

# 多目标推荐场景下的深度学习实践

罗景 机器学习与推荐算法 5天前

嘿，记得给“机器学习与推荐算法”添加星标

分享嘉宾：罗景 58同城 高级架构师

编辑整理：陈维美

出品平台：DataFunTalk

**导读：**58 同城作为一个 UV 千万量级的APP，覆盖了租房，招聘，二手房，二手车，本地生活，新房，新车等多个业务。在首页这个复杂的业务推荐场景下，我们面临着多业务融合挑战。如何利用从双通道到多通道到多场景适配的深度学习模型，实现涉及连接效率提升，增加商业收入，提高用户体验，优化用户流程，支持运营活动的各种目标的平衡。本文将以58同城首页猜你喜欢为例，将具体介绍实践多业务融合推荐场景下的深度学习模型应用，以及相关的工程实践。

## 01

### 多业务融合的挑战

在多业务融合的推荐场景下，召回策略，重排策略，展示样式，排序模型都会面临多业务所带来的挑战：

- ① 召回策略：会涉及到多个召回通道的业务适配。
- ② 重排策略：需保证各个业务间流量的均衡分配。
- ③ 展示样式：展示元素与样式的一些适配问题。
- ④ 排序模型（传统使用XGBoost模型），面临着两大困难：

- 特征对齐困难：由于58 业务多样性，各类业务贴的属性信息都存在显著的差异，导致了样本中的帖子特征和用户兴趣特征无法对齐。

- 特征工程复杂度：针对不同的业务需构建相应用户兴趣，各个业务均存在特定的特征工程逻辑，部分特征包含了复杂的流程和逻辑。同时，需要对不同业务做匹配交叉，这导致整个特征工程逻辑非常复杂和庞大。

## 02

### 定制化通道和序列化通道的双通道模型

#### 1. 行为序列兴趣模型验证

近两年来，用户序列化兴趣深度模型已经被广泛应用在推荐广告上，取得了显著的效果。它是以用户行为序列作为模型的核心输入，具备比较优秀的兴趣建模能力，同时又较好地解决了特征对齐的困难。由于它输入的是用户的行为序列，显著的降低了特征工程的复杂度。我们早期的用户序列化兴趣深度模型，主要是涉及以下两个方面内容：

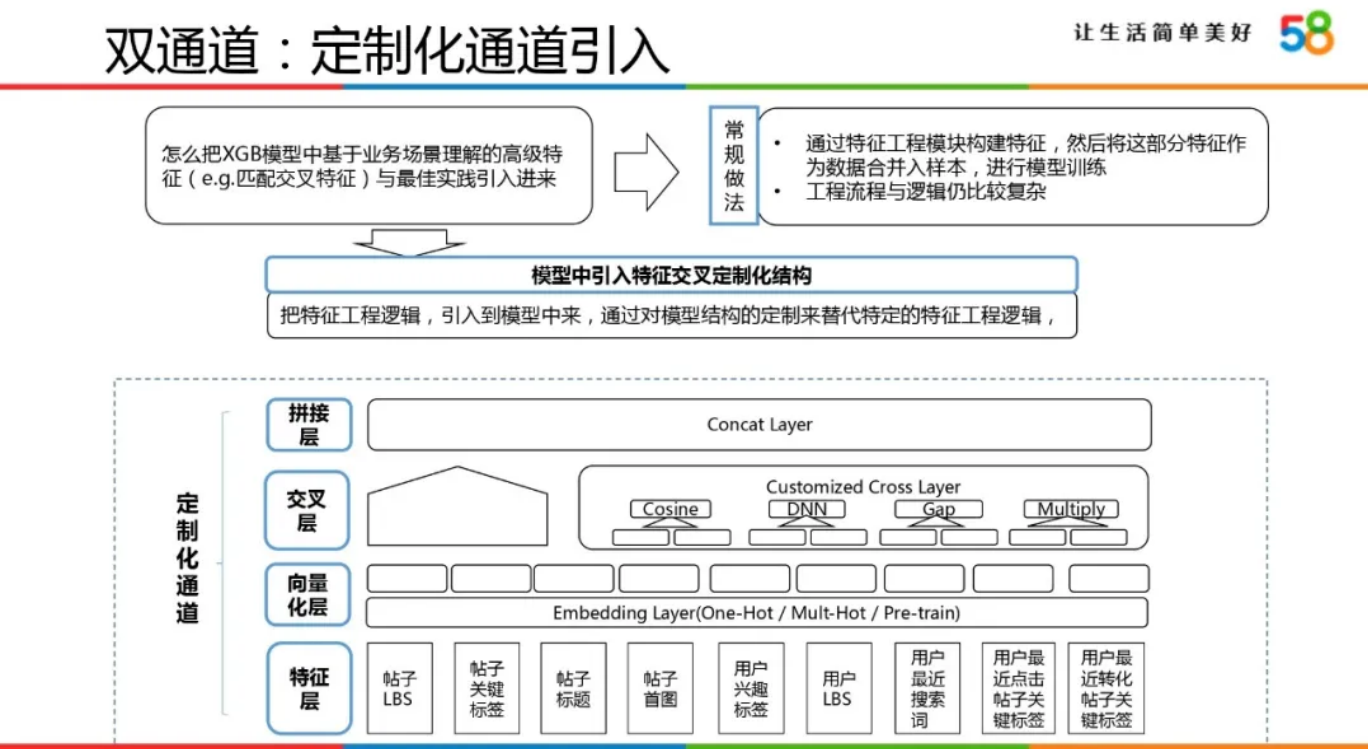
① 用户行为序列的构建：我们主要采用用户的点击和转化行为来构成用户行为序列，在具体行为的表征方式上，采用了帖子，类目，地域等关键属性信息来作为序列的表征，而整个序列长度是基于业务周期的。

