



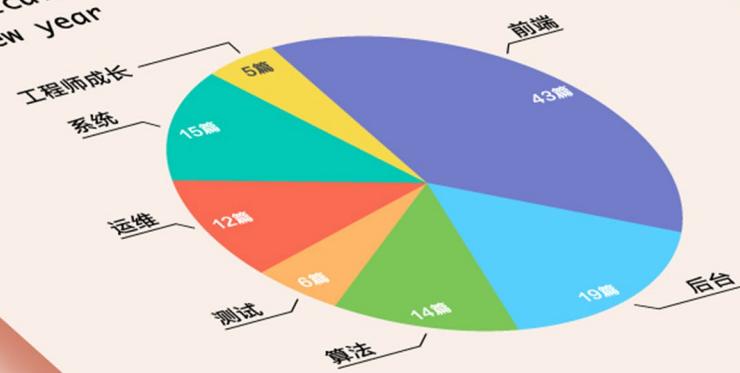
美团点评 2018 技术年货

CODE A BETTER LIFE



2018 美团技术团队答卷

System.out.println
("114 technical articles for you in 2018");
//Happy new year



序

春节已近，年味渐浓。

又到了我们献上技术年货的时候。

不久前，我们已经给大家分享了技术沙龙大套餐，汇集了过去一年我们线上线下技术沙龙 [99位讲师、85个演讲、70+小时](#) 分享。

今天出场的，同样重磅——技术博客全年大合集。

2018年，是美团技术团队官方博客第5个年头，[博客网站](#) 全年独立访问用户累计超过300万，微信公众号（meituantech）的关注数也超过了15万。

由衷地感谢大家一直以来对我们的鼓励和陪伴！

在2019年春节到来之际，我们再次精选了114篇技术干货，制作成一本厚达1200多页的电子书呈送给大家。

这本电子书主要包括前端、后台、系统、算法、测试、运维、工程师成长等7个板块。疑义相与析，大家在阅读中如果发现Bug、问题，欢迎扫描文末二维码，通过微信公众号与我们交流。

也欢迎大家转给有相同兴趣的同事、朋友，一起切磋，共同成长。

最后祝大家，新春快乐，阖家幸福。



目录 - 算法篇

美团“猜你喜欢”深度学习排序模型实践	4
美团大脑：知识图谱的建模方法及其应用	16
深度学习在美团搜索广告排序的应用实践	37
美团深度学习系统的工程实践	47
美团餐饮娱乐知识图谱——美团大脑揭秘	56
美团在O2O场景下的广告营销	68
美团外卖骑手背后的AI技术	85
基于TensorFlow Serving的深度学习在线预估	102
强化学习在美团“猜你喜欢”的实践	112
深度学习在文本领域的应用	125
机器学习在美团配送系统的实践：用技术还原真实世界	143
如何基于深度学习实现图像的智能审核？	154
深度学习在OCR中的应用	163
AI技术在智能海报设计中的应用	176

美团“猜你喜欢”深度学习排序模型实践

作者: 绍哲, 刘锐

引言

推荐作为解决信息过载和挖掘用户潜在需求的技术手段, 在美团点评这样业务丰富的生活服务电子商务平台, 发挥着重要的作用。在美团App里, 首页的“猜你喜欢”、运营区、酒店旅游推荐等重要的业务场景, 都是推荐的用武之地。



图1 美团首页“猜你喜欢”场景

目前, 深度学习模型凭借其强大的表达能力和灵活的网络结构在诸多领域取得了重大突破, 美团平台拥有海量的用户与商家数据, 以及丰富的产品使用场景, 也为深度学习的应用提供了必要的条件。本文将主要介绍深度学习模型在美团平台推荐排序场景下的应用和探索。

深度学习模型的应用与探索

美团推荐场景中每天活跃着千万级别的用户, 这些用户与产品交互产生了海量的真实行为数据, 每天能够提供十亿级别的有效训练样本。为处理大规模的训练样本和提高训练效率, 我们基于PS-Lite研发了分布式训练的DNN模型, 并基于该框架进行了很多的优化尝试, 在排序场景下取得了显著的效果提升。

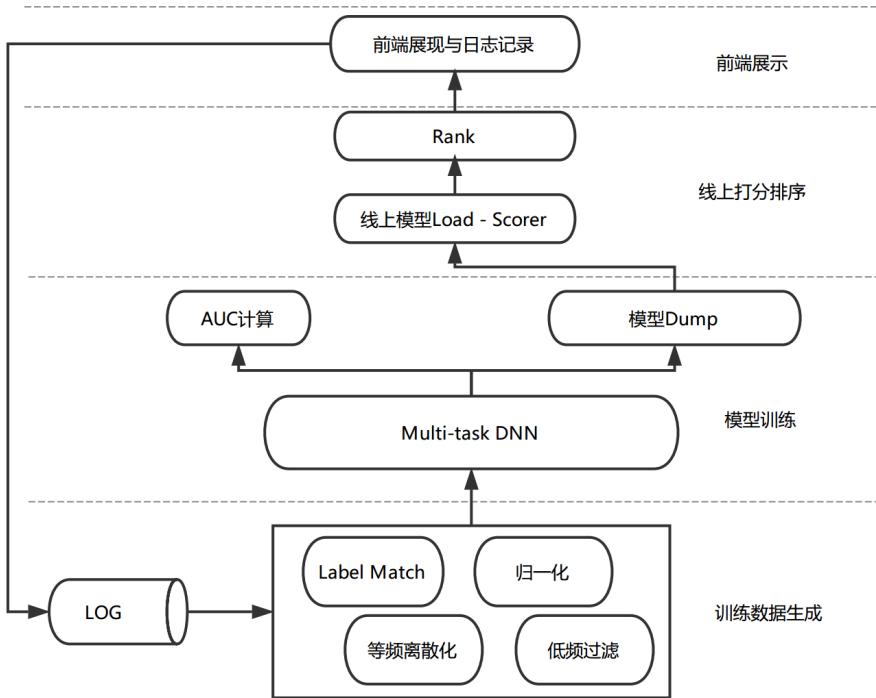


图2 模型排序流程图

如上图所示，模型排序流程包括日志收集、训练数据生成、模型训练和线上打分等阶段。当推荐系统对浏览推荐场景的用户进行推荐时，会记录当时的商品特征、用户状态与上下文信息，并收集本次推荐的用户行为反馈。在经过标签匹配和特征处理流程后生成最终的训练数据。我们在离线运用PS-Lite框架对Multi-task DNN模型进行分布式训练，通过离线评测指标选出效果较好的模型并加载到线上，用于线上排序服务。

下面将着重介绍我们在特征处理和模型结构方面所做的优化与尝试。

特征处理

美团“猜你喜欢”场景接入了包括美食、酒店、旅游、外卖、民宿、交通等多种业务，这些业务各自有着丰富的内涵和特点，同时各业务的供给、需求与天气、时间、地理位置等条件交织，构成了O2O生活服务场景下特有的多样性和复杂性，这就给如何更高效地组织排序结果提出了更高的要求。构造更全面的特征、更准确高效地利用样本一直是我们优化的重点方向。

特征种类

- User特征：用户年龄，性别，婚否，有无孩子等
- Item特征：价格，折扣，品类和品牌相关特征，短期和长期统计类特征等
- Context特征：天气，时间，地理位置，温度等
- 用户行为：用户点击Item序列，下单Item序列等

除上述列举的几类特征外，我们还根据O2O领域的知识积累，对部分特征进行交叉，并针对学习效果对特征进行了进一步处理。具体的样本和特征处理流程如下：

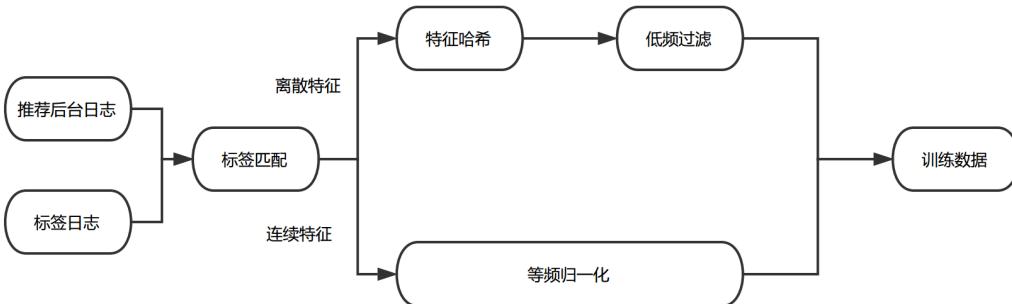


图3 训练数据处理流程

标签匹配

推荐后台日志会记录当前样本对应的User特征、Item特征与Context特征，Label日志会捕获用户对于推荐项的行为反馈。我们把两份数据按照唯一ID拼接到一起，生成原始的训练日志。

等频归一化

通过对训练数据的分析，我们发现不同维度特征的取值分布、相同维度下特征值的差异都很大。例如距离、价格等特征的数据服从长尾分布，体现为大部分样本的特征值都比较小，存在少量样本的特征值非常大。常规的归一化方法（例如 **min-max**, **z-score**）都只是对数据的分布进行平移和拉伸，最后特征的分布仍然是长尾分布，这就导致大部分样本的特征值都集中在非常小的取值范围内，使得样本特征的区分度减小；与此同时，少量的大值特征可能造成训练时的波动，减缓收敛速度。此外也可以对特征值做对数转化，但由于不同维度间特征的分布不同，这种特征值处理的方式并不一定适用于其他维度的特征。

在实践中，我们参考了Google的Wide & Deep Model[^6]中对于连续特征的处理方式，根据特征值在累计分布函数中的位置进行归一化。即将特征进行等频分桶，保证每个桶里的样本量基本相等，假设总共分了n个桶，而特征 x_i 属于其中的第 b_i ($b_i \in \{0, \dots, n - 1\}$)个桶，则特征 x_i 最终会归一化成 b_i/n 。这种方法保证对于不同分布的特征都可以映射到近似均匀分布，从而保证样本间特征的区分度和数值的稳定性。

低频过滤

过多的极为稀疏的离散特征会在训练过程中造成过拟合问题，同时增加参数的储存数量。为避免该问题，我们对离散特征进行了低频过滤处理，丢掉小于出现频次阈值的特征。

经过上述特征抽取、标签匹配、特征处理后，我们会给特征分配对应的域，并对离散特征进行Hash处理，最终生成LIBFFM格式的数据，作为Multi-task DNN的训练样本。下面介绍针对业务目标所做的模型方面的优化尝试。

模型优化与尝试

在模型方面，我们借鉴工业界的成功经验，在MLP模型的基础上，针对推荐场景进行模型结构方面的优化。在深度学习中，很多方法和机制都具有通用性，比如Attention机制在机器翻译，图像标注等方向上取得了显著的效果提升，但并不是所有具体的模型结构都能够直接迁移，这就需要结合实际业务问题，对引入的模型网络结构进行了针对性调整，从而提高模型在具体场景中的效果。

Multi-task DNN

推荐场景上的优化目标要综合考虑用户的点击率和下单率。在过去我们使用XGBoost进行单目标训练的时候，通过把点击的样本和下单的样本都作为正样本，并对下单的样本进行上采样或者加权，来平衡点击率和下单率。但这种样本的加权方式也会有一些缺点，例如调整下单权重或者采样率的成本较高，每次调整都需要重新训练，并且对于模型来说较难用同一套参数来表达这两种混合的样本分布。针对上述问题，我们利用DNN灵活的网络结构引入了Multi-task训练。

根据业务目标，我们把点击率和下单率拆分出来，形成两个独立的训练目标，分别建立各自的Loss Function，作为对模型训练的监督和指导。DNN网络的前几层作为共享层，点击任务和下单任务共享其表达，并在BP阶段根据两个任务算出的梯度共同进行参数更新。网络在最后一个全连接层进行拆分，单独学习对应Loss的参数，从而更好地专注于拟合各自Label的分布。

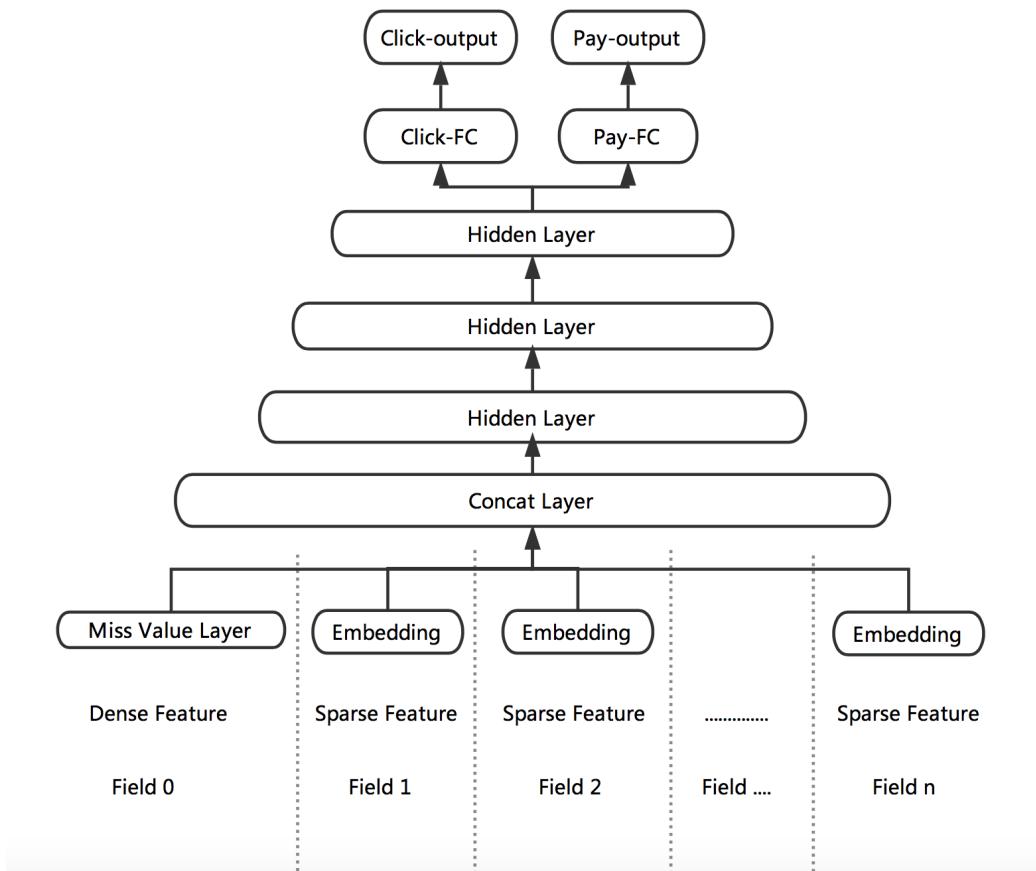


图4 点击与下单多目标学习

Multi-task DNN的网络结构如上图所示。线上预测时，我们将Click-output和Pay-output做一个线性融合。

在此结构的基础上，我们结合数据分布特点和业务目标进行了进一步的优化：针对特征缺失普遍存在的情况我们提出Missing Value Layer，以用更合理的方式拟合线上数据分布；考虑将不同task的物理意义关联起来，我们提出KL–divergence Bound，以减轻某单一目标的Noise的影响。下面我们就这两块工作做具体介绍。

Missing Value Layer

通常在训练样本中难以避免地有部分连续特征存在缺失值，更好地处理缺失值会对训练的收敛和最终效果都有一定帮助。通常处理连续特征缺失值的方式有：取零值，或者取该维特征的平均值。取零值会导致相应权重无法进行更新，收敛速度减慢。而取平均值也略显武断，毕竟不同的特征缺失所表示的含义可能不尽相同。一些非神经网络的模型能比较合理的处理缺失值，比如XGBoost会通过Loss的计算过程自适应地判断特征缺失的样本被划分到左子树还是右子树更优。受此启发，我们希望神经网络也可以通过学习的方式自适应地处理缺失值，而不是人为设置默认值。因此设计了如下的Layer来自适应的学习缺失值的权重：

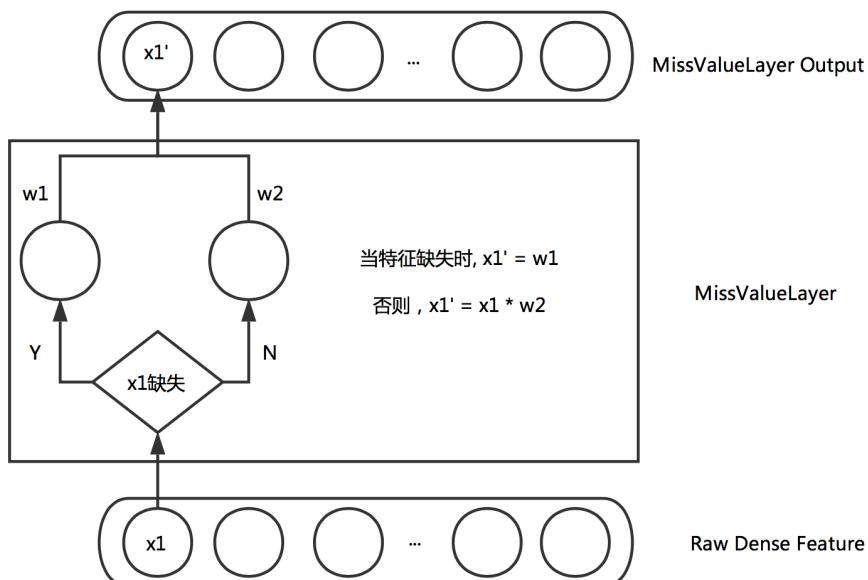


图5 Miss Value Layer

通过上述的Layer，缺失的特征可以根据对应特征的分布去自适应的学习出一个合理的取值。

通过离线调研，对于提升模型的训练效果，自适应学习特征缺失值的方法要远优于取零值、取均值的方式，模型离线AUC随训练轮数的变化如下图所示：

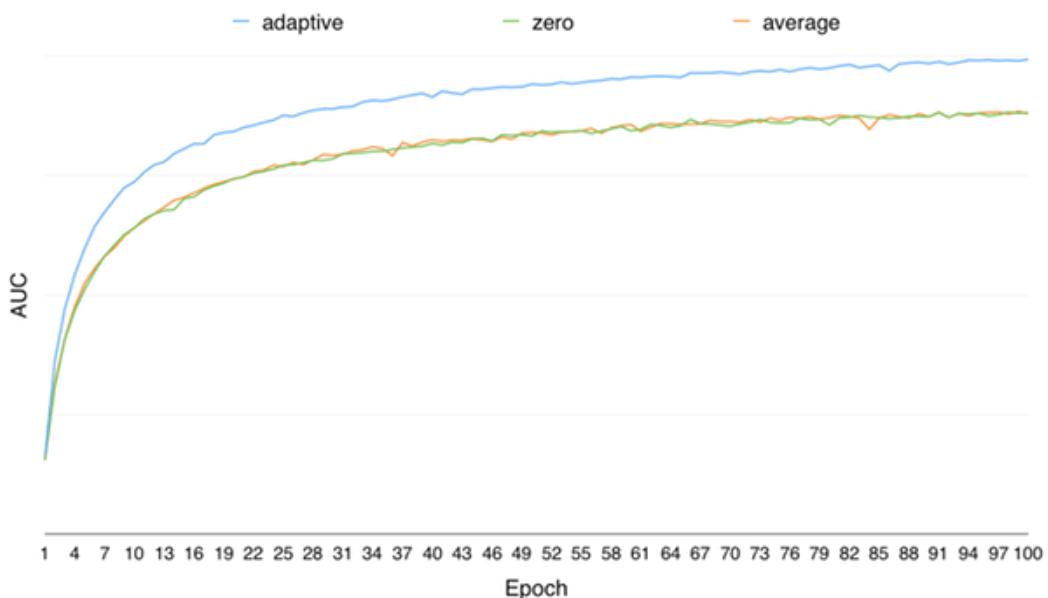


图6 自适应学习特征缺失值与取0值和均值效果对比

AUC相对值提升如下表所示：

	AUC
Zero	baseline
Average	0%
Adaptive	+ 0.4%

图7 自适应学习特征缺失值AUC相对值提升

KL–divergence Bound

我们同时考虑到，不同的标签会带有不同的Noise，如果能通过物理意义将有关系的Label关联起来，一定程度上可以提高模型学习的鲁棒性，减少单独标签的Noise对训练的影响。例如，可以通过MTL同时学习样本的点击率，下单率和转化率(下单/点击)，三者满足 $p(\text{点击}) * p(\text{转化}) = p(\text{下单})$ 的意义。因此我们又加入了一个KL散度的Bound，使得预测出来的 $p(\text{点击}) * p(\text{转化})$ 更接近于 $p(\text{下单})$ 。但由于KL散度是非对称的，即 $\text{KL}(p\|q) \neq \text{KL}(q\|p)$ ，因此真正使用的时候，优化的是 $\text{KL}(p\|q) + \text{KL}(q\|p)$ 。

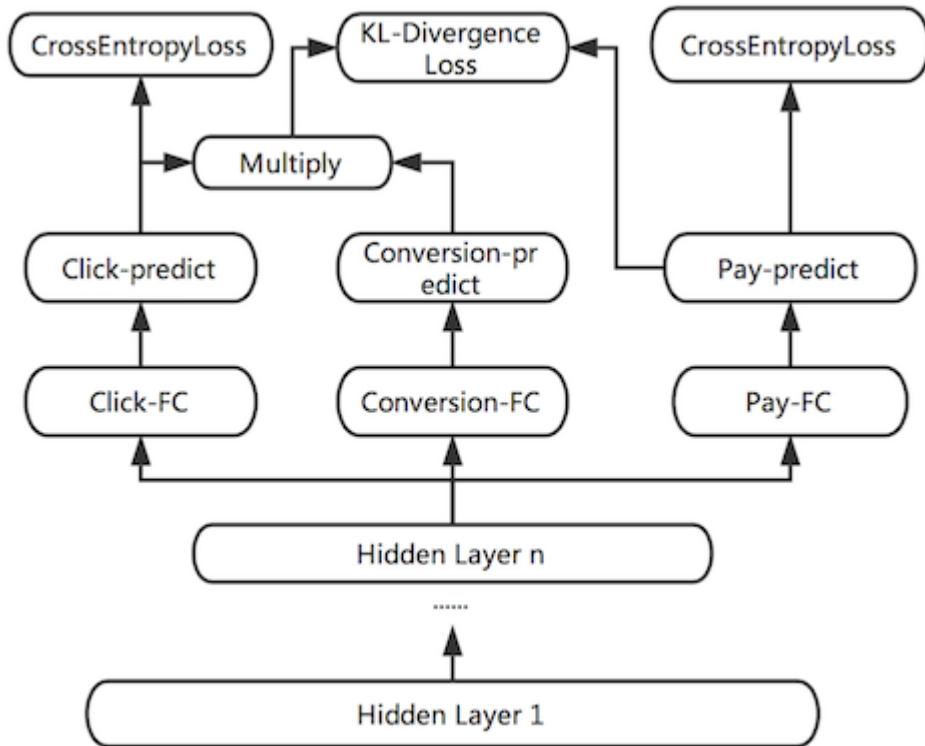


图8 KL-divergence Bound

经过上述工作，Multi-task DNN模型效果稳定超过XGBoost模型，目前已经在美团首页“猜你喜欢”场景全量上线，在线上也取得了点击率的提升：

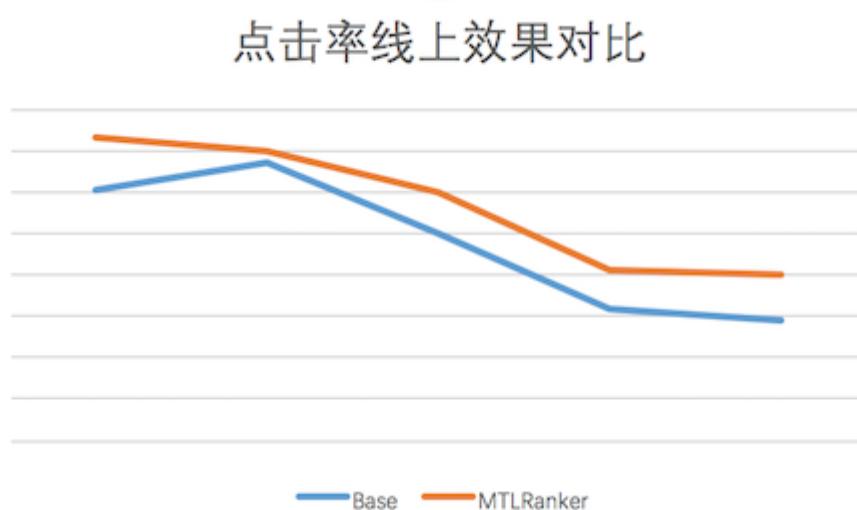


图9 线上CTR效果与基线对比图

线上CTR相对值提升如下表所示：

	Online CTR
Base	baseline
Multi-task DNN	+1.23%

图10 线上CTR效果相对值提升

除了线上效果的提升，Multi-task训练方式也很好的提高了DNN模型的扩展性，模型训练时可以同时考虑多个业务目标，方便我们加入业务约束。

更多探索

在Multi-task DNN模型上线后，为了进一步提升效果，我们利用DNN网络结构的灵活性，又做了多方面的优化尝试。下面就NFM和用户兴趣向量的探索做具体介绍。

NFM

为了引入Low-order特征组合，我们在Multi-task DNN的基础上进行了加入NFM的尝试。各个域的离散特征首先通过Embedding层学习得到相应的向量表达，作为NFM的输入，NFM通过Bi-Interaction Pooling的方式对输入向量对应的每一维进行2-order的特征组合，最终输出一个跟输入维度相同的向量。我们把NFM学出的向量与DNN的隐层拼接在一起，作为样本的表达，进行后续的学习。

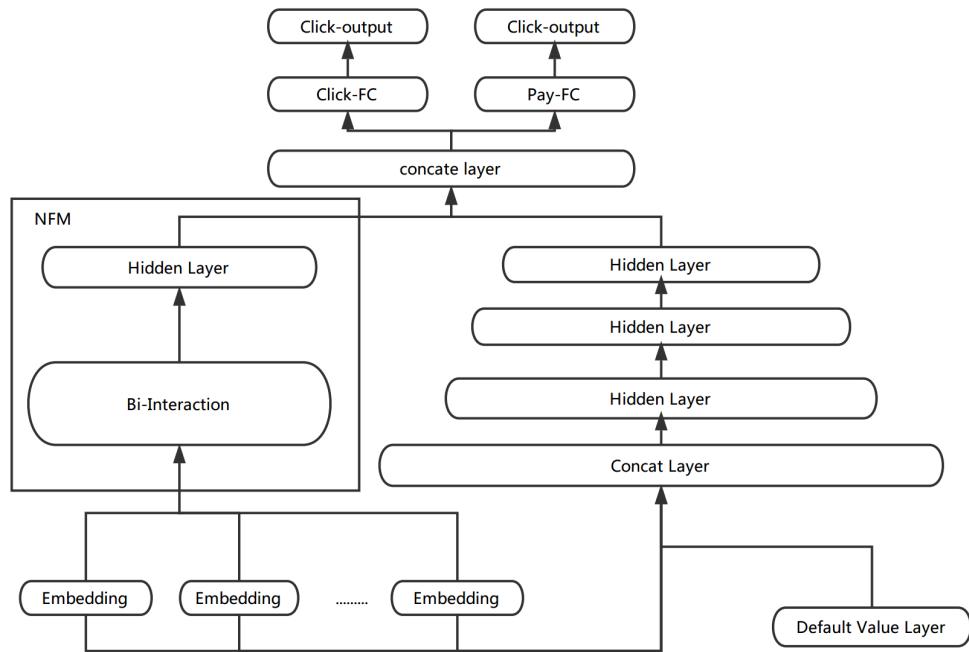


图11 NFM + DNN

NFM的输出结果为向量形式，很方便和DNN的隐层进行融合。而且从调研的过程中发现，NFM能够加快训练的收敛速度，从而更有利于Embedding层的学习。因为DNN部分的层数较多，在训练的BP阶段，当梯度传到最底层的Embedding层时很容易出现梯度消失的问题，但NFM与DNN相比层数较浅，有利于梯度的传递，从而加快Embedding层的学习。

通过离线调研，加入NFM后，虽然训练的收敛速度加快，但AUC并没有明显提升。分析原因是由于目前加入NFM模型部分的特征还比较有限，限制了学习的效果。后续会尝试加入更多的特征域，以提供足够的信息帮助NFM学出有用的表达，深挖NFM的潜力。

用户兴趣向量

用户兴趣作为重要的特征，通常体现在用户的历史行为中。通过引入用户历史行为序列，我们尝试了多种方式对用户兴趣进行向量化表达。

1. Item的向量化表达：线上打印的用户行为序列中的Item是以ID的形式存在，所以首先需要对Item进行Embedding获取其向量化的表达。最初我们尝试通过随机初始化Item Embedding向量，并在训练过程中更新其参数的方式进行学习。但由于Item ID的稀疏性，上述随机初始化的方式很容易出现过拟合。后来采用先生成item Embedding向量，用该向量进行初始化，并在训练过程中进行fine tuning的方式进行训练。
2. 用户兴趣的向量化表达：为生成用户兴趣向量，我们对用户行为序列中的Item向量进行了包括Average Pooling、Max Pooling与Weighted Pooling三种方式的融合。其中Weighted Pooling参考了DIN的实现，首先获取用户的行为序列，通过一层非线性网络(Attention Net)学出每个行为Item对于当前要预测Item的权重(Align Vector)，根据学出的权重，对行为序列进行Weighted Pooling，最终生成用户的兴趣向量。计算过程如下图所示：

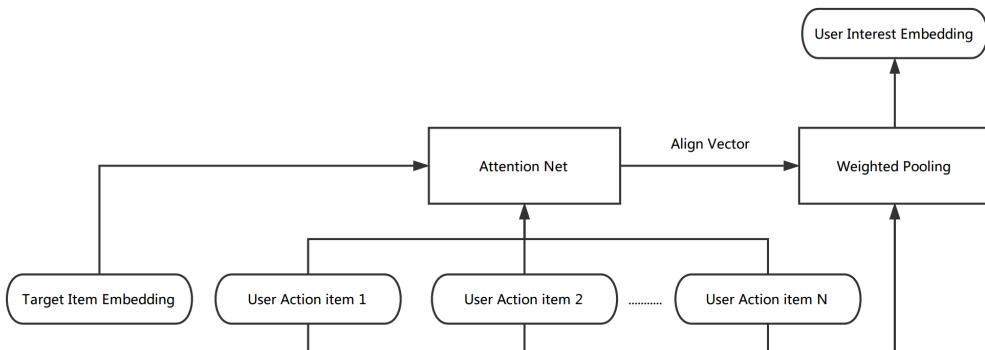


图12 Weighted Pooling

通过离线AUC对比，针对目前的训练数据，Average Pooling的效果为最优的。效果对比如下图所示：

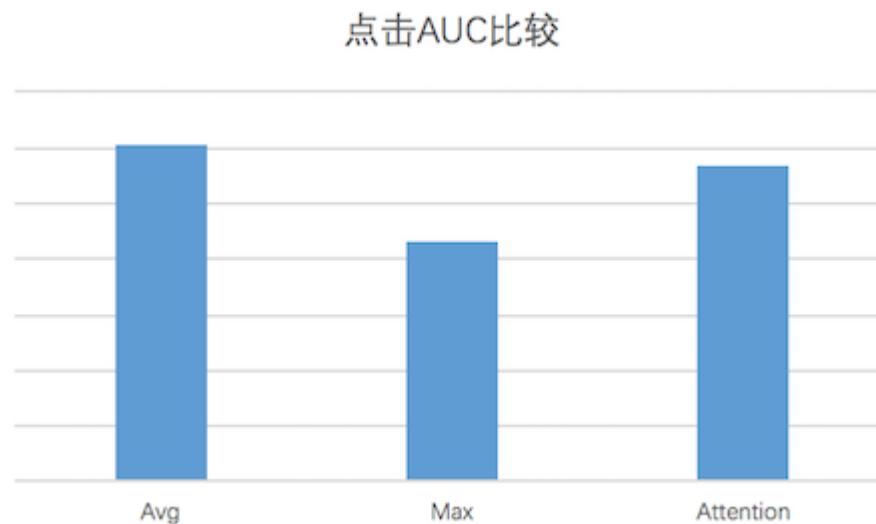


图13 不同Pooling方式点击AUC对比

以上是我们在模型结构方面的优化经验和尝试，下面我们将介绍针对提高模型训练效率所做的框架性能优化工作。

训练效率优化

经过对开源框架的广泛调研和选型，我们选择了PS-Lite作为DNN模型的训练框架。PS-Lite是DMLC开源的Parameter Server实现，主要包含Server和Worker两种角色，其中Server端负责模型参数的存储与更新，Worker端负责读取训练数据、构建网络结构和进行梯度计算。相较于其他开源框架，其显著优点在于：

- PS框架：PS-Lite的设计中可以更好的利用特征的稀疏性，适用于推荐这种有大量离散特征的场景。
- 封装合理：通信框架和算法解耦，API强大且清晰，集成比较方便。

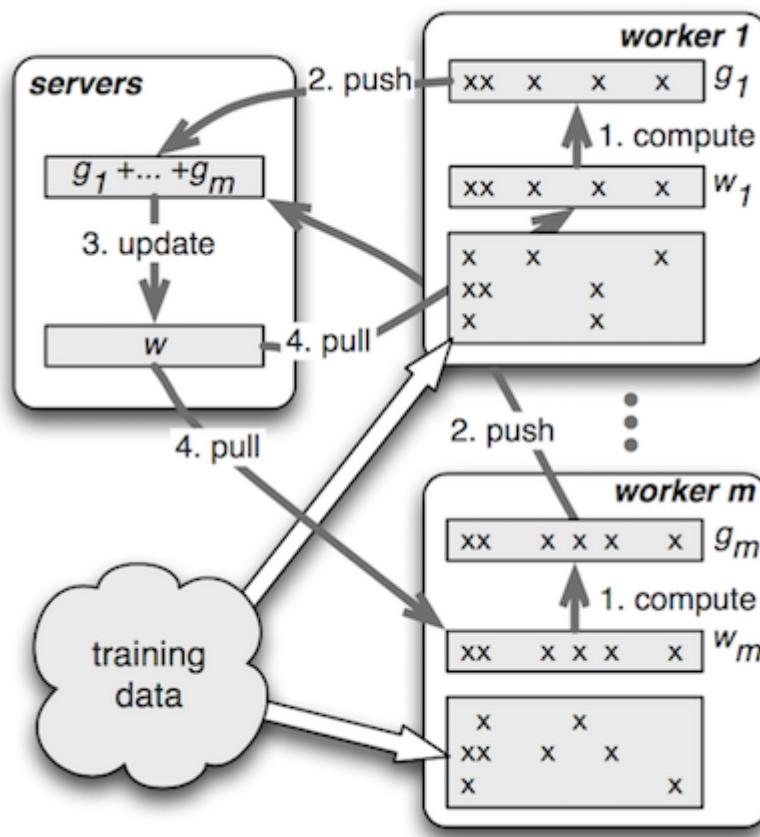


图14 Parameter Server

在开发过程中，我们也遇到并解决了一些性能优化问题：

1. 为了节约Worker的内存，通常不会将所有的数据储存在内存中，而是分Batch从硬盘中Pre-fetch数据，但这个过程中存在大量的数据解析过程，一些元数据的重复计算(大量的key排序去重等)，累计起来也是比较可观的消耗。针对这个问题我们修改了数据的读取方式，将计算过的元数据也序列化到硬盘中，并通过多线程提前将数据Pre-fetch到对应的数据结构里，避免了在此处浪费大量的时间来进行重复计算。

2. 在训练过程中Worker的计算效率受到宿主机实时负载和硬件条件的影响，不同的Worker之间的执行进度可能存在差异(如下图所示，对于实验测试数据，大部分Worker会在700秒完成一轮训练，而最慢的Worker会耗时900秒)。而通常每当训练完一个Epoch之后，需要进行模型的Checkpoint、评测指标计算等需要同步的流程，因此最慢的节点会拖慢整个训练的流程。考虑到Worker的执行效率是大致服从高斯分布的，只有小部分的Worker是效率极低的，因此我们在训练流程中添加了一个中断机制：当大部分的机器已经执行完当前Epoch的时候，剩余的Worker进行中断，牺牲少量Worker上的部分训练数据来防止训练流程长时间的阻塞。而中断的Worker在下个Epoch开始时，会从中断时的Batch开始继续训练，保证慢节点也能利用所有的训练数据。

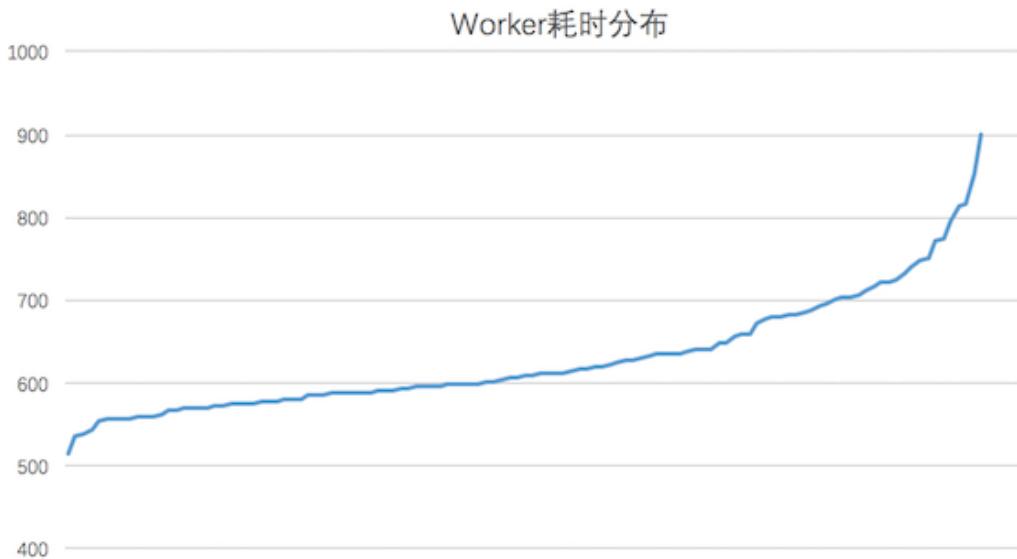


图15 Worker耗时分布

总结与展望

深度学习模型落地到推荐场景后，对业务指标有了明显的提升，今后我们还将深化对业务场景的理解，做进一步优化尝试。

在业务方面，我们将尝试对更多的业务规则进行抽象，以学习目标的方式加入到模型中。业务规则一般是我们短期解决业务问题时提出的，但解决问题的方式一般不够平滑，规则也不会随着场景的变化进行自适应。通过Multi-task方式，把业务的Bias抽象成学习目标，在训练过程中对模型的学习进行指导，从而可以比较优雅的通过模型解决业务问题。

在特征方面，我们会继续对特征的挖掘和利用进行深入调研。不同于其他推荐场景，对于O2O业务，Context特征的作用非常显著，时间，地点，天气等因素都会影响用户的决策。今后会继续尝试挖掘多样的Context特征，并利用特征工程或者模型进行特征的组合，用于优化对样本的表达。

在模型方面，我们将持续进行网络结构的探索，尝试新的模型特性，并针对场景的特点进行契合。学术界和工业界的成功经验都很有价值，给我们提供了新的思路和方法，但由于面临的业务问题和场景积累的数据不同，还是需要进行针对场景的适配，以达到业务目标的提升。

参考文献

- [1]: Mu Li, David G. Andersen, Alexander Smola, and Kai Yu. Communication Efficient Distributed Machine Learning with the Parameter Server. NIPS, 2014b.
- [2]: Rich Caruana. Multitask Learning. Betascript Publishing, 1997.
- [3]: Xiangnan He and Tat-Seng Chua. Neural Factorization Machines for Sparse Predictive Analytics. Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval(2017).
- [4]: Guorui Zhou, Chengru Song, et al. Deep Interest Network for Click-Through Rate Prediction.arXiv preprint arXiv:1706.06978,2017.
- [5]: Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. ICLR '15, May 2015.
- [6]: H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, et al. Wide & deep learning for recommender systems. arXiv preprint arXiv:1606.07792, 2016.

作者简介

- 绍哲，2015年校招入职原美团，主要从事推荐排序模型相关的工作。
- 刘锐，先后在百度、阿里巴巴从事推荐业务相关工作，现就职于美团平台推荐技术中心，专注深度学习模型的研发工作。

招聘

美团平台推荐技术中心负责美团首页“猜你喜欢”、运营区、酒店旅游推荐等公司重要的业务场景，长期致力于推荐场景的深度优化，在挖掘用户吃喝玩乐需求的同时提供更好的产品体验。美团平台推荐技术中心高端算法职位持续招聘中！期待你的加入，与我们一起解决技术与业务问题，快乐工作，共同成长。
简历请投递至：caohao#meituan.com

美团大脑：知识图谱的建模方法及其应用

作者: 仲远

作为人工智能时代最重要的知识表示方式之一，知识图谱能够打破不同场景下的数据隔离，为搜索、推荐、问答、解释与决策等应用提供基础支撑。美团大脑围绕吃喝玩乐等多种场景，构建了生活娱乐领域超大规模的知识图谱，为用户和商家建立起全方位的链接。我们美团希望能够通过对应用场景下的用户偏好和商家定位进行更为深度的理解，进而为大众提供更好的智能化服务，帮大家吃得更好，生活更好。

近日，美团 AI 平台部 NLP 中心负责人、大众点评搜索智能中心负责人王仲远博士受邀在 AI 科技大本营做了一期线上分享，为大家讲解了美团大脑的设计思路、构建过程、目前面临的挑战，以及在美团点评中的具体应用与实践，其内容整理如下：

知识图谱的重要性



近年来，人工智能正在快速地改变人们的生活，可以看到各家科技公司都纷纷推出人工智能产品或者系统，比如说在 2016 年，谷歌推出的 AlphaGo，一问世便横扫整个围棋界，完胜了人类冠军。又比如亚马逊推出的 Amazon Go 无人超市，用户只需下载一个 App，走进这家超市，就可以直接拿走商品，无需排队结账便可离开，这是人工智能时代的“新零售”体验。又比如微软推出的 Skype Translator，它能够帮助使用不同语言的人群进行实时的、无障碍的交流。再比如说苹果推出的 Siri 智能助理，它让每一个用苹果手机的用户都能够非常便捷地完成各项任务。所有这些人工智能产品的出现都依赖于背后各个领域技术突飞猛进的进展，包括机器学习、计算机视觉、语音识别、自然语言处理等等。

美团点评在人工智能的布局

美团点评AI平台部NLP（自然语言处理）中心于2018年2月正式成立
愿景：用人工智能帮大家吃得更好，生活更好
NLP (Natural Language Processing)：语言是人类智慧的结晶，自然语言处理是人工智能中最为困难的问题之一，其核心是让机器像人类一样理解和使用语言

人工智能助理

用户餐饮娱乐的决策需求
基于对用户评价和商家的深刻理解，帮助用户快速完成决策

作为全球领先的生活服务电子商务平台，美团点评在人工智能领域也在积极地进行布局。今年2月份，AI平台部NLP中心正式成立，我们的愿景是用人工智能帮大家吃得更好，生活更好。语言是人类智慧的结晶，而自然语言处理是人工智能中最为困难的问题之一，其核心是让机器能像人类一样理解和使用语言。

我们希望在不久的将来，当用户发表一条评价的时候，能够让机器阅读这条评价，充分理解用户的喜怒哀乐。当用户进入大众点评的一个商家页面时，面对成千上万条用户评论，我们希望机器能够代替用户快速地阅读这些评论，总结商家的情况，供用户进行参考。未来，当用户有任何餐饮、娱乐方面的决策需求的时候，美团点评能够提供人工智能助理服务，帮助用户快速的进行决策。

人工智能两大技术驱动力

	深度学习	知识图谱
场景示例	人脸识别	语音助手
目前进展	在一些任务上已经接近或超过人类	在知识量上超过人类，在知识推理上不如人类
任务范围	面向具体任务，难以迁移	广泛适用于不同任务
可解释性	较难解释	可解释性强
数据量	海量训练数据	海量知识数据
未来趋势	未来深度融合	

深度学习（隐性模型）

- 面向某一个具体任务（如下围棋，识别猫，人脸识别等）
- 需要海量训练数据
- 需要强大的计算力

知识图谱（显性模型）

- 可广泛用于不同任务
- 从海量数据中进行知识学习和挖掘
- 可理解、可解释，类似人类的思考方式

所有这一切，都依赖于人工智能背后两大技术驱动力：**深度学习**和**知识图谱**。我们可以将这两个技术进行一个简单的比较：

我们将深度学习归纳为隐性的模型，它通常是面向某一个具体任务，比如说下围棋、识别猫、人脸识别、语音识别等等。通常而言，在很多任务上它能够取得非常优秀的结果，同时它也有非常多的局限性，比如说它需要海量的训练数据，以及非常强大的计算能力，难以进行任务上的迁移，而且可解释性比较差。

另一方面，知识图谱是人工智能的另外一大技术驱动力，它能够广泛地适用于不同的任务。相比深度学习，知识图谱中的知识可以沉淀，可解释性非常强，类似于人类的思考。

机器和人类分别是如何识别猫？

The diagram illustrates the machine learning process for cat recognition. It shows a grid of many cat images, followed by a neural network architecture with three partitions (Partition 1, Partition 2, Partition 3) processing the input. Arrows point from the network to two outcome nodes: '是猫' (Cat) and '不是猫' (Not Cat). Below this, a section titled '人如何识别“猫”' (How humans identify 'cat') shows two children looking at books, which then lead to images of a cat and a dog. A dashed arrow labeled '知识' (Knowledge) connects the cat and dog images, indicating the conceptual leap humans make based on prior knowledge.

