简化基于内容的神经新闻推荐：关于用户建模和训练目标

Andreea Iana 曼海姆大学 德国曼海姆
andreea.iana@uni-mannheim.de

Goran Glavaš CAIDAS，维尔茨堡大学 德国维尔茨堡
g
oran.glavas@uni-wuerzburg.de

Heiko Paulheim 曼海姆大学 德国曼海姆
heiko.paulheim@uni-mannheim.de

arXiv:2304.03112v1 [cs.IR] 6 Apr 2023

摘要

个性化新闻推荐的兴起催生了日益复杂的推荐系统架构。大多数神经新闻推荐系统依赖用户点击行为，并通常引入专门的用户编码器，将已点击新闻的内容聚合为用户嵌入（早期融合）。这些模型主要采用标准的逐点分类目标进行训练。现有研究主要存在两个缺陷：(1) 尽管整体设计具有同质性，但由于评估数据集和协议各不相同，模型之间的直接比较受到阻碍；(2) 替代性的模型设计和训练目标仍未得到充分探索。在本研究中，我们提出了一个统一的新闻推荐框架，使得能够围绕几个关键设计维度，对神经新闻推荐系统进行系统且公平的比较：(i)用户建模中的候选感知，(ii) 点击行为融合，以及(iii)训练目标。我们的研究结果对神经新闻推荐领域的现状提出了挑战。我们表明，以候选新闻嵌入与已点击新闻嵌入之间参数高效的点积（晚期融合）取代大规模用户编码器，往往能够带来显著的性能提升。此外，我们的结果表明，对比学习训练是逐点分类目标的一种可行替代方案。

CCS 概念

• **信息系统 →推荐系统**。

关键词

神经新闻推荐、用户建模、晚期融合、训练目标、对比学习、评估

ACM 参考格式：

Andreea Iana、Goran Glavaš 和 HeikoPaulheim。2023。简化基于内容的神经新闻推荐：关于用户建模和训练目标。载于*SIGIR '23* 会议论文集。ACM，纽约州纽约市，共 5 页。https://doi.org/XXXXXXX.XXXXXXX

1 引言

近年来，基于内容的新闻推荐中出现了日益复杂的神经推荐架构，旨在根据用户的兴趣定制推荐内容 [13, 34]。大多数神经新闻推荐模型（NNR）通常由两部分组成：(i) 专门的新闻编码器（NE）和 (ii) 用户编码器（UE） [34, 36]。新闻编码器（NE）——具体形式包括卷积网络 [26, 29, 31], 、自注意力网络 [19, 32, 35], 、图注意力网络 [18], , 或最近的预训练变换器网络 [36, 41] ——将输入特征（例如标题、类别、实体）转换为新闻嵌入。用户编码器（UE）通过序列编码器 [1, 21, 27]或注意力编码器 [26, 29, 32]聚合已点击新闻的嵌入，生成用户级表示；这些编码器根据点击行为模式 [1, 15, 37]，为已点击新闻的嵌入提供上下文信息。我们将这一主流范式称为早期融合（EF），因为它在与推荐候选项进行比较之前，先聚合已点击新闻的表示（即构建用户表示）。

大多数神经新闻推荐模型会以一种与候选项无关的方式分别对用户和候选新闻进行编码 [1, 29, 32]。与之相反，候选项感知的模- 型 [20, 22, 25, 42], 认识到，并非所有已点击新闻都能同等有效地反映候选项的相关性（例如，候选项通常只代表用户偏好的一个子集），并在使用用户编码器进行用户级聚合时，利用候选项的嵌入对已点击新闻的表示进行上下文处理。最后，将候选项的嵌入（新闻编码器的输出）与用户嵌入（用户编码器的输出）进行比较：候选项的推荐分数直接计算为两个嵌入的点积 [29] , 或通过前馈评分器计算 [26]。神经新闻推荐模型主要通过带有负采样的标准分类目标进行训练[26, 29, 32, 36] [7, 30]。

现有研究主要存在两个缺点。首先，尽管总体设计具有同质性，但由于缺乏透明度以及采用临时性的评估协议 [8, 23]，近期神经新闻推荐系统之间的直接比较受到阻碍。具体而言，绝大多数个性化新闻推荐器都在专有数据集上进行评估（例如MSN 新闻 [29, 32],必应新闻 [26], NewsApp [20]）

即使是少数使用 Adressa [6] 或 MIND [38]等公开数据集进行评估的模型，也无法直接比较，因为它们采用了不同的数据集划分和评估协议（例如，模型选择策略） [5, 27, 36, 42]。其次，更简单且或许更直观的设计方案在很大程度上仍未得到探索。

首先，现有工作采用射血分数作为默认架构，提出了日益复杂的用户编码组件 [1, 20],，但通常很少为增加的复杂性提供实证依据。其次，只有一小部分神经新闻推荐系统利用对比学习目标 [33,41], , 尽管已有证明表明，这类训练标准在密切相关的检索和推荐任务中非常有效 [14, 28, 39, 40]

