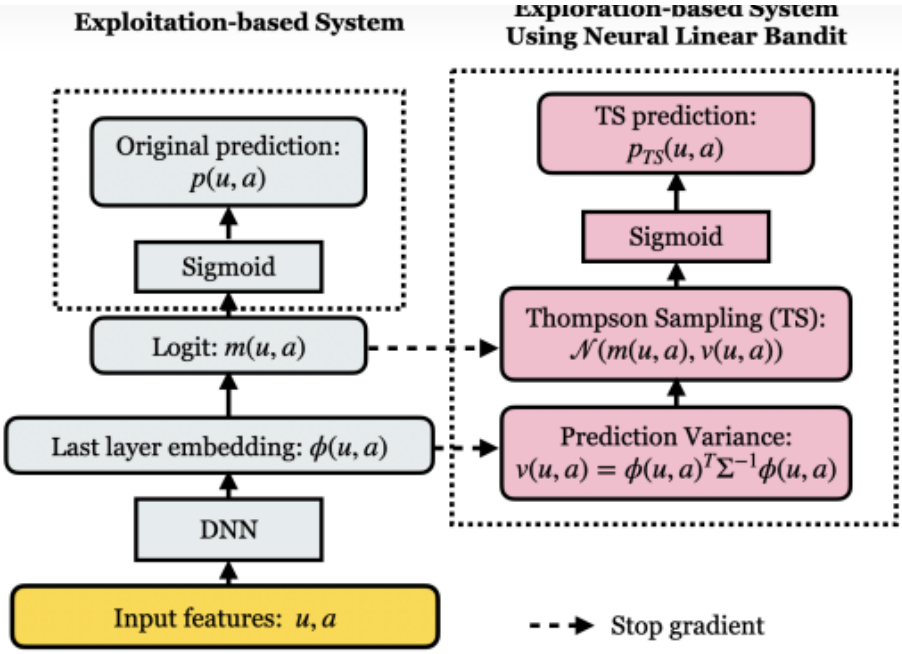


谷歌2024年度探索：推荐系统中的长期价值挖掘——详述衡量、发现与算法策略

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

原文《Long-Term Value of Exploration: Measurements, Findings and Algorithms》

Introduction

推荐系统已经广泛应用于人们的日常生活，在推荐平台上向用户推送相关的内容。当前推荐系统被训练来预测和利用用户的即时反应，如点击、停留时间和购买，取得了显著的个性化成果。然而，这种依赖于当前反馈环效果的系统存在一些问题，即推荐系统和用户互相强化选择。当用户接受推荐的项目并给予反馈时，系统会根据带有偏见的反馈数据进一步加强并固化用户的档案，使其更倾向于他们以前互动过的内容。这导致用户陷入了一个越来越狭窄的内容集合中，而平台上的大量内容仍未被发现。

对于打破依赖当前反馈环的效果，探索是一种关键方法。通过让用户接触到不确定的内容，系统可以积极获取关于未知用户内容的学习信号来填补知识差距。这种行为被称为用户探索，它可以引导用户接触新鲜的内容，同时也使得更多的新鲜和尾部内容能够在平台上得到发现，被称为物品探索。

在博弈论和强化学习文献中，已经有许多关于有效探索技术的研究，但在实际工业环境中使用这些技术面临着很多挑战。主要的问题是如何衡量探索的具体收益，这将是判断切换到基于探索的系统是否优于目前单纯依赖现有系统的具体和可衡量证据。

虽然UPPER-CONFIDENCE-BOUND和Thompson采样等探索技术已经被数学证明可以获得比贪婪策略更好的后悔，但不清楚这些优势是否能够在工业推荐环境中，面对具有噪音和延迟反馈以及无法测试的建模假设的环境。我们的研究提出了一个系统性的方法来研究探索对内容库的影响，以及这个影响如何转化为长期用户的参与度增长。