- 2012年谷歌X实验室宣布使用“深度学习”技术让机器成功识别图片中的“猫”
 - 1000台服务器，16000个处理器连接而成，内部共有10亿个节点
 - 机器“阅读”了1000万张从YouTube从抽取的图片
- 人如何识别“猫”
 - 3岁的小朋友看了几张猫的图片，就可以识别出猫

我们可以通过上面的例子，来观察深度学习技术和人类是如何识别猫的，以及它们的过程有哪些区别。

2012 年，Google X 实验室宣布使用深度学习技术，让机器成功识别了图片中的猫。它们使用了 1000 台服务器，16000 个处理器，连接成一个 10 亿节点的人工智能大脑。这个系统阅读了 1000 万张从 YouTube 上抽取的图片，最终成功识别出这个图片中有没有猫。

我们再来看看人类是如何做的。对于一个 3 岁的小朋友，我们只需要给他看几张猫的图片，他就能够很快识别出不同图片中的猫，而这背后其实就是大脑对于这些知识的推理。

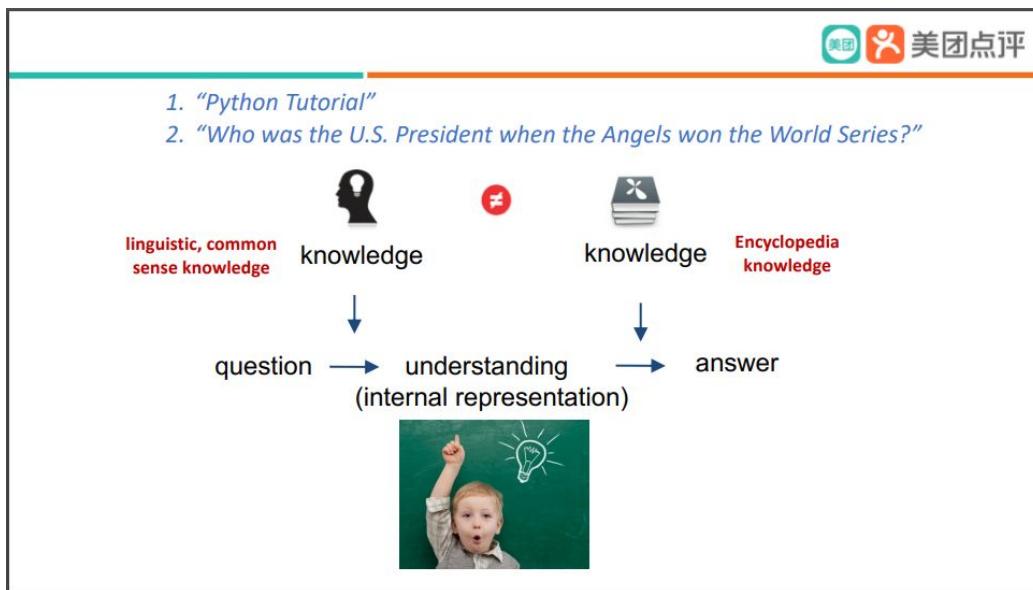
2011 年，Science 上有一篇非常出名的论文叫《How to Grow a Mind》。这篇论文的作者来自于 MIT、CMU、UC Berkeley、Stanford 等美国名校的教授。在这篇论文里，最重要的一个结论就是：**如果我们的思维能够跳出给定的数据，那么必须有 Another Source Of Information 来 Make Up The Difference。**

这里的知识语言是什么？对于人类来讲，其实就是我们从小到大接受的学校教育，报纸上、电视上看到的信息，通过社交媒体，通过与其他人交流，不断积累起来的知识。

近年来，不管是学术界还是工业界都纷纷构建自家的知识图谱，有面向全领域的知识图谱，也有面向垂直领域的知识图谱。其实早在文艺复兴时期，培根就提出了“知识就是力量”，在当今人工智能时代，各大科技公司更是纷纷提出：**知识图谱就是人工智能的基础。**



全球的互联网公司都在积极布局知识图谱。早在 2010 年微软就开始构建知识图谱，包括 Satori 和 Probbase。2012 年，Google 正式发布了 Google Knowledge Graph，现在规模已经达到 700 亿左右。目前微软和 Google 拥有全世界最大的通用知识图谱，Facebook 拥有全世界最大的社交知识图谱，而阿里巴巴和亚马逊则分别构建了商品知识图谱。



如果按照人类理解问题和回答问题这一过程来进行区分，我们可以将知识图谱分成两类。我们来看这样一个例子，如果用户看到这样一个问题，“Who was the U.S. President when the Angels won the World Series？”相信所有的用户都能够理解这个问题，也就是当 Angels 队赢了 World Series 的时候，谁是美国的总统？

这是一个问题理解的过程，它所需要的知识通常我们称之为 Common Sense Knowledge（常识性知识）。另外一方面，很多网友可能回答不出这个问题，因为它需要另外一个百科全书式的知识。

因此，我们将知识图谱分成两大类，一类叫 Common Sense Knowledge Graph（常识知识图谱），另外一类叫 Encyclopedia Knowledge Graph（百科全书知识图谱）。这两类知识图谱有很明显的区别。针对 Common Sense Knowledge Graph，通常而言，我们会挖掘这些词之间的 Linguistic

Knowledge；对于 Encyclopedia Knowledge Graph，我们通常会在乎它的 Entities 和这些 Entities 之间的 Facts。

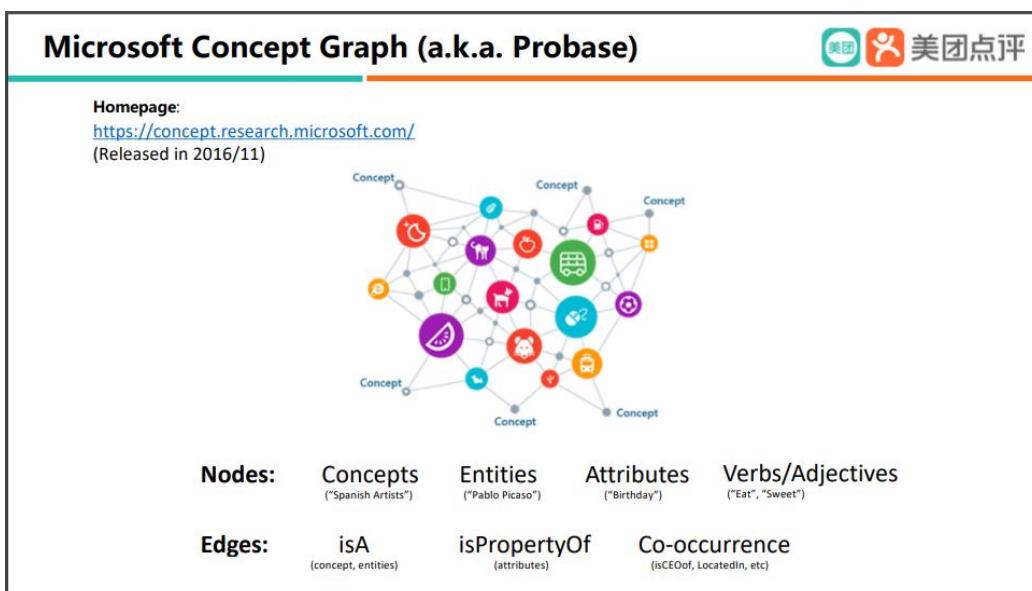
对于 Common Sense Knowledge Graph，一般而言我们比较在乎的 Relation 包括 isA Relation、isPropertyOf Relation。对于 Encyclopedia Knowledge Graph，通常我们会预定义一些谓词，比如说 DayOfbirth、LocatedIn、SpouseOf 等等。

对于 Common Sense Knowledge Graph 通常带有一定的概率，但是 Encyclopedia Knowledge Graph 通常就是“非黑即白”，那么构建这种知识图谱时，我们在乎的就是 Precision（准确率）。

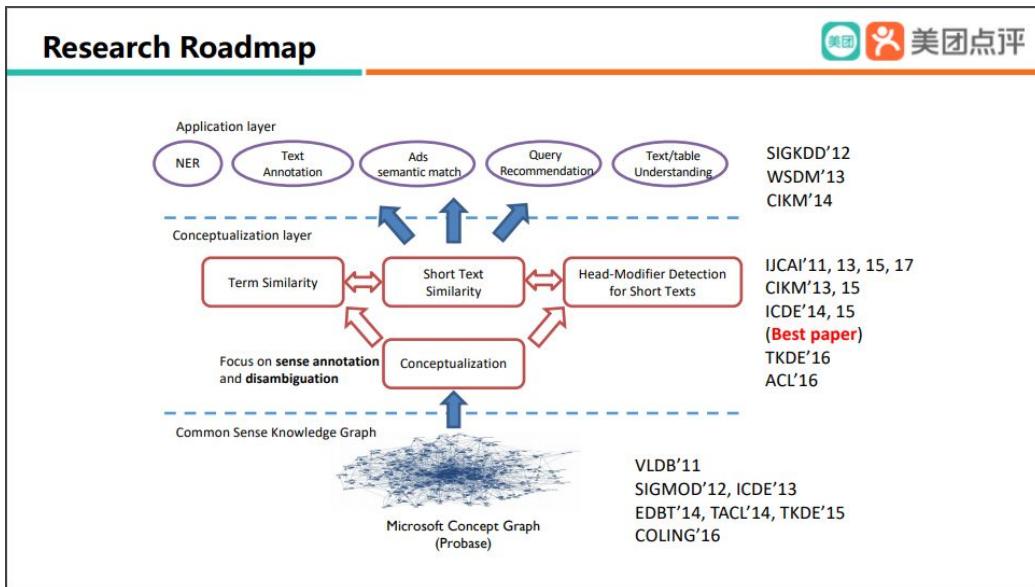
Common Sense Knowledge Graph 比较有代表性的工作包括 WordNet、KnowItAll、NELL 以及 Microsoft Concept Graph。而 Encyclopedia Knowledge Graph 则有 Freebase、Yago、Google Knowledge Graph 以及正在构建中的“美团大脑”。

这里跟大家介绍两个代表性工作：1) Common Sense Knowledge Graph: Probable；2) Encyclopedia Knowledge Graph: 美团大脑。

常识性知识图谱 (Common Sense Knowledge Graph)



Microsoft Concept Graph 于 2016 年 11 月正式发布，但是它早在 2010 年就已经开始进行研究，是一个非常大的图谱。在这个图谱里面有上百万个 Nodes（节点），这些 Nodes 有 Concepts（概念），比如说 Spanish Artists（西班牙艺术家）；有 Entities（实体），比如说 Picasso（毕加索）；有 Attributes（属性），比如 Birthday（生日）；有 Verbs（动词），有 Adjectives（形容词），比如说 Eat、Sweet。也有很多很多的边，最重要的边，是这种 isA 边，比如说 Picasso，还有 isPropertyOf 边。对于其他的 Relation，我们会统称为 Co-occurrence。



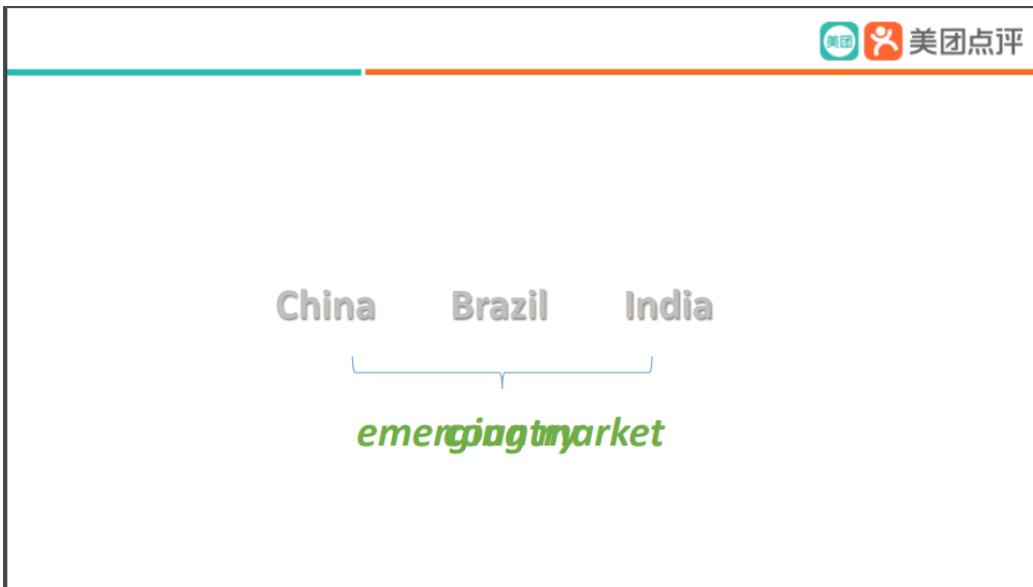
这是我们在微软亚洲研究院期间对 Common Sense Knowledge Graph 的 Research Roadmap（研究路线图）。当我们构建出 Common Sense Knowledge Graph 之后，重要的是在上面构建各种各样的模型。我们提出了一些模型叫 Conceptualization（概念化模型），它能够支持 Term Similarity、Short Text Similarity 以及 Head–Modifier Detection，最终支持各种应用，比如 NER、文本标注、Ads、Query Recommendation、Text Understanding 等等。

到底什么是 Short Text Understanding? 常识怎么用在 Text Understanding 中? 下面我们可以看一些具体的例子：

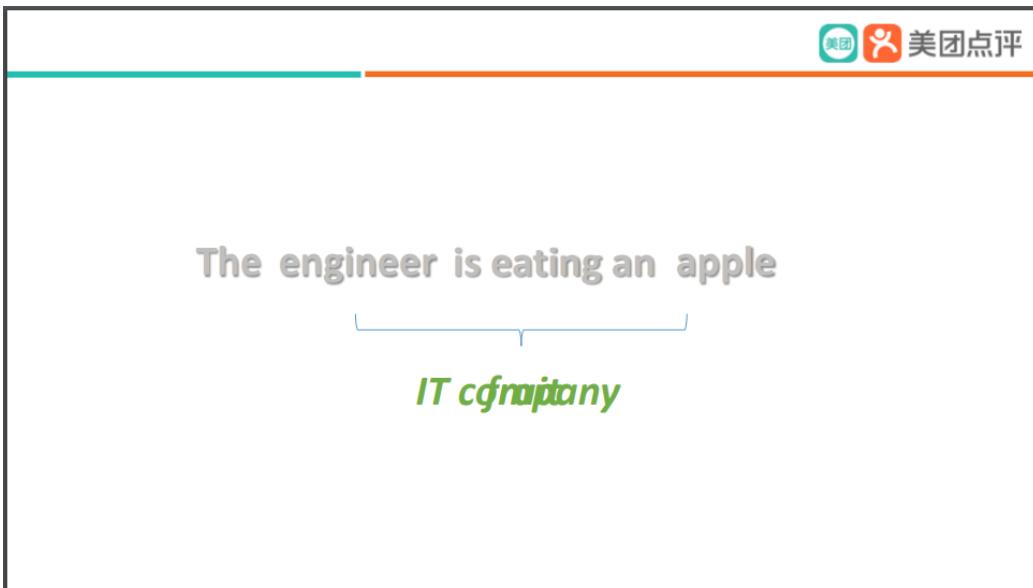
Add Common Sense to Computing

Pablo Picasso | 25 Oct 1881 | Spanish

当大家看到上面中间的文本时，相信所有人都能够认出这应该是一个日期，但是大家没办法知道这个日期代表什么含义。但如果再多给一些上下文信息，比如 Picasso、Spanish等等，大家对这个日期就会有一些常识性的推理。我们会猜测这个日期很可能是 Picasso 的出生日期，或者是去世日期，这就是常识。



比如说当我们给定 China 和 India 这两个 Entity 的时候，我们的大脑就会做出一些常识性的推理，我们会认为这两个 Entity 在描述 Country。如果再多给一个 Entity: Brazil，这时候我们通常会想到 Emerging Market。如果再加上 Russia，大家可能就会想到“金砖四国”或者“金砖五国”。所有这一切就是常识性的推理。



再比如，当我们看到 Engineer 和 Apple 的时候，我们会对 Apple 做一些推理，认为它就是一个 IT Company，但是如果再多给一些上下文信息，在这个句子里面由于 eating 的出现，我相信大家的大脑也会一样地做出常识推理，认为这个 Apple 不再是代表 Company，而是代表 Fruit。

Conceptualization

• Conceptualization: An explicit representation for the short text

$P(\text{concept} \mid \text{short text})$

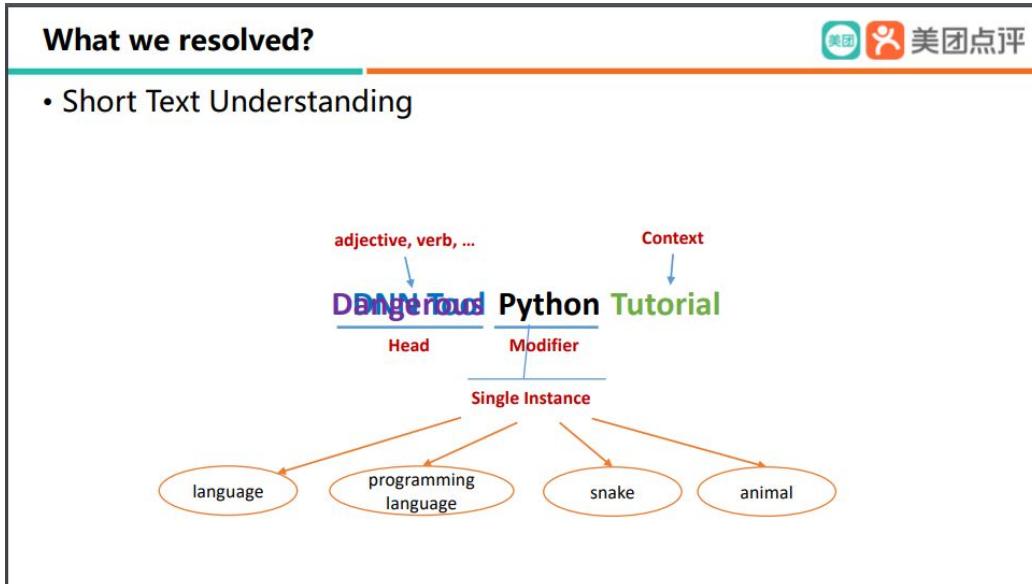
- Short text is sparse, noisy, and ambiguous
- Explicit means
 - Conceptualization results can be easily understood by human beings
 - Conceptualization model can be easily customized for different scenarios

所以，这就是我们提出来的 Conceptualization Model，它是一个 Explicit Representation。我们希望它能够将 Text，尤其是 Short Text，映射到 Millions Concepts，这样的 Representation 能够比较容易让用户进行理解，同时能够应用到不同场景当中。

Conceptualization

• Conceptualization: An explicit representation for the short text

ShortText: [pear apple]	Show Parameters	Conceptualize																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
Elapsed Time : 00:00:00.0140014																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
<table border="1"> <thead> <tr> <th></th> <th>pear</th> <th>apple</th> </tr> <tr> <th>[25/fruit]</th> <th>[25/fruit]</th> <th>[25/fruit]</th> </tr> </thead> <tbody> <tr> <td>fresh fruit</td> <td>0.5724769</td> <td>0.5718007</td> </tr> <tr> <td>tree fruit</td> <td>0.02916807</td> <td>0.02943546</td> </tr> <tr> <td>tree fruit</td> <td>0.01260049</td> <td>0.01677355</td> </tr> <tr> <td>dried fruit</td> <td>0.01165293</td> <td>0.01932771</td> </tr> <tr> <td>seasonal fruit</td> <td>0.01160144</td> <td>0.01539114</td> </tr> <tr> <td>juice</td> <td>0.01062348</td> <td>0.01546957</td> </tr> <tr> <td>hard fruit</td> <td>0.01011514</td> <td>0.01090830</td> </tr> <tr> <td>climacteric fruit</td> <td>0.00925454</td> <td>0.01297377</td> </tr> <tr> <td>fruit juice</td> <td>0.00904666</td> <td>0.01062579</td> </tr> <tr> <td>sweet fruit</td> <td>0.00892432</td> <td>0.01033749</td> </tr> <tr> <td>9405/food</td> <td>0.10539999</td> <td>0.00515157</td> </tr> <tr> <td>food</td> <td>0.03129783</td> <td>0.00844553</td> </tr> <tr> <td>high fiber food</td> <td>0.00863075</td> <td>0.01038461</td> </tr> <tr> <td>ingredient</td> <td>0.007349507</td> <td>0.009757149</td> </tr> <tr> <td>high-fiber food</td> <td>0.004699597</td> <td>0.004135749</td> </tr> <tr> <td>fresh food</td> <td>0.004038753</td> <td>0.004845139</td> </tr> <tr> <td>hard food</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low fat</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low carb</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sodium</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low cholesterol</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low sugar</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low calorie</td> <td>0.00355773</td> <td>0.003733393</td> </tr> <tr> <td>low salt</td> <td>0.003557</td></tr></tbody></table>		pear	apple	[25/fruit]	[25/fruit]	[25/fruit]	fresh fruit	0.5724769	0.5718007	tree fruit	0.02916807	0.02943546	tree fruit	0.01260049	0.01677355	dried fruit	0.01165293	0.01932771	seasonal fruit	0.01160144	0.01539114	juice	0.01062348	0.01546957	hard fruit	0.01011514	0.01090830	climacteric fruit	0.00925454	0.01297377	fruit juice	0.00904666	0.01062579	sweet fruit	0.00892432	0.01033749	9405/food	0.10539999	0.00515157	food	0.03129783	0.00844553	high fiber food	0.00863075	0.01038461	ingredient	0.007349507	0.009757149	high-fiber food	0.004699597	0.004135749	fresh food	0.004038753	0.004845139	hard food	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.00355773	0.003733393	low fat	0.00355773	0.003733393	low carb	0.00355773	0.003733393	low sodium	0.00355773	0.003733393	low cholesterol	0.00355773	0.003733393	low sugar	0.00355773	0.003733393	low calorie	0.00355773	0.003733393	low salt	0.003557
	pear	apple																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
[25/fruit]	[25/fruit]	[25/fruit]																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
fresh fruit	0.5724769	0.5718007																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
tree fruit	0.02916807	0.02943546																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
tree fruit	0.01260049	0.01677355																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
dried fruit	0.01165293	0.01932771																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
seasonal fruit	0.01160144	0.01539114																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
juice	0.01062348	0.01546957																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
hard fruit	0.01011514	0.01090830																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
climacteric fruit	0.00925454	0.01297377																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
fruit juice	0.00904666	0.01062579																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
sweet fruit	0.00892432	0.01033749																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
9405/food	0.10539999	0.00515157																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
food	0.03129783	0.00844553																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
high fiber food	0.00863075	0.01038461																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
ingredient	0.007349507	0.009757149																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
high-fiber food	0.004699597	0.004135749																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
fresh food	0.004038753	0.004845139																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
hard food	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low fat	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low carb	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sodium	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low cholesterol	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low sugar	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low calorie	0.00355773	0.003733393																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
low salt	0.003557																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												

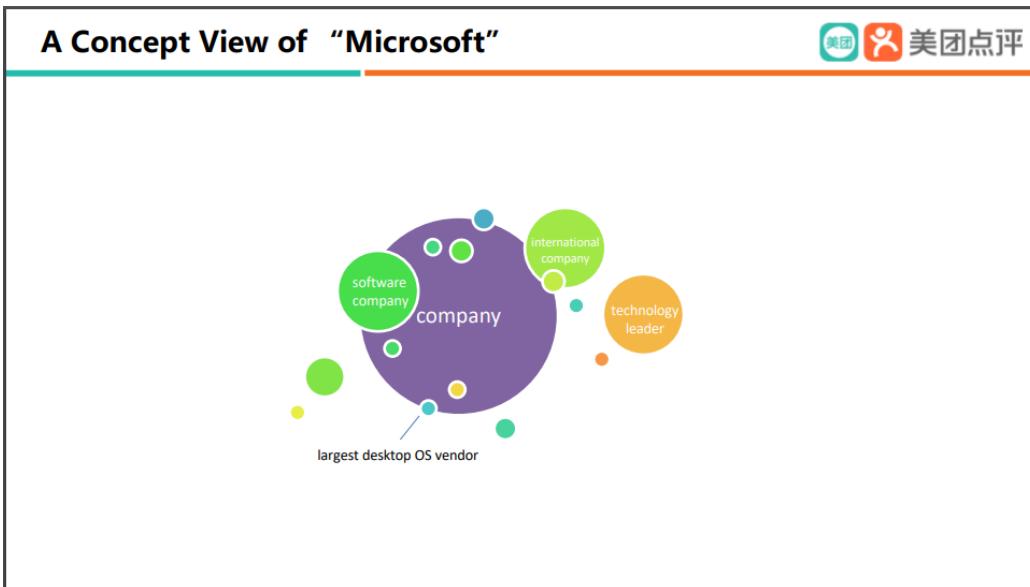


通过这样的一个 Conceptualization Model，我们能够解决什么样的文本理解问题？我们可以看这样一个例子。比如说给定一个非常短的一个文本 Python，它只是一个 Single Instance，那么我们会希望将它映射到至少两大类的 Concept 上，一种可能是 Programming Language，另外一种是 Snake。当它有一些 Context，比如说 Python Tutorial 的时候，那么这个时候 Python 指的应该是 Programming Language，如果当它有其他的 Adjective、Verb，比如有 Dangerous 时，这时候我们就会将 Python 理解为 Snake。

同时如果在一个文本里面包含了多个的 Entity，比如说 DNN Tool、Python，那么我们希望能够检测出在这个文本里面哪一个是比较重要的 Entity，哪一个是用来做限制的 Entity。

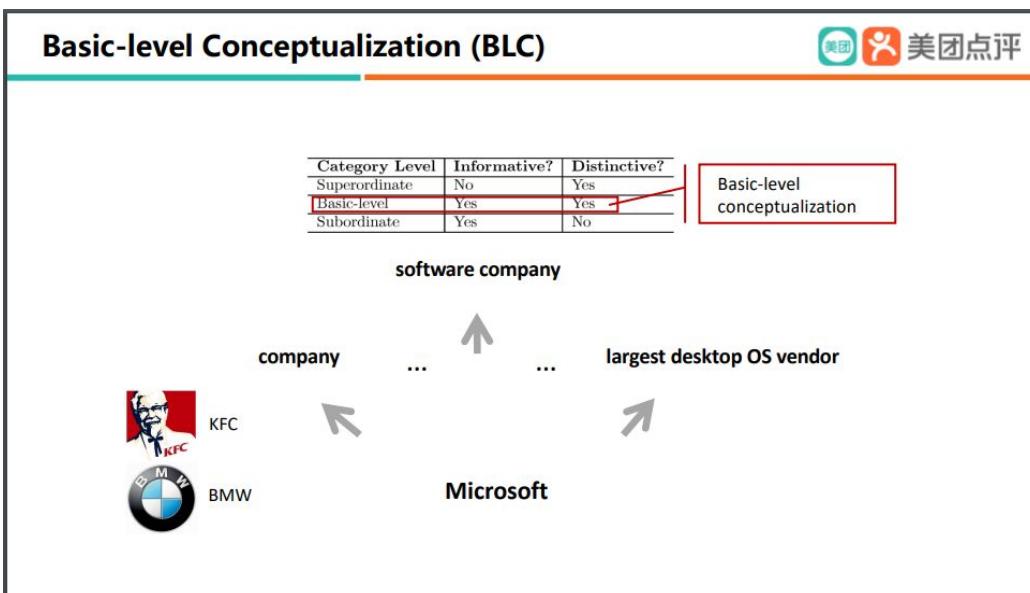
The screenshot shows a Google search results page for "Microsoft". The main search results include links to Microsoft's official website and various news articles. On the right side, there is a prominent "Knowledge Panel" for Microsoft Corporation. The panel includes the company's logo, a brief description ("Microsoft Corporation is an American multinational technology company with headquarters in Redmond, Washington. Wikipedia"), stock information (Stock price: MSFT (NASDAQ) US\$104.40 +0.84 (+0.78%), Last updated: 24 Aug, 4:00 PM GMT+4 - Disclaimer), CEO (Satya Nadella (Feb 4, 2014-)), Headquarters (Redmond, Washington, United States), and founders (Bill Gates, Paul Allen). It also lists partners like SAP, IBM, and Cisco Systems. Below the panel, there is a "Knowledge Panel" section with a list of entities related to Microsoft.

下面我们将简单地介绍一下，具体应该怎么做。当我们在 Google 里搜一个 Single Instance 的时候，通常在右侧会出现这个 Knowledge Panel。对于 Microsoft 这样一个 Instance，我们可以看到这个红色框所框出来的 Concept，Microsoft 指向的是 Technology Company，这背后是怎么实现的？



我们可以看到，Microsoft 实际上会指向非常非常多的 Concept，比如说 Company, Software Company, Technology Leader 等等。我们将它映射到哪一个 Concept 上最合适？

如果将它映射到 Company 这个 Concept 上，很显然它是对的，但是我们却没办法将 Microsoft 和 KFC、BMW 这样其他类型的产品区分开来。另外一方面，如果我们将 Microsoft 映射到 Largest Desktop OS Vendor 上，那么这是一个非常 Specific 的 Concept，这样也不太好，为什么？因为这个 Concept 太 Specific，太 Detail，它可能只包含了 Microsoft 这样一个 Entity，那么它就失去了 Concept 的这种抽象能力。



所以我们希望将 Microsoft 映射到一个既不是特别 General（抽象），又不是一个特别 Specific（具体）的 Concept 上。在语言学上，我们将这种映射称之为 Basic-level，我们将整个映射过程命名为 Basic-level Conceptualization。

我们提出了一种计算 Basic-level Conceptualization 的方法，其实它非常简单而且非常有效。就是将两种的 Typicality 做了一些融合，同时我们也证明了它们跟 PMI 和 Commute Time 之间的一些关联。并且在一个大规模的数据集上，我们通过 Precision 和 NDCG 对它们进行了评价。最后证明，我们所提出

来的 Scoring 方法，它在 NDCG 和 Precision 上都能达到比较好的结果。最重要的是，它在理论上是能够对 Basic-Level 进行很好的解释。

Using $Rep(e, c)$ for BLC

 美团点评

- Our measure $Rep(e, c) = P(c|e) * P(e|c)$ means:

Given e , the c should be its typical concept (shortest distance)

Given c , the e should be its typical entity (shortest distance)

A process of finding concept nodes having shortest expected distance with e
- (With PMI) If we take the logarithm of our scoring function, we get:

$$\log Rep(e, c) = \log P(c|e) * P(e|c) = \log \frac{P(e, c)}{P(e) * P(c)} = \log \frac{P(e, c)^2}{P(e)P(c)} = PMI(e, c) + \log P(e, c) = PMI^2$$
- (With Commute Time) The commute time between an instance e and a concept c is:

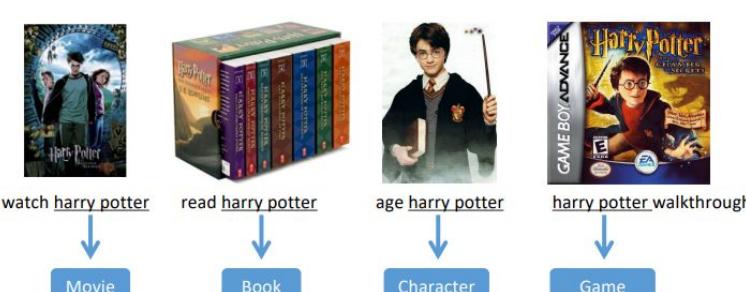
$$\begin{aligned} Time(e, c) &= \sum_{k=1}^{\infty} (2k) * P_k(e, c) = \sum_{k=1}^T (2k) * P_k(e, c) + \sum_{k=T+1}^{\infty} (2k) * P_k(e, c) \\ &\geq \sum_{k=1}^T (2k) * P_k(e, c) + 2(T+1) * (1 - \sum_{k=1}^T P_k(e, c)) = 4 - 2 * Rep(e, c) \end{aligned}$$

下面我们来看一下，当 Instance 有了一些 Context 之后，我们应该怎么去进行处理。我们通过一个例子，来简单地解释一下这背后最主要的思想。

比如说 iPad、Apple，其中 iPad 基本上是没有歧义的，它会映射到 Device、Product。但是对于 Apple 而言，它可能会映射到至少两类的 Concept 上，比如说 Fruit、Company。那么我们怎么用 iPad 对 Apple 做消歧呢？

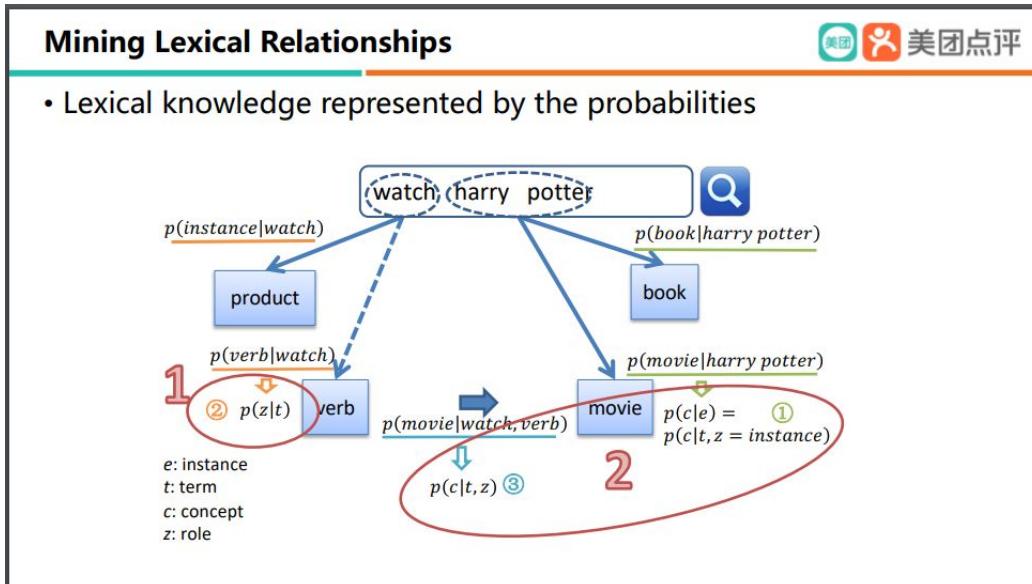
方法其实也挺直观的。我们会通过大量的统计去发现像 iPad 这样的 Entity，通常会跟 Company、Product 共同出现。比如说 iPad 有可能会跟三星共同出现，有可能会跟 Google 共同出现，那么我们就发现它会经常跟 Brand、Company、Product 共同出现。于是我们就利用新挖掘出来的 Knowledge 对 Apple 做消歧，这就是背后最主要的思想。

 美团点评



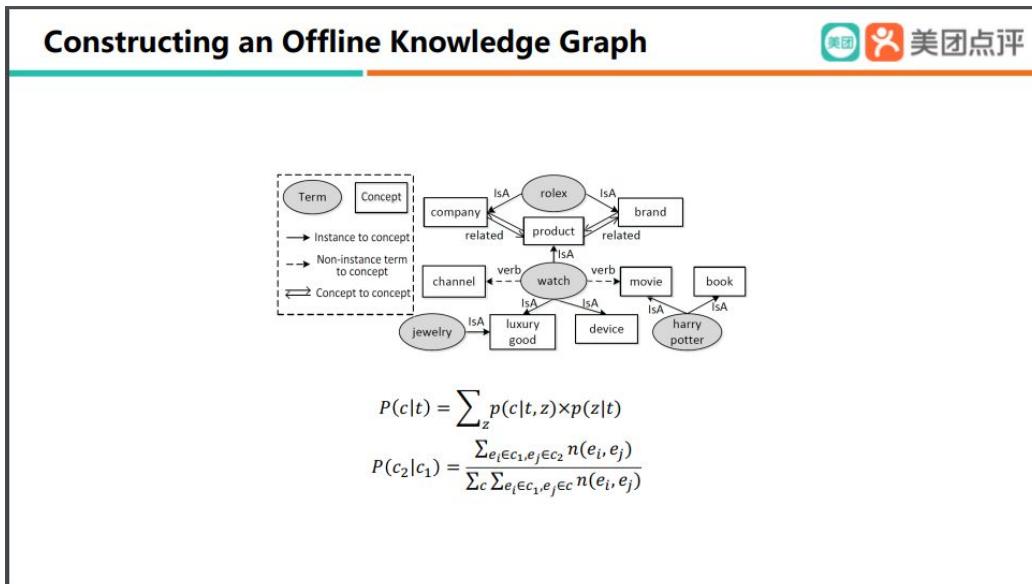
			
watch harry potter	read harry potter	age harry potter	harry potter walkthrough
			

除了刚才这样一个 General Context 以外，在很多时候这些 Text 可能还会包含很多一些特殊的类型，比如说 Verb、Adjective。具体而言，我们希望在看到 Watch Harry Potter 时，能够知道 Harry Potter 是 Movie，当我们看到 Read Harry Potter 时，能够知道 Harry Potter 是 Book。同样的，Harry Potter 还有可能是一个角色名称，或者是一个游戏名称。

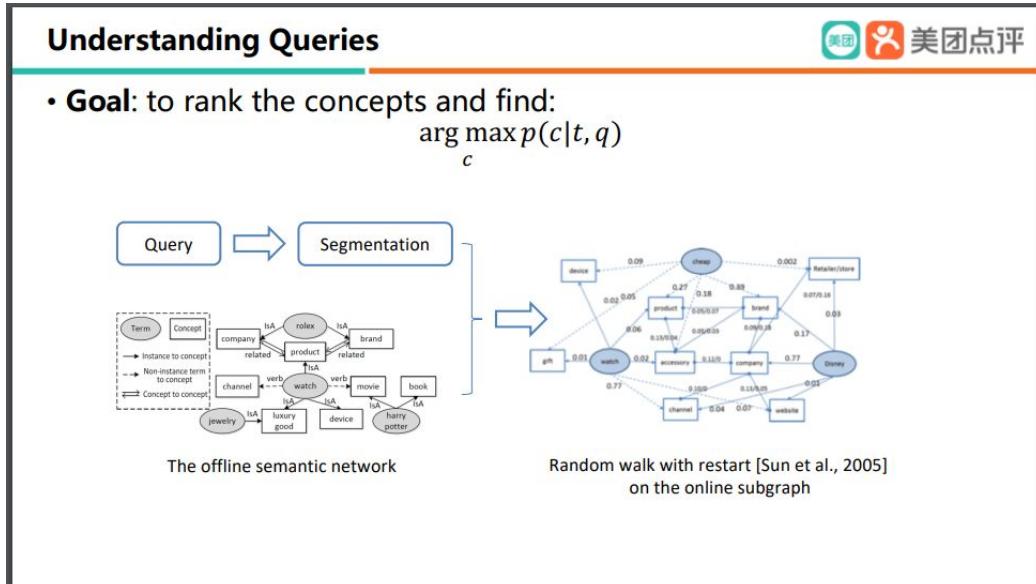


那么我们来看一看应该怎样去解决这样一件事情。当我们看到 Watch Harry Potter 时，我们首先要知道，Harry Potter 有可能是一本 Book，也有可能是一部 Movie。我们可以算出一个先验概率，这通常要通过大规模的统计。同时我们要知道，Watch 它有可能是一个名词，同时它也有可能是一个动词，并且我们还需要去挖掘，当 Watch 作为动词的时候，它和 Movie 有非常紧密的关联。

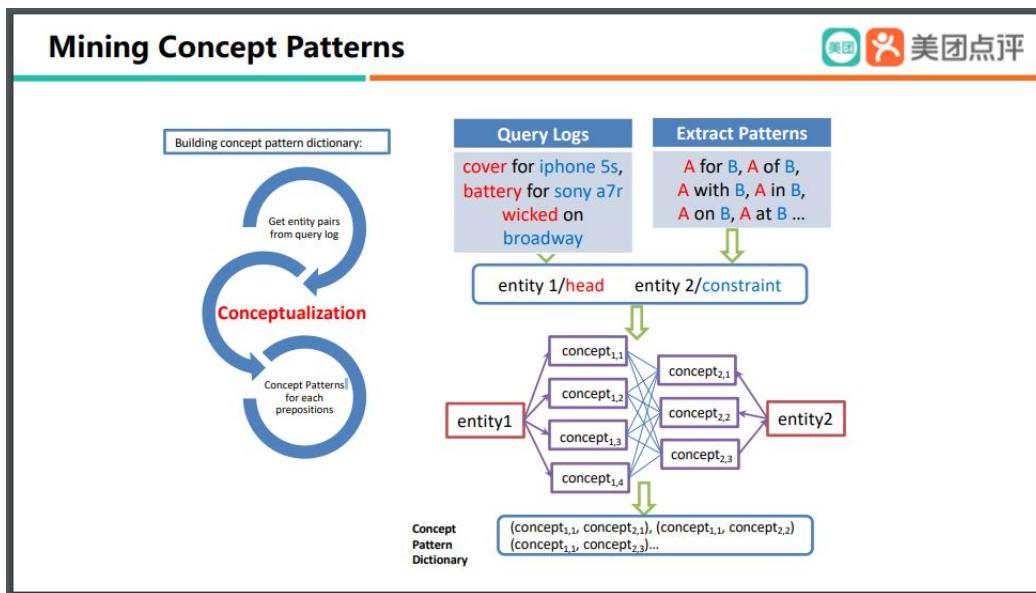
所以我们本质上是要去做一些概率上的推理，不仅要将条件概率做非常细粒度的分解，最后还要做概率计算。



通过概率计算的方法，我们实际上就可以构建出一个非常大的离线知识图谱，那么我们在这个上面，就可以有很多的 Term，以及它们所属的一些 Type，以及不同 Term 之间的一些关联。



当我们用这样一个非常大的离线知识图谱来做 Text Understanding 的时候，我们可以首先将这个 Text 进行分割处理，在分割之后，我们实际上是可以从这个非常大的离线知识图谱中截取出它的一个子图。最后我们使用了 Random Walk With Restart 的模型，来对这样一个在线的 Subgraph 进行分类。



我们再来看一下，如果一个文本里包含了 Multiple Entities，要怎样处理？我们需要做知识挖掘，怎么做？首先我们可以得到非常多的 Query Log，然后我们也可以去预定一些 Pattern，通过这种 Pattern 的定义，可以抽取出非常多 Entity 之间 Head 和 Modifier 这样的 Relation，那么在接下来我们可以将这些 Entity 映射到 Concept 上，之后得到一个 Pattern。

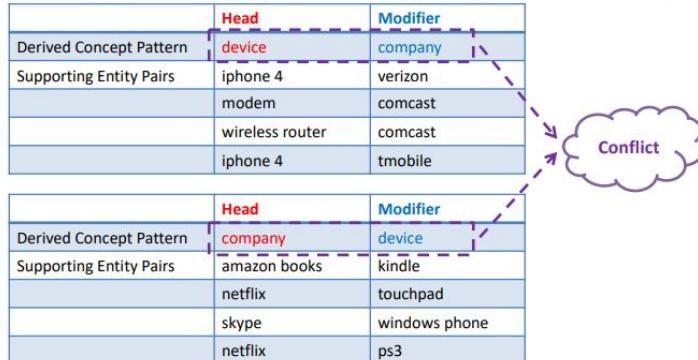
Why Concepts Can't Be Too General

美团 美团点评

- It may cause too many concept pattern conflicts: can't distinguish head and modifier for general concept pairs

	Head	Modifier
Derived Concept Pattern	device	company
Supporting Entity Pairs	iphone 4	verizon
	modem	comcast
	wireless router	comcast
	iphone 4	tmobile

	Head	Modifier
Derived Concept Pattern	company	device
Supporting Entity Pairs	amazon books	kindle
	netflix	touchpad
	skype	windows phone
	netflix	ps3



在这个过程之中，我们要将 Entity 映射到 Concept 上，那么这就是前面所提到的 Conceptualization。我们希望之后的映射不能太 General，避免 Concept Pattern 冲突。

Why Concepts Can't Be Too Specific

美团 美团点评

- It may generate concepts with less representation

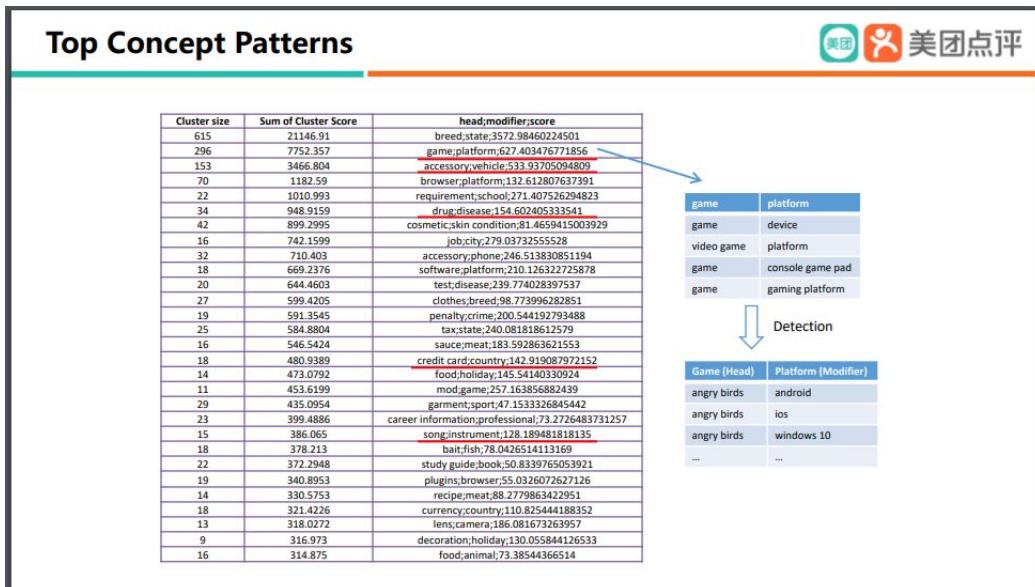
...	...
device	largest desktop OS vendor
device	largest software development company
device	largest global corporation
device	latest windows and office provider
...	...

