提升大语言模型在多文档摘要中的公平性

Haoyuan Li[1], Rui Zhang[2], Snigdha Chaturvedi[1]
[1]University of North Carolina at Chapel Hill, [2]Pennsylvania State University
{ haoyuanl, snigdha } @cs.unc.edu, { rmz5227 } @psu.edu

Abstract

多文档摘要的公平性对于提供具有不同社会属性值的文档的全面观点至关重要，这可能对决策产生重大影响。例如，一个倾向于过度代表产品负面评论的摘要系统可能会误导客户忽略好的产品。先前的工作在两个层面测量多文档摘要的公平性：摘要级和语料库级。虽然摘要级公平性关注于个体摘要，语料库级公平性关注于摘要语料库。近期的方法主要集中在摘要级公平性。我们提出了 FairPO，一种关注多文档摘要中摘要级和语料库级公平性的偏好调优方法。为了提高摘要级公平性，我们建议通过扰动文档集生成偏好对。为了提高语料库级公平性，我们建议通过动态调整偏好对的权重来进行公平感知偏好调优。我们的实验表明，FairPO 在保持摘要关键质量的同时，优于强基线。代码可在 https://github.com/leehaoyuan/coverage_fairness 获取。

1 引言

多文档摘要 (MDS) 旨在总结关于某一实体的多个文档中的重要信息，例如一款产品的评论。每个文档通常与社会属性相关联，如评论中的情感。这些具有不同社会属性值的文档，例如正面情感或负面情感，往往包含多样的信息或相互矛盾的观点。至关重要的是，摘要要公正地代表矛盾信息，因为这会显著影响决策。

之前的工作 (Shandilya et al., 2018; Olabisi et al., 2022; Huang et al., 2024) 在两个层次上测量 MDS 的公平性：摘要层次或语料库层次。摘要层次的公平性衡量一个摘要在多大程度上公平地代表具有不同社会属性值的文档。语料库层次的公平性衡量作为整体的摘要语料库在多大程度上公平地代表不同的社会属性值。

最近的研究 (Zhang et al., 2023; Li et al., 2024) 发现，现代的摘要方法如 LLMs 在摘要级和语料库级公平性方面都存在困难。为了提高摘要级的公平性，Zhang et al. (2023) 提示 LLMs 根据文档中社会属性的分布生成摘要。然而，这依赖于用户对公平性问题和社会属性的事先了解，限制了其在实践中的有效性。Huang et al. (2024) 通过策略梯度改进了 T5 (Raffel et al., 2020) 的摘要级公平性，但他们的方法可能无法推广到像 LLMs 这样的现代模型。此外，两个方法都专注于摘要级公平性，而忽视了语料库级公平性。

我们提出了 FairPO（公平偏好优化），这是一种偏好调整 (Ziegler et al., 2019) 方法，着重于 LLMs 在多文档总结中的摘要级和语料库级公平性。尽管之前的工作 (Stiennon et al., 2020; Roit et al., 2023) 使用偏好调整来提高摘要的其他质量，FairPO 是第一个在多文档总结中使用偏好调整以实现公平性的方法。FairPO 基于直接偏好优化（DPO）(Rafailov et al., 2024) 。为了优化摘要级的公平性，FairPO 通过移除具有某些社会属性值的小部分文档，生成扰动输入文档集的偏好对。为了进一步改善语料库级的公平性，FairPO 通过动态调整偏好对的权重来执行具备公平意识的偏好调整。

我们使用三个大型语言模型对 FairPO 进行实证评估：Llama3.1 (AI@Meta, 2024) 、Mistral (Jiang et al., 2023) 和 Gemma2 (Team et al., 2024) ，并在亚马逊 (Ni et al., 2019) 、MITweet (Liu et al., 2023) 和 SemEval 数据集 (Mohammad et al., 2016) 上进行实验。我们的实验表明，FairPO 优于强基线，同时保持摘要的其他关键质量，比如相关性和准确性。

我们的贡献如下：

毫米我们提出了 FairPO 来提高 MDS 中大语言模型的公平性；我们建议通过基于扰动的偏好对生成和公平性感知的偏好调优来改善摘要级和语料库级的公平性；我们进行了全面的实验以展示 FairPO 的有效性。

2 背景

在本节中，我们提供了关于 MDS 公平性的背景知识。设 G 表示 MDS 语料库中的所有文档集。每个文档集 $D \in G$ 包含多个文档 $\{d_1, ..., d_n\}$ ，其中每个文档 d_i 都标有一个社会属性 $a_i \in \{1, ..., K\}$ 。对于每个文档集 D ，

www.xueshuxiangzi.com

MDS 系统应生成一个摘要 S 。

为了评估 MDS 中的公平性，我们使用 Equal Coverage $EC(D,S)$（一个摘要级别的度量）和 Coverage Parity $CP(G)$（一个语料库级别的度量），由 Li et al. (2024) 提出。下面，我们总结了这些概念在原始论文中被引入的方式。

等效覆盖 检查每个社会属性值是否具有被文档集 D 的摘要 S 覆盖的相同概率。具体来说，它首先定义覆盖概率差 $c(d_i,S)$ 为文档 d_i 的覆盖概率与 $p(d_i,s)$ 之间的差异。它还定义了所有文档的平均覆盖概率，$p(d,s)$ 。为了估计文档 d_i 的覆盖概率 $p(d_i,s)$ ，FairPO 估计文档 d_i 被摘要句子 s_j 覆盖的概率 $p(d_i,s_j)$ 。具体来说，概率 $p(d_i,s_j)$ 被估计为文档 d_i 的任何文档块 $d_{i,l}$ 与摘要句子 s_j 之间的最大蕴涵概率 $p(d_{i,l},s_j)$ ，并使用一个蕴涵模型： 。然后，文档 d_i 的覆盖概率 $p(d_i,s)$ 被估计为概率 $p(d_i,s_j)$ 的平均值：