② 序列模型的选择：在模型的选择上，我们分别验证了DIN，DIEN 和Transformer模型在用户兴趣序列建模中的效果。在整体结构上，Transformer 模型的扩展性会更加好一点。所以我们最终选择Transformer 作为我们整个序列兴趣建模的基础组件。Transformer最早是谷歌提出的一个序列建模组件，利用自注意力与多头机制，通过位置的Embedding来进行序列化兴趣建模。

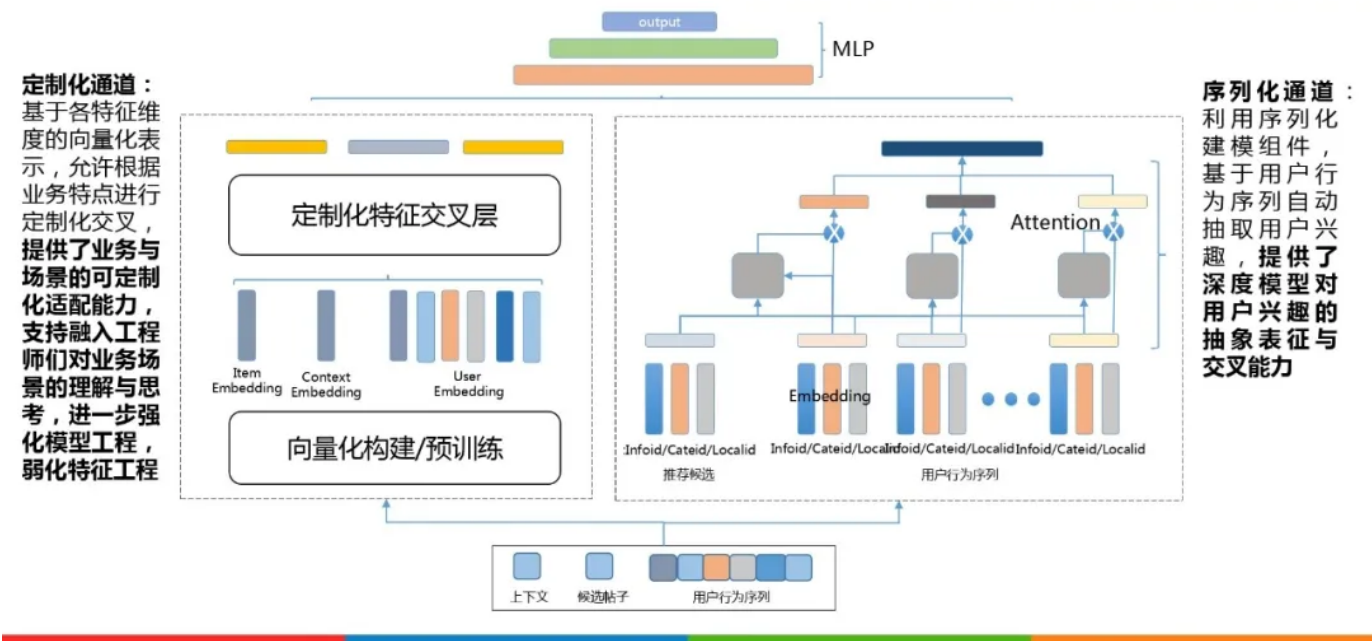
我们发现，纯序列兴趣模型的效果并不能超越特征工程比较完善的XGBoost模型。对比两边模型发现，传统的XGBoost 模型有很多高级特征，包括匹配交叉特征，以及在特征工程上的最佳实践。虽然纯序列模型通过深度学习提供了比较强的兴趣表达能力，也支持帖子交叉，但是缺乏业务定制能力。因此，我们尝试把XGBoost 模型中的对业务场景有理解的高级特征引入模型中。通过特征工程模块构建特征，然后把这部分特征作为数据一部分合并到样本中，然后进行模型训练。这种方法的优点比较明显，使我们能够复用现有的特征工程流程，但是相应的工程流程与逻辑的仍然比较复杂。在下一小节中，我们将引入定制化通道。

