课程学习引导的心理健康线上帖子抽象式摘要生成

Sajad Sotudeh [1]*, **Nazli Goharian**[1], **Hanieh Deilamsalehy**[2], 和**Franck Dernoncourt**[2]

[1]IRLab，乔治城大学
{sajad,nazli}@ir.cs.georgetown.edu[2]Adobe Research{deilamsa,franck.dernoncourt}@adobe.com

arXiv:2302.00954v1 [cs.CL] 2 Feb 2023

摘要

自动生成用户心理健康帖子摘要，有助于节省咨询师的阅读时间并减轻其疲劳，从而使他们能够及时回应那些寻求帮助、希望改善心理状态的人。近年来，基于 Transformer 的摘要模型为抽象式摘要生成提供了一种颇具前景的方法。它们超越了句子选择和抽取式策略，能够处理词语创新生成和句子改写等更复杂的任务。然而，这些模型存在一个显著缺陷：其训练策略效率不高，限制了模型的性能。本文采用课程学习方法对训练样本重新加权，从而实现高效的学习过程。我们将模型应用于 MENTSUM 心理健康帖子数据集中的极限摘要任务；该数据集收录了来自 Reddit 社交媒体的心理健康相关帖子。与最先进的模型相比，我们提出的方法在 ROUGE 和 BERTSCORE 评估指标上取得了显著提升，相对提升幅度分别为 3.5%（ROUGE-1）、10.4%（ROUGE-2）、4.7%（ROUGE-L）和 1.5%（BERTSCORE）。

1 引言

心理健康网络帖子的摘要生成是一项新兴任务，旨在为用户在 Reddit [1] 和 Reachout [2]等在线社交网络中发布的求助帖子生成摘要，以帮助改善其心理状态。帖子可能涉及用户关注的多个问题，也可能只是详细描述用户的心理和情绪状况。根据用户的偏好，每篇用户撰写的帖子都可以附上一段简洁的摘要（称为太长不看摘要 [3]），用于概括主要内容。这一太长不看摘要旨在让咨询师能够先快速阅读用户发布的内容，再通读整篇帖子；因此，咨询师可以及时作出回应。在此，我们旨在提升该任务中 (Sotudeh、Goharian 和 Young，2022) 所报告的最先进结果。

大规模深度神经模型通常难以训练，依赖复杂的启发式设置；这些设置的调优可能既耗时又昂贵（Gong 等人，2019年；Chen 等人，2021）。对于基于 Transformer 的摘要生成模型而言，这一点尤为明显。研究表明，在经过严格调优后，它们的表现始终优于循环神经网络（Popel 和 Bojar，2018年），但也需要采用专门的学习率和大批量训练等启发式方法（Platanios 等人，2019年）。本文尝试通过为 BART (Lewis 等人，2020年) Transformer 摘要生成模型引入一种课程学习（*CL*）策略（Bengio 等人，2009年），以克服上述问题，从而改善摘要模型的收敛时间和性能。

受人类教学方式的启发，课程学习提出教学过程应从较简单的样本逐步过渡到较难的样本，其历史可追溯至20世纪90年代（Elman，1993）。这一方法的核心思想是，按照特定且明确的顺序将训练样本输入网络，从较简单的样本逐步过渡到较难的样本，可以帮助网络更好地学习任务（Chang 等人，2021）。在神经网络领域，这一过程可视为一种技术，能够增强网络的鲁棒性，避免其陷入局部最优点；这种情况在训练初期更容易发生。鉴于摘要网络面临上述挑战，我们采用 SUPERLOSS (Castells 等人，2020年）函数，该函数属于置信度感知课程学习技术，并向网络引入一个名为置信度的新参数（即 σ）。

* 部分工作是在 Adobe Research 实习期间完成的。
[1]https://www.reddit.com/[2]https://au.reachout.com/[3]太长不看摘要（TL;DR）是“Too Long, Didn’t Read”（太长，没读）的缩写。本文中，我们交替使用“TL;DR”和“摘要”这两个说法。

