

[论文阅读]阿里DIEN深度兴趣进化网络之总体解读

原创 罗西的思考 罗西的思考 10月31日

[论文阅读]阿里DIEN深度兴趣进化网络之总体解读

- 0x00 摘要
- 0x01 论文概要
 - 1.1 文章信息
 - 1.2 基本观点
 - 1.3 名词解释
- 0x02 总体概要
 - 2.1 模型架构
- 0x03 兴趣抽取层
 - 3.1 之前工作
 - 3.2 GRU
 - 3.3 辅助损失
 - 3.4 总结
- 0x04 兴趣进化层
 - 4.1 演化规律
 - 4.2 AUGRU
 - 4.3 attention
 - 4.4 特点
- 0x05 总结
- 0xFF 参考

0x00 摘要

之前我们介绍了阿里的深度兴趣网络(Deep Interest Network, 以下简称DIN), 一年后阿里再次升级其模型到深度兴趣进化网络(Deep Interest Evolution Network, 以下简称DIEN)。

本系列文章通过解读DIN & DIEN论文以及源码, 顺便梳理一些深度学习相关概念和TensorFlow的实现。

本文是系列第六篇：DIEN 论文解读，参考了大量文章，衷心感谢各位兄弟的分享，具体参见文末链接。

0x01 论文概要

1.1 文章信息

- 论文标题: Deep Interest Evolution Network for Click-Through Rate Prediction
- 论文地址: <https://arxiv.org/pdf/1809.03672.pdf>
- 代码地址: <https://github.com/mouna99/dien>

1.2 基本观点

1.2.1 DIN的问题

DIN忽略兴趣的变化。

- 用户的兴趣是不断变化的。例如用户对衣服的喜好，会随季节、时尚风潮以及个人品味的变化而变化，呈现一种连续的变迁趋势。
- 在淘宝平台中，用户的兴趣是丰富多样的，且每个兴趣的演变基本互不影响。
- 此外，影响最终行为的仅仅是与目标商品相关的兴趣。

1.2.2 DIEN创新

作者提出，以前的CTR预估方法都是直接将用户表现的表示向量当作兴趣，而没有通过具体的表现对隐藏的兴趣进行建模。因此提出DIEN【关键 - 兴趣直接导致了连续的行为，所以需要用户对用户兴趣及其演变建模，从用户历史行为中挖掘用户与目标商品相关的兴趣及演变】。

DIEN有两个关键模块：

- 一个是兴趣抽取层：
 - 从具体的用户表现中通过模拟用户兴趣迁移过程，抽取潜在的兴趣，主要是利用 GRU + 一个辅助损失（auxiliary loss）。即DIN没有考虑用户历史之间的时序关系，而DIEN则使用了GRU来建模用户历史的时间序列；
 - 直接使用GRU的缺陷：隐藏状态 只是表示捕捉行为之间依赖关系，并不能代表兴趣。且目标物体的点击是由最终的兴趣触发的，所以GRU只能学习行为之间的依赖，并不能很好反映用户兴趣；

- **创新**：由于每一步的兴趣状态都会直接导致下一个连续的行为。因此作者提出：辅助损失，使用下一个行为 来监督兴趣状态 的学习；
- 一个是兴趣进化层：
 - 兴趣的多样性会导致兴趣偏移的现象。在相邻的访问中，用户的意图可能非常不同，用户的一个行为可能依赖于很久以前的行为。
 - 基于从兴趣抽取层获得的兴趣序列基础上加入注意力机制，模拟与当前目标广告相关的兴趣进化过程。利用AUGRU（带有注意力机制更新门的GRU，attentional update gate，简称AUGRU）建模兴趣变化的过程；
 - AUGRU 增强了在 兴趣演化 中 相关兴趣 的影响，同时削弱了兴趣漂移所产生的非相关兴趣效应。通过在更新门中引入注意机制，AUGRU 可以实现针对不同目标物体的特定兴趣演化过程。

可以这样说，在训练的每一步中，我们为 interest extractor layer 引入了辅助loss。在 interest evolving layer 中加入了attention机制。

