

112 | 阿里巴巴的广告点击率预估模型

2018-06-20 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 07:07 大小 3.26M



今天，我们继续来进行点击率预估的案例分析，结合三篇核心论文，来看一看阿里巴巴的广告预估又有哪些值得我们学习的地方。

多段线性模型

我们之前介绍了多个公司关于点击率或者转化率预估的案例。从这些案例中，你可能已经发现有两个非常重要的特征需要机器学习模型来处理。

第一，就是**数据中呈现的非线性化的关系**。也就是说，我们的模型必须在某一个地方考虑到特性之间的非线性表征，以及对于目标标签的非线性关系。

第二，就是**数据的不均衡以及数据的稀疏性**。有很多广告商是新广告商，很多广告是新广告。在这样的情况下，我们就必须要处理“冷启动”和“热启动”这两种局面。

在《从广告点击率预估的大规模数据中学习多段线性模型》（Learning Piece-wise Linear Models from Large Scale Data for Ad Click Prediction）[1] 这篇文章中，作者们提出了一种**多段线性模型**来解决我们刚刚说的这两个问题，这个模型简称为**LS-PLM**（Large Scale Piecewise Linear Model）。

LS-PLM 的核心思路其实非常直观。既然数据在整个空间里可能呈现非线性的关系，那么我们是否能够**把整个空间分割成较小的区域**，使得每个区域内依然可以使用**线性模型**来逼近这个区域内的数据点呢？其实在统计学习中，这种模型常常被叫作“**混合模型**”。在很多机器学习教科书中都会讲授的一种混合模型是“**高斯混合模型**”（Gaussian Mixture Model）。

LS-PLM 在这篇论文的实际应用中，基本上可以被理解成为一种**混合线性模型**。这个模型的一个子模型叫作“**分割函数**”，也就是模型需要学习每一个数据点到底是依赖于哪一个线性模型来进行预测的。当然，这个分割是一种概率的分割。实际上，每一个数据点都依赖所有的线性模型来进行预测，只不过对每个模型的依赖程度不一样。对于每一个不同的线性模型来说，最大的不同就是每一个模型有自己的系数。也就是说，之前只有一个全局模型并且只有一组系数，相比之下，这里有多组系数来决定模型的预测效果。很明显，**对于 LS-PLM 来说，每一个局部都是线性的，但是在整体上依然是一个非线性的模型**。

LS-PLM 还借助了**两种正则化机制**。一种叫作**L1 正则**，这种正则化主要是希望模型保留尽可能少的特性，从而达到对于模型特性的选择。另外，模型还采用了一种**L2,1 正则**的方法，这种方法的目的也是特性选择，但是希望能够把一组特性全部选择或者全部置零。

在实际的实验中，作者们尝试了不同数目的数据分割，从 2 个到 36 个不等。最终，他们发现当数据分割为 12 个的时候，模型的效果达到最优，而之后，模型效果并没有明显提升。最终推出模型的 AUC 比直接使用一个对数概率回归的全局模型，效果要好 1.4%。

广告点击率预估和图像处理的结合

我们在电商上购物，对于商品的图像会不会影响我们的点击或者购买，应该有一个直观的感受。那么在广告的点击率预估上，商品的图像特征对于模型性能上的提高到底有没有帮助呢？我们再来看一篇论文 [2]，在这篇文章中，阿里巴巴的工程师就尝试对这个问题进行回答。

这篇文章结合了近期好几个利用深度学习来进行图像处理和广告点击率预估的工作。首先，就是所有的特性都利用一个“嵌入层”（Embedding Layer）**把原始的特性转换成为数值特性**。这种思路我们在之前介绍文本处理，特别是 Word2Vec 的时候曾经进行了详细的讲解。而在这里，不管是文本信息还是图像信息，都根据自己的特点转换成为了数值特性。

这里我们要解决的一个核心问题，就是用户和广告之间的匹配问题，这篇论文模型是这么处理的。首先，对所有广告 ID 及其图像进行单独的嵌入。然后对用户过去的喜好，特别是对图像的喜好进行了另外的嵌入，然后这些嵌入向量形成用户的某种“画像”。用户的画像和广告信息的嵌入被直接串联起来，形成最终的特征向量。在此之上，利用一个多层的神经网络来学习最后的点击率的可能性。

在深度学习建模中，这种把多种来源不同的信息通过简单的拼接，然后利用多层神经网络来进行学习的方法非常普遍和实用。

在这篇论文的介绍中，除了在模型上对图像进行处理以外，还有一个创新，就是提出了一个叫“**高级模型服务器**”（Advanced Model Server），简称**AMS**的架构理念。AMS 是针对深度学习模型的大计算量而专门打造的计算体系。总体来说，**AMS 的目的是把深度学习模型中的很多基础步骤进行拆分，然后把这些步骤部署到不同的服务器上，从而能够把复杂的模型拆分成细小的可以互相交流的步骤。**