我们在实际工业环境中面临的主要挑战包括衡量探索的收益、实验设计和设计探索型系统。为了应对这些挑战，我们采用了神经线性Bandit算法，这是一种可扩展的探索算法，可以在深度神经网络学习的表示作为上下文特征的基础上进行线性回归，以估计不确定性。这个算法可以很好地适应现代深度学习驱动的推荐系统，同时实现了准确的不确定性估计的简单计算。

我们研究了Neural Linear在探索方面可能的优势，并在服务数十亿用户的大规模短视频推荐平台上进行了为期三个月的研究。这是一个多阶段系统，如图所示。第一个阶段包括多个检索系统[+]来识别和提名总体内容库中的数十亿个潜在候选内容。第二个阶段涉及一个评分和排序系统来对候选内容的池（按数百个顺序排列），然后是最终包装阶段以实现不同的业务目标和多样化的整体列表。除非另有说明，否则我们将生产推荐系统的非明确探索策略作为在线A/B测试的所有对照组。我们将分别描述相应的实验手臂在不同的实验中。

Long-term Value of Exploration through Enlarged Corpus

我们通过语料库的变化研究探索的好处。总结起来，不确定区域的探索增加了新鲜和尾部内容的曝光率和可发现性，并改变了总体语料库的分布，从而提高了长期用户的体验。首先定义了语料库的度量标准，即可发现的语料库；然后介绍了一个新的"用户-语料库-代码转换"实验框架来衡量探索对可发现语料库的好处。最后，我们展示了语料库变化对用户体验的影响的长期研究。

Corpus Metrics

理想情况下，系统的探索能力越强，对各种X区间的Discoverable Corpus@X, Y天数越大，同时保持相对中立的用户体验作为基本条件。评估所使用的时窗，即Y要求允许新探索的语料库增长的时间窗口。在我们的实验中，我们使用7天的时窗来捕捉短期的语料库增长，使用3个月的时窗来捕捉长期的增长。

User-Corpus-CoDiverted Experiment

传统的用户转移AB测试提供了衡量推荐更改对用户侧向影响的强大工具。然而，这种测试无法捕捉到任何语料库的变化。为了解决这个问题，我们提出了一种名为用户-语料库-代码剥离的AB测试。这是一种多随机化设计（MRD）的实例，通过随机将x%的语料库分发到对照组和实验组，以及随机将x%的用户按比例分配到对照组和实验组来实现。

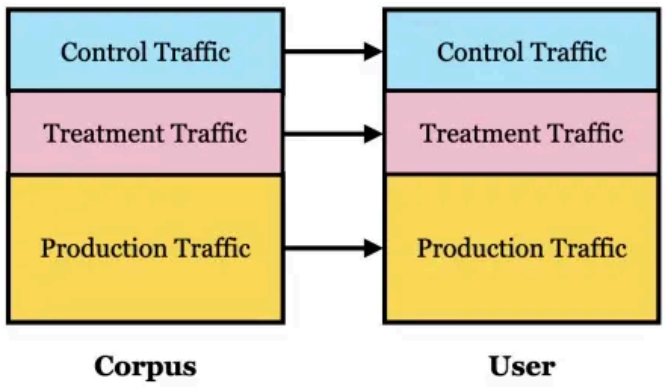


Figure 2: User-Corpus-CoDiverted experiment diagram

相比于传统的用户转移实验，这种设计能够避免实验泄露，并且允许衡量基于语料库的指标的实验效果。在实验中，我们保持用户和语料库的比例，例如5%的用户探索5%的语料库，这样探索实验的效果与完全部署时的50%用户探索整个语料库一致。否则，使用5%的用户流量探索整个语料库（100%）会导致对语料库分布的影响最小。