在这项工作中，我们弥补了当前神经新闻推荐模型的上述不足，并为用户建模和训练目标带来了新的见解。[1] **1)**

允许为个人或课堂使用而免费制作本作品全部或部分内容的数字或纸质副本，但不得为盈利或商业利益而制作或分发副本，且副本必须在首页附上本声明和完整引文。必须尊重本作品中归 ACM 以外其他方所有的组成部分的版权。注明出处后允许摘录。若要以其他方式复制、再版、发布到服务器或再分发至列表，须事先获得明确许可和/或支付费用。请通过 permissions@acm.org 申请许可。*SIGIR '23，2023*年7月*23*日至*27*日，中国台湾台北© 2023 计算机协会。
ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00
https://doi.org/XXXXXXX.XXXXXXX

[1]**免责声明：** 在这项工作中，我们专门研究不依赖基于图的建模来表示用户之间关系的神经新闻推荐模型。

具体而言，我们提出了一个统一的神经新闻推荐框架，以便在三个关键设计维度上对神经新闻推荐模型进行系统且公平的比较：(i) 用户建模中的候选项感知，(ii) 点击行为融合，以及 (iii) 训练目标。2) 我们提出用对候选新闻与已点击新闻嵌入之间的点积得分进行简单池化（即晚期融合），取代采用复杂用户编码器的用户建模（即早期融合）。我们表明，尽管概念简单，晚期融合仍比基于早期融合的神经新闻推荐带来显著的性能提升，因此，复杂的用户嵌入在经验上并无必要。3) 最后，我们展示了监督式对比训练的优势，证明它是逐点分类的一种可行替代方案。我们的工作通过为基于复杂用户建模的既有范式引入更简单、更有效的替代方案，从根本上挑战了神经新闻推荐领域的现状。

2 方法

图1展示了我们面向神经新闻推荐（NNR）的统一评估框架，聚焦于新闻推荐的三个关键维度。给定由新闻和用户行为组成的输入数据，我们分析：（i）候选项无关（C-AG）与候选项感知（C-AW）的用户建模；（ii）两种点击行为融合策略，即射血分数（EF）和晚期融合（LF）；以及（iii）通过最小化标准交叉熵损失（CE）或监督式对比目标（SCL）来训练各模型。接下来，我们将介绍选用于评估的模型，并对具体的设计选择进行形式化描述。

2.1 用户建模

候选项无关（C-AG）模型。对于这些模型，用户编码器（UE）根据已点击新闻的嵌入生成用户嵌入，而不针对候选项进行上下文建模。我们评估以下 C-AG 模型，它们的新闻编码器（NE）组件各不相同（即嵌入已点击新闻的方式不同）：(1) *NPA* [30] 使用个性化注意力模块聚合用户已点击新闻的表示，并将用户 ID 的投影嵌入用作注意力查询；(2) *NAML* [29] 使用加性注意力 [2] 对用户偏好进行编码；(3) 归一化均方根 [32] 使用由两层编码器学习用户表示，该编码器由多头自注意力 [24] 和加性注意力组成；(4) *LSTUR* [38] 使用循环网络学习用户表示：通过 GRU [4],从已点击新闻生成短期用户嵌入，并将其与长期嵌入相结合；长期嵌入由随机初始化并经过微调的部分构成。随后，通过以下任一方式获得最终用户嵌入：(i) 将长期嵌入作为初始状态，取短期 GRU 的最终隐藏状态（*LSTURini*）；或 (ii) 直接拼接短期和长期用户嵌入（*LSTURcon*）；(5)*Cen-NewsRec* [21] 采用与 LSTUR 类似的用户编码器（UE）架构，但它通过多头自注意力和注意力池化网络序列从已点击新闻中学习长期用户向量，而不是为每位用户存储显式嵌入；(6)*MINS* [27] 通过结合多头自注意力、多通道 GRU 循环网络和加性注意力来编码用户。

候选感知（C-AW）模型。候选无关模型中的用户嵌入会生成相同的用户嵌入，不受候选新闻内容的影响。相比之下，候选感知模型中的用户嵌入（我们在实证分析中纳入了其中两种）会生成用户

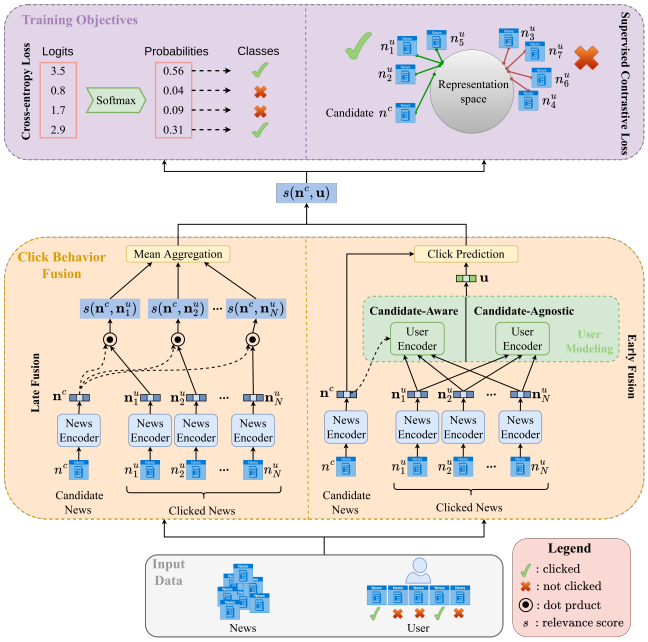

图 1：统一神经新闻推荐框架示意图，重点展示三个关键设计维度：（i）用户建模中的候选感知（绿色框）；（ii）点击行为融合（橙色框）；以及（iii）训练目标（紫色框）。