- Concept level may regress to entity level
 - Large storage space: up to (million * million) patterns

We should use Basic-level Conceptualization (BLC)

但是它也不能太 Specific，因为如果太 Specific，可能就会缺少表达能力。最坏的情况，它有可能就会退化到 Entity Level，而 Entity 至少都是百万的规模，那么整个 Concept Patterns 就有可能变成百万乘以百万的级别，显然是不可用的。

所以我们就用到了前面介绍的 Basic-Level Conceptualization 的方法，将它映射到一个既不是特别 General，也不是特别 Specific 的 Concept 上。



大家看一下我们能够挖掘出来的一些 Top 的 Concept Patterns，比如说 Game 和 Platform，就是一个 Concept 和一个 Pattern。它有什么用？举一个具体的例子，当用户在搜 Angry Birds、iOS 的时候，我们就可以知道用户想找的是 Angry Birds 这款游戏，而 iOS 是用来限制这款游戏的一个 Platform。苹果公司每年都会推出新版本的 iOS，那么我们挖掘出这样的 Concept Pattern 之后，不管苹果出到 iOS 15或者 iOS 16，那么我们只需要将它们映射到 Platform，那么我们的 Concept Patterns 就仍然有效，这样可以很容易地进行知识扩展。

所以 Common Sense Knowledge Mining 以及 Conceptualization Modeling，可以用在很多的应用上，它可以用来算 Short Text Similarity，可以用来做 Classification、Clustering，也可以用来做广告的 Semantic Match、Q/A System、Chatbot 等等。

美团大脑——百科全书式知识图谱 (Encyclopedia Knowledge Graph)

在介绍完 Common Sense Knowledge Graph 之后，给大家介绍一下 Encyclopedia Knowledge Graph。这是美团的知识图谱项目——美团大脑。

美团大脑是什么？美团大脑是我们正在构建中的一个全球最大的餐饮娱乐知识图谱。我们希望能够充分地挖掘关联美团点评各个业务场景里的公开数据，比如说我们有累计 40 亿的用户评价，超过 10 万条个性化标签，遍布全球的 3000 多万商户以及超过 1.4 亿的店菜，我们还定义了 20 级细粒度的情感分析。

我们希望能够充分挖掘出这些元素之间的关联，构建出一个知识的“大脑”，用它来提供更加智能的生活服务。

知识提炼：美团大脑的构建

美团点评

- 大数据：**
 - 累计40亿的公开评价数据
 - 3,450万全球商家
 - 1.4亿店菜
 - 10万个性化标签
- NLP算法模型：**
 - 统计语言模型 (Language Model)
 - 主题生成模型 (Topic Model)
 - 深度学习模型 (Deep Learning Model)

我们简单地介绍一下美团大脑是如何进行构建的。我们会使用 Language Model (统计语言模型)、Topic Model (主题生成模型) 以及 Deep Learning Model (深度学习模型) 等各种模型，希望能够做到商家标签的挖掘，菜品标签的挖掘和情感分析的挖掘等等。

商户标签挖掘

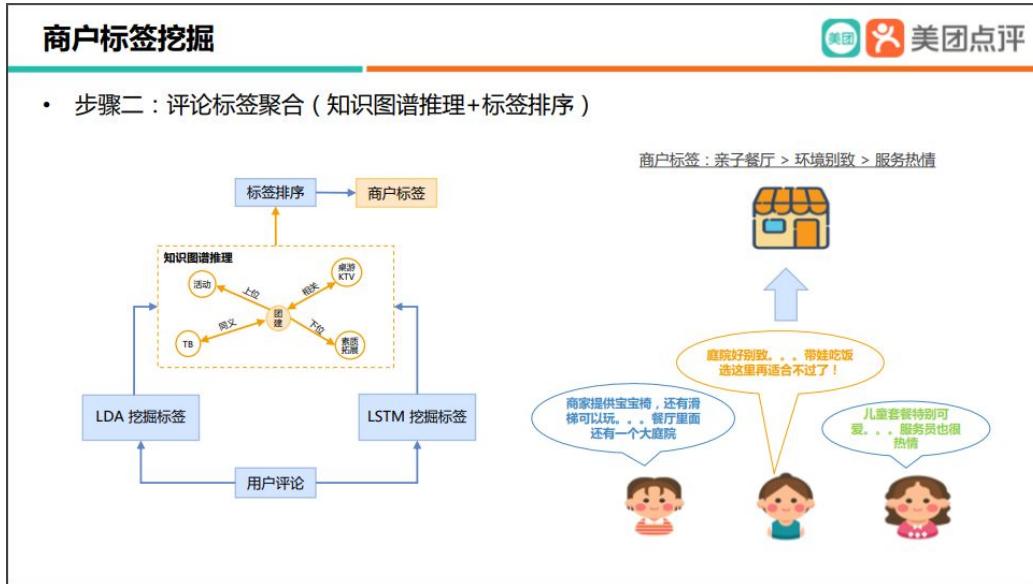
美团点评

- 步骤一：用户评论内容挖掘标签 (无监督模型+有监督深度学习模型)**
 - 无监督模型: LDA**
 - 成本低，无需标注数据
 - 准确性不可控，需要严格筛选分数高的标签
 - 有监督深度学习模型 : LSTM**
 - 成本高，需要大量标注数据
 - 准确性较高

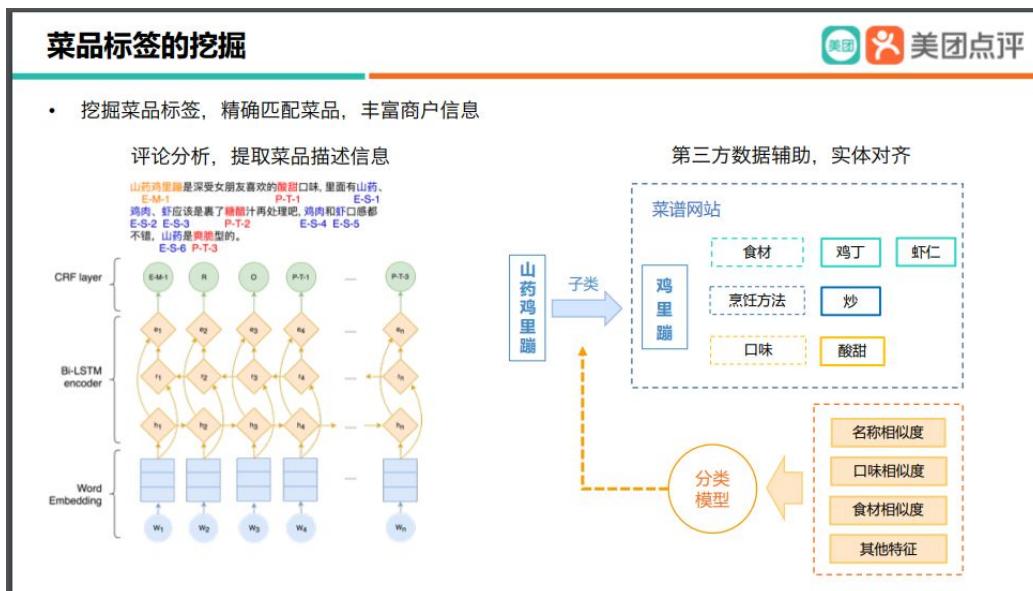
为了挖掘商户标签，首先我们要让机器去阅读评论。我们使用了无监督和有监督的深度学习模型。

无监督模型我们主要用了LDA，它的特点是成本比较低，无需标注的数据。当然，它准确性会比较不可控，同时对挖掘出来的标签我们还需要进行人工的筛选。至于有监督的深度学习模型，那么我们用了LSTM，它的特点是需要比较大量的标注数据。

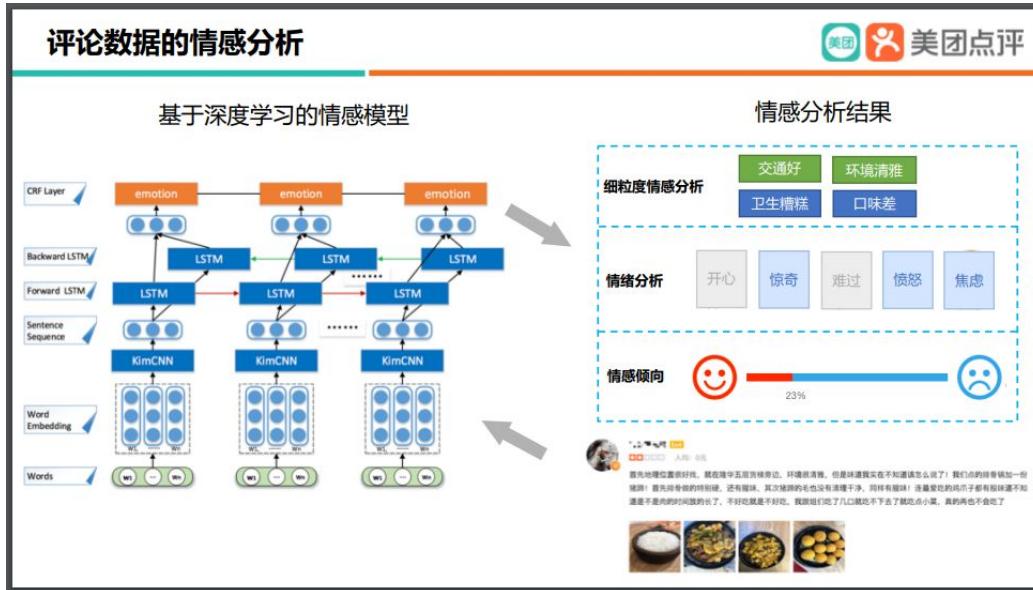
通过这两种模型挖掘出来的标签，我们会再加上知识图谱里面的一些推理，最终构建出商户的标签。



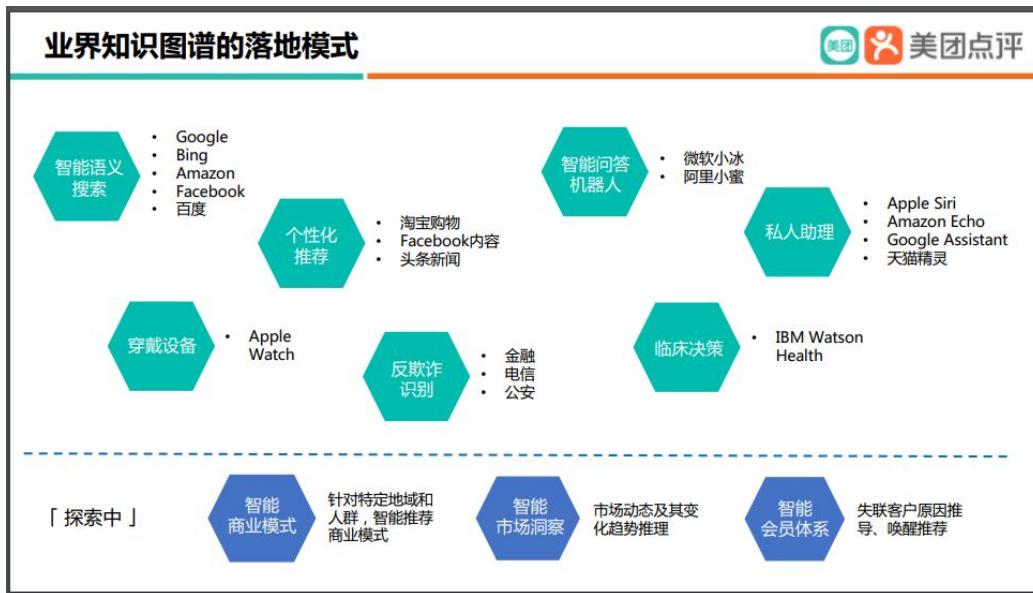
如果这个商户有很多的评价，都是围绕着宝宝椅、带娃吃饭、儿童套餐等话题，那么我们就可以得出很多关于这个商户的标签。比如说我们可以知道它是一个亲子餐厅，它的环境比较别致，服务也比较热情。



下面介绍一下我们如何对菜品进行标签的挖掘？我们使用了 Bi-LSTM 以及 CRF 模型。比如说从这个评论里面我们就可以抽取出这样的 Entity，再通过与其他的一些菜谱网站做一些关联，我们就可以得到它的食材、烹饪方法、口味等信息，这样我们就为每一个店菜挖掘出了非常丰富的口味标签、食材标签等各种各样的标签。



下面再简单介绍一下，我们如何进行评论数据的情感挖掘。我们用的是 CNN+LSTM 的模型，对于每一个用户的评价我们都能够分析出他的一些情感的倾向。同时我们也正在做细粒度的情感分析，我们希望能够通过用户短短的评价，分析出他在不同的维度，比如说交通、环境、卫生、菜品、口味等方面的不同的情感分析的结果。值得一提的是，这种细粒度的情感分析结果，目前在全世界范围内都没有很好的解决办法，但是美团大脑已经迈出了非常重要的一步。



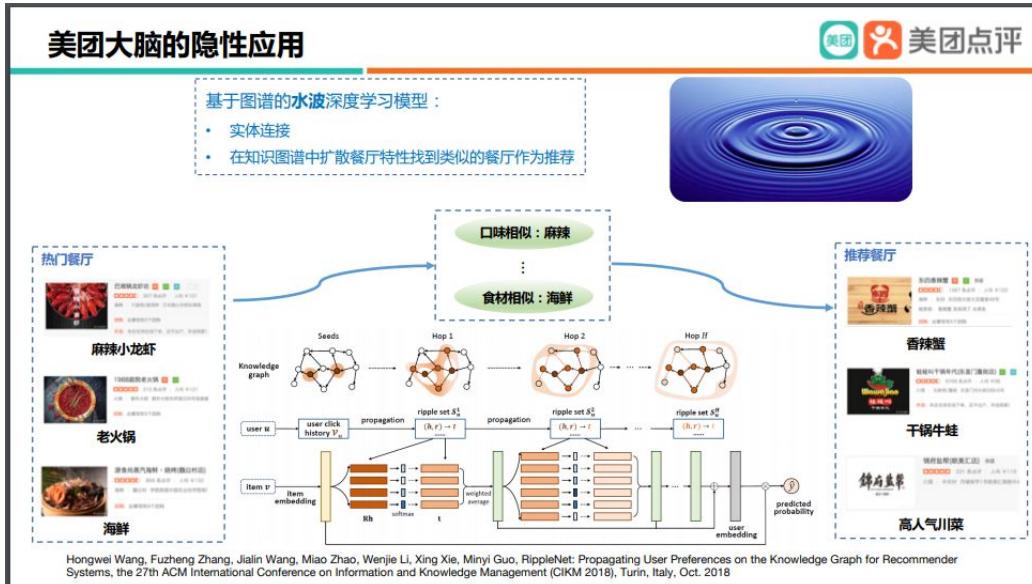
下面介绍一下我们的知识图谱是如何进行落地的。目前业界知识图谱已经有非常多的成熟应用，比如搜索、推荐、问答机器人、智能助理，包括在穿戴设备、反欺诈、临床决策上都有非常好的应用。同时业界也有很多的探索，包括智能商业模式、智能市场洞察、智能会员体系等等。



如何用知识图谱来改进我们的搜索？如果大家现在打开大众点评，搜索某一个菜品时，比如说麻辣小龙虾，其实我们的机器是已经帮大家提前阅读了所有的评价，然后分析出提供这道菜品的商家，我们还会根据用户评论的情感分析结果来改进这些搜索排序。

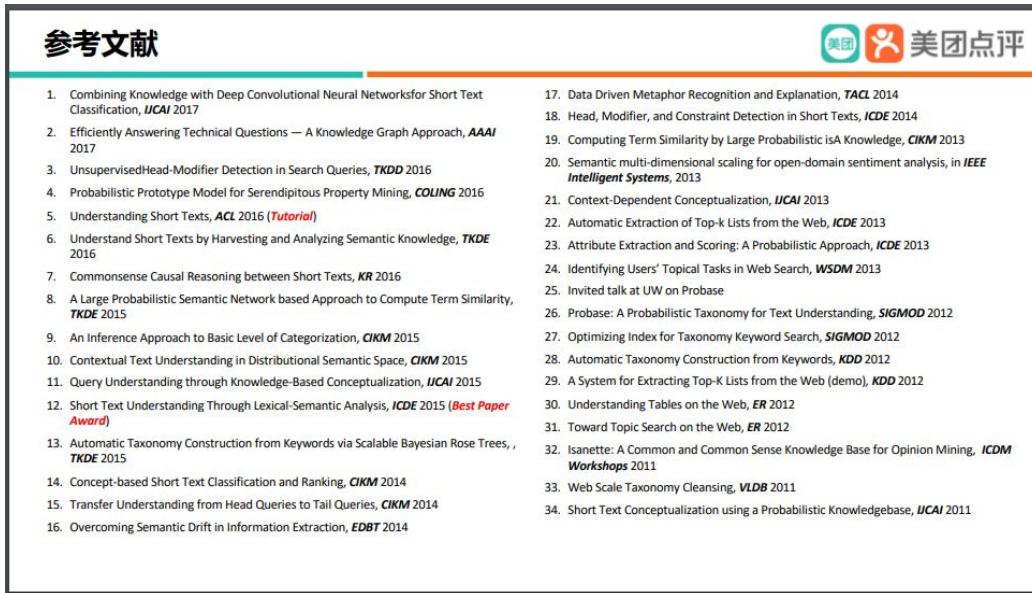


此外，我们也将它用在商圈的个性化推荐。当大家打开大众点评时，如果你现在位于某一个商场或者商圈，那么大家很快就能够看到这个商场或者商圈的页面入口。当用户进入这个商场和商户页面时，通过知识图谱，我们就能够提供“千人千面”的个性化排序和个性化推荐。



在这背后其实使用了一个“水波”的深度学习模型，关于这个深度学习模型更详细的介绍，大家可以参见我们在 CIKM 上的一篇论文。

所有的这一切，其实还有很多的技术突破等待我们去解决。比如整个美团大脑的知识图谱在百亿的量级，这也是世界上最大的餐饮娱乐知识图谱，为了支撑这个知识图谱，我们需要去研究千亿级别的图存储和计算引擎技术。我们也正在搭建一个超大规模的 GPU 集群，来支持海量数据的深度学习算法。未来，当所有的这些技术都成熟之后，我们还希望能够为所有用户提供“智慧餐厅”和“智能助理”的体验。



文章转载自 AI 科技大本营 (rgznai100)，部分内容有修正。

作者简介

- 仲远，博士，美团点评高级研究员、高级总监，美团 AI 平台部 NLP 中心负责人、大众点评搜索智能中心负责人。加入美团点评前，担任美国 Facebook 公司 Research Scientist，负责 Facebook 产品级 NLP Service。在 Facebook 之前，担任微软亚洲研究院的主管研究员，负责微软研究院知识图谱项目和对话机器人项目。多年来专注于自然语言处理、知识图谱及其在文本理解方面的研究，在国际顶级学术会议如 VLDB、ICDE、IJCAI、CIKM 等发表论文30余篇，获得 ICDE 2015 最佳论文奖，并是 ACL 2016 Tutorial “Understanding Short Texts”的主讲人，

出版学术专著3部，获得美国专利5项。在 NLP 和 KG 研究领域及实际产品系统中均有丰富经验，研究领域包括：自然语言处理、知识图谱、深度学习、数据挖掘等。

招聘信息

美团点评 NLP 团队招聘各类算法人才，Base 北京上海均可。NLP 中心使命是打造世界一流的自然语言处理核心技术和服务能力，依托 NLP（自然语言处理）、Deep Learning（深度学习）、Knowledge Graph（知识图谱）等技术，处理美团点评海量文本数据，打通餐饮、旅行、休闲娱乐等各个场景数据，构建美团点评知识图谱，搭建通用 NLP Service，为美团点评各项业务提供智能的文本语义理解服务。我们的团队既注重AI技术的落地，也开展中长期的NLP及知识图谱基础研究。目前项目及业务包括美团点评知识图谱、智能客服、语音语义搜索、文章评论语义理解、美团点评智能助理等。真正助力于“帮大家吃得更好，生活更好”企业使命的实现，优化用户的生活体验，改善和提升消费者的生活品质。欢迎各位朋友推荐或自荐至 hr.ai@meituan.com。

算法岗： [NLP算法工程师/专家/研究员](#)、[知识图谱算法工程师/专家/研究员](#)

工程岗： [C++/Java研发专家/工程师](#)、[AI平台研发工程师/专家](#)

产品岗： [AI产品经理/专家](#) (NLP、数据方向)

深度学习在美团搜索广告排序的应用实践

作者: 薛欢 姚强 玉林 王新

一、前言

在计算广告场景中，需要平衡和优化三个参与方——用户、广告主、平台的关键指标，而预估点击率CTR (Click-through Rate) 和转化率CVR (Conversion Rate) 是其中非常重要的一环，准确地预估CTR和CVR对于提高流量变现效率，提升广告主ROI (Return on Investment)，保证用户体验等都有重要的指导作用。

传统的CTR/CVR预估，典型的机器学习方法包括人工特征工程 + LR (Logistic Regression)¹²、GBDT (Gradient Boosting Decision Tree)²³ + LR、FM (Factorization Machine)³⁴和FFM (Field-aware Factorization Machine)⁴等模型。相比于传统机器学习方法，深度学习模型近几年在多领域多任务（图像识别、物体检测、翻译系统等）的突出表现，印证了神经网络的强大表达能力，以及端到端模型有效的特征构造能力。同时各种开源深度学习框架层出不穷，美团集团数据平台中心也迅速地搭建了GPU计算平台，提供GPU集群，支持TensorFlow、MXNet、Caffe等框架，提供数据预处理、模型训练、离线预测、模型部署等功能，为集团各部门的策略算法迭代提供了强有力的支持。

美团海量的用户与商家数据，广告复杂的场景下众多的影响因素，为深度学习方法的应用落地提供了丰富的场景。本文将结合广告特殊的业务场景，介绍美团搜索广告场景下深度学习的应用和探索。主要包括以下两大部分：

- CTR/CVR预估由机器学习向深度学习迁移的模型探索
- CTR/CVR预估基于深度学习模型的线下训练/线上预估的工程优化

二、从机器学习到深度学习的模型探索

2.1 场景与特征

美团搜索广告业务囊括了关键词搜索、频道筛选等业务，覆盖了美食、休闲、酒店、丽人、结婚、亲子等200多种应用场景，用户需求具有多样性。同时O2O模式下存在地理位置、时间等独特的限制。

结合上述场景，我们抽取了以下几大类特征：

- 用户特征
 - 人口属性：用户年龄，性别，职业等。
 - 行为特征：对商户/商圈/品类的偏好（实时、历史），外卖偏好，活跃度等。
 - 建模特征：基于用户的行为序列建模产生的特征等。
- 商户特征
 - 属性特征：品类，城市，商圈，品牌，价格，促销，星级，评论等。
 - 统计特征：不同维度/时间粒度的统计特征等。
 - 图像特征：类别，建模特征等。
 - 业务特征：酒店房型等。
- Query特征
 - 分词，意图，与商户相似度，业务特征等。
- 上下文特征
 - 时间，距离，地理位置，请求品类，竞争情况等。
 - 广告曝光位次。

结合美团多品类的业务特点及O2O模式独特的需求，着重介绍几个业务场景以及如何刻画：

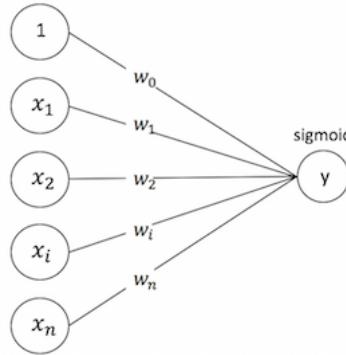
- 用户的消费场景
 - “附近”请求：美团和大众点评App中，大部分用户发起请求为“附近”请求，即寻找附近的美食、酒店、休闲娱乐场所等。因此给用户返回就近的商户可以起到事半功倍的效果。“**请求到商户的距离**”特征可以很好地刻画这一需求。
 - “**指定区域（商圈）**”请求：寻找指定区域的商户，这个区域的属性可作为该流量的信息表征。
 - “**位置**”请求：用户搜索词为某个位置，比如“五道口”，和指定区域类似，识别位置坐标，计算商户到该坐标的距离。
 - “**家/公司**”：用户部分的消费场所为“家”或“公司”，比如寻找“家”附近的美食，在“公司”附近点餐等，根据用户画像得到的用户“家”和“公司”的位置来识别这种场景。
- 多品类
 - 针对美食、酒店、休闲、丽人、结婚、亲子等众多品类的消费习惯以及服务方式，将数据拆分成三大部分，包括**美食**、**酒店**、**综合**（休闲、丽人、结婚、亲子等）。其中**美食**表达用户的餐饮需求，**酒店**表达用户的旅游及住宿需求，**综合**表达用户的其他生活需求。
- 用户的行为轨迹
 - 实验中发现用户的实时行为对表达用户需求起到很重要的作用。比如用户想找个餐馆聚餐，先筛选了美食，发现附近有火锅、韩餐、日料等店，大家对火锅比较感兴趣，又去搜索特定火锅等等。用户点击过的商户、品类、位置，以及行为序列等都对用户下一刻的决策起到很大作用。

2.2 模型

搜索广告CTR/CVR预估经历了从传统机器学习模型到深度学习模型的过渡。下面先简单介绍下传统机器学习模型（GBDT、LR、FM & FFM）及应用，然后再详细介绍在深度学习模型的迭代。

GBDT又叫MART (Multiple Additive Regression Tree)，是一种迭代的决策树算法。它由多棵决策树组成，所有树的结论累加起来作为最终答案。它能自动发现多种有区别的特征以及特征组合，并省去了复杂的特征预处理逻辑。Facebook实现GBDT + LR ⁵ 的方案，并取得了一定的成果。

LR



$$y(\mathbf{x}) = \text{sigmoid}(w_0 + \sum_{i=1}^n w_i x_i)$$

LR可以视作单层单节点的“DNN”，是一种宽而不深的结构，所有的特征直接作用在最后的输出结果上。模型优点是简单、可控性好，但是效果的好坏直接取决于特征工程的程度，需要非常精细的连续型、离散型、时间型等特征处理及特征组合。通常通过正则化等方式控制过拟合。

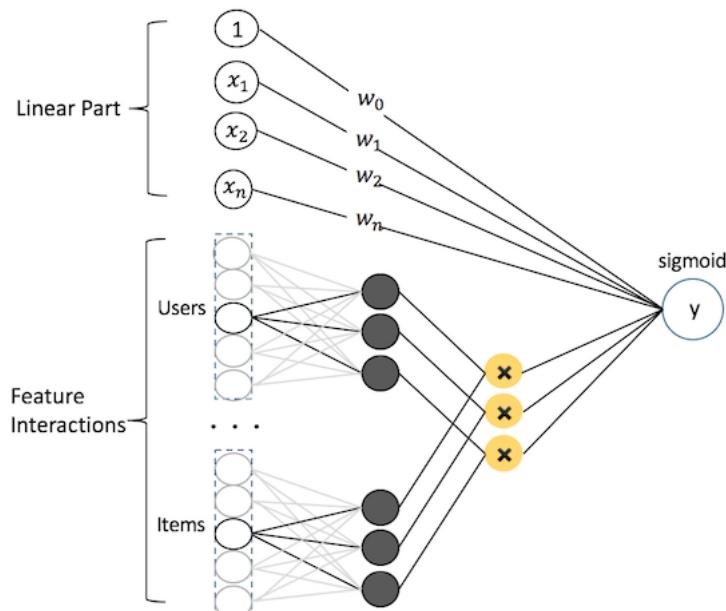
FM & FFM

FM可以看做带特征交叉的LR，如下图所示：

$$y_{FM}(\mathbf{x}) = \text{sigmoid}(w_0 + \sum_{i=1}^N w_i x_i + \sum_{i=1}^N \sum_{j=i+1}^N \langle v_i, v_j \rangle x_i x_j)$$

Linear Part Feature Interactions

从神经网络的角度考虑，可以看做下图的简单网络搭建方式：



模型覆盖了LR的宽模型结构，同时也引入了交叉特征，增加模型的非线性，提升模型容量，能捕捉更多的信息，对于广告CTR预估等复杂场景有更好的捕捉。

在使用DNN模型之前，搜索广告CTR预估使用了FFM模型，FFM模型中引入field概念，把n个特征归属到f个field里，得到nf个隐向量的二次项，拟合公式如下：

$$y(\mathbf{x}) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_{i,f_j}, \mathbf{v}_{j,f_i} \rangle x_i x_j$$

上式中， f_j 表示第j个特征所属的field。设定隐向量长度为k，那么相比于FM的nk个二次项参数，FFM有nkf个二次项参数，学习和表达能力也更强。

例如，在搜索广告场景中，假设将特征划分到8个Field，分别是用户、广告、Query、上下文、用户-广告、上下文-广告、用户-上下文及其他，相对于FM能更好地捕捉每个Field的信息以及交叉信息，每个特征构建的向量长度 $8 \times k$ ，整个模型参数空间为 $8k + n + n + 1$ 。

Yu-Chin Juan实现了一个C++版的[FFM模型工具包](#)，但是该工具包只能在单机训练，难以支持大规模的训练数据及特征集合；并且它省略了常数项和一次项，只包含了特征交叉项，对于某些特征的优化需求难以满足，因此我们开发了[基于PS-Lite的分布式FFM训练工具](#)（支持亿级别样本，千万级别特征，分钟级完成训练，已经在公司内部普遍使用），主要添加了以下新的特性：

- 支持FFM模型的分布式训练。
- 支持一次项和常数项参数学习，支持部分特征只学习一次项参数（不需要和其他特征做交叉运算），例如广告位次特征等。拟合公式如下：

$$y(\mathbf{x}) = w_0 + \sum_{i=1}^n w_i x_i + \frac{1}{2} \sum_{i \in \text{group}} \sum_{j \in \text{group and } j \neq i} \langle \mathbf{v}_{i,f_j}, \mathbf{v}_{j,f_i} \rangle x_i x_j$$

- 支持多种优化算法。

从GBDT模型切到FFM模型，积累的效果如下所示，主要的提升来源于对大规模离散特征的刻画及使用更充分的训练数据：

模型	AUC	CTR
GBDT	baseline	baseline
FFM	+10%+	+10%+

DNN

从上面的介绍大家可以看到，美团场景具有多样性和很高的复杂度，而实验表明从线性的LR到具备非线性交叉的FM，到具备Field信息交叉的FFM，模型复杂度（模型容量）的提升，带来的都是结果的提升。而LR和FM/FFM可以视作简单的浅层神经网络模型，基于下面一些考虑，我们在搜索广告的场景下把CTR模型切换到深度学习神经网络模型：

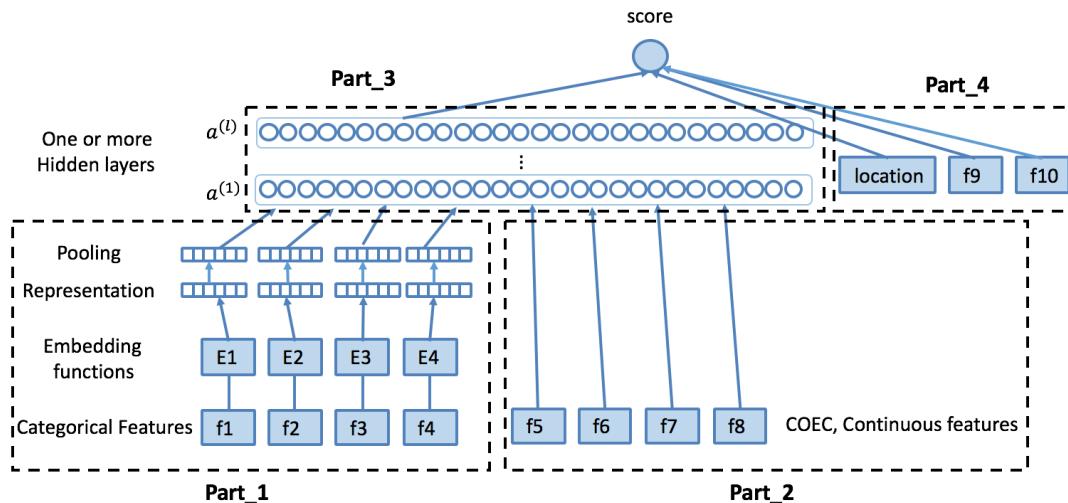
- 通过改进模型结构，加入深度结构，利用端到端的结构挖掘高阶非线性特征，以及浅层模型无法捕捉的潜在模式。
- 对于某些ID类特别稀疏的特征，可以在模型中学习到保持分布关系的稠密表达（embedding）。
- 充分利用图片和文本等在简单模型中不好利用的信息。

我们主要尝试了以下网络结构和超参调优的实验。

Wide & Deep

首先尝试的是Google提出的经典模型Wide & Deep Model [\(1\)](#)，模型包含Wide和Deep两个部分，其中Wide部分可以很好地学习样本中的高频部分，在LR中使用到的特征可以直接在这个部分使用，但对于没有见过的ID类特征，模型学习能力较差，同时合理的人工特征工程对于这个部分的表达有帮助。Deep部分可以补充学习样本中的长尾部分，同时提高模型的泛化能力。Wide和Deep部分在这个端到端的模型里会联合训练。

在完成[场景与特征](#)部分介绍的特征工程后，我们基于Wide & Deep模型进行结构调整，搭建了以下网络：



在搜索广告的场景中，上图的Part_1包含离散型特征及部分连续型特征离散化后的结果（例如用户ID、广告ID、商圈ID、品类ID、GEO、各种统计类特征离散化结果等等）。离散化方式主要采用等频划分或MDLP [\(2\)](#)。每个域构建自己的embedding向量（缺失特征和按照一定阈值过滤后的低频特征在这里统一视作Rare特征），得到特征的Representation，然后通过Pooling层做采样，并拼接在一起进行信息融合。

右侧的Part_2部分主要包含我们场景下的统计类特征及部分其他途径建模表示后输入的特征（例如图片特征、文本特征等），和Part_1的最后一层拼接在一起做信息融合。

$$a^{(1)} = concat(concat(Pooling(Emb_i(feature_i))), feature_{continuous})$$

Part_3为多个全连接层，每个Layer后面连接激活函数，例如ReLU, Tanh等。

$$a^{(l+1)} = f(W^{(l)} a^{(l)} + b^{(l)})$$

右上的Part_4部分主要包含广告曝光位次 (Position Bias) 及部分离散特征，主要为了提高模型的记忆性，具有更强的刻画能力。Wide和Deep部分结合，得到最终的模型：

$$y(\mathbf{x}) = \sigma(W_{wide}^T x_{wide} + W_{deep}^T a^{(l)} + b)$$

深度学习模型在图像语音等数据上有显著作用的原因之一是，我们在这类数据上不太方便产出能很好刻画场景的特征，人工特征+传统机器学习模型并不能学习出来全面合理的数据分布表示，而深度学习end-to-end的方式，直接结合Label去学习如何从原始数据抽取合适的表达 (representation)。但是在美团等电商的业务场景下，输入的数据形态非常丰富，有很多业务数据有明确的物理含义，因此一部分人工特征工程也是必要的，提前对信息做一个合理的抽取表示，再通过神经网络学习进行更好的信息融合和表达。

在美团搜索广告的场景下，用户的实时行为有非常强的指代性，但是以原始形态直接送入神经网络，会损失掉很多信息，因此我们对它进行了不同方式描述和表示，再送入神经网络之中进行信息融合和学习。另一类很有用的信息是图像信息，这部分信息的一种处理方式是，可以通过end-to-end的方式，用卷积神经网络和DNN进行拼接做信息融合，但是可能会有网络的复杂度过高，以及训练的收敛速度等问题，也可以选择用CNN预先抽取特征，再进行信息融合。

下面以这两类数据特征为例，介绍在Wide & Deep模型中的使用方式。

- 用户实时行为

- 行为实体 用户的实时行为包括点击商户 (C_P)、下单商户 (O_P)、搜索 (Q)、筛选品类 (S) 等。商户的上层属性包括品类 (Type: C_Type, O_Type)、位置 (Loc: C_Loc, O_Loc) 等。
- Item Embedding 对用户的行为实体构建embedding向量，然后进行Sum/Average/Weighted Pooling，和其他特征拼接在一起。实验发现，上层属性实体 (C_Type, O_Type, C_Loc, O_Loc) 的表现很正向，离线效果有了很明显的提升。但是C_P, O_P, Q, S这些实体因为过于稀疏，导致模型过拟合严重，离线效果变差。因此，我们做了两方面的改进：
 1. 使用更充分的数据，单独对用户行为序列建模。例如LSTM模型，基于用户当前的行为序列，来预测用户下一时刻的行为，从中得到当前时刻的“Memory信息”，作为对用户的embedding表示；或Word2Vec模型，生成行为实体的embedding表示，Doc2Vec模型，得到用户的embedding表示。实验发现，将用户的embedding表示加入到模型Part_2部分，特征覆盖率增加，离线效果有了明显提升，而且由于模型参数空间增加很小，模型训练的时间基本不变。
 2. 使用以上方法产生的行为实体embedding作为模型参数初始值，并在模型训练过程中进行fine tuning。同时为了解决过拟合问题，对不同域的特征设置不同的阈值过滤。
- 计数特征 即对不同行为实体发生的频次，它是对行为实体更上一层的抽象。
- Pattern特征 用户最近期的几个行为实体序列（例如A-B-C）作为Pattern特征，它表示了行为实体之间的顺序关系，也更细粒度地描述了用户的行为轨迹。

- 图片

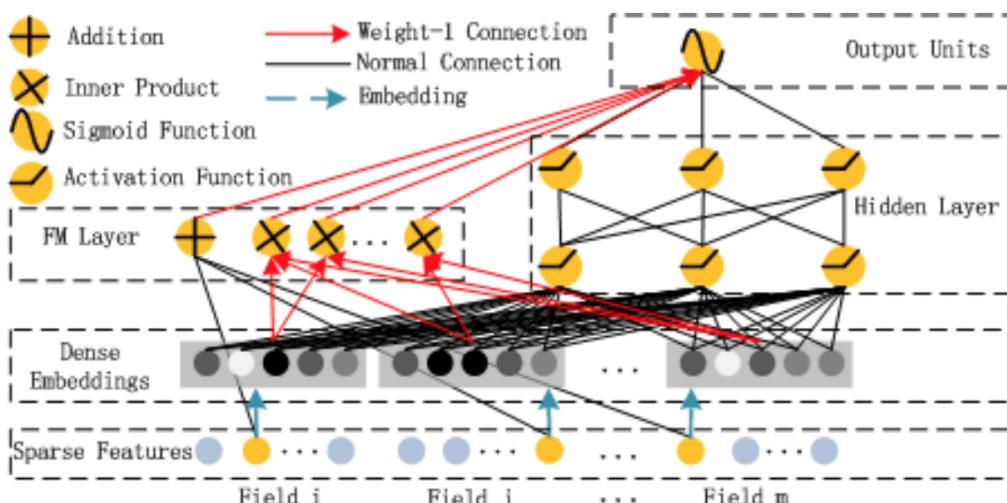
- 描述 商户的头图在App商品展示中占据着很重要的位置，而图片也非常吸引用户的注意力。
- 图片分类特征 使用VGG16、Inception V4等训练图片分类模型，提取图片特征，然后加入到CTR模型中。
- E2E model 将Wide & Deep模型和图片分类模型结合起来，训练端到端的网络。

从FFM模型切到Wide & Deep模型，积累到目前的效果如下所示，主要的提升来源于模型的非线性表达及对更多特征的更充分刻画。

模型	AUC	CTR
FFM	baseline	baseline
Wide & Deep	+2.4%	+4%

DeepFM

华为诺亚方舟团队结合FM相比LR的特征交叉的功能，将Wide & Deep部分的LR部分替换成FM来避免人工特征工程，于是有了DeepFM^②，网络结构如下图所示。



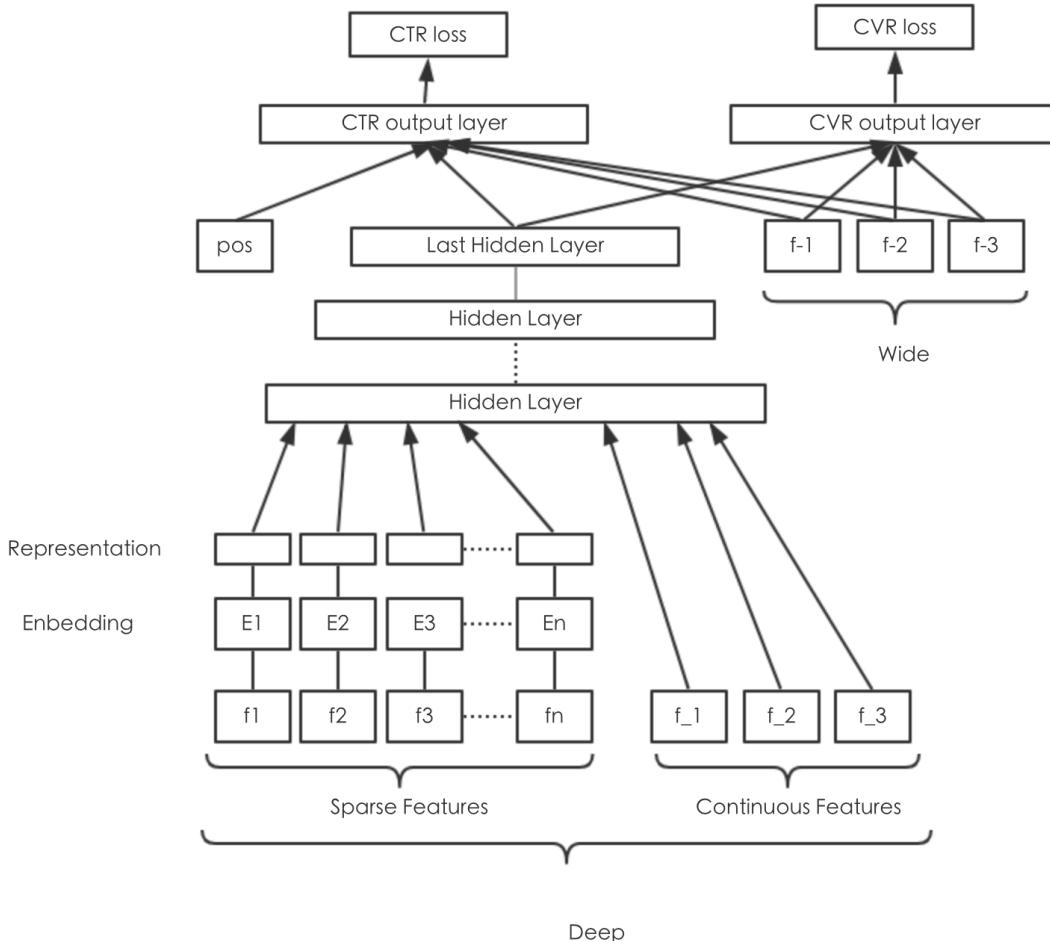
比起Wide & Deep的LR部分，DeepFM采用FM作为Wide部分的输出，在训练过程中共享了对不同Field特征的embedding信息。