我们在 MENTSUM (Sotudeh、Goharian 和 Young，2022）数据集上验证了我们的模型。该数据集包含从 Reddit 社交媒体上 43 个与心理健康相关的社群中收集的超过 2.4 万个样本。实验结果表明，在 BART 摘要生成模型上应用课程学习目标效果显著，达到了新的最先进性能。

2 相关工作

尽管心理健康研究中的大多数工作都侧重于通过分类和预测任务研究用户的行为模式（Choudhury 等人，2013年；Resnik 等人，2013年；Coppersmith 等人，2014年；Yates 等人，2017年；Cohan 等人，2017年、2018年；MacAvaney 等人，2018年），但近期，线上心理健康帖子的摘要生成已成为可行方向。Sotudeh、Goharian 和 Young（2022）率先朝这一方向迈进，推出了线上心理健康帖子 MENTSUM 数据集，并给出了基线结果。过去十年间，课程学习（CL）逐渐受到研究界的关注（Tay 等人，2019年；MacAvaney 等人，2020年；Xu 等人，2020年）。Bengio 等人（2015年）率先将这一策略应用于序列预测，通过计划采样方法，在解码过程中逐步将训练过程从使用真实标注词元转变为使用模型生成的词元。在这一方案中，样本难度是一个关键概念，用于区分容易样本和困难样本。研究人员采用了多种文本特征作为“难度度量”，包括 n 元语法频率（Haffari 等人，2009年）、词语稀有度和句子长度（Platanios 等人，2019年）。近期的研究（Saxena 等人，2019年；Cachola 等人，2020年）采用了置信度感知方法，学习训练样本的难度，并在训练过程中动态调整样本权重。

3 我们的方法

在本节中，我们将详细介绍所提出模型，其中在基于 BART 的 Transformer 框架中加入了课程学习架构，提高较容易训练样本的权重；因此，在学习阶段增加了这些样本的贡献。

面向 B 的课程学习器ART。 考虑到课程学习在训练大规模网络中的适用性，我们旨在将其应用于摘要框架。在将课程学习策略纳入模型的训练阶

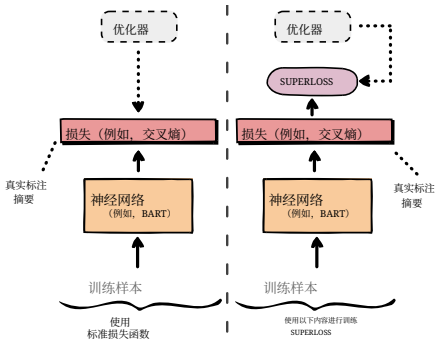

图 1：使用标准损失函数进行训练（左侧）和使用 SuperLoss准则进行训练（右侧）

段之前，我们首先需要定义难度指标，以区分样本的难易程度。在实践中，为每个样本估计先验难度是一项复杂的任务，因此我们提出在训练过程中利用渐进式信号来区分样本——例如每次训练迭代时相应样本的损失。在此背景下，课程学习通过在各次训练迭代中以权重的形式预测每个样本的难度来实现，使困难样本在训练早期获得较低权重，反之亦然。为建模课程学习，我们提出使用 SUPERLOSS (Castells 等人，2020年），这是一种基于任务损失函数构建的通用损失准则，如图 1所示。

更具体地说，SUPERLOSS 是一种与任务无关且感知置信度的损失函数，其输入包括两个参数：(1) 任务损失 $\mathcal{L}_i = \ell(y_i,\ \widehat{y_i})$，其中 y_i 是神经网络（即 BART 生成的摘要）的输出，$\widehat{y_i}$ 是金标准标签（即真实摘要）；以及 (2) σ_i，作为第 i 个样本的置信度参数。SUPERLOSS 被表示为 $\mathrm{L}_\lambda(\mathcal{L}_i,\ \sigma_i)$，其计算方式如下：

$$\mathrm{L}_\lambda(\mathcal{L}_i, \sigma_i) = (\mathcal{L}_i - \tau)\sigma_i + \lambda(\log \sigma_i)^2 \quad (1)$$