#### 2. 定制化通道引入

定制化通道引入是指引入特征交叉的定制化结构，把特征工程逻辑引入到模型中。通过对模型结构的定制来代替特定的特征工程逻辑。它的基本思路就是首先引入原始的基础特征，然后通过序列化进行表征。在序列化表征的基础上，基于向量来做定制化特征的交叉。



# 双通道：模型架构



简单来说它具备了两个通道：

- ① 定制化通道：基于各个特征维度的序列化表示，允许工程师们根据业务特点进行定制化交叉，为业务与场景提供了定制能力，同时也支持融入工程师对业务与场景的理解。
- ② 序列化通道：利用序列化建模组件。基于用户行为序列自动抽取的用户兴趣，为模型提供了用户信息的抽象表征与交叉能力。

引入定制化通道后，最上层再连接MLP，这样就形成了下面的双通道模型架构。

## 4. 总结

双通道模型在提升模型效果的同时，使得特征工程的流程大幅减少，复杂度也明显降低了。主要可以细分为以下四点：

- 实现了基线最优模型的稳定超越：其中点击率提升了3%，曝光转化率提升5%。
- 显著降低特征工程复杂度：通过在定制化通道引入特征交叉能力，和基于用户行为序列的建模方式，使得特征工程流程大幅减少，复杂度明显降低。

- 建立了序列化兴趣模型线上线下的工程架构基础：解决离线在线性能问题，成为主流排序模型。
- 进一步优化的基础：后续的多通道模型升级以及多场景适配探索都是以双通道模型为基础。

