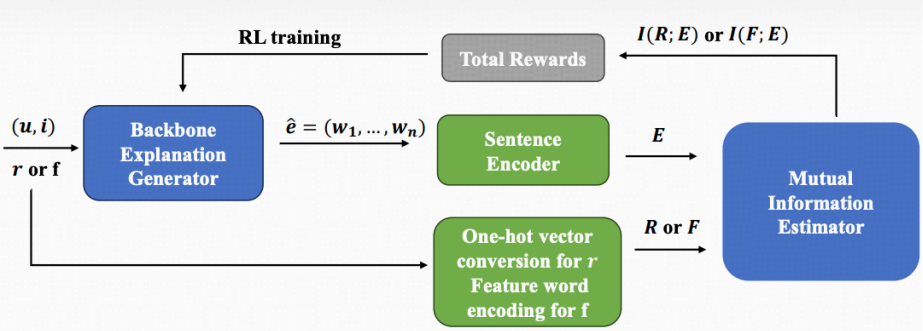


美团2024：个性化推荐理由 —— 让每次选择都有理由

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

Introduction

生成基于自然语言的推荐理由在近年来受到了广泛关注。一系列研究显示，在推荐平台上提供理由具有潜在益处，如提高推荐接受率、增加用户满意度和信任度。为了生成流畅且个性化的理由，大多数研究利用现有用户评论作为基准，并采用最大似然估计+（MLE）方法训练模型，以生成与用户评论相似的理由。尽管如此，这样的做法在传统文本生成指标（如BLEU、ROUGE）方面有潜力生成高质量文本，但在满足推荐平台上的用户对理由的额外要求（如信任和有效性）方面遇到了挑战。

理由作为推荐内容的辅助信息+，旨在帮助用户在平台上做出更加明智的决策。一个合格的理由应具备以下两个属性：

与预测评级的一致性：理由应支持推荐者预测的评级，因为它可以帮助用户理解特定推荐背后的逻辑。如果理由传达的情感与预测评级之间存在明显的不一致（例如，对于评分2/5星的内容，理由说"装饰很好，员工友好"），这将误导用户做出错误的决策，并增加他们对推荐平台的怀疑。

与内容特征的一致性：理由需要包含与推荐内容特定特征或方面相关的高度具体信息。通用理由如"美食和服务都很好"不足以让用户获得关于推荐内容的详细知识，因此不能帮助他们决定是否接受或拒绝相应的推荐。

用户评论中的噪音和MLE训练目标的平均性质，使得仅仅模仿用户评论不足以实现与预测评级的一致性和与内容特征的一致性这两个目标。这种倾向会加剧理由与评分和特征一致性的问题。一方面，数据集中的评论大多情感上是积极的，即使实际预测的评级较低，现有的理由生成器+也会不断生成积极的句子。另一方面，数据集中内容特征的分布不均，理由生成器因此倾向于提及一些常见但不那么具体的特征（例如，Yelp数据集中的食物和服务）。直观地利用用户评论作为理由具有局限性，表中展示了两种对比案例：具有高度对齐属性的理由案例与被认为在大多数先前工作中表现完美的用户评论案例进行比较。这些理由在自然语言生成（NLG）指标上对用户评论数据集具有高价值。

	Predicted Rating	Explanation perfectly reconstructing the review	Explanation aligning with predicted rating/item feature
Case A	2.0	the staff was very good	the sauce was **bland** and the texture was **too thick**
Case B	3.0	it was okay .	the **prices** are very reasonable

在案例A中，平台预测用户将给予商品低评分，暗示不满。然而，与评论高度相似的理由却与预测的负面情绪相矛盾。相反，与预测评分更好的对齐理由则揭示了商品的令人失望之处，如酱汁平淡寡味，质地过厚。这种理由更有可能帮助用户理解推荐平台的预测评分背后的原因，从而可能增加用户对推荐平台的感知透明度。

