

小红书场景下的NoteLLM：探讨LLM如何助力提升I2I召回效果



SmartMindAI

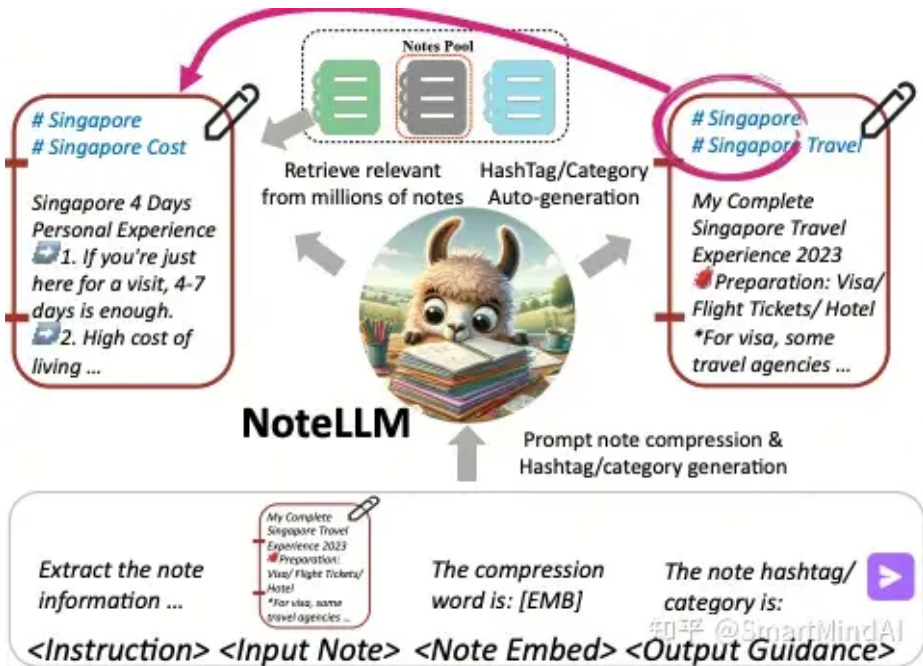
专注搜索、广告、推荐、大模型和人工智能最新技术，欢迎关注我

已关注

40 人赞同了该文章

Introduction

社交平台如小红书和Lemon8因其鼓励用户分享UGC而受欢迎。个性化笔记推荐能提升用户参与度，尤其在（I2I）推荐中从海量笔记库找出用户感兴趣内容是非常关键的。当前方法虽用BERT处理嵌入相似度，但未充分利用笔记的标签信息。



研究表明，笔记标签如"#新加坡"揭示了内容关联性，类似嵌入生成过程。因此，利用大型语言模型⁺（LLMs）来学习生成标签和分类能优化嵌入质量。然而，这方面研究将LLMs应用到I2I推荐中还较少。

本文提出NoteLLM框架，利用LLMs改进I2I推荐，有望从大量笔记中精准推荐相关内容。

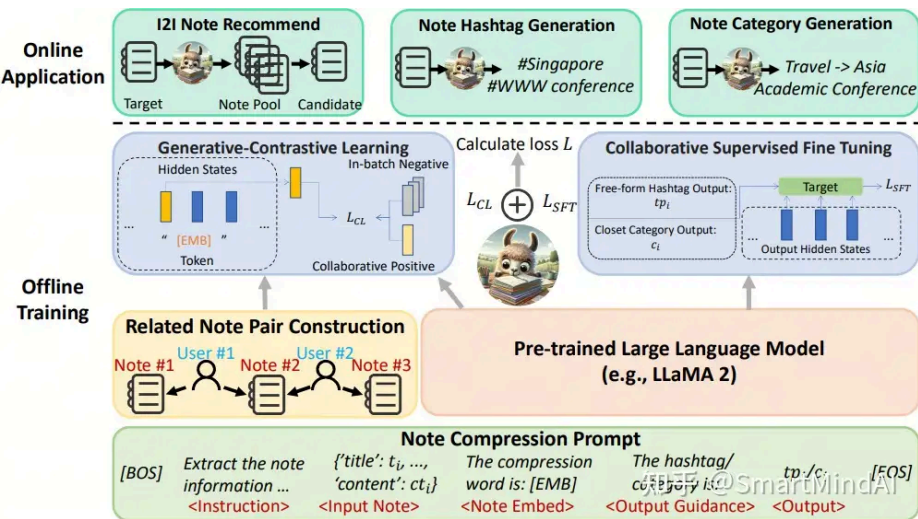
1. 我们构建了NoteLLM模型，结合了个性化笔记推荐和生成标签的任务，以充分利用信息。