我们在部分业务上尝试了DeepFM模型，并进行了超参的重新调优，取得了一定的效果。其他业务也在尝试中。具体效果如下：

模型	AUC	CTR
Wide & Deep	baseline	baseline
DeepFM	+0.3%	+1.4%

Multi-Task

广告预估场景中存在多个训练任务，比如CTR、CVR、交易额等。既考虑到多个任务之间的联系，又考虑到任务之间的差别，我们利用Multi-Task Learning的思想，同时预估点击率、下单率，模型结构如下图所示：



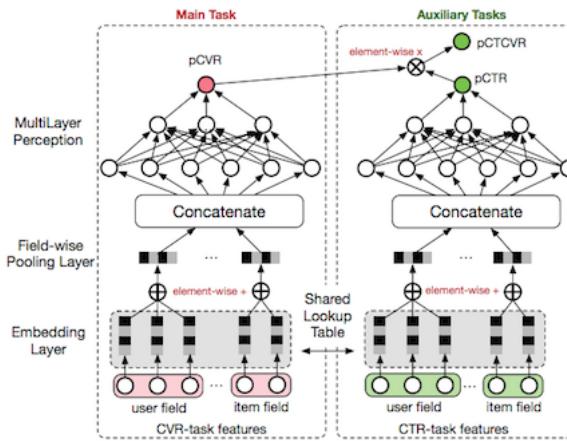
- 由于CTR、CVR两个任务非常类似，所以采用“Hard Parameter Sharing”的结构，完全共享网络层的参数，只在输出层区分不同的任务。
- 由于下单行为受展现位次的影响非常小，所以下单率的输出层不考虑位次偏差的因素。
- 输出层在不同任务上单独增加所需特征。
- 离线训练和线上预估流程减半，性能提升；效果上相对于单模型，效果基本持平：

模型	AUC	CTR	时间
Wide & Deep	baseline	baseline	baseline
Multi-Task	基本持平	基本持平	50%

近期，阿里发表论文“Entire Space Multi-Task Model”[\[1\]](#)，提出目前CVR预估主要存在Sample Selection Bias (SSB) 和Data Sparsity (DS) 两个问题，并提出在全局空间建模（以pCTCVR和pCTR来优化CVR）和特征Transform的方法来解决。具体的Loss Function是：

$$\begin{aligned}
 L(\theta_{ctr}, \theta_{cvr}) = & \sum_{i=1}^N l(y_i, f(\mathbf{x}_i; \theta_{ctr})) \\
 & + \sum_{i=1}^N l(y_i \& z_i, f(\mathbf{x}_i; \theta_{ctr}) \times f(\mathbf{x}_i; \theta_{cvr}))
 \end{aligned}$$

网络结构是：



超参调优

除了以上对网络结构的尝试，我们也进行了多组超参的调优。神经网络最常用的超参设置有：隐层数及节点数、学习率、正则化、Dropout Ratio、优化器、激活函数、Batch Normalization、Batch Size等。不同的参数对神经网络的影响不同，神经网络常见的一些问题也可以通过超参的设置来解决：

- 过拟合
 - 网络宽度深度适当调小，正则化参数适当调大，Dropout Ratio适当调大等。
- 欠拟合
 - 网络宽度深度适当调大，正则化参数调小，学习率减小等。
- 梯度消失/爆炸问题
 - 合适的激活函数，添加Batch Normalization，网络宽度深度变小等。
- 局部最优解
 - 调大Learning Rate，合适的优化器，减小Batch Size等。
- Covariate Shift
 - 增加Batch Normalization，网络宽度深度变小等。

影响神经网络的超参数非常多，神经网络调参也是一件非常重要的事情。工业界比较实用的调参方法包括：

- **网格搜索/Grid Search**：这是在机器学习模型调参时最常用到的方法，对每个超参数都敲定几个要尝试的候选值，形成一个网格，把所有超参数网格中的组合遍历一下尝试效果。简单暴力，如果能全部遍历的话，结果比较可靠。但是时间开销比较大，神经网络的场景下一般尝试不了太多的参数组合。
- **随机搜索/Random Search**：Bengio在“Random Search for Hyper-Parameter Optimization”¹⁰中指出，Random Search比Grid Search更有效。实际操作的时候，可以先用Grid Search的方法，得到所有候选参数，然后每次从中随机选择进行训练。这种方式的优点是因为采样，时间开销变小，但另一方面，也有可能会错过较优的超参数组合。
- **分阶段调参**：先进行初步范围搜索，然后根据好结果出现的地方，再缩小范围进行更精细的搜索。或者根据经验值固定住其他的超参数，有针对性地实验其中一个超参数，逐次迭代直至完成所有超参数的选择。这个方式的优点是在优先尝试次数中，拿到效果较好的结果。

我们在实际调参过程中，使用的是第3种方式，在根据经验值初始化超参数之后，按照隐层大小→学习率→Batch Size→Drop out/L1/L2的顺序进行参数调优。

在搜索广告数据集上，不同超参的实验结果如下：

隐层大小	AUC	Learning Rate	AUC	Batch Size	AUC
128,128	0%	0.0005	0%	500	0%
512,256	+0.24%	0.001	-0.06%	1000	+0.46%
512,256,128	+0.34%	0.005	-0.93%	3000	+0.78%
512,256,256,256,128	+0.37%	0.01	-16.10%	5000	+0.92%
				10000	+0.96%
Drop Out (keep_prob)	AUC	L1	AUC	L2	AUC
0.8	0%	0.001	0%	0.001	0%
0.9	+0.13%	0.0001	+0.01%	0.0001	-0.005%
1	+0.48%	0	+0.01%	0	+0.01%

2.3 小结

搜索广告排序模型经历了从GBDT → FFM → DNN的迭代，同时构建了更加完善的特征体系，线下AUC累积提升13%+，线上CTR累积提升15%+。

三、基于深度学习模型的工程优化

3.1 线下训练

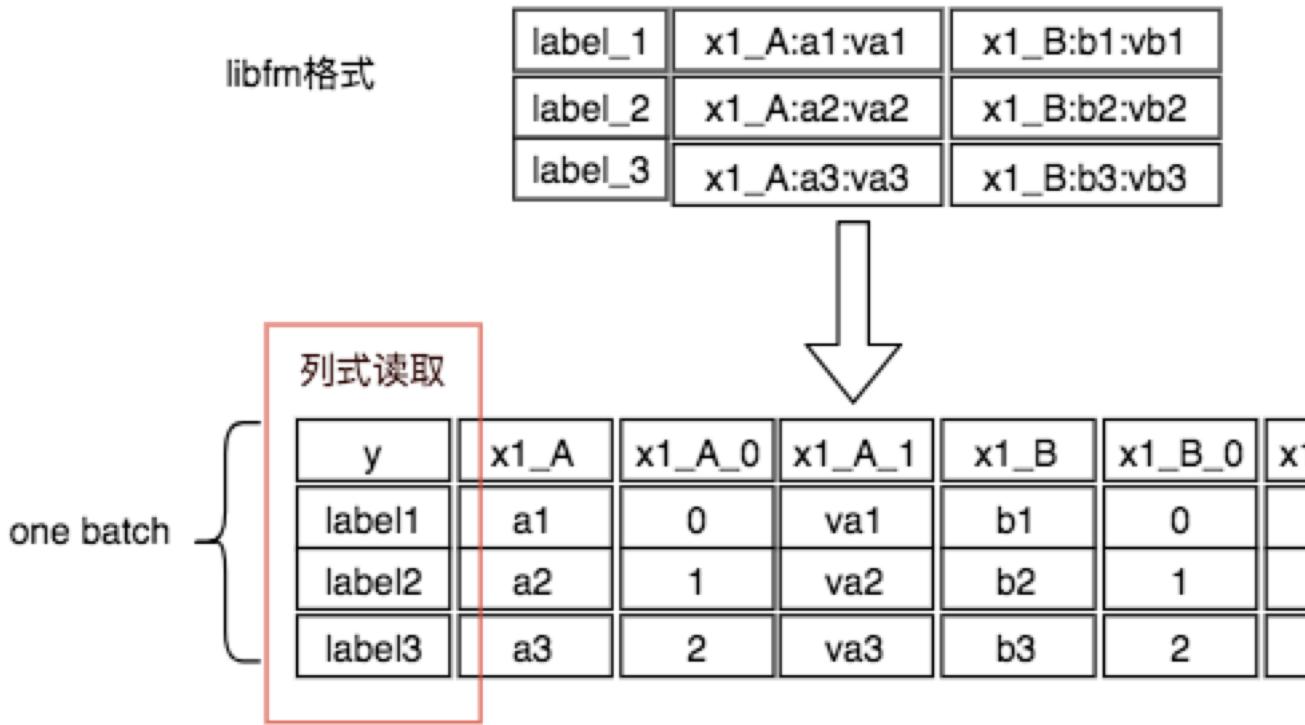
TensorFlow程序如果单机运行中出现性能问题，一般会有以下几种问题：

1. 复杂的预处理逻辑耦合在训练过程中。
2. 选择正确的IO方式。

剥离预处理流程

在模型的试验阶段，为了快速试验，数据预处理逻辑与模型训练部分都耦合在一起，而数据预处理包含大量IO类型操作，所以很适合用HadoopMR或者Spark处理。具体流程如下：

1. 在预处理阶段将查表、join字典等操作都做完，并且将查询结果与原始数据merge在一起。
2. 将libfm格式的数据转为易于TensorFlow操作的SparseTensor方式：



3. 将原始数据转换为TensorFlow Record。

选择正确的IO方式

TensorFlow读取数据的方式主要有2种，一般选择错误会造成性能问题，两种方式为：

1. **Feed_dict** 通过feed_dict将数据喂给session.run函数，这种方式的好处是思路很清晰，易于理解。缺点是性能差，性能差的原因是feed给session的数据需要在session.run之前准备好，如果之前这个数据没有进入内存，那么就需要等待数据进入内存，而在实际场景中，这不仅仅是等待数据从磁盘或者网络进入内存的事情，还可能包括很多前期预处理的工作也在这里做，所以相当于一个串行过程。而数据进入内存后，还要串行的调用PyArrayToTF_Tensor，将其copy成tensorflow的tensorValue。此时，GPU显存处于等待状态，同时，由于tf的Graph中的input为空，所以CPU也处于等待状态，无法运算。
2. **RecordReader** 这种方式是tf在Graph中将读取数据这个操作看做图中一个operation节点，减少了一个copy的过程。同时，在tf中还有batch与threads的概念，可以异步的读取数据，保证在GPU或者CPU进行计算的时候，读取数据这个操作也可以多线程异步执行。静态图中各个节点间的阻塞：在一个复杂的DAG计算图中，如果有任何一个点计算比较慢时，会造成阻塞，下游节点不得不等待。此时，首先要考虑的问题是图中节点参数所存储的位置是否正确。比如如果某个计算节点是在GPU上运算，那么如果这个节点所有依赖的variable对象声明在CPU上，那么就要做一次memcpy，将其从内存中copy到GPU上。因为GPU计算的很快，所以大部分时间花在拷贝上了。总之，如果网络模型比较简单，那么这种操作就会非常致命；如果网络结构复杂，比如网络层次非常深，那么这个问题倒不是太大的问题了。

在这个Case中，因为需要提升吞吐，而不仅仅是在试验阶段。所以需要用RecordReader方式处理数据。

优化过程

- 将整体程序中的预处理部分从代码中去除，直接用Map–Reduce批处理去做（因为批处理可以将数据分散去做，所以性能非常好，2亿的数据分散到4900多个map中，大概处理了15分钟左右）。
- MR输出为TensorFlow Record格式，避免使用Feed_dict。
- 数据预读，也就是用多进程的方式，将HDFS上预处理好的数据拉取到本地磁盘（使用joblib库+shell将HDFS数据用多进程的方式拉取到本地，基本可以打满节点带宽2.4GB/s，所以，拉取数据也可以在10分钟内完成）。
- 程序通过TensorFlow提供的TFRecordReader的方式读取本地磁盘上的数据，这部分的性能提升是最为明显的。原有的程序处理数据的性能大概是1000条/秒，而通过TFRecordReader读取数据并且处理，性能大概是18000条/秒，性能大概提升了18倍。
- 由于每次run的时候计算都要等待TFRecordReader读出数据，而没用利用batch的方式。如果用多线程batch可以在计算期间异步读取数据。在TensorFlow所有例子中都是使用TFRecordReader的read接口去读取数据，再用batch将数据多线程抓过来。但是，其实这样做加速很慢。需要使用TFRecordReader的read_up_to的方法配合batch的enqueue_many=True的参数，才可以做到最大的加速比。使用tf.train.batch的API后，性能提升了38倍。

此时，性能已经基本达到我们的预期了。例如整体数据量是2亿，按照以前的性能计算1000条/秒，大概需要运行55个小时。而现在大概需要运行87分钟，再加上预处理（15分钟）与预拉取数据（10分钟）的时间，在不增加任何计算资源的情况下大概需要2个小时以内。而如果是并行处理，则可以在分钟级完成训练。

3.2 线上预估

线上流量是模型效果的试金石。离线训练好的模型只有参与到线上真实流量预估，才能发挥其价值。在演化的过程中，我们开发了一套稳定可靠的线上预估体系，提高了模型迭代的效率。

模型同步

我们开发了一个高可用的同步组件：用户只需要提供线下训练好的模型的HDFS路径，该组件会自动同步到线上服务机器上。该组件基于HTTPFS实现，它是美团离线计算组提供的HDFS的HTTP方式访问接口。同步过程如下：

- 同步前，检查模型md5文件，只有该文件更新了，才需要同步。
- 同步时，随机链接HTTPFS机器并限制下载速度。
- 同步后，校验模型文件md5值并备份旧模型。

同步过程中，如果发生错误或者超时，都会触发报警并重试。依赖这一组件，我们实现了在2min内可靠的将模型文件同步到线上。

模型计算

当前我们线上有两套并行的预估计算服务。

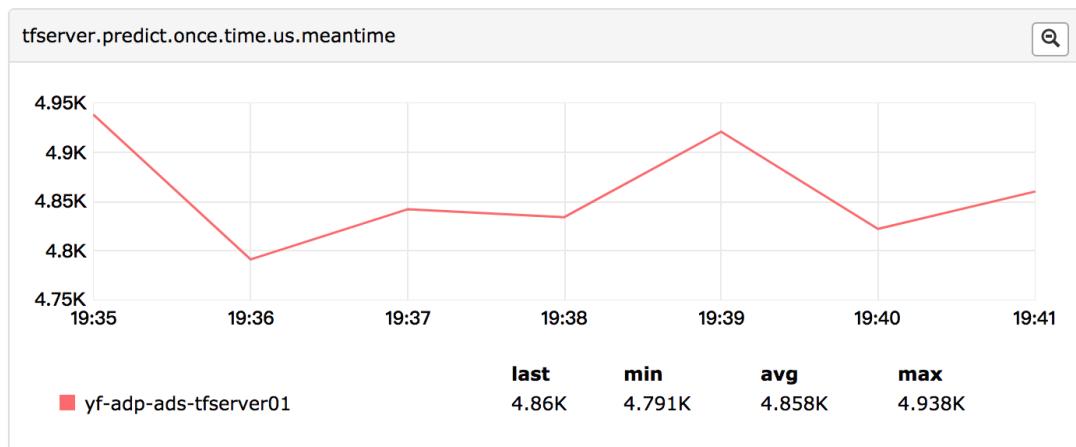
基于TF Serving的模型服务

TF Serving是TensorFlow官方提供的一套用于在线实时预估的框架。它的突出优点是：和TensorFlow无缝链接，具有很好的扩展性。使用TF serving可以快速支持RNN、LSTM、GAN等多种网络结构，而不需要额外开发代码。这非常有利于我们模型快速实验和迭代。

使用这种方式，线上服务需要将特征发送给TF Serving，这不可避免引入了网络IO，给带宽和预估时延带来压力。我们尝试了以下优化，效果显著。

- 并发请求。一个请求会召回很多符合条件的广告。在客户端多个广告并发请求TF Serving，可以有效降低整体预估时延。
- 特征ID化。通过将字符串类型的特征名哈希到64位整型空间，可以有效减少传输的数据量，降低使用的带宽。

TF Serving服务端的性能差强人意。在典型的五层网络（512*256*256*256*128）下，单个广告的预估时延约4800μs，具体见下图：



定制的模型计算实现

由于广告线上服务需要极高的性能，对于主流深度学习模型，我们也定制开发了具体计算实现。这种方式可以针对性的优化，并避免TF Serving不必要的特征转换和线程同步，从而提高服务性能。

例如全连接DNN模型中使用Relu作为激活函数时，我们可以使用滚动数组、剪枝、寄存器和CPU Cache等优化技巧，具体如下：

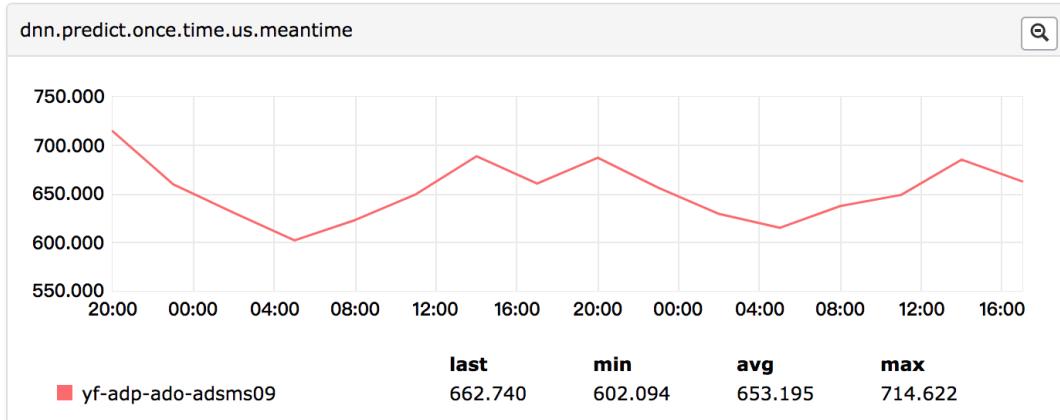
```
// 滚动数组
int nextLayerIndex = currentLayerIndex ^ 1 ;
System.arraycopy(bias, bOff, data[nextLayerIndex], 0, nextLayerSize);
```

```

for (int i = 0; i < currentLayerSize; i++) {
    float value = data[currentLayerIndex][i];
    // 剪枝
    if (value > 0.0) {
        // 寄存器
        int index = wOff + i * nextLayerSize;
        // CPU 缓存友好
        for (int j = 0; j < nextLayerSize; j++) {
            data[nextLayerIndex][j] += value * weights[index + j];
        }
    }
}
for (int i = 0; i < nextLayerSize; k++) {
    data[nextArrayIndex][i] = ReLu(data[nextArrayIndex][i]);
}
arrayIndex = nextArrayIndex;

```

优化后的单个广告预估时延约650μs，见下图：



综上，当前线上预估采取“两条腿走路”的策略。利用TF Serving快速实验新的模型结构，以保证迭代效率；一旦模型成熟切换主流量，我们会开发定制实现，以保证线上性能。

模型效果

借助于我们的分层实验平台，我们可以方便的分配流量，完成模型的小流量实验上线。该分层实验平台同时提供了分钟粒度的小流量实时效果数据，便于模型评估和效果监控。

四、总结与展望

经过一段时间的摸索与实践，搜索广告业务在深度学习模型排序上有了一定的成果与积累。接下来，我们将继续在特征、模型、工程角度迭代优化。特征上，更深度挖掘用户意图，刻画上下文场景，并结合DNN模型强大的表达能力充分发挥特征的作用。模型上，探索新的网络结构，并结合CNN、RNN、Attention机制等发挥深度学习模型的优势。持续跟进业界动态，并结合实际场景，应用到业务中。工程上，跟进TensorFlow的新特性，并对目前实际应用中遇到的问题针对性优化，以达到性能与效果的提升。我们在持续探索中。

作者简介

- 薛欢，2016年3月加入美团，主要从事搜索广告排序模型相关的工作。
- 姚强，2016年4月加入美团，主要从事搜索广告召回、机制与排序等相关算法研究应用工作。
- 玉林，2015年5月加入美团，主要从事搜索广告排序相关的工程优化工作。
- 王新，2017年4月加入美团，主要从事GPU集群管理与深度学习工程优化的工作。

招聘

美团广告平台全面负责美团到店餐饮、到店综合（结婚、丽人、休闲娱乐、学习培训、亲子、家装）、酒店旅游的商业变现。搜索广告基于数亿用户、数百万商家和数千万订单的真实数据做挖掘，在变现的同时确保用户体验和商家利益。欢迎有意向的同学加入搜索广告算法组。简历请投递至：leijun#meituan.com

参考文献

- Chapelle, O., Manavoglu, E., & Rosales, R. (2015). Simple and scalable response prediction for display advertising. ACM Transactions on Intelligent Systems and Technology (TIST), 5(4), 61. [\[return\]](#)
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189–1232. [\[return\]](#)
- Rendle, S. (2010, December). Factorization machines. In Data Mining (ICDM), 2010 IEEE 10th International Conference on (pp. 995–1000). IEEE. [\[return\]](#)
- Juan, Y., Zhuang, Y., Chin, W. S., & Lin, C. J. (2016, September). Field-aware factorization machines for CTR prediction. In Proceedings of the 10th ACM Conference on Recommender Systems (pp. 43–50). ACM. [\[return\]](#)
- He, X., Pan, J., Jin, O., Xu, T., Liu, B., Xu, T., ... & Candela, J. Q. (2014, August). Practical lessons from predicting clicks on ads at facebook. In Proceedings of the Eighth International Workshop on Data Mining for Online Advertising (pp. 1–9). ACM. [\[return\]](#)
- Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., ... & Anil, R. (2016, September). Wide & deep learning for recommender systems. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems (pp. 7–10). ACM. [\[return\]](#)

7. Dougherty, J., Kohavi, R., & Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. In Machine Learning Proceedings 1995 (pp. 194–202). [\[return\]](#)
8. Guo, H., Tang, R., Ye, Y., Li, Z., & He, X. (2017). Deepfm: A factorization-machine based neural network for CTR prediction. arXiv preprint arXiv:1703.04247. [\[return\]](#)
9. Ma, X., Zhao, L., Huang, G., Wang, Z., Hu, Z., Zhu, X., & Gai, K. (2018). Entire Space Multi-Task Model: An Effective Approach for Estimating Post-Click Conversion Rate. arXiv preprint arXiv:1804.07931. [\[return\]](#)
10. Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13(Feb), 281–305. [\[return\]](#)

美团深度学习系统的工程实践

作者: 剑鹏

背景

深度学习作为AI时代的核心技术，已经被应用于多个场景。在系统设计层面，由于其具有计算密集型的特性，所以与传统的机器学习算法在工程实践过程中存在诸多的不同。本文将介绍美团平台在应用深度学习技术的过程中，相关系统设计的一些经验。

本文将首先列举部分深度学习算法所需的计算量，然后再介绍为满足这些计算量，目前业界比较常见的一些解决方案。最后，我们将介绍美团平台在NLU和语音识别两个领域中，设计相关系统的经验。

深度学习的计算量

Model	Input Size	Param Size	Flops
AlexNet	227 × 227	233 MB	727 MFLOPs
CaffeNet	224 × 224	233 MB	724 MFLOPs
VGG–VD–16	224 × 224	528 MB	16 GFLOPs
VGG–VD–19	224 × 224	548 MB	20 GFLOPs
GoogleNet	224 × 224	51 MB	2 GFLOPs
ResNet–34	224 × 224	83 MB	4 GFLOPs
ResNet–152	224 × 224	230 MB	11 GFLOPs
SENet	224 × 224	440 MB	21 GFLOPs

数据来源

上表列举了，ImageNet图像识别中常见算法的模型大小以及单张图片一次训练（One Pass）所需要的计算量。

自2012年，Hinton的学生Alex Krizhevsky提出AlexNet，一举摘下ILSVRC 2012的桂冠后，ILSVRC比赛冠军的准确率越来越高。与此同时，其中使用到的深度学习算法也越来越复杂，所需要的计算量也越来越大。SENet与AlexNet相比，计算量多了近30倍。我们知道，ImageNet大概有120万张图片，以SENet为例，如果要完成100个epoch的完整训练，将需要 2.52×10^{18} 的计算量。如此庞大的计算量，已经远远超出传统的机器学习算法的范畴。更别说，Google在论文 [《Revisiting Unreasonable Effectiveness of Data in Deep Learning Era》](#) 中提及的、比ImageNet大300倍的数据集。

物理计算性能

面对如此庞大的计算量，那么，我们业界当前常用的计算单元的计算力是多少呢？

- CPU 物理核：一般浮点运算能力在 10^{10} FLOPS量级。一台16 Cores的服务器，大致上有200 GFLOPS的运算能力。实际运行，CPU 大概能用到80%的性能，那就160 GFLOPS的运算能力。完成上述SENet运行，需要182天。
- NVIDIA GPGPU：目前的V100，单精度浮点运算的峰值大概为14 TFLOPS，实际运行中，我们假设能用到50%的峰值性能，那就是7 TFLOPS，需要4天。

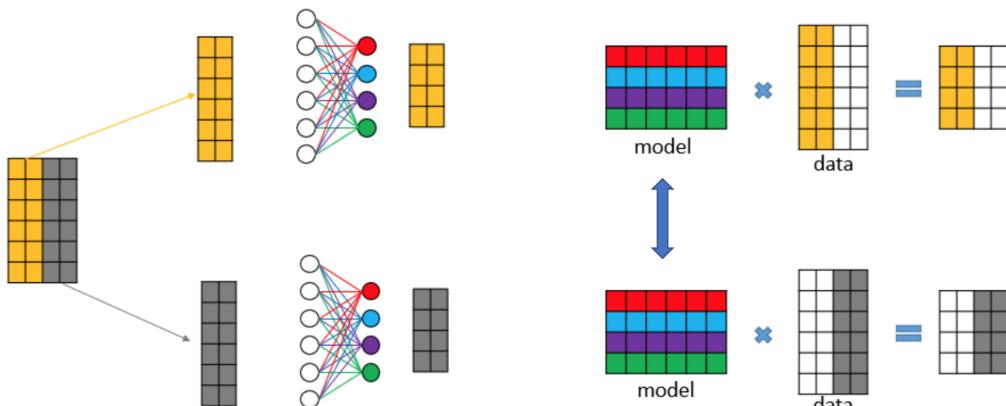
根据以上数据结果可以看出：在深度学习领域，GPU训练数据集所需要耗费的时间，远远少于CPU，这也是当前深度学习训练都是采用GPU的重要原因。

业界的解决方案

从前面的计算可知，即使使用GPU来计算，训练一次ImageNet 也需要4天的时间。但对于算法工程师做实验、调参而言，这种耗时数天的等待是难以忍受的。为此，目前业界针对深度学习训练的加速，提出了各种各样的解决方案。

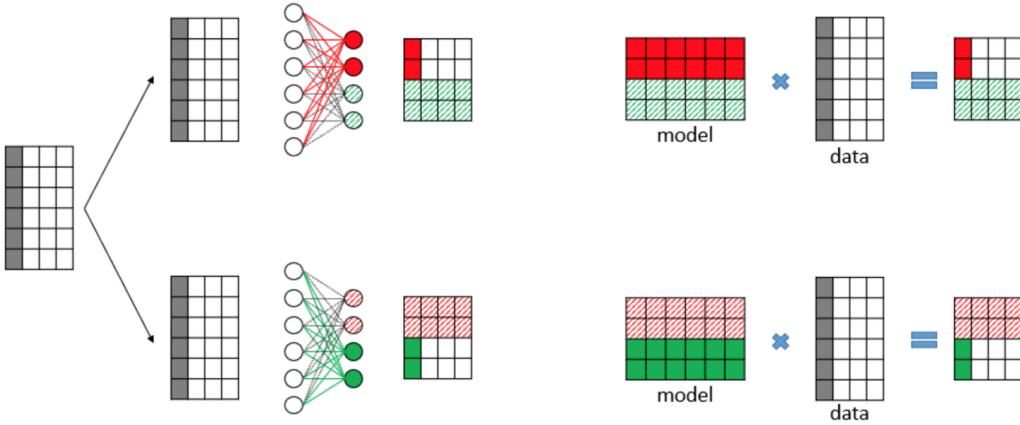
异构计算的并行方案

数据并行 (Data Parallelism)



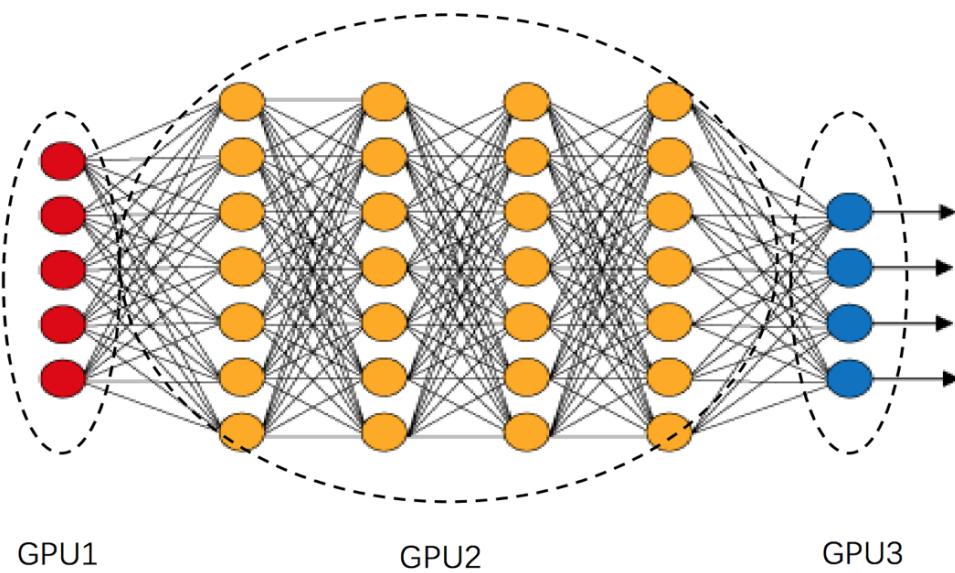
数据并行，即每个计算单元都保留一份完整的模型拷贝，分别训练不同的数据，经过一个Iteration或若干个Iteration后，把各个计算单元的模型做一次同步。这是最常见的深度学习训练方式，好处在于逻辑简单、代码实现方便。

模型并行 (Model Parallelism)



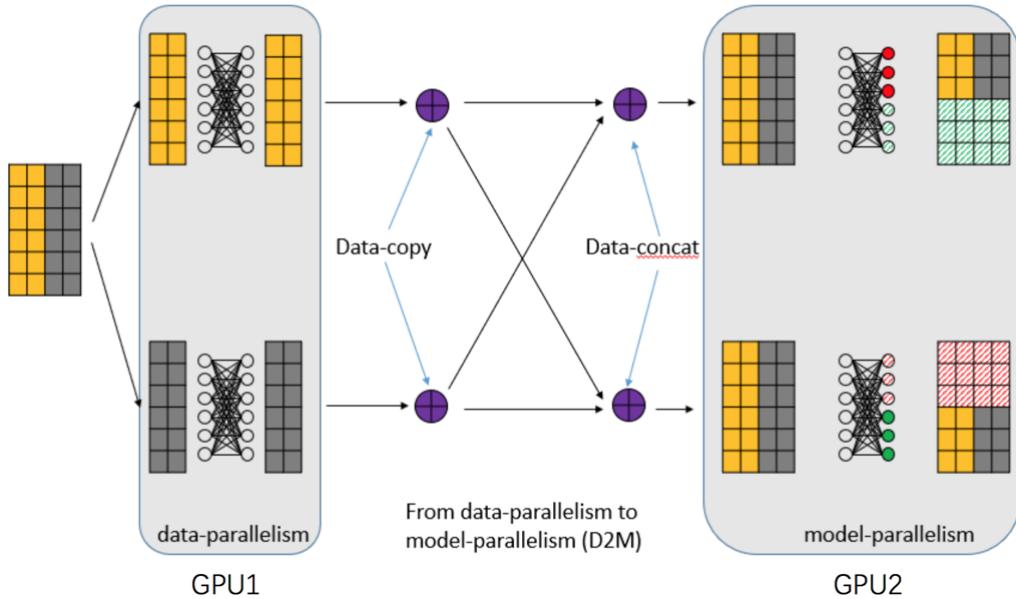
模型并行，即各个计算单元存储同一层模型数据的不同部分，训练相同的数据。相对于数据并行，因为各个运算单元每训练完一层神经网络，就必须要同步一次，频繁的同步通信导致系统不能充分地利用硬件的运算能力，所以更为少见。但是在一些业务场景下，Softmax层需要分类的类别可能会有很多，导致Softmax层太大，单个计算单元无法存储，这个时候，需要把模型切割成若干部分，存储在不同的运算单元。模型并行常见于NLU、推荐、金融等领域。

流式并行 (Stream Parallelism)



流式并行，即每个计算单元都存储不同层的模型数据，训练相同的数据。如上图所示，GPU1只负责第一层神经网络的计算，GPU2只负责2~5层神经网络的计算，GPU3只负责第6层的计算。流式并行的好处在于每个运算单元之间的通信和计算重叠（overlap），如果配置得当，可以非常充分地利用硬件资源。缺点在于，根据不同的模型，需要平衡好各个计算单元的计算量，如果配置不好，很容易形成“堰塞湖”。如上图所示，很有可能出现GPU1 负责的运算量太少，而GPU2 负责的运算量太多，导致GPU1 和GPU2 之间堵塞住大量的Mini-batch，更常见于线上环境。

混合并行 (Hybrid Parallelism)



混合并行，即上面提到的并行方式的混合。如对于一些图像识别任务来说，可能前几层使用数据并行，最后的Softmax层，使用模型并行。

异构计算的硬件解决方案

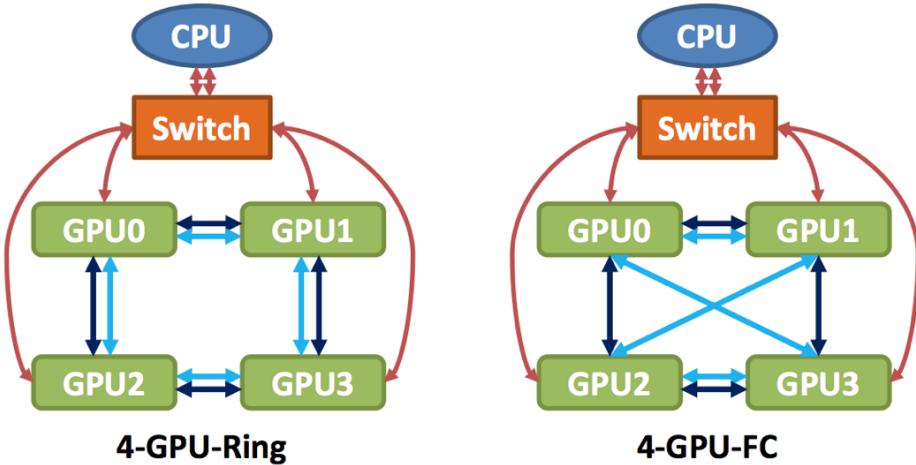
- 单机单卡：一个主机内安装上一块GPU运算卡。常见于个人计算机。
- 单机多卡：一个主机内安装上多块GPU运算卡。常见的有：1机4卡，1机8卡，甚至有1机10卡。一般公司都采取这种硬件方案。
- 多机多卡：多台主机内安装多块GPU运算卡。常见于公司内部的计算集群，一般多机之间采取Infiniband 来实现网络的快速通信。
- 定制化：即类似于Google的TPU解决方案。常见于“巨无霸”公司内部。

异构计算的通信解决方案

根据上面的硬件解决方案，我们以ResNet为例：模型的大小为230M，单张图片运算量为11 GFLOPS，Mini-batch假设为128。可以计算出各个硬件模块在深度学习训练中的耗时比较：

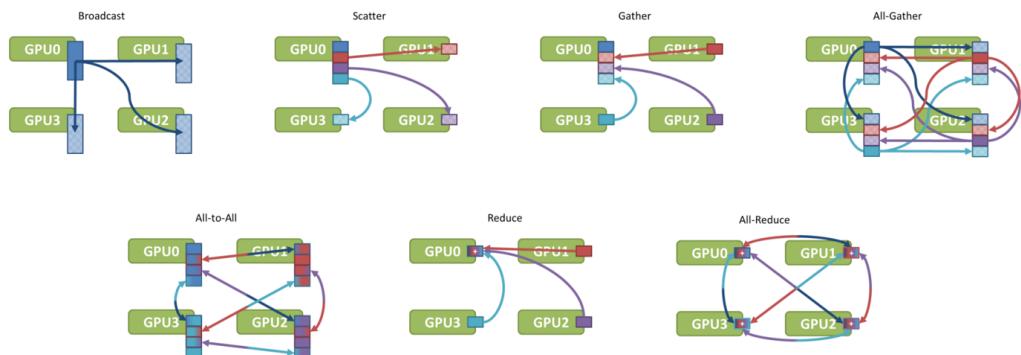
- GPU：对于V100，假设有6 TFLOPS，一次Mini-batch 理论耗时：0.23s。
- PCI-E：常见PCI-E 3.0 * 16，速度为10 GB/s，传输一个模型的理论耗时为：0.023s。
- 网络：假设为10 GB/s的高速网络，传输一个模型的理论耗时：0.023s。
- Disk：普通的磁盘，我们假设200M/s的读取速度，读取一次Mini-batch所需要的图片耗时：0.094s。

根据上面的数据结果，我们似乎可以得出一个结论：PCI-E和网络的传输耗时，相对于GPU来说，整整少了一个数量级，所以网络通信同步的时间可以忽略不计。然而问题并没有那么简单，上面例子中的耗时只是单个模型的耗时，但是对于8卡的集群来说，如果使用数据并行，每次同步就需要传输8份模型，这就导致数据传输的时间和GPU的计算时间“旗鼓相当”。这样的话，GPU就得每训练完一个Mini-batch，都得等候很久的一段时间（采取同步更新），这会浪费很多计算资源。因此，网络通信也需要制定对应的解决方案。下面我们将以Nvidia NCCL中单机多卡的通信解决方案为例介绍，而多机多卡的通信解决方案其实是类似的。

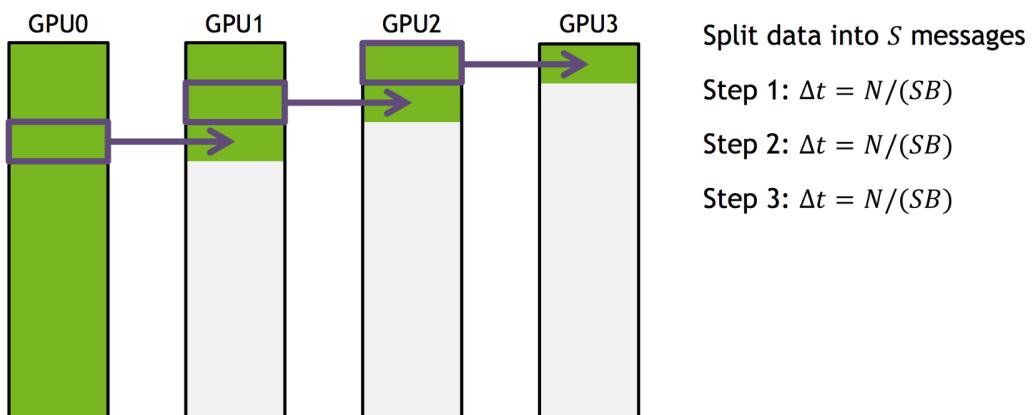


上图是单机4卡机器，在硬件上，两种不同的通信体系。左边为普通的PCI-E通信，即4个GPU之间组成一个环状。右边为NVLink通信，即两两之间相互连接。

常见的通信类型如下图所示：



对于深度学习训练而言，关键的两种通信类型为：Broadcast和Reduce。Broadcast用于Master分发最新的模型给各个GPU。Reduce 用于各个GPU计算完Mini-batch后，把模型更新值汇总到Master上。以 Broadcast为例，最简单的通信方式是Master往各个GPU上发送数据，这样的耗时就是4次模型传输的时间，通信时间就会太长，一种简单的优化方法如下图所示：



即把所需要传输的数据分成若干块，然后通过接力的方式逐个传递，每个GPU都把自己最新的一块数据发送到下一个GPU卡上。这种传输方式能充分利用硬件层面的通信结构，使得需要的耗时大幅缩减。与

此类似的，Reduce的通信优化也可以采取相同的方式进行提速。

美团的定制化深度学习系统

尽管目前在业界已经推出了很多著名的深度学习训练平台，通用的训练平台如TensorFlow、MxNet等等，还有领域专用的训练平台，如语音识别中的Kaldi，但是我们经过调研后，决定内部自主开发一套深度学习系统，理由如下：

- 通用的训练平台，缺乏了领域特色的功能。如语音识别中的特征提取模块和算法。
- 通用的训练平台，通常是基于Data-flow Graph，来对计算图中的每个operator进行建模，所以颗粒度很小，需要调度的单元多，导致任务调度复杂。
- 领域特色的训练平台，如Kaldi，在神经网络训练的时候，性能不足。
- 线上业务存在很多特殊性，如果使用TensorFlow之类作为训练平台，不太适合线上业务的情景。

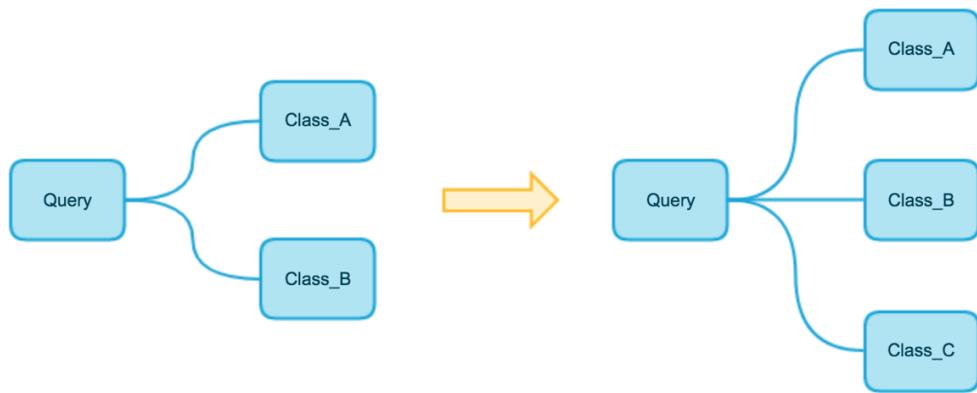
NLU线上系统

线上系统的业务特点

我们在设计NLU线上系统时，考虑了NLU业务的一些特性。发现其具备如下的一些特点： * 随着业务和技术的变化，算法流程也经常发生变化。 * 算法流程是多个算法串联组成的，不单纯的只有深度学习算法。如分词等算法就不是DL算法。 * 为了能够快速响应一些紧急问题，需要经常对模型进行热更新。 * 更重要的是，我们希望构建一个能以“数据驱动”的自动迭代闭环。

业务多变

NLU任务的算法流程是多层次级的，并且业务经常发生变化。如下图所示：



即随着业务要求的变化，NLU系统一开始的算法流程，只需要把一个Query分为两个类，但是到后面，极有可能会变成需要分为三个类别。

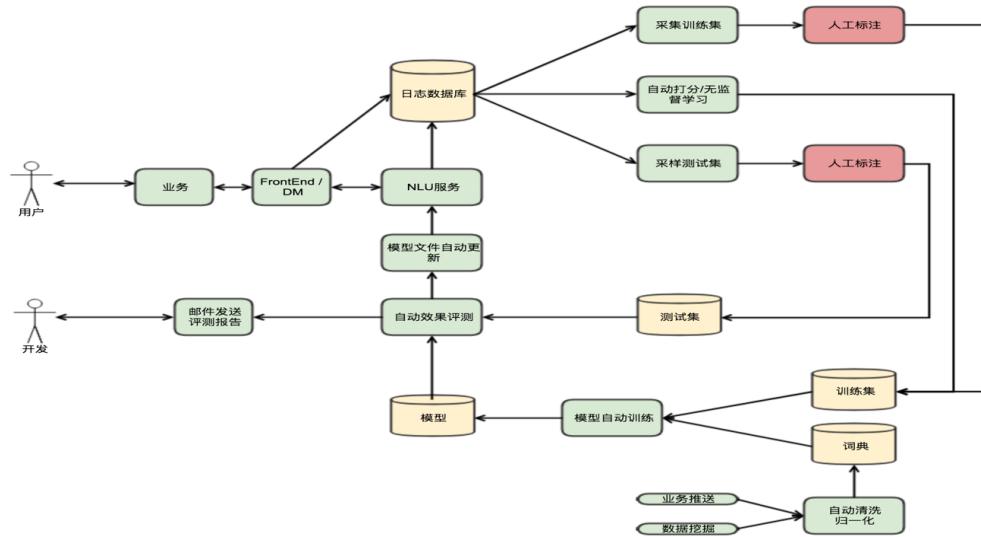
热更新

根据业务需求，或者为了紧急处理一些特殊问题，NLU线上系统经常需要做出快速响应，热更新算法模型。如最近的热点词“skr”，几乎是一夜之间，突然火爆起来。如下图所示的微博，如果不能正确理解“skr”的正确语义，可能就不能准确理解这条微博想要表达的意思。