依赖于候选项的嵌入。(7) *DKN* [26]将用户的候选项感知表示计算为其已点击新闻嵌入的加权和，权重由注意力网络生成，该网络以候选项和已点击新闻的嵌入作为输入，这些嵌入由新闻编码器生成。最近，(8) *CAUM* [20] 结合了(i)候选项感知的自注意力网络，用于在候选项条件下建模已点击新闻之间的长程依赖关系，以及(ii)候选项感知的卷积网络（卷积神经网络），用于从相邻点击中捕捉短期用户兴趣，同样以候选项的内容为条件；最终通过对长程表示和短期表示进行注意力计算，得到候选项感知的用户嵌入。

新闻编码器。我们评估的神经新闻推荐模型主要以新闻标题作为输入，通常使用预训练词嵌入 [17]对其进行嵌入。NAML、LSTUR、MINS 和 CAUM 还利用类别信息，并通过线性层对类别进行嵌入。CAUM 还会编码标题实体，而 DKN 则利用知识图谱嵌入 [9]。浅层词嵌入和实体嵌入通过以下方式之一融入上下文信息：结合多头自注意力（用于 NRMS、MINS 和 CAUM），或采用由卷积神经网络 [11] 和加性注意力网络组成的序列（用于 NAML 和 LSTUR）。NPA [30] 也使用卷积神经网络为词嵌入融入上下文信息，随后接入个性化注意力模块，其结构类似于用户编码器中的模块；DKN 则采用词实体对齐的知识感知卷积神经网络 [26]。CenNewsRec[21] 将卷积神经网络与多头自注意力和加性注意力模块相结合。具有多个特征向量的模型通过简单拼接（LSTUR、CAUM）或对这些向量进行注意力运算（NAML、MINS）来生成最终的新闻嵌入。

2.2 点击行为融合

我们质疑早期融合（EF）的设计和计算复杂度是否合理；早期融合即当前最先进的神经新闻推荐模型中是否有必要设置专门的用户编码器。为此，我们提出一种轻量级替代方案：晚期融合（LF）方法。该方法以候选项 n^c 嵌入与已点击新闻 n_i^u 嵌入之间点积得分的均值池化，取代复杂的用户编码器。给定一条候选新闻 n^c 和用户点击的一系列新闻 $H = n_1^u, \ldots\ldots n_N^u$，我们计算候选新闻相对于用户 u 历史记录的相关性得分，表示为 $s(\mathbf{n}^c, u) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{n}^c \cdot \mathbf{n}_i^u$，其中 $\mathbf{n}$ 表示由新闻编码器学习得到的新闻嵌入，N 表示历史记录长度。

尽管晚期融合表明，对于点击预测而言，显式编码用户行为可能并非必要，但协同过滤模型中仍然需要用户嵌入 [13]。请注意，上述晚期融合公式等价于候选项嵌入 $\mathbf{n}^c$ 与用户已点击新闻嵌入的均值 $\mathbf{n}_i^u$，$s(\mathbf{n}^c, u) = \mathbf{n}^c \cdot \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{n}_i^u\right)$ 之间的点积。这意味着，如果需要，晚期融合也可以无缝地提供用户嵌入（只需将已点击新闻嵌入取平均值即可）。因此，晚期融合可以被视为一种无参数用户编码器，即现有射血分数模型中复杂参数化用户编码器的计算高效替代方案。由于 (i) 我们独立生成候选项和已点击新闻的嵌入，并且 (ii)将已点击新闻嵌入的平均值作为用户嵌入，因此，晚期融合模型是候选无关的（C-AG）。

2.3 训练目标

现有神经新闻推荐（NNR）研究绝大多数无论采用何种具体的用户建模架构，都是通过最直接明了的分类目标——交叉熵损失（结合负采样；见图1）来调整参数，并且基本没有探索有效的替代方案，尤其是对比目标 [16, 33]。这使得我们无法了解模型在不同训练方案下的效果。为解决这一局限，我们不仅使用（1）常见的交叉熵损失（结合负采样），还使用（2）对比学习目标，具体而言是监督式对比损失 [10]，来训练所有模型（见 §2.1）。

3 实验设置

数据。我们在 Wu等人 [38]提出的 MINDsmall 和 MINDlarge 数据集上开展实验。表1总结了它们的主要统计信息。由于 Wu等人 [38] 未公开测试集标签，我们使用相应的验证集部分进行测试，并将各自的训练集划分为时间不相交的训练集（前四天的数据）和验证集部分（最后一天）。