Exploration Increases Discoverable Corpus

实验结果表明，探索式系统可以有效提高短时间内的内容成功数量，如图所示。随着时间的推移，实验组和对照组之间的差距继续扩大，这可能是因为实验组的内容提供商比对照组创建更多的可发现内容。然而，这并不意味着长期内容增长会一直持续下去。理想的探索系统应能识别出有可能病毒化的高质量内容，以促进初始启动后的增长。为了评估探索系统在长期内对内容的影响，我们分析了不同X_l桶中的发现可及内容库@$X_l, 3$-月期。结果显示，无论是在哪个X_l桶中，探索实验系统都增加了发现可及内容库@$X_l, 3$-月期的数量，增量百分比大致保持在50%左右。

知乎

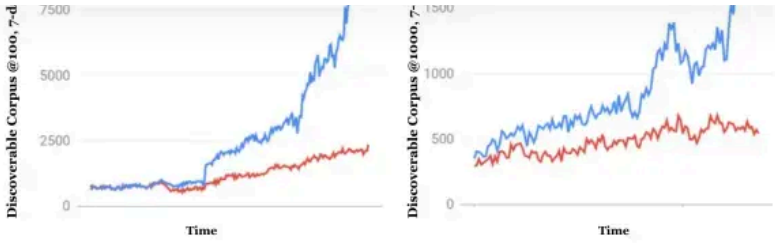
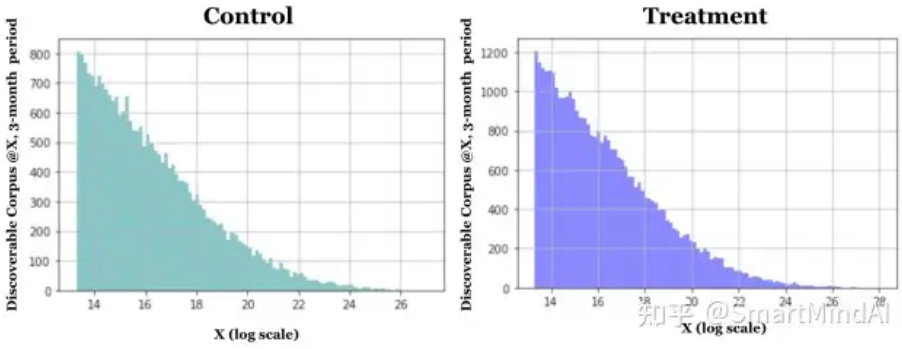


Figure 3: Discoverable Corpus @100, 7-day period (left) and Discoverable Corpus @1000, 7-day period (right) for both control and treatment arms.

为了全面考察质量分布，我们使用Discoverable Corpus \@X, 3-month期间作为质量指标。



图表显示了对Discoverable Corpus \@X, 3-month期间的对照（左）和实验（右）臂的直方图[+]。我们关注至少有10k次post-exploration正面反馈的内容，以便于缩放到corpus的"高质量"区域。x轴表示logarithmic尺度上的post-exploration用户交互数量，y轴绘制特定桶中的内容数量。

Table 1: The change in Discoverable Corpus @X_l, 3-month period between control and treatment arms.

X_l	100 (in 7 days)	1000 (in 7 days)	10K	1M	10M
Change	+119.4%	+58.5%	+48.2%	+51.0%	+53.8%

结果显示，虽然实验组的数字更高，但我们观察到两者的分布非常相似，这意味着探索系统发现的质量分布与原始的质量分布相当。换句话说，探索不仅帮助实现初始成功，而且还发现了最终会达到高质量 。

Long Term Value of Enlarged Discoverable Corpus

我们将这个框架通过闭合论证与长期用户体验联系起来。为了衡量用户的满意度，我们使用一种度量每日活跃用户满意互动数量（基于满意度调查预测）的指标，在整个论文中我们将其称为"满意的每日活跃用户"。为了让每个用户都能访问一个固定的减小了的语料库，我们将相同的种子s_u用于来自同一用户的不同请求（参见算法）。但是，为了避免任何特定的内容对于所有用户流量被删除，我们在不同的用户之间选择不同的种子。我们的研究表明，允许每个用户访问一个固定的减小了的语料库