- 4. 通过加入GCL和CSFT，增强了笔记嵌入的深度和相关性。
- 5. 研究首次探索了如何利用LLMs在I2I推荐中的作用，填补了现有技术的空白。

Methodology

Framework of NoteLLM

我们设计了Note Compression Prompt (NCP)，这是一种策略，用于结合I2I推荐和标签/类别生成任务。首先，我们利用NCP来整合了推荐系统*所需的协同信号和语义信息*。然后，我们通过将这些压缩后的提示输入LLMs进行处理。



GCL (Generative Contrastive Learning) 在这个过程中发挥了关键作用，它利用生成的压缩词的隐藏状态进行对比学习，从LLMs的输出中捕获协同信息。CSFT (Category-Specific Feature Transfer) 则利用笔记的语义和协同特征，由LLMs自动生成标签和类别，确保推荐和分类的准确性。

Note Compression Prompt

我们构建了'Note Compression Prompt' (NCP)，这是一种创新方法，结合了I2I推荐和标签/类别生成任务。在NCP中，我们首先压缩'*NoteContent*'，用'\<compress>'作为标志，然后明确指明任务类型，如'*TaskType*'为'HashTag/Category'。

word is: "[EMB]". <Output Guidance> <Output>[EOS]

In this template, [BOS], [EMB], and [EOS] are special tokens, while <Instruction>, <Input Note>, <Output Guidance>, and <Output> are placeholders replaced by specific content. The specific content for category generation is defined as follows:

Note Compression Prompt for Category Generation.

<Instruction>: Extract the note information in json format, compress it into one word for recommendation, and generate the category of the note.

<Input Note>: {'title': t_i , 'topic': tp_i , 'content': ct_i }.

<Output Guidance>: The category is:

<Output>: c_i

The template for hashtag generation is presented below:

Note Compression Prompt for Hashtag Generation.

<Instruction>: Extract the note information in json format, compress it into one word for recommendation, and generate <j> topics of the note.

<Input Note>: {'title': t_i , 'content': ct_i }.

<Output Guidance>: The <j> topics are:

<Output>: <j> topics from tp_i

知乎 @SmartMindAI

这种方法不仅利用LLMs的语义信息，还通过协作信号的整合，提高了推荐系统的精准度和适应性，确保了推荐结果的多样性和质量。

$$s_{n_A \rightarrow n_B} = \sum_{i=1}^U \frac{1}{N_i}$$

在共现分数集 $S_{n_A \rightarrow n_B}$ 中，我们计算从笔记 n_A 到所有笔记 n_B 的关联分数 $S_{n_A \rightarrow n_B}$ ，其中 S 是基于用户行为数据 U 和用户行为记录 N_i 。我们设定阈值 u 和 l 以排除异常点击，仅保留分数在区间内的笔记对。接着，我们从这些笔记中选取与笔记 n_i 相关性最强的 t 个作为它的相关笔记集 S_{n_i} 。