# 03

## 多通道模型探索

### 1. 探索方向

如何进一步优化双通道模型，更精准地去刻画用户的兴趣。我们总结了两个简单的方向：

① 第一个是在深度上，怎么结合业务特点，在单行为序列上做更深度的优化。列出几个在实践上效果比较好的模型：

IDSIN(深度会话兴趣模型)：基于行为序列的会话切分对用户兴趣进行结构化建模；

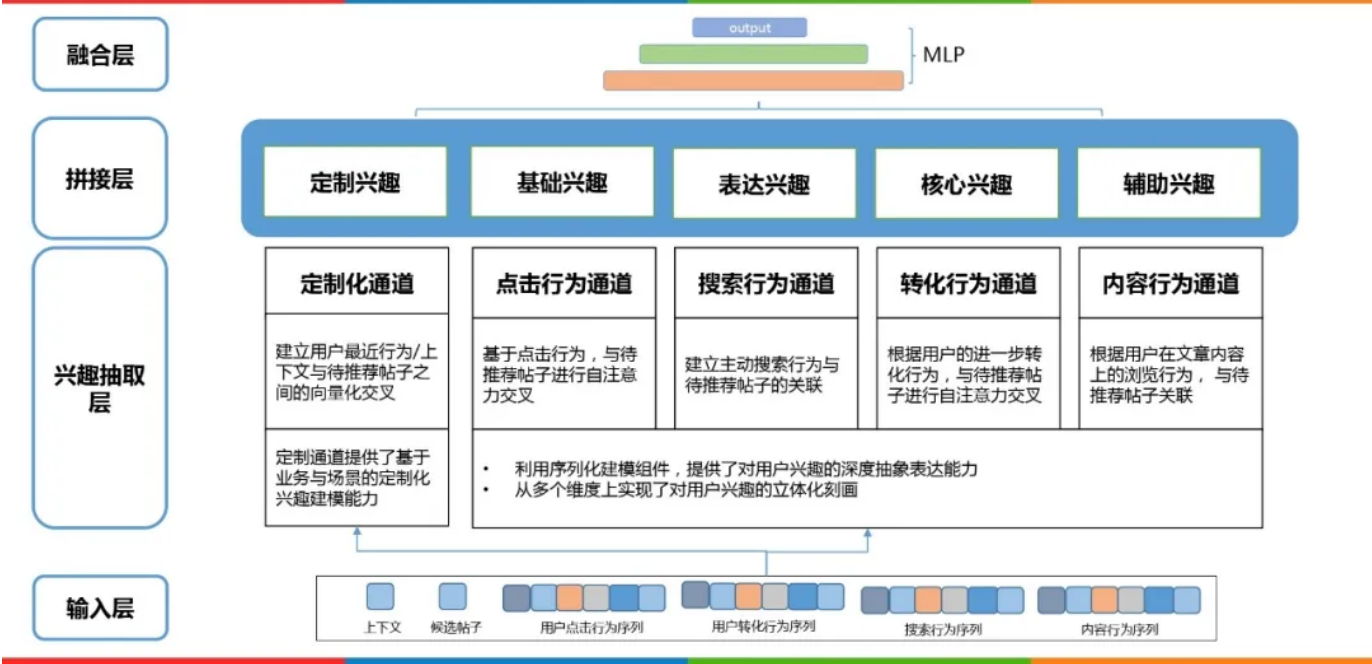
IMIMN(长序列的用户行为建模)：以NTM为基础进行长用户行为序列建模；

ISIM(基于搜索的终身行为序列建模)：采用两阶段的方式，首先从用户的历史行为中找到推荐商品的相近行为，然后再做更精细化的建模。但是在低频业务场景下，该模型的收益有限。

② 第二在广度上，引入多种异构的用户行为序列，以实现对用户兴趣更全面和立体化的刻画。我们提出了一个“1+N”的多通道深度兴趣模型，“1”代表是的定制化通道，“N”代表用户的点击搜索转化内容浏览等行为的序列通道。通过这样的模型结构，最终能够形成定制兴趣，基础兴趣，表达兴趣，核心兴趣，辅助兴趣等多种形式的兴趣表征。

### 2. 用户兴趣的全面立体化刻画

# 多通道模型：用户兴趣的全面立体化刻画



我们将整个模型结构分为以下四层：

- ① 第一层是以用户的行为序列作为核心的，包括点击搜索内容内容浏览等多种的行为序列。
- ② 第二层是兴趣抽取层，通过五个通道来学习相应的兴趣表征：
  - 定制化通道，是建立用户最近行为以及上下文与待推荐帖子间的序列化交叉，形成的定制化兴趣。
  - 点击行为通道，基于用户在帖子上的点击行为或浏览行为，与待推荐帖子做自注意力的交叉，从而形成基础用户兴趣。
  - 搜索行为通道，强化了核心兴趣。

通过建立搜索行为与待推荐帖子的关联，形成用户的表达兴趣。在输入用户行为序列，建立了query和帖子的表征后，再进行序列化的兴趣抽取。搜索行为序列，关键在于如何填平query和帖子间的异构表征差异。因为Query通常是文本序列，而帖子是结构化的信息体，怎么将其映射到统一的表征空间，保证特征的可交叉性，是亟待我们解决的问题。



目前我们采用了基于词向量的语义表征方式。首先我们训练了字词向量数据，在这个基础上，再对帖子抽取标题和核心标签来表示帖子。其中有两个关键点，一是出示什么样的核心标签。根据业务的特点，我们发现用户在不同的内部上搜索的关注点非常的聚焦，比如说租房主要搜索商圈小区地域信息，二手车以搜索车的品牌车型车系作为主要的搜索。为了简化，我们只把搜索关注点纳入到核心标签中，作为结构化的字段，相对容易抽取。采用基于词向量的语义表征方式，用不同的手段，比如LSTM语言模型。目前我们采用比较简单的Average Pooling，把query和帖子文本片段的词向量求和平均。使用该方法一方面是由于标题和query 间的词顺序对相关性的影响不是很大，另一方面它的工程实践相对比较简单。

- 转化行为通道：体现用户主动意图。

根据用户的转化行为，与待推荐帖子做self-attention 交叉，构建了核心的用户兴趣。它以转化行为序列为输入，以transformer作为建模组件。在58复杂的业务场景下，它能体现用户的进一步的交易意图。相比于点击浏览行为，这类数据更稀疏，但更能体现用户的真实需求。比如说租房，用户打了电话，说明用户对该房子的兴趣度很高。

那么转化行为中帖子具体是怎么表征的？如果继续沿用帖子ID，数据过于稀疏，模型的效果不是很理想。针对转化行为数据稀疏的特点，我们引入了一个聚类ID，来代替该帖子ID。聚类ID在泛化能力与宣传度上，能够做到较好的平衡，它的信息强度弱于帖子ID，但是它泛化性会更好。具体就是把每个帖子的关键业务属性组合起来进行聚类，相应的关键业务属性组合值，就是聚类ID。