实现与优化细节。我们使用300维预训练的GloVe词嵌入 [17] 和在Wikidata上预训练的100维TransE嵌入 [3]，分别初始化参与比较的神经新闻推荐模型的词嵌入和实体嵌入。我们将最大历史长度设为50。按照Wu等人 [33], 的方法，我们的负采样为每个正样本创建四个负样本。我们根据验证集表现确定SCL的最佳温度，在区间 $[0.08, 0.3]$ 内以0.02为步长进行搜索。对于所有C-AG模型，我们使用512的批大小进行训练；对于DKN，使用256；对于CAUM，由于计算限制，仅使用64。我们将所有其他模型特定超参数采用各论文报告的最优值。所有模型均采用混合精度训练，并使用固定计算预算：在 MINDsmall 上训练 25 个周期，在 MINDlarge 上训练 10 个周期。我们使用 Adam 算法 [12], 进行优化，学习率设为 $1e$-4。每项实验重复五次（使用不同的随机种子），并报告常见指标的平均值（及标准差）：曲线下面积、平均倒数排名、归一化折损累计增益@5 和归一化折损累计增益@10。每个模型均在一块配备 32GB 显存的 NVIDIA Tesla V100 GPU 上训练。我们的实现已公开发布。[2]

表 1：MINDsmall 和 MINDlarge 数据集的统计信息。

	MINDsmall		MINDlarge	
统计项	训练	Test	训练	Test
# 新闻	51,282	42,416	101,527	72,023
# 用户	49,108	48,593	698,365	248,973
# 印象	153,727	70,938	2,186,683	365,201
# 类别	17	17	18	17
# 子类别	264	252	285	269

4 结果与讨论

表 2 展示了在我们比较评估框架的四种不同配置下，C-AG（NPA、NAML、NRMS、LSTUR、CenNewsRec 和 MINS）与 C-AW 模型（DKN、CAUM）在 MINDsmall 和 MINDlarge 上的表现：(i) 采用 EF 与 LF 进行用户建模，以及 (2) 使用交叉熵与 SCL 目标进行训练。接下来，我们将沿着框架的三个维度（§2）剖析结果：用户建模、点击行为融合和训练目标。

候选项无关与候选项感知神经新闻推荐模型。由于采用 LF 后所有模型都会变为候选项无关模型，因此我们在默认的 EF 配置下分析 C-AG 与 C-AW 模型。CAUM 拥有最复杂且具备候选项感知能力的用户编码器，在交叉熵和 SCL 两种训练方式下，以及大多数评估指标上，通常都优于其他模型。在大型训练数据集 MINDlarge 上，就曲线下面积指标而言，差距尤为显著。仅凭这一结果，可能会得出这样的结论：要获得更好的推荐，必须采用更复杂、具备候选项感知能力的用户建模。然而，(1) 在我们的评估中，作为另一种 C-AW 模型的 DKN 通常远逊于 C-AG 模型；(2) 我们采用简单、无参数用户编码器的 LF 模型变体，其表现与采用 EF 的 CAUM 相当，甚至更优。这些事实都削弱了这一结论。除 DKN 外，所有其他模型在更大的 MINDlarge 数据集上训练时，表现都更好。不过，NAML 和 MINS 在 MINDsmall 上也具有竞争力（曲线下面积指标除外），但在 MINDlarge 上不及 CAUM，这表明 CAUM 精巧的用户编码器从更多训练数据中获益最大。

然而，有一个我们未加以控制、因此在比较结果时需要谨慎考虑的混杂因素：模型之间的差异不仅体现在用户嵌入上，也体现在新闻编码器组件上，即它们如何编码新闻以及使用哪些特征作为输入。例如，NAML 和 MINS 使用相同的新闻编码器，在 MINDsmall 上取得了相近的性能。然而，在 MINDlarge 上，MINS 更复杂的用户编码器相较于 NAML 较简单的用户编码器带来了显著提升（但仅限于采用标准射血分数融合和交叉熵训练的情况）。