1.3 名词解释

隐式兴趣（latent interest）：用户与系统交互的行为是表达兴趣的载体。

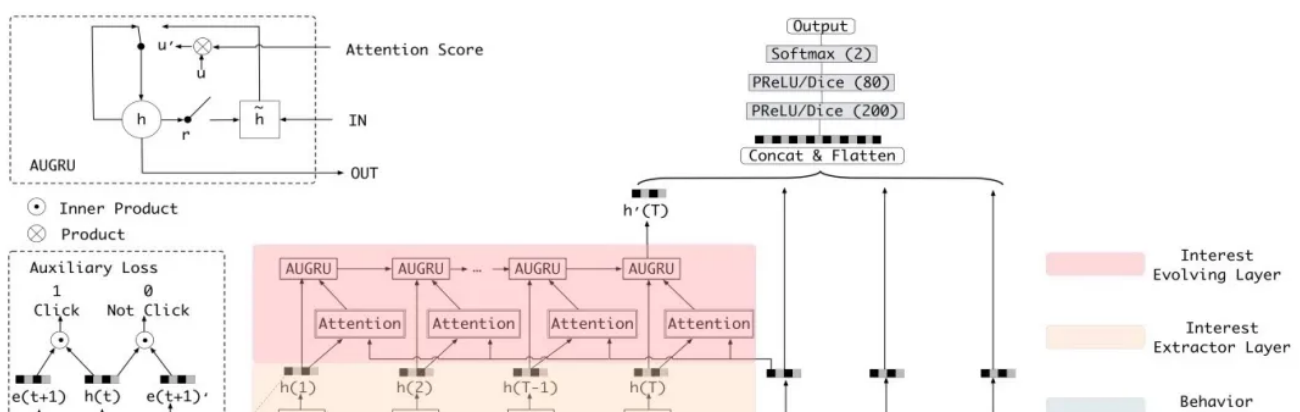
兴趣演变（Interest evolving）：受外在环境与内在认知变化的影响，用户的兴趣往往会随时间不断变化。以购买衣服为例，用户对兴趣的喜好，会随季节、时尚风潮以及个人品味的变化而变化。

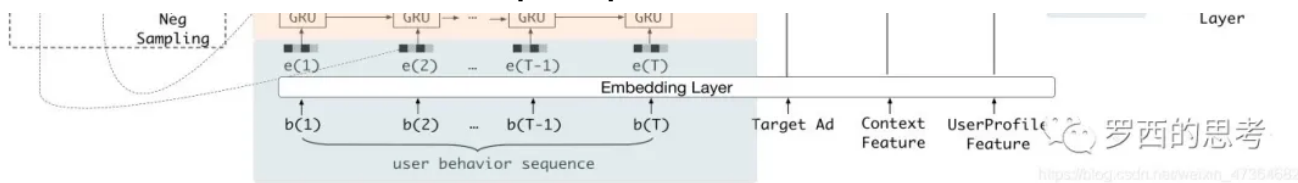
因此，想要做好点击率预估，必须在挖掘用户兴趣的同时，抓住用户兴趣的变化过程。

0x02 总体概要

2.1 模型架构

DIN架构如下：





在这里插入图片描述

与DIN类似，模型架构同样是由 输入层 + Embedding层 + 连接层 + 多层全连接神经网络 + 输出层的整体架构；

和DIN不同的是，DIEN把 user behavior 组织成序列数据的形式，并把简单的使用外积完成的 activation unit 变成一个 attention-based GRU 网络。

深度兴趣进化网络分为几层，从下到上依次是：

- 行为序列层（Behavior Layer）：主要作用是将用户浏览过的商品转换成对应的 embedding，并且按照浏览时间做排序，即把原始的id类行为序列特征转换成Embedding行为序列。
- 兴趣抽取层（Interest Extractor Layer）：主要作用是通过模拟用户的兴趣迁移过程，基于行为序列提取用户兴趣序列。
- 兴趣进化层（Interest Evolving Layer）：主要作用是通过在兴趣抽取层基础上加入 Attention 机制，模拟与当前目标广告相关的兴趣进化过程，对与目标物品相关的兴趣演化过程进行建模；
- 将兴趣表示和 ad、user profile、context 的 embedding 向量进行拼接。最后使用 MLP 完成最后的预测；