- 内容行为通道：对兴趣发现与补充。

根据用户在文章内容上的浏览行为，与待推荐帖子关联形成补充的辅助兴趣。多种用户行为序列通道利用序列化建模组件，提供了对用户兴趣的深度抽象表达能力。我们发现越来越多的互联网公司都在针对自己的业务构建自己的内容体系，以提升用户的使用时长忠诚度和留存。在提升用户留存忠诚度的同时，可以辅助用户做决策，包括业务的引流工厂。内容和业务其实是强相关的，如果我们把用户在内容上的浏览行为应用到业务推荐中，可以辅助发现和补充一些用户兴趣。

在用户的行为相对比较稀疏的情况下，常规的做法就是构建用户信息标签，根据用户的内容浏览行为在内容上提示用户在业务上的偏好，然后将其应用到推荐中来。但是这种做法的局限在于，需要进行内容理解，将它映射到相应的业务属性上。在实践中我们考虑了两种方案，第一种方案是语义空间映射，基

本的思路是把文本内容进行关键词与类目提取，然后在空间上进行交叉。这种方式的突出问题在于内容体与业务体在语义相关性通常比较弱。而且随着内容呈现的多媒体化，很多内容都以视频形式呈现，给我们的语义表征带来很大的困难。目前，我们最近在研究探索一个新的思路：基于行为窗口的协同表征。

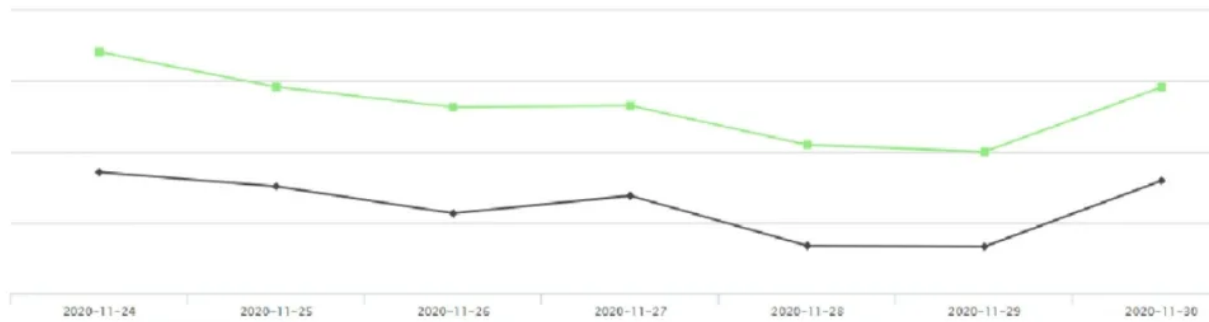
③ 第三层是拼接层，将多个通道进行拼接。

④ 第四层是融合层，通过MLP网络对信息进行融合。从用户的点击搜索转化内容浏览等多种行为中來抽出兴趣，有很多优点。一方面比较适合58的相对低频的业务场景，能从多种维度上去补齐信息。另一方面能从不同的角度来刻画用户的兴趣。

### 3. 多通道深度兴趣模型效果

## 模型效果：多通道深度兴趣模型

让生活简单美好 58



**整体效果上：**多通道深度兴趣模型在曝光转化率上实现了进一步提升，目前线上相比基线模型，提升幅度超过10%

**通道效果上：**点击行为序列 > 搜索行为序列 > 转化行为序列，内容行为序列还在探索，还没有明确的效果

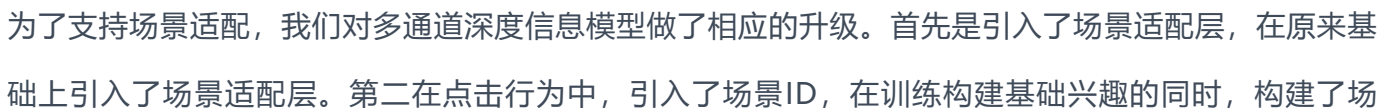
多通道深度兴趣模型，关键在对不同类型的用户行为，采用合适的表征方式建立行为与待推荐帖子的交叉关联。我们在点击行为通道上采用帖子的关键属性和帖子 ID来进行表征。转化行为通道是根据行为的稀疏程度，权衡表示特征的泛化能力与信号强度的基础上，采用聚类ID与关键属性来进行表征。搜索行为通道是基于词向量将搜索词与帖子文本片段表征到统一的语义空间行为通道。利用点击搜索转化内容浏览等多种的行为，聚焦低频需求的业务场景和特点，比较好地补齐低频导致的兴趣刻画的短板，



# 04

## 1. 迁移能力探索

## 2. 场景适配层引入



景表征。第三是在定制化通道中引入了更丰富的场景特征的交叉逻辑，以覆盖不同场景的需求。

简单来说，我们基于场景的表征，通过注意力机制来对定制兴趣，基础兴趣，表达兴趣这些场景的注意力适配，把相应的通道兴趣与常见表征进行拼接，然后再交叉学习，最终输出相应适配系数。通过点击行为序列通道引入点击行为，在点击行为的表征中，引入场景ID信息作为表征的重要组成部分，从而来学习场景向量化表示。