[2]代码可在此处获取：https://github.com/andreeaiana/simplifying_nnr

表2：在点击行为融合（CBF）与训练目标的不同组合下，各对比模型的推荐性能。我们报告了五次不同运行结果的平均值和标准差。

		MINDsmall								MINDlarge							
		AUC		MRR		归一化折损累计增益@5		归一化折损累计增益@10		AUC		MRR		归一化折损累计增益@5		归一化折损累计增益@10	
模型	CBF	CE	SCL	CE	SCL	CE	SCL	CE	SCL	CE	SCL	CE	SCL	CE	SCL	CE	SCL
NPA	EF	54.7±0.6	56.5±0.7	29.0±0.7	28.4±0.6	26.9±0.8	26.6±0.6	33.2±0.8	32.6±0.4	56.8±0.2	58.1±0.8	31.4±0.5	30.0±0.6	29.5±0.4	27.7±0.6	35.9±0.4	34.3±0.6
	LF	55.1±0.9	57.3±1.2	28.6±0.3	27.5±1.0	26.4±0.4	25.5±1.1	32.9±0.4	31.8±0.9	61.2±0.5	58.7±0.8	32.1±0.6	28.3±0.5	30.2±0.7	26.0±0.7	36.6±0.6	32.7±0.7
NAML	EF	50.1±0.0	57.1±1.1	33.6±0.5	32.2±0.7	31.6±0.6	30.4±0.7	38.0±0.5	36.8±0.7	50.1±0.0	60.4±0.6	33.2±0.4	34.2±0.4	31.3±0.5	32.5±0.3	37.9±0.4	38.9±0.3
	LF	50.0±0.0	62.7±0.5	33.7±0.8	32.0±0.7	31.8±0.9	30.3±0.7	38.1±0.8	36.6±0.6	50.0±0.0	65.4±0.5	32.7±0.5	33.5±0.5	31.0±0.5	31.7±0.5	37.6±0.4	38.2±0.4
NRMS	EF	52.6±1.3	59.9±0.6	27.6±0.8	29.2±0.7	25.7±0.5	27.2±0.9	32.3±0.5	33.7±0.7	54.6±1.4	62.8±0.7	31.9±1.0	32.4±0.5	30.0±1.1	30.5±0.7	36.6±1.0	36.9±0.6
	LF	58.9±1.0	60.2±1.1	31.8±0.7	30.7±0.6	29.9±0.7	28.7±0.6	36.3±0.6	35.1±0.6	56.1±2.1	63.6±1.1	32.9±0.7	32.4±0.6	31.7±1.1	30.6±0.8	37.8±0.4	37.1±0.7
LSTUR	联合分数	53.5±1.2	55.4±0.5	29.6±0.5	28.1±0.7	28.0±0.5	26.5±0.7	34.4±0.4	32.9±0.6	50.0±0.1	56.9±1.5	32.5±2.4	31.3±1.5	30.9±2.4	29.8±1.8	37.4±2.4	36.2±1.6
	EF_{con}	50.2±0.0	59.8±1.4	31.8±0.7	31.3±0.8	30.1±0.8	30.3±1.3	36.4±0.7	36.2±0.7	51.4±0.4	54.3±0.4	27.7±0.4	26.5±0.2	25.9±0.5	24.6±0.2	32.3±0.5	31.1±0.2
	LF	50.0±0.0	50.0±0.0	33.8±0.6	33.9±0.6	31.9±0.7	32.0±0.7	38.0±0.6	38.1±0.6	50.0±0.0	50.0±0.0	34.7±0.6	33.1±0.2	33.2±0.6	31.6±0.4	39.2±0.5	37.7±0.3
CenNewsRec	EF	54.0±0.8	60.0±0.4	28.3±0.5	30.6±0.8	26.5±0.4	28.5±0.8	32.9±0.3	34.8±0.7	56.4±0.8	64.5±0.4	33.7±0.3	33.5±0.4	31.9±0.3	31.8±0.5	38.3±0.2	38.1±0.4
	LF	59.3±0.6	61.9±0.7	32.8±0.8	32.2±0.8	30.9±0.8	30.4±0.8	37.1±0.6	36.6±0.7	53.3±0.7	64.2±0.6	33.2±0.4	33.3±0.4	31.4±0.5	31.7±0.4	37.9±0.4	38.1±0.4
MINS	EF	50.6±0.3	62.9±1.7	33.7±1.0	32.4±0.3	31.9±1.1	30.7±0.4	38.3±0.9	37.1±0.3	51.7±0.2	65.8±0.5	34.3±0.2	34.4±0.5	32.5±0.4	32.6±0.5	39.1±0.4	39.1±0.5
	LF	59.1±1.2	64.2±0.7	35.0±0.5	34.1±0.6	33.2±0.6	32.3±0.7	39.4±0.6	38.5±0.6	53.8±0.6	66.7±0.8	34.9±0.2	34.8±0.7	33.0±0.2	33.1±0.7	39.5±0.2	39.6±0.6
DKN	EF	50.0±0.0	51.0±2.3	26.4±0.6	25.9±0.9	24.4±0.7	23.9±1.1	31.0±0.6	30.5±0.9	50.0±0.0	50.0±0.0	25.2±0.4	24.8±0.3	23.4±0.7	22.6±0.3	30.0±0.5	29.1±0.3
	LF	50.0±0.0	50.0±0.0	27.5±0.6	26.4±0.8	25.0±0.5	24.0±0.8	31.7±0.6	30.8±0.8	50.0±0.0	50.0±0.0	29.1±0.4	27.8±1.1	26.3±0.3	25.4±1.0	33.2±0.4	32.1±1.0
CAUM	EF	61.4±1.0	63.2±0.9	33.8±0.6	33.7±0.8	32.0±0.6	31.8±0.9	38.4±0.5	38.2±0.8	67.1±0.8	66.4±0.9	35.3±0.5	35.1±0.5	33.6±0.6	33.4±0.6	40.1±0.5	39.9±0.5
	LF	62.4±0.8	63.5±0.8	33.7±0.6	33.7±0.7	31.8±0.5	31.8±0.8	38.2±0.5	38.0±0.7	53.1±0.3	65.9±0.2	34.5±0.4	34.5±0.1	32.6±0.3	32.8±0.1	39.2±0.3	39.3±0.1

