Simplifying Content-Based Neural News Recommendation: On User Modeling and Training Objectives

Andreea Iana, Goran Glavaš and Heiko Paulheim

摘要

个性化新闻推荐的出现带来了愈加复杂的新闻推荐模型。大多数神经新闻推荐依赖用户点击行为，引入专门的用户编码器，将被点击新闻的内容向量聚合到用户嵌入中，并以标准的逐点分类方式来训练模型。现存工作主要存在两方面的不足：(1) 尽管新闻推荐模型的架构大体相同，但是由于数据集、数据集划分以及评价标准的差异而无法直接、公平地进行比较；(2) 由于推荐模型设计趋于同质化，大量可选择的模型设计与训练方法未得到充分的研究探索。针对上述问题，论文提出了一个统一的新闻推荐模型评价框架，从用户建模的候选感知性、点击行为融合策略和训练目标三个关键维度上对不同模型进行了系统、公平的比较。复现实验结果表明，用候选新闻嵌入与被点击新闻嵌入的点积代替相当规模参数的用户编码器不仅使模型变得简单，而且能够带来较大的性能提升。此外，复现实验结果发现，对比学习能够取代逐点分类成为模型训练的可行选择。

关键词： 神经新闻推荐，用户建模，后期融合，对比学习

1 引言

近年来，基于内容的神经新闻推荐的模型架构变得越来越复杂了，旨在实现根据用户兴趣来定制推荐列表。大多数神经新闻推荐 (Neural News Recommendation, NNR) 模型通常包含一个专门的新闻编码器 (News Encoder, NE) 和用户编码器 (User Encoder, UE)。NE 通常实例化为卷积神经网络、自注意力网络、图注意力网络或者是预训练好的 Transformer 网络，将输入特征，如新闻标题，目录，实体...，转化为新闻嵌入。UE 则通过序列编码器或者是注意力编码器来将被点击新闻嵌入在用户级层面上进行聚合来生成用户嵌入。论文作者将这种主流的模型范式称为早期融合 (Early Fusion, EF)，因为它先将被点击新闻嵌入进行聚合 (即构建用户嵌入)，再将用户嵌入与候选新闻嵌入进行匹配。

大多数 NNR 模型以一种候选不可知 (Candidate-agnostic, C-AG) 的方式单独地进行用户嵌入和候选新闻嵌入的编码。相反，候选感知 (Candidate-aware, C-AW) 模型承认，并不是所有的被点击新闻对于与候选新闻相关性都具备同等的提供有效信息的能力 (例如，一条候选新闻通常只代表用户兴趣的一个子集)，并利用候选新闻嵌入将被点击新闻嵌入上下文化后再由 UE 在用户级层面上进行聚合。最后，候选新闻嵌入 (即 NE 的输出) 与用户嵌入 (即 UE

的输出) 进行比较：候选新闻的推荐分数直接计算为两个向量的点积或者是一个由前馈神经网络组成的计分器的输出。NNR 模型常常通过负采样的标准逐点分类方进行训练。

现存工作主要存在两方面的不足：(1) 尽管推荐模型总体框架设计具有同质性，但是由于缺乏透明度和采用不同的评价标准，不同 NNR 模型间无法进行直接比较。特别地，绝大多数个性化新闻推荐都是在专有数据集上进行性能评估的。即使是使用公开的数据集 (如 addressa 或 MIND [1]) 评估的少数模型，由于不同的数据集划分和评价标准，也无法直接进行比较。(2) 更简单、更直观的推荐系统模型设计选择在很大程度上没有得到充分探索。首先，现有推荐模型常默认采用 EF 架构，并设计出越来越复杂的 UE 部件，然而并无经验或理论来证明 UE 额外的复杂性与系统性能之间的关系。其次，尽管对比学习在相关的检索和推荐任务中被证明非常有效，只有一小部分 NNR 模型使用对比损失进行模型训练。