# 05

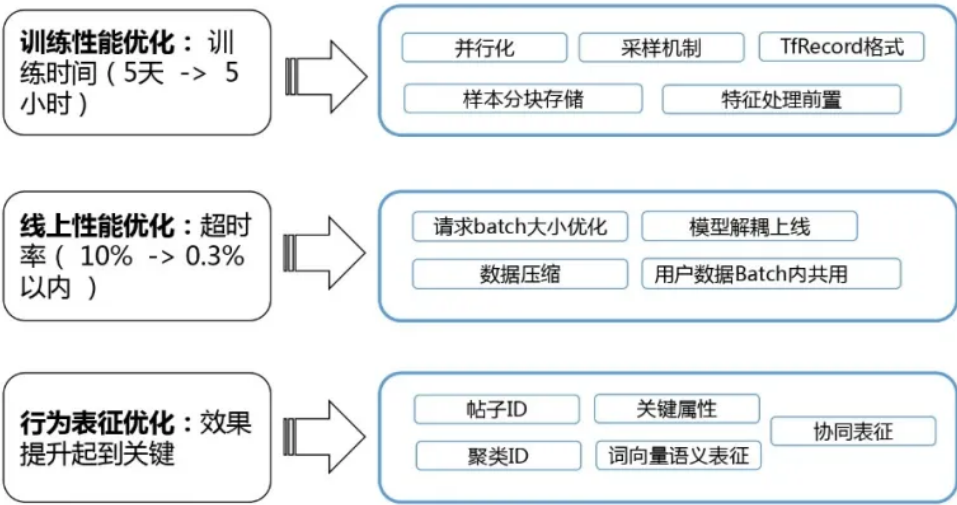
## 多通道模型工程化实践落地

接下来简单介绍一下实践落地过程中的工程化的工作。

### 实践落地：工程化

让生活简单美好

58



① 训练性能的优化。刚开始训练时长长达五天。通过持续的优化，目前控制在五个小时内，按天更新。期间的优化涉及到数据格式的改造，存储方式的优化，特征处理前置，还包括采样机制等优化手段。

② 线上性能的优化，刚上线时模型的超时率达到了10%。通过逐步调优，控制超时率在0.3%。其中的优化包括请求batch大小的优化，通过性能的逐步调试，目前一个batch的大小大控制在15~20之间。模型的解耦上线与目前平台无法支持大模型的上限，所以我们采用向量化的方式，把向量层剥离出来，

将参数存入Redis中来支持模型上线。另外还包括用户数据batch 内共用，以及数据压缩等方式实现性能优化。

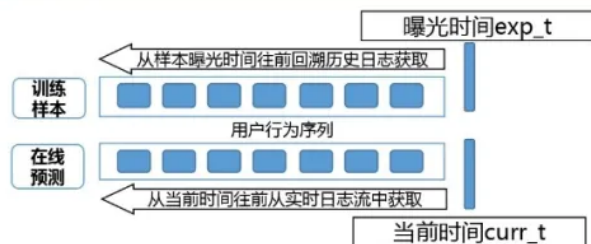
③ 模型调优中，行为表征优化也对模型效果提升起了关键作用，包括从帖子ID，关键属性到聚类ID以及基于词向量语义表征方式等。

④ 用户行为序列时效对齐问题

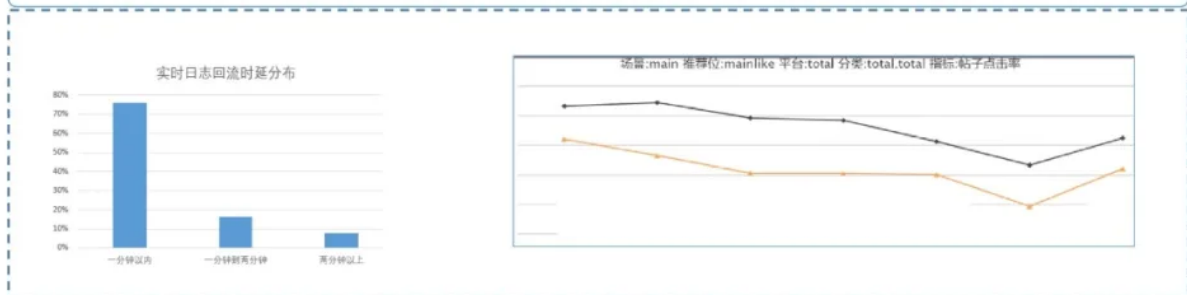
## 数据流：用户行为序列时效对齐

让生活简单美好 58

- **现象：**离线训练效果显著优于基线模型，上线后效果却并不理想
- **问题：**诊断发现用户行为序列，在训练阶段与预测阶段时间上未能对齐 -> **离线在线数据分布不一致**



