

基于 Mega 模型的文本分类与长文本生成研究

许惠惠¹
XU Huihui

摘要

随着社交媒体、新闻平台和电商评论等领域中长文本数据的激增,传统模型如 RNN 和 LSTM 在建模长距离依赖方面存在局限,而 Transformer 在处理超长文本时计算开销巨大。为此,文章提出基于 Mega (moving average equipped gated attention) 模型的长文本分类与生成方法。通过引入指数移动平均 (exponential moving average, EMA) 和门控注意力机制, Mega 模型增强了长距离依赖建模能力,并通过 Mega-chunk 机制提升计算效率。在文本分类任务中, Mega 在 AG News、IMDB 和 Amazon Reviews 数据集上的表现优于 RNN、LSTM、Transformer 和 Longformer,尤其在长文本分类中展现了更高的准确率和计算效率。综上, Mega 模型通过创新设计,在长文本处理任务中实现了性能与效率的平衡,适用于智能问答、新闻摘要等实际场景。

关键词

Mega 模型; 文本分类; 长文本生成; 门控注意力机制; 指数移动平均

doi: 10.3969/j.issn.1672-9528.2025.04.029

0 引言

近年来,随着深度学习技术的快速发展,文本分类与文本生成任务取得了显著进展。这两类任务在自然语言处理 (natural language processing, NLP) 领域中扮演着至关重要的“角色”^[1-3],广泛应用于情感分析、舆情监测、新闻分类等实际场景,同时在自动写作、内容生成和智能对话系统等方面展现了巨大的创新潜力。然而,在处理海量数据和长文本时,传统模型如循环神经网络 (recurrent neural network, RNN)^[4]以及标准 Transformer^[5]受限于弱归纳偏置和高计算复杂度等问题,难以高效建模长序列信息,表现出明显的不足。

为应对上述挑战, Ma 等人^[6]提出了 Mega (moving average equipped gated attention) 模型。该模型通过引入指数移动平均 (exponential moving average, EMA)^[7]和门控注意力机制 (gated attention)^[8],在降低计算复杂度的同时显著增强了对长距离依赖关系的建模能力。Mega 模型在自注意

力机制中融入了时间衰减特性,不仅能够高效捕捉文本的局部依赖关系,还擅长建模全局语义关系,在长文本生成与理解任务中表现出色。此外, Mega 模型的 Mega-chunk 变体采用序列分块策略,具备线性复杂度,进一步优化了计算效率,使其特别适用于处理长文本数据的场景^[9]。

本文旨在深入探讨 Mega 模型在文本分类与长文本生成任务中的应用,重点评估其在准确性、计算效率以及长依赖关系处理上的优势。具体而言,本文首先针对文本分类任务,通过不同文本长度下的实验,验证 Mega 模型在情感分析和话题分类等任务中的鲁棒性;随后,在长文本生成任务中,通过新闻生成与摘要生成等实际场景,考察该模型在生成质量和上下文连贯性方面的表现。本文不仅扩展了 Mega 模型的应用场景,还为长文本处理提供了一种创新的解决方案,为未来的长序列文本分类与生成任务提供了理论与实验支持。

1 基于 Mega 的文本分类方法

为了克服传统方法在长文本分类任务中的局限性,本文基于 Mega 模型设计了一种高效的文本分类方法。Mega 模型通过其独特的指数移动平均和门控注意力机制,能够在较低计算复杂度的条件下,捕捉文本中的复杂依赖关系和上下文信息,为文本分类任务提供了强有力的支持。

文本预处理与嵌入层输入文本首先经过标准化处理,包括去除停用词、标点符号和特殊字符。假设输入序列为 $X = [x_1, x_2, \dots, x_n]$,通过预训练的词向量或上下文嵌入模型,

1. 山西药科职业学院素质教育教学研究部 山西太原 030031
[基金项目] 教育部职业院校信息化教学指导委员会 2024 年度全国高等职业院校信息技术课程教学改革研究项目课题“医药类高职院校‘信息技术’课程数字化项目实践教学建设研究”的阶段性成果 (KT2024176); 2021 年度山西省高等学校哲学社会科学研究项目 (思想政治教育专项)“新时代高职大学生群体画像构建研究”的阶段性成果 (2021zsszxsx207); 2021 年度山西省教育科学规划课题“基于神经网络模型的校企协同顶岗实习的评价研究”的阶段性成果 (PJ-21045)

得到嵌入矩阵:

$$E = [e_1, e_2, \dots, e_n], e_i \in \mathbf{R}^d \quad (1)$$

本文提出的文本分类方法主要包括以下几个模块:

(1) 基 Mega 的特征提取模块

在特征提取阶段, Mega 模型作为核心组件用于建模文本的局部和全局依赖关系。通过 EMA 机制, 模型能够捕捉输入文本中的时间序列特性, 使得上下文信息能够以递归方式自然衰减。这种设计可以有效增强远距离依赖的捕捉能力, 同时避免无关信息对分类任务的干扰。门控注意力机制则进一步强化了模型对关键特征的选择性关注能力。与传统自注意力机制不同, 门控机制通过动态权重分配过滤低重要度信息, 使得模型在文本长度增加时依然能够保持对重要上下文的关注, 从而提升分类性能。

EMA 通过引入时间衰减特性对序列建模, 其核心计算公式为:

$$h_t = \alpha h_{t-1} + (1 - \alpha) e_t \quad (2)$$

式中: h_t 为第 t 个时间步的隐藏状态; $\alpha \in [0, 1]$ 为时间衰减因子。

门控注意力机制通过门控操作动态调整注意力权重。假设注意力权重为 $A \in \mathbf{R}^{n \times n}$, 通过门控机制得到修正后的权重 A' :

$$A'_{i,j} = \sigma(g_{i,j}) \cdot A_{i,j} \quad (3)$$

$$g_{i,j} = W_g [e_i; e_j] + b_g \quad (4)$$

式中: $\sigma(\cdot)$ 为 Sigmoid 激活函数; W_g 、 b_g 为可学习参数。

(2) 分类器模块

Mega 模型提取的特征表示经过池化操作 (如平均池化、最大池化或自适应池化) 后, 被送入全连接层进行分类。池化操作在保持文本全局语义信息的同时降低了特征维度, 增强了分类器的效率和泛化能力。最后, 通过 Softmax 激活函数生成分类概率分布, 并将类别概率最高的输出作为预测结果。

Mega 提取的特征表示经过池化操作得到全局向量表示:

$$z = \text{Pooling}(H) \quad (5)$$

式中: $z \in \mathbf{R}^d$ 表示文本的全局语义表示。随后将 z 输入全连接层:

$$y = \text{Softmax}(W_c z + b_c) \quad (6)$$

式中: W_c 、 b_c 为分类器的权重和偏置参数; $y \in \mathbf{R}^C$ 为类别概率分布; C 为分类任务的类别数。

(3) 模型优化与训练策略

采用交叉熵损失函数作为优化目标, 公式为:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}) \quad (7)$$

式中: $\hat{y}_{i,c}$ 表示样本 i 在类别 c 上的预测概率; $y_{i,c}$ 表示真实标签 (One-hot 编码)。

2 实验与结果分析

2.1 数据集

在文本分类和长文本生成任务中, 数据集的选择和构建至关重要。选择 3 个数据集用于文本分类任务, 两个数据集用于长文本生成任务。

(1) 文本分类任务

① AG News 数据集: 用于新闻分类, 包含 4 个类别 (世界、体育、商业、科技), 文本短小且信息密集, 适合评估多类别分类模型性能^[10]。

② IMDB 数据集: 聚焦情感分类, 包含大量电影评论, 文本长且复杂, 挑战在于情感分析和上下文理解^[11]。

③ Amazon Reviews 数据集: 多类别商品评论数据集, 面临类别不平衡和噪声问题, 适用于测试模型在真实世界数据中的鲁棒性。

(2) 长文本生成任务

① CNN/DailyMail 数据集: 用于新闻摘要生成, 挑战在于提取关键信息并生成流畅简洁的摘要。

② WikiText-103: 大规模维基百科文章数据集, 专注于长文本生成任务, 考验模型在保持连贯性和风格一致性方面的能力。

通过对以上数据集的深入分析和实验, 可以更好地设计和优化文本分类与生成模型, 推动自然语言处理技术的发展。

2.2 性能评估指标

F_1 分数综合查准率和召回率, 适用于类别不平衡情况, 能公平评估各类别表现。BLEU 分数评估生成文本与参考文本在词汇和短语上的相似性, ROUGE 分数评估生成文本与参考文本的重叠信息量, 特别是召回率, 常用于摘要生成任务。困惑度衡量语言模型对下一个词的预测能力, 较低的困惑度表明模型理解更深入, 生成文本更流畅。

2.3 实验结果分析

为全面评估 Mega 模型的性能, 本文设计了一系列实验, 涵盖文本分类和长文本生成任务, 并将其与当前主流模型进行对比。实验采用多种评估指标, 以确保结果的公平性和权威性。表 1 详细描述实验结果与分析。