本工作弥补了 NNR 模型的上述缺点，并对用户建模和训练目标提供了新的想法。具体地，论文首先引入了一个统一的神经新闻推荐模型评价框架，对不同 NNR 模型在三个关键设计维度上进行了系统、公平的比较，三个维度分别为 (i) 用户建模的候选感知性，(ii) 点击行为融合策略，(iii) 训练目标。其次，论文还提出用使用候选新闻嵌入和被点击新闻嵌入的点积的简单池化来取代复杂的用户建模过程 (如 EF)，论文把这种方式归属为后期融合策略 (Lately Fusion, LF)，与 EF 不同，其仅在进行用户——候选新闻匹配时才聚合被点击新闻嵌入，而不是先显式地单独求出用户嵌入，再与候选新闻嵌入进行匹配。尽管直观上看 LF 过于简单，实验表明，相比于 EF，使用 LF 给 NNR 模型带来了实质性的性能提升，使得当前 UE 不合理的结构复杂性进一步缺乏经验理论支撑。最后，实验结果证明，监督对比训练有望取代逐点分类成为可行的模型训练替代方案。本工作为当前基于复杂的用户建模的推荐模型范式引入了更简单、更有效的替代方案，从根本上挑战了 NNR 的研究现状，具有极大的研究意义。

2 相关工作

为了对不同的 NNR 模型进行系统、公平的直接比较，论文着眼于三个关键设计维度:(1) 候选不可知方式进行用户建模与候选感知方式进行用户建模的对比；(2) 用户点击行为早期融合和后期融合策略的对比;(3) 使用标准交叉熵损失函数 (Cross-Entropy Loss, CE) 训练模型与使用监督对比损失函数 (Supervised Contrastive Loss, SCL) 训练模型的对比，并通过在三种维度上的不同组合来构建推荐系统，最后由推荐模型在测试集上的评价指标得分来比较不同推荐模型的性能。下面将仔细介绍所选用的推荐模型并阐述具体的模型设计选择。

2.1 候选不可知 (C-AG) 模型

忽略 C-AG 模型间 NE 的差别，这些模型的 UE 使用被点击新闻嵌入生成用户嵌入的过程中均没有使用候选新闻嵌入进行上下文化。论文采用了以下的 C-AG 模型用于对比评估:(1)NPA [2] 以用户 ID 嵌入作为查询，利用一个注意力模块来聚被点击新闻嵌入得到用户嵌入。(2)NAML [3] 使用 Additive 注意力机制来编码用户偏好；(3)NRMS [4] 通过一个由多头自注意力和 Additive 注意力组成的两层编码器来学习用户嵌入；(4)LSTUR [5] 通过循环神经网络学习用户嵌入：用户的短期兴趣嵌入由被点击新闻嵌入经过门控循环神经网络 (GRU) 来产生，而用户的长期兴趣嵌入则由随机初始化部分和微调部分组成。最终的用户表征可通

过两种方式得到：(i) 用长期兴趣嵌入初始化 GRU 的输入，以短期兴趣 GRU 的最终隐藏状态作为用户表征 ($LSTUR_{ini}$)；(ii) 简单地将短期和长期兴趣嵌入连接起来得到用户表征 ($LSTUR_{con}$)；(5)MINS [6] 则通过组合多头自注意力、基于 GRU 的多通道循环神经网络和 Additive 注意力来进行用户嵌入编码。

2.2 候选感知 (C-AW) 模型

C-AW 模型的 UE 总是学习到相同的用户嵌入，与候选新闻的内容无关。相反，C-AW 模型的 UE 则依赖候选新闻来学习用户嵌入，论文中用于分析的模型为：(1)DKN [7] 计算候选感知的用户表征为用户点击新闻嵌入的加权和，权重由一个注意力网络产生，该网络以候选新闻嵌入和被点击新闻嵌入作为输入；(2)CAUM [8] 组合了一个模拟被点击新闻间的长期依赖关系的自注意力网络与一个从用户相邻的点击新闻中捕捉短期兴趣的候选感知卷积网络，两个网络均以候选新闻的内容为基础；然后由长期兴趣嵌入和短期兴趣嵌入得到最终的候选感知的用户嵌入。