为了避免影响用户体验，我们可能会对NLU系统，马上进行热更新，把新模型紧急进行上线。

数据驱动的自动迭代闭环

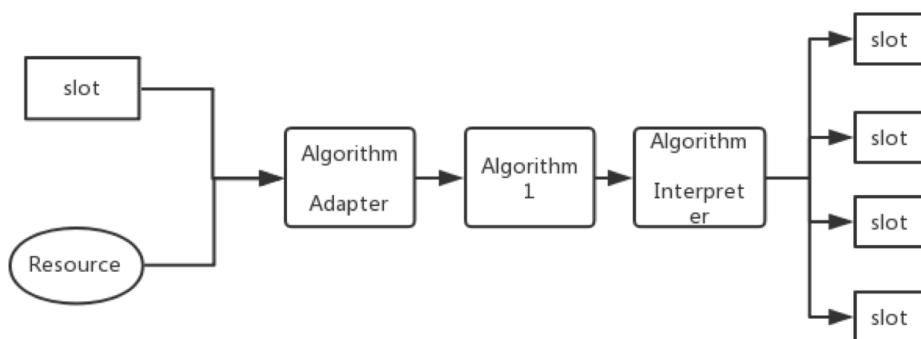


对于线上系统而言，构建如上图所示的自动迭代闭环，能更好地利用业务数据来提升服务质量。

NLU线上系统的核心设计

算法流程的抽象

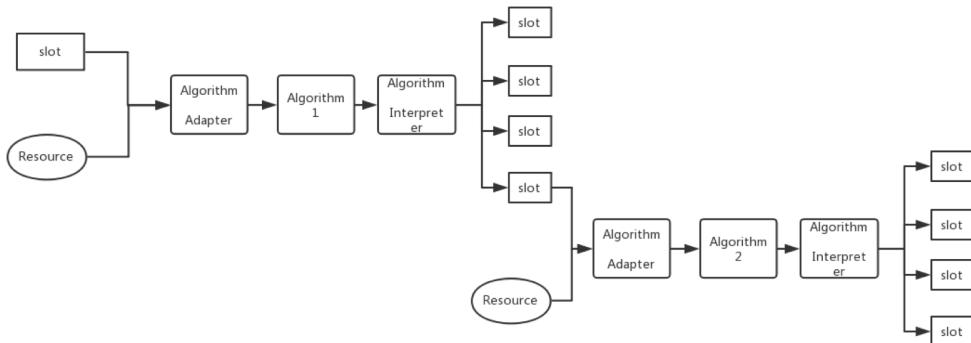
为了适应线上系统串联、多变的算法流程，我们把线上的算法进行抽象，如下图所示：



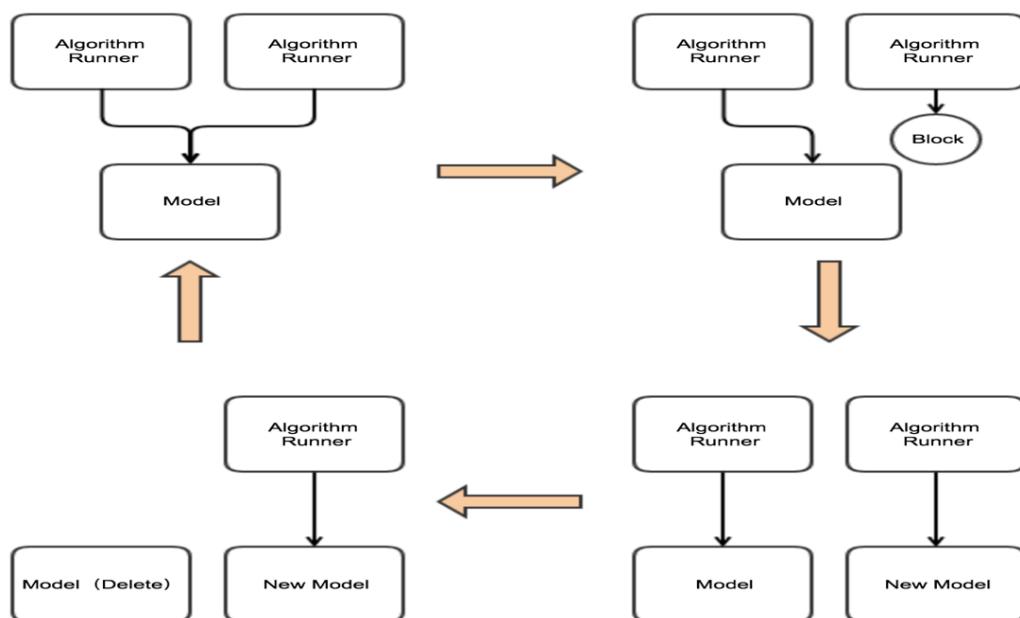
即每一个算法，都依赖于若干个槽位（Slot）和资源（Resource），一旦槽位和资源就位，就会触发对应的算法执行。算法的执行先通过算法适配器，来适配槽位和资源中的数据，转换成算子的输入格式。然

后算子执行算法本身，执行完算子后，再经过算法解析器。算法解析器主要用于解析算法执行的结果，触发对应的槽位。如根据算法的结果，触发Top 3的结果。

多个算法串联起来，就构建成如下结果：



热更新流程的设计



如上图所示，我们把算法的热更新流程设计如上。初试状态为左上角，即多个Query使用同一份模型数据。当遇到模型更新的请求后，系统将会block住新的query（右上角状态）。然后更新模型完后，新的query使用新的模型，旧query依然使用旧模型（右下角状态）。最后，当使用旧模型的query结束后，把旧的模型从内存中删除（左下角），然后系统恢复到初始状态。

声学模型训练系统

因为TensorFlow等通用深度学习训练平台，缺乏了特征提取等业务相关的领域功能，而Kaldi的声学模型训练过程又太慢。所以美团开发了一个声学模型训练系统——Mimir，其具备如下特性：

- 使用比TensorFlow更粗颗粒度的建模单元，使得任务调度、优化更简单方便易行。
 - 使用数据并行的并行方案，单机多卡可达到近线性加速。（采取同步更新策略下，4卡加速比达到3.8）
 - 移植了Kaldi的一些特有的训练算法。
 - 速度上为Kaldi的6~7倍。（800个小时的训练数据，单机单卡的条件下，Kaldi需要6~7天，Mimir只需20个小时）

- 业务上，移植了Kaldi的特征提取等领域的相关模块。

参考资料

- [NCCL: ACCELERATED MULTI-GPU COLLECTIVE COMMUNICATIONS](#)
- [【深度学习】老师木讲架构：深度学习平台技术演进](#)

作者简介

- 剑鹏，美团点评算法专家。2017年加入美团，目前作为语音识别团队的声学模型负责人，负责声学模型相关的算法和系统设计与开发。

美团餐饮娱乐知识图谱——美团大脑揭秘

作者: 仲远

前言

“I’m sorry. I can’t do that, Dave.”这是经典科幻电影《2001: A Space Odyssey》里HAL 9000机器人说的一句话，浓缩了人类对终极人工智能的憧憬。让机器学会说这样简单一句话，需要机器具备情感认知、自我认识以及对世界的认识，来辅助机器处理接收到的各种信息，了解信息背后的意思，从而生成自己的决策。而这些认知模块的基础，都需要机器具备知识学习组织推理的能力，知识图谱就是为实现这些目标而生。

今年5月，美团NLP中心开始构建大规模的餐饮娱乐知识图谱——美团大脑，它将充分挖掘关联各个场景数据，用AI技术让机器“阅读”用户评论数据，理解用户在菜品、价格、服务、环境等方面喜好，挖掘人、店、商品、标签之间的知识关联，从而构建出一个“知识大脑”。美团大脑已经在公司多个业务中初步落地，例如智能搜索推荐、智能金融、智能商户运营等。

此前，《[美团大脑：知识图谱的建模方法及其应用](#)》一文，介绍了知识图谱的分类及其具体应用，尤其是常识性知识图谱及百科全书式知识图谱分别是如何使用的。之后我们收到非常多的反馈，希望能进一步了解“美团大脑”的细节。为了让大家更系统地了解美团大脑，NLP中心会在接下来一段时间，陆续分享一系列技术文章，包括知识图谱相关的技术，美团大脑背后的算法能力，千亿级别图引擎建设以及不同应用场景的业务效果等等，本文是美团大脑系列的第一篇文章。

迈向认知智能

“

海量数据和大规模分布式计算力，催生了以深度学习为代表的第三次（1993–目前）人工智能高潮。Web 2.0产生的海量数据给机器学习和深度学习技术提供了大量标注数据，而GPU和云计算的发展为深度学习的复杂数值计算提供了必要算力条件。深度学习技术在语音、图像领域均取得了突破性的进展，这表示学习技术成果使得机器首次在感知能力上达到甚至超越了人类的水平，人工智能已经进入感知智能阶段。

然而，随着深度学习被广泛应用，其局限性也愈发明显。

- **缺乏可解释性：**神经网络端到端学习的“黑箱”特性使得很多模型不具有可解释性，导致很多需要人去参与决策，在这些应用场景中机器结果无法完全置信而需要谨慎的使用，比如医学的疾病诊断、金融的智能投顾等等。这些场景属于低容错高风险场景，必须需要显示的证据去支持模型结果，从而辅助人去做决策。
- **常识（Common Sense）缺失：**人的日常活动需要大量的常识背景知识支持，数据驱动的机器学习和深度学习，它们学习到的是样本空间的特征、表征，而大量的背景常识是隐式且模糊的，很难在样本数据中进行体现。比如下雨要打伞，但打伞不一定都是下雨天。这些特征数据背后的关联逻辑隐藏在我们的文化背景中。
- **缺乏语义理解。**模型并不理解数据中的语义知识，缺乏推理和抽象能力，对于未见数据模型泛化能力差。

- 依赖大量样本数据：**机器学习和深度学习需要大量标注样本数据去训练模型，而数据标注的成本很高，很多场景缺乏标注数据来进行冷启动。

能力表现	数据驱动的 AI	知识驱动的 AI
目前进展	在一些任务上已经接近或超过人类	在知识量上超过人类，在知识推理上不如人类
可解释性	弱	强
鲁棒性	弱	强
泛化能力	弱	强
重大错误	大	小
应用场景	数据丰富	知识丰富

图1 数据知识驱动AI能力对比

从人工智能整体发展来说，综上的局限性也是机器从感知智能向认知智能的迁跃的过程中必须解决的问题。认知智能需要机器具备推理和抽象能力，需要模型能够利用先验知识，总结出人可理解、模型可复用的知识。机器计算能力整体上需要从数据计算转向知识计算，知识图谱就显得必不可少。知识图谱可以组织现实世界中的知识，描述客观概念、实体、关系。这种基于符号语义的计算模型，一方面可以促成人和机器的有效沟通，另一方面可以为深度学习模型提供先验知识，将机器学习结果转化为可复用的符号知识累积起来。

知识究竟是什么呢？知识就是有结构的信息。人从数据中提取有效信息，从信息中提炼有用知识，信息组织成了结构就有了知识。知识工程，作为代表人工智能发展的主要研究领域之一，就是机器仿照人处理信息积累知识运用知识的过程。而知识图谱就是知识工程这一领域数十年来的代表性研究方向。在数据还是稀有资源的早期，知识图谱的研究重点偏向语义模型和逻辑推理，知识建模多是自顶向下的设计模式，语义模型非常复杂。其中典型工作，是在1956年人工智能学科奠基之会——达特茅斯会议上公布的“逻辑理论家”(Logic Theorist)定理证明程序，该程序可以证明《数学原理》中的部分定理。伴随着Web带来前所未有的数据之后，知识图谱技术的重心从严谨语义模型转向海量事实实例构建，图谱中知识被组织成<主，谓，宾>三元组的形式，来表征客观世界中的实体和实体之间的关系。比如像名人的维基百科词条页面中，InfoBox卡片都会描述该名人的国籍信息，其结构就是<人，国籍，国家>这样的三元组。



图2 互联网公司知识图谱布局

目前，知识图谱已被广泛应用在问答、搜索、推荐等系统，已涉及金融、医疗、电商等商业领域，图谱技术成为“兵家必争”之地。微软于2010年开始构建Satori知识图谱来增强Bing搜索；Google在2012年提出Knowledge Graph概念，用图谱来增强自己的搜索引擎；2013年Facebook发布Open Graph应用于社交网络智能搜索；2015年阿里巴巴开始构建自己的电商领域知识图谱；2016年Amazon也开始构建知识图谱。



图3 美团大脑

2018年5月，美团点评NLP中心开始构建大规模的餐饮娱乐知识图谱——美团大脑。美团点评作为中国最大的在线本地生活服务平台，覆盖了餐饮娱乐领域的众多生活场景，连接了数亿用户和数千万商户，积累了宝贵的业务数据，蕴含着丰富的日常生活相关知识。在建的美团大脑知识图谱目前有数十类概念，数十亿实体和数百亿三元组，美团大脑的知识关联数量预计在未来一年内将上涨到数千亿的规模。

美团大脑将充分挖掘关联各个场景数据，用AI技术让机器“阅读”用户评论和行为数据，理解用户在菜品、价格、服务、环境等方面的好恶，构建人、店、商品、场景之间的知识关联，从而形成一个“知识大脑”。相比于深度学习的“黑盒子”，知识图谱具有很强的可解释性，在美团跨场景的多个业务中应用性非常强，已经在搜索、金融等场景中初步验证了知识图谱的有效性。近年来，深度学习和知识图谱技术

都有很大的发展，并且存在一种互相融合的趋势，在美团大脑知识构建过程中，我们也会使用深度学习技术，把数据背后的知识挖掘出来，从而赋能业务，实现智能化的本地生活服务，帮助每个人“Eat Better, Live Better”。

知识图谱技术链

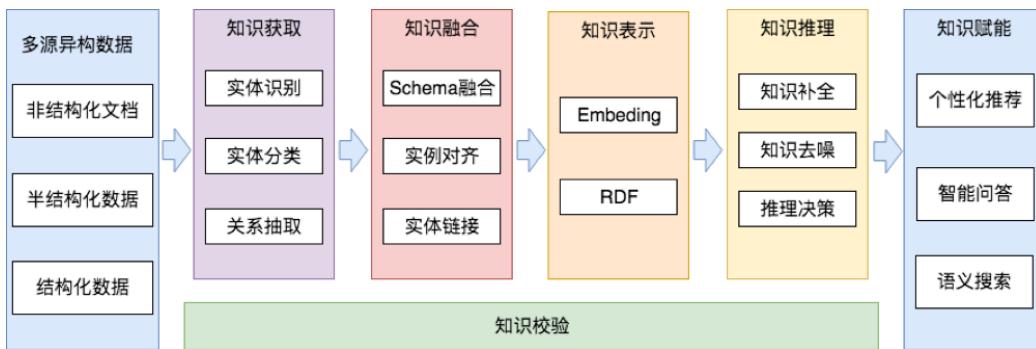


图4 知识图谱技术链

知识图谱的源数据来自多个维度。通常来说，结构化数据处理简单、准确率高，其自有的数据结构设计，对数据模型的构建也有一定指导意义，是初期构建图谱的首要选择。世界知名的高质量的大规模开放知识库如Wikidata、DBpedia、Yago是构建通用领域多语言知识图谱的首选，国内有OpenKG提供了诸多中文知识库的Dump文件或API。工业界往往基于自有的海量结构化数据，进行图谱的设计与构建，并同时利用实体识别、关系抽取等方式处理非结构化数据，增加更多丰富的信息。

知识图谱通常以实体为节点形成一个大的网络，图谱的Schema相当于数据模型，描述了领域下包含的类型 (Type)，与类型下描述实体的属性 (Property)，Property中实体与实体之间的关系为边 (Relation)，实体自带信息为属性 (Attribute)。除此之外Schema也会描述它们的约束关系。

美团大脑围绕用户打造吃喝玩乐全方面的知识图谱，从实际业务需求出发，在现有数据表之上抽象出数据模型，以商户、商品、用户等为主要实体，其基本信息作为属性，商户与商品、与用户的关联为边，将多领域的信息关联起来，同时利用评论数据、互联网数据等，结合知识获取方法，填充图谱信息，从而提供更加多元化的知识。

知识获取

知识获取是指从不同来源、不同结构数据中，抽取相关实体、属性、关系、事件等知识。从数据结构划分可以分为结构化数据、半结构化数据和纯文本数据。结构化数据指的关系型数据库表示和存储的二维形式数据，这类数据可以直接通过Schema融合、实体对齐等技术将数据提取到知识图谱中。半结构化数据主要指有相关标记用来分隔语义元素，但又不存在数据库形式的强定义数据，如网页中的表格数据、维基百科中的Infobox等等。这类数据通过爬虫、网页解析等技术可以将其转换为结构化数据。现实中结构化、半结构化数据都比较有限，大量的知识往往存在于文本中，这也和人获取知识的方式一致。对应纯文本数据获取知识，主要包括实体识别、实体分类、关系抽取、实体链接等技术。

实体作为知识图谱的核心单位，从文本中抽取实体是知识获取的一个关键技术。文本中识别实体，一般可以作为一个序列标注问题来进行解决。传统的实体识别方法以统计模型如HMM、CRF等为主导，随着深

深度学习的兴起，BiLSTM+CRF[1]模型备受青睐，该模型避免了传统CRF的特征模版构建工作，同时双向LSTM能更好地利用前后的语义信息，能够明显提高识别效果。在美团点评-美食图谱子领域的建设中，每个店家下的推荐菜（简称店菜）是图谱中的重要实体之一，评论中用户对店菜的评价，能很好地反映用户偏好与店菜的实际特征，利用知识获取方法，从评论中提取出店菜实体、用户对店菜的评价内容与评价情感，对补充实体信息、分析用户偏好、指导店家进行改善有着非常重要的意义。

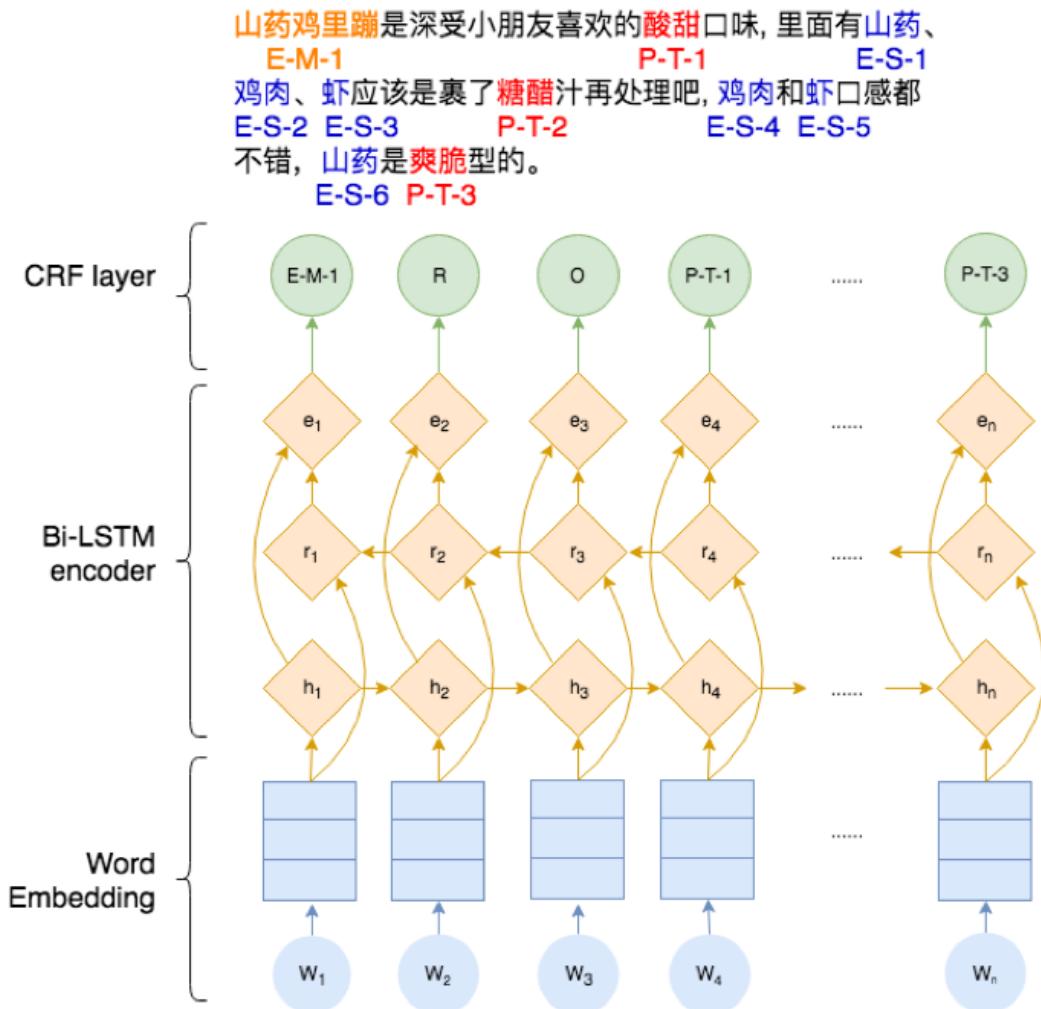


图5 BiLSTM+CRF模型

实体分类则是对抽取出来的实体进行归类。当从文本中发现一个新的实体，给实体相应的Type是实体概念化的基本目标。比如用该实体的上下文特征与其他Type下的实体特征进行对比，将新实体归入最相似的Type中。此外，在Schema不完善的情况下，对大量实体进行聚类，进而抽象出每个簇对应的Type，是自底向上构建图谱的一个常用方法，在补充Type层的同时，也顺便完成了实体归类。

关系抽取，是从文本中自动抽取实体与实体之间的特定的语义关系，以补充图谱中缺失的关系，例如，从“干酪鱼原来是奶酪做的”中抽取出<干酪鱼，食材，奶酪>。关系抽取可以通过定义规则模版来获取，如匹配某种表达句式、利用文法语义特征等，但规则类方法消耗大量人力，杂质较多。基于Bootstrap Learning的方法利用少量种子实例或模版抽取新的关系，再利用新的结果生成更多模版，如此迭代，KnowItAll[2]、TextRunner[3]基于这类思想；远程监督（Distant Supervision）方法[4]把现有的三元组信息作为种子，在文本中匹配同时含有主语和宾语的信息，作为关系的标注数据。这两种方法解决了人力耗费问题，但准确率还有待提高。近期的深度学习方法则基于联合模型思想，利用神经网络的端对端模

型，同时实现实体识别和关系抽取[5][6]，从而避免前期实体识别的结果对关系抽取造成的误差累积影响。

知识校验

知识校验贯穿整个知识图谱的构建过程。在初期的Schema设计过程中，需要严格定义Type下的Property，Property关联的是属性信息还是实体，以及实体所属的Type等等。Schema若不够规范，会导致错误传达到数据层且不易纠错。在数据层，通过源数据获取或者通过算法抽取的知识或多或少都包含着杂质，可以在Schema层面上，添加人工校验方法与验证约束规则，保证导入数据的规范性，比如对于<店A, 包含, 店菜B>关系，严格要求主语A的Type是POI，宾语B的Type是Dish。而对于实体间关系的准确性，如上下位关系是否正确、实例的类型是否正确，实例之间的关系是否准确等，可以利用实体的信息与图谱中的结构化信息计算一个关系的置信度，或看作关系对错与否的二分类问题，比如<店A, 适合, 情侣约会>，对于“情侣约会”标签，利用店A的信息去计算一个权重会使得数据更有说服力。此外，如果涉及到其他来源的数据，在数据融合的同时进行交叉验证，保留验证通过的知识。当图谱数据初步成型，在知识应用过程中，通过模型结果倒推出的错误，也有助于净化图谱中的杂质，比如知识推理时出现的矛盾，必然存在知识有误的情况。

知识融合

知识融合主要解决多源异构数据整合问题，即从不同来源、不同结构但表达统一实体或概念的数据融合为一个实体或概念。融入来自多源数据的知识，必然会涉及知识融合工作，实体融合主要涉及Schema融合、实体对齐、实体链接等技术。

Schema是知识图谱的模型，其融合等价于Type层的合并和Property的合并。在特定领域的图谱中，Type与Property数量有限，可以通过人工进行合并。对于实例的对齐，可以看作一个寻找Top匹配的实例的排序问题，或者是否匹配的二分类问题，其特征可以基于实体属性信息、Schema结构化信息、语义信息等来获取。

实体对齐是多源数据融合中的重要过程。当数据来自于不同的知识库体系，需要分辨其描述的是同一个实体，将相关信息融合，最终生成该知识库中唯一的实体。这通常是一个求最相似问题或判断两个实体是否是同一个的二分类问题，实体名称、实体携带属性以及其结构化信息，都可以作为有用特征。同时，通过Type或规则限制，缩小匹配的实体范围。

一旦图谱构建完成，如何从文本中准确匹配上图谱中相应的实体，进而延伸出相关的背景知识，则是一个实体链接问题。实体链接[7] 主要依赖于实体Entity与所有Mention（文本文档中实体的目标文本）的一个多对多的映射关系表，如“小龙虾”这个Mention在图谱中实际对应的实体Entity可能是“麻辣小龙虾”的菜，也可能是“十三香小龙虾”的菜。对于从文本中识别出的Mention，利用上下文等信息，对其候选Entity进行排序，找出最可能的Entity。实体链接可以正确地定位用户所提实体，理解用户真实的表达意图，从而进一步挖掘用户行为，了解用户偏好。

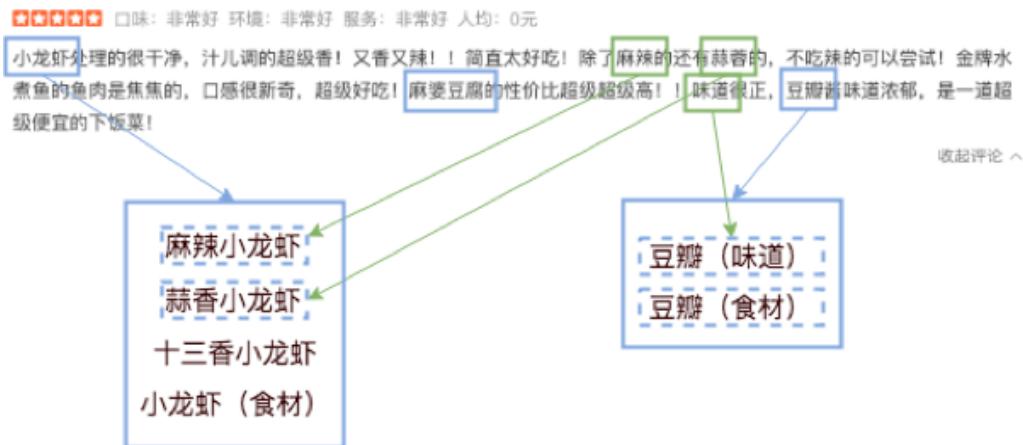


图6 实体链接 (Entity Linking)

美团大脑也参考并融入了多源的数据信息，知识融合是构建图谱的一个重要步骤。以美食领域子图谱为例，该图谱是由结构化数据和文本挖掘出来的知识融合而成，首要任务是将图谱中已构建的菜品通过菜名、口味、食材等方面相似度将菜品与文本挖掘出来的菜品知识进行关联，其次还要对无法关联的菜品知识聚类抽象成一个菜品实体。知识的融合很大程度上增加了菜品的数量，丰富了菜品信息，同时为实体链接的映射关系表提供了候选对，有助于我们在搜索过程中，支持更多维度（如口味、食材）的查询。

知识表示

知识表示是对知识数据的一种描述和约定，目的是让计算机可以像人一样去理解知识，从而可以让计算机进一步的推理、计算。大多数知识图谱是以符号化的方法表示，其中RDF是最常用的符号语义表示模型，其一条边对于一个三元组<主语Subject，谓语Predicate，宾语Object>，表达一个客观事实，该方法直观易懂，具备可解释性，支持推理。

而随着深度学习的发展，基于向量表示的Embedding算法逐渐兴起，其为每个实体与关系训练一个可表征的向量，该方法易于进行算法学习，可表征隐形知识并进一步发掘隐形知识。常用的Embedding模型有Word2Vec与Trans系列[8][9]，将会在之后的文章里进一步讲解。美团大脑参考Freebase的建模思想，以< Subject, Predicate, Object>的三元组形式将海量知识存储在分布式数据仓库中，并以CVT(Compound Value Type)设计承载多元数据，即抽象一个CVT的实例来携带多元信息，图为一个知识表示的例子。与此同时，美团大脑基于上亿节点计算Graph Embedding的表征，并将结果应用到搜索领域中。

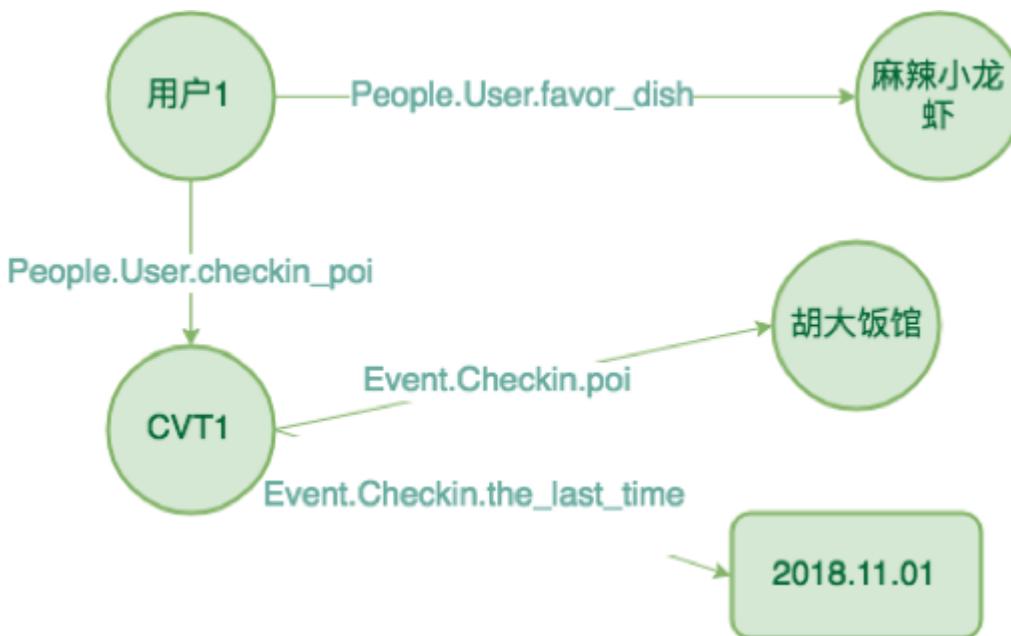


图7 美团大脑知识表示

知识推理

基于知识图谱的推理工作，旨在依据现有的知识信息推导出新知识，包括实体关系、属性等，或者识别出错误关系。可以分为基于符号的推理与基于统计的推理，前者一般根据经典逻辑创建新的实体关系的规则，或者判断现有关系的矛盾之处，后者则是通过统计规律从图谱中学到新的实体关系。

利用实体之间的关系可以推导出一些场景，辅助进行决策判断。美团大脑金融子图谱利用用户行为、用户关系、地理位置去挖掘金融领域诈骗团伙。团伙通常会存在较多关联及相似特性，图谱中的关系可以帮助人工识别出多层、多维度关联的欺诈团伙，再利用规则等方式，识别出批量具有相似行为的客户，辅助人工优化调查，同时可以优化策略。

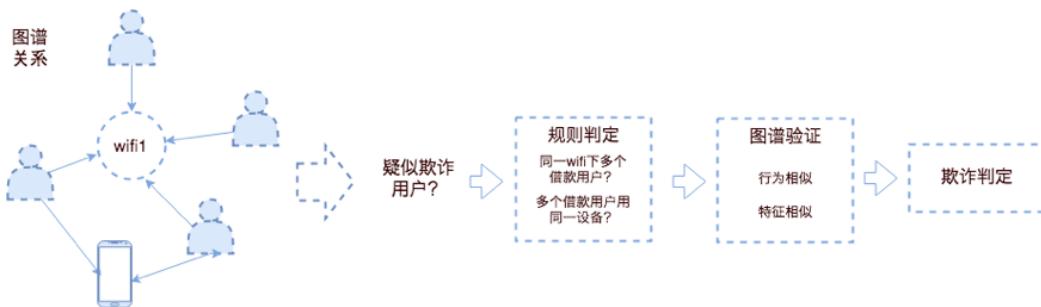


图8 知识推理在金融场景应用

知识赋能

知识图谱含有丰富的语义信息，对文本有基于语义的更为深入的理解，在推荐、搜索、问答等领域能提供更加直接与精确的查询结果，使得服务更加智能化。

个性化推荐通过实体与实体之间的关系，利用用户感兴趣的实体，进一步扩展用户偏好的相似的实体，提供可解释性的推荐内容。一方面，图谱提供了实体在多个维度的特征信息，另一方面，表示学习向量带有一定的语义信息，使得寻找推荐实体更接近目标实体或更偏向用户喜好。

语义搜索，是指搜索引擎对Query的处理不再拘泥于字面本身，而是抽象出其中的实体、查询意图，通过知识图谱直接提供用户需要的答案，而不仅是提供网页排序结果，更精准的满足用户的需求。当前Google、百度、神马搜索都已经将基于知识图谱的语义搜索融入到搜索引擎中，对于一些知识性内容的查找，能智能地直接显示结果信息。

美团大脑的业务应用

依托深度学习模型，美团大脑充分挖掘、关联美团点评各个业务场景公开数据（如用户评价、菜品、标签等），正在构建餐饮娱乐“知识大脑”，并且已经开始在美团不同业务中进行落地，利用人工智能技术全面提升用户的生活体验。

智能搜索：帮助用户做决策

知识图谱可以从多维度精准地刻画商家，已经在美食搜索和旅游搜索中应用，为用户搜索出更适合Ta的店。基于知识图谱的搜索结果，不仅具有精准性，还具有多样性，例如：当用户在美食类目下搜索关键词“鱼”，通过图谱可以认知到用户的搜索词是“鱼”这种“食材”。因此搜索的结果不仅有“糖醋鱼”、“清蒸鱼”这样的精准结果，还有“赛螃蟹”这样以鱼肉作为主食材的菜品，大大增加了搜索结果的多样性，提升用户的搜索体验。并且对于每一个推荐的商家，能够基于知识图谱找到用户最关心的因素，从而生成“千人千面”的推荐理由，例如在浏览到大董烤鸭店的时候，偏好“无肉不欢”的用户A看到的推荐理由是“大董的烤鸭名不虚传”，而偏好“环境优雅”的用户B，看到的推荐理由就是“环境小资，有舞台表演”，不仅让搜索结果更具有解释性，同时也能吸引不同偏好的用户进入商家。



图9 知识图谱在点评搜索中应用

对于场景化搜索，知识图谱也具有很强的优势，以七夕节为例，通过知识图谱中的七夕特色化标签，如约会圣地、环境私密、菜品新颖、音乐餐厅、别墅餐厅等等，结合商家评论中的细粒度情感分析，为美团搜索提供了更多适合情侣过七夕节的商户数据，用于七夕场景化搜索的结果召回与展示，极大的提升了用户体验和用户点击转化。

在NLP中心以及大众点评搜索智能中心两个团队的紧密合作下，依赖知识图谱技术和深度学习技术对搜索架构进行了整体的升级。经过5个月时间，点评搜索核心指标在高位基础上，仍然有非常明显的提升。

ToB商户赋能：商业大脑指导店老板决策

美团大脑正在应用于SaaS收银系统专业版，通过机器智能阅读每个商家的每一条评论，可以充分理解每个用户对于商家的感受，针对每个商家将大量的用户评价进行归纳总结，从而可以发现商家在市场上的竞争优势/劣势、用户对于商家的总体印象趋势、商家的菜品的受欢迎程度变化。进一步，通过细粒度用户评论全方位分析，可以细致刻画商家服务现状，以及对商家提供前瞻性经营方向。这些智能经营建议将通过美团SaaS收银系统专业版定期触达到各个商家，智能化指导商家精准优化经营模式。

传统给店老板提供商业分析服务中主要聚焦于单店的现金流、客源分析。美团大脑充分挖掘了商户及顾客之间的关联关系，可以提供围绕商户到顾客，商户到所在商圈的更多维度商业分析，在商户营业前、营业中以及将来经营方向，均可以提供细粒度运营指导。

在商家服务能力分析上，通过图谱中关于商家评论所挖掘的主观、客观标签，例如“服务热情”、“上菜快”、“停车免费”等等，同时结合用户在这些标签所在维度上的Aspect细粒度情感分析，告诉商家在哪些方面做的不错，是目前的竞争优势；在哪些方面做的还不够，需要尽快改进。因而可以更准确地指导商家进行经营活动。更加智能的是，美团大脑还可以推理出顾客对商家的认可程度，是高于还是低于其所在商圈的平均情感值，让店老板一目了然地了解自己的实际竞争力。

在消费用户群体分析上，美团大脑不仅能够告诉店老板来消费的顾客的年龄层、性别分布，还可以推理出顾客的消费水平，对于就餐环境的偏好，适合他们的推荐菜，让店老板有针对性的调整价格、更新菜品、优化就餐环境。

金融风险管理与反欺诈：从用户行为建立征信体系

知识图谱的推理能力和可解释性，在金融场景中具有天然的优势，NLP中心和美团金融共建的金融好用户扩散以及用户反欺诈，就是利用知识图谱中的社区发现、标签传播等方法来对用户进行风险管理，能够更准确的识别逾期客户以及用户的不良行为，从而大大提升信用风险管理能力。

在反欺诈场景中，知识图谱已经帮助金融团队在案件调查中发现并确认多个欺诈案件。由于团伙通常会存在较多关联及相似特性，关系图可以帮助识别出多层、多维度关联的欺诈团伙，能通过用户和用户、用户和设备、设备和设备之间四度、五度甚至更深度的关联关系，发现共用设备、共用Wi-Fi来识别欺诈团伙，还可在已有的反欺诈规则上进行推理预测可疑设备、可疑用户来进行预警，从而成为案件调查的有力助手。

未来的挑战

知识图谱建设过程是美团第一次摸索基于图的构建/挖掘/存储/应用过程，也遇到了很多挑战，主要的挑战和应对思路如下：

(1) 数据生成与导入

- 难点：Schema构建和更新；数据源多，数据不一致问题；数据质检。
- 应对思路：通过针对不同的数据进行特定清洗，元数据约束校验、业务逻辑正确性校验等，设置了严格的数据接入和更新规范。

(2) 知识挖掘

- 难点：知识的融合、表征、推理和验证。
- 应对思路：通过借鉴文本中的词向量表征，为知识建立统一的语义空间表征，使得语义可计算，基于深度学习和知识表示的算法进行推理。

(3) 百亿图存储及查询引擎

- 难点：数据的存储、查询和同步，数据量极大，没有成熟开源引擎直接使用。
- 应对思路：构建分层增量系统，实时增量、离线增量、全量图三层Merge查询，减少图更新影响范围。同时建设完整的容灾容错、灰度、子图回滚机制。基于LBS等业务特点合理切分子图View，构建分布式图查询索引层。

(4) 知识图谱应用挑战

- 难点：算法设计，系统实现难和实时应用。
- 应对思路：知识图谱的应用法则需要有效融合数据驱动和知识引导，才能提升算法效果和提供更好的解释性，属于研究前沿领域。百亿甚至千亿关系规模下，需要设计和实现分布式的图应用算法，这对算法和系统都有重大的挑战。

总而言之，为打造越来越强大的美团大脑，NLP中心一方面利用业界前沿的算法模型来挖掘关联以及应用知识，另一方面，也在逐步建立国内领先的商业化分布式图引擎系统，支撑千亿级别知识图谱的实时图查询、图推理和图计算。在未来的系列文章中，NLP中心将一一揭秘这背后的创新性技术，敬请期待。

参考文献

- [1] Huang, Zhiheng, Wei Xu, and Kai Yu. “Bidirectional LSTM–CRF models for sequence tagging.” arXiv preprint arXiv:1508.01991 (2015).
- [2] Etzioni, Oren, et al. “Unsupervised named–entity extraction from the web: An experimental study.” Artificial intelligence165.1 (2005): 91–134.
- [3] Banko, Michele, et al. “Open information extraction from the web.” IJCAI. Vol. 7. 2007.
- [4] Mintz, Mike, et al. “Distant supervision for relation extraction without labeled data.” Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2–Volume 2. Association for Computational Linguistics, 2009.
- [5] Zheng, Suncong, et al. “Joint entity and relation extraction based on a hybrid neural network.” Neurocomputing 257 (2017): 59–66.
- [6] Zheng, Suncong, et al. “Joint extraction of entities and relations based on a novel tagging scheme.” arXiv preprint arXiv:1706.05075 (2017).
- [7] Shen, Wei, Jianyong Wang, and Jiawei Han. “Entity linking with a knowledge base: Issues, techniques, and solutions.” IEEE Transactions on Knowledge and Data Engineering 27.2 (2015): 443–460.

- [8] Bordes, Antoine, et al. “Translating embeddings for modeling multi-relational data.” Advances in neural information processing systems. 2013.
- [9] Wang, Zhen, et al. “Knowledge Graph Embedding by Translating on Hyperplanes.” AAAI. Vol. 14. 2014.

