NLP_双关语

摘要

- LLM生成幽默句子的能力仍然存在局限（ChatGPT的幽默生成能力仅限于创作25条不重复的笑话。）
- 工作简介
  - 我们提出一种多阶段课程偏好学习框架，同步优化双关结构偏好与幽默偏好。具体而言，我们改进了直接偏好优化（DPO）算法以解决多目标对齐难题
- 新数据集
  - 为推进该领域发展，我们构建了包含2.1k条双关语及对应标注的中文双关数据集

简介

- 双关语
  - 双关语生成旨在根据给定的双关词与替代词组合（如表1所示）生成具有幽默效果的语句，其核心目标包含双重维度：双关结构与幽默效果
- 为从新维度提升LLMs能力，偏好学习/对齐已成为当前主流方法（Casper等人，2023）
  - 然而，在双关生成任务中直接应用直接偏好优化（DPO）（Rafailov等人，2023）等偏好学习方法效率较低——这源于我们需在较小规模数据集上同步实现双关结构偏好与幽默偏好的双重对齐。
- 文章提出的解决办法：多阶段课程学习框架
  - 第一阶段优先学习双关结构偏好
  - 第二阶段进阶学习幽默偏好
    - 为解决灾难性的遗忘问题：在该阶段引入一种改进的幽默偏好对齐算法
  - 与标准DPO方法（仅从正负样本对中学习）不同，我们通过为每个训练样本添加第三元——即使用相同输入提示从第一阶段模型获得的额外生成输出——将其构建为三元组样本
    - 提出一种新型损失函数，既能适配这种多目标对齐任务，又可缓解灾难性遗忘效应。
- 数据集
  - 英语双关语数据集：English pun dataset from SemEval 2017 task 7 (SemEval) (Miller et al., 2017)
    - 包含1298个同形双关语
    - 1098 个用双关语和替代词注释的谐音双关语
  - 中文数据集 ChinesePun

Type	sentences	words	pun words min	pun words max	pun words mean
homophonic	1049	96212	1	4	1.11
homographic	1057	91100	1	3	1.05

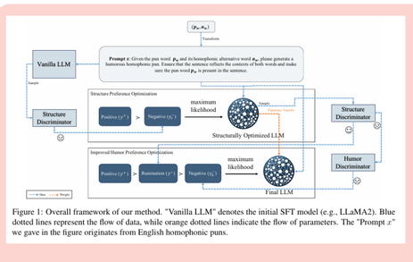
Table 2: Statistics of the ChinesePun dataset

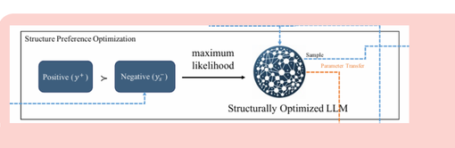
    - 本数据集包含1,049条谐音双关（homophonic puns）和1,057条同形异义双关（homographic puns），每条双关语句均标注有对应的双关词及其替代词
- 贡献总结
  - 我们提出了一种多阶段课程学习框架，结合创新的三元组偏好学习方法，显著提升了大语言模型生成幽默双关语的能力。
  - 我们发布了首个面向中文双关语生成任务的数据集，旨在推动中文幽默理解与生成领域的研究进展。
  - 我们在中英文双关数据集上进行了全面实验，结果充分证明了所提出方法的优越性。

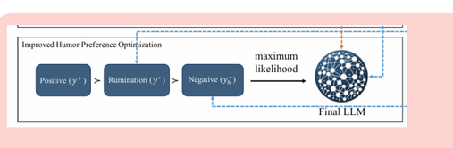
相关工作

- 双关语模型上
  - 然而，上述所有方法都是专门为小规模模型设计的。据我们所知，我们的工作是首次增强 LLM 的双关语生成能力。
- 偏好学习上
  - 直接偏好优化（DPO）（Rafailov等人，2023）通过配对偏好实现模型对齐，显著提升效率与稳定性；
  - 基于DPO和SLiC（Zhao等人，2022）改进的统计拒绝采样优化（RSO）（Liu等人，2023）采用拒绝采样实现更有效的优化
  - Xu等人（2023）则提出渐进式对比训练课程，通过从简单到复杂的偏好对逐步提升模型性能。
  - 本文工作创新
    - 我们创新性地提出多阶段课程学习框架和三元组偏好学习损失函数，专门针对双关语生成任务中的多目标偏好对齐稳定性进行优化