3 本文方法

3.1 本文方法概述

图 1 描绘了一个统一的 NNR 模型评价框架，论文将从三个关键的设计维度入手比较不同 NNR 的模型性能，分别是 (i) 用户建模过程中的候选感知性，见图 1 绿色方框部分；(ii) 点击行为融合策略，见图 1 橙色方框部分；(iii) 模型训练目标，见图 1 紫色方框部分。给定新闻和用户点击行为作为输入，论文通过分析 C-AG 模型和 C-AW 模型在不同点击行为融合策略下分别使用 CE 和 SCL 作为损失函数进行模型训练后的性能来进行实验。

3.2 点击行为融合策略

论文对目前被各种最先进 NNR 模型所广泛采用的早期融合策略 (如 EF) 提出了质疑：NNR 模型中专用的用户编码器是否有存在的必要以及其 UE 结构日益增加的复杂性是否有利于推荐系统性能的提升？为此，论文提出了一种轻量级的替代性方案————后期融合策略 (Lately Fusion, LF)，即用候选新闻嵌入与被点击新闻嵌入点积的平均池化来替代 NNR 模型中复杂的 UE。LF 的数学形式为：给定候选新闻 $\boldsymbol{n}^c$ 和用户点击新闻序列 $H = \boldsymbol{n}_1^u, ..., \boldsymbol{n}_N^u$，候选新闻关于用户的相关度分数计算为 $s(\boldsymbol{n}^c, u) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{n}^c \cdot \boldsymbol{n}_i^u$，其中$\boldsymbol{n}$ 表示由 NE 学习得到的新闻嵌入，N 表示用户点击过的新闻的数量。

值得注意的是，上文的公式与候选新闻嵌入 $\boldsymbol{n}^c$ 和用户被点击新闻嵌入 $\boldsymbol{n}_i^u$ 的平均值的点积是等价的，即 $s(\boldsymbol{n}^c, u) = \boldsymbol{n}^c \cdot (\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{n}_i^u)$。这意味着 LF 也可以根据实际需要生成用户嵌入 (简单地取作被点击新闻嵌入的平均值)。因此，LF 可以被看作是一个无参数的 UE，也就是说，它是现存 NNR 模型中复杂参数化 UE 的一种高计算效率的替代方案。由于 LF 独立生成候选新闻和点击新闻的嵌入，并以被点击新闻嵌入的平均值作为用户嵌入，因此采用 LF 后所有模型均变为候选不可知 (C-AG)。