$$p(d_i,s)=\frac{1}{|S|}\sum_{s_j\in S}p(d_i,s_j), \tag{1}$$

。通过对文档集 D 中所有文档的覆盖概率 $p(d_i,s)$ 进行平均，得出平均覆盖概率 $p(d,s)$ 。利用这些值，Equal Coverage 计算覆盖概率差 $c(d_i,S)=p(d_i,s)-p(d,s)$ 。Equal Coverage 值 $EC(D,S)$ 然后被计算为每个社会属性值的文档绝对平均覆盖概率差 $c(d_i,S)$ 的平均值：

$$EC(D,S)=\frac{1}{K}\sum_{k=1}^{K}|\mathbb{E}(\{c(d_i,S)|a_i=k\})| \tag{2}$$

。较低的 $EC(D,S)$ 指示更公平的摘要 S 。为了评估系统的公平性，我们使用语料库 G 中的平均 Equal Coverage 值。

覆盖均衡 检查某些社会属性值在整个语料库 G 中是否系统性地被过度代表或不足代表。覆盖公平性将这些覆盖概率差异 $c(d_i,S)$ 从数据集 G 的所有输入文档中，收集到一个集合 C_k ，其中社会属性值是 k 。然后，通过计算每个社会属性值对应文档的绝对平均覆盖概率差异 $c(d_i,S)$ 的平均值，来得到覆盖公平性值 $CP(G)$:

$$CP(G)=\frac{1}{K}\sum_{k=1}^{K}|\mathbb{E}(C_k)|, \tag{3}$$

较低的 $CP(G)$ 表示一个更公平的系统。更多细节请参考 Li et al. (2024) 。

在本节中，我们描述了我们提出的偏好调优方法，FairPO。

在本节中，我们描述了如何基于扰动生成偏好对。FairPO 的一个偏好对包含一个被选择的摘要 S_c 和一个被拒绝的摘要 S_r ，针对文档集 D 。理想情况下，选择的和拒绝的摘要在表示具有不同社会属性值的文档时应显著不同。为此，FairPO 为扰动输入文档集生成摘要，其中去除了具有特定社会属性值的文档的小子集 $(\alpha\,\%)$。

具体而言，FairPO 首先为输入文档集 D 生成摘要 S ，并识别其最为过度代表的 k^+ 和代表不足的 k^- 社会属性值。为了信息的完整性，FairPO 仅考虑出现在超过 $\alpha\%$ 的文档中的社会属性值（详细信息见附录 **??** ）。这些是基于最高或最低平均覆盖概率差异 $\mathbb{E}(\{c(d_i,S)|a_i=k\})$ 确定的。然后，FairPO 为扰动后的输入文档集生成摘要 S^+ 和 S^- ，其中具有社会属性值 a_i 的 $\alpha\%$ 比例的随机抽样文档从 k^+ 和 k^- 中移除。在摘要 S 、S^+ 、S^- 中，FairPO 选择具有最低等覆盖值的摘要，表示最佳摘要层级公平性，作为选定的摘要 S_c 。具有最高等覆盖值的摘要被选择为被拒绝的摘要 S_r 。

在本节中，我们描述了能够优化概要级和语料库级公平性的公平感知偏好调整。为实现这一目标，FairPO 在训练期间根据估计的语料库级公平性动态地为选择的概要 S_c 和被拒绝的概要 S_r 分配单独的权重。

FairPO 修改了 DPO 目标（更多说明见附录 .1 ），并为选择的摘要 S_c 和拒绝的摘要 S_r 分别引入了独立的权重 w_c 和 w_r ：

$$\sigma(-m)\beta(w_r log\frac{\pi_\theta(S_r|D)}{\pi_{ref}(S_r|D)}-w_c log\frac{\pi_\theta(S_c|D)}{\pi_{ref}(S_c|D)}) \tag{4}$$

其中 σ 是 sigmoid 函数，π_θ 是策略模型，π_{ref} 是参考模型，m 是 DPO 中的奖励边界：

$$m=\beta log\frac{\pi_\theta(S_c|D)}{\pi_{ref}(S_c|D)}-\beta log\frac{\pi_\theta(S_r|D)}{\pi_{ref}(S_r|D)} \tag{5}$$

方程 4 中的项 $\sigma(-m)$ 用作缩放因子，FairPO 不考虑其梯度。

FairPO 根据其对语料库层面公平性的影响为总结分配权重 w_c 和 w_r 。它为那些通过平衡社会属性值的过度代表和不足代表来改善语料库层面公平性的选择总结分配高权重 w_c 。相反，它为那些损害语料库层面公平性的被拒绝总结分配高权重 w_r 。为了估计语料库层面公平性，FairPO 计算具有社会属性值 k 、$C_k(D,S_*)=\sum_{d\in\{d_i|a_i=k\}}c(d,S_*)$ 的文档在每个选择或拒绝的总结 S_* 上的覆盖概率差异之和。一个总结 S_* 被认为是过度代表或不足代表社会属性值 k 的，如果覆盖概率差异之和 $C_k(D,S_*)$ 分别大于或小于零。在每个训练步骤中，FairPO 估计社会属性值 k 的过度代表 $O(k)$:

www.xueshuxiangzi.com

	Domain	Soci. Attr.	Soci. Attr. Val.	Doc. Set Size	Doc. Len
Amazon	Review	Sentiment	negative, neutral, positive	8	40
MiTweet	Tweet	Ideology	left, center, right	20	34
SemEval	Tweet	Stance	support, against	30	17

Table 1: 数据集统计。文档集大小指的是文档集的大小。文档长度指的是文档的平均长度。

$$O(k) = \frac{\sum_{(D,S)\in T_k^+} |C_k(D,S)| \cdot \pi_\theta(S|D)/|S|}{\sum_{(D,S)\in T_k^+} \pi_\theta(S|D)/|S|} \tag{6}$$

其中 T_k^+ 是文档集 D 以及对应的选择或拒绝总结集，在最近的训练步骤中，这些总结过度代表了社会属性值 k （$C_k(D,S_*) > 0$）。类似地，FairPO 使用文档集和总结集合 T_k^- 来估计不足代表 $U(k)$，这些集合在最近的训练步骤中不足代表社会属性值 k （$C_k(D,S_*) < 0$），如方程 6 所示。