在LLMs的语义和协作信号处理阶段，我们不直接使用聚合词，而是构建虚拟词。通过LLMs的自回归特性，我们提取每个相关笔记的最后一个隐状态作为其语义和协作信息的代表，将其投影到维度为 d 的嵌入空间，得到嵌入向量 n_i 。这里 n_i^+ 表示与笔记 n_i 相关的另一个笔记的嵌入。在每个小批量中，包含 $2B$ 个笔记，包括两对相关笔记。GCL的损失函数⁺基于这些嵌入，通过比较 n_i 和 n_i^+ 来评估它们的相似性，以优化模型对笔记相关性的理解和推荐。具体的损失计算方式如下：

$$L_{cl} = -\frac{1}{2B} \sum_{i=1}^{2B} \log \frac{e^{sim(n_i, n_i^+) \cdot e^{\tau}}}{\sum_{j \in [2B] \setminus \{i\}} e^{sim(n_i, n_j) \cdot e^{\tau}}}$$

在GCL（相关性学习）中， L_{cl} 是损失函数，它衡量的是笔记嵌入之间的相似性。通过LLMs，我们获取每个相关笔记的嵌入向量，具体是取每个笔记最后一个隐状态的向量，如 n_i 和 n_i^+ 。计算方式是使用余弦相似度⁺

$$sim(n_i, n_i^+) = \frac{n_i \cdot n_i^+}{\|n_i\| \|n_i^+\|}$$

这个过程温度参数 τ 是一个可调整的超参数⁺，它控制了相似度的软化程度。高 τ 可能导致更平滑的相似度分布，而低 τ 则可能保持更硬的判别。GCL的目标是优化模型，使其能准确识别不同笔记间的真正相关性，从而提升推荐的精准性和用户满意度。

Collaborative Supervised Fine-Tuning

$$L_{cft} = -\sum_{(n_A, n_B)} \log P_{\theta}(c|n_A, n_B; \tau) + \lambda \cdot sim(c, c_{gt})$$

这个概率由模型输出的软化相似度决定，即

$$\text{sim}(\mathbf{n}_i, \mathbf{n}_i^+) / (\|\mathbf{n}_i\| \|\mathbf{n}_i^+\|)$$

其中 θ 是模型参数。其次，我们引入了一个正则化项 λ ，它控制了生成标签的准确度与真实标签 \mathbf{c}_{gt} 的相似度。这个项旨在鼓励模型生成更接近真实标签的结果，但同时避免过度拟合。总的来说，CSFT通过优化这个损失函数，提升了LLMs在推荐任务中的标签生成能力和对用户兴趣的捕捉能力，从而提高推荐质量。

$$L_{gen} = -\frac{1}{T} \sum_{i=1}^T \log(p(o_i | o_{<i}, i))$$

$$L = \frac{L_{cl} + \alpha L_{gen}}{1 + \alpha}$$

$$L_{total} = L_{csft} + \beta \cdot L_{task_specific}$$

其中 L_{total} 是NoteLLM总损失，它将协同标注学习（CSFT）的泛化能力与针对特定任务如I2I推荐或标签生成的额外任务损失 $L_{task_specific}$ 相结合。参数 β 控制了对任务特定损失的权重，确保模型既能有效进行多笔记推荐，又能精确地生成标签和类别信息。通过这种加权优化，NoteLLM在笔记推荐领域实现了语义理解和目标标签生成的双重目标，提升推荐质量和用户满意度。

Experiments

Dataset and Experiment Setting

我们在小红书产品数据集上进行了离线实验，以确保模型训练的均衡。我们按照每周产品数据，从每个类别中选取固定数量的笔记对作为训练样本，以避免过拟合⁺。测试集则来自未来一个月的新数据，但排除了已出现在训练集中记录的笔记。

Table 1: Detailed statistics of training and testing dataset.

training dataset			
#notes	458,221	#note pairs	312,564
avg. #words per title	11.54	avg. #hashtag per note	3.02
avg. #words per hashtag	4.19	avg. #words per content	47.67
testing dataset			
#notes	257,937	#note pairs	27,999
avg. #words per title	13.70	avg. #hashtag per note	5.49
avg. #words per hashtag	4.53	avg. #words per content	182.45