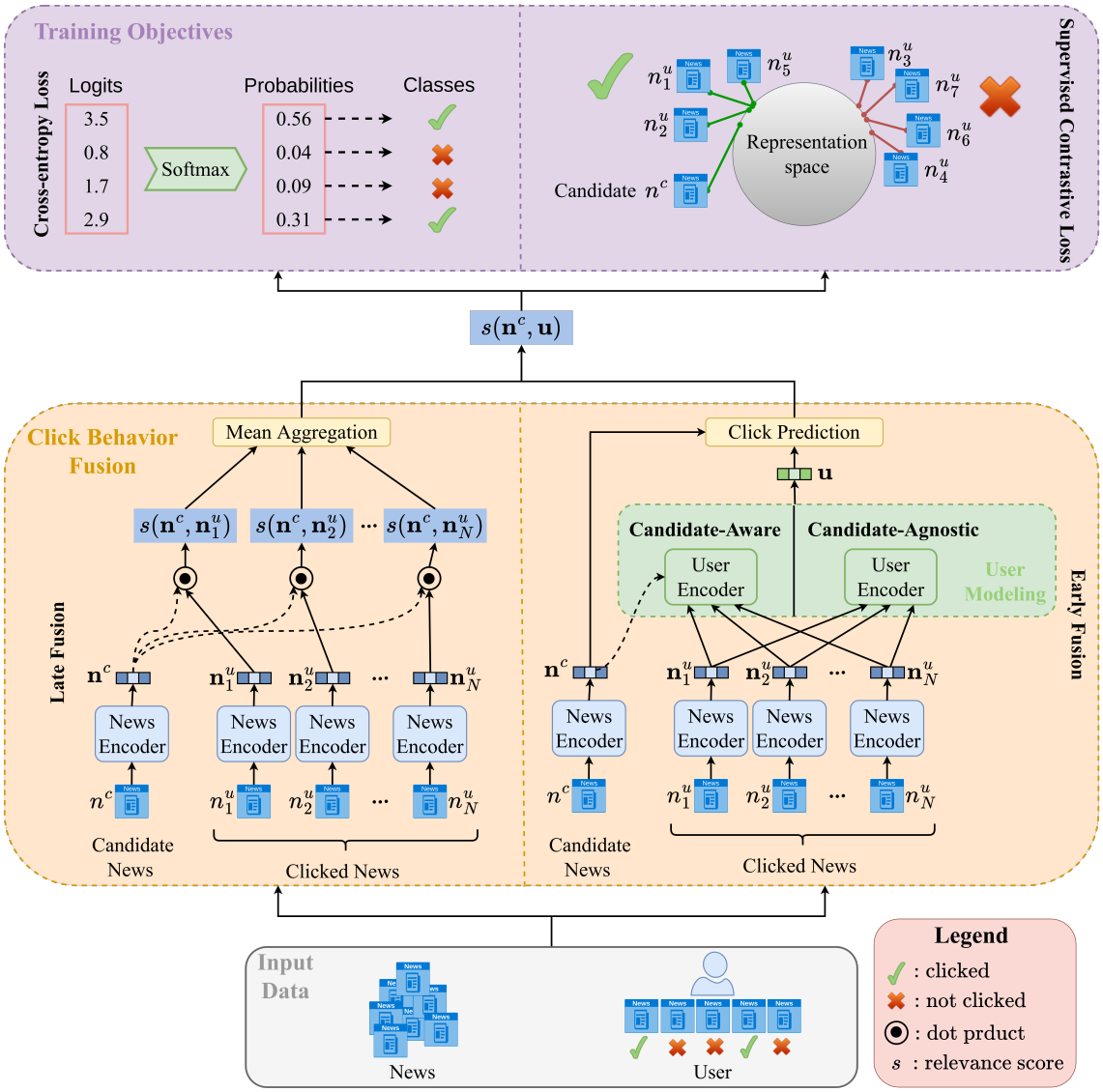

图 1. NNR 统一评价框架示意图

3.3 损失函数定义

尽管对比学习已经被证实在密切相关的检索和推荐任务中训练模型是十分高效的，但是目前绝大多数 NNR 模型都是简单直接地通过最小化交叉熵损失函数 (Cross Entropy Loss, CE) 来调整模型参数，CE 存在一些缺点，例如缺乏鲁棒性、对噪声标签敏感... 为了研究不同训练目标函数对模型有效性的影响，论文不仅通过常见的交叉熵损失函数 (负采样) 来训练所有模型，而且还增加了采用对比学习为训练目标的对照组，具体地以监督对比损失 (Supervised Contrastive Loss, SCL) [9] 为损失函数来训练模型。与自监督对比损失 (Self-supervised Contrastive Loss) 中每个锚点 (Anchor) 匹配一个正样本不同，SCL 为每个锚点匹配多个正样本，正样本与锚点属于同一分类，其数学表达式为

$$L_{out}^{sup} = \sum_{i \in I} L_{out}^{sup} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} log \frac{exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_p / \tau)}{\sum_{a \in A(i)} exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_a / \tau)}$$

其中，$\boldsymbol{z}$ 表示样本的嵌入；$\cdot$ 表示两个向量的内积；τ 是温度系数，为标量；I 表示批量样本，索引 i 代表锚点，P(i) 表示与锚点 i 属于同一分类的所有正样本集合，A(i) 表示与锚