利用过度代表 $O(k)$ 和代表不充分 $U(k)$，FairPO 分配权重 w_c 和 w_r。选择能够帮助平衡过度代表 $O(k)$ 和代表不充分 $U(k)$ 的摘要会获得更高的权重，反之，对于被拒绝的摘要也是如此。例如，如果一个系统性被低估的社会属性值 k （$U(k) > O(k)$）在选定的摘要 S_c（$C_k(D,S_c) > 0$）中被过度代表，则权重 w_c 应更高。对于社会属性值 k，FairPO 为选定的摘要 S_c 计算一个中间权重 $w_{c,k}$：

$$w_{c,k} = \frac{2}{1 + (O(k)/(U(k))^{C_k(D,S_c)/\tau}} \tag{7}$$

其中 τ 为温度。选择的摘要的权重 w_c 是跨所有社会属性值的平均中间权重 $w_{c,k}$。用于被拒绝摘要的权重 w_r 与中间权重 $w_{r,k}$ 相似地计算：

$$w_{r,k} = \frac{2}{1 + (U(k)/(O(k))^{C_k(D,S_r)/\tau}} \tag{8}$$

该设计确保优先考虑提高语料库公平性的摘要。

在本节中，我们描述了用 FairPO 微调模型的实验。

我们在三个数据集上进行了实验：Amazon、MITweet 和 SemEval 数据集。每个数据集包括用于训练的 1000 个样本，用于验证的 300 个样本，以及用于测试的 300 个样本。训练集、验证集和测试集的划分基于社会属性值和话题的分层采样。表 1 显示了这些数据集的统计信息。摘要长度为 50 个单词。数据预处理的详细信息在附录 **??** 中。

我们使用三个大型语言模型（LLMs）进行实验：Llama3.1-8b-Instruct (AI@Meta, 2024)、Mistral-7B-Instruct-v0.3 (Jiang et al., 2023)、Gemma-2-9b-it (Team et al., 2024)。每个 LLM 使用 LoRA (Hu et al., 2021) 进行 2 个 epoch 的训练，学习率为 $5e-5$，批次大小为 16。为了生成偏好对，FairPO 移除了 $\alpha = 10\%$ 的文档。在 MITweet 数据集上的温度是 1，对于 Mistral 在 Amazon 数据集上是 2，对于其他 LLM 是 1，在 SemEval 数据集上对于 Mistral 是 3，对于其他 LLM 是 2。所有的超参数都在验证集上进行了调优。更多细节见附录 **??**。

	Amazon		MITweet		SemEval		Overall	
	$EC\downarrow$	$CP\downarrow$	EC	$CP\downarrow$	$EC\downarrow$	$CP\downarrow$	$\overline{EC}\downarrow$	$\overline{CP}\downarrow$
Llama3.1	7.95	1.89	4.50	0.59	2.98	1.41	5.14	1.30
+DPO	7.23	1.27	4.25	0.47	2.66	1.09	4.72	0.94
+OPTune	6.70	0.62	4.33	0.51	2.60	0.95	4.54	0.69
+Prompt	7.42	1.64	4.36	0.45	2.62	0.29	4.80	0.79
+Policy G.	7.73	1.88	4.51	0.55	2.97	1.38	5.07	1.27
+FairPO	6.87	0.42	4.24	0.42	2.49	0.66	4.53	0.50
Mistral	8.36	2.83	4.16	0.61	2.83	1.27	5.12	1.57
+DPO	7.20	1.82	3.55	0.34	2.41	0.93	4.39	1.03
+OPTune	6.85	0.88	3.58	0.51	2.07	0.57	4.17	0.65
+Prompt	7.74	1.92	3.97	0.37	2.35	0.36	4.68	0.88
+FairPO	6.32	0.46	3.70	0.40	2.10	0.43	4.04	0.43
Gemma2	8.32	2.48	4.20	0.60	2.81	0.96	5.11	1.35
+DPO	6.90	0.91	4.04	0.40	2.44	0.56	4.46	0.62
+OPTune	6.84	0.88	3.89	0.57	2.32	0.49	4.35	0.65
+Prompt	7.28	1.16	4.33	0.32	2.73	0.48	4.78	0.65
+FairPO	6.18	0.44	3.76	0.48	2.50	0.45	4.15	0.46

Table 2: 由不同方法生成的总结的总结级公平性（EC）和语料库级公平性（CP）。表现最好的方法以粗体显示。表现第二好的方法是 underlined。FairPO 具有最佳的整体性能。

2.1 FairPO 的自动评估

我们将 FairPO 与以下基线进行比较：(i) DPO (Rafailov et al., 2024)，其中选择和拒绝的摘要是在三个随机抽样摘要中基于与 FairPO 相同的 EC 值进行选择的，以进行公平比较；(ii) OPTune (Chen et al., 2024)，它像 DPO 一样选择选择和拒绝的摘要，并根据 EC 值差异加权偏好对；(iii) 策略梯度 (Lei et al., 2024) 和 (iv) 一种提示方法 (Zhang et al., 2023)。这些基线的实现细节在附录 **??** 中。在评估中，我们使用相等覆盖（EC）和覆盖平衡（CP）(Li et al., 2024) 来考虑摘要级和语料库级的公平性。对于这些指标，值越低越好。我们在表格 2 中报告了训练、验证和测试的三种拆分的平均结果。我们还在附录 .2 中报告了每种拆分的结果。我们观察到，FairPO 在所有数据集的大多数 LLM 上表现优于其他方法，并为所有 LLM 提供了最佳的整体性能。结果表明，FairPO 改善了摘要级和语料库级的公平性。