其中，λ 是正则化参数，τ 是训练过程中任务损失（即 $\mathcal{L}$）的运行平均值或静态平均值。尽管 SUPERLOSS 为课程学习策略提供了一种定义明确的方法，但对于文本摘要等训练样本丰富的任务，学习 σ 参数并不可行。为了解决这一问题并避免引入新的可学习参数，SUPERLOSS 建议使用 σ_i 在极限情况下的收敛值，

$$\sigma_\lambda^*(\ell_i) = \arg\min_{\sigma_i} \mathrm{L}_\lambda(\ell_i, \sigma_i)$$
$$\mathrm{SL}_\lambda(\ell_i) = \mathrm{L}_\lambda(\ell_i, \sigma_\lambda^*(\ell_i, \sigma_i)) = \min_{\sigma_i} \mathrm{L}_\lambda(\ell_i, \sigma_i), \quad (2)$$

采用这种技术，无需在训练过程中学习置信度参数。Castells 等人(2020年) 发现 $\sigma_\lambda^*(\ell_i)$ 存在闭式解，计算如下：

$$\sigma_\lambda^*(\ell_i) = e^{-W(\frac{1}{2}\max(-\frac{2}{e}, \beta))}, \beta = \frac{\ell_i - \tau}{\lambda} \quad (3)$$

其中， W 是 Lambert W 函数。有鉴于此，SUPERLOSS 会在训练过程中动态提高较简单样本的权重，为摘要生成提供一种课程学习方法。在实验中，我们将该模型称为 CURRSUM。

4 实验设置

4.1 数据集

我们在实验中使用 MENTSUM 数据集。该数据集包含超过 2.4 万组帖子TL；DR 配对，分为 21,695 个训练样本、1209 个验证样本和 1215 个测试样本。这些数据通过在 Reddit 上依据严格的筛选规则，从 43 个心理健康子版块社区中挖掘而得。关于该数据集的更多详情，请读者参阅论文正文（Sotudeh、Goharian 和 Young，2022）。

4.2 对比

我们将模型与 BART（Lewis 等人，2020年）基线进行比较，该基线未使用课程学习目标。BART 是一种抽象式模型，采用预训练编码器-解码器架构，而 BERT 仅使用预训练编码器。如（Sotudeh、Goharian 和 Young，2022）所示，BART 是最强的基线；因此，我们在其上应用课程学习，以评估其对摘要生成的影响。有关更多抽取式和抽象式基线，请参阅原论文。

4.3 实现细节

我们使用 Huggingface 的 Transformers 库（Wolf 等人，2020年）[4] 来实现我们的模型。所有模型均训练 8 个周期，每隔 0.5 个周期进行一次评估，并使用 ROUGE-L 得分最高的检查点

[4]https://github.com/huggingface/transformers

模型	R-1	R-2	R-L	BS
ORACLEEXT	35.98	11.59	23.21	82.72
BART (2020年)	29.13	7.98	20.27	85.01
CURRSUM（我们的方法）	**30.16**	**8.82**	**21.24**	**86.32**

表 1：MENTSUM 数据集测试集上的 ROUGE 和 BERTSCORE 结果。由于 BART 是（Sotudeh、Goharian 和 Young，2022）中表现最佳的基线模型，因此本文评估课程学习在 BART 上的有效性。其他基线模型请参阅论文正文。

用于推理的验证集得分。使用学习率为 $3e-5$、$(\beta_1$、$\beta_2)$ $=(0。9$、$0。98)$的 AdamW优化器（Loshchilov 和 Hutter，2019年）以及 0.01 的权重衰减，用于我们所有的摘要模型以及 BART。所有模型均使用交叉熵损失。为跟踪学习过程，我们使用 Weights & Bi- ases（Biewald，2020年）工具包 [5]。