点 i 属于不同分类的所有负样本集合。

与自监督对比损失相比，SCL 为每个锚点匹配多个正样本，使得编码器能够将同属一类的所有样本在特征空间中的距离更为紧密，因此其特征空间下的聚类效果更加健壮。

4 复现细节

4.1 与已有开源代码对比

论文作者开源了 NNR 模型的代码在 https://github.com/andreeaiana/simplifying_nnr 上，因此复现实验工作是基于开源代码来进行的。尽管论文作者公开了源代码，但是具体的模型超参数、随机数种子等并没有给出，为此复现的首要工作是从各种 NNR 模型的对应参考文献 [2] [3] [4] [5] [6] [7] [8] 中找到关于模型超参数的最优值，并依据这些最优超参数来训练模型。此外，NNR 模型各自的模型训练时使用的损失函数都是 CE，而复现工作中增加了使用 SCL 为损失函数的对照，而对于以 SCL 为损失函数的推荐系统，复现实验时使用大小为 0.02 的步伐扫描区间 [0.08,0.3]，根据推荐模型在验证集上的性能来找到 SCL 的最佳温度系数 τ，并以该最优温度系数 τ 来作为推荐模型在测试集上运行的最终参数。

4.2 实验环境搭建

实验所使用的数据集是 Wu 等人 [1] 构建的 MINDsmall 和 MINDlarge 公开数据集，下载地址为 https://msnews.github.io/index.html。MIND 数据集是根据微软新闻自 2019 年 10 月 12 日至 11 月 22 日间，总计 6 周的用户日志样本构建的，其中使用前 5 周用户日志构建训练集和验证集，使用第 6 周用户日志构建测试集。在 MIND 数据集中，新闻标题中的实体被自动提取并链接到 Wikidata。基于从 WikiData 中提取的知识元组，通过 TransE 方法对它们的嵌入进行训练。由于 Wu 等人 [1] 没有发布测试集标签，因此论文使用 MINDlarge 和 MINDsmall 的验证集作为测试集，并将两者的训练集分成不相交的训练集 (前四天的数据) 和验证集 (最后一天的数据)。

实验使用了在 Wikidata 上预训练的 300 维 Glove 嵌入和 100 维 TransE 嵌入分别作为 NNR 模型的输入的单词嵌入和实体嵌入。最大的用户新闻浏览历史记录长度设置为 50。近似训练方式为负采样，并为每个正样本匹配 4 个负样本。所有 C-AG 模型的训练样本批量大小为 512，DKN 为 256，CAUM 为 64(受 GPU 显存限制)。其他所有模型专有的超参数设置为参考文献中的最优值。以混合精度训练所有模型: 在 MINDsmall 上模型训练轮数为 25 次，在 MINDlarge 上模型训练轮数为 10 次；优化算法采用 Adam 算法，学习率设置为 1e-4；采用常见的推荐系统性能指标作为评价标准，分别为:AUC、MRR、nDCG@5 和 nDCG@10。每个模型都是使用一块 40GB 内存的 NVIDIA A100 GPU 进行训练。

5 实验结果分析

表 1 记录了在统一模型评价框架、不同配置 (点击行为融合策略 EF/LF 和训练损失函数 CE/SCL 分别组合) 下 C-AG (NPA, NAML, NRMS, LSTUR 和 MINS) 和 C-AW 模型 (DKN, CAUM) 分别在 MINDsmall 和 MINDlarge 测试集上的推荐性能。接下来，报告将沿着框架的三个设计维度，即用户建模、点击行为融合策略和训练目标，来对复现结果进行分析。