实验中，我们选用Meta LLaMA 2作为基础LLMs。构建相关笔记对时，我们设置了共现得分阈值：上限 $u = 30$ ，下限 $l = 0.01$ ，并且取 $t = 10$ 。笔记嵌入维度我们设置为 $d = 128$ 。批量大小为 $B = 64$ 。通过这样的设置，NoteLLM能够在进行多笔记推荐的同时，通过学习笔记的语义和生成目标标签，实现对推荐质量和多样性的优化。我们在离线环境下评估无监督推荐模型，以全面理解其性能。我们使用全类别输入作为提示，针对每对笔记，以第一条笔记为目标，另一条作为真实标签。我们对测试集中的笔记按顺序排序，不包括目标笔记，通过Recall@100、Recall@1k等指标考察推荐的一对一准确性。

对于最接近类别的生成，我们使用精度(Acc.)和错位比例(III.)来度量，III.表示误分类的比例。而对于自由形式的标签生成，我们使用BLEU4、ROUGE1、ROUGE2和ROUGEL来评估模型⁺的生成质量。这些指标帮助我们评估模型在生成标签时的精确性和多样性。

Offline Performance Evaluation

1. 文本到文本 (I2I) 推荐：这类方法主要基于笔记内容的理解，包括内容基础推荐 (CBC)、话题基础推荐 (TBR) 和基于语义相似性推荐 (SRR)，它们通过比较笔记内容进行推荐。



3. 非LLM策略：还有一些方法不直接依赖LLMs，而是运用传统模型，如TF-IDF和词嵌入，进行相似性计算。

Table 2: Performance of different methods in I2I recommendation tasks (%).

	Model Size	Recall@100	Recall@1k	Recall@10k	Recall@100k	Avg.
LLaMA 2 zero-shot	7B	11.94	19.44	32.53	68.81	33.18
PromptEOL zero-shot [10]	7B	55.27	74.47	88.71	98.04	79.12
SentenceBERT (Online) [36]	110M	70.72	87.88	96.29	99.62	88.63
PromptEOL+CSE [10]	7B	83.28	95.26	99.20	99.96	94.43
RepLLaMA [26]	7B	83.63	95.10	99.27	99.94	94.49
NoteLLM	7B	84.02	95.23	99.23	99.96	94.66

实验将证明，我们的（相关性学习）和CSFT（协同标注学习），能提供更优质的嵌入，并在 Recall@系列指标上展现出优异性能。同时，我们会评估在最接近类别生成和自由形式标签生成任务上，。

研究表明，尽管零样本方法有潜力，但它们在笔记推荐任务中并未超越微调方法，显示出特定领域知识的重要性。LLaMA 2相对于SentenceBERT有明显优势，证明大型语言模型理解笔记内容的能力更强。使用提示的PromptEOL+CSE在提升零样本检索表现上与无提示的RepLLaMA相当，但微调后此优势减弱。我们的NoteLLM（未提及具体名称）通过CSFT整合总结能力，通过压缩高质量笔记嵌入，显著超越了其他LLMs方法，证明了有效利用任务相关知识的重要性。

Effect on Different Exposure Notes

我们根据笔记的曝光程度将其分为低曝光（不足1,500次，30%，0.5%累积）和高曝光（超过75,000次，10%，75%总体）。通过对比，我们计算了召回率。结果显示，CSFT模块具有普遍适用性，不受曝光程度影响。

Table 3: Performance of different methods for low exposure notes and high exposure notes in I2I recommendation tasks (%).

	Low Exposure		High Exposure		Overall	
	Recall@100	Recall@1k	Recall@100	Recall@1k	Recall@100	Recall@1k
SentenceBERT (Online) [36]	75.00	90.54	59.03	81.91	70.72	87.88
PromptEOL+CSE [10]	86.28	96.63	72.46	91.40	83.28	95.26
RepLLaMA [26]	86.54	96.18	72.64	91.37	83.63	95.10
NoteLLM	87.85	96.63	73.46	91.26	84.02	95.23

尤其值得注意的是，尽管在处理低曝光笔记时表现优秀，但当面对高曝光笔记时，其他方法性能有所下滑。这揭示了模型对流行度偏见的适应性-----它能更准确地针对内容而非数量，对冷启动笔记的检索尤其有利。这鼓励用户增加新笔记的发布，以丰富社区内容。