5 伦理与隐私

虽然我们使用公开可用的 Reddit 数据来训练和验证我们的模型，但我们认识到，由于此类数据具有敏感性以及用户隐私问题，在处理这类数据时应格外谨慎（Benton 等，2017年；Nicholas 等，2020年）。因此，我们没有尝试识别或联系用户，也没有探究用户关系来追踪其社交媒体账户。需要指出的是，MENTSUM 数据集通过数据使用协议（DUA） [6] document 进行分发，以进一步确保用户身份永不被披露。

6 结果

自动评估。表1 报告了基线模型和我们的模型在 MENTSUM 数据集上的表现，并采用 ROUGE分数变体（Lin，2004）和 BERTSCORE（Zhang 等人，2020年）进行评估。如表所示，表现最佳的是我们的 CURRSUM，它直接在 BART 模型基础上采用 SUPERLOSS课程学习；与 BART 相比，该模型在所有指标上均有明显提升，达到了当前最先进的性能。具体而言，与不采用课程学习的基准模型（即 BART）相比，我们的模型在 ROUGE-1、ROUGE-2、ROUGE-L 和 BERTSCORE 上分别提升了 3.5%、10.4%、4.7% 和 1.5%。查看 ORACLEEXT 的分数后可以发现，

[5]https://wandb.ai/[6] https://ir.cs.georgetown.edu/resources/mentsum.html

人工撰写	BART生成	CURRSUM生成
……生孩子……我的注意缺陷多动障碍……对为人父母的看法——关于为人父母的看法……	我认为，与其现在不生孩子，不如永远不生生孩子这件事。	我在考虑要个孩子，而且我不想成为一个糟糕的爸爸，因为我有注意缺陷多动障碍。有没有家长有注意缺陷多动障碍相关的经历?
……抑郁症状……去寻求帮助……	我觉得自己患有临床抑郁，而且我几乎所有抑郁症状都有，所以我在寻求帮-助	我觉得自己患有临床抑郁*xxx*。我几乎所有都有这份清单上的症状。
……十年来一直没有焦虑……症-症状……	我已经 摆脱焦虑还不到一周 而且我担心这只是暂时的解决办法。 我不知道该怎么办。	我已经摆脱焦虑十年了而且我担心如果症状开始复发，我会能够再次陷入其中。
…约她出去……我们几乎没怎么聊过……把我的机会搞砸……	我喜欢上了工作中的一个女孩，想约她出去，但我不想在这个过程中毁掉这个机会。	我喜欢上了工作中的一个女孩，但我不知道该怎么开口她出去。

表2：人工撰写的、BART生成的以及 CURRSUM生成的 TL；DR 摘要。人工撰写的样本仅部分展示，以保护用户隐私。也就是说，我们只展示了重要的人工撰写短语，以便在生成的摘要中追踪这些短语。与帖子内容不符（即用户帖子不支持）的文本以灰色显示，而摘要生成系统提取的显著信息则以**粗体**显示。

理想抽取式摘要器所能达到的性能上限，似乎表明抽取式设置仍有提升空间，以实现最先进的性能。研究者可以投入更多精力，开发更复杂的抽取式或混合式摘要模型，例如（Gehrmann 等人，2018年；MacAvaney 等人，2019年；Sotudeh 等人，2020年）。

案例研究与分析。 尽管我们提出的模型显著优于 BART 基线模型，但我们也认识到 ROUGE 指标在评估摘要生成系统时存在局限性（Cohan 和 Goharian，2016）。为了探究我们工作的优点和局限性，我们分析了人工撰写的太长不看摘要，以及 BART 和我们模型生成的结果，并对它们进行比较。表 2 展示了四个人工撰写和系统生成的太长不看摘要样例。如图所示，在第一、第二和第三个样例中，我们的模型能够提高摘要的忠实度 [7]。在考察研究中的其他案例后，我们发现，课程学习似乎能够有效减少忠实度方面的错误。这可能是因为在学习过程的每次迭代中，通过控制各样本的贡献，摘要生成器能够更好地理解 源文档。观察第二个样例可以发现，我们的模型能够提高摘要的简洁性；也就是说，能够在摘要中简明扼要地传达要点。比较系统生成的摘要时