5.1 用户建模的候选感知性

因为在 LF 架构下，所有 NNR 模型变成候选不可知，因此将在 EF 架构下比较 C-AG 和 C-AW 模型。CAUM 具有最复杂的候选感知的 UE 部件，在两种训练目标 (CE 和 SCL) 下评估指标下都优于所有其他模型，尤其是在 MINDlarge 上的 AUC 指标最为突出。这个结果可能会误导人们得出这样的结论: 为了获得更好的推荐推能表现，NNR 模型需要更复杂的、候选感知的用户编码器。然而两个事实否定了这个结论，(1) 首先作为评估中的另一个 C-AW 模型的 DKN 比其他 C-AG 模型表现差劲得多；(2) 其次，存在部分 NNR 模型采用 LF 架构后其性能表现接近甚至超过了 CAUM(EF 架构) 的性能表现。除了 DKN 之外，所有其他模型采用 LF 设计后在 MINDlarge 数据集上训练时都取得了更好的性能。故 NNR 模型中 UE 的复杂性与系统性能并不存在正相关关系。

Dataset		MINDsmall								MINDlarge							
Criterion		AUC		MRR		nDCG@5		nDCG@10		AUC		MRR		nDCG@5		nDCG@10	
Model	CBF	CE	SCL	CE	SCL	CE	SCL	CE	SCL	CE	SCL	CE	SCL	CE	SCL	CE	SCL
NPA	EF	54.32	55.72	28.495	28.077	25.93	26.443	32.46	32.542	57.116	59.246	32.12	33.614	30.219	31.141	36.498	37.248
	LF	55.187	59.055	28.316	30.67	26.07	28.048	32.694	34.157	60.483	60.484	31.706	29.265	29.941	26.687	36.166	33.573
NAML	EF	53.268	55.729	30.393	26.515	28.405	23.885	34.709	30.866	61.71	60.23	32.56	31.44	30.68	29.28	37.08	35.79
	LF	52.845	56.455	28.526	26.713	26.121	24.31	32.705	30.989	62.315	61.398	32.063	31.092	30.282	28.756	36.653	35.25
NRMS	EF	53.084	60.223	26.995	30.157	25.504	28.002	32.103	34.385	56.168	61.216	32.255	33.727	30.7	31.718	37.157	38.303
	LF	58.477	60.75	32.812	31.078	30.802	29.069	36.887	35.241	54.647	63.85	33.446	32.89	31.593	31.11	37.973	37.57
LSTUR	EF_{ini}	58.22	56.82	29.35	27.8	27.29	25.44	33.734	32.1	59.2	62.26	31.24	31.63	29.33	29.82	35.8	36.24
	EF_{con}	52.84	52.67	27.57	26.75	25.32	24.51	31.61	30.79	56.89	55.51	27.69	24.595	25.75	21.761	32.47	29.16
	LF	55.49	59.25	28.99	27.69	26.78	25.5	33.37	32.22	62.12	61.8	31.95	31.5	30.23	29.16	36.58	35.68
MINS	EF	57.718	57.973	31.643	28.454	29.531	26.542	35.827	32.817	55.65	63.015	32.963	31.657	31.14	29.79	37.618	36.34
	LF	58.477	60.75	32.812	31.078	30.802	29.069	36.887	35.241	60.553	63.266	33.354	32.214	31.495	30.446	37.964	36.882
DKN	EF	50	54.584	25.847	25.629	23.71	22.989	30.325	30.054	50	58.382	25.73	25.03	23.882	23.188	30.857	30.489
	LF	50	51.339	27.952	28.18	25.725	25.88	32.376	32.732	50	59.415	30.064	28.744	26.872	26.266	33.857	33.1
CAUM	EF	62.23	63.799	33.73	34.2	32.06	32.531	38.55	38.703	67.424	66.333	35.947	35.322	34.259	33.727	40.686	40.0054
	LF	62.61	64.093	35.765	34.188	33.923	32.368	40.039	38.458	59.253	66.196	34.919	34.486	33.245	32.907	39.653	39.328