表 1 不同模型在分类任务中的表现

模型	AG News (accuracy)	IMDB (F_1)	Amazon reviews (F_1)	平均时间 / 批次
RNN	88.5	82.4	75.6	48
LSTM	89.7	84.1	77.8	72
Transformer	91.3	85.2	81.5	154
Longformer	92.1	86.7	82.3	120
Mega	92.9	87.8	83.5	85

(1) 文本分类任务

从 AG News 到 Amazon Reviews 的对比中可以看出，Mega 在长文本数据集上的表现尤为突出。传统的 RNN 和 LSTM 虽然能够处理一定长度的文本，但在捕捉长距离依赖关系方面存在不足，导致其性能较低。而 Mega 引入的 EMA 和门控注意力机制有效地缓解了这一问题，使得 Mega 能够更好地处理长文本中的复杂依赖。

Mega 模型在多个标准文本分类数据集上的表现均优于传统的循环神经网络（RNN）、LSTM 和当前先进的 Transformer 模型。Mega 通过高效的长文本依赖建模与优化的计算设计，在保证分类准确率的同时，降低了计算复杂度，特别适用于大规模文本分类任务。

(2) 长文本生成任务

通过表 2 中 ROUGE-L 和 BLEU 指标来看，Mega 能够生成质量更高的长文本，尤其在关键信息提取和语言流畅性方面超越了 LSTM、Transformer 和 Longformer。特别是在生成新闻摘要和长篇文章时，Mega 不仅能够准确捕捉文本的主要信息，还能在生成的文本中保持更高的连贯性和一致性。

表 2 生成任务的评估结果

模型	CNN/DailyMail (ROUGE-L)	WikiText-103 (BLEU)	Perplexity (PPL)	平均时间 / 批次
LSTM	35.6	24.3	42.1	115
Transformer	38.9	26.8	18.5	240
Longformer	39.4	27.1	17.9	195
Mega	41.2	29.5	16.7	160

Mega 模型在长文本生成任务中的表现显著优于传统的 LSTM 和当前流行的 Transformer、Longformer 模型。Mega 不仅在 ROUGE-L、BLEU 和困惑度等关键指标上都取得了领先的成绩，还在计算效率方面具备一定的优势。尤其在生成文本的流畅性、一致性和信息覆盖度方面，Mega 表现出了更强的能力，这使其在实际应用中更具竞争力。

总体来看，Mega 模型通过其创新的 EMA 和门控注意力机制，在长文本生成中成功地平衡了生成质量和计算效率，特别适用于需要生成序列、处理复杂上下文的自然语言生成任务。在实际应用中，Mega 可用于新闻摘要生成、自动写作和大规模内容创作等任务，展现了其广泛的应用前景。

(3) 消融实验

为了进一步分析 Mega 模型中各个组成部分的贡献，本文设计的消融实验，逐步去除不同模块并观察模型性能的变化。具体实验配置包括：去除 EMA（指数移动平均）、去除门控注意力机制、去除 Mega-chunk 机制，以及完整模型的对比，结果如图 1 所示。

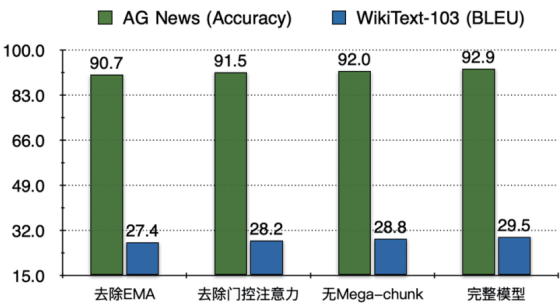


图 1 完整模型对比图

在 IMDB 数据集上，去除 EMA 和门控注意力对模型性能的影响较大，尤其在捕捉长距离依赖和关注关键信息方面，去除这些模块会导致模型性能显著下降。相较之下，去除 Mega-chunk 机制对模型性能的影响较小，表明 Mega 模型的核心能力主要来源于 EMA 和门控注意力机制，而 Mega-chunk 机制则主要在提高计算效率上起到作用。在 WikiText-103 数据集上，去除 Mega-chunk 模块后，BLEU 分数为 28.8，较完整模型的 29.5 下降了 0.7。Mega-chunk 模块主要优化计算效率，通过分块处理输入序列，降低了长文本生成中的计算开销。去除 Mega-chunk 后，模型生成质量略有下降，但相比去除 EMA 或门控注意力的影响，Mega-chunk 对生成质量的影响较小。

3 结语

本文针对长文本分类和生成任务，提出并深入探讨了基于 Mega 模型的创新应用。通过与主流自然语言处理（NLP）模型（如 RNN、LSTM、Transformer、Longformer 等）进行对比，并开展一系列消融实验。

Mega 模型的优势使其在多个实际应用场景中具有广泛的潜力，包括但不限于智能问答、新闻摘要生成、长篇文章自动生成、话题分类等任务。尤其在处理大规模文本数据和复杂上下文时，MEGA 的高效性和优异的长依赖关系建模能力，使其成为解决长文本处理挑战的理想选择。

参考文献：

[1] RESNIK P, LIN J. Evaluation of NLP systems[J/OL]. The handbook of computational linguistics and natural language processing,2010[2024-09-13].<https://doi.org/10.1002/9781444324044.ch11>.