早期与晚期点击行为融合。 将基于复杂射血分数的用户嵌入替换为我们提出的简单、无参数的晚期融合，在各方面都带来了显著的性能提升。综合所有模型和两种训练目标来看，晚期融合在 MINDsmall 和 MINDlarge 上分别带来了 5.58 和 4.63 个平均倒数排名点的大幅提升。同样重要的是，采用晚期融合——即使用相同的无参数用户嵌入——时，各模型的性能彼此更加接近，其他模型通常也缩小了与 CAUM 的差距。这表明，晚期融合降低了不同模型间新闻编码器架构差异的影响，因此不仅通过对已点击新闻嵌入进行无参数平均简化了用户编码器，也使得采用更简单的新闻编码器成为可能。

交叉熵与监督式对比损失。 SCL 是常见的基于负采样的交叉熵分类方法的一种可行替代方案（比较表 2 中各评估指标下的 CE 和 SCL 列）。在曲线下面积方面，SCL 相较于 CE 有显著提升（在 MINDsmall 上提升 +8.26 个百分点，在 MINDlarge 上提升 +12.14 个百分点；均为所有模型在 EF 和 LF 变体下的平均值）。这表明，SCL 能够在表示空间中更好地区分已点击新闻和未点击新闻。相比之下，根据平均倒数排名和归一化折损累计增益这两项排序指标，SCL 略逊于 CE（在 MINDsmall 和 MINDlarge 上分别低 1.54 和 1.78 个平均倒数排名点）。我们推测，这是因为存在困难负样本——用户未点击但与其已点击新闻相似的新闻——而 CE 对这类新闻会更直接地表明其不相关；对于经 SCL 训练的模型，这些新闻可能会成为排名靠前的假阳性。

模型规模。 最后，我们量化 LF 相较于 EF 带来的模型参数减少量。图2展示了在 MINDsmall 上原始 EF 配置中的可训练参数数量。[3] 虽然在大多数模型中，NE 占据了参数的大部分，但图中显示，若干模型的 UE 参数占比不可忽视，而 LSTUR 的这一占比则遥遥领先。采用无参数的 UE 后，除性能提升外，LF 还使 CenNewsRec、CAUM 和 LSTURini 的模型规模分别相对减少了 14.7%、18.1% 和高达 82.3%。

[3]对于某些模型（例如使用用户嵌入矩阵的 LSTUR），参数数量取决于训练数据的规模。

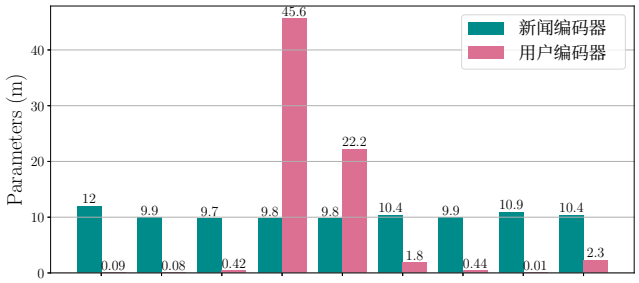

图2：模型参数数量（百万）。

5 结论

个性化神经新闻推荐器的快速发展，阻碍了公平的比较性模型评估和对设计选择的系统分析。在本研究中，我们提出了一个统一的神经新闻推荐框架，聚焦于 NNR 的三个关键设计维度：(i) 用户建模中的候选感知，(ii) 点击行为融合，以及 (iii) 训练目标。对众多最新最先进模型进行广泛评估后发现，NNR 可以大幅简化：用对已点击新闻嵌入进行无参数聚合的方式取代复杂的用户编码器，能够全面显著提升性能，同时降低模型复杂度。此外，我们还表明，对比学习可以作为标准分类式（交叉熵）损失的可行替代方案。我们希望这些发现能够推动更透明的 NNR 评估，包括开展系统的模型消融，以揭示驱动性能提升的组件。