作者简介

- 仲远，博士，美团AI平台部NLP中心负责人，点评搜索智能中心负责人。在国际顶级学术会议发表论文30余篇，获得ICDE 2015最佳论文奖，并是ACL 2016 Tutorial “Understanding Short Texts”主讲人，出版学术专著3部，获得美国专利5项。此前，博士曾担任微软亚洲研究院主管研究员，以及美国Facebook公司Research Scientist。曾负责微软研究院知识图谱、对话机器人项目和Facebook产品级NLP Service。
- 富峥，博士，美团AI平台NLP中心研究员，目前主要负责美团大脑项目。在此之前，博士在微软亚洲研究院社会计算组担任研究员，并在相关领域的顶级会议和期刊上发表30余篇论文，曾获ICDM2013最佳论文大奖，出版学术专著1部。张富峥博士曾担任ASONAM的工业界主席，IJCAI、WSDM、SIGIR等国际会议和TKDE、TOIS、TIST等国际期刊的评审委员。
- 王珺，博士，美团AI平台NLP中心产品和数据负责人。在此之前，王珺在阿里云负责智能顾问多产品线，推动建立了阿里云智能服务体系。
- 明洋，硕士，美团AI平台NLP中心知识图谱算法工程师。2016年毕业于清华大学计算机系知识工程实验室。
- 思睿，硕士，美团AI平台NLP中心知识图谱算法专家。此前在百度AIG知识图谱部负责知识图谱、NLP相关算法研究，参与了百度知识图谱整个构建及落地过程。
- 一飞，负责AI平台NLP中心知识图谱产品。目前主要负责美团大脑以及知识图谱落地项目。
- 梦迪，美团AI平台NLP中心知识图谱算法工程师，此前在金融科技公司文因互联任高级工程师及开放数据负责人，前清华大学知识工程实验室研究助理，中文开放知识图谱联盟OpenKG联合发起人。

招聘信息

美团点评 NLP 团队招聘各类算法人才，Base 北京上海均可。NLP 中心使命是打造世界一流的自然语言处理核心技术和服务能力，依托 NLP（自然语言处理）、Deep Learning（深度学习）、Knowledge Graph（知识图谱）等技术，处理美团点评海量文本数据，打通餐饮、旅行、休闲娱乐等各个场景数据，构建美团点评知识图谱，搭建通用 NLP Service，为美团点评各项业务提供智能的文本语义理解服务。我们的团队既注重AI技术的落地，也开展中长期的NLP及知识图谱基础研究。目前项目及业务包括美团点评知识图谱、智能客服、语音语义搜索、文章评论语义理解、美团点评智能助理等。真正助力于“帮大家吃得更好，生活更好”企业使命的实现，优化用户的生活体验，改善和提升消费者的生活品质。欢迎各位朋友推荐或自荐至 hr.ai@meituan.com。

算法岗： [NLP算法工程师/专家/研究员](#)、[知识图谱算法工程师/专家/研究员](#)

工程岗： [C++/Java研发专家/工程师](#)、[AI平台研发工程师/专家](#)

产品岗： [AI产品经理/专家](#) (NLP、数据方向)

美团在O2O场景下的广告营销

作者: 亦平

美团作为中国最大的在线本地生活服务平台，覆盖了餐饮、酒店、旅行、休闲娱乐、外卖配送等方方面面生活场景，连接了数亿用户和数百万商户。如何帮助本地商户开展在线营销，使得他们能快速有效地触达目标用户群体提升经营效率，是美团的核心问题之一，而机器学习相关技术在本地在线营销场景下发挥着非常关键作用。

本文将从5个方面来介绍。首先，介绍O2O场景下广告业务的特点，及其与B2B和B2C广告业务的差别；其次，从商户效果感知、用户体验和媒体平台收益三个维度，介绍O2O广告业务的最重要的考量指标；第三，从前两节阐述的业务特点和考量指标出发，介绍O2O场景下在线广告营销的机制设计；第四，介绍O2O特有的实时场景化下的推送广告；最后，简要介绍O2O广告系统相关的工具。

O2O场景下的广告业务特点

在O2O业务模式和相关的平台崛起之前，大品牌的商家由于收入规模大、营销费用充分，为了提升品牌知名度依靠如下的途径开展宣传：传统媒体（如电视、广播、报纸）；互联网流量（如传统搜索引擎、门户网站等）；户外广告（如公交地铁车身、广告牌、灯箱等）。通过上述媒介，商户能快速接触大量的用户，宣传品牌形象。上述营销方式也有其局限。首先广告投放的资金门槛较高，营销预算有限的商户无法承担相关费用；其次对于以直接效果为导向的商家来说，上述投放形式过于粗放且无法形成直接的购买转化效果闭环。对于大多数的中小商家来说，他们营销预算有限且更加注重直接的购买转化，获取潜在客户的主要途径是散发传单、派发礼物、沿街喇叭广告。但是，这些线下的营销手段覆盖到的潜在消费者较为有限，并且这些方式无法长期持续开展。

以美团为代表的O2O本地生活服务平台快速成长壮大，逐渐成为广大本地服务商户在线营销的最重要的手段之一。美团平台上聚集了上亿的消费者，他们使用平台寻找商家、查询优惠信息、浏览评论。对于商户而言，他们是最直接的潜在消费者。通过在美团平台开展在线营销，商户能够获得更多的展示机会吸引更多客户到店消费。借助于便利的在线咨询、预定和支付手段，平台上的广告业务可以形成了效果闭环，商户能清晰准确掌握广告投放的效果并以此优化广告投放策略。

对于美团而言，平台可以基于对用户大数据的挖掘和分析，在由时间、地点、用户和关系构成的特定场景下，连接用户线上和线下行为，理解并判断用户情感、态度和需求，为用户提供实时、定向、创意的信息和内容服务。O2O场景下的在线营销广告相较于传统的B2C和B2B商业模式下的广告有其独特属性，独特性主要体现在移动化、本地化、场景化以及多样性4个维度。

随着宽带无线接入技术和移动终端技术的飞速发展，人们逐渐开始使用手机等移动设备随时随地从互联网上获取信息和服务。在这个时代，无论是新闻阅读、社交通信还是电子购物，人们都习惯于通过手机应用来直接满足自身的需求。事实上，美团在移动互联网发展的初期就主动适应了这一历史潮流，大力发展战略服务能力，目前已经有超过90%的交易行为是通过移动互联网服务达成的。O2O广告作为连接人和服务的本地化生活服务营销模式，它有鲜明的移动化和本地化的特点。

- **移动化。**它主要体现在精确性、即时性和互动性三个方面。通过移动设备的传感器，我们能精确了解用户所处的地理位置，推送更加精准的广告。绝大多数用户随时都把手机带在身边，所以广告信息能及时推送给用户。功能强大的各种移动应用，为广告提供了多种互动可能性，例如在美团App上，用户直接可以完成推广商户的信息查询、排队和交易。
- **本地化。**以转化效果为导向的O2O广告营销，营销的目标用户是提供服务的本地商户附近的人群。在淘宝上，一双皮鞋可以对全国的用户开展推广和售卖，无论消费者在何地，物流和快递都会准确地把货物送达消费者的手中。而在美团上，一家在五道口的火锅店最佳的推广对象是五道口附近的食客，这些食客才最有可能直接到火锅店来消费。实际上，通过观测实际的交易数据，我们发现超过90%的交易中用户和商户的距离小于3公里。营销活动要取得好的效果，必须针对性地选择目标群体，在O2O广告中目标群体就是本地化的用户人群。移动设备的精确定位为商户发现目标人群提供了保证。
- **场景化。**消费者、移动设备、时间、空间构成了用户消费需求的精准场景。PC时代，用户的标识以Cookie为载体，但Cookie极易清除，同时一台电脑可能会被多人使用，这导致用户信息很难有效串联，连受众年龄、居住地等基础信息都无法准确把握。而在移动互联网时代一机一人的模式下，通过分析和挖掘用户在平台上留下的各种行为足迹，我们能对用户方方面面的属性和偏好进行解析和重构，产出十分精准的用户画像。在了解用户的地理位置、消费意图和行为轨迹等用户信息前提下，O2O广告营销能在由时间、地点、用户和需求构成的特定场景下，为用户提供实时、定向和富有创意的营销内容，连接用户线上和线下的行为。例如在一个阳光明媚的下午，对一个在CBD上班并有喝下午茶习惯的白领，平台可以适时地推送下午茶或者咖啡店商户。
- **多样性。**O2O商业模式面对的是各式各样的本地生活服务业务，不同的业务有着不同的特点，并对O2O广告营销也提出了不同的需求。举个简单的例子，不同的服务业务对目标用户的本地性要求也大相径庭：餐饮类服务对距离比较敏感，这一类服务商家的目标用户群体是商户周边的食客；婚纱摄影类服务对距离就没那么敏感了，这一类服务商家的目标群体是全城的新婚夫妇。

商户、用户和平台三者利益平衡

广告系统和搜索系统、推荐系统，有着十分相似的系统架构：它们大都采用了检索加排序的流程体系。基于这一点，有很多人认为广告业务和搜索推荐业务没有区别。实际上，广告业务有其独特规律。广告首先是一项商业活动，它的出现远远早于互联网。作为一种商业活动，商户、消费者和媒体平台三者的利益都要被重视和考虑，这些利益指标是广告业务得以可持续健康发展的启明灯。本节将从商业活动的角度出发，分析美团O2O广告营销中的商户感知、用户体验和平台收益这三项重要指标。

商户效果感知

商户在美团广告平台上进行广告营销的根本目的，是通过美团触达更多的潜在消费者，获得最大的增量利益。本地生活服务类型的商家的成本可以分为两个部分：变动成本和固定成本。变动成本是随着业务量变动而线性变动的成本，主要来自原料消耗。而固定成本是在一定时期内是不会随着业务量的改变而改变的成本，如门面装修的投入、店铺的租金、店铺服务人员的基本工资等。商户如果没有足够的业务量，不能招揽足够多的消费者，则单位业务量的成本会居高不下，导致严重亏损。因此，对于餐饮行业，商家的

首要目标是提升翻桌率、减小空座率，而对于酒店行业，商家的首要目标是提升满房率、减少空房情况。固定成本的存在是本地商户开展O2O广告营销的基本前提。

从商户的角度出发，O2O广告营销的效果可以从三个维度来衡量：广告的可见性、广告带来的线上增量收益和广告带来的整体增量收益。

对于商户来说，可见性是最初步且最直接的营销结果，是商户得到的最快的效果反馈。广告的可见性表示商户的营销信息已经开始通过媒体平台去触达潜在的消费群体。因此，稳定可靠的广告展现预期是赢得广大商户对O2O广告营销信任的最基础要求。

广告带来的线上增量收益是指通过在美团等媒体平台上的广告投放带来的线上收益。这一部分收益可以分为两类：一类是直接的在线订单带来的收入，例如团购、酒店预订等；另一类则是在线预约等非直接交易带来的收益。对这一部分收益，平台方能给出准确的统计、分析并反馈给广告商户。对于外卖、婚纱摄影和酒店旅游等对线上流量、线上交易依附度很高的行业来说，它们的线上收益占整体收益的比例非常大，这个比例直接反映了商户的经营活动的状况。

除了直接使用在线交易，用户使用美团的另一种场景是通过平台查看商户菜品、评价和地理位置等信息，然后直接到店进行消费。广告给商户带来的整体增量收益即包含了这部分离线客户引流带来的收益。餐饮类的商户的线上交易只占门店整体收益的一小部分，因此，对广告效果的衡量需要综合考虑在线和离线两部分收益。离线引流部分收益相对于在线交易收益较难准确统计，但是平台可以通过用户的实时地理位置准确统计部分用户的到店情况，或者通过曝光、点击至到店的数据漏斗模型对到店数据进行估计。未来随着电子化支付方式的普及，平台将能更好地对商户的整体收益进行统计。

了解了O2O广告营销效果的主要衡量指标后，要确定商户的广告投放成本是否真的较低，需要用到投入产出比 (Return over Investment, ROI) 这一常用的评价指标，即某次广告活动的总产出与总投入的比例。对应于两种广告收益指标，ROI也可以分为在线支付ROI和整体ROI：在线支付ROI等于在线增量交易额除以广告费消耗，整体支付ROI等于整体门店收入增量除以广告费消耗。在广告费预算有限的情况下，商户总是寻求优化广告投放，提升ROI。

用户体验

有效地保障用户体验，是美团开展O2O广告营销的基本前提条件。平台只有保障用户体验、对用户有用，它的价值才能得到体现。美团通过让更多的用户留存并活跃在平台上，才能吸引更多的本地生活服务商户来进行广告投放，才能生成更大的流量用以广告变现。

美团主要从短期和长期两个维度来进行用户体验指标的设计和度量。从信息曝光、用户点击和用户交易这个用户行为漏斗出发，短期用户体验指标主要考虑了点击和交易情况。第一个短期用户体验指标是点击率 (Click through Rate, CTR)，其数学表达为点击次数 (Click) 除以曝光次数 (Impression)。点击率反映了给用户展示的商户信息的质量和相关性，与用户意向无关的、与用户所处时间地点场景不匹配的广告信息展示，不能满足用户的需求、吸引用户的点击，从而导致较低的点击率。点击率这一指标又细分为广告曝光的点击率和整体页面的点击率，前者度量了广告本身的优劣，后者反映了广告对整体信息呈现效果（自然结果加广告结果）的影响。劣质的广告除自身点击率较低之外，还会搅扰用户整体浏览行为，使得用户不能愉悦获取需要的本地生活服务信息。

CTR=点击次数/曝光次数

为了获得真实的曝光，一般会在移动端进行埋点监控每个POI在手机屏幕上实际展现的比例和时间，将超过一定展示比例和时间阈值的POI纳入曝光次数的统计。

第二个短期用户体验指标是转化率（Conversion Rate, CVR），其数学表达为交易次数（Order）除以点击次数（Click）。转化率同样反映了商户信息展示的相关性和质量，和用户需求不相匹配的商户展示将不能促成交易的达成，从而导致较低的转化率。和点击率指标类似，转化率指标亦可分为广告转化率和整体页面转化率。其中广告转化率还和商户的在线交易ROI成正比，准确有效的广告投放，不仅可以提升用户体验，还能提升商户的ROI。

CVR = 交易次数/点击次数

长期用户体验指标以更长的时间跨度为出发点，评价广告对用户的长期持续影响。长期用户体验指标主要包括回访率和复购率两个指标。回访率是一个反映用户长期留存的指标，其意义为一定时期内用户是否还会重新登录和使用美团平台。回访率指标包括周回访率、月回访率等。低质量的广告投放，搅扰了用户使用平台方便获取商户信息的感受和体验，使得用户脱离平台以致流失，从而导致回访率降低。复购率则反映了用户消费体验的指标，其意义是一定时期内用户是否会重新购买某一个商家的服务。同样，低质量的商户服务会损害了用户的消费体验，使得用户不再进行同样的消费，进而导致了复购率的下降。

为了准确衡量广告投放带来的用户体验影响，除了进行策略变更对比测试之外，平台会长期保留一小部分流量作为对照组，不对这部分用户开展广告投放，通过比较整体流量和对照组上相关用户体验指标的差异，来确定广告对用户体验的长期影响，进而督促和指导平台优化广告投放策略。

平台收益

美团作为媒体平台的目标是，在保障商户ROI和用户体验的情况下，优化流量变现效率，实现商户营销诉求和用户消费诉求的最佳连接。

前两节已经介绍了商户ROI和用户体验的基本概念。我们知道只有保障商户的ROI，才会有更多的商户、更多的预算进入到广告投放系统内；只有保障用户的体验，才会有更多的用户、更多的流量用于广告变现。这两者决定了广告业务这一块奶酪的大小。

流量变现效率衡量单位流量所能带来广告收益。对于展示广告业务，流量变现效率主要用千次广告展示收益（Revenue per Mille, RPM）来表示。对于搜索广告，流量变现效率主要用单次搜索广告收益（Revenue per Search, RPS）来表示。

广告收入=曝光次数×CTR×CPC 广告收入=广告主数×ARPU

从流量供给端来看，广告收入（Revenue）是广告曝光次数、点击率和点击单价（CPC）的乘积；从流量需求端来看，广告收入是广告主数量和每用户平均收入（Average Revenue Per User, ARPU）的乘积。在广告商户数、预算和流量情况稳定的前提条件下，流量变现效率的提高主要通过点击率和点击单价两个关键指标驱动，而这两个指标的良性提高依赖于广告投放的机制设计和投放算法，详细内容将在下文展开陈述。

O2O广告机制设计

前面介绍了美团O2O广告营销的特点，分析了商户、用户和平台三者的利益情况。本节将从上述特点和利益情况出发，阐述美团实际业务中O2O广告机制的设计原理，包括广告位设定、广告召回机制和广告排序机制。

广告位设定

在移动端，美团的自然结果以列表的样式进行信息呈现，而广告占用列表中的固定位置（区间浮动固定位置）进行展现。从商户的效果感知角度出发，固定位广告形式能给商户以较为确定的广告展现预期，使得商户有明确的竞价标的（即固定展现位置）。

广告位的设定，需要综合考虑和平衡商户、用户和平台三者的利益关系。过于密集的广告位置设计和广告展现会降低用户寻找商户信息的效率，影响用户使用体验。过于稀疏的广告位置设计导致广告展现机会过少，导致平台流量变现效率能力不足。头部广告位对用户体验影响较大，但是能获取更多的曝光，更有价值，更能激发商户的出价意愿。腰尾部广告位对用户体验影响较小，但是广告位曝光概率小，不能有效刺激商户出价。美团的实际广告位设定，一方面考虑了各个展位和业务的自身特点，另一方面通过A/B测试进行多种方案的比较和选择，最终选择能有效兼顾用户体验、商户效果和平台收入的设计方案。

广告召回机制

广告召回在技术上与搜索和推荐十分相似。搜索场景广告会使用用户的查询词去广告商户索引中去寻找匹配的商户，推荐场景广告会根据用户的意图、位置等场景信息去匹配合适的商户。

搜索广告匹配中，一项重要技术是查询改写。一方面，我们使用传统的自然语言处理方法，对查询进行有效分析（例如成分分析），完成同义和近义改写；另一方面，我们使用深度语义相似度神经网络模型（DSSM）和序列到序列模型（Sequence to Sequence）进行查询的改写，进一步提升广告匹配的覆盖率和准确性。

针对O2O商业模式的特点和广告业务各方的利益，广告召回机制在传统搜索推荐召回机制基础上进行了优化改进。我们在召回中引入了逐层召回的理念，各层依次设置由紧到松的相关性水准（Match Level）控制召回广告的质量，在当前相关性水准已经召回足够数量广告候选的情况下，不再进行后续召回。相关性水准考虑多种相关性因素：查询匹配模式、距离和星级等。例如针对Query匹配模式，广告召回时会优先使用Query精确匹配模式召回，其次选择模糊匹配模式，最后才尝试采用语义匹配模式。针对距离因素，广告召回会优先召回距离3公里内的商户，其次选择5公里内的商户，最后尝试全城召回。

相关性水准的设置应该充分考虑到不同O2O业务的特点。例如距离的设置上，对于餐饮类流量，系统会优先召回3公里内的商户，而对于距离相对不敏感的婚纱摄影类流量，系统则会放宽限制，优先召回10公里内的商户，或者直接采用全城召回策略。

广告排序机制

和传统的搜索广告业务一样，美团的广告是按点击计费（Cost Per Click, CPC）广告，广告主依据广告的点击价值进行出价（bid），广告系统按照RankScore（RankScore为出价和广告质量度的乘积）进行

广告排序。在广告系统中，广告质量度一般用广告的预估点击率来衡量。

RankScore=bid×CTR

广告按照RankScore排序后，会依据广义第二价格（Generalized Second Price）进行计费。

$$\text{charge}_i = (\text{CTR}_{i+1} \times \text{bid}_{i+1}) / \text{CTR}_i$$

由此可知，准确的预测广告的点击率是保障广告收入和用户体验的前提。广告点击率预估问题是一个典型的监督机器学习问题，它的目标是在给定广告商户、用户和查询上下文的前提下准确预测点击行为发生的概率。这个监督学习问题的特征我们用 x 表示，目标用 $y \in \{1, -1\}$ 表示（广告曝光后获得点击为1，否则为-1）。通过收集线上的广告曝光和点击日志，我们可以获得大量的标注样本 $\{(x_i, y_i)\}$ 作为监督学习的训练数据。

我们使用参数模型拟合这个概率：

$$\text{CTR} = \text{Probability}(\text{Click}|\text{AD, User, Query}) = f(x, w) = \frac{1}{1 + e^{-\phi(x, w)}}$$

其中， w 监督学习问题即是搜索 w 使得目标损失函数最小的一个优化问题：

$$L(y, f(x, w)) = \log(1 + e^{-y\phi(x, w)})$$

$$\min_w \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i, w)) \quad (\text{公式 1})$$

$$\min_w \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i, w)) + R(w) \quad (\text{公式 2})$$

其中， $L(y, f(x, w))$ 是模型的损失函数，在点击率预估问题中一般使用负Log似然函数（Negative Log-Likelihood）作为损失函数。优化问题（公式1）是原始的点击率预估问题，优化问题（公式2）引入了正则项 $R(w)$ ，用以控制模型的复杂度，防止模型过拟合。此外当我们选择L1范数作为正则项的时候，我们能获得稀疏解，缩减模型大小，进而减少线上服务加载模型的内存需求，提升模型的预测速度。下面我们简要介绍几种常用的点击率预估模型。

逻辑回归模型

$$\phi(x, w) = w^T x$$

逻辑回归模型是广泛应用的点击率预估模型，它是一种线性模型，相应的优化问题有非常好的性质。它是一个无约束的凸优化问题，有全局唯一的最优解。它支持大规模的特征，通过常用的梯度下降方法能较快收敛。

到最优解。逻辑回归模型的可解释性十分优良，通过特征对应的权重我们能很好分析各个特征的重要性以及它们对点击率的影响关系。

逻辑回归也有它的劣势：首先，作为线性模型，它的表达能力相对较弱，需要通过大量的特征工程工作（例如，特征组合）来弥补和提升模型的表达能力；其次，它需要进行大量的特征预处理工作，例如特征归一化、离散化等。

逻辑回归作为基础模型和其他模型相结合，扬长避短，充分发挥其作用。例如逻辑回归和梯度提升决策树结合，通过梯度提升决策树解决特征离散化和特征组合问题，并充分发挥逻辑回归对大规模特征的支持和良好的优化问题性质。

因子分解机FM模型和场感知因子分解机FFM模型

$$\text{FM: } \phi(x, w) = \sum_{i,j} \langle w_i, w_j \rangle x_i x_j$$

$$\text{FFM: } \phi(x, w) = \sum_{i,j} \langle w_{i,f_j}, w_{j,f_i} \rangle x_i x_j$$

FM模型和FFM模型是非线性模型，它们对特征进行两两组合，提升了模型的表达能力。此外FM和FFM模型都对特征进行向量化的表达和学习 (w_i, w_{i,f_j}) 提升模型的泛化能力。FFM相对于FM引入了域的概念，在FM中特征*i*和其他特征组合用的是同一个向量表示，而在FFM中特征*i*和不同域的特征组合会使用不同的向量表示，进一步提升了模型的复杂度和表达力。

人工神经网络（Artificial Neural Network, ANN）

近几年神经网络模型强势复兴，以深度神经网络为代表的方法，在图像识别、语音识别以及自然语言处理等领域超越传统浅模型，取得了突破性的进展。在点击率预估这个任务上，最近也涌现出一批深度神经网络模型，取得了明显的效果，其中典型的模型是Wide & Deep模型。

Wide & Deep模型包含Wide和Deep两个部分。Wide部分可以类比逻辑回归模型，能对相关特征的作用进行很好的记忆。Deep部分类似FM模型和FFM模型，它们都对相关特征进行了向量化的表示（Embedding）和学习，但是Deep部分通过复杂的网络结构可以表达更复杂的特征交互和组合关系，提供了更好的泛化能力和表达力。

$$w_t = w_{t-1} + \eta \frac{1}{N} \sum_{i=1}^N \frac{\partial L(y_i, f(x_i, w))}{\partial w} \quad (\text{公式 3})$$

$$w_t = w_{t-1} + \eta \frac{1}{b} \sum_{i=1}^b \frac{\partial L(y_i, f(x_i, w))}{\partial w} \quad (\text{公式 4})$$

梯度方法是模型的优化（优化问题的求解）的基础方法。公式3是使用标准梯度方法求解点击率预估优化问题公式1的迭代步骤。在点击率预估问题中，由于训练样本数量庞大（十亿、百亿），直接应用公式3计算量巨大，迭代速度受限。因此在点击率预估问题中，我们一般使用随机梯度下降法（Stochastic Gradient Descent, SGD）进行问题求解。在SGD方法（公式4）中，我们使用一小部分样本上的梯度对整体优化目标的梯度进行近似估计，加快参数迭代速度。其中b是样本集合大小，当b=1时，我们用单样本的梯度来近似整体目标函数的梯度值。

在美团我们使用参数服务器框架（Parameter Server）实现模型并行和数据并行，以解决包含大规模训练数据和特征的复杂模型求解问题，如图1所示：

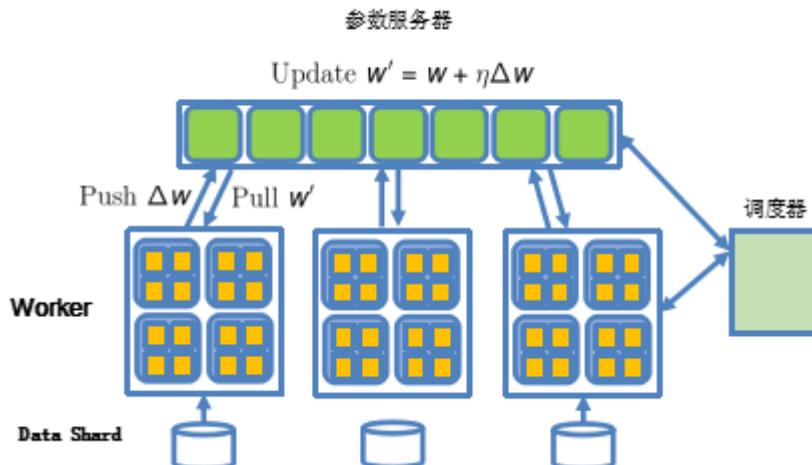


图1 参数服务器框架

特征工程方面，点击率预估特征主要从广告、用户和查询这三个方面来挖掘和刻画广告展现场景（见图2），特征需要包含影响点击率的方方面面，是模型成败的重要因素。特征的选取需要从业务场景出发。在O2O场景下，一个影响点击率的重要特征就是商户和用户之间的距离。



图2 点击率预估的特征

O2O推送广告

在O2O场景下，除了搜索推荐广告，推送广告也非常重要。推送广告就是媒体在合适的时机将合适的广告以消息的形式推送给合适的人群。推送广告的主要目标是：提升用户活跃度、实现人群精准触达。美团注册用户有3.5亿，但是日活跃用户只有3000万，年活跃用户也只有1亿，还有很大一部分用户平时不登录美团App，或者登录次数很少。给这部分用户推送广告，引导用户打开App，有助于提高用户活跃度。另一方面，推送广告通过丰富的人群定向，实现精准投放。要实现精准触达，需要做到两点：有完整的用户画像，用户画像包含属性标签、偏好标签和行为标签，我们以此来判断用户对广告的兴趣；智能匹配技术，将广告精准定位到合适的用户上。

推送广告的特点是：主动触达、用户意图不明确。理论上来说，推送广告可以在任意时间给任意用户推送任意广告，而搜索广告只能在用户搜索或者筛选的时候给用户展示广告。但推送广告的劣势是用户意图不明确，而搜索广告具有搜索词或者明确的筛选条件，这些都是明确的用户意图。所以相比搜索广告，推送广告更需要精准的受众定向。

受众定向

常用的定向方式有下面几种：

- **时间定向**。时间定向能够让品牌根据消费者行为、营业时间，甚至是季节性活动或特殊事件来进行广告投放。举个例子，美发沙龙只有白天营业，如果定向时间包括了晚上非营业时段，那么用户在晚上非营业时间看到广告后，无法通过打电话来预约，这样就没有转化。
- **重定向**。它指根据用户的历史行为，将曾在商家发生过浏览、收藏、购买等行为的用户作为商家的精准定向人群，进行广告推送，拉回用户完成转化。通常情况下，消费者不会看过就能记住，你需要重定向。根据重定向推送广告是通过视觉方式提醒消费者有关商家产品信息的好方式。消费者看到后可能会想：“啊，我忘了要买这双鞋……”而这种面包屑式的提醒方式往往能够诱使他们点击并购买。重定向方式是所有定向方式中最精准、投资回报率最高的。
- **地理位置类定向**。它指的是根据用户实时地理位置（一般是蜂窝信息或者GPS经纬度）做一些定向，有助于帮助商家触达那些正在前往商家所在区域的消费者，包括距离定向、商圈定向等。这种定向方式在移动设备上投放广告时有着非常重要的作用。比如本地的一家美发沙龙，想要招揽本地生意，那么就可以使用地理位置定向技术在特定半径内进行宣传。如果这家店在三角区内有发廊的特许经营权，那么它就可以使用该技术进行一个以上定位。当然针对各区域进行定向的时候，商家可以根据区域内业务发展状况调整出价。
- **人口属性定向**。人口属性标签包括性别、年龄、收入水平、婚姻状况、是否有车、是否有小孩等。通过人口属性标签，可以将广告推送给相关消费者，就是可能购买的人群，具体选择什么标签主要取决于商家销售的是什么产品。例如婚纱摄影类商家会选择婚姻状况标签为“未婚”的人群进行广告投放，美甲美睫类商家会选择性别标签为“女性”的人群进行广告投放。这些标签里面，性别、年龄这种标签比较容易得到，因为用户注册的时候就提供了相关信息；而收入水平这种标签需要通过预估得到。使

用人口属性定向的时候，标签既不能过于笼统，也不能太过细分。例如美甲美睫商家选择年龄标签的时候，一方面，不能选择0~60岁，这种人群太泛了，低年龄段和高年龄段人群可能没有很强烈的美甲美睫需求。另一方面，也要防止对人群太过细分，例如，尽管最终可能需要选择一个更细化的年龄标签，但却不能仅仅定位为一个具体年龄，若只选择22岁的人群，这可能导致人群覆盖不完整。时刻记住目标受众，但在定位的时候得找到一个折中的办法。

- **行为定向。**它是从用户的行为数据中挖掘用户兴趣偏好，从而推送相应的广告。行为数据包括频道、商家详情页、团单详情页的浏览和点击，用户评论和打分等。兴趣偏好一般分为长期、短期和实时偏好。当我们挖掘用户长期偏好的时候，使用的是“一段时间内的行为”，需要对不同时间的行为计算不同的权重，因为用户的兴趣是动态变化的，三个月前用户商圈偏好是A，可能现在搬家了商圈偏好变成了B。为了衡量不同时间行为权重，将行为累计控制在一段时间内，一般使用滑动窗口法和时间衰减法。兴趣偏好包括品类偏好、价格偏好、商圈偏好等。品类偏好指的是用户偏好的商品和服务的品类，例如，用户喜欢吃川湘菜还是江浙菜，喜欢火锅还是自助餐；价格偏好指的是用户的消费水平，例如，用户点外卖价格区间是偏好0~20元价位还是偏好20~40元价位。
- **新客推荐。**即Look-alike，以广告主的老顾客作为种子信息，结合广告平台的大数据，寻找出老顾客具有的某种特征或规律，为广告主找到具有相同特征或规律的潜在顾客。这种方式可在保证精准定向效果的同时，扩大用户覆盖面。例如一个川菜馆广告主想投放广告，目标人群除了在本店消费过的顾客外，还可以选择在别的川菜馆或者湘菜馆消费过的顾客，因为他们可能口味差不多。

用户在平台上的搜索、浏览、收藏、购买等行为会被记录下来，形成用户日志。通过对用户日志的分析和挖掘得到用户画像，包括用户基本属性、兴趣偏好、行为标签等。广告定向是广告和用户匹配的过程，为每个广告找到适合的受众群体。广告投放后，需要统计定向效果，包括定向精准程度和覆盖率。精准定向广告的运作流程如图3所示。

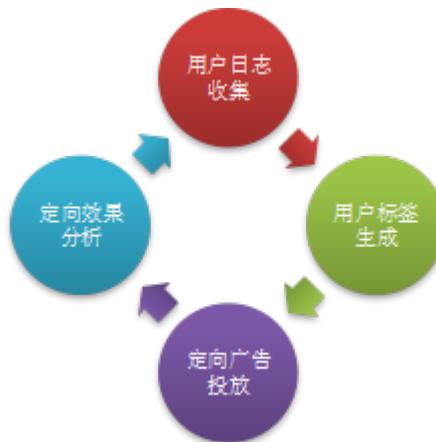


图3 精准定向广告流程

为了实现广告和用户的匹配，首先需要为广告的受众人群进行初步假设，即确定广告感兴趣的人群，并将其与用户画像标签映射起来，这一步要靠产品调研和分析得到。然后根据这个初步假设确定广告投放的定向条件，匹配到符合条件的人群。

- 单个定向条件的表示。每一个定向条件都用一个对来表达。比如，职业分为学生和白领两种，定向条件为学生时表示为 <“Professional”，“Student”>。

- 组合定向条件的表示。广告主设定的定向条件组合往往非常复杂，是各种不同定向条件的组合，涉及交、并、取反等操作。我们采用析取范式（Disjunctive Normal Form, DNF）的形式来存储广告的定向条件。下面以几个例子来说明DNF的表达方式。

- DNF1: (30岁 男性) \cup (25岁 女性)
- DNF2: (广东人 广东男性) \cup (北京人 北京品牌新客)
- DNF3: (非男性) \cup (男性 实时位置在店铺周围2公里范围) \cup (喜欢美食的)

在这样的表达形式中，有两点需要说明：第一，每个DNF可以分解成一个或者多个合取范式（Conjunction Normal Form, CNF）， $DNF1 = C1 \cup C2$ ，其中， $C1 = (30岁 男性)$ ， $C2 = (25岁 女性)$ ；第二，每个CNF可以分解成一个或者多个条件的交。上例中的 $C1 = A1 \cap A2$ ，其中 $A1 = 30岁$ ， $A2 = 男性$ 。

- **定向条件匹配。**定向匹配过程如图4所示。一个定向请求包括用户ID和广告投放ID，首先根据用户ID去取用户标签，根据广告投放ID取定向包，将定向包解析表达成DNF的形式，然后与用户标签进行匹配。
- **定向效果评价。**定向效果一般从质和量两个方面进行评估。质指的是精准程度，主要指标是点击率和转化率。量指的是覆盖程度，主要指标是用户覆盖率、广告主使用率以及定向方式对应的流量占比。

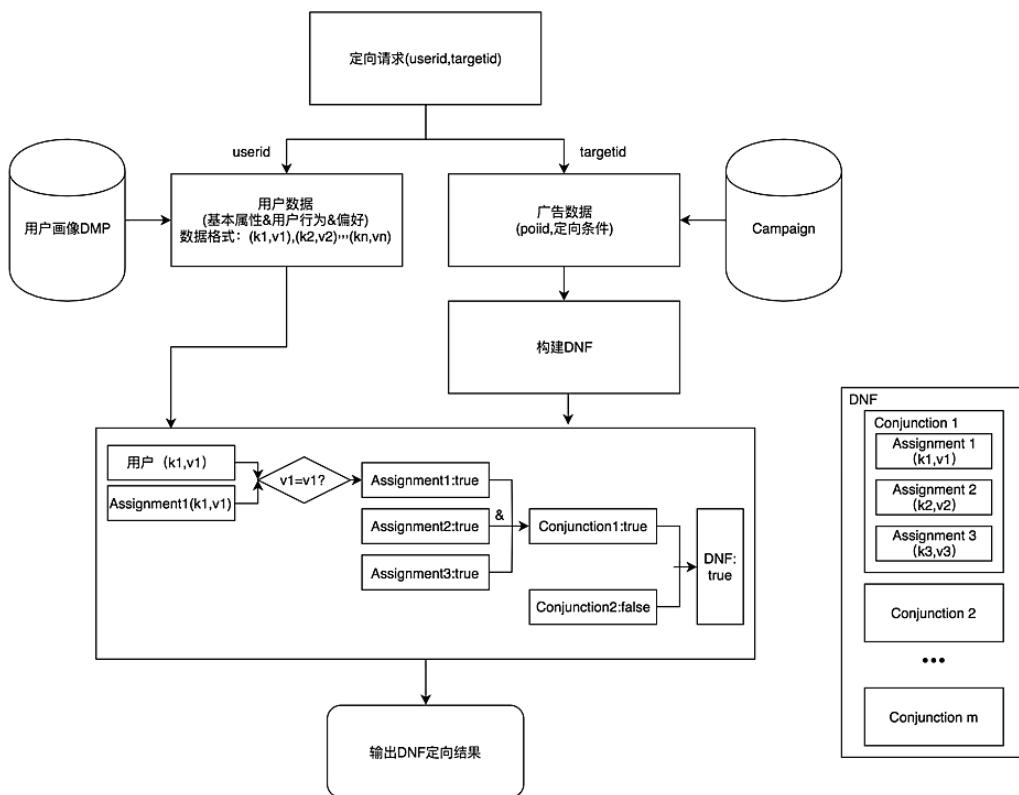


图4 定向匹配过程

在美团推送广告中，重定向方式点击率和转化率最好，但覆盖率最低；地理位置定向和人口属性标签拥有更广泛的人群，效果相对较差。实际采用哪种定向，需要看广告主的推广需求，广告主需要综合考虑精准程度和覆盖率的平衡。

O2O广告系统工具

“工欲善其事必先利其器”，有效的工具是一个优秀高效的广告生态的重要组成部分。本节我们从面向开发人员、面向广告主和运营人员两个角度进行简述。

面向开发人员的系统工具

面向开发者的工具主要包含三个方面：离线数据分析工具、实时数据分析工具以及在线广告系统调试工具。

离线数据分析工具支持从各个维度（广告位、广告类型、时间、区域、算法策略等）统计广告业务的各项关键指标（召回率、点击率、转化率、RPS/RPM、CPC等），检视广告系统的短板和漏洞，帮助广告算法和工程团队发现问题和寻找潜力。

实时数据分析工具从时效性角度弥补离线数据分析的缺点，帮助开发者尽早发现数据异常，更快地响应和修复问题。在背后支撑这些分析方法的是Hive、Spark、Elasticsearch和Druid等大数据处理工具。如图5所示是实时消耗数据分析工具。



图5 实时消耗数据分析工具

在线广告系统调试工具是针对单个广告主或单个查询等具体问题的排查。通过调试工具可方便构造模拟请求并查看单个服务处理详细信息，收集各个广告流程步骤（召回、排序和创意优选等）的信息，跟踪和定位线上实时环境中各步骤的问题。除了线上问题排查之外，调试工具也是开发阶段用于验证策略效果和算法正确性不可或缺的手段。如图6所示是在线广告调试工具的基本界面。

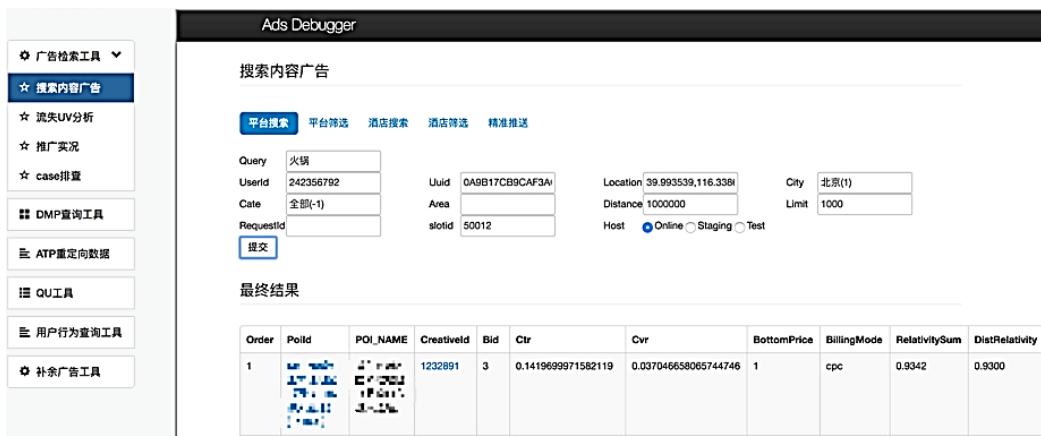


图6 在线广告调试工具界面

面向广告主和运营人员的工具

面向广告主和运营人员的工具包括广告主出价预估和排名预估、商户效果漏斗分析、账户诊断等相关工具。面向广告主的工具帮助广告主更好地衡量和感知广告效果，让其了解市场竞争情况，协助其有效主动地优化广告投放效果。面向运营人员的工具能让运营人员对广告主的投放情况有更清晰的了解，进而帮助其更好地指导和服务广告主。

1. 效果漏斗分析工具

如前文所述，O2O广告从在线展示到用户进店消费需要经过点击和转化多个流程，为了帮助广告主优化整体投放效果，我们在推广后台提供了效果漏斗分析工具。效果漏斗分析工具主要包括曝光/访问量/感兴趣/到店三层漏斗，同时给出相应的问题诊断和优化建议，如图7所示：



图7 效果漏斗分析工具

2. 推广实况工具

感知广告展示位置以及竞价实况是投放中的广告主核心需求之一。但个性化智能排序技术体系和带有地理位置限制属性的O2O广告场景下，由于用户个性标签、地理位置等原因会导致广告主看不到自己投放中的广告在客户端曝光，广告主难以分析原因，也不知道如何优化现有的广告投放。

推广实况工具提供查看排名、模拟出价和诊断优化功能。广告主可以查看选定商圈、类目、地理位置等特定条件下的实时排名，也可以查看去个性化后一般情况下的平均排名。同时工具对于广告展现位次过低或得不到展示的情况会给出具体的原因和相应提示。广告主可根据提示调整投放设置，比如对出价过低导致排名靠后情况建议调高出价，通过工具可以实时查看调整后的新排名情况，如图8所示：

The screenshot shows the 'Promotion Status' tool interface. At the top, there are two sections: '必选字段:' (Required Fields) and '可选字段:' (Optional Fields). Under '必选字段:', 'poiid' is set to '159839642' and 'query' is set to '东来顺火锅'. Under '可选字段:', 'cityid' is set to 'sinai获取(0)', 'areaid' is set to '朝阳区(14)', 'location' is set to '39.993539,116.3386', and 'cate' is set to 'sinai获取(0)'. Below these fields are 'launchid' (set to '5580474'), 'bid' (set to '1.5'), and a 'Host' dropdown with 'Online' selected. A '提交' (Submit) button is at the bottom left. The second section, '诊断原因' (Diagnosis Reason), shows a table with one row for target '5580474' with status '满足ATP时间定向' (Meets ATP time targeting). The table has columns: target, campaign status, and atp_status.

target	campaign status	atp_status
5580474		满足ATP时间定向

图8 推广实况工具

3. 流失订单分析工具

流失订单分析工具基于门店流失订单记录提供对比分析功能。流失订单是指最近一周内用户对商家A进行了点击，但实际去B、C商家下单的流量算作A的流失订单。分析工具根据用户的点击下单行为数据帮商家分析自身与用户最终下单商家之间的差距在哪里。比如对酒店商家，工具会提供商家平均房价、平均评分、商家首图等信息对比，广告主从中可以分析出订单流失原因，如图9所示：

最近一周的流失分析结果									
分析项		酒店自身信息	流失对比结果		订单流失到的酒店信息				
所属类型	指标	酒店	差距对比	结果反馈	酒店	酒店	酒店	酒店	酒店
基础信息	美团poi	9970	---	---	487%	5924	4175	160737	6694
	名称	双流国际机场店(双流国际机场店)	---	---	法式主题酒店	海景酒店	商务酒店	商务酒店	酒店
	商圈1	NULL	---	---	双流机场	双流机场	NULL	双流机场	NULL
	流失订单	0	-31	✖危险:近7天浏览过本店的客人,最终在竞争商家有-31个下单	8	8	7	4	4
	流失uv	0	-5	---	1	1	1	1	1
	物理库存	45	-20	---	90	98	38	21	80
	历史消费人数	5371	2379	✓真棒! 您此项指标优于其他酒店,请继续加油。	2528	10216	412	234	1570
	支付订单量(周)	84	7	---	86	175	24	44	56
	支付夜间量(周)	93	4	---	105	200	24	52	66
	支付订单金额(周)	17792	-632.88	---	28749	43919.42	2730	6616	10110
酒店定性	酒店类型	[商务酒店]	---	✖危险:您的订单流失到了主题,情侣,经济型类型酒店,请正确定位自家酒店类型。	[主题酒店,情侣酒店]	[商务酒店]	[经济型]	[经济型]	[商务酒店]
图片	头图		---	---					
	图片量	49	10	✓真棒! 您此项指标优于其他酒店,请继续加油。	68	38	37	21	35
	房型平均图片数	2.875	0.65	✓真棒! 您此项指标优于其他酒店,请继续加油。	1.3077	2.5714	4.2	1.75	1.3

图9 流失订单分析工具

4. 广告收益模拟器

为吸引潜在新广告客户入驻，广告收益模拟器为商家提供广告收益预估功能。该工具基于门店的历史非广告时期点击转化率、门店所在商圈流量以及竞争对手状态等信息，预估门店投放广告后能够带来的新增流量和订单量，帮助新客户快速了解广告产品，建立投资回报预期。同时通过该工具，商户可以方便地跳转到推广通平台进行注册和投放。此外，模拟器也可协助销售人员对商圈流量以及商圈可承载广告数进行预估，让销售人员更有针对性地开拓市场，提升新签成功率，如图10所示：



图10 广告收益模拟器

总结

本章从O2O广告的特性出发，介绍了O2O广告利益相关方的主要关注指标。O2O广告是美团的核心问题之一。本章着重介绍了如何应用机器学习方法提升广告投放的效果和效率，以及本地场景化的推送广告。此外，还简要展示了O2O广告平台相关的工具。

作者简介

- 亦平，2013年6月加入美团，目前负责美团搜索广告算法策略，曾负责美团搜索排序工作。

招聘信息

美团广告平台全面负责美团到店餐饮、到店综合（结婚、丽人、休闲娱乐、学习培训、亲子、家装）、酒店旅游的商业变现。搜索广告基于数亿用户、数百万商家和数千万订单的真实数据做挖掘，在变现的同时确保用户体验和商家利益。欢迎有意向的同学加入搜索广告算法组。简历请投递至：
zhouyayue#meituan.com

参考文献

- [1] Huang, P.-S., He, X., Gao, J., Deng, L., Acero, A., & Heck, L. (2013). Learning deep structured semantic models for web search using clickthrough data. In Proceedings of the 22nd ACM international conference on Conference on information & knowledge management (pp. 2333–2338).
- [2] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104–3112).
- [3] Juan, Y., Zhuang, Y., Chin, W.-S., & Lin, C.-J. (2016). Field-aware Factorization Machines for CTR Prediction. In Proceedings of the 10th ACM Conference on Recommender Systems (pp. 43–50).
- [4] Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., . others. (2016). Wide & deep learning for recommender systems. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems (pp. 7–10).
- [5] Li, M., Zhou, L., Yang, Z., Li, A., Xia, F., Andersen, D. G., & Smola, A. (2013). Parameter server for distributed machine learning. In Big Learning NIPS Workshop (Vol. 1).
- [6] Edelman, B., Ostrovsky, M., & Schwarz, M. (2005). Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords.