再详细点说：

- 用户历史肯定是一个时间序列，将其喂入 RNN，则最后一个状态可以认为包含了所有历史信息。因此，作者用一个双层的 GRU 来建模用户兴趣。
 - 将用户历史接触过的 item embedding 微量，喂进第一层 GRU，输出的就是用户各时刻的兴趣。这一层被称为 **Interest Extraction Layer**
 - 将第一层的输出，喂进第二层 GRU，并用 attention score（基于第一层的输出向量与候选物料计算得出）来控制第二层的 GRU 的 update gate。这一层叫做 **Interest Evolving Layer**。
- **Interest Evolving Layer** 的最后一个状态作为用户兴趣的向量表示，与 ad, context 的特征一同喂入 MLP，预测点击率。

0x03 兴趣抽取层

3.1 之前工作

DIEN论文中提到，之前的一些算法，通过引入RNN来发掘和利用行为序列（也就是浏览过的商品序列）中的依赖关系，比直接对行为序列做pooling效果更好。但是这些算法存在的问题就是直接将RNN的隐层输出作为用户的兴趣表示。而商品embedding才是对商品的真实表达，也是对用户兴趣的直接反应，RNN的隐层输出向量未必能够真正表达用户的兴趣。

因为观测到当前时刻的兴趣直接影响了下一时刻行为的发生，而GRU对兴趣表示的针对性弱。所以兴趣提取层引入了一个有监督学习，设计了auxiliary loss 损失函数，目的是用下一时刻的行为监督当前时刻兴趣的学习。这样就强行将RNN输出隐层与商品embedding发生交互，如架构图中左侧的Auxiliary Loss所示。

即：

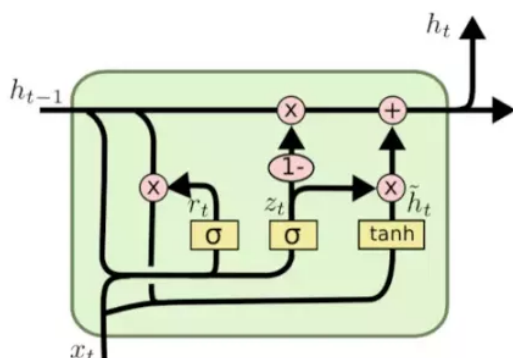
- 用户的行为是依时间产生的序列数据，所以使用GRU结构的RNN；
- 用户当前的兴趣直接导致下一行为的产生，所以设计了辅助损失函数，用下一刻的行为监督当前时刻兴趣的学习；

3.2 GRU

兴趣进化层的基本结构是GRU（Gated Recurrent Unit）网络，如架构中黄色区域所示，即用GRU来对用户行为之间的依赖性进行建模。

电子商务系统中的用户行为丰富，即使在很短的如两周这样的时间，其历史行为序列也很长。为了在效率和性能之间进行平衡，我们采用GRU对行为进行建模。

GRU的输入是用户按时间排序的行为序列，也就是行为对应的商品（按照时间步排列的商品embedding向量）。相比传统的序列模型RNN和LSTM，GRU解决了RNN的梯度消失问题；与LSTM相比，GRU的参数更少，训练收敛速度更快。



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

罗茜的思考
https://blog.csdn.net/weixin_47354682

在这里插入图片描述

参数解释如下：

- 假设第t个时间步输入e(t)，GRU输出隐单元h(t)；

- 令下一个时间步的输入向量 $\mathbf{e}(t+1)$ 作为正样本，随机采样负样本 $\mathbf{e}(t+1)'$ ，且 $\mathbf{e}(t+1)' \neq \mathbf{e}(t)$ ；
- $\mathbf{h}(t)$ 与正负样本向量分别做内积；

3.3 辅助损失

为了使序列结构模型的隐藏状态有效地表示潜在兴趣，应该对隐藏状态进行额外的监督，例如引入排名信息。在推荐系统中，排名损失已被广泛用于排名任务。

3.3.1 辅助损失

辅助loss来源于全部的点击记录，而非针对目标广告，有利于避免梯度消失，将有点击的作为正样本，未点击的作为负样本。

DIEN定义辅助损失如下：

$$L_{aux} = -\frac{1}{N} \left(\sum_{i=1}^N \sum_t \log \sigma(\mathbf{h}_t^i, \mathbf{e}_b^i[t+1]) + \log(1 - \sigma(\mathbf{h}_t^i, \hat{\mathbf{e}}_b^i[t+1])) \right)$$