致谢

作者感谢巴登-符腾堡州通过 bwHPC 提供的支持。AndreeaIana获得了 ReNewRS 项目资助，该项目由巴登-符腾堡基金会的负责任人工智能计划资助。

参考文献

[1] Mingxiao An、Fangzhao Wu、Chuhan Wu、Kun Zhang、Zheng Liu 和 Xing Xie。2019。基于长短期用户表示的神经新闻推荐。载于计算语言学协会第57届年会论文集。336–345。[2] Dzmitry Bahdanau、Kyunghyun Cho 和 Yoshua Bengio。2014。通过联合学习对齐与翻译实现神经机器翻译。ICLR（2014）。[3] Antoine Bordes、Nicolas Usunier、Alberto Garcia-Duran、Jason Weston 和 Ok- sana Yakhnenko。2013。用于建模多关系数据的嵌入翻译。载于第26届神经信息处理系统国际会议论文集——第2卷，第26卷。[4] Kyunghyun Cho、Bart van Merriënboer、Caglar Gulcehre、Dzmitry Bahdanau、Fethi Bougares、Holger Schwenk 和 Yoshua Bengio。2014。使用RNN编码器–解码器学习用于统计机器翻译的短语表示。载于2014年自然语言处理实证方法会议（EMNLP）论文集。1724–1734。[5] Shansan Gong 和 Kenny Q Zhu。2022。正向、负向与中性：在基于会话的新闻推荐中建模隐式反馈。载于第45届ACM SIGIR信息检索研究与发展国际会议论文集。1185–1195。[6] Jon Atle Gulla、Lemei Zhang、Peng Liu、Özlem Özgöbek 和 Xiaomeng Su。2017。用于新闻推荐的Adressa数据集。载于网络智能国际会议论文集。1042–1048。[7] Po-Sen Huang、Xiaodong He、Jianfeng Gao、Li Deng、Alex Acero 和 Larry Heck。2013。利用点击数据学习用于网页搜索的深度结构化语义模型。载于第22届ACM信息与知识管理国际会议论文集。2333–2338。[8] Andreea Iana、Mehwish Alam 和 Heiko Paulheim。2022。知识感知新闻推荐系统综述。语义网预印本（2022），1–62。[9] Guoliang Ji、Shizhu He、Liheng Xu、Kang Liu 和 Jun Zhao。2015。通过动态映射矩阵实现知识图谱嵌入。载于第53届计算语言学协会年会暨第7届自然语言处理国际联合会议论文集（第1卷：长篇论文）。687–696。[10] Prannay Khosla、Piotr Teterwak、Chen Wang、Aaron Sarna、Yonglong Tian、Phillip Isola、Aaron Maschinot、Ce Liu 和 Dilip Krishnan。2020。监督式对比学习。载于第34届神经信息处理系统国际会议论文集，第33卷。18661–18673。[11] Yoon Kim。2014。用于句子分类的卷积神经网络。载于2014年自然语言处理实证方法会议（EMNLP）论文集。计算语言学协会，多哈，卡塔尔，1746 1751 10 3115 1 14 1181 – 。https://doi.org/ 。/v / [12] D - Diederik P Kingma 和 Jimmy Ba。2014。。Adam：一种随机优化方法。ICLR（2014）。[13] Miaomiao Li 和 Licheng Wang。2019。个性化新闻推荐技术综述。IEEE Access 7（2019），145861–145879。[14] Yizhi Li、Zhenghao Liu、Chenyan Xiong 和 Zhiyuan Liu。2021。通过对比式双重学习提升稠密检索的鲁棒性。载于2021年ACM SIGIR信息检索理论国际会议论文集。287–296。[15] Shumpei Okura、Yukihiro Tagami、Shingo Ono 和 Akira Tajima。2017。面向数百万用户的基于嵌入的新闻推荐。载于第23届ACM SIGKDD知识发现与数据挖掘国际会议论文集。1933–1942。[16] Aaron van den Oord、Yazhe Li 和 Oriol Vinyals。2018。基于对比预测编码的表示学习。arXiv预印本 arXiv:1807.03748（2018）。[17] Jeffrey Pennington、Richard Socher 和 Christopher D Manning。2014。GloVe：词表示全局向量。载于2014年自然语言处理实证方法会议（EMNLP）论文集。1532–1543。[18] Tao Qi、Fangzhao Wu、Chuhan Wu 和 Yongfeng Huang。2021。基于知识感知交互匹配的个性化新闻推荐。载于第44届ACM SIGIR信息检索研究与发展国际会议论文集。61–70。[19] Tao Qi、Fangzhao Wu、Chuhan Wu 和 Yongfeng Huang。2021。PP-Rec：基于个性化用户兴趣和时间感知新闻热度的新闻推荐。载于计算语言学协会第59届年会暨第11届自然语言处理国际联合会议论文集（第1卷：长篇论文）。5457–5467。[20] Tao Qi、Fangzhao Wu、Chuhan Wu 和 Yongfeng Huang。2022。基于候选项感知用户建模的新闻推荐。载于第45届ACM SIGIR信息检索研究与发展国际会议论文集。1917–1921。[21] Tao Qi、Fangzhao Wu、Chuhan Wu、Yongfeng Huang 和 Xing Xie。2020。保护隐私的新闻推荐模型学习。载于计算语言学协会论文集：EMNLP 2020。1423–1432。[22] Tao Qi、Fangzhao Wu、Chuhan Wu、Peiru Yang、Yang Yu、Xing Xie 和 Yongfeng Huang。