详细做法

- 问题定义
  - 同音异意
    - 我们系统的输入由双关词对构成，包含一个双关词（pw，如weak）及其对应替代词（aw，如week）
  - 同形异意
    - 在同形异义双关的情形下，当双关词在特定语境中具有逻辑连贯的双重含义时，我们遵循Tian等人（2022）的方法，采用相同表征形式（即设pw = aw）
  - 合格的生成结果需满足两个条件：符合双关结构（如包含双关词）且具有幽默性。
    - I lift weights only on Saturday and Sunday because Monday to Friday are weak days
      - 虚弱日和工作日
- 

  Pun Generation with Curriculum Learning (PGCL)
  - 蓝色虚线表示数据流
  - 黄色虚线表示参数流
- 方法
  - 1.多阶段"课程"学习（Curriculum Learning——从易到难学习样本）
    - 1.我们分两个阶段使用直接偏好优化（DPO）方法，以引导模型优化两个关键偏好：双关语结构和幽默。
      - 

        （1）结构偏好优化模块（上图），用于在第一阶段增强大语言模型满足双关结构要求的能力
      - 

        2）幽默偏好优化模块（下图），旨在第二阶段实现更具挑战性的幽默偏好对齐。
    - 2.结构首选项对齐
      - 对于输入pw和aw，通过模板转换为提示词x ③
      - DPO（直接偏好优化）是一种离线偏好优化技术，其输入为预先计算的正样本y⁺与负样本y⁻对，二者均对应相同提示x。我们将训练数据集中标注的双关语作为正样本y⁺，并利用监督微调大语言模型（如LLaMA2）生成负样本y⁻。具体实现包括三个关键步骤:
        1. 样本生成机制：随机生成提示x对应的双关语句，通过结构判别器验证其合规性:
           - 英语/中文同形异义双关（pw=aw）：仅需包含pw
           - 英语谐音双关（基于SemEval数据集设定）：须含pw且不含aw
           - 中文谐音双关：必须同时包含pw与aw
        2. 负样本筛选标准：若生成语句不符合上述结构规则，则采纳为负样本y⁻
        3. 偏好对构建：最终形成结构偏好对(y⁺,y⁻)用于执行结构DPO优化

        正负样本生成——不满足3条规则其中之一的样本做负样本
        - DPO——旨在增加正类可能性，降低负类的可能性
      - DPO损失

$$r^+(\theta) = \beta(\log \pi_\theta(y^+|x) - \log \pi_{\text{sft}}(y^+|x)) \quad (1)$$

$$r_s^-(\theta) = \beta(\log \pi_\theta(y_s^-|x) - \log \pi_{\text{sft}}(y_s^-|x)) \quad (2)$$

$$\mathcal{L}_{\text{structure-dpo}}(\theta) = -\log \sigma(r^+(\theta) - r_s^-(\theta)) \quad (3)$$

        - 参数介绍：其中β为温度超参数；πθ表示待优化的语言模型，πsft代表监督微调（SFT）语言模型；σ为sigmoid函数；r⁺和r⁻s为两个伪奖励函数，其作用类似于RLHF中的奖励机制。DPO算法需要从冻结的参考模型（即SFT模型）中获取正负样本序列的对数概率，并通过扩大r⁺与r⁻s之间的差异来提升模型性能。
          - 冻结模型：指在模型训练过程中固定其参数不更新的预训练模型，仅作为参考或特征提取器使用。
    - 幽默参数对齐
      - 使用与第一阶段相同的提示x，和正样例y+,并用第一阶段优化后的LLM生成候选句子
        - 判别器—— fine-tune RoBERTa-large model
          - 英语幽默判别器：
            - 基于Mittal等人（2022）方法
            - 使用ColBERT数据集（含20万条笑话/非笑话样本）微调RoBERTa-large模型（Liu等人，2019）
          - 中文幽默判别器：
            - 采用Chen等人（2023）的幽默识别数据集（18,709条笑话+7,709条非笑话）
            - 基于RoBERTa-large中文版（Cui等人，2020, 2021）微调
          - 3. 负样本筛选机制
            - 利用分类模型输出的概率值筛选幽默度不足的语句作为负样本y⁻ₕ
            - 最终构建幽默偏好对(y⁺, y⁻ₕ)用于训练
      - 损失函数

$$r_\phi^+(\theta) = \beta(\log \pi_\theta(y^+|x) - \log \pi_\phi(y^+|x)) \quad (4)$$

$$r_h^-(\theta) = \beta(\log \pi_\theta(y_h^-|x) - \log \pi_\phi(y_h^-|x)) \quad (5)$$

$$\mathcal{L}_{\text{humor-dpo}}(\theta) = -\log \sigma(r_\phi^+(\theta) - r_h^-(\theta)) \quad (6)$$