表 1. 不同点击行为融合策略 (CBF) 与训练目标 (CE/SCL) 组合下 NNR 的推荐性能

5.2 点击行为融合策略

采用 LF 架构来替代包含复杂用户编码器的 EF 架构可以提高推荐模型的性能表现。对所有模型在两种训练目标下性能表现得分取平均值，LF 系统在 MINDsmall 和 MINDlarge 上的 MRR 分别得到了 4.66% 和 1.76% 的提升，对于其他指标如 AUC、nDCG 也均有超过 1% 的提升。此外，相比于 EF，使用 LF（相当于相同的无参数 UE）后不同 NNR 模型间的性能更加接近，同时其他 NNR 模型与 CAUM 的性能差距也变得减小。这表明 LF 使得不同 NNR 模型间的 NE 结构差异对推荐性能的影响变小，因此 LF 不仅通过对被点击新闻嵌入的平均池化来实现 UE 的简化，而且揭示了 NE 存在简化设计的可能性。

5.3 模型训练损失函数

监督对比损失是可行的模型训练替代方法，可以取代基于负采样的交叉熵损失函数来训练模型，理由如下：由表 1 可以得出，使用 SCL 训练模型后，推荐系统的 AUC 指标得分提高 (对所有 EF 和 LF 下系统的 AUC 取平均值)，在 MINDsmall 和 MINDlarge 上分别得到了 4.13%、5.62% 的提升。这表明，SCL 在特征空间中可以更好地对被点击和未被点击的新闻进行分类。相比之下，在 MRR 和 nDCG 指标上, 使用 SCL 训练得到的推荐系统性能得分则落

后于使用 CE 训练模型的推荐系统性能得分 (MRR 在 MINDsmall 和 MINDlarge 上分别下降了 2.68% 和 2.26%)。初步假设是因为 CE 能更直接区分一些强负样本 (Hard Negatives)——用户没有点击，但与被点击新闻相似的新闻样本，而 SCL 则会令所有相似的新闻样本在特征空间的距离彼此更加接近，因此这些强负样本可能会成为推荐列表中排名靠前的假阳样本。

5.4 模型规模

根据上文结论可知，替换掉主流的 EF 架构设计，采用更加简单的 LF 设计后并不会导致推荐系统性能的下降，相反这样做不仅带来性能上的提升，更重要的是由于 LF 无没有额外的复杂用户编码器，它能够降低模型的规模。图 2 展示了 NNR 模型在原始 EF 架构下可训练参数量的统计情况。可以看出，对于大多数 NNR 模型而言，NE 的可训练参数规模在整个模型的参数组成中占据主要部分，唯有 LSTUR 模型不同，其 UE 的可训练参数量远远大于 NE 的参数量。由于 LF 架构可以看作使用了零参数的 UE 来进行用户嵌入编码，由图 2 可知，LF 架构分别使 LSTUR_{ini}、LSTUR_{con}、CAUM 模型的规模减少了 93.8%, 88.5%, 14.3%。