从最终的实验结果上来看，基于深度学习的模型要比对数几率回归的模型好 2~3%。整体上来看，利用了图像的模型要比没有利用图像的模型都要好，哪怕是线性模型也是一样的效果。但是，这个好的程度非常之小，基本上可以忽略不计。看来如何好好利用图像的信息，依然是一个比较大的挑战。

深度兴趣网络

我们刚才介绍的这种把其他信息和图像信息进行结合的方法，最近在一篇文章 [3] 中有一个总结。在这篇论文中，作者们提出了一种叫“**深度兴趣网络**”，或者简称**DIN**的架构。

DIN 依靠一种基本的模型架构，那就是先把所有的特性变换成嵌入向量，然后针对不同的特性进行划组，一些特性得以直接进入下一轮，另一些特性经过类似图像中的池化

（Pooling）操作抽取到更加高级的特性。之后，所有的特性都被简单串联起来，然后再经过多层的深度神经网络的操作。

DIN 在这个架构的基础上，提出了一种新的“**激活函数**”（Activation Function），叫 **DICE**，目的是可以在不同的用户数据中灵活选择究竟更依赖于哪一部分数据。可以说，在某种意义上，这个架构非常类似深度学习中比较火热的**Attention 架构**，其目的也是要看究竟那部分数据对于最终的预测更有效果。

从最后的实验中看，不管是在内部数据还是外部公开的例如 MovieLens 或者 Amazon 的数据上，基于 DIN 的模型都比线性模型和其他的深度学习的模型有显著的提高。

总结

今天我为你介绍了阿里巴巴广告点击率预估的核心算法。一起来回顾下要点：第一，我们讲了如何利用混合线性模型来引入非线性的因素从而提高预测效果。第二，我们聊了如何利用深度学习模型来对数据进行建模，谈到了图像在这里面起到的作用。

最后，给你留一个思考题，深度学习模型在点击率预估方面的最大优势是什么？又有什么劣势呢？

欢迎你给我留言，和我一起讨论。

参考文献

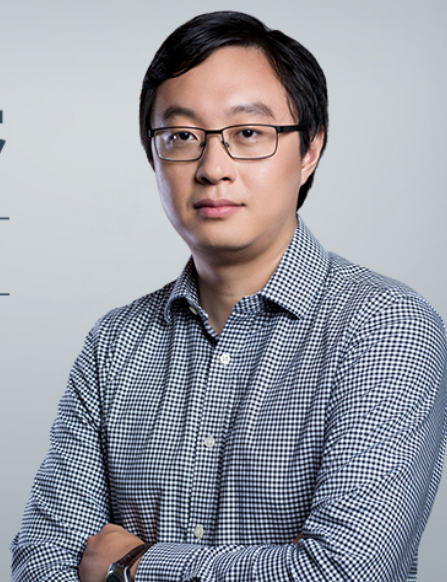
1. Kun Gai, Xiaoqiang Zhu, Han Li, Kai Liu, Zhe Wang. Learning Piece-wise Linear Models from Large Scale Data for Ad Click Prediction. CoRR abs/1704.05194 , 2017.
 2. Tiezheng Ge, Liqin Zhao, Guorui Zhou, Keyu Chen, Shuying Liu, Huiming Yi, Zelin Hu, Bochao Liu, Peng Sun, Haoyu Liu, Pengtao Yi, Sui Huang, Zhiqiang Zhang, Xiaoqiang Zhu, Yu Zhang, Kun Gai. Image Matters: Jointly Train Advertising CTR Model with Image Representation of Ad and User Behavior. CoRR abs/1711.06505 , 2017.
 3. Guorui Zhou, Chengru Song, Xiaoqiang Zhu, Xiao Ma, Yanghui Yan, Xingya Dai, Han Zhu, Junqi Jin, Han Li, Kun Gai. Deep Interest Network for Click-Through Rate Prediction. CoRR abs/1706.06978 , 2017.
-

AI 技术内参

你的360度人工智能信息助理

洪亮劼

Etsy 数据科学主管
前雅虎研究院资深科学家



新版升级：点击「👤 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 111 | Twitter的广告点击率预估模型

下一篇 113 | 什么是“基于第二价位的广告竞拍”？

精选留言 (2)

写留言



yy

2018-06-20



深度学习优势在特征工程方面可以学习到高阶特征，还可以单模型多目标优化业务，劣势是调参门槛高



yy

2018-06-20



深度学习的优势是在特征工程上能省去很多人为操作，模型能学习到高阶特征，很多业务场景还可以通过深度学习来实现单模型多目标优化。劣势是模型调参数门槛高。

展开 ▾