美团外卖骑手背后的AI技术

作者: 何仁清

背景

随着数字化时代的到来，外卖市场近年来发展非常迅猛。对外卖物流系统而言，配送效率和用户体验至关重要。而实际配送过程是由配送员（骑手）最终完成的，因此，想要真正提升配送效率，不但要在智能调度系统（订单指派、路径规划、ETA）上下功夫，还要不断提升配送员的“附加”能力，让他们越送越“熟”，越送越“顺”，越送越“快”。以此为出发点，美团点评研发团队设计了骑手智能助手，全面提升骑手的各方面能力。

在1月份的AICon全球人工智能与机器学习技术大会上，美团点评配送人工智能方向负责人何仁清分享了《美团骑手智能助手的技术与实践》。讲解如何在使用环境复杂、用户群体多元化的情况下，以智能耳机和语音交互为载体，并通过大数据挖掘、机器学习、自然语言处理等技术，让智能助手具备复杂场景精准识别、服务智能推送，智能引导、全语音操作等能力。最终在智能、安全、便捷、精准等多个维度上，全面提升骑手配送能力，从而提升整个配送效率和用户体验。以下系演讲内容整理：

AI技术对同城配送的业务价值

总体而言，物流业务是一个比较传统的行业，但是随着整个电商、移动互联网和移动支付的兴起，近些年整个物流行业实现了持续和高速的发展。



DataFrom : 中国物流与采购联合会 《中国物流业2016年发展回顾与2017年展望》

AiCon
全球人工智能与机器学习技术大会

主办: Geekbang InfoQ

物流行业

上图系中国物流与采购联合会在2016年发布的一个报告，调研数据表明，全国物流件数环比增长超过50%，达到300多亿件。

同时整个物流的费用占比也很高，从图中可以看到，物流成本已经占据GDP的15%。而在欧美国家以及日本，这个比例大概只有8%~9%左右，所以中国的物流行业还有很大的优化空间。这也是很多公司大力投入去做物流行业的一个很重要的原因：行业正处于高速发展阶段，而且体验、效率和成本方面都有巨大的优化空间，大有可为。

下图主要介绍了美团外卖现在的发展情况：



02美团外卖发展情况

美团外卖从2013年启动，目前大概能够服务2.5亿用户，已经覆盖1300多个城市，能够为200多万商户提供服务，日峰值订单超过1800万。美团外卖智能配送调度系统每天匹配50多万外卖小哥，基于海量数据和人工智能算法，确保平均配送时长不超过28分钟。这也是目前世界上规模最大、复杂度最高的多人、多点实时智能配送调度系统。

我们对美团配送的定位是：**做成最大的即时配送平台**。

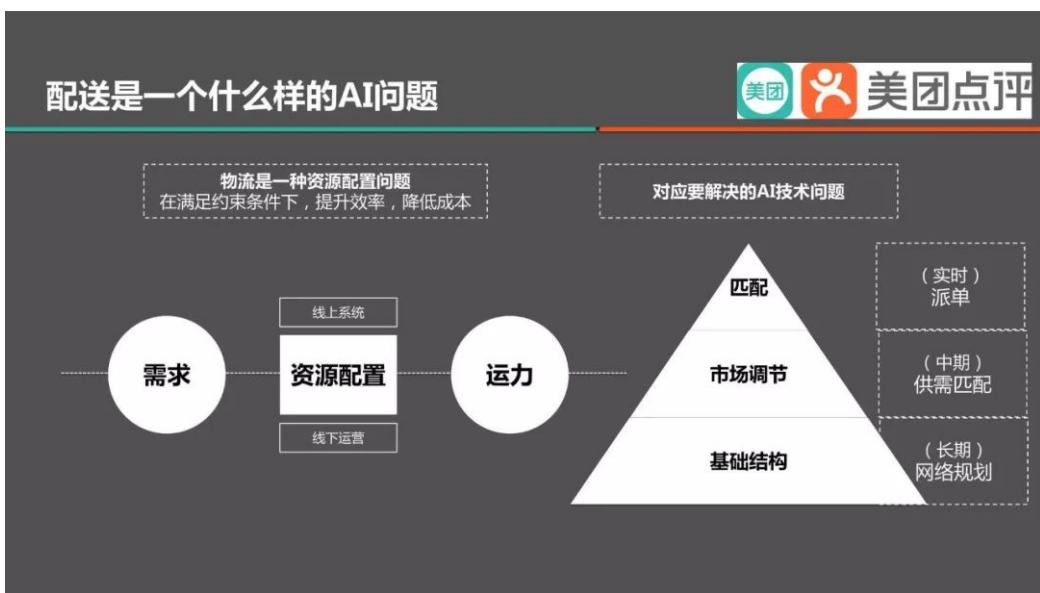


03美团配送定位

相比传统物流，即时配送包括以下几个优势：

- 第一点，非常快。从商家发单，比如说一个外卖订单，从下单到用户收到，平均要在 30分钟内能完成，最慢的也应该在一小时左右。快，是最重要的一一个特点，快，也能够使整个服务的要求和服务质量得到巨大提升。
- 第二点，能够直接联系用户和商户。之前的物流基本是从商家接单，要经过很多环节，包括仓储、运输调度、人员配送等等，最后再送到用户，中间几经转手，甚至由不同的公司配送，或者有不同的加盟商。但是即时配送直接将用户和商户联系起来，进而直接影响目标人群，这是很大的一项价值。
- 第三点，能够承担多种配送场景，不仅仅可以送外卖，还可以送商超、生鲜等等，基本上所有的同城快件，都可以纳入其配送服务范围。

总体来说，配送是一个非常复杂的业务，为了能够便于大家理解，我把这个业务模型进一些抽象和简化，可以用下面这张图来进行说明。



04配送ai问题

从本质上来讲，配送主要是把用户的配送需求和线下的各种运力（比如说骑手或车辆之类）进行匹配的过程。匹配分为线下匹配和线上匹配，线下主要靠运营，线上就是我们技术部门所构建一些系统。从这个层面而言，我们要解决的主要是在这个需求和运力之间，如何实现最优匹配的问题。

这其实也是一个相对比较传统的问题，像做广告或者推荐，都会面临这个问题，需求是要推荐的产品，供给是广告位置，但位置并不是无限多，如何在需求和供给之间达到最好的匹配，这本身就是效率优化问题，只不过广告和推荐使用的 CTR 预估，而物流中使用的方法更加复杂一些。

配送中的复杂性，具体来说有几点：

- 这是一个NP-Hard问题，计算复杂度随着规模呈指数级增加。比如是骑手身上 N个订单的路径规划问题，或者是 M 个订单与 K个骑手的订单分配问题，这两个都是指数级复杂度，而且相互关联。
- 这不但一个多点取多点送问题，而且随时有新订单增加，具有非常强的实时计算要求，当一个新订单生成后，需要在几十毫秒内完成调度运算，相比传统物流中有几十分钟以上的计算时间，即时配送系统设计的难度要大得多。
- 配送场景非常复杂，涉及天气、路况、骑手熟练程度、商家出餐速度等多达几十个因素，极大增加了解空间的随机性和复杂度，对配送算法的稳定性和适应力挑战极大。

对美团配送来说，要完成这个任务，需要分为大概三个层次，如上图最右侧所示。

- 第一层，物流基础结构建设。包括在城市里如何建设站点，如何配备人力，如何配备商家的供给情况。这些基础结构不但深刻影响配送的规模、成本、效率，而且是物流管理和运营的基础，比如加盟商管理、骑手运营等都需基于这个结构进行展开，因此这些基础结构的作用非常重要，而且它们较难进行即时调整，非常考验技术的长期预测和规划能力。
- 第二层，供需匹配的动态均衡，通过定价机制进行市场调节，包括几个方面：一个是基础定价，比如一个定单来了，到底向用户收多少钱，向商家收多少钱，给骑手多少补贴，这需要考虑很多因素，保证定价的合理、公平。另一个是供需平衡，当遇到恶劣天气等突发情况，通过动态调价方式，实时调节用户需求和运力供给，保证整个系统的稳定与用户体验。
- 第三层，订单和骑手的实时匹配，也就是派单，在订单出现后在几十毫秒内分配到一个最合适骑手，并完成多个订单的路径规划。这是一个NP-Hard问题，而且由于不断有新订单生成，需要实时计算，对并行计算引擎的要求很高。派单的优化目标是：提升整体配送的效率，并保证用户体验，是整个配送系统的核心模块之一。

以上，主要是我们对整个配送的理解，接下来讲述如何使用技术手段来进行落地和实践。

对于 AI 问题来说，整个配送在 AI 问题中的分类应该是什么样？下图给出了一个解释。



05ai位置

我们可以从两个维度来看AI问题。一个维度，是看机器与人工的对比，速度上是不是比人工更快，是不是比人工的效果更好。

另一个维度是AI所发挥的作用。首先是不是能够感知世界，比如说现在做得图像识别、语音识别以及OCR，都是像人一样能够感知这个世界。其次是不是能做到认知，比如说了一句话，“今天天气怎么样”，不但要把语音翻译成文本，这里讲的是“天气”这个实体，还有“今天”这些限定因素。第三就是要做决策，现在比较火的人工智能应用都在“如何做决策”这个层面，而且要做比人做更好的决策。一些代表性应用，比如智能助手，特别是辅助人进决策权的（聊天机器人会差一些），可以帮你完成更好的任务；比如无人驾驶；比如在物流领域，如何分配订单，并通过无人车或别的方式交付订单；还有在游戏和医疗里面，AI辅助医生做决策，在游戏里面，当用户掉线时，游戏AI可以帮助用户打怪升级。

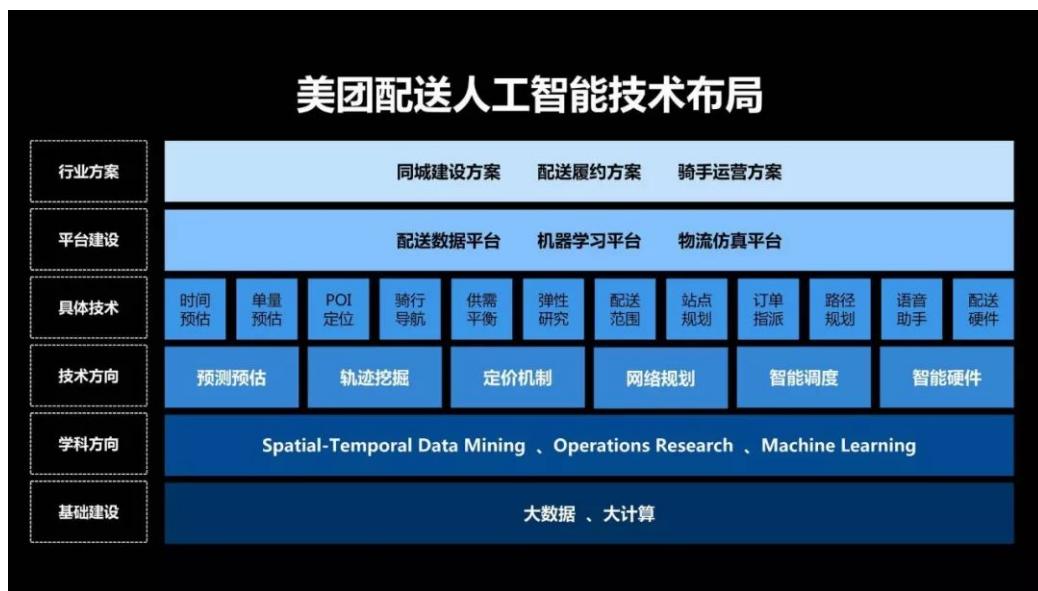
可以看到在配送层面，我们会涉及智能助手、智慧物流、无人驾驶等多个维度，而为了提升配送的整体智能化程度，我们构建了自己的“美团配送 AI”，具体来说分为两大部分：



07 智慧物流

- 第一部分是信息化，也就是数据收集。举个例子，要收集到什么样的数据？我们要收集到一个商圈的数据，这个商圈可能要精细到小区和楼宇级别，一个楼在什么地方，这个小区是不是让骑手进来，同时还要收集天气数据，比如风速、温度，是否有雾霾，因为所有数据会影响到配送的效率，用户下单情况，比如今天雾霾，北京的外卖订单量估计会上涨。
- 第二部分是智能化，也就是构建一整套智能化模块，构成一个智能配送系统，覆盖配送的各个环节。

为了完成这个“美团配送 AI”的具有挑战的目标，并考虑整个行业的长期发展，我们在整个人工智能上的布局如下：



08 技术布局

- 首先是广度方面的建设。我们的目标是配送整体流程和环节进行 AI化，从用户下单开始的每个配送步骤都要覆盖，为此我们整体技术方向的面非常广，不但横跨三个大学科，而且从预测、挖掘、定价、规划、调度和硬件等都要进行技

术研究和业务落地。

- 其次是深度方面的建设。这不单单是指技术方面，比如基础计算框架和模型研究等，还包括技术与配送业务的深度整合，比如配送仿真平台建设，具备进行多配送场景的仿真能力，无需上线就能够对不同业务策略效果进行准确预估。同时还要结合行业情况，提供行业的智能化解决方案，比如在骑手运营方面，更有效的骑手激励和骑手留存的机制设计。

而美团外卖语音助手就属于我们在广度和深度结合比较好的案例。接下来就和大家分享一下我们在整个智能助手的实践和设计过程中，以及在整个物流业务中，如何将人工智能技术更好的落地的一些经验。

美团外卖智能语音助手定位



AiCon
全球人工智能与机器学习技术大会

主办: Geekbang InfoQ

09外卖订单配送过程

我们为什么要智能语音助手？骑手到底在什么情况下需要智能助手服务，整个服务里面的关键是什么？先解释一下这个问题。如上图所示，这个是整个骑手在配送过程中遇到的一些环节，可以分为两大部分。

第一部分是线上的决策，而且涉及的决策各式各样。举个例子，这个骑手有定单，要送到一个用户那里，他可能要做几个决策，比如说要不要给用户打电话，因为有些地方是不用打电话的，像住宅楼里面，骑手有很大概率知道这个用户应该在家里的，不用打电话；有些必须打，比如写字楼，因为骑手上不去，所以需要提前打电话让用户下来。

但需要提前多长时间呢？是提前一分钟，两分钟，还是五分钟？这个问题很关键，如果打电话时间比较早，用户就会提前下来，会造成用户等待骑手的问题，用户体验不好，可能会有投诉。如果这个骑手非常保守，到楼下再打，但用户住在10层，那么用户下来包括等电梯的时间可能要需要10分钟，效率会变得非常低。

第二个部分是骑手操作过程，因为骑手会频繁和手机交互。他要查看一个定单，步骤非常复杂，把手机拿出来，解锁，打开App，查看信息，做操作（比如说点击完成），最后放回手机，大概需要五到六个过程。如果操作快，也需要10到20秒钟。而且很多骑手是在骑行过程中做这些操作的，这样会非常危险。

总结一下，配送骑手遇到的困难可以总结为三个大的层面：

骑手遇到的三大难题



任务复杂

熟悉程度：在哪？怎么去？上楼多少时间？
最佳决策：先去哪？出餐快慢？什么时候打电话？

操作繁琐

因为不熟悉和信息实时更新，频繁操作手机
一个操作需要5~6个步骤完成

安全隐患

频繁查看手机，注意力不集中
骑手安全问题比较突出



主办方: Geekbang InfoQ

10骑手三大难题

- 第一，任务复杂，需要做很多决策，不过复杂度会随着骑手的熟练度有所变化。
- 第二，操作繁琐，大概需要五到六个步骤，至少需要 10 到 20 秒，或者更长时间。
- 第三，骑手在骑行过程中操作手机非常危险。对于有 50 万骑手的平台，我们必须考虑骑手在整个驾驶过程中的安全。

基于这些考虑，我们做了美团外卖语音助手，它的定位主要包括以下三点：

美团外卖语音助手的定位



安全

全流程语音交互

- 语音替代传统操作，让交互更自然
- 解放骑手双手，让配送更加安全



高效

极简操作步骤

- 在全场景下，将复杂（5~6步）的操作，通过1~2步，简单快捷的完成
- 通过命令词快速完成操作



智能

全场景智能引导

- 针对时间、地点、事件，识别场景
- 根据场景线索，主动发起交互
- 场景识别+预测，给出最优操作建议
- 场景识别，预测接下来最可能发生的事件
- 根据最佳时机主动发起操作提醒



接收派单 到店上报 取餐上报 拨打电话 送达上报

主办方: Geekbang InfoQ

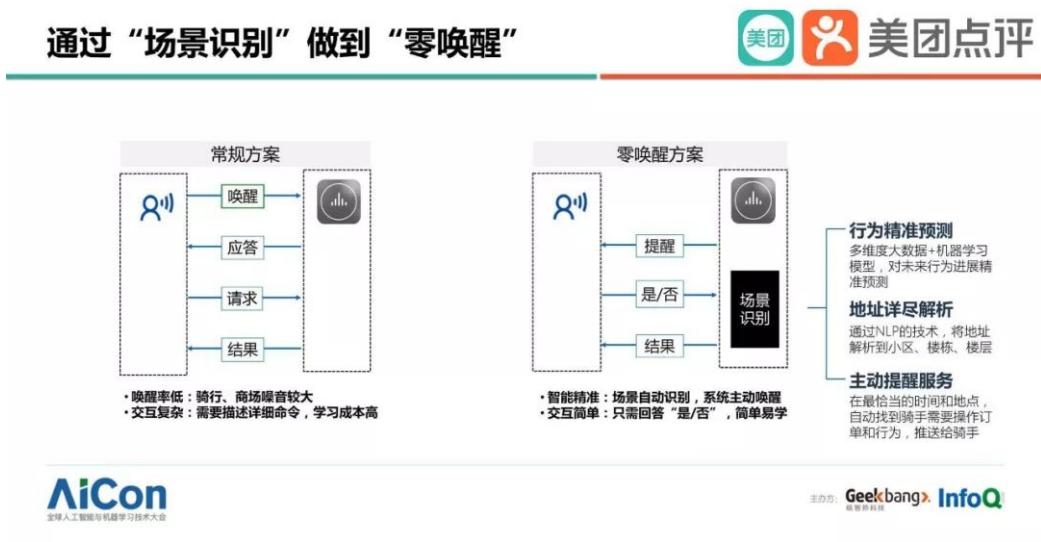
11语言助手定位

- 第一点就是要求安全，要做一套全流程的语音交互方案，配送过程中的各个环节都能用语音操作，不需要骑手看手机，解放双手，让骑手更加安全。比如在行驶过程中，有个订单过来了，系统问骑手要不要接单，只要通过指令回答，“是”或“否”，或者“OK”这种，整个过程就完成了；不需要像以前那样的，把手机掏出来再进行操作，这个场景非常受骑手欢迎。
- 第二个，设计极简的步骤，所有操作能在一到两个步骤里完成，第一个步骤是信息播报，第二个步骤通过语音命令完成操作，将原来的五到六个步骤，精简到现在的一到两个。
- 第三个，提供很多智能化服务。最典型的，刚才说的骑手要去一个楼，用户可能在 5 楼，可能在 4 楼，这个用户下来需要多长时间，做智能化推荐，根据用户的地址信息，系统智能推荐打电话时机，当然还包括像导航之类的基础功

能。

上文的分析，基本上将我们怎么把智能语音助手在场景里落地的最关键的点分析出来了。我们要落地，最核心的就是要帮助骑手完成配送任务，而不是“聊天”或者“问答”。这就要求语音交互整个过程要非常便捷，同时也非常智能。

而我们遇到的第一个挑战，就是交互模式如何设计的问题。



12 零唤醒

如上图所示，左侧是一般的语音助手方案，需要唤醒、应答、请求和再应答四个步骤，但是并不符合配送场景的要求。首先，骑手所在的场景，噪音很大，比如风噪、汽车噪音以及商场噪音等等，唤醒比较难实现。其次，需要四个步骤，还要考虑骑手的工作状态，这个操作过程太繁琐。

那怎么办？我们思考，是否能做到一套不需要唤醒的解决方案呢？答案很肯定，可以做！

- 第一点，我们的数据非常多。包括骑手、用户和商家，这些数据都是实时的，我们能够了解比骑手多得多的全局配送信息。
- 第二点，我们能做到精准预测，利用机器学习、智能调度等技术，可以对骑手下一个操作场景进行识别。

举个例子，一个骑手身上可能有几个订单，他正在朝一个地方前进，通过场景分析，我们知道他要给具体哪个用户配送，而且我们能了解用户在这个楼里的几层，下来大概需要几分钟，所以能够推算出来，大概在哪个时间点提醒骑手打电话比较好。这样我们就可以省略唤醒和应答流程，直接给骑手发提醒，骑手只要回答是或否够就可以了。这样设计才符合骑手线下的实际配送情况，能够真正给骑手解决实际问题，才能够真正称之为“智能”。

AI核心技术

具体技术分为几个主要的部分。第一个部分是基础设施，包括语音识别和语义理解，现在这方面开源的东西非常多，做通用的语音识别不是很难。

基础设施：语音识别 + 语义理解



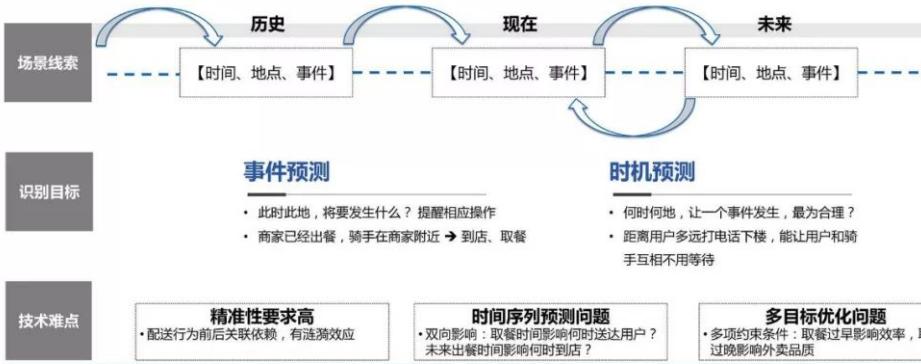
主办方: Geekbang InfoQ

13基础设施

在我们场景中，要解决各种环境噪音的问题，可能骑手并没有说话，但旁边有些噪音，车的噪音或者别的噪音，甚至路上正在放一个歌，都会被识别为是骑手在说话，所以 VAD（静音检测）方面需要做很多工作。

另一个基本的组件是NLU，自然语言理解。举个例子，骑手要给尾号 6551打电话，首先系统要知道，骑手的意图是要打电话，后面要调起打电话的操作；其次要知道打电话的对象是谁，是用户，而不是商户，这就要找出用户信息；第三，要做检测，比如骑手已经送完某个订单，再打电话可能是错误操作，需要提醒骑手。

精准的场景识别



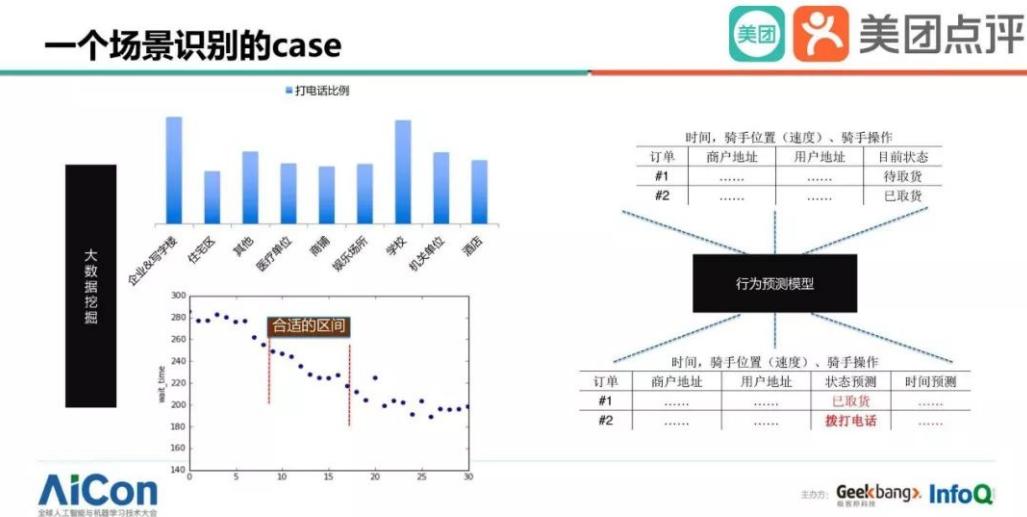
主办方: Geekbang InfoQ

14精准场景识别

即时配送场景是一个典型的时间序列问题。从上面的图可以看出，场景包含前后关联，一个骑手历史的行为和决策会影响现在，同时现在的决策和行为会影响未来，这是个典型的时间序列问题。

场景识别要解决的两个主要目标，一个是事件预测，要知道下一时刻大概会发生什么事情，比如骑手是不是已到商家，商家是不是已经出餐；另一个是时机预测，未来要打电话，到底什么时候打更合适？

为了更好的说明，举个打电话的案例。



15识别案例

首先，要判断是否需要打电话，如果在不需要的场景也频繁提醒打电话，对骑手和用户都是骚扰。上图列举了不同地址类型下骑手打电话的比例，可以看到，像在企业和写字楼里面比例很高，但是住宅区就很低了，因为在住宅区，很大概率用户都是在家的。

其次，要针对每一个小区和楼宇类型，给一个合适的打电话时机，即提前多久打电话，对骑手和用户是最好的体验。打电话太早，用户在楼下等骑手，体验比较差。打电话太晚，骑手在楼下等用户，效率太低。我们有精准的骑车轨迹数据，我们知道针对每一栋楼、每个小区，骑手在不同时刻打电话时会在楼下停留多久，所以可以画出一个曲线。合适的区间就在两条红线之间。

前两个主要是大数据分析，最后要实时决策，哪个订单，什么时刻需要打电话。这里就要根据骑手的实时数据了，包括订单状态、轨迹状态、环境情况等等，结合前面的大数据分析进行实时的预测骑手下一个配送地点和配送任务，并在合适的时机通过语音助手给出提醒。

具体到实现方面，场景识别需要三方面的技术：骑手轨迹挖掘、机器学习和数据挖掘。

场景识别的技术支撑

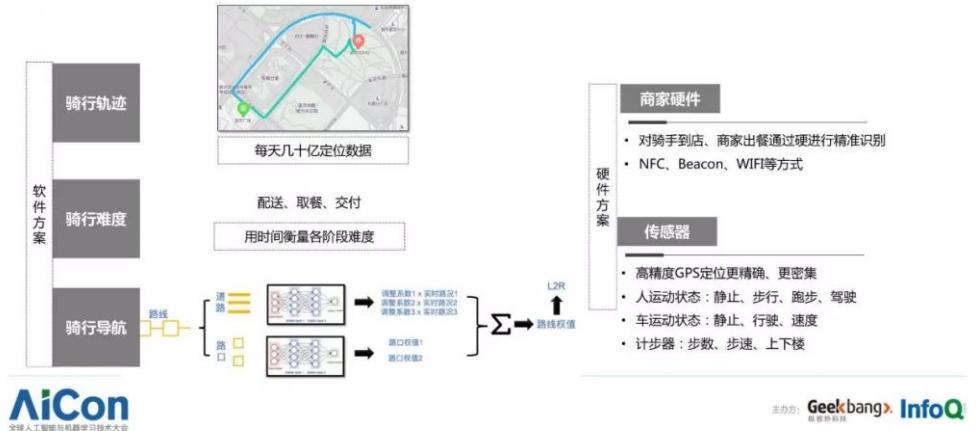


主办方: Geekbang InfoQ

16场景识别技术支撑

先介绍一下轨迹，我们每天能有几十亿次的定位数据，进而可以基于这些数据做很多事情。

轨迹挖掘：软硬件结合的AI解决方案



17轨迹挖掘

- 第一，可以精准知道 A、B 两点间最好的导航方式，相比第三方地图，可以挖掘到 A 和 B 间可能有可能有更好的骑行通过方式。
- 第二，光有轨迹数据还不够，我们还需要解决室内定位问题，室内 GPS 定位已经不够用了，需要新的技术体系。比如 WiFi 定位，同时还需要设计硬件，比如在商家部署硬件，判断骑手是否到店
- 第三，传感器的使用，无论在室内还是在室内时候，我们不但要知道骑手的精准定位，还要知道运动方式，比如是停留、步行、骑行，是爬楼还是坐电梯，这些信息不但判断骑手在到底做什么。而且能够精细刻画配送难度，在定价和调度上非常有价值。

我们可以通过骑行轨迹来修正导航和定位。来看两个例子。

案例：导航和定位的修正结果



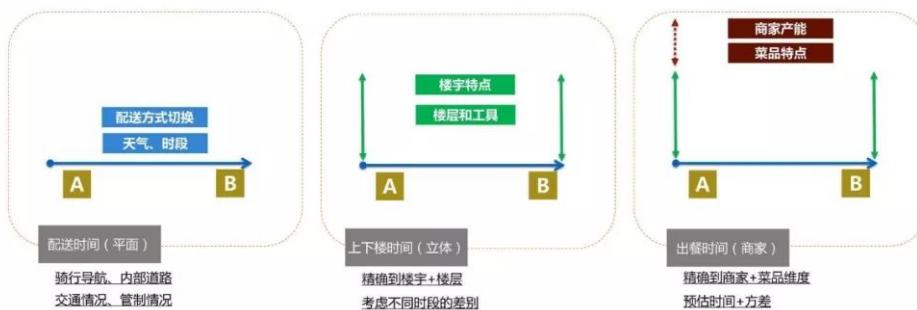
18修正结果

第一个例子（左侧）用户在下单时定位的分布，因为大家在室内下单，定位偏离是非常大的。但通过骑手轨迹的修正，实际上大概只有四个点，每个点可以认为是这个这栋楼的一个门口，这大幅提升了用户的定位精度，让骑手配送更容易。

第二个例子（右侧）通过骑手轨迹对 AB两个点的骑行路径进行修正，上图中轨迹分析发现了更短路径，穿过小区更节省时间；下图中，原地图导航要跨过中间过街天桥，但通过轨迹发现更多骑手是绕行通过，这才更符合真实的情况。

下面介绍一些机器学习相关技术，主要是应用在各种时间预估层面。

考虑三个维度的精细预估



主办方: Geekbang > InfoQ

19精细预估

只有高精度的 ETA (预计到达时间) 预估，这样才能更加准确的预测骑手行为，我们会做三个维度的精细预估，包括平面的配送时长、上下楼时长以及商家出餐时长。这样才能比较全面和精细的刻画骑手的配送过程。

为此，我们做了很多基础工作，比如实时特征平台，机器学习平台，包括深度学习在内模型等各种机器学习相关工作。同时我们还会做比较精细的配送知识图谱建设工作，比如精细化地址解析。

数据挖掘：精细化地址解析



自然语言处理

- 实体识别：小区、楼宇、单元、楼层
- 实体消歧：基于上下文的语义消歧
- 地址归一：结合文本和地图的地址的归一+聚类



实时地址解析

- 通过NLP实时解析地址
- 通过CART进行Mapping，获取相应统计数据
- 通过Model进行平滑（数据稀疏）



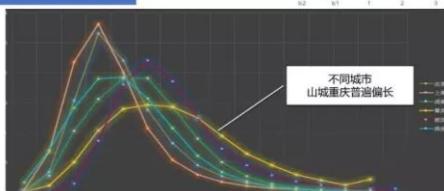
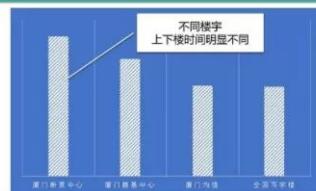
主办: Geekbang InfoQ

20精细挖掘

地址对配送来说是非常重要的信息，通过 NLP 和地图搜索的方法，解析成层次结构，对分析商圈、楼宇维度的画像非常有帮助。我们把一个地址分解为四个层次，小区、楼宇、单元号和楼层等。其中要解决很多实际问题，比如用户填写的信息完全不标准、存在歧义等问题。

做了这些工作之后，能实际产生的效果还是很有意思的。我们通过“上下楼时间”这个具体场景来进行分析。

案例：上下楼时间的结果分析



主办: Geekbang InfoQ

21上下楼分析

- 第一张图，是不同楼宇的上下楼时长，左侧两个是厦门的两个楼宇的时间，右侧两个厦门平均值和全国的均值。可以看到，不同楼宇的上下楼时长还是存在很大差异，无法简单利用城市或者全国维度的均值进行替代。
- 第二张图，是不同楼层的上下楼时间，从 B2开始到 8楼。有个很有意思的是，上下楼时长与高度不是线性关系，大概在二楼、三楼和四楼时，相隔的时间很长，但是到了五楼、六楼、七楼，时间差就很小了。原因很简单，楼层较低

时，骑手可能会选择爬楼。高层则选择乘电梯。不同楼层之间停留时间很短，越往上时间间隔越小。

- 第三张图，是不同城市的上下楼时长分布，最有意思的是黄色的线，也就是重庆的整体上下楼时长明显偏长。因为重庆是山城，房子经常在半山腰上，与平原比起来其上下楼的难度当然更大。

整体效果

上面整体介绍了语音助手依赖的场景识别技术，现在介绍一下语音助手的整体效果。首先语音助手提供了四个核心功能，包括定制耳机、语音交互、场景识别、智能引导等。

提供四项核心功能



定制蓝牙耳机



语音交互



场景识别



智能引导



主办方: Geekbang x InfoQ

22四项核心

为什么要定制耳机呢？在骑手的使用环境中，需要克服很多噪音，很难通过软件和程序去做，而必须通过硬件去做。所以我们和厂商进行合作，定制去噪效果好的硬件。

定制蓝牙耳机



MT-R1

- 高防风降噪**
有效滤除骑行风噪和环境噪声
- 高户外防水**
IPX-7级防水，中雨使用无压力
- 充足续航**
8小时+高频持续工作
- 佩戴舒适**
左右耳可调防止不适，保护双耳
- 兼容适配**
兼容骑手TOP50大部分机型
- 快速响应**
多APP调用快速稳定响应



六大技术专为服务骑手



主办方: Geekbang x InfoQ

23定制蓝牙耳机

第二个功能是语音交互，它可以在派单、查询、取餐、拨打电话等配送全流程中实现语音交互，骑手整个过程中不需要看手机，只要耳机提醒就可以完成智能配送。

全流程语音交互



主办方: Geekbang & InfoQ

24全流程语音

第三个是智能引导功能，包括安全驾驶提醒，信息播报，任务地图引导等，主要是让骑手行驶更加安全，提供全面的信息服务，让骑手配送更加方便和高效。

全场景智能引导



主办方: Geekbang & InfoQ

25全场景智能引导.

下图是智能语音在线下推广中的一些实际数据。

智能语音助手的效果-更安全



有效降低操作app的次数，让骑手配送更安全



主办: Geekbang InfoQ

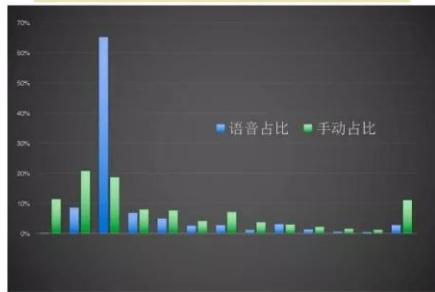
26语音助手效果

蓝色的线是使用语音助手的骑手的操作次数，绿色的线是不使用的操作次数。可以看到，操作次数明显下降。但是还没有降为 0，有两个原因：骑手在静止状态下，不需要使用语音助手；有些骑手的蓝牙耳机还没有下发到位。再来看下一张图：

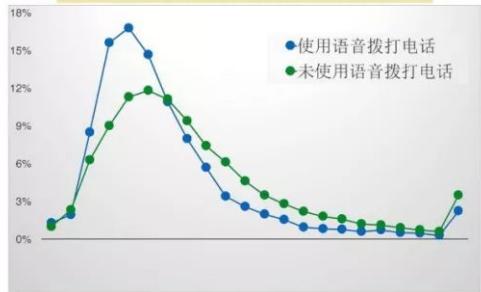
智能语音助手的效果-更高效



语音接单，接单时间更短，提高工作效率



打电话智能提醒，效缩短交付时长，降低配送成本



主办: Geekbang InfoQ

27语音助手效果更安全

左图是骑手接单时间时长分布，越往右骑手接单的时间越长，用户体验越差。绿色的线就是之前骑手手动接单的一个分布，长尾情况比较严重，通过语音接单，接单时长明显向左侧靠拢，整体接单时长明显缩小，比较好的提升了用户体验。

右图是骑手在用户交付外卖所花费的时间的比例，横轴是骑手在楼下等待用户的时长，越往右，骑手在楼下等用户的时间越长。通过语音的提醒后，可以明显降低骑手长时间等待的情况，节省了大量骑手的时间。

写在最后

总结一下，语音识别和语音助手在实际落地过程中面临很多挑战，而且大多和场景有关系，场景识别非常重要，甚至比语音识别更为重要。

因为语音识别现在已经是比较通用的技术了，而且有很多专业厂商提供服务，硬件也是如此，进行定制化相对比较容易。因此目前做一个软硬件结合的语音助手，从基础技术来讲都不是问题，想做一个 DEMO 并不会存在太大的技术障碍。

反而在具体的业务中，如何结合业务场景，把语音助手落地，才是我们需要真正考虑的。也就是说，如何将语音助手从“能用”做到“好用”，再做到让用户“愿意用”，这些才是未来语音助手面对的真正挑战。

语音识别和语音助手在实际落地过程中有很多挑战，而且和场景有关系，场景识别比较重要的，甚至比语音识别更为重要，因为语音识别现在已经是比较通用的技术了，如何结合业务场景，把语音助手落地、用好，可能是未来一段时间的挑战。

为了实现配送的全面智能化，美团点评在其中做了大量工作和尝试，这里不单单是要做好机器学习，还包括如何进行更好的实时运筹优化、实时空间数据挖掘以及人机交互等多个方面的技术内容。

作者简介

- 仁清，美团点评配送算法策略方向负责人。2016 年加入美团点评，负责美团配送业务的算法整体方向。包括智能调度系统，智能网络规划系统、机器学习平台、配送仿真平台等，全面支持美团专送、快送、跑腿等多个业务方向发展。加入美团点评前，曾任百度凤巢团队 T9 架构师，从事搜索广告 NLP、数据挖掘、检索技术方向研究。

基于TensorFlow Serving的深度学习在线预估

作者: 仲达 鸿杰 廷稳

一、前言

随着深度学习在图像、语言、广告点击率预估等各个领域不断发展，很多团队开始探索深度学习技术在业务层面的实践与应用。而在广告CTR预估方面，新模型也是层出不穷：Wide and Deep^[^1]、DeepCross Network^[^2]、DeepFM^[^3]、xDeepFM^[^4]，美团很多篇深度学习博客也做了详细的介绍。但是，当离线模型需要上线时，就会遇见各种新的问题：离线模型性能能否满足线上要求、模型预估如何镶嵌到原有工程系统等等。只有准确的理解深度学习框架，才能更好地将深度学习部署到线上，从而兼容原工程系统、满足线上性能要求。

本文首先介绍下美团平台用户增长组业务场景及离线训练流程，然后主要介绍我们使用TensorFlow Serving部署WDL模型到线上的全过程，以及如何优化线上服务性能，希望能对大家有所启发。

二、业务场景及离线流程

2.1 业务场景

在广告精排的场景下，针对每个用户，最多会有几百个广告召回，模型根据用户特征与每一个广告相关特征，分别预估该用户对每条广告的点击率，从而进行排序。由于广告交易平台（AdExchange）对于DSP的超时时间限制，我们的排序模块平均响应时间必须控制在10ms以内，同时美团DSP需要根据预估点击率参与实时竞价，因此对模型预估性能要求比较高。

2.2 离线训练

离线数据方面，我们使用Spark生成TensorFlow^[^5]原生态的数据格式tfrecord，加快数据读取。

模型方面，使用经典的Wide and Deep模型，特征包括用户维度特征、场景维度特征、商品维度特征。Wide部分有80多特征输入，Deep部分有60多特征输入，经过Embedding输入层大约有600维度，之后是3层256等宽全连接，模型参数一共有35万参数，对应导出模型文件大小大约11M。

离线训练方面，使用TensorFlow同步 + Backup Workers^[^6]的分布式框架，解决异步更新延迟和同步更新性能慢的问题。

在分布式ps参数分配方面，使用GreedyLoadBalancing方式，根据预估参数大小分配参数，取代Round Robin取模分配的方法，可以使各个PS负载均衡。

计算设备方面，我们发现只使用CPU而不使用GPU，训练速度会更快，这主要是因为尽管GPU计算上性能可能会提升，但是却增加了CPU与GPU之间数据传输的开销，当模型计算并不太复杂时，使用CPU效果会更好些。

同时我们使用了Estimator高级API，将数据读取、分布式训练、模型验证、TensorFlow Serving模型导出进行封装。使用Estimator的主要好处在于：

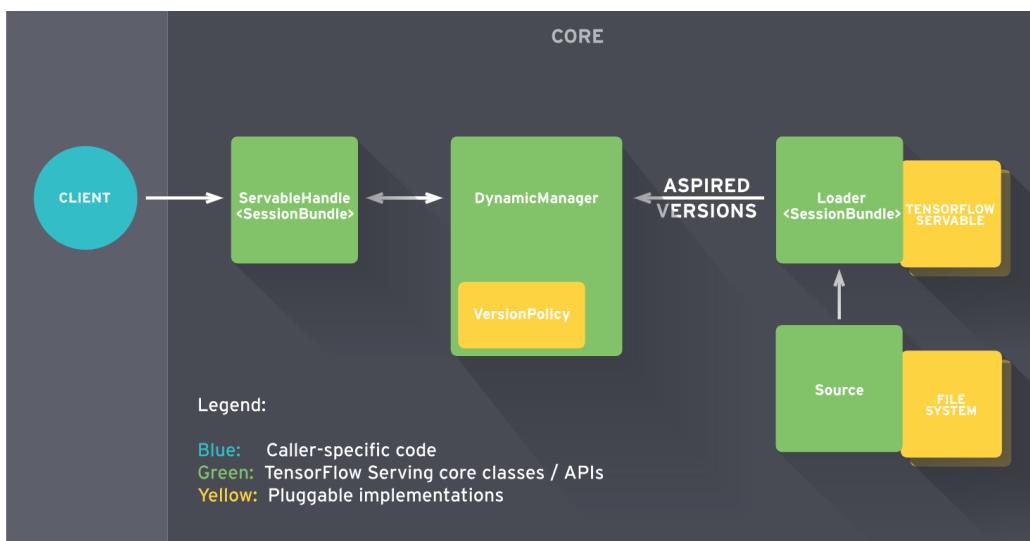
1. 单机训练与分布式训练可以很简单的切换，而且在使用不同设备：CPU、GPU、TPU时，无需修改过多的代码。
2. Estimator的框架十分清晰，便于开发者之间的交流。
3. 初学者还可以直接使用一些已经构建好的Estimator模型：DNN模型、XGBoost模型、线性模型等。

三、TensorFlow Serving及性能优化

3.1 TensorFlow Serving介绍

TensorFlow Serving是一个用于机器学习模型Serving的高性能开源库，它可以将训练好的机器学习模型部署到线上，使用gRPC作为接口接受外部调用。TensorFlow Serving支持模型热更新与自动模型版本管理，具有非常灵活的特点。

下图为TensorFlow Serving整个框架图。Client端会不断给Manager发送请求，Manager会根据版本管理策略管理模型更新，并将最新的模型计算结果返回给Client端。



TensorFlow Serving架构，图片来源于TensorFlow Serving官方文档

美团内部由数据平台提供专门TensorFlow Serving通过YARN分布式地跑在集群上，其周期性地扫描HDFS路径来检查模型版本，并自动进行更新。当然，每一台本地机器都可以安装TensorFlow Serving进行试验。

在我们站外广告精排的场景下，每来一位用户时，线上请求端会把该用户和召回所得100个广告的所有信息，转化成模型输入格式，然后作为一个Batch发送给TensorFlow Serving，TensorFlow Serving接受请求后，经过计算得到CTR预估值，再返回给请求端。

部署TensorFlow Serving的第一版时，QPS大约200时，打包请求需要5ms，网络开销需要固定3ms左右，仅模型预估计算需要10ms，整个过程的TP50线大约18ms，性能完全达不到线上的要求。接下来详细介绍下我们性能优化的过程。

3.2 性能优化

3.2.1 请求端优化

线上请求端优化主要是对一百个广告进行并行处理，我们使用OpenMP多线程并行处理数据，将请求时间性能从5ms降低到2ms左右。

```
#pragma omp parallel for
for (int i = 0; i < request->ad_feat_size(); ++i) {
    tensorflow::Example example;
    data_processing();
}
```