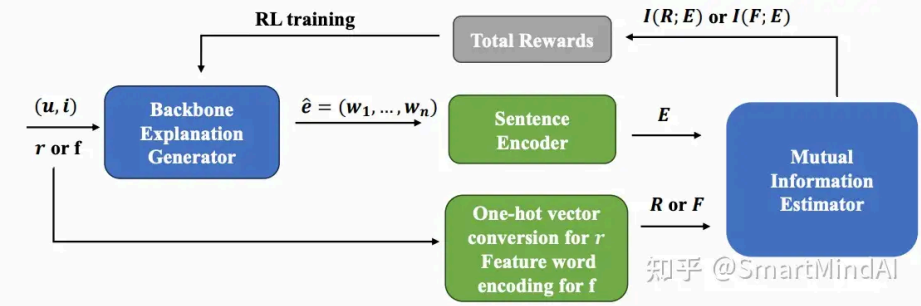
在案例B中，复制对应评论的理由提供了一般性声明，缺乏对商品特性的具体描述。同时，与商品特性更好的对齐理由则强调了用户在决定接受或拒绝推荐时可能认为有价值的特定商品特性（价

知乎

为了解决上述限制，我们提出了一种模型无关的通用MMI框架，用于加强当前理由生成模型的对齐能力。最大互信息作为两个变量之间相互依赖的主要度量，我们利用它来衡量理由与预测评级或内容特征之间的对齐程度。MMI框架包括以下特点：

1. 神经MI估计器：用于估计基于文本的理由与预测评级/内容特征之间的对齐。
2. 基于强化学习的微调过程：将现有的MLE训练的理由生成模型作为骨干，并通过MI估计器输出的基于MI的奖励对其进行微调。为了防止潜在的奖励作弊并保持骨干模型模仿用户评论的能力，我们还整合了KL散度和熵作为正则化器，以使微调生成器在与评级/特征的对齐能力和生成流畅、自然、类似于用户评论的文本之间取得良好的平衡。

Preliminary



Generating Explanation for Recommendation

当前的理由生成方法被我们分为事后理由生成器和多任务学习模型。

Post-hoc Generation

事后理由生成器，例如，仅致力于为给定的用户-内容对(u, i)生成理由，同时包含额外属性，如评分或内容预定义特征。它们通常采用Seq2Seq模型架构，将给定的属性对$A = (a_1, a_2, \ldots, a_n)$作为输入，并使用负对数似合度（NLL）损失来最大化在给定属性A条件下生成真实评价e的概率。

Multi-task Learning

多任务学习模型同时进行评分与理由生成。模型的训练目标旨在最小化预测评分$\hat{r}$与实际评分r之间的均方误差，并基于真实评论e使用相同的负对数似然（NLL）损失来生成理由。

$$L_e = -\sum_{w \in e} \log \hat{s}(w)$$

其中$\hat{s}$是预测的词分布于词汇集合。

Mutual Information and its Estimation

Mutual Information

互信息是衡量两个随机变量基于熵的依赖性的指标。对于X和Y这两个随机变量，它们之间的互信息定义如下：

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = I(Y;X)$$

其中信息熵$H(X)$是X的香农熵，条件熵$H(X|Y)$是在给定Y的情况下X的条件熵。因此，互信息度量了在Y给定的情况下X的不确定性降低。直观上X和Y之间的高互信息值意味着X和Y之间的依赖性更强，因为知道Y会降低X的不确定性。这样的特性激发我们通过最大化互信息（MI）来模拟理由与评分或特征的匹配度，进一步强化这种关系。

知乎

MI的定义等效地通过KL离散表示为两个变量X和Y的联合分布与边缘分布P_X和P_Y的乘积之间的差异：

$$MI = D_{KL}(P(X,Y)||P_X \cdot P_Y)$$

$$I(X;Y) = D_{KL}(\mathbb{P}_{XY}||\mathbb{P}_X \otimes \mathbb{P}_Y)$$

其中，我们可以看出直接计算MI是不可行的，因为我们通常只能获取样本，而无法直接访问底层分布。因此，近期的研究工作结合了不同的MI变分界限以及深度学习，以实现MI的可微和可计算估计。在本工作中，我们采用了一种最先进的方法，名为神经估计器互信息（MINE），来估计两个给定变量X和Y之间的MI。

MINE的核心思想是通过以下唐斯克-瓦拉达汉界限推导MI的下界：

$$D_{KL}(P||Q) \geqslant \sup_{T \in F} \mathbb{E}_P[T] - \log(\mathbb{E}_Q[e^T])$$