[7]忠实性定义为生成有用户帖子内容支持的文本。

在第四个样例中，可以看到我们的系统通过混合用户帖子中的不同部分生成了一个短语（以灰色显示）。令人惊讶的是，“我不知道该怎么做/该说什么”似乎是大多数寻求社区建议的人工撰写摘要中常用的短语。实验所用的摘要生成系统（即 BART 和我们的系统）都倾向于过度生成这一短语。

7 结论

根据社区用户的在线帖子生成简短摘要，可以节省咨询师的阅读时间并减轻他们的疲劳。在此基础上，他们能够更快地回应社区用户。尽管基于神经 Transformer 的摘要模型前景可期，但其训练过程效率低下，阻碍了模型发挥潜力并取得良好性能。为解决这一问题，本文将一种置信度感知课程学习方法整合到摘要框架中，该方法使用与任务无关的 SUPERLOSS，旨在提升模型的泛化能力，最终改善模型性能。我们在心理健康帖子 MENTSUM 数据集上进行的自动评估表明，该模型效果显著：与 BART 摘要器相比，在 ROUGE-1、ROUGE-2、ROUGE-L 和 BERTSCORE 上分别取得了 3.5%、10.4%、4.7% 和 1.5% 的相对提升。我们的模型在 MENTSUM 数据集上创下了新的最先进性能。我们还展示了多个系统生成的摘要，以呈现所提模型的优点与局限。

参考文献

Samy Bengio、Oriol Vinyals、Navdeep Jaitly 和 Noam Shazeer。2015年。利用循环神经网络进行序列预测的计划采样。载于《神经信息处理系统进展》第*28*卷：*2015*年神经信息处理系统年度会议，*2015*年*12*月*7*日至*12*日，加拿大魁北克省蒙特利尔，第1171–1179页。

Yoshua Bengio、Jérôme Louradour、Ronan Collobert 和 Jason Weston。2009年。课程学习。载于第*26*届国际机器学习年度会议论文集，*ICML 2009*，蒙特利尔，魁北克，加拿大，*2009*年*6*月*14*日至*18*日，第382卷*ACM*国际会议论文集系列，第41–48页。ACM。

Adrian Benton、Glen A. Coppersmith 和 Mark Dredze。2017年。社交媒体健康研究的伦理研究规范。载于*EthNLP@EACL*。

Lukas Biewald。2020年。使用 weights and biases 进行实验跟踪。软件可从 wandb.com 获取。

Isabel Cachola、Kyle Lo、Arman Cohan 和 Daniel Weld。2020年。TLDR：科学文档的极限摘要生成。载于计算语言学协会论文集：*EMNLP 2020* 精选论文，第4766–4777页，线上。计算语言学协会。

Thibault Castells、Philippe Weinzaepfel 和 Jérôme Revaud。2020年。SUPERLOSS：一种用于稳健课程学习的通用损失函数。载于《神经信息处理系统进展》第*33*卷：神经信息处理系统年度会议*2020*，*NeurIPS 2020*，*2020*年*12*月*6*日至*12*日，线上。Ernie Chang、Hui-Syuan Yeh 和 Vera Demberg。2021年。训练样本的顺序重要吗？使用课程学习改进神经数据到文本生成。载于《计算语言学协会欧洲分会第*16*届会议论文集：主会场论文集》，第727–733页，线上。计算语言学协会。Xiaohan Chen、Yu Cheng、Shuohang Wang、Zhe Gan、Zhangyang Wang 和 Jingjing Liu。2021年。Early-BERT：通过早鸟彩票券实现高效 BERT 训练。载于《计算语言学协会第*59*届年会暨自然语言处理第*11*届国际联合会议论文集》（第*1*卷：长篇论文），第2195–2207页，线上。计算语言学协会。Munmun De Choudhury、Michael Gamon、Scott Counts 和 Eric Horvitz。2013年。通过社交媒体预测抑郁症。载于*ICWSM*。Arman Cohan、Bart Desmet、Andrew Yates、Luca Soldaini、Sean MacAvaney 和 Nazli Goharian。2018年。Smhd：用于探索多种心理健康状况在线语言使用的大规模资源。载于*COLING*。