为了验证基于扰动的偏好对生成和公平感知偏好调整的效果，我们将 FairPO 与其消融版本进行比较。我们考虑以下消融版本：(i) (无扰动) 中，选择和拒绝的摘要是在三个随机抽样的摘要中根据相等覆盖值选出的；(ii) (无公平)

www.xueshuxiangzi.com

	Llama3.1			Mistral			Gemma2		
	flu. ↑	rel. ↑	fac. ↑	flu. ↑	rel. ↑	fac. ↑	flu. ↑	rel. ↑	fac. ↑
DPO	7.56	8.33	2.78	5.11	11.56	11.56	5.11	1.11	8.67
OPTune	1.00	0.44	-6.89	-0.78	6.78	8.89	7.00	11.67	11.67
Prompt	-15.33	-19.22	-24.44	-0.44	-6.00	-5.56	-42.67	-50.78	-51.44
FairPO	5.78	3.11	2.89	2.11	5.33	9.11	11.44	16.11	9.44

Table 3: 调优前后由 LLM 生成的摘要之间质量的成对比较。根据成对自助重采样 (Koehn, 2004)，具有统计显著差异的 ($p < 0.05$) 部分下划线标出。FairPO 不影响摘要质量。

中，使用 DPO 目标进行偏好调整，而不是公平感知的偏好调整；(iii) (无奖励) 中，在 DPO 目标中直接分配权重 w_c 和 w_r (方程 12)，这削弱了奖励边距的效果（更多解释在附录 .1 中）。表 **??** 报告了每个数据集的结果以及 Overall 分数，即所有数据集的平均值。较低的值表示更好的公平性。

从表中我们观察到，FairPO 相比于其裁减版本提供了最佳的整体性能。结果表明基于扰动的偏好对生成和公平意识的偏好调整的有效性。它也为 FairPO 目标的设计选择提供了实证证据。

我们进行了一项人工评估，以比较使用 DPO 和 FairPO 调整的 LLM 生成的摘要的公平性。对于每个 LLM，我们随机选择 10 对由使用 DPO 或 FairPO 调整的 LLM 生成的摘要，总共得到 30 对。每对摘要由三名来自 Amazon Mechanical Turk 的标注者进行标注。标注者被要求阅读所有相应的文件并选择更公平的摘要。我们选择了 Amazon 数据集，因为每个文档集仅包含八个评论（表格 1），判断观点的情感对普通用户来说比较容易。三位标注者之间标注的 Randolph’s Kappa 是 (Randolph, 2005)，显示出中等的相关性。考虑到任务的主观性，这种相关性是预期的。更多细节在附录 **??** 中。

在 30 对中，FairPO 调整的 LLM 生成的摘要在 18 对中更公平，而 DPO 调整的 LLM 生成的摘要在 9 对中更公平。使用引导法 (Koehn, 2004)，该差异在统计上显著（$p < 0.05$）。结果表明，FairPO 在提高公平性方面表现优于 DPO。我们在附录 **??** 中还展示了 FairPO 生成的摘要示例。

2.2 总结质量评估

为了评估 FairPO 对摘要质量的影响，我们比较了 LLM 调整前后生成的摘要，以改善公平性。具体而言，对于一对摘要，我们指示 Prometheus 2 (7B) (Kim et al., 2024) 从三个维度选择更好的摘要：流畅性、相关性和真实性。为减轻位置偏见 (Huang et al., 2023)，我们以不同的摘要顺序进行了两次成对比较，仅考虑一致的结果。表 3 报告了不同方法的胜率和败率之差。正值表示摘要质量较原始 LLM 更好。

从表中可以看出，经过 FairPO 微调的大型语言模型生成的摘要质量与原始大型语言模型生成的摘要质量相当。相反，提示显著降低了摘要的质量。结果表明，FairPO 在提高摘要公平性的同时保持了其质量。

3 结论

我们提出了 FairPO，一种偏好调整方法，用于优化多文档摘要中的摘要级公平性和语料库级公平性。具体来说，FairPO 通过使用扰动的文档集生成偏好对，以提高摘要级公平性，并进行公平感知的偏好调整以提高语料库级公平性。我们的实验表明，FairPO 在保持摘要关键质量的同时，优于强基线。

4 限制

我们的实验展示了 FairPO 在提高单个领域内摘要的总结层面和语料库层面公平性方面的有效性。虽然这项工作专注于优化单个领域内的公平性，但扩展 FairPO 以同时提升具有多种社会属性的多个领域的公平性是一个充满前景的未来方向。此外，FairPO 目前在生成的三个摘要中选择公平性差异最大的两个进行偏好调节，这是遵循 DPO 常用的做法。探索如何利用 FairPO 生成的全部三个摘要也可以是另一个有趣的未来方向。

5 致谢

本工作得到了 NSF 奖项 DRL-2112635 和 2338418 的支持。

6 伦理考量

我们使用的数据集都是公开可用的。我们没有自行标注任何数据。本论文中使用的所有模型都是公开可访问的。模型的推理和微调在一台 Nvidia A6000 或 Nvidia A100 GPU 上进行。

我们在 Amazon Mechanical Turk 上进行人工评价实验。注释者的报酬为每小时 $ 20。在评价过程中，人工注释者未接触到任何敏感或露骨的内容。

References

AI@Meta. 2024. Llama 3 model card.

Lichang Chen, Jiuhai Chen, Chenxi Liu, John Kirchenbauer, Davit Soselia, Chen Zhu, Tom Goldstein, Tianyi Zhou, and Heng Huang. 2024. Optune: Efficient online preference tuning. *arXiv preprint arXiv:2406.07657*.

www.xueshuxiangzi.com

Edward J Hu, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2021. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Kung-Hsiang Huang, Philippe Laban, Alexander R Fabbri, Prafulla Kumar Choubey, Shafiq Joty, Caiming Xiong, and Chien-Sheng Wu. 2023. Embrace divergence for richer insights: A multi-document summarization benchmark and a case study on summarizing diverse information from news articles. *arXiv preprint arXiv:2309.09369*.

Nannan Huang, Haytham Fayek, and Xiuzhen Zhang. 2024. Bias in opinion summarisation from pretraining to adaptation: A case study in political bias. *arXiv preprint arXiv:2402.00322*.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Seungone Kim, Juyoung Suk, Shayne Longpre, Bill Yuchen Lin, Jamin Shin, Sean Welleck, Graham Neubig, Moontae Lee, Kyungjae Lee, and Minjoon Seo. 2024. Prometheus 2: An open source language model specialized in evaluating other language models. *Preprint*, arXiv:2405.01535.

Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 388–395, Barcelona, Spain. Association for Computational Linguistics.

Yuanyuan Lei, Kaiqiang Song, Sangwoo Cho, Xiaoyang Wang, Ruihong Huang, and Dong Yu. 2024. Polarity calibration for opinion summarization. *arXiv preprint arXiv:2404.01706*.

Haoyuan Li, Yusen Zhang, Rui Zhang, and Snigdha Chaturvedi. 2024. Coverage-based fairness in multi-document summarization. *Preprint*, arXiv:2412.08795.

Songtao Liu, Ziling Luo, Minghua Xu, LiXiao Wei, Ziyao Wei, Han Yu, Wei Xiang, and Bang Wang. 2023. Ideology takes multiple looks: A high-quality dataset for multifaceted ideology detection. In *The 2023 Conference on Empirical Methods in Natural Language Processing*.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California. Association for Computational Linguistics.

Jianmo Ni, Jiacheng Li, and Julian McAuley. 2019. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 188–197, Hong Kong, China. Association for Computational Linguistics.

Olubusayo Olabisi, Aaron Hudson, Antonie Jetter, and Ameeta Agrawal. 2022. Analyzing the dialect diversity in multi-document summaries. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6208–6221, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

Justus J Randolph. 2005. Free-marginal multirater kappa (multirater k [free]): An alternative to fleiss' fixed-marginal multirater kappa. *Online submission*.

Paul Roit, Johan Ferret, Lior Shani, Roee Aharoni, Geoffrey Cideron, Robert Dadashi, Matthieu Geist, Sertan Girgin, Leonard Hussenot, Orgad Keller, Nikola Momchev, Sabela Ramos Garea, Piotr Stanczyk, Nino Vieillard, Olivier Bachem, Gal Elidan, Avinatan Hassidim, Olivier Pietquin, and Idan Szpektor. 2023. Factually consistent summarization via reinforcement learning with textual entailment feedback. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6252–6272, Toronto, Canada. Association for Computational Linguistics.

Anurag Shandilya, Kripabandhu Ghosh, and Saptarshi Ghosh. 2018. Fairness of extractive text summarization. In *Companion Proceedings of the The Web Conference 2018*, pages 97–98.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. 2020. Learning to summarize with human feedback. *Advances in Neural Information Processing Systems*, 33:3008–3021.

www.xueshuxiangzi.com

Gemma Team, Thomas Mesnard, Cassidy Hardin, Robert Dadashi, Surya Bhupatiraju, Shreya Pathak, Laurent Sifre, Morgane Rivière, Mihir Sanjay Kale, Juliette Love, et al. 2024. Gemma: Open models based on gemini research and technology. *arXiv preprint arXiv:2403.08295*.

Yusen Zhang, Nan Zhang, Yixin Liu, Alexander Fabbri, Junru Liu, Ryo Kamoi, Xiaoxin Lu, Caiming Xiong, Jieyu Zhao, Dragomir Radev, et al. 2023. Fair abstractive summarization of diverse perspectives. *arXiv preprint arXiv:2311.07884*.

Daniel M Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. 2019. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*.

在本节中，我们描述了如何预处理数据集。

亚马逊 (Ni et al., 2019) 由带有不同产品评分标签的评论组成。我们过滤掉那些非英文或没有评分的评论。我们根据数据集中提供的评分来获取每条评论的社会属性。如果评论的评分是 4 或 5 ，则其社会属性为正；如果评分是 3 ，则为中性；如果评分是 1 或 2 ，则为负。为了构建训练集、验证集和测试集，我们基于每个集合中文档集合的社会属性值分布进行分层抽样。因此，每个集合中每种社会属性值主导的文档集合都有相等的比例。我们为训练采样 1000 个产品及其对应的评论，为验证采样 300 个产品，为测试采样 300 个产品。

微推 (Liu et al., 2023) 由关于不同主题的不同方面的政治意识形态标签的推文组成。推文的社会属性在大多数方面为左就为左，在大多数方面为右就为右，否则为中立。首先，我们将每个主题的所有推文平均分成两部分，以使两个部分之间的主题分布相同。对于每个部分，我们基于推文的 TFIDF 相似性将相同主题的推文聚类成集群。然后我们将这些集群分成同一主题的 20 条推文的输入文档集。我们从推文的第一部分生成 1000 个训练输入文档集。同样地，我们从推文的第二部分生成 300 个验证输入文档集和 300 个测试输入文档集。在生成训练、验证和测试集合的输入文档集时，我们还根据社会属性值的分布进行分层抽样，以使每个集合的文档集中由每个社会属性值占主导的文档集 D 的比例相等。

推文立场 (Mohammad et al., 2016) 由具有针对目标短语（如气候变化或希拉里 · 克林顿）的立场标签的推文组成。首先，我们将每个主题的所有推文均匀地分为两部分，以确保目标短语的分布在两部分之间相同。然后，我们根据推文的 TFIDF 相似性将关于同一目标短语的推文聚类为多个簇。接着，我们将这些簇分为 30 条关于同一目标短语的推文输入文档集。我们从推文的第一部分生成 1000 个用于训练的输入文档集。类似地，我们从推文的第二部分生成 300 个用于验证的输入文档集和 300 个用于测试的输入文档集。当生成训练、验证和测试集的输入文档集时，我们还根据社会属性值的分布进行分层抽样，以确保每个集中的文档集 D 由每个社会属性值主导的比例相同。

我们进行了一次人工评估，以比较经过 DPO 和 FairPO 调整的 LLM 所生成摘要的公平性。对于每个 LLM，我们随机选择由经过 DPO 或 FairPO 调整的 LLM 生成的 10 对摘要，总计 30 对。为了进一步简化评估，我们仅考虑包含负面和正面评论的文档集。每对摘要由三名来自亚马逊机械土耳其平台招募的标注者进行标注。标注者应来自英语国家，并且 HIT 批准率大于 98% 。对于每对摘要，标注者首先被要求阅读相应的评论和通过 GPT-4o-mini 自动提取的独特观点 (Ouyang et al., 2022) 。然后，他们评估每个摘要是否反映这些观点，并将摘要分类为倾向于负面、公平或倾向于正面。最终，他们被要求选择每对中较为公平的摘要。人工评估的界面如图 1 所示。