在这里插入图片描述

3.3.2 全局损失

DIEN使用的全局损失函数如下：

- Label target 是CTR任务的loss函数；
- 将CTR的loss和辅助loss相加作为整个网络的loss进行优化；
- alpha 是平衡最终CTR预测和兴趣表示的超参数；

$$L = L_{target} + \alpha * L_{aux}$$

在这里插入图片描述

3.3.3 辅助损失作用



- [//mp.weixin.qq.com/s/IIVZCVtDco3hWuvnsUmekq](https://mp.weixin.qq.com/s/IIVZCVtDco3hWuvnsUmekq)

3.4 总结

经过GRU组成的兴趣抽取层后，用户的行为向量 $b(t)$ 被进一步抽象化，形成了兴趣状态向量 $h(t)$ 。

再用一句话总结下，兴趣提取层的作用是挖掘行为序列中商品之间的联系，对用户的兴趣进行提取和表达。

0x04 兴趣进化层

兴趣进化层 Interest Evolution Layer 的主要目标是刻画用户兴趣的进化过程。

用户兴趣是不断变化的：

- 用户在某一段时间的喜好具有一定的集中性。比如用户可能在一段时间内不断买书，在另一段时间内不断买衣服；
- 每种兴趣都有自己的演变趋势，不同种类的兴趣之间很少相互影响，例如买书和买衣服的兴趣基本互不相关；

用户这种变化会直接影响用户的点击决策。建模用户兴趣的进化过程有两方面的好处：

- 追踪用户的interest可以使我们学习final interest的表达时包含更多的历史信息；
- 可以根据interest的变化趋势更好地进行CTR预测；

4.1 演化规律

随着外部环境和内部认知的变化，用户兴趣也在不断变化，因此用户的行为受不同兴趣的影响。兴趣进化层相比兴趣抽取层的最大特点在于引入Attention机制，为了更有针对性地模拟与目标广告相关的兴趣进化路径。

推荐模型永远不能脱离具体业务场景，在阿里这种电商环境下，用户非常有可能同时对多品类商品感兴趣，比如在购买“机械键盘”的同时也在查看“衣服”品类下的商品。这样当目标商品是某电子产品时，“机械键盘”相关的兴趣演化路径就要比“衣服”相关的演化路径重要。

用户兴趣的演化具有如下规律：

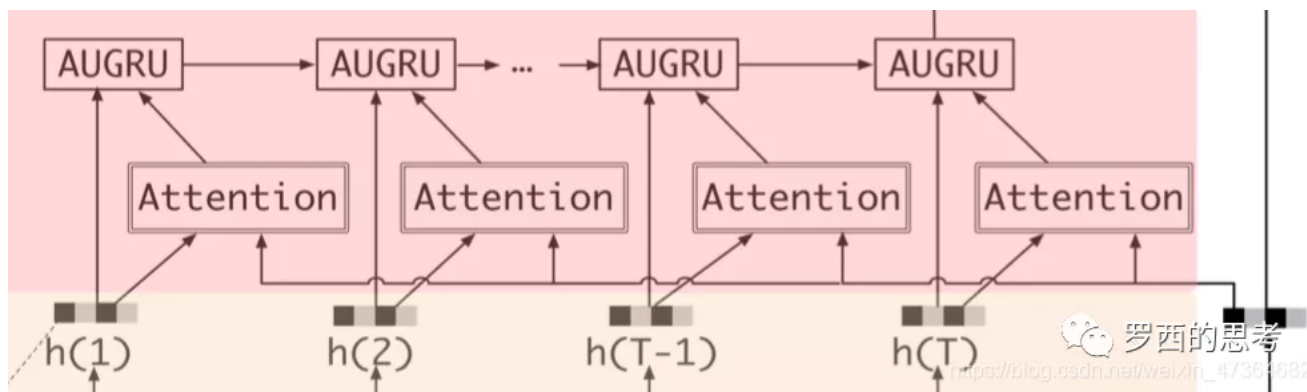
- Interest Drift: 由于兴趣的多样性，兴趣可能会漂移。 用户在某一段时间的interest会有一定的集中性。比如用户可能在一段时间内不断买书，在另一段时间内不断买衣服。
- Interest Individual: 一种interest有自己的发展趋势，不同种类的interest之间很少相互影响，例如买书和买衣服的interest基本互不相关。我们只关注与目标物品相关的演化过程。

