的背景信息和面临的挑战或问题

板块二:接着描述人物是如何通过创新或特殊的方法解决问题的。

板块三:最后展示人物的努力和创新带来的积极成果和影响。

榜样类儿童新闻的撰写特点:

注重人物故事:通过讲述个人的故事和经历,使新闻更加生动和吸引人。

强调正面价值:通过展示主人公面对困难的勇气、持之以恒的精神传递正面的价值观和榜样的力量。

启示与反思:通过榜样的故事,引发读者的启示和反思

寓教于乐:在吸引儿童注意力的同时,传递积极的价值观和行为模范,鼓励儿童学习和模仿。

可读性强:通过生动的故事情节和具体的人物形象,要善于使用一些增强语言生动效果的修辞手法,比如比喻,使得新闻内容富有趣味性和吸引力,容易被儿童接受和理解。

环境保护类儿童新闻:

板块原则

每个板块的内容要完整,每个板块的字数约在200左右,每个板块有小标题

板块一:首先介绍背景和问题的存在,让读者了解目前面临的挑战或问题的严重性。

板块二:接着,深入探讨问题带来的具体影响或潜在危害。这一部分旨在增强读者对问题紧迫性的认识。

板块三:最后,提出解决问题的方法、已经采取的措施或呼吁采取的行动。这部分旨在激励读者或社会采取行动,对问题进行缓解或解决。

环境类儿童新闻的撰写特点:

互动性和参与感:通过提问或呼吁读者思考,激发儿童对环境问题的兴趣和参与感。如通过展示积极的改变激发他们对保护环境的兴趣和责任感。

科普性:在故事化的叙述中融入科学知识和环保理念,既娱乐又教育,帮助儿童读者在轻松愉快的阅读过程中学习到有关环境保护的知识。

生动的比喻和描绘:通过将复杂的环境问题比作人类熟悉的情景(如“地球之肺”生病),使儿童更容易理解和感受到问题的紧迫性。

简化的数据和事实:使用容易理解的数据(如“每过8秒,就有足球场大小的森林消失”)来说明问题,避免了复杂和抽象的统计数据,使信息更加亲近儿童的认知水平。

经济类儿童新闻:

板块原则:

每个板块的内容要完整,每个板块的字数约在200左右,每个板块有小标题

板块一:以生活中熟悉的事物或现象作为引入,或通过讲述一个具体的例子来吸引儿童的兴趣激发儿童的兴趣。

板块二:在吸引了读者的注意力之后,文本转而介绍了背景信息和当前的问题,这样做能帮助儿童理解实验的背景和目的。

板块三:探讨解决方案,引发了关于其好处或坏处的讨论。这一部分旨在激发儿童对结果的好奇心,并鼓励他们思考和讨论。

经济类儿童新闻的撰写特点:

语言简洁生动: 使用易于理解和生动的语言来解释经济概念, 避免了复杂的术语和概念, 使儿童能够更容易地理解。

情境化教学: 通过创造一个具体的情境或故事来介绍经济概念, 使抽象的经济理论变得具体化, 更容易吸引儿童的注意力。

鼓励思考和讨论: 提出了问题和可能的后果, 鼓励儿童进行思考和讨论, 培养他们的批判性思维能力。

关联现实生活: 通过与儿童的现实生活相关联的例子来解释经济概念, 或介绍具体的行动和政策, 强调理论知识在实际生活中的应用, 使得内容更加贴近儿童的生活经验, 增加了学习的实用价值。

科学类儿童新闻:

板块原则

每个板块的内容要完整, 每个板块的字数约在200左右, 每个板块有小标题

板块一: 以引人入胜的方式提出一些日常生活中的问题或现象, 激发儿童的好奇心。

板块二: 介绍针对提出的问题的科学实验过程和结果 / 科技或方法 / 科技产品的功能和作用

板块三: 给出了实验的结论, 介绍这些发现或发明的实际应用, 这不仅满足了儿童对问题答案的好奇心, 还向他们展示了科学研究对日常生活的影响。

科技类儿童新闻的撰写特点:

科学与价值观教育相结合: 不仅传递科学知识, 还通过探索的故事传达了勇于探索未知、追求知识的价值观

趣味性: 文章选题有趣, 语言风格幽默, 使得复杂的科学问题变得易于理解和亲近, 适合儿童的阅读习惯。

教育性: 通过介绍真实的科学研究和实验, 文章传达了科学探索的方法和精神, 鼓励儿童保持好奇心, 培养他们的科学思维

实用性: 文章通过介绍科学研究的实际应用, 如解决日常生活中的小问题, 让儿童了解科学知识的实用价值, 增强他们学习科学的动力。

社会类儿童新闻:

板块原则

每个板块的内容要完整, 每个板块的字数约在200左右, 每个板块有小标题

板块一: 通过生动的描述引入事件的背景和情境,

板块二: 介绍事件中的关键角色和行动, 详细叙述事件的过程和影响。

板块三: 揭示故事背后的深层意义, 强调正面价值观和教育意义。

社会类儿童新闻的撰写特点:

生动形象: 通过具体的场景描写和比喻, 让儿童能够直观感受到故事情境, 增强阅读的吸引力。

情感共鸣: 通过故事化的方式, 让儿童能够与新闻中的人物或事件产生情感上的共鸣, 从而激发他们的同情心和正义感。

启发式教育: 在报道中穿插启发性问题或思考, 引导儿童思考事件的原因、结果以及可能的解决方案, 培养他们的批判性思维和问题解决能力。

行动倡议: 提供简单、实际的行动建议或参与方式, 让儿童明白他们也能为社会问题做出贡献, 增强他们的社会参与感和责任感。

文化多样性: 展现不同文化和社会背景下的事件和人物, 增进儿童对多元文化的认识和尊重, 培养他们的全球视野和包容心态。

安全意识: 在报道涉及风险或危机的事件时, 强调安全措施和预防方法, 教育儿童如何在类似情况下保护自己 and 他人。

体育类儿童新闻:

板块原则

每个板块的内容要完整，每个板块的字数约在200左右，每个板块有小标题

板块一：以一个独特、有趣的场景或事件作为开头，介绍事件的背景、地点、时间和一些吸引人的信息，为读者建立一个场景。

板块二：深入讲述事件的具体内容，包括人物、地点、特殊情境等，使读者能够更加详细地了解事件。

板块三：在故事的结尾，传达一个正面的信息或价值观，让读者从故事中获得启发。

体育类儿童新闻的撰写特点:

道德和价值观的强调：通过故事传递积极的价值观，如勤奋、公平竞争、团队精神、坚持不懈、社区参与、多元文化的重要性等，鼓励儿童积极参与体育活动，培养团队精神和坚持不懈的精神

情感丰富有趣：通过描述人物、情境和细节，激发读者的情感共鸣，增加故事的吸引力。

互动性：鼓励读者参与和思考，如通过提问或建议读者想象自己在场景中的行为。

易于理解的规则解释：对于体育项目的规则和术语，用简单易懂的语言进行解释，帮助儿童更好地理解比赛，增加他们的参与感。

健康和安全意识的普及：在报道中强调运动的安全意识和健康习惯，教育儿童如何在参与体育活动时保护自己，避免受伤。

文化类儿童新闻:

板块原则

每个板块的内容要完整，每个板块的字数约在200左右，每个板块有小标题

板块一：通过一个引人入胜的故事介绍文化背景或历史起源。这部分可以通过一个具体、生动的事例或者角色引入主题。

板块二：详细阐述文章的主题内容，通过具体的例子或故事来展开讨论。

板块三：最后一部分对文章进行总结，并提供更广泛的视角或思考，有时也可以提出问题，引导读者进行进一步的思考。

文化类儿童新闻的撰写特点:

故事化的叙述：使用简洁、生动的语言描述复杂的概念，通过讲故事的方式介绍文化现象或历史背景，增加文章的吸引力和记忆点。

历史文化知识的融入：通过故事介绍相关的历史文化知识，让儿童在阅读故事的同时增长知识。

鼓励探索和思考：通过展示文化现象的变迁和多样性，鼓励儿童读者思考文化的意义和城市的变化，激发他们对探索和学习的热情。、互动性：通过提问或小活动，让儿童参与到文化学习中来，例如，可以邀请他们想象自己是历史事件的见证者或参与者，或者设想自己如何将学到的文化知识应用到日常生活中。

政治类儿童新闻:

板块原则

每个板块的内容要完整，每个板块的字数约在200左右，每个板块有小标题

板块一：首先介绍事件的背景，为孩子们提供足够的信息来理解事件的重要性。例如，介绍基辛格为何要进行秘密访问中国，以及这次访问的历史背景。

板块二：详细叙述事件的经过，突出重要的转折点和细节。这部分内容帮助孩子们理解事件的发展过程，如基辛格的秘密行动细节、与中国官员的会谈等。

板块三：最后，阐述事件的结果和对未来的影响。这有助于孩子们理解事件的意义，例如中美关系的改善，以及基辛格对两国友谊所做出的贡献。

政治类儿童新闻撰写的特点:

简化复杂性：将复杂的政治事件简化，使用易于理解的语言和概念，同时保留事件的核心要素。

强调教育价值：突出事件的教育意义，例如友谊、和平、合作等正面价值观。

避免偏见：在描述政治事件时保持中立，避免传达任何形式的偏见或立场，让儿童自己形成看法。

鼓励思考和讨论：在新闻的结尾提出问题或思考点，鼓励孩子们思考讨论，培养他们的批判性思维能力