.1 FairPO 与 DPO 之间的关系

FairPO 目标（方程 4 ）受到 DPO 目标相对于模型参数 θ 的导数的启发：

$$\begin{aligned}\sigma(-m)\beta(\pi_\theta(S_r|D)^{-1}\frac{\partial\pi_\theta(S_r|D)}{\partial\theta}\\ -\pi_\theta(S_c|D)^{-1}\frac{\partial\pi_\theta(S_c|D)}{\partial\theta})\end{aligned} \tag{9}$$

其中 σ 是 sigmoid 函数，π_θ 是策略模型，π_{ref} 是参考模型，m 是 DPO 中的奖励边界：

$$\beta log\frac{\pi_\theta(S_c|D)}{\pi_{ref}(S_c|D)} - \beta log\frac{\pi_\theta(S_r|D)}{\pi_{ref}(S_r|D)} \tag{10}$$

奖励边界 m 可以视为模型区分所选摘要 S_c 和被拒绝摘要 S_r 能力的一种度量。m 的较大值表明模型在区分 S_c 和 S_r 方面已经很熟练。因此，DPO 将较低的权重，$\sigma(-m)$ ，分配给模型对它们的区别有信心的所选和被拒绝的摘要，而将更高的权重分配给区别更具挑战性的所选和被拒绝的摘要。术语 $\sigma(-m)$ 可以帮助模型更专注于困难的案例。

FairPO 的目标被设计为使选择和拒绝的摘要具有单独的权重，同时保留方程 9 中的项 $\sigma(-m)$ 的影响。FariPO 目标关于模型参数 θ

www.xueshuxiangzi.com

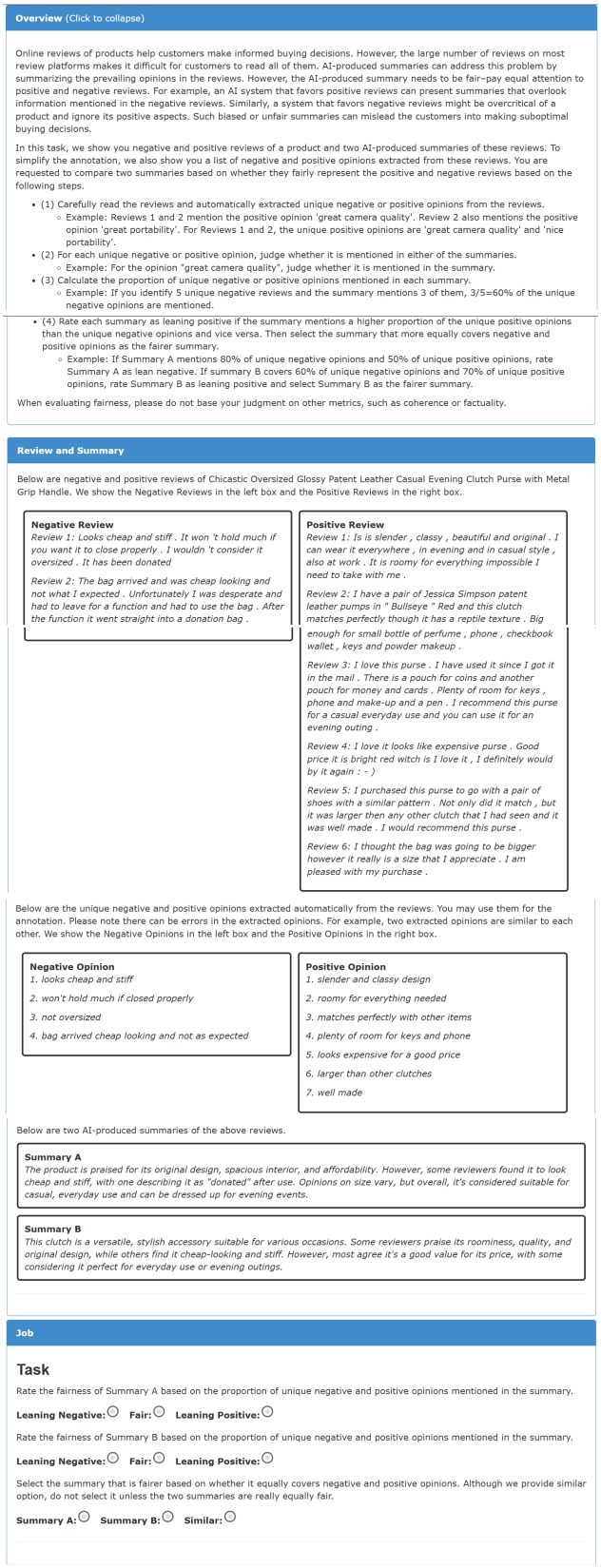

Figure 1: 用于人类评估的界面

www.xueshuxiangzi.com

Below is a list of product reviews:

1.This is a card reader that does everything I needed it to . My adapters for the micro SD cards were defective so I have no complaints only praise . It reads any Compact Flash , Memory Stick , SD , and XD cards . Well that is all I wanted to say except this is a great product overall , and thank you .

2.The pins in the CF slot are very flimsy and get bent out of alignment easily , making it impossible to insert the card (until you perform delicate surgery on the pins with small tweezers) . Do not buy this product if you will ever use the CompactFlash slot . It will just lead to frustration .

3.So far I only use this for SM and SD cards , but it installed (USB) quickly , easily and reads the cards I need read .

4.Initially it worked great but after the 5th time it stopped working . It also helped fry my SD-card will all my pictures and video clips . Not happy at all with this product .

5.Reads 64 cards is quite deceiving . It only reads four types of cards made by 64 different manufacturers . Also , the connector port is difficult to plug in .