        - \pi_theta为待优化模型：其中，r_ϕ^+ 和 r_ϕ^- 为对应的伪奖励函数，π_ϕ 表示已在第一阶段完成结构优化的语言模型。
  - 改进的具备幽默偏好的DPO
    - 由于训练数据有限，多阶段偏好对齐面临灾难性遗忘问题——即第二阶段中结构偏好对齐的效果会逐渐衰减，而提出三元组幽默对齐DPO方法。（ triplet humor alignment DPO）
    - $(y^+, y_{h^*}^-, y^*)$.

      够减幽默偏好三元组
      - y*，4.1节要求——前面提到的三条规则
        1. 锚点样本y*生成
           - 使用阶段1结构优化后的LLMs，基于相同提示x生成候选句
           - 通过4.1节所述结构判别器筛选符合双关结构的语句（保留率≈78%）
           - 称此类样本为"反刍样本y*"（Rumination Sample），其特性包括：
             ✓ 严格满足pw/aw位置约束
             ✓ 未经过幽默度优化
      - y-
        2. 负样本y⁻ₕ*筛选
           - 从候选句中联合使用两个判别器筛选:
             - 结构判别器：排除不符合双关语法规则的语句
             - 幽默判别器：剔除幽默评分>0.7的语句（阈值经网格搜索确定）
           - 最终y⁻ₕ*需同时满足:
             ✓ 结构错误（如缺失pw）
             ✓ 幽默度低下（p(humor)<0.3）
      - y+：来自标注数据（结构&幽默双达标）
    - 损失函数

$$r_{h^*}^-(\theta) = \beta(\log \pi_\theta(y_{h^*}^-|x) - \log \pi_\phi(y_{h^*}^-|x)) \quad (7)$$

$$r^*(\theta) = \beta(\log \pi_\theta(y^*|x) - \log \pi_\phi(y^*|x)) \quad (8)$$

$$\mathcal{L}_{\text{I-humor-dpo}}(\theta) = -\log \sigma(r_\phi^+(\theta) - r^*(\theta)) - \log \sigma(r^*(\theta) - r_{h^*}^-(\theta)) \quad (9)$$

      - 参数解释：其中，r^* 表示反刍样本（rumination samples）的伪奖励值；r_ϕ^+ 的定义与公式4保持一致。该优化目标旨在实现双重差异最大化:
        1. 正样本与反刍样本的奖励差值
        2. 反刍样本与负样本的奖励差值

        通过上述损失函数优化，最终可获得同时对齐结构偏好与幽默偏好的大语言模型。

实验

- 数据集（187,315 words in total）

Dataset	SemEval phonic	SemEval graphic	ChinesePun phonic	ChinesePun graphic
Train Examples	879	1039	839	845
Test Examples	219	259	210	212
Total Examples	1098	1298	1049	1057

  - phoinc：谐音双关
  - graphic：同形异义双关
  - 我们发现大多数双关语语句只包含一个双关语对。因此，我们的研究将重点完全缩小到这种情况。
- 指标
  - 自动评估体系
    - structure success rate
      1. 结构成功率（Structure Succ.）
         - 定义：生成结果中正确包含给定双关词对（pw, aw）的比率
         - 计算方法:

$$\text{Structure Succ.} = \frac{\sum_{i=1}^{N} \mathbb{I}(\text{pw} \in y_i \wedge \text{aw} \in y_i)}{N}$$

         - 依据：遵循Sun等人（2022a,b）的双关结构评估框架
    - diversity
      2. 多样性指标
         - Dist-1/Dist-2:
           - 句子级：计算单句内unique 1-gram/2-gram占比
           - 语料级：统计全测试集unique n-gram的覆盖率
           - 公式:

$$\text{Dist-n} = \frac{\text{unique n-grams}}{\text{total n-grams}}$$

           - 基准方法：沿用Luo等人（2019）、Yu等人（2018）和Mittal等人（2022）的标准
    - Average sentence length
      3. 平均句长（Avg-Length）
         - 意义：反映生成结果的复杂度（过短可能缺乏语境，过长易偏离双关结构）
         - 计算：测试集所有生成句token数量的算术平均
  - 人工评估
    - 评估数据准备
      - 从测试集随机抽取50组双关词对（phonic/graphic各25组）
      - 每组词对生成:
        - 本模型输出
        - ChatGPT输出
        - 人类专家创作（作为黄金标准）
    - 感觉这部分通过机器结果参超下就行了
- 基准模型
  - AmbiPun
  - LLMs LLaMA2-7B and Baichuan2-7B
    - 源自HuggingFace Transformers库
  - ChatGPT
    - gpt-3.5-turbo版本
- 实验设置
  - 优化器
    - 优化器：AdamW（Loshchilov & Hutter, 2017）
    - 基础参数:
      - 学习率：5e-5
      - 批量大小：4
      - 微调策略：LoRA（Hu等, 2021），仅训练0.1%参数
  - DPO参数设置
    - DPO专项配置:
      - 温度系数β=0.5
      - 解码策略:
        - 温度（temperature）=0.95
        - top-p=0.95
        - top-k=5
  - 偏好数据规模
    - 偏好数据规模:
      - 偏好对/三元组样本量≥10,000组
      - 采样细则参见附录B
    - 结构偏好数据：如第4.1节所述，我们首先使用SFT阶段设计的专用提示模板，从监督微调（SFT）模型中采样生成初始输出。随后通过结构判别器对这些输出进行分类，将其划分为符合结构与不符合结构的两组语句。最终，我们随机选取10,000组结构偏好对（正样本：结构合规句；负样本：结构违规句）进行结构偏好对齐训练。
    - 幽默偏好数据
      - 数据采样与分类
        1. 数据采样与分类
           - 输入源：从已完成结构优化的模型中生成候选句
           - 双重判别:
             - 幽默判别器：筛选幽默度评分（p(humor) < 0.3为负样本）
             - 结构判别器：验证双关结构合规性（pw/aw位置约束）
           - 样本分类:
             - 反刍样本（Rumination）：结构合规但幽默不足
             - 负样本（Negative）：结构违规且缺乏幽默
      - 三元组构建
      - We randomly selected 10,000 triplets for humor preference alignment.
  - 实验环境
    3. 实验环境
       - 开发框架：基于LLaMA-Factory构建
       - 硬件配置：2×RTX 3090（24GB显存/卡）
       - 其他参数：保持库默认设置
- 实验结果

Dataset	Model	Avg-Length	Corpus-Div % Dist-1	Corpus-Div % Dist-2	Sentence-Div % Dist-1	Sentence-Div % Dist-2	Humor A/B-test vs. ChatGPT % Win	Humor A/B-test vs. ChatGPT % Lose	Structure Succ. %	Pun Succ. %
ChinesePun	ChatGPT	53.94	**5.51**	**41.88**	**77.87**	**95.17**	—	—	**93.07**	28.00
	AmbiPun	79.31	2.68	15.50	64.50	85.06	28.00	42.00	53.84	16.00
	Baichuan2-7B	91.12	3.00	26.92	70.84	92.36	18.00	44.00	78.69	14.00
	$Baichuan2_{sft}$	85.57	3.16	30.94	62.61	85.27	20.00	46.00	74.30	16.00
	$Baichuan2_{dpo}$	79.26	3.22	30.25	60.04	82.74	30.00	52.00	80.41	20.00
	PGCL (ours)	85.20	2.78	27.52	58.73	81.39	**64.00**	**22.00**	89.10	**44.00**
	Human	88.94	1.85	30.19	65.92	89.13	—	—	87.01	—
SemEval	ChatGPT	16.35	24.29	68.69	92.64	99.78	—	—	91.42	34.00
	AmbiPun	14.0	21.50	64.96	91.15	98.75	45.00	48.00	94.12	30.00
	LLaMA2-7B	59.95	9.56	32.72	79.99	98.46	32.00	36.00	89.12	22.00
	$LLaMA2_{sft}$	11.71	24.84	66.49	**95.84**	99.76	32.00	38.00	84.52	22.00
	$LLaMA2_{dpo}$	43.65	13.90	49.31	89.36	98.09	42.00	34.00	94.56	28.00
	PGCL (ours)	22.51	20.09	61.64	94.26	99.19	**68.00**	**14.00**	**98.95**	**56.00**
	Human	11.66	**31.76**	**79.05**	95.36	**99.86**	—	—	86.82	—

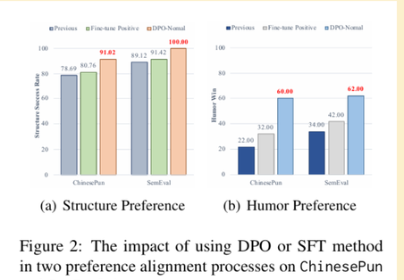
Table 4: The results of pun generation. The boldface denotes the best performance and the underline denotes the second-best performance among systems. The paired t-test shows that the difference between our model and the baseline methods is statistically significant ($p < 0.05$).