$$\mathcal{C}' \subset \mathcal{C}$$

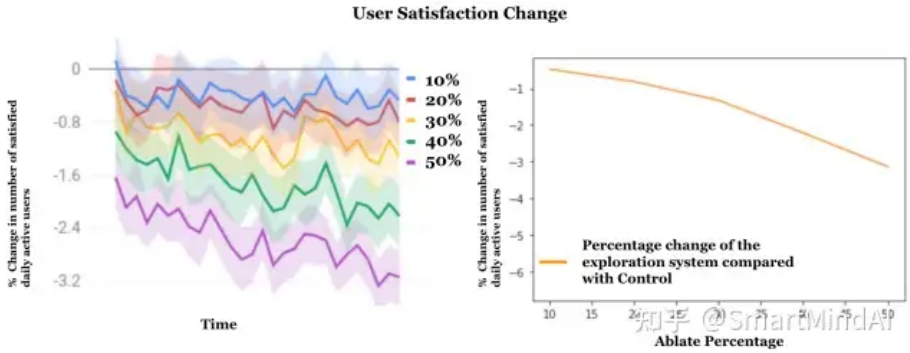
其中去除

$$\mathcal{C} \backslash \mathcal{C}'$$

增加了提名的数量。

我们进行了为期四周的剥离研究，对照组和实验组都运行图所示的多阶段推荐系统。每臂接收提名者输出的所有候选内容，而实验组随机从平台上过滤$x\%$的语料库。除了上述直接在探索性内容上删除各种百分比的方法外，还有其他方法可以调整内容语料库并观察用户体验的变化。

另一种替代方法是直接从平台中删除各种比例的"探索性"内容，但这可能会受到所使用的探索系统的选择影响。在这里，我们提出这种方法的更一般形式------随机过滤，以降低可发现语料库的大小。研究表明，随着ablation大小的增大，每天活跃用户满意度显著降低，并且这种影响随着时间的推移而增强。



右侧的图表显示了可发现语料库大小的变化与满意每日活跃用户数量之间存在单调的关系，因此，我们推测增加可发现语料库将带来良好的用户体验。然而，这种线性关系可能只存在于特定范围内的语料库大小之内。此外，增长语料库可能会在可发现语料库达到一定程度时出现饱和效应。确定这种关系的确切性质是未来研究的一个重要方向。总的来说，通过增加可发现语料库的大小，可以进一步提高长期用户的满意度。

Neural Linear Bandit Based Ranking System

为了考虑长期的价值，我们深入研究了在实际世界和大规模工业系统中有效运行的高效探索技术。其中，我们重点关注了神经线性Bandit (NLB) 这种算法，它基于 NeuralLinear 算法 。我们选择了 NeuralLinear 作为基础算法✦，因为它是许多工业系统的基础，并且具有计算方差高的优点。我们的研究重点在于如何将这个算法纳入现有的工业系统管道，以及如何构建基于探索的推荐系统面临的挑战和未来的机遇。

Neural Linear Bandit

$$R_T(\pi) := \mathbf{u}_t \sim P_{\mathcal{U}}, a_t \sim \pi(\cdot|ub_t), rb_t \sim P(\cdot|ub_t, ab_t) \left[\sum_{t=1}^{T} rb(ub_t, ab_t^*) - rb(ub_t, ab_t) \right]$$

线性模型无法完全捕捉到数据的复杂性，因此我们引入了神经线性模型。神经线性模型是一种结合了线性模型和深度神经网络✦的模型。这个模型的假设包括两个条件：（1）随着时间的推移，线性模型的预测结果逐渐收敛到真实的相关性函数$f(ub^*, ab^*)$；（2）每次的选择都有接近于真实回报的趋势。这种模型通过这种方式改善了线性模型的表现力。在具体操作中，我们将一个动作a_t与所有可能的状态x_t联系起来，这样就形成了一个二元树形结构。每个节点代表一个状态，每个叶子代表一个动作。我们使用采样方法从树中随机选取一个动作，然后将这个动作应用于状态。最后，我们从当前状态出发沿着这个树移动，直到达到最终状态，并且计算出最终的回报。这个过程就是UCB或Thompson Sampling算法的核心思想。