4.2 AUGRU

基于以上规律，兴趣进化层通过 AUGRU（GRU with Attentional Update gate）来引入注意力机制，通过使用兴趣状态和 **target item** 计算得到相关性，AUGRU 增强相关兴趣的影响，同时减弱不相关兴趣的影响，进而去捕捉与目标商品相关的兴趣及其演变。

即通过分析兴趣演化的特征，作者结合注意力机制的局部激活能力和GRU的序列学习能力来对兴趣演化进行建模。这样，在GRU的每个步骤中，注意力机制都可以增强相对兴趣的影响，并减弱来自兴趣漂移的干扰。

有了用户的兴趣表示，兴趣发展层的作用就是捕获与candidate相关的兴趣发展模式，如架构图中红色区域所示，这里使用了第2个GRU。将 candidate 的 embedding 向量与第1个GRU的输出隐向量发生交互，生成attention 分数。需要注意的是，与 DIN 不同，这里的 attention 分数采用 softmax 做了归一化。attention 分数 反应了目标物品 和当前兴趣状态 的关系，相关性越大，则分数越大。



在这里插入图片描述

4.3 attention

attention计算如下：

$$a_t = \frac{\exp(\mathbf{h}_t W \mathbf{e}_a)}{\sum_{j=1}^T \exp(\mathbf{h}_j W \mathbf{e}_a)},$$

在这里插入图片描述

如何将attention机制加到GRU中呢？文中尝试了3种方法

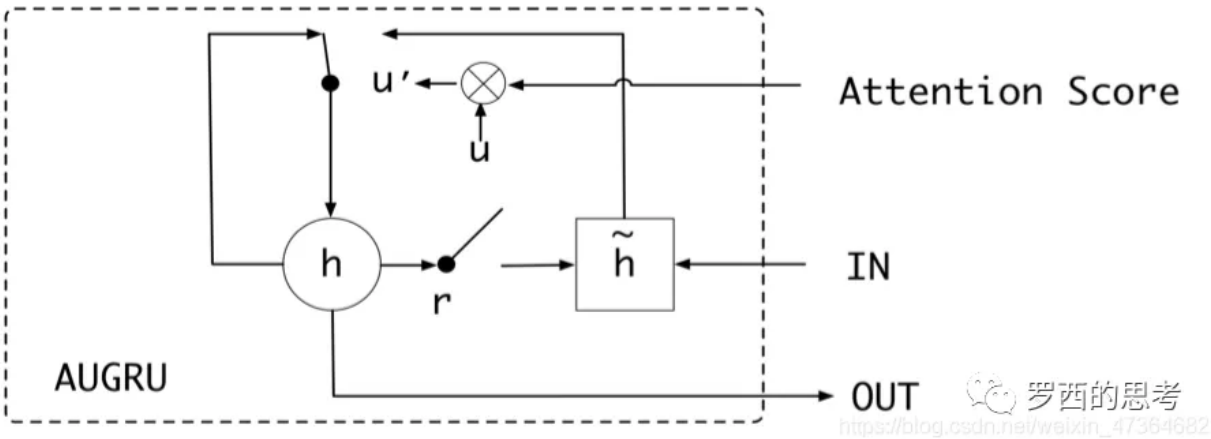
- **GRU with attentional input (AIGRU):** 在输入中结合注意力机制。

AIGRU使用注意力评分影响兴趣进化层的输入。即直接将attention系数和输入相乘。在理想的情况下，相关兴趣越少，输入值越小，那么我们就可以对与目标项目相关的兴趣进化趋势进行建模。然而，AIGRU的表现并不好。因为即使是零输入也会改变GRU的隐藏状态，所以相对兴趣越少，也会影响兴趣进化的学习。

- **Attention based GRU(AGRU):** 将注意力分数替换GRU的更新门（利用注意力score来控制hidden state的更新），直接地改变了隐藏状态。即直接将attention系数来替换GRU的update gate，直接对隐状态进行更新。