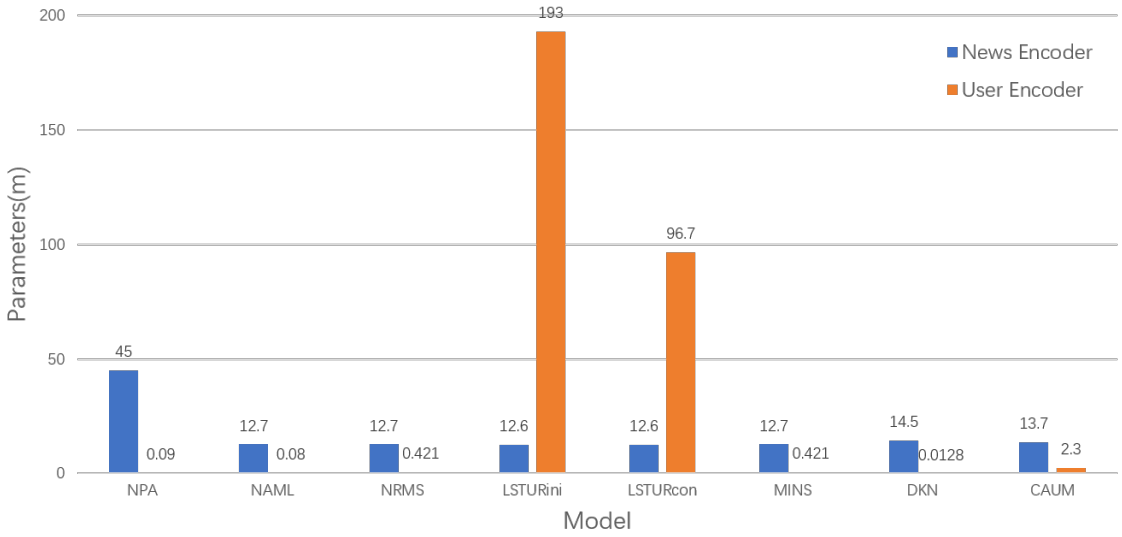

图 2. NNR 模型原始 EF 架构下的参数规模 (单位: 百万)

6 总结与展望

个性化神经新闻推荐的快速发展带来了越来越复杂的推荐模型，这阻碍了不同模型间进行公平的评估和结构设计的系统性分析。回顾整个论文复现工作过程，通过建立了统一的神经新闻模型框架，从三个关键设计维度:(i) 用户建模过程的候选感知性；(ii) 用户点击行为融合策略，以及 (iii) 训练目标，来对 7 种 NNR 模型进行了直接、公平的比较。复现实验结果表明，用被点击新闻嵌入的无参数聚合取代 NNR 模型种复杂的用户编码器不仅可以提高推荐系统性能，还可以降低模型规模与复杂性，这对未来推荐模型的设计具有重要意义，颠覆了“更复杂的推荐模型往往具备更高的性能”这一缺乏佐证的观念。这一发现将激发未来 (i) 关于如何对神经新闻推荐模型进行更加透明、公平地评估的研究，以及 (ii) 揭示驱动推荐性能提升的组件的推荐模型消融实验的研究。此外，监督对比损失有望在后续的神经新闻推荐模型研究中成为更加高效的模型训练方法。

参考文献

[1] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian andDanyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, and Ming Zhou. Mind: A large-scale dataset for news recommendation. In *Proc. Association for Computational Linguistics*, page 3597–3606, 2020.

[2] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. Npa: Neural news recommendation with personalized attention. In *Proc. ACM Conf. on Knowledge Discovery and Data Mining*, pages 2576–2584, 2019.

[3] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. Neural news recommendation with attentive multi-view learning. In *Proc. Int. Conf. on Artificial Intelligence*, page 3863–3869, 2019.

[4] Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. Neural news recommendation with multi-head self-attention. In *Proc.f Conf. on Empirical Methods in Natural Language Processing and Int. Conf. on Natural Language Processing*, page 6389–6394, 2019.

[5] Mingxiao An, Fangzhao Wu, Chuhan Wu, Kun Zhang, Zheng Liu, and Xing Xie. Neural news recommendation with long- and short-term user representations. In *Proc. Association for Computational Linguistics*, page 336–345, 2019.

[6] Rongyao Wang, Shoujin Wang, Wenpeng Lu, and Xueping Peng. Recommendation via multi-interest news sequence modelling. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, page 7942–7946, 2022.

[7] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. Dkn: Deep knowledge-aware network for news recommendation. In *Proc. on World Wide Web Conference*, page 1835–1844, 2018.

[8] Tao Qi, Fangzhao Wu, Chuhan Wu, and Yongfeng Huang. News recommendation with candidate-aware user modeling. In *Proc. ACM Int. Conf. on Research and Development in Information Retrieval*, page 1917–1921, 2022.

[9] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, and and Dilip Krishnan Ce Liu. Supervised contrastive learning. In *Proc. Int. Conf. on Neural Information Processing Systems*, page 18661–18673, 2020.