[$assum: linear\_last\_layer$] There exists a representation function

$$\phi : RR^{d_u} \times RR^{d_a} \to RR^d$$

and an unknown parameterRR^d, such that for all user-content pair(ub, ab), the mean reward$rb(ub, ab)$is linear in the representation$\phi(ub, ab)$with

知乎

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \frac{rb_t - \phi(ub_t, ab_t)^T \beta_{t-1}}{\sigma^2}$$

$$\sigma^2 = \sigma^2 + \frac{rb_t - \phi(ub_t, ab_t)^T \hat{\beta}_{t-1}}{\nu_t}$$

$$\nu_t = \nu_{t-1} + \frac{1}{\nu_t}$$

$$PP(\beta|rb_t) \propto PP(\beta)PP(rb_t|\beta) \propto Ncal(\hat{\beta}_t, \sigma^2\Sigma_t^{-1})$$

$$\Sigma_t = \sum_{s=0}^{t-1} \beta_s \nabla \log p(rb_s|ub_s) \nabla \log p(ub_s)^\top$$

而参数估计$^+$$\hat{\beta}_t$则是通过以下方式更新的：

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \frac{rb_t - \phi(ub_t, b_t)^T \hat{beta}_{t-1}}{\nu_t}$$

$$Ncal\big(\phi(ub, ab)^T \hat{\beta}_t, \sigma^2 \phi(ub, ab)^T \Sigma_t^{-1} \phi(ub, ab)\big)$$

神经线性Bandit可以通过从后验分布$^+$中抽取每个动作的奖励并选择获得最高奖励的动作进行拉动来进行操作。这种策略称为基于最大奖励的拉取策略。

Implementation

尽管直接将算法融入工业训练和服务管道具有简单性，但仍然面临着挑战，其中包括： 1. 隐私保护问题：短名算法需要处理大量的个人数据，如果这些数据被泄露，将会对用户的隐私造成威胁。 2. 计算效率问题：短名算法通常需要大量的计算资源才能运行，这对于一些资源有限的企业来说是一大挑战。 3. 模型解释性问题：短名算法往往具有很强的复杂度，这使得其结果难以理解和解释，这对企业和用户来说都是一个挑战。 4. 结果可靠性问题：短名算法的结果可能会受到各种因素的影响，比如数据质量、特征选择$^+$等，这使得其结果可能不那么可靠。

$$\begin{aligned}\sigma^2 \phi(ub, ab)^T \Sigma^{-1} \phi(ub, ab) &= \sigma^2 \phi(ub, ab)^T (LL^T)^{-1} \phi(ub, ab) \\ &= \sigma^2 \phi(ub, ab)^T L^{-T} L^{-1} \phi(ub, ab) \\ &= \sigma^2 z(ub, ab)^T z(ub, ab)\end{aligned}$$

对于

$$z(ub, ab) = L^{-1} \phi(ub, ab)$$

的求解，只需使用$Ocal(d^2)$的复杂度。同样，如果已知Cholesky分解，可以连续地通过求解两个下三角矩阵L的线性系统$^+$来计算参数β。图示了预测奖励与神经网络预测$\hat{r}(ub, ab)$之间的平均绝对差异，即