**方案：**90%以上的行为日志回流时延在2分钟以内，训练样本中用户行为序列从曝光时间往前推两分钟



在离线训练时，效果优于基线模型的，但上线效果并不理想。在诊断时发现用户的行为序列在训练阶段与预测阶段时间是未能对齐的，即离线数据和在线数据分布不一致。但机器学习的核心假设，要求训练数据，预测数据是同分布的。我们可以看到训练样本是从样本曝光时间往前回溯历史日志获取到的用户行为序列。而在线预测时，用户的行为序列也是从当前的时间往前推，但是在对实时日志分析时发现，客户日志回流存在一个不确定性。这样的延时，说明预测时我们无法拿到最近的用户行为序列，而训练中使用的是过往数据，不存在日志回流延时的问题。换句话说，训练样本与在线预测数据并不是同分布的。为了解决这个问题，我们统计了日志回流的实验分布，发现90%以上的行为日志回流都在两分钟左右。

基于此发现，在构建训练样本时，我们将用户行为序列从曝光时间往前推两分钟。整体模型效果有比较明显的一个提升。

⑤ 模型工程化下的特征工程

实践：模型工程下的特征工程



行为序列中每个行为有不同的表征方式，我们用帖子ID，类目地域属性信息来综合表征帖子。其中帖子ID 的信息强度很强，区分度很高，但是数据比较稀疏，尤其是在转化行为上。类目地域属性信息的泛化能力很好，但是区分度又比较欠缺。因此我们在结合业务实践中，构建了一个聚类ID，作为对帖子重要的表征方式。基本的思路是对业务选择相应的核心属性，将核心属性的组合起来作为聚类ID，列出了几个主要业务的关键组合属性。以租房为例，我们把地域，价格，户型作为关键组合属性，其中价格会进行一个分层的操作，例如下面展示的帖子，它的关键属性组合就是九星桥6000~8000的一个两室一厅的房子，这就是聚类ID。这种表征方式在泛化能力和区分度上，能够做到比较好的平衡。这可以被我们认为是一种特征工程，虽然深度序列化这样一个模型，能够大幅降低特征工程的代价，把注意力更多地聚焦到模型上。但是特征工程在效果优化过程中仍发挥着不可忽视地作用。在模型中对行为的表征方式以及特征通道中特征交叉，仍然是属于比较广义的特征工程的范畴。

06

动力与展望

1. 动力：业务结合

# 关键驱动因素：业务结合

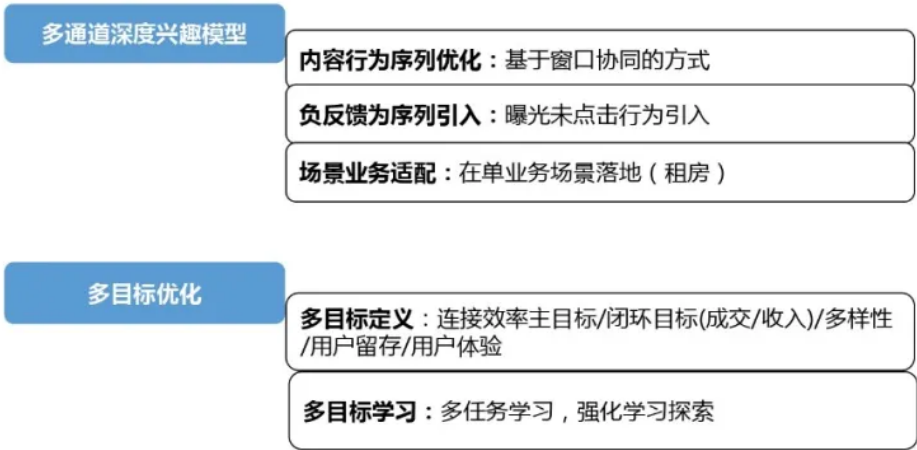
	用户需求集中度	用户行为追溯度	物料结构化程度	物料时效性特点	用户频次周期
电商	可发散性，逛逛买点别的	闭环	中（结构化信息与非结构化信）	生命期通常较长	中频，长期
资讯	兴趣偏好宽泛，有意思就好	引流	低（以非结构的文本图像视频）	更新速度较快	高频，长期
生活服务	聚焦，就是来租房/求职	半闭环	高（核心信息是结构化好的）	业务周期	低频，活跃周期