Arman Cohan 和 Nazli Goharian。2016年。重新审视科学文章的摘要评估。载于《第十届国际语言资源与评估会议论文集（*LREC'16*）》，第806–813页，波多罗日，斯洛文尼亚。欧洲语言资源协会（ELRA）。

Arman Cohan、Sydney Young、Andrew Yates 和 Na-zli Goharian。2017年。在线心理健康论坛中的内容严重程度分诊。信息科学与技术协会期刊，68。

Glen A. Coppersmith、Craig Harman 和 Mark Dredze。2014年。推特上的创伤后应激障碍测量。收录于*ICWSM*。

J. Elman。1993年。神经网络中的学习与发展：从小处着手的重要性。认知，48：71–99。

Sebastian Gehrmann、Yuntian Deng 和 Alexander Rush。2018年。自底向上抽象式摘要。收录于*2018*年自然语言处理实证方法会议论文集，第4098–4109页，比利时布鲁塞尔。计算语言学协会。

Linyuan Gong、Di He、Zhuohan Li、Tao Qin、Liwei Wang 和 Tie-Yan Liu。2019年。通过逐步堆叠高效训练 BERT。载于第*36*届国际机器学习会议（*ICML 2019*），*2019*年*6*月*9*日至*15*日，美国加利福尼亚州长滩，第97卷《机器学习研究论文集》，第2337–2346页。PMLR。Gholamreza Haffari、Maxim Roy 和 Anoop Sarkar。2009年。用于统计式短语机器翻译的主动学习。载于《人类语言技术：北美计算语言学协会*2009*年年会论文集》，第415–423页，美国科罗拉多州博尔德。计算语言学协会。Mike Lewis、Yinhan Liu、Naman Goyal、Marjan Ghazvininejad、Abdelrahman Mohamed、Omer Levy、Veselin Stoyanov 和 Luke Zettlemoyer。2020年。BART：面向自然语言生成、翻译和理解的去噪序列到序列预训练。载于《计算语言学协会第*58*届年会论文集》，第7871–7880页，线上。计算语言学协会。Chin-Yew Lin。2004年。ROUGE：自动评估摘要的工具包。载于《文本摘要的新进展》，第74–81页，西班牙巴塞罗那。计算语言学协会。Ilya Loshchilov 和 Frank Hutter。2019年。解耦权重衰减正则化。载于第七届国际学习表征会议（*ICLR 2019*），美国路易斯安那州新奥尔良，*2019*年*5*月*6*日至*9*日。OpenReview.net。

Sean MacAvaney、BartDesmet、ArmanCohan、Luca Soldaini、Andrew Yates、Ayah Zirikly 和 Nazli Goharian。2018年。Rsdd-time：自我报告的心理健康诊断的时间标注。载于*CLPsych@NAACL-HTL*。

Sean MacAvaney、Franco Maria Nardini、Raffaele Perego、Nicola Tonellotto、Nazli Goharian 和 Ophir Frieder。2020年。开放域问答重排序的训练课程。载于 第*43*届国际 *ACM SIGIR* 信息检索研究与发展会议论文集，*SIGIR 2020*，线上会议，中国，*2020*年*7*月*25*日至*30*日，第529–538页。ACM。

Sean MacAvaney、Sajad Sotudeh、Arman Cohan、Nazli Goharian、Ish A. Talati 和 Ross W. Filice。2019年。本体感知临床抽象式摘要。载于 第*42*届国际 *ACM SIGIR* 信息检索研究与发展会议论文集，*SIGIR 2019*，法国巴黎，*2019*年*7*月*21*日至*25*日，第1013–1016页。ACM。