通过将等式（3）和（4）结合起来，并选择深度神经网络⁺参数$\theta \in \Theta$参数化⁺的一组函数 $T_\theta : X \times Y \to R$

作为F，MINE定义了以下真实互信息的下界：

真实互信息的下界 = 定义

$$I(X;Y) \geqslant I_\theta(X;Y) = \sup_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}_{XY}}[T_\theta] - \log(\mathbb{E}_{\mathbb{P}_X \otimes \mathbb{P}_Y}[e^{T_\theta}])$$

其中T_θ在 MINE 中被称为统计模型⁺。它接受两个变量X, Y作为输入，并输出一个实数值。公式中的期望值是通过从联合分布$\mathbb{P}_{\mathbf{XY}}$和分别的边缘分布⁺$\mathbb{P}_{\mathbf{X}}$和$\mathbb{P}_{\mathbf{Y}}$中获取的经验样本来估计的。 直观上，下界值（lower bound value）越高，真实 MI 的估计就越准确。这意味着我们可以将下界值视为优化目标，并采用常见的梯度下降方法（gradient descent method），如 SGD，来迭代更新统计模型T_θ。一旦统计模型T_θ收敛，我们就可以使用它来推导 MI 的估计值。

MMI framework

在我们提出的最大互信息（MMI）框架中，我们以任意预先训练的理由生成模型为起点，将其称为主干理由生成模型。这个模型通过最大似然估计（MLE）在评论数据上进行训练，因此它具有生成类似于用户评论的文本的强大能力。我们的目标是通过微调来进一步增强其对齐能力。这个微调框架的核心理念是，通过估计理由与评分/特征之间的互信息（MI）作为度量标准，来衡量当前生成的理由与评分/特征之间的关系。由于估计的MI值是非可微的，将其视为奖励，利用强化学习（RL）来指导主干理由生成模型学习更好的对齐，是自然的选择。此外，为了保持主干模型能够模拟用户评论的能力，我们还引入了KL散度奖励和熵奖励，以补偿仅优化的MI值导致的文本质量的下降。

RL for Fine-tuning Backbone models

在基于RL的理由生成模型中，主干模型被视为一个代理，其行动是在位置t上，根据前一个位置$t-1$上的单词序列$w_{1:t-1}$，生成下一个单词w_t。生成概率$p_\theta(w_t|w_{1:t-1})$代表了一个随机策略⁺。我们定义了一个自定义的奖励 $\pi_{\hat{e}} = \pi_{w_{1:T}}$

在生成序列结束时，其中T是预定义的生成句子的最大长度。生成器θ的优化目标是最大化总奖励的期望值，这产生了以下损失函数⁺：

$$\begin{aligned} L_{RL} &= -\sum_{\hat{e}} p_\theta(\hat{e})\pi(\hat{e}) \\ &= -\sum_{\hat{e}} \prod_{t=1}^{T-1} p_\theta(w_{t+1}|w_{1:t})\pi(\hat{e}) \end{aligned}$$

我们采用策略梯度⁺方法来达成上述优化目标。

知乎

MI Reward for Enhancing Alignment

为了加强理由与评分或特征的对齐，我们提出了一种基于互信息（MI）的奖励机制。MI奖励$\pi_{MI}(\hat{e})$的计算如下：

1）使用句子编码器将理由生成器生成的样本$\hat{e}$转换为句子嵌入E。

2）对于与评分任务的对齐，我们将5级评分分数转换为5维的独热向量；对于与特征任务的对齐，我们将预定义的内容特征单词f编码为单词嵌入F。

3）我们将E与R或F的拼接作为MI估计器的输入，估计器的输出将是MI奖励。对于与评分任务的对齐，我们计算$I(R;E)$；对于与特征任务的对齐，我们计算$I(F;E)$。

如公式(5)所示，我们计算与评分任务对齐的MI奖励$\pi_{MI}(\hat{e})$为：
$\pi_{MI}(\hat{e}) = I(R;E)$ 或 $I(F;E)$