关键驱动因素是业务结合。从实践的经历中来看，我们发现模型优化与用户的需求集中度，行为可追溯度，物料的结构化程度，时效特点，用户使用频次与周期等都存在密切的关系。理解这些特点与业务形态特点，非常有助于我们在具体的业务场景下做更好的模型优化。多通兴趣模型提出和优化过程，是不断结合业务与融合业务特点的过程。我们58的业务是用户的频次低，没有足够的信息来刻画用户的兴趣，那么多行为序列的引入就成为我们主要探索的一个方向。58有固定的业务周期，以及明确的用户需求，所以我们并没有做过多的研究在怎么去扩展序列的长度。同时，58的帖子的信息结构化程度很高，在用户行为表征上，能比较方便去引入聚类ID的表征。在内容挖掘上，我们可以直接用核心属性来做表征。

## 2. 两个方向的规划



# 展望：后续工作



- ① 多目标优化，这也是我们未来重点的探索方向。虽然首页猜你喜欢显出了多目标平衡的特点，但是由于58的业务是非闭环的。所以我们无法直接跟踪到用户最终的交易行为，目标定义存在巨大的挑战。这一方面我们还在持续的探索，但是在模型上我们会进一步探索多任务的学习模型和强化学习。
- ② 多通道深度兴趣模型：包括内容行为序列优化，基于窗口协同方式的持续优化。引入负反馈行为序列，增加曝光未点击行为引入，以及在长期适配上能力的落地。

## 3. 总结

# 总结：首页猜你喜欢

让生活简单美好



首页猜你喜欢场景，过去一年半时间里，大盘效果，曝光转化率提升50+%

召回	排序	重排	展示
<ul style="list-style-type: none"><li>• <b>特色召回通道引入</b>：再营销召回，冷启动召回优化，场景化召回</li><li>• <b>向量化召回</b>：W2V → 双塔模型</li><li>• <b>召回配置化</b>：手工配置 → 数据驱动（半闭环）→ 自动调节（探索中）</li></ul>	<ul style="list-style-type: none"><li>• <b>模型架构优化</b>：双通道 → 多通道深度兴趣模型 → 多场景适配（探索中）</li><li>• <b>多目标优化</b>：现有模型架构基础上，采用ESSM多目标模型训练</li></ul>	<ul style="list-style-type: none"><li>• <b>打散</b>：规则打散 → 基于相似度的柔性打散策略</li><li>• <b>流量分配</b>：基于用户兴趣集中度的流量分配机制</li></ul>	<ul style="list-style-type: none"><li>• <b>样式对齐</b>：无图业务引入特色展示元素</li><li>• <b>推荐解释</b>：引入推荐解释，策略优化（生成策略，密度控制，维度优选）</li></ul>

在效果上来说，整体的大盘效果大概提升了50%。主要在四个层次上做了优化：

## ① 召回策略：

- 引入了特色的召回通道，比如说再营销召回，冷启动召回优化，场景化召回。
- 向量化召回，逐步从 Word2Vec 过渡到双塔模型。
- 召回配置化，从手工配置各个通道的额度和优先级，到逐步的是基于数据驱动热加载的方式进行配置，逐步探索自动调节和动态调节的可能性。

## ② 排序模型：

- 模型架构优化：从是双通道到多通道深度兴趣模型，再到多场景适配。
- 多目标优化：基于现有的模型架构，采用ESSM多目标模型来做训练，

## ③ 重排：从规则打散到基于相似度的柔性打散策略

## ④ 展示优化：

- 样式对齐：我们引入了特色展示元素。例如在招聘帖没有图片展示无法与其他的业务做对齐，那么用户体验就会比较差。所以经过多轮迭代之后，我们在招聘帖上引入图像元素信息来进行补充。最终发现在体验和最终的点击转化效果都有明显的提升。
- 推荐解释：在推进解释上的生成策略，密度控制，维度优化上做了进一步的优化。

今天的分享就到这里，谢谢大家。

---

在文末分享、点赞、在看，给个3连击呗~

---

### 嘉宾介绍：



罗景

58同城 | 高级架构师

58同城TEG推荐技术团队负责人/高级算法架构师。TEG推荐技术团队以提升连接效率/打磨关键技术能力/服务业务产品为团队愿景，在工程架构/模型算法/产品形态上持续迭代升级。目前，排序模型上，基于序列化模型组件等构建了适配58业务特点的多通道深度模型架构，成为了主要场景的主流线上模型，仍在进一步结合业务探索新模型架构/强化学习/迁移学习等的落地应用，期待和大家进一步探讨交流。

### 推荐阅读

深度学习技术在社会化推荐场景中的总结  
深度总结 | 多任务学习方法在推荐中的演变  
社会化推荐浅谈