$$\phi(ub, ab)^T \hat{\beta} - \hat{r}(ub, ab)$$

结果显示，Cholesky分解在解决方案的准确性和稳定性方面优于伪逆，但由于训练速度更快，故我们在我们的设置下选择了伪逆。

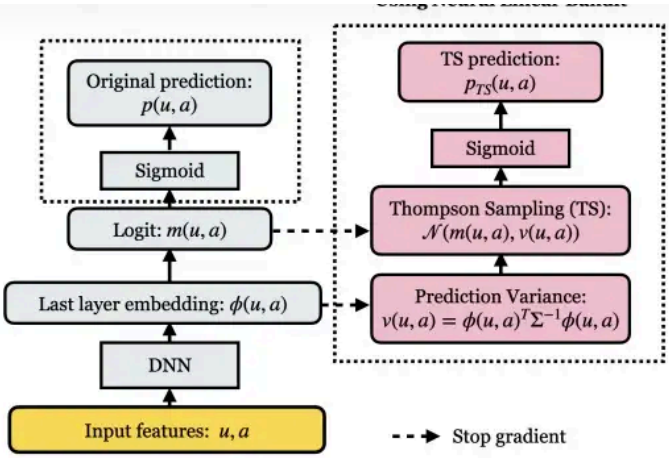


Figure 6: The model architecture for the exploitation-based system (control) and the exploration-based system using Neural Linear Bandit (treatment), on a classification task.

Input: Total number of training runs$T$; Total number of batches per training run$H$; Initialize the model$f$with parameter$\theta_0$and last layer representationϕ_0; Initialize the dataset$Dcal_0 = \emptyset$; Initialize Neural Linear Bandit parameter

$$\Sigma_{1,0} = \Sigma_0 := \epsilon \mathbf{I}$$

$\beta_0 = \mathbf{0}$; noise parameterσ^2\

调查3探讨了分类任务的扩展，尤其是在预测完成、点击和喜欢等情境下。相比于回归任务，广义线性模型（如逻辑回归）在返回值为二元时表现更佳。以前的研究已经研究过如何有效地探索一种通用形式的广义线性模型，该模型接受一个关于上下文特征的广义线性模型的链接函数作为输入，并输出一个奖励。这种方法的一个常见例子是GLM-UCB和GLM-TSL，其中参数β需要通过不断迭代的方式来估计最大似然估计+。然而，在GLM中，与在线性模型不同的是，最大似然估计不能以一次性的闭合形式计算出来，因此每次更新都需要通过求解目标函数来获取β_t。

$$\sum_{\tau=1}^{t-1}(rb_\tau - \mu(\phi(ub_\tau, ab_\tau)^T\beta))\phi(ub_\tau, ab_\tau) = 0$$

该方法利用每轮所有的先前观测并产生了昂贵的样例梯度更新。但是，可以明显看出$\hat{\beta}$只需用于预测后验分布的奖赏均值

$$\mu(\phi(ub, ab)^T\hat{\beta})$$

此外，还存在一个便宜的近似物，即原始二进制标签预测的逻辑值$hatb(ub, ab)$，这是当前系统的副产品，提供了一致的估计。为了选择最佳行动，我们选择使$\phi(ub, ab)^T\beta$最大的动作ab，因为当μ是严格增函数时，它等于

$$\mathbf{argmax}\ \mu(\phi(ub, ab)^T\beta)$$

这个思想类似于 @ding2021efficient提出的SGD-TS算法，后者在考虑上下文特征的多样性假设下，表明在线SGD和TS探索可以实现有限臂GLM问题的$\tilde{O}(\sqrt{T})$的后悔。不同之处在于，我们的计算矩阵伪逆以更准确地估计不确定性，而不仅仅是对角矩阵进行近似。此外，不确定性估计与线性情况相同，可以通过简单地维护协方差矩阵来计算。线性逻辑空间采样结束后，我们可以通过链接函数μ将样本转换回原空间。

Experiments

我们在一个最大的短视频推荐平台上对神经网络线性带标记排名系统进行了线上AB测试，以评估其性能。我们研究了不确定性的性质和可靠性。首先，在生产环境中运行了为期六周的用户偏离AB测试，其中对照组为原始排名模型，实验组为使用神经网络线性带标记的探索式排名系统。对

为了确保矩阵求逆的稳定性，我们设置正则化参数$\epsilon = 1e^{-6}$（公式）。为了挑选噪声参数σ^2，我们通过不同训练模型的集合（作为昂贵的实证标准）计算出不确定性，并选择常数超参数$\sigma^2 = 10$，使得由群体和神经网络线性带标记获得的不确定性大致在同一数量级上。

Table 2: The gain in various freshness related metric.