6.good product , reads quite fast. only issue is that the card reader does not have a satisfying ' click ' when the card is inserted. you kinda have to stick the card in the slot and hope it is lodged properly .

7.I can get it to read SD cards , but I bought it to read my CF 's and it won 't read a single one . My experience is in line with others . Go check out similar reviews on newegg.com.

8.The card reader comes in retail packaging and totally lacks instructions on how best to put 68 types of cards into 4 slots . It did read an SD card successfully . The micro usb plug on the usb cord broke after 1 use .

Please write a single summary around 50 words for all the above reviews.

Figure 2: 亚马逊数据集的总结提示。

的导数如下：

$$\sigma(-m)\beta(w_r\pi_\theta(S_r|D)^{-1}\frac{\partial\pi_\theta(S_r|D)}{\partial\theta} - w_c\pi_\theta(S_c|D)^{-1}\frac{\partial\pi_\theta(S_c|D)}{\partial\theta}) \tag{11}$$

。与 DPO 目标的导数（方程 9 ）相比，项 $\sigma(-m)$ 在 FairPO 目标的导数中保持一致。

假设我们直接为选择的摘要和被拒绝的摘要向 DPO 目标添加分别的权重 w_c 和 w_r 。相应的目标如下：

$$-log\sigma(\beta w_c log\frac{\pi_\theta(S_c|D)}{\pi_{ref}(S_c|D)} - \beta w_r log\frac{\pi_\theta(S_r|D)}{\pi_{ref}(S_r|D)}) \tag{12}$$

。相应的导数如下：

$$\sigma(-m')\beta(w_r\pi_\theta(S_r|D)^{-1}\frac{\partial\pi_\theta(S_r|D)}{\partial\theta} - w_c\pi_\theta(S_c|D)^{-1}\frac{\partial\pi_\theta(S_c|D)}{\partial\theta}) \tag{13}$$

，其中 m' 是加权奖励边界：

$$\beta w_c log\frac{\pi_\theta(S_c|D)}{\pi_{ref}(S_c|D)} - \beta w_r log\frac{\pi_\theta(S_r|D)}{\pi_{ref}(S_r|D)} \tag{14}$$

。与 m 相比，m' 作为衡量模型区分选择的摘要 S_c 和被拒绝的摘要 S_r 的能力的指标效果较差，因为项 $log\frac{\pi_\theta(S_c|D)}{\pi_{ref}(S_c|D)}$ 和 $log\frac{\pi_\theta(S_r|D)}{\pi_{ref}(S_r|D)}$ 具有不同的权重。在附录 **??** 中，我们还提供了实证证据。

为了减少训练成本，我们进行了 LoRA (Hu et al., 2021) 微调。具体来说，LoRA 微调的秩为 16 ，缩放因子也是 16 。所有模型都被量化为 8 位以进一步减少训练成本。

当对每个文档集进行扰动以生成偏好对时，我们观察到某些社会属性值在某些文档集中极为罕见。如果 FairPO 移除这些罕见社会属性值的 α 百分比的文档，这些社会属性值将从文档集中完全消失。因此，在进行扰动时，我们仅考虑出现在超过 α 百分比文档中的社会属性值。在最极端的情况下，如果只有一个社会属性值满足此要求，FairPO 将对具有该社会属性值的 α 百分比文档的不同子集进行采样。通过这样做，我们确保在扰动后社会属性值的完整性。

我们提示这些大型语言模型为不同数据集的输入文档集生成摘要。提示经过调整，使生成的摘要的平均长度为 50 个单词。我们在图 2 中展示了用于 Amazon 数据集的摘要提示。所有大型语言模型的生成温度均为 0.6 。

方程式 6 中的集合 T_k^+ 进行更新，以便最近的训练步骤具有更高的影响力。具体而言，在每个训练步骤结束时，集合 T_k^+ 中已经存在的所有样本的影响力都会根据折扣因子 γ 进行降低。然后，将当前训练步骤中所有过度表现社会属性值 k （$C_k(D,S_*)>0$）的样本添加到集合 T_k^+ 中。对于 Llama3.1，折扣因子 γ 为 0.75 ，对于其他 LLM 则为 0.5 。

在方程 7 和方程 8 中，$C_k(D,S_*)$ 的目标在于调整权重 w_c 和 w_r ，以便当 $C_k(D,S_*)$ 与 0 偏离更多时，它与 1 偏离更多。因此，FairPO 不会直接使用覆盖概率差异之和的原始值 $C_k(D,S_*)$ 作为指数。而是，FairPO 在所有训练样本中单独归一化 $C_k(D,S_*)$ ，其中 $C_k(D,S_*)$ 大于零或小于零。

我们实现了 Lei et al. (2024) 提出的政策梯度

www.xueshuxiangzi.com

	Amazon		MITweet		SemEval		Overall	
	$EC\downarrow$	$CP\downarrow$	EC	$CP\downarrow$	$EC\downarrow$	$CP\downarrow$	$\overline{EC}\downarrow$	$\overline{CP}\downarrow$
Llama3.1	7.90	1.92	4.43	0.26	2.94	1.33	5.09	1.17
+DPO	6.87	1.04	4.03	0.31	2.55	0.91	4.49	0.75
+OPTune	6.58	0.75	4.22	0.23	2.50	0.81	4.43	0.60
+Prompt	7.71	1.84	4.33	0.38	2.53	0.26	4.86	0.83
+Policy G.	7.71	2.10	4.46	0.31	2.95	1.32	5.04	1.24
+FairPO	6.57	0.37	4.20	0.26	2.39	0.56	4.39	0.39
Mistral	8.18	2.98	3.98	0.42	2.67	1.07	4.94	1.49
+DPO	7.17	1.55	3.60	0.28	2.21	0.64	4.33	0.82
+OPTune	7.48	1.56	3.60	0.25	2.00	0.67	4.36	0.83
+Prompt	7.67	1.93	4.02	0.23	2.38	0.38	4.69	0.85
+FairPO	6.98	0.89	3.56	0.21	1.97	0.36	4.17	0.49
Gemma2	8.44	2.75	4.17	0.34	2.74	0.91	5.12	1.33
+DPO	6.87	1.04	4.04	0.29	2.42	0.70	4.44	0.68
+OPTune	6.90	1.15	3.86	0.45	2.40	0.65	4.39	0.75
+Prompt	7.21	1.13	4.28	0.24	2.62	0.30	4.70	0.56
+FairPO	6.09	0.33	3.84	0.47	2.53	0.59	4.15	0.46