Jennifer Nicholas、Sandersan Onie 和 Mark Erik Larsen。2020年。心理健康社交媒体研究中的伦理与隐私。*CurrentPsychiatry Reports*，22：1–7。

Emmanouil Antonios Platanios、Otilia Stretcu、Graham Neubig、Barnabas Poczos 和 Tom Mitchell。2019年。用于神经机器翻译的基于能力的课程学习。载于计算语言学协会北美分会 *2019* 年会议论文集：人类语言技术，第 *1* 卷（长篇与短篇论文），第 1162–1172 页，明尼阿波利斯，明尼苏达州。计算语言学协会。

Martin Popel 和 Ondrej Bojar。2018年。Transformer 模型的训练技巧。布拉格数学语言学公报，110：43–70。

Philip Resnik、Anderson Garron 和 Rebecca Resnik。2013年。使用主题建模改进对大学生神经质和抑郁症的预测。载于*EMNLP*。

Shreyas Saxena、Oncel Tuzel 和 Dennis DeCoste。2019年。数据参数：用于学习可微课程的一类新参数。载于神经信息处理系统进展 *32*：*2019* 年神经信息处理系统年会，*NeurIPS 2019*，*2019*年*12*月*8*日至*14*日，温哥华，不列颠哥伦比亚省，加拿大，第 11093–11103 页。

Sajad Sotudeh、Nazli Goharian 和 Ross Filice。2019年。关注医学本体：临床抽象式摘要的内容选择。载于第*58*届计算语言学协会年会论文集，第1899–1905页，线上。计算语言学协会。

Sajad Sotudeh、Nazli Goharian 和 Zachary Young。2022 年。MENTSUM：用于探索心理健康摘要的资源——心理健康网络帖子的摘要。载于语言资源与评估会议论文集，第 2682–2692 页，法国马赛。欧洲语言资源协会。

Yi Tay、Shuohang Wang、Anh Tuan Luu、Jie Fu、Minh C. Phan、Xingdi Yuan、Jinfeng Rao、Siu Che-ung Hui 和 Aston Zhang。2019年。用于长篇叙事文本阅读理解的简单高效课程指针生成网络。载于第*57*届计算语言学协会年会论文集，第4922–4931页，意大利佛罗伦萨。计算语言学协会。

Thomas Wolf、Lysandre Debut、Victor Sanh、Julien Chaumond、Clement Delangue、Anthony Moi、Pierric Cistac、Tim Rault、Remi Louf、Morgan Funtowicz、Joe Davison、Sam Shleifer、Patrick von Platen、Clara Ma、Yacine Jernite、Julien Plu、Canwen Xu、Teven Le Scao、Sylvain Gugger、Mariama Drame、Quentin Lhoest 和 Alexander Rush。2019年。Transformers：自然语言处理领域的先进技术。载于*2020*年自然语言处理实证方法会议：系统演示论文集，第38–45页，线上。计算语言学协会。

Benfeng Xu、Licheng Zhang、Zhendong Mao、Quan Wang、Hongtao Xie 和 Yongdong Zhang。2020年。面向自然语言理解的课程学习。载于计算语言学协会第 *58* 届年会论文集，第 6095–6104 页，线上。计算语言学协会。

Andrew Yates、Arman Cohan 和 Nazli Goharian。2017年。在线论坛中的抑郁症与自残风险评估。载于*EMNLP*。

Tianyi Zhang、Varsha Kishore、Felix Wu、Kilian Q. Weinberger 和 Yoav Artzi。2020年。BERTSCORE：使用 BERT 评估文本生成。载于第八届国际学习表征会议（*ICLR 2020*），亚的斯亚贝巴，埃塞俄比亚，*2020*年*4*月*26*日至*30*日。OpenReview.net。