Case Study

知乎

(a)	<div>Category: Dressing Hashtag: #SummerOutfit #Cleanfit#SimpleOutfit #Minimalism Content: ① Tight clothes...</div>	<div>Category: Dressing Hashtag: #SummerOutfit Content: Share I'm here to share my May outfit collection ☺</div>	<div>Category: Dressing Hashtag: #summerwea Content: Summer outfit guide, a total of 96 looks! 🌟 ...</div>	<div>Category: Dressing Hashtag: #summerwea Content: Summer outfit guide, a total of 96 looks! 🌟 ...</div>	<div>Category: Dressing Hashtag: #Simple living #Streamlined living #Cleanfit Content: Why downsize your wardrobe? One is that I hope I won't buy clothes randomly...</div>	
	<div>Query Note Title: Don't let strangers into your own rabbit!! Category: Toys Hashtag: #Keep strangers away Content: I went to the mall and bought two rabbits toy...</div>	<div>SentenceBERT(online) Title: Raising rabbits !! Category: Rabbits Hashtag: #Rabbitdaily Content: Changed to a larger cage and kept cleaning it 2-3 times a day. With rabbit food ...</div>	<div>PromptEOL Title: How to raise rabbits? Category: Rabbits Hashtag: #Rabbit Content: It's a very small one, it feels like it's just one month old...</div>	<div>RepLLaMA Title: Mother Rabbit plucked all her fur, maternal love is the greatest Category: Rabbits Hashtag: #petRabbit Content: Pregnant mother rabbit pluck...</div>	<div>NoteLLM Title: Keep strangers away from the circus sell or change Category: Toys Hashtag: #Keep strangers away #circus performer #blind box Content: Brand new and unopened rabbit toy box for sell...</div>	
	(b)					
<div>Note Title: Does anyone know about abandoned places in Huicheng? Content: Does anyone know about abandoned places in Huicheng, such as abandoned factories and unfinished buildings, and would like to take photos? RedHashtag (online) #Summer job #summer job entering the factory #summer job for college students #working in Chongqin NoteLLM #Photography #street sweeping #travel #Yuzhong District #Chongqing. GT #Photography #Huizhou</div>						
(c)						
(d)						
	<div>Note Title: Xu Wei's Endless Light Chongqing Concert on 5.20 Content: There is no utopia for young people, and it becomes clear that the heart is far away. RedHashtag (online) #Music live sharing #Cool LIVE scene #Band performance #Livehouse NoteLLM #Xu Wei #Xu Wei Endless Light #Chongqing #Chongqing performance #Chongqing concert GT #Xu Wei Concert #Xu Wei Endless Light National Tour</div>					
	(d)					

最后部分通过实例展示了CSFT在推荐和生成任务中的表现。图(a)显示，相较于基线推荐，，体现了其对实际需求的理解。图(b)说明了CSFT能理解并避免误解上下文，特别是在rabbit的案例中。图(c)和(d)则展示了在标签生成任务中，（如factories）并生成有针对性的标签，但同时也存在一些幻觉问题，这突显了其在理解和生成任务中的优势以及未来可能的改进方向。

Online Experiments

我们在小红书进行了为期一周的在线I2I推荐实验，使用NoteLLM模型替代SentenceBERT，结果显示，我们的方法显著提升了16.20%的点击率⁺。同时，推荐的召回性能增强，带来了1.10%的评论增长和0.41%的WAP数量提升，这证实了LLMs对提升推荐质量和用户互动的积极作用。

尤其值得注意的是，新笔记在冷启动阶段评论增长了3.58%，显示出LLMs在处理这类非活跃内容时的泛化能力提升。目前，我们已将NoteLLM模型成功应用于小红书的I2I笔记推荐任务中。

发布于 2024-04-23 10:38 · IP 属地北京

LLM 小红书 推荐系统



理性发言，友善互动



还没有评论，发表第一个评论吧