我们采用MINE作为互信息估计器，用θ_{MIR}表示MINE的统计模型用于$I(R;E)$，用θ_{MIF}表示用于$I(F;E)$。

$$\pi_{MI}(\hat{e}) = I_{\theta_{MIR}}(R;E) = \mathbb{E}_{\mathbb{P}_{RE}}[T_{\theta_{MIR}}] - \log(\mathbb{E}_{\mathbb{P}_R \otimes \mathbb{P}_E}[e^{T_{\theta_{MIR}}}])$$

与特征任务对齐的MI奖励$\pi_{MI}(\hat{e})$是：

$$\pi_{MI}(\hat{e}) = I_{\theta_{MIF}}(F;E) = \mathbb{E}_{\mathbb{P}_{FE}}[T_{\theta_{MIF}}] - \log(\mathbb{E}_{\mathbb{P}_F \otimes \mathbb{P}_E}[e^{T_{\theta_{MIF}}}])$$

我们通过两种策略确保MI奖励模型能够为生成器提供指导：1）在数据集的训练部分，我们预先训练MI估计器，将用户评论作为E，真实评级和特征作为R和F。2）在GAN的框架下，我们交替更新奖励模型和生成器，以提升奖励模型捕捉新输出样本的性能。

KL and Entropy Reward for Regularization

原始策略指的是基础模型的预训练版本，新策略是经过微调的，KL奖励$\pi_{KL}(\hat{e})$被定义：

$$\pi_{KL}(\hat{e}) = -D_{KL}[q(\hat{e})||p_\theta(\hat{e})]$$

其中$q(\hat{e})$代表预训练版本的主干模型生成当前理由$\hat{e}$的概率。通过最大化KL散度奖励，我们可以减少微调模型与预训练模型之间的偏差，确保微调后的策略具有安全的基准。

此外，为了进一步增加生成结果的多样性，促进RL训练期间的更好探索，我们将熵奖励作为另一个正则化目标添加。熵奖励$\pi_{Entropy}$的计算通常涉及对策略π的熵值[+]的计算，这是强化学习中常用的一种衡量策略多样性的方法。

$$\pi_{Entropy}(\hat{e}) = H(\hat{e}) = -\sum_{w_t \in \hat{e}} p_\theta(w_t|w_{t-1}) \log p_\theta(w_t|w_{t-1})$$

最后，总奖励是加权的 MI（互信息）、KL（Kullback-Leibler 散度）和熵的和：

$$\pi(\hat{e}) = \pi_{MI}(\hat{e}) + \alpha\pi_{KL}(\hat{e}) + \beta\pi_{Entropy}(\hat{e})$$

Dynamic Weighting Mechanism for Multi-objective Rewards

根据公式（12），总奖励需要在三个不同的目标奖励之间取得良好的平衡。为了避免对权重参数进行耗时的搜索，我们提出了一种动态权重机制，该机制灵感源自动态加权平均（Dynamic Weighted Average，DWA）算法在相关领域的应用。这种动态权重机制学习随着时间的推移，对每个奖励的变化率进行考虑，以学习对奖励权重进行平均的方法。具体来说，在时间t时，奖励k的权重γ被定义为： $\gamma_k(t) =$ 动态权重机制，其中$\gamma_k(t)$表示在时间t时，奖励k的权重。

$$\gamma_k(t) = \frac{Ke^{\frac{h_k(t-1)}{\tau}}}{\sum_i e^{\frac{h_i(t-1)}{\tau}}}, h_k(t-1) = \frac{\pi_k(t-2)}{\pi_k(t-1)}$$

知乎

$$\pi_t(\hat{e}) = \gamma_{MI}(t) \cdot \pi_{MI}(\hat{e}) + \gamma_{KL}(t) \cdot \pi_{KL}(\hat{e}) + \gamma_{Entropy}(t) \cdot \pi_{Entropy}(\hat{e})$$