3.2.2 构建模型OPS优化

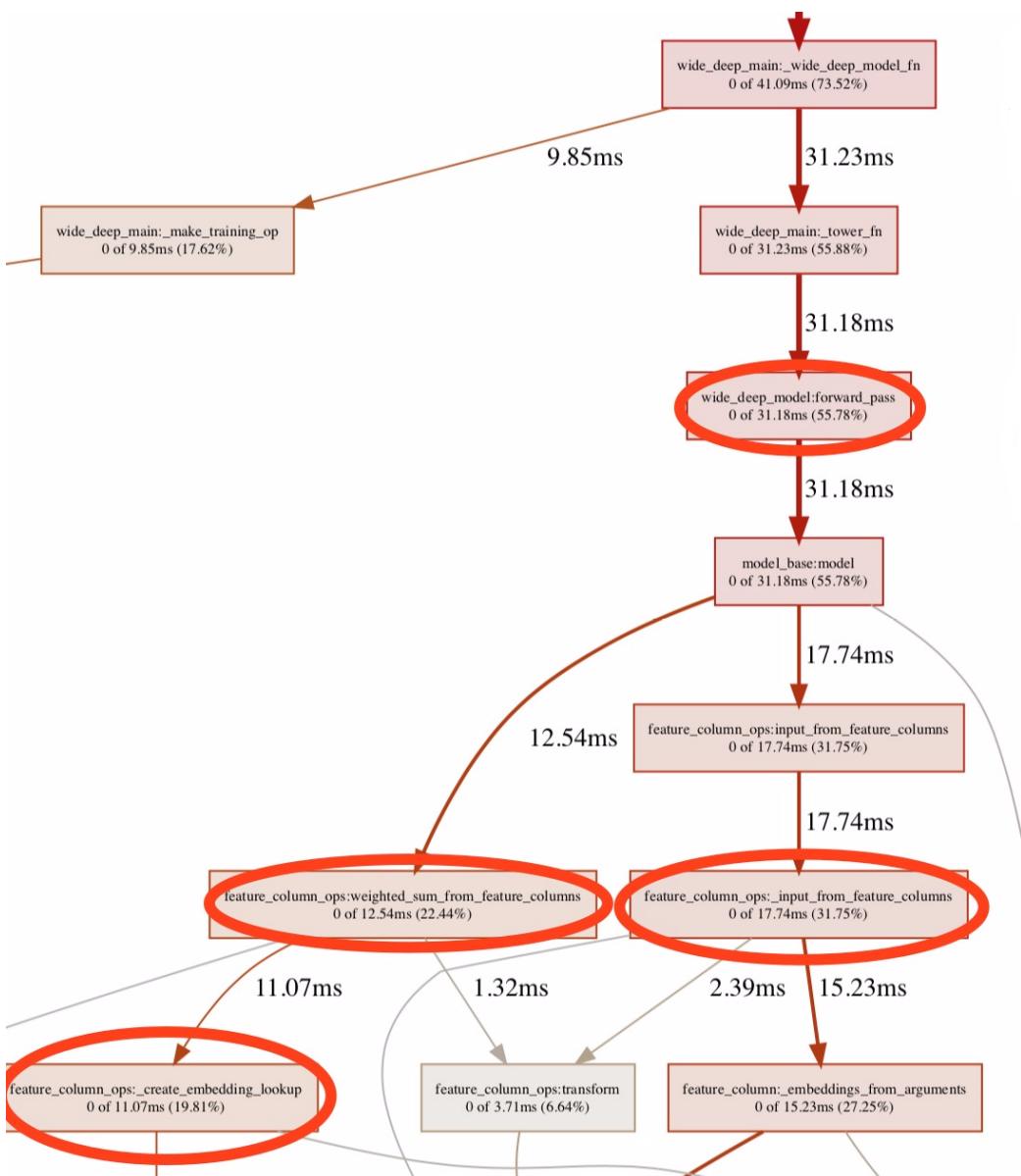
在没有进行优化之前，模型的输入是未进行处理的原格式数据，例如，渠道特征取值可能为：'渠道1'、'渠道2' 这样的string格式，然后在模型里面做One Hot处理。

最初模型使用了大量的高阶tf.feature_column对数据进行处理，转为One Hot和embedding格式。使用tf.feature_column的好处是，输入时不需要对原数据做任何处理，可以通过feature_column API在模型内部对特征做很多常用的处理，例如：tf.feature_column.bucketized_column可以做分桶，tf.feature_column.crossed_column可以对类别特征做特征交叉。但特征处理的压力就放在了模型里。

为了进一步分析使用feature_column的耗时，我们使用tf.profiler工具，对整个离线训练流程耗时做了分析。在Estimator框架下使用tf.profiler是非常方便的，只需加一行代码即可。

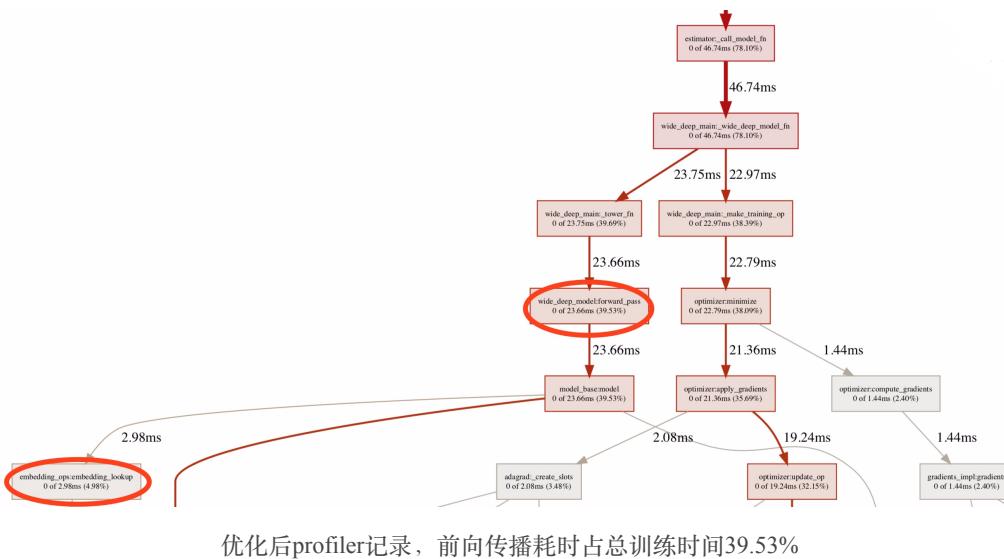
```
with tf.contrib.tfprof.ProfileContext(job_dir + '/tmp/train_dir') as pctx:
    estimator = tf.estimator.Estimator(model_fn=get_model_fn(job_dir),
                                         config=run_config,
                                         params=hparams)
```

下图为使用tf.profiler，网络在向前传播的耗时分布图，可以看出使用feature_column API的特征处理耗费了很大时间。



优化前profiler记录，前向传播的耗时占总训练时间55.78%，主要耗费在feature_column OPS对原始数据的预处理

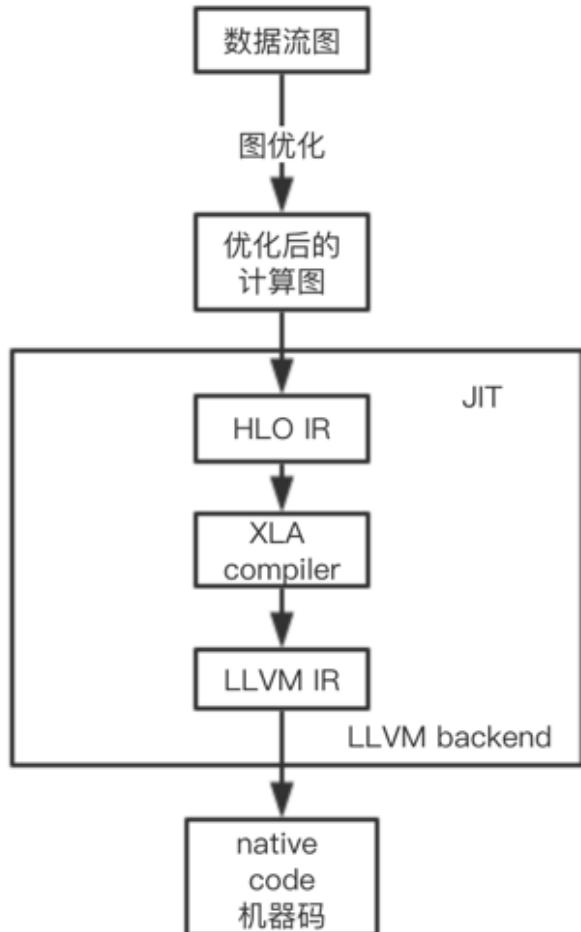
为了解决特征在模型内做处理耗时大的问题，我们在处理离线数据时，把所有string格式的原生数据，提前做好One Hot的映射，并且把映射关系落到本地feature_index文件，进而供线上线下使用。这样就相当于把原本需要在模型端计算One Hot的过程省略掉，替代为使用词典做O(1)的查找。同时在构建模型时候，使用更多性能有保证的低阶API替代feature_column这样的高阶API。下图为性能优化后，前向传播耗时在整个训练流程的占比。可以看出，前向传播的耗时占比降低了很多。



3.2.3 XLA, JIT编译优化

TensorFlow采用有向数据流图来表达整个计算过程，其中Node代表着操作（OPS），数据通过Tensor的方式来表达，不同Node间有向的边表示数据流动方向，整个图就是有向的数据流图。

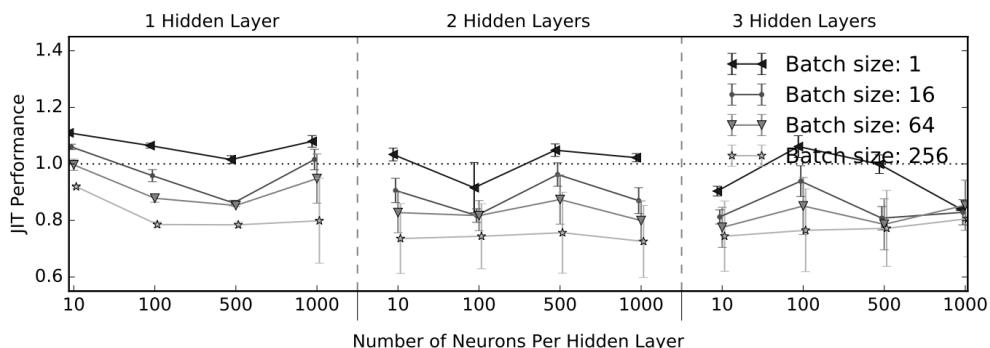
XLA (Accelerated Linear Algebra) 是一种专门对TensorFlow中线性代数运算进行优化的编译器，当打开JIT (Just In Time) 编译模式时，便会使用XLA编译器。整个编译流程如下图所示：



TensorFlow计算流程

首先TensorFlow整个计算图会经过优化，图中冗余的计算会被剪掉。HLO (High Level Optimizer) 会将优化后的计算图生成HLO的原始操作，XLA编译器会对HLO的原始操作进行一些优化，最后交给LLVM IR根据不同的后端设备，生成不同的机器代码。

JIT的使用，有助于LLVM IR根据 HLO原始操作生成 更高效的机器码；同时，对于多个可融合的HLO原始操作，会融合成一个更加高效的计算操作。但是JIT的编译是在代码运行时进行编译，这也意味着运行代码时会有一部分额外的编译开销。



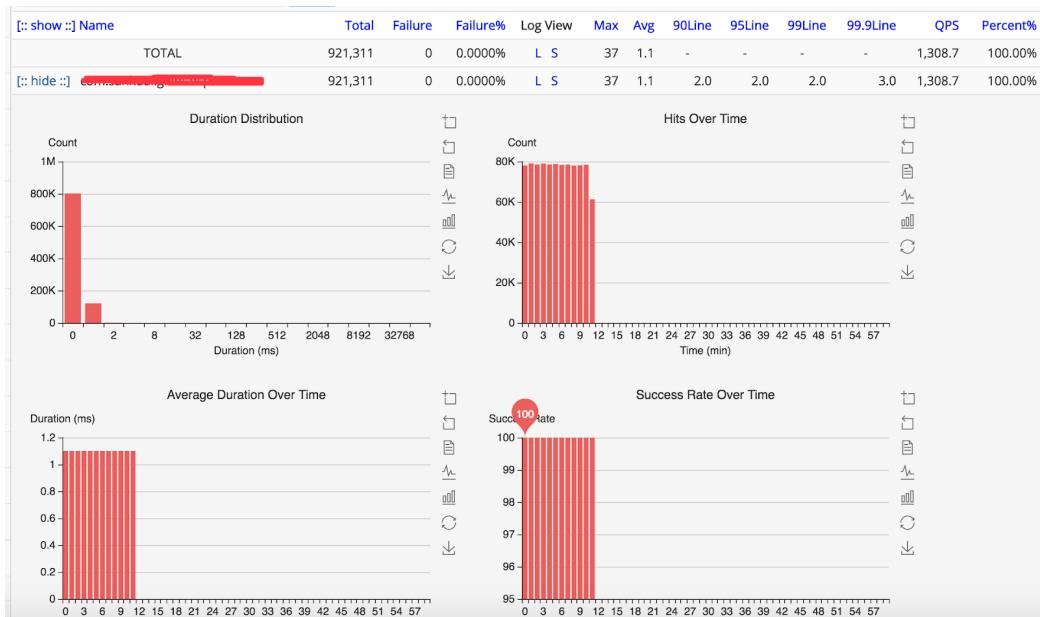
网络结构、Batch Size对JIT性能影响[7]

上图显示为不同网络结构，不同Batch Size下使用JIT编译后与不使用JIT编译的耗时之比。可以看出，较大的Batch Size性能优化比较明显，层数与神经元个数变化对JIT编译优化影响不大。

在实际的应用中，具体效果会因网络结构、模型参数、硬件设备等原因而异。

3.2.4 最终性能

经过上述一系列的性能优化，模型预估时间从开始的10ms降低到1.1ms，请求时间从5ms降到2ms。整个流程从打包发送请求到收到结果，耗时大约6ms。

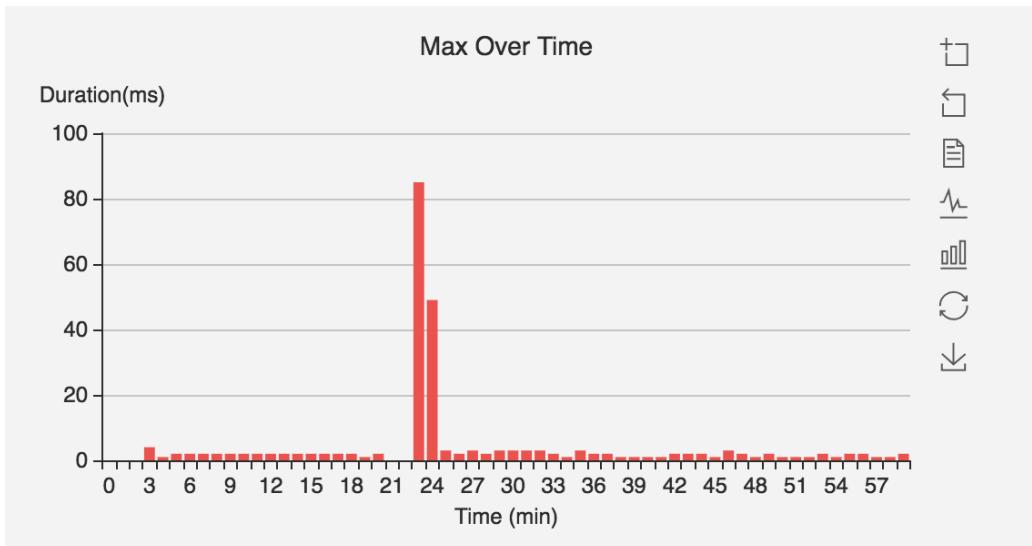


模型计算时间相关参数：QPS:1308，50line:1.1ms，999line:3.0ms。下面四个图分别为：耗时分布图显示大部分耗时控制在1ms内；

请求次数显示每分钟请求大约8万次，折合QPS为1308；平均耗时时间为1.1ms；成功率为100%

3.3 模型切换毛刺问题

通过监控发现，当模型进行更新时，会有大量的请求超时。如下图所示，每次更新都会导致有大量请求超时，对系统的影响较大。通过TensorFlow Serving日志和代码分析发现，超时问题主要源于两个方面，一方面，更新、加载模型和处理TensorFlow Serving请求的线程共用一个线程池，导致切换模型时候无法处理请求；另一方面，模型加载后，计算图采用Lazy Initialization方式，导致第一次请求需要等待计算图初始化。



模型切换导致请求超时

问题一主要是因为加载和卸载模型线程池配置问题，在源代码中：

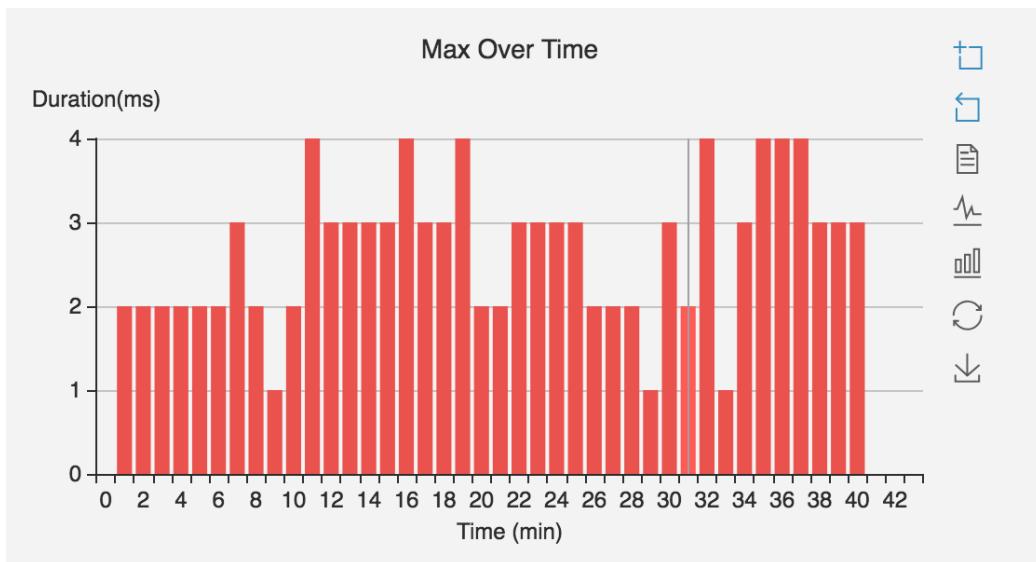
```
uint32 num_load_threads = 0; uint32 num_unload_threads = 0;
```

这两个参数默认为 0，表示不使用独立线程池，和Serving Manager在同一个线程中运行。修改成1便可以有效解决此问题。

模型加载的核心操作为RestoreOp，包括从存储读取模型文件、分配内存、查找对应的Variable等操作，其通过调用Session的run方法来执行。而默认情况下，一个进程内的所有Session的运算均使用同一个线程池。所以导致模型加载过程中加载操作和处理Serving请求的运算使用同一线程池，导致Serving请求延迟。解决方法是通过配置文件设置，可构造多个线程池，模型加载时指定使用独立的线程池执行加载操作。

对于问题二，模型首次运行耗时较长的问题，采用在模型加载完成后提前进行一次Warm Up运算的方法，可以避免在请求时运算影响请求性能。这里使用Warm Up的方法是，根据导出模型时设置的Signature，拿出输入数据的类型，然后构造出假的输入数据来初始化模型。

通过上述两方面的优化，模型切换后请求延迟问题得到很好的解决。如下图所示，切换模型时毛刺由原来的84ms降低为4ms左右。



优化后模型切换后，毛刺降低

四、总结与展望

本文主要介绍了用户增长组基于Tensorflow Serving在深度学习线上预估的探索，对性能问题的定位、分析、解决；最终实现了高性能、稳定性强、支持各种深度学习模型的在线服务。

在具备完整的离线训练与在线预估框架基础之后，我们将会加快策略的快速迭代。在模型方面，我们可以快速尝试新的模型，尝试将强化学习与竞价结合；在性能方面，结合工程要求，我们会对TensorFlow的图优化、底层操作算子、操作融合等方面做进一步的探索；除此之外，TensorFlow Serving的预估功能可以用于模型分析，谷歌也基于此推出What-If-Tools来帮助模型开发者对模型深入分析。最后，我们也会结合模型分析，对数据、特征再做重新的审视。

作者简介

- 仲达，2017年毕业于美国罗彻斯特大学数据科学专业，后在加州湾区Stentor Technology Company工作，2018年加入美团，主要负责用户增长组深度学习、强化学习落地业务场景工作。
- 鸿杰，2015年加入美团点评。美团平台与酒旅事业群用户增长组算法负责人，曾就职于阿里，主要致力于通过机器学习提升美团点评平台的活跃用户数，作为技术负责人，主导了美团DSP广告投放、站内拉新等项目的算法工作，有效提升营销效率，降低营销成本。
- 廷稳，2015年加入美团点评。在美团点评离线计算方向先后从事YARN资源调度及GPU计算平台建设工作。

招聘信息

美团DSP是美团在线数字营销的核心业务方向，加入我们，你可以亲身参与打造和优化一个可触达亿级用户的营销平台，并引导他们的生活娱乐决策。同时，你也会直面如何精准，高效，低成本营销的挑战，也有机会接触到计算广告领域前沿的AI算法体系和大数据解决方案。你会和美团营销技术团队一起推动建立流量运营生态，支持酒旅、外卖、到店、打车、金融等业务继续快速的发展。

我们诚邀有激情、有想法、有经验、有能力的你，和我们一起并肩奋斗！参与美团点评站外广告投放体系的实现，基于大规模用户行为数据，优化在线广告算法，提升DAU, ROI, 提高在线广告的相关度、投放效果。欢迎邮件wuhongjie#meituan.com咨询。

参考文献

- [^1]: Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., ... & Anil, R. (2016, September). Wide & deep learning for recommender systems. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems (pp. 7–10). ACM.
- [^2]: Wang, R., Fu, B., Fu, G., & Wang, M. (2017, August). Deep & cross network for ad click predictions. In Proceedings of the ADKDD’17 (p. 12). ACM.
- [^3]: Guo, H., Tang, R., Ye, Y., Li, Z., & He, X. (2017). Deepfm: a factorization–machine based neural network for ctr prediction. arXiv preprint arXiv:1703.04247.
- [^4]: Lian, J., Zhou, X., Zhang, F., Chen, Z., Xie, X., & Sun, G. (2018). xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems. arXiv preprint arXiv:1803.05170.
- [^5]: Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016, November). TensorFlow: a system for large-scale machine learning. In OSDI (Vol. 16, pp. 265–283).
- [^6]: Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., ... & He, K. (2017). Accurate, large minibatch SGD: training imagenet in 1 hour. arXiv preprint arXiv:1706.02677.
- [^7]: Neill, R., Drebes, A., Pop, A. (2018). Performance Analysis of Just-in-Time Compilation for Training TensorFlow Multi-Layer Perceptrons.

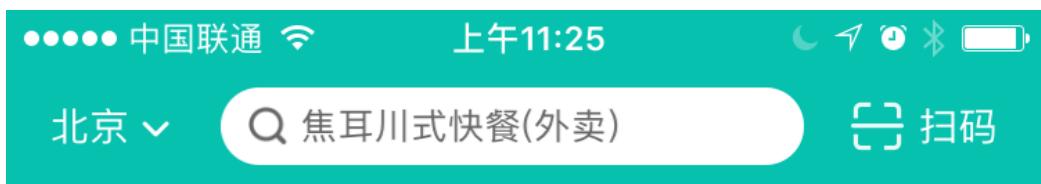
强化学习在美团“猜你喜欢”的实践

作者: 段瑾

1 概述

“猜你喜欢”是美团流量最大的推荐展位，位于首页最下方，产品形态为信息流，承担了帮助用户完成意图转化、发现兴趣、并向美团点评各个业务方导流的责任。经过多年迭代，目前“猜你喜欢”基线策略的排序模型是业界领先的流式更新的Wide&Deep模型^[1]。考虑Point-Wise模型缺少对候选集Item之间的相关性刻画，产品体验中也存在对用户意图捕捉不充分的问题，从模型、特征入手，更深入地理解时间，仍有推荐体验和效果的提升空间。近年来，强化学习在游戏、控制等领域取得了令人瞩目的成果，我们尝试利用强化学习针对以上问题进行优化，优化目标是在推荐系统与用户的多轮交互过程中的长期收益。

在过去的工作中，我们从基本的Q-Learning着手，沿着状态从低维到高维，动作从离散到连续，更新方式从离线到实时的路径进行了一些技术尝试。本文将介绍美团“猜你喜欢”展位应用强化学习的算法和工程经验。第2节介绍基于多轮交互的MDP建模，这部分和业务场景强相关，我们在用户意图建模的部分做了较多工作，初步奠定了强化学习取得正向收益的基础。第3节介绍网络结构上的优化，针对强化学习训练不稳定、难以收敛、学习效率低、要求海量训练数据的问题，我们结合线上A/B Test的线上场景改进了DDPG模型，取得了稳定的正向收益。第4节介绍轻量级实时DRL框架的工作，其中针对TensorFlow对Online Learning支持不够好和TF serving更新模型时平响骤升的问题做了一些优化。



霸夫Buff(望京店) 外卖

51分钟送达

起送价¥20|配送费¥5|美团专送

满26减10, 满40减15

月售715



水木锦堂铁板烧自助餐厅 2.5km

[2店通用] 铁板自助午餐

¥198 美食畅享

已售19208



源咖啡·简餐 外卖

47分钟送达

起送价¥20|配送费¥5|美团专送

新用户立减16元,首次使用银行... 月售57



绿茶餐厅 (北京望京新世... 外卖

63分钟送达

起送价¥20|配送费¥6|美团专送

满50减10, 满80减12

月售9228



水木锦堂铁板烧自助餐厅

2.5km



首页



附近



逛一逛



订单



我的

图1 美团首页“猜你喜欢”场景

在“猜你喜欢”展位中，用户可以通过翻页来实现与推荐系统的多轮交互，此过程中推荐系统能够感知用户的实时行为，从而更加理解用户，在接下来的交互中提供更好的体验。“猜你喜欢”用户-翻页次数的分布是一个长尾的分布，在图2中我们把用户数取了对数。可知多轮交互确实天然存在于推荐场景中。

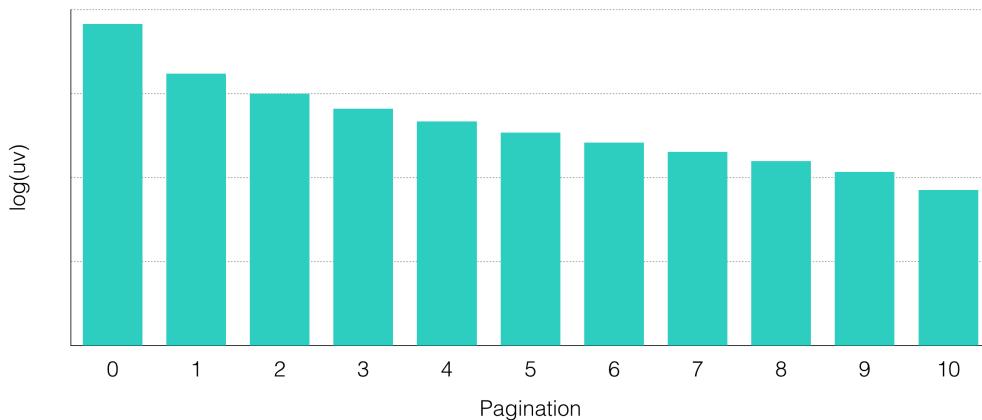


图2 “猜你喜欢”展位用户翻页情况统计

在这样的多轮交互中，我们把推荐系统看作智能体 (Agent)，用户看作环境 (Environment)，推荐系统与用户的多轮交互过程可以建模为MDP：

- State: Agent对Environment的观测，即用户的意图和所处场景。
- Action: 以List-Wise粒度对推荐列表做调整，考虑长期收益对当前决策的影响。
- Reward: 根据用户反馈给予Agent相应的奖励，为业务目标直接负责。
- $P(s,a)$: Agent在当前State s 下采取Action a 的状态转移概率。

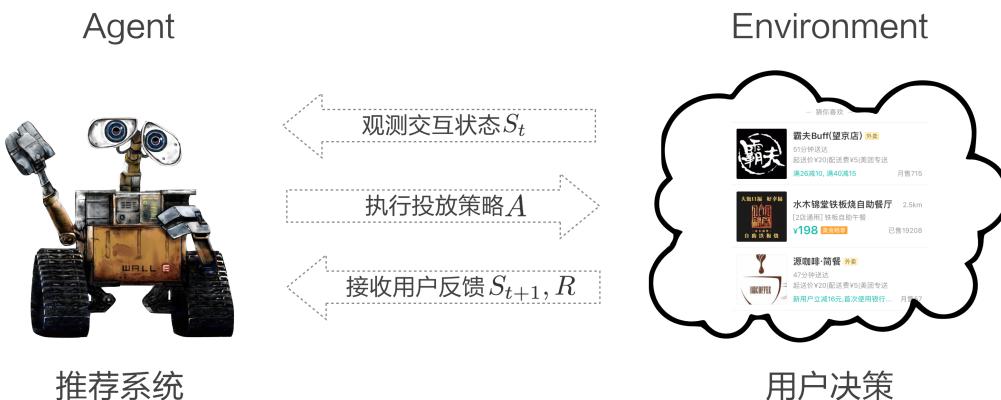


图3 推荐系统与用户交互示意图

我们的优化目标是使Agent在多轮交互中获得的收益最大化：

$$Q^*(s, a) = \max_{\pi} \mathbb{E} \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k} \mid s_t = s, a_t = a \right\}, \forall s \in S, \forall a \in A, \forall t \geq 0$$

具体而言，我们把交互过程中的MDP建模如下：

2.1 状态建模

状态来自于Agent对Environment的观察，在推荐场景下即用户的意图和所处场景，我们设计了如图4所示的网络结构来提取状态的表达。网络主要分为两个部分：把用户实时行为序列的Item Embedding作为输入，使用一维CNN学习用户实时意图的表达；推荐场景其实仍然相当依赖传统特征工程，因此我们使用Dense和Embedding特征表达用户所处的时间、地点、场景，以及更长时间周期内用户行为习惯的挖掘。

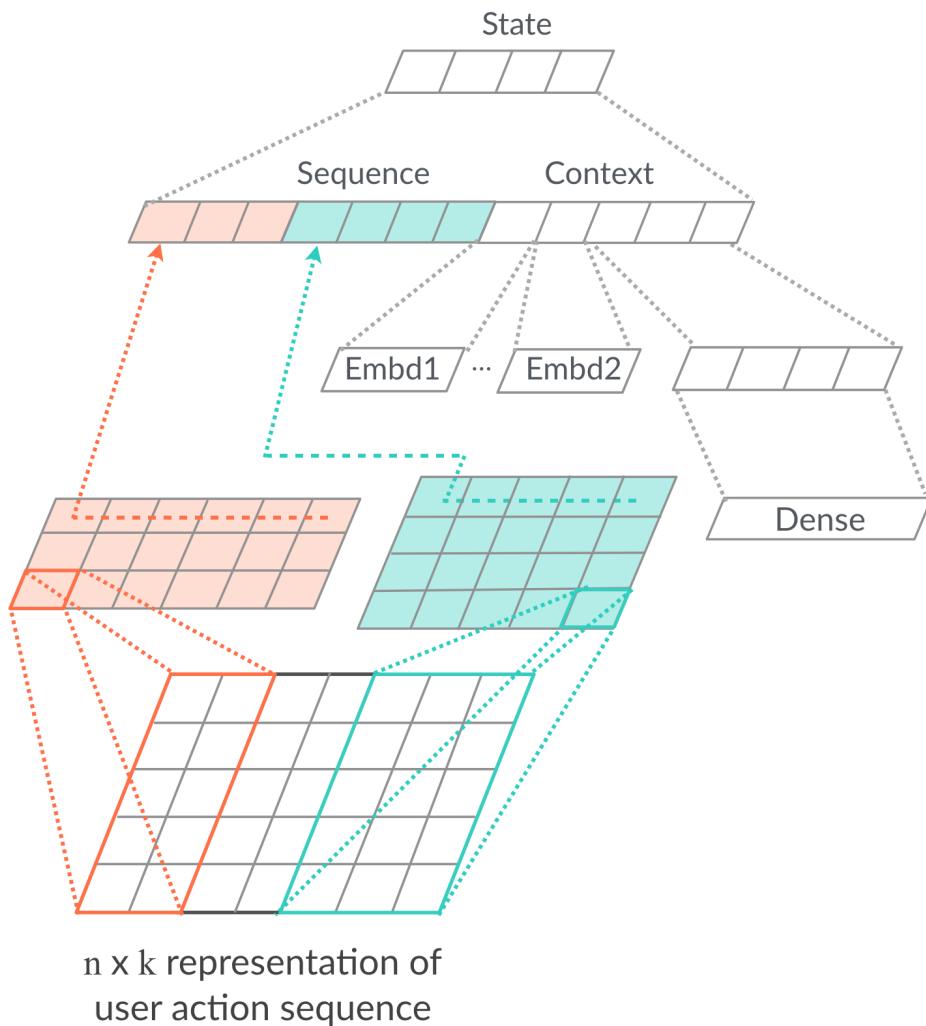


图4 状态建模网络结构

这里我们介绍一下使用Embedding特征表达用户行为习惯挖掘的Binary Sequence^[2]方法。我们通过特征工程对用户行为序列做各种维度的抽象，做成一些列离散的N进制编码，表示每一位有N种状态。例如统计用户在1H/6H/1D/3D/1W不同时间窗口内是否有点击行为编码成5位2进制数，把这些数字作为离散特征学习Embedding表达，作为一类特征处理方法。除此之外，还有点击品类是否发生转移、点击间隔的gap等等，在“猜你喜欢”场景的排序模型和强化学习状态建模中都取得了很不错的效果。原因是在行为数据非常丰富的情况下，序列模型受限于复杂度和效率，不足以充分利用这些信息，Binary Sequence可以作为一个很好的补充。

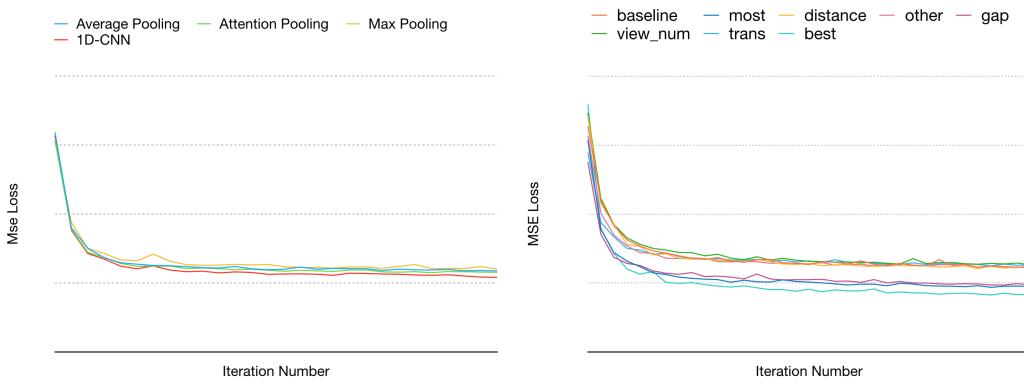


图5 序列模型和特征工程效果对照

图5左侧是序列模型的部分，分别使用不同的Pooling方式和一维CNN离线效果的对比，右侧是Dense和Embedding特征的部分，分别加入用户高频行为、距离、行为时间间隔、行为次数、意图转移等特征，以及加入所有显著正向特征的离线效果。

2.2 动作设计

“猜你喜欢”目前使用的排序模型由两个同构的Wide&Deep模型组成，分别以点击和支付作为目标训练，最后把两个模型的输出做融合。融合方法如下图所示：

$$score = click * (pay + \phi)$$

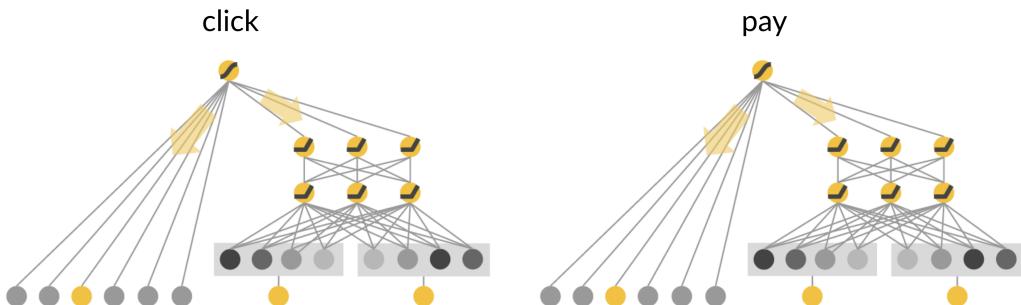


图6 排序模型示意图

超参数 ϕ 的物理意义是调整全量数据集中点击和下单模型的Trade Off，通过综合考虑点击和下单两个任务的AUC确定，没有个性化的因素。我们以此为切入点，使用Agent的动作调整融合超参数，令：

$$score = click * (pay + \phi)$$

a 是由Agent的策略生成Action，这样做有两个好处：其一，我们知道一个较优解是 $a=1$ ，这种情况下强化学习策略和基线的排序策略保持一致，由于强化学习是个不断试错的过程，我们可以很方便地初始化Agent的策略为 $a=1$ ，从而避免在实验初期伤害线上效果。其二，允许我们根据物理意义对Action做Clip，从而减轻强化学习更新过程不稳定造成的影响。

2.3 奖励塑形

“猜你喜欢”展位的优化核心指标是点击率和下单率，在每个实验分桶中分母是基本相同的，因此业务目标可以看成优化点击次数和下单次数，我们尝试将奖励塑形如下：

$$r = w_c * \sum I_{click} + w_p * \sum I_{pay}$$

相对于关注每个Item转化效率的Point Wise粒度的排序模型，强化学习的目标是最大化多轮交互中的奖励收益，为业务目标直接负责。

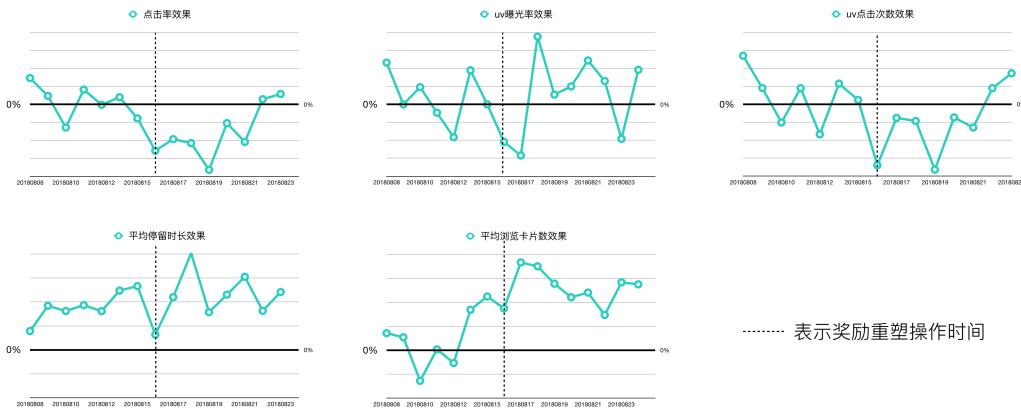


图7 加入惩罚项前后的相对效果变化

在实验过程中我们发现，强化学习的策略可能上线初期效果很好，在点击和下单指标上都取得了一定的提升，但在后续会逐渐下降，如图7前半段所示。在逐层转化效率的分析中，我们发现强化学习分桶的设备曝光率和UV维度点击率有所降低，而用户停留时长和浏览深度稳定提升，这说明Agent学习到了让用户与推荐系统更多交互，从而获取更多曝光和转化机会的策略，但这种策略对于部分强烈下单意图用户的体验是有伤害的，因为这部分用户意图转化的代价变高了，因而对展位的期望变低。针对这种情况，我们在奖励塑形中加入两个惩罚项：

1. 惩罚没有发生任何转化（点击/下单）行为的中间交互页面（penalty1），从而让模型学习用户意图转化的最短路；
2. 惩罚没有发生任何转化且用户离开的页面（penalty2），从而保护用户体验。

修正后的奖励为：

$$r = w_c * \sum I_{click} + w_p * \sum I_{pay} + penalty1 + penalty2$$

由于用户体验是时间连续的，UV维度的效果在报表上有一定的滞后性，约一周后点击率和万订单恢复到正向水平，同时用户停留时长和浏览深度有进一步提升，说明Agent确实学到了在避免伤害用户的前提下，从多轮交互中获取更多转化的策略，如图7后半段所示。

这一节我们介绍了MDP建模相关的工作。MDP跟业务场景是强相关的，经验不是很容易迁移。就本文的场景而言，我们花了较多精力做状态表达的特征，这部分工作使强化学习得到了在自己的目标上取得正向收益的能力，因此对这部分介绍比较细致。动作设计是针对多目标模型融合的场景，是个业界普遍存在并且监督学习不太适用的场景，也能体现强化学习的能力。奖励塑形是为了缩小强化学习的目标和业务目标之间的Gap，需要在数据洞察和业务理解上做一些工作。完成上述工作后强化学习在自己的目标和业务指

标上已经能取得了一些正向效果，但不够稳定。另外由于策略迭代是个Online Learning的过程，实验上线后需要实时训练一周才能收敛并观察效果，这也严重影响了我们的迭代效率。针对这些情况我们针对模型做了一些改进。

3 改进的DDPG模型

在模型方面，我们在不断改进MDP建模的过程中先后尝试了Q-Learning、DQN^[3]和DDPG^[4]模型，也面临着强化学习中普遍存在更新不够稳定、训练过程容易不收敛、学习效率较低（这里指样本利用效率低，因此需要海量样本）的问题。具体到推荐场景中，由于List-Wise维度的样本比Point-Wise少得多，以及需要真实的动作和反馈作为训练样本，因此我们只能用实验组的小流量做实时训练。这样以来训练数据量相对就比较少，每天仅有几十万，迭代效率较低。为此我们对网络结构做了一些改进，包括引入具体的Advantage函数、State权值共享、On-Policy策略的优化，结合线上A/B Test框架做了十数倍的数据增强，以及对预训练的支持。接下来我们以DDPG为基石，介绍模型改进的工作。

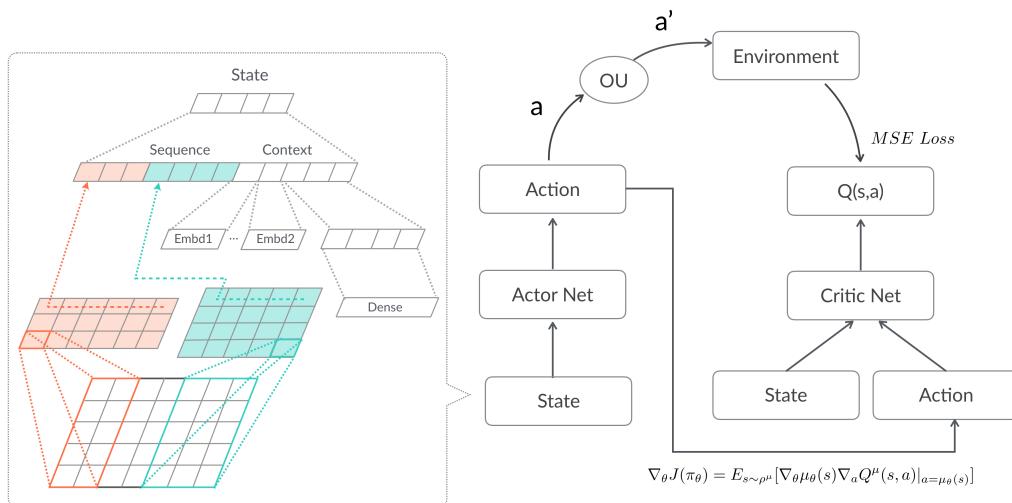


图8 DDPG模型

如图8所示，基本的DDPG是Actor–Critic架构。线上使用Actor网络，预测当前State下最好的动作a，并通过Ornstein–Uhlenbeck过程对预测的Action加一个随机噪声得到a'，从而达到在最优策略附近探索的目的。将a'作用于线上，并从用户(Environment)获得相应的收益。训练过程中，Critic学习估计当前状态s下采取动作a获得的收益，使用MSE作为Loss Function：

$$L = \mathbb{E}[(r + \gamma \max_{a'} Q(s', a') - Q(s, a))^2]$$

对参数求导：

$$\frac{\partial L(w)}{\partial w} = \mathbb{E}[(r + \gamma \max_{a'} Q(s', a') - Q(s, a)) \frac{\partial Q(s, a, w)}{\partial w}]$$

Actor使用Critic反向传播的策