2021。HieRec：用于个性化新闻推荐的层次化用户兴趣建模。载于计算语言学协会第59届年会论文集。计算语言学与第11届自然语言处理国际联合会议（第1卷：长篇论文）。5446–5456。[23] ShainaRaza和Chen Ding。2022。新闻推荐系统：近期进展、挑战与机遇综述。ArtificialIntelligence Review（2022），1–52。[24] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser 和 Illia Polosukhin。2017。Attention Is All YouNeed。在神经信息处理系统第31届国际会议论文集第30卷中。[25] Heyuan Wang、Fangzhao Wu、Zheng Liu 和 Xing Xie。2020。神经新闻推荐的细粒度兴趣匹配。在计算语言学协会第58届年会论文集。836–845。[26] Hongwei Wang、FuzhengZhang、XingXie 和Minyi Guo。2018。DKN：用于新闻推荐的深度知识感知网络。在2018年万维网大会论文集。1835–1844。[27] Rongyao Wang、Shoujin Wang、WenpengLu 和 Xueping Peng。2022。通过多兴趣新闻序列建模进行新闻推荐。在ICASSP 2022—2022年IEEE声学、语音与信号处理国际会议（ICASSP）中。IEEE，7942–7946。[28] Yinwei Wei、Xiang Wang、Qi Li、Liqiang Nie、Yan Li、XuanpingLi 和 Tat-SengChua。2021。用于冷启动推荐的对比学习。在第29届ACM多媒体国际会议论文集。5382–5390。[29] Chuhan Wu、FangzhaoWu、Mingxiao An、JianqiangHuang、Yongfeng Huang 和 XingXie。2019。采用注意力多视图学习的神经新闻推荐。在第28届人工智能国际联合会议论文集。3863–3869。[30] Chuhan Wu、FangzhaoWu、Mingxiao An、JianqiangHuang、Yongfeng Huang 和 Xing Xie。2019。NPA：采用个性化注意力的神经新闻推荐。在第25届ACMSIGKDD知识发现与数据挖掘国际会议论文集。2576–2584。[31] Chuhan Wu、Fangzhao Wu、Mingxiao An、Yongfeng Huang 和 Xing Xie。2019。采用主题感知新闻表示的神经新闻推荐。在计算语言学协会第57届年会论文集。1154–1159。[32] Chuhan Wu、Fangzhao Wu、Suyu Ge、Tao Qi、YongfengHuang 和 Xing Xie。2019。采用多头自注意力的神经新闻推荐。在2019年自然语言处理实证方法会议与第9届自然语言处理国际联合会议（EMNLP-IJCNLP）论文集。6389–6394。[33] ChuhanWu、Fangzhao Wu 和YongfengHuang。2022。重新审视InfoNCE：你需要多少负样本？在第31届人工智能国际联合会议论文集（IJCAI-22）中，Lud DeRaedt（编）。人工智能国际联合会议组织，2509–2515。[34] Chuhan Wu、Fangzhao Wu、Yongfeng Huang 和 XingXie。2023。个性化新闻推荐：方法与挑战。ACM信息系统汇刊41，1（2023），1–50。[35] Chuhan Wu、FangzhaoWu、Tao Qi和Yongfeng Huang。2019。Sentirec：情感多样性感知的神经新闻推荐。在亚太地区计算语言学协会第1届会议与第10届自然语言处理国际联合会议论文集。44–53。[36] Chuhan Wu、FangzhaoWu、Tao Qi 和 YongfengHuang。2021。利用预训练语言模型增强新闻推荐。在第44届ACMSIGIR信息检索研究与发展国际会议论文集。1652–1656。[37] Chuhan Wu、Fangzhao Wu、Tao Qi、ChenliangLi 和 Yongfeng Huang。2022。新闻推荐是一项序列推荐任务吗？在第45届ACMSIGIR信息检索研究与发展国际会议论文集。2382–2386。[38] Fangzhao Wu、Ying Qiao、Jiun-Hung Chen、Chuhan Wu、Tao Qi、Jianxun Lian、Danyang Liu、Xing Xie、JianfengGao、Winnie Wu 和 Ming Zhou。2020。MIND：用于新闻推荐的大规模数据集。在计算语言学协会第58届年会论文集。3597–3606。[39] Xu Xie、Fei Sun、Zhaoyang Liu、Shiwen Wu、Jinyang Gao、JiandongZhang、Bolin Ding和 Bin Cui。2022。用于序列推荐的对比学习。在2022年第38届IEEE数据工程国际会议（ICDE）中。IEEE，1259–1273。[40] Chun Yang、Jianxiao Zou、JianHuaWu、Hongbing Xu 和 Shicai Fan。2022。用于推荐的监督式对比学习。Knowledge-Based Systems 258（2022），109973。[41] Yang Yu、Fangzhao Wu、Chuhan Wu、Jingwei Yi和Qi Liu。2022。Tiny-Ne wsRec：有效且高效的基于预训练语言模型的新闻推荐。在2022年自然语言处理实证方法会议论文集。5478–5489。[42] QiZhang、Qinglin Jia、Chuyuan Wang、Jingjie Li、Zhaowei Wang 和Xiuqiang He。2021。AMM：用于新闻推荐的注意力多字段匹配。在第44届ACMSIGIR信息检索研究与发展国际会议论文集。1588–1592。