  - 性能突破
    1. 性能突破
       - 在中英文双关数据集上均实现SOTA性能:
         - 双关成功率提升：中文ChinesePun（+16%）、英文SemEval（+22%）
         - 幽默度与结构成功率持续领先所有基线模型
  - 多样性（Div）ChatGPT显著优于其他模型
    - 其他模型都在特定数据集微调过
  - 模型的dpo版本性能要由于sft(指令微调版本)
    - 这证实了偏好对齐方法相比指令微调能更有效地帮助大语言模型掌握特定能力（如双关结构或幽默生成）
  - 与基准模型的DPO方法相比，本方法最重要的改进在于采用多阶段课程学习策略——通过"由易到难"的分阶段方式独立优化两个偏好目标（结构→幽默），该设计的有效性得到了实验数据的充分验证。
- 消融实验
  - 评估PGCL各组成部分对其的影响——实验其与各变体比较
    - w/o Improved Humor DPO——使用标准的DPO算法而非改进后的
    - w/o Humor——移除幽默对齐阶段（第二阶段），仅学习结构偏好
  - 结果

Dataset	Model	Avg-Length	Corpus-Div % Dist-1	Corpus-Div % Dist-2	Sentence-Div % Dist-1	Sentence-Div % Dist-2	Humor A/B-test vs. ChatGPT % Win	Humor A/B-test vs. ChatGPT % Lose	Structure Succ. %	Pun Succ. %
ChinesePun	PGCL (ours)	85.20	2.78	27.52	58.73	81.39	**64.00**	**22.00**	89.10	**44.00**
	w/o Improved Humor DPO	86.94	2.30	22.96	55.51	78.39	60.00	32.00	80.46	38.00
	w/o Humor	88.28	**2.82**	**27.73**	**59.94**	**83.06**	22.00	48.00	**91.02**	26.00
SemEval	PGCL (ours)	22.51	20.09	**61.64**	94.26	**99.19**	**68.00**	**14.00**	98.95	**56.00**
	w/o Improved Humor DPO	29.15	20.68	59.55	93.02	97.64	62.00	24.00	91.64	46.00
	w/o Humor	17.39	**24.11**	61.39	**95.45**	98.76	34.00	44.00	**100.00**	26.00

Table 5: Architecture ablation analysis on SemEval and ChinesePun dataset

    - 与第二阶段使用标准DPO算法微调的LLM相比，PGCL的结构成功率分别高出11.02%和8.36%。这表明，标准DPO算法在进行幽默偏好对齐时，难以保持对结构偏好的良好维持。
    - 新颖的三元组损失带来显著性能提升（对比PGCL与PGCL w/o Improved Humor DPO）：
      - 在两个数据集上的结构成功率分别提升约8.64%和7.31%；
      - 双关结构满意度的改进进一步使双关成功率提升约6%和10%。

      新的三元组对齐损失能有效指导大语言模型（LLMs）在结构偏好与幽默偏好之间保持平衡。
- 相关前期探索
  - 

    (a) Structure Preference (b) Humor Preference

    Figure 2: The impact of using DPO or SFT method in two preference alignment processes on ChinesePun and SemEval datasets.

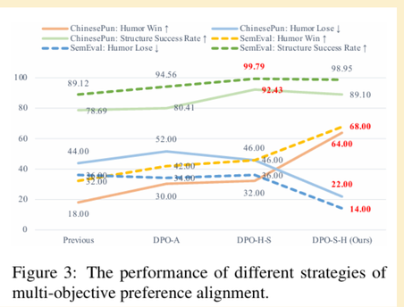
    DPOvs. instruction tuning
    - 结论：相较于指令微调，DPO方法能显著提升大语言模型的性能。
  - 

    Figure 3: The performance of different strategies of multi-objective preference alignment.

    不同策略在提升模型多目标偏好对齐能力上的性能表现（基于中文双关语数据集和SemEval数据集）
    - 多阶段课程学习使LLM能专注攻克困难任务（如幽默生成）而不牺牲整体性能
    - 而DPO-A因双关语生成的固有难度，模型难以直接同步优化，性能提升有限。