AGRU利用注意力得分直接控制隐藏状态的更新，削弱了兴趣演变过程中较少相关兴趣的影响。将注意力嵌入到GRU中可以改善注意力机制的影响力，并有助于AGRU克服AIGRU的缺陷。虽然AGRU可以使用注意分数直接控制隐藏状态的更新，但它使用一个标量（attention score）代替一个向量，忽略了不同维度间的重要性差异。

- **GRU with attentional update gate (AUGRU):** 在GRU的更新门中加入注意力分数:



在这里插入图片描述

$$\tilde{u}'_t = a_t * u'_t,$$
$$h'_t = (1 - \tilde{u}'_t) \circ h'_{t-1} + \tilde{u}'_t \circ \tilde{h}'_t,$$

罗西的思考

在这里插入图片描述

在AUGRU中，保留更新门的原始尺寸信息，通过注意力分数 来缩放更新门的所有维度，从而导致相关度较小的兴趣对隐藏状态的影响也较小。AUGRU可以更有效地避免利益漂移带来的干扰，并推动相对兴趣平稳发展。

AUGRU是效果最好的一种。文中将attention分数与update gate相乘，替换原始的update gate，称为AUGRU，其中A指的是Attention，U指的是Update gate。

我们看一下DIEN这个结构有什么问题。GRU是串行计算结构，要按照时间步一步一步进行计算。DIEN有两个GRU，第2个GRU还要基于第1个GRU的结果来做attention，所以第2个GRU必

须要等到第1个GRU全部计算完成才能开始计算，两个GRU单元无法并行计算，所以可能存在时延大的问题，序列越长，时延可能就越长。文中介绍说工业场景输入的序列长度是50，累积两个GRU的时延，相当于序列长度为100的时延。

4.4 特点

对兴趣进化建模的优点如下：

- 兴趣进化模块可以为最终兴趣的表示提供更多的相关历史信息
- 根据兴趣进化趋势预测目标项目的点击率比较好

兴趣进化层结合了注意力机制中的局部激活能力和GRU的序列学习能力来实现建模兴趣演化的目标。

0x05 总结

DIEN的主要贡献如下：

- 在电商系统中关注了兴趣进化现象，提出了一个新的网络架构对兴趣进化过程进行建模。兴趣进化模型使兴趣表示更丰富，CTR预测更准确。
- 与直接以行为为兴趣不同，DIEN中专门设计了兴趣抽取层。针对GRU的隐藏状态对兴趣表示较差的问题，提出了一种辅助损失。
- 设计了兴趣进化层。兴趣进化层有效地模拟了与目标项目相关的兴趣进化过程。

下一篇将介绍模型源码整体架构，敬请期待。

0xFF 参考

<https://blog.csdn.net/John159151/article/details/91377508>

【paper reading】Deep Interest Evolution Network for Click-Through Rate Prediction

也评Deep Interest Evolution Network

<https://zhuanlan.zhihu.com/p/134170462>

【读书笔记】Deep Interest Evolution Network for Click-Through Rate Prediction

人人都能看懂的LSTM

从动图中理解 RNN，LSTM 和 GRU

台大李宏毅机器学习(一)——RNN&LSTM

李宏毅机器学习(2016)

推荐系统遇上深度学习(二十四)-深度兴趣进化网络DIEN原理及实战！

from google.protobuf.pyext import _message，使用tensorflow出现 ImportError: DLL load failed

DIN 深度兴趣网络介绍以及源码浅析

https://blog.csdn.net/qq_35564813/article/details/90714056

阿里CTR预估三部曲(2)：Deep Interest Evolution Network for Click-Through Rate Prediction 简析

推荐系统遇上深度学习(二十四)-深度兴趣进化网络DIEN原理及实战！

【论文导读】2019阿里CTR预估模型—DIEN(深度兴趣演化网络)

阅读原文

喜欢此内容的人还喜欢

[从源码学设计]蚂蚁金服SOFARegistry之推拉模型

罗西的思考

川、拜大选之战，到底哪一方的支持者是低智？

徐某56

许倬云 | 现代文明正在走入秋季，寒冬已经不远

智纲智库