Applying MMI framework on Different Types of Backbone Models

应用MMI方法到主干模型的一般流程如图所示。然而，与常规流程不同，当我们应用此方法到多任务学习模型以更好地与评分对齐时，需要进行特殊的适应。这是因为，与事后调整的评分不同，模型本身预测的评分$\hat{r}$是一个非固定值。这意味着在微调过程中，评分$\hat{r}$也会被更新，这可能会影响多任务模型的推荐性能。因此，为了确保推荐性能不受与评分对齐任务的影响，我们将主干模型的原损失$L_{backbone}$和RL目标函数*L_{RL}结合起来，作为在多任务学习主干模型上执行与评分对齐任务的优化目标。

$$L = \lambda L_{RL} + (1 - \lambda) L_{Backbone}$$

Experimental Setup

Datasets

实验在三个不同领域的真实数据集上进行：TripAdvisor（酒店）、Yelp（餐厅）和 Amazon-MoviesAndTV。我们基于 [@NETE; @pepler; @peter; @erra] 中预处理的版本构建数据集，过滤掉评论少于 5 条的用户。评论中的内容特征由 Sentires 提取。此外，我们使用 Spacy 工具包对每条评论进行句子依存分析，移除主语为 "I" 或 "We" 的评论。这是因为这些评论通常缺乏对内容的客观描述，不适合用于生成解释。最后，我们将整个数据集按 8:1:1 的比例划分为训练集、验证集和测试集*。

	TripAdvisor	Amazon	Yelp
# users	9,765	7,506	27,146
# items	6,280	7,360	20,266
# reviews	269,491	374,081	950,960
# records per user	27.60	49.84	35.03
# records per item	42.91	50.83	46.92

Evaluation Metrics

对于评级对齐： **归一化互信息（Normalized Mutual Information）**

我们通过将生成理由的句子表示与表示预测评级的一热向量进行连接，作为输入数据训练一个MINE模型。当模型收敛*时，方程（5）中的最终下界值将是对$I(R; E)$的估计。然而，由于不同模型预测的评级R不同（对于后验模型Att2Seq，我们遵循先前的工作直接使用真实评级作为预测），我们采用归一化的互信息（NMI）来使值可比较：$\frac{I(R;E)}{H(R)}$。NMI的范围在$0, 1$之间，值越高，理由与评级的对齐越强。

对于情感准确性： **情感匹配度（Sentiment Matching）**

我们还对生成的理由执行情感分类任务，以测量理由预测的情感是否与预测评级的情感相匹配。我们分别进行精细粒度（标签是预测评级的1-5级）和粗粒度（标签是负、中性和正）评估。

对于特征对齐*： **互信息（Mutual Information）** 类似于估计$I(R; E)$，我们通过将生成理由的句子表示与所有模型预先定义的内容特征的单词表示进行连接，来训练一个MINE模型。我们可以直接比较不同模型的估计$I(F; E)$，因为所有模型的预先定义特征都是相同的。

FMR（特征匹配比率） 我们检查分配的特征：

$$FMR = \frac{1}{N} \sum_{u,i} \mathbb{I}(f_{u,i} \in E_{u,i})$$

由与用户评论在相似性方面的质量。

RQ1: Alignment with ratings/features

Table 3: Performance of explanation generation methods in terms of Alignment with Rating

	TripAdvisor			Amazon			Yelp		
	$\frac{I(R;E)}{H(R)}$	Sentiment Accuracy		$\frac{I(R;E)}{H(R)}$	Sentiment Accuracy		$\frac{I(R;E)}{H(R)}$	Sentiment Accuracy	
		5-class	3-class		5-class	3-class		5-class	3-class
NRT	0.15	45.44	87.12	0.13	44.24	75.12	0.10	47.91	69.63
PEPLER+MF	0.02	35.80	65.82	0.01	29.36	59.09	0.00	34.09	61.27
DualPC	0.35	65.84	89.33	0.04	31.08	64.25	0.43	67.32	77.86
SAER	0.24	52.46	87.45	0.12	43.21	73.01	0.47	64.54	82.64
Att2Seq	0.22	47.69	74.66	0.27	49.13	73.35	0.31	51.58	75.11
PETER	0.18	50.68	89.94	0.19	44.24	75.12	0.20	51.88	71.69
Att2Seq + MMI	**0.93**	**76.53**	88.89	**0.88**	**73.79**	89.96	0.52	61.84	82.50
PETER + MMI	0.44	70.55	**93.17**	0.51	71.39	**96.57**	0.61	76.78	88.58