Fresh pos. feedback gain	Mean	95% CI
1h	1.49%	[1.20, 1.77]%
3h	1.51%	[1.24, 1.77]%
12h	1.45%	[1.19, 1.71]%
1d	1.43%	[1.18, 1.69]%
3d	2.55%	[2.30, 2.81]%
12d	1.16%	[0.92, 1.41]%

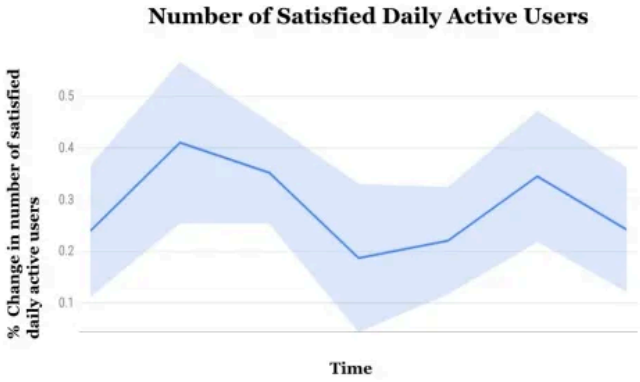


Figure 8: The gain (with the 95% confidence interval) for the Neural Linear Bandit based ranking system in terms of user satisfaction, compared with the the control production system, over a 6-week period.

NLB通过对新鲜和尾部内容给予更多曝光机会，改变整体的内容语料库分布，并从中获取有价值的信号进行学习，从而产生用户参与的增长。 表显示在不同时间段发布的新鲜内容上的积极互动增加，有助于理解不同新鲜程度下的效果。随着新鲜程度的变化，用户积极参与度明显提高，证明了探索系统的有效性。此外，系统能够帮助用户发现新的兴趣点，因此每天活跃用户数量也出现了稳定的增长。

	Neural Linear Bandit
Content Age	-0.35 ± 0.003
Content Popularity	-0.26 ± 0.003
User Activity	0.02 ± 0.008

（1）内容发布+的日期距今（即内容年龄）；（2）内容生命周期内的正面互动次数（即内容流行度）和一个用来捕捉用户属性：（3）用户平台上的总互动次数（即用户活动）。通过斯皮尔曼+秩相关系数衡量这三个特征与不确定性之间的关系。使用Neural Linear Bandit来计算不确定性，并发现在新鲜且不太流行的内容上，当前系统对于不确定性有较高的估算。而对于用户活动水平不同的用户，则没有明显的区别。同时，对比使用堆叠模型和神经线性Bandit获得的不确定性，其结果分别为负0.3和相似，表明不确定性估计具有一定的可靠性和一致性。

知乎

升。此外，相较于仅依赖exploitation的系统，神经线性Bandit更公平地分配了内容，表现为post-exploration指标的改善以及更多尾部内容的发现。

发布于 2024-03-22 11:41 · IP 属地北京

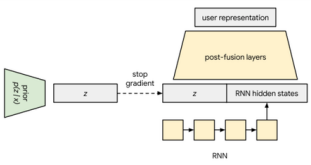
推荐系统 谷歌 (Google) 探索

理性发言，友善互动

发布

还没有评论，发表第一个评论吧

推荐阅读



谷歌2023-揭秘序列推荐中的用户潜在意图-论文深度解析

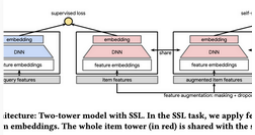
SmartMindAI

10分钟读懂谷歌分析GA（Google Analytics）

Eliza

干货--谷歌大神手把手教你做用研下集

杨老助你求... 发表于思考的杨咩...



【Google Paper】对比于解决推荐系统长尾问题

吴家丫头