Table 4: 不同方法在第一次训练、验证和测试分割中生成摘要的摘要级别公平性（EC）和语料库级别公平性（CP）。表现最佳的方法以粗体显示。表现第二好的方法是 underlined 。FairPO 在第一次分割中总体表现最佳。

	Amazon		MITweet		SemEval		Overall	
	$EC\downarrow$	$CP\downarrow$	EC	$CP\downarrow$	$EC\downarrow$	$CP\downarrow$	$\overline{EC}\downarrow$	$\overline{CP}\downarrow$
Llama3.1	7.90	2.05	4.50	0.63	2.90	1.41	5.10	1.36
+DPO	7.27	1.37	4.30	0.37	2.70	1.12	4.76	0.95
+OPTune	6.92	0.40	4.30	0.52	2.82	1.00	4.68	0.64
+Prompt	7.28	1.67	4.41	0.44	2.74	0.51	4.81	0.87
+Policy G.	7.75	1.85	4.47	0.48	2.80	1.30	5.02	1.21
+FairPO	6.96	0.44	4.26	0.29	2.69	0.59	4.64	0.44
Mistral	8.60	2.74	4.18	0.73	2.91	1.28	5.23	1.58
+DPO	7.24	1.79	3.39	0.26	2.70	1.15	4.44	1.07
+OPTune	6.59	0.52	3.57	0.53	2.04	0.58	4.07	0.54
+Prompt	7.90	1.76	3.74	0.51	2.43	0.52	4.69	0.93
+FairPO	6.06	0.11	3.83	0.39	2.13	0.33	4.01	0.28
Gemma2	8.31	2.33	4.30	0.80	2.97	1.03	5.19	1.38
+DPO	7.04	0.98	4.07	0.44	2.43	0.48	4.51	0.63
+OPTune	6.91	0.56	3.94	0.86	2.35	0.56	4.40	0.66
+Prompt	7.33	1.26	4.49	0.44	2.91	0.85	4.91	0.85
+FairPO	6.09	0.44	3.82	0.65	2.70	0.32	4.20	0.47

Table 5: 在第二次划分出的训练集、验证集和测试集上，不同方法生成的摘要的总结层面公平性（EC）和语料库层面公平性（CP）。表现最好的方法以粗体显示。表现第二好的方法是 underlined 。FairPO 在第二次划分中的总体表现最佳。

方法作为基线。在原始实现中，除策略梯度外，还有一个最大化参考摘要概率的损失。由于本文使用的数据集不包含参考摘要，我们只考虑策略梯度。此外，为了与其他方法进行公平比较，我们在离线环境中实现了策略梯度方法。政策梯度的学习率为 $1e-6$ ，遵循原始论文的设置。我们仅对 Llama3.1 实现了策略梯度方法，因为即使将学习率降低至 $1e-9$ ，对于 Mistral 和 Gemma2 的训练仍然非常不稳定。对于 OPTune 和 DPO，它们使用与 FairPO 相同的超参数。

	Amazon		MITweet		SemEval		Overall	
	$EC\downarrow$	$CP\downarrow$	EC	$CP\downarrow$	$EC\downarrow$	$CP\downarrow$	$\overline{EC}\downarrow$	$\overline{CP}\downarrow$
Llama3.1	8.06	1.70	4.57	0.87	3.11	1.51	5.25	1.36
+DPO	7.55	1.39	4.43	0.74	2.73	1.24	4.90	1.12
+OPTune	6.61	0.72	4.47	0.79	2.48	1.03	4.52	0.85
+Prompt	7.26	1.42	4.35	0.53	2.59	0.11	4.74	0.69
+Policy G.	7.72	1.69	4.60	0.85	3.16	1.53	5.16	1.36
+FairPO	7.07	0.44	4.25	0.71	2.38	0.82	4.57	0.66
Mistral	8.29	2.76	4.32	0.67	2.90	1.46	5.17	1.63
+DPO	7.20	2.11	3.65	0.48	2.32	0.99	4.39	1.19
+OPTune	6.47	0.55	3.58	0.76	2.18	0.46	4.08	0.59
+Prompt	7.66	2.07	4.14	0.36	2.23	0.19	4.68	0.87
+FairPO	5.92	0.38	3.71	0.61	2.21	0.59	3.95	0.53
Gemma2	8.21	2.36	4.14	0.67	2.72	0.96	5.02	1.33
+DPO	6.80	0.70	4.01	0.49	2.46	0.49	4.42	0.56
+OPTune	6.72	0.94	3.86	0.41	2.21	0.25	4.27	0.53
+Prompt	7.29	1.09	4.21	0.28	2.66	0.28	4.72	0.55
+FairPO	6.35	0.56	3.64	0.32	2.29	0.43	4.09	0.44

Table 6: 总结级公平性（EC）和语料库级公平性（CP），生成的摘要由不同的方法在训练、验证和测试的第三次分割中生成。表现最好的方法以粗体显示。第二好的方法是 underlined 。FairPO 在第三次分割中表现最好。

.2 使用不同数据集划分的结果

为了验证 FairPO 在三种不同数据集拆分上的稳定性，我们使用不同的随机种子生成训练集、验证集和测试集，并在这些拆分上运行自动评估。每种拆分的结果分别显示在表格 4 、5 、6 中。从表中可以看出，FairPO 仍然表现出最佳的整体性能，这表明 FairPO 在不同数据集拆分上的稳定性。

我们在图 **??** 中展示了通过 DPO 和 FairPO 调优的 LLM 在 Amazon 数据集上生成的示例摘要。从图中可以看出，经过 FairPO 调优的 LLM 生成的摘要倾向于更平衡地呈现负面和正面信息。

www.xueshuxiangzi.com