如表所示，报告了不同生成方法的对齐性能。从表中，我们可以得出结论，通过在Att2Seq和PETER上应用MMI框架，我们在所有设置下的NMI和情感准确性方面实现了更优的性能。配备了MMI框架，Att2Seq和PETER模型的微调版本相较于它们的预训练版本，获得了更强的对齐能力，这表明MMI框架对多任务学习模型和后续生成模型都有益处。

除了我们的MMI方法之外，SAER和DualPC在TripAdvisor和Yelp数据集上击败了其他基线模型。这是因为它们设计了内在模型机制，将理由生成与评分预测联系得更紧密，而其他模型仅通过共享潜在空间或简单地将预测的评分作为理由生成器的初始状态来松散地连接这两个任务。
PEPLER+MF在与评分对齐方面的能力最差，这表明预训练LLM的提示微调的局限性。

然后，我们分析了与特征对齐的结果。如表所示，我们的MMI框架使ApRef2Seq和PETER+能够获得与它们的预训练版本相比更强的对齐能力，并且在最强大的竞争对手PEPLER-D之外，还超过了大多数基线模型。然而，我们注意到尽管PEPLER-D能够有效地生成包含指定特征的句子，但由于其直接将特征作为提示词的做法，这些句子大多缺乏特色，缺乏多样性（例如，"食物很好。"，"服务很好。"）。这样的观察理由了PEPLER-D在文本生成方面的不佳表现，如表所示。

总之，全面评估结果表明，所提出的MMI框架在增强理由的对齐特性方面是有效的。基于不同的主干架构的改进在一定程度上反映了框架的灵活性和普适性。

Human Evaluation

我们招募了25名参与者，并基于Yelp数据集设计了两个任务。在第一个任务中，我们将具有不同评分的内容配对，并要求参与者根据生成的理由选择他们认为更好的内容。我们对比了5种方法：注意力序列模型（Att2Seq）、注意力序列模型+MMI、SAER、双路径对比模型（DualPC）以及一个参考方法，该方法直接将对应用户的评论作为理由。每位参与者需要标注60个记录，每个理由方法包含12个记录。比较结果呈现在表中，并根据两个内容预测评分之间的差异值进行分组。在所有情况下，注意力序列模型+MMI都取得了最高的共识率，这说明评分对齐的理由能帮助用户更好地理解预测评分，同时能更有效地区分不同内容。同时，用户评论的相对表现不佳表明了将用户评论作为理由生成基准的局限性。

	$\Delta r = 1$	$\Delta r = 2$	$\Delta r = 3$	$\Delta r = 4$
Att2Seq+MMI	**67.61**	**76.00**	**93.33**	**93.5**
Att2Seq	31.43	56.41	57.83	51.35
DualPC	49.43	71.23	68.89	72.86
SAER	40.28	45.57	42.68	63.01
User Review	30.00	57.89	64.90	63.41

在第二个任务中，我们从数据集中抽取用户-内容对，并收集由ApRef2Seq、ApRef2Seq+MMI、ERRA、PEPLER-D生成的特征和理由。我们要求参与者从三个方面对理由进行注释：**信息性**（生成的理由包含了具体信息，而非模糊描述）、**相关性**（生成理由中的细节与分配给业务的特征一致且相关）和**满意度**（生成的理由使得推荐系统*的使用变得有趣）。我们为每个参与者分配25个记录，并确保每个记录至少被3个参与者注释。注释结果见表。ApRef2Seq+MMI在所有方面都比其

知乎

子提供了很少的内容细节。

原文《Aligning Explanations for Recommendation with Rating and Feature via Maximizing Mutual Information》

发布于 2024-08-07 15:12 · IP 属地北京

个性化推荐 美团 推荐理由

推荐阅读

美团优选3个月要铺20省，目标1亿家庭用户

商业观察家 发表于商业观察家

美团优选是什么?

美团优选是美团旗bai下的社区团购du业务，采取“预购+自提”的模式，进入社区zhi团购赛道，进一步探索社区生鲜dao零售业态，满足差异化消费需求，推动生鲜零售线上线下加速融合。 用户可在...

朴一

美团优选的优势？持续更新

乐驰短剧

美团优选是什么？怎么做

有干货