[2] KANG Y, CAI Z, TAN C W, et al. Natural language processing (NLP) in management research: a literature review[J]. Journal of management analytics, 2020, 7(2): 139-172.

[3] 姚志新, 姜伟, 王河山, 等. 基于 NLP 的智能问答服务系统的设计与实现 [J]. 信息技术与信息化, 2019(9): 64-68.

基于可重构智能面的双空间调制方案设计

卜祥燕¹ 刘传举¹

BU Xiangyan LIU Chuanju

摘要

随着 5G 物联网技术的快速发展,降低系统能耗、提高信号传输可靠性已成为无线通信的首要问题。文章提出了一种基于可重构智能面(reconfigurable intelligent surfaces, RIS)的双空间调制(double spatial modulation, DSM)方案,简称 DSM-RIS。因其在发送端进行两次独立的空间调制,显著提升了系统的频谱效率。同时,为了克服无线信道传输的不确定性,通过可重构智能面对信息进行转发,通过优化可重构智能面的相位角提高系统传输可靠性。文章推导了 DSM-RIS 方案的平均比特错误概率(average bit error probability, ABEP)并进行了蒙特卡洛仿真,仿真结果证明了所提方案性能的优越性。

关键词

可重构智能面;双空间调制;相位优化;平均比特错误概率;5G

doi: 10.3969/j.issn.1672-9528.2025.04.030

0 引言

作为一种特殊的多输入多输出(multiple input multiple output, MIMO)技术,空间调制(spatial modulation, SM)^[1-2]技术已得到了广泛的研究和应用。在每个时隙,SM 使用附加信息比特仅激活一根发射天线传输信息,因此,相比 MIMO,SM 提高了频谱效率,降低了接收端的

检测复杂度。而正交空间调制(orthogonal spatial modulation, QSM)^[3-4]作为一种新型的 SM 技术,从另一角度提高了频谱效率和系统性能。在每个时隙, QSM 在发送端激活两个发射天线,分别传输相互正交的调制符号实部和虚部。因此, QSM 降低了天线间的相互干扰。在文献[5]中,一种双空间调制(double spatial modulation, DSM)技术被提出。DSM 通过在发送端独立地进行两次空间调制将系统的频谱效率提高两倍,同时联合星座优化问题提高了整

1. 山东劳动职业技术学院 山东济南 250301

- [4] 曾强,刘晓群,郝娟.基于 RNN 与级联损失函数的图像超分辨率研究[J].信息技术与信息化,2024(10):81-84.
- [5] 夏雪,闫恩来,李喜武.Transformer 在时间序列预测中的应用综述[J].信息技术与信息化,2024(3):124-128.
- [6] MA X Z, ZHOU C T, KONG X, et al. Mega: moving average equipped gated attention[DB/OL].(2023-01-28)[2024-06-11]. <https://doi.org/10.48550/arXiv.2209.10655>.
- [7] KLINKER F. Exponential moving average versus moving exponential average[DB/OL].(2020-01-20)[2024-10-12]. <https://doi.org/10.1007/s00591-010-0080-8>.
- [8] QIU X R, ZHU R J, CHOU Y H, et al. Gated attention coding for training high-performance and efficient spiking neural networks[C]//Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence. Palo Alto: AAAI, 2024: 601-610.

- [9] REN L L, LIU Y, WANG S H, et al. Sparse modular activation for efficient sequence modeling[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. New York: ACM, 2023: 19799-19822.
- [10] ZHANG X, ZHAO J B, LECUN Y. Character-level convolutional networks for text classification[C]//Proceedings of the 29th International Conference on Neural Information Processing Systems. New York: ACM, 2015: 649-657.
- [11] TRIPATHI S, MEHROTRA R, BANSAL V, et al. Analyzing sentiment using IMDb dataset[C]//2020 12th International Conference on Computational Intelligence and Communication Networks (CICN). Piscataway: IEEE, 2016: 520-523.

【作者简介】

许惠惠(1983—),女,山西洪洞人,硕士,讲师,研究方向:计算机应用、数据分析与挖掘。

(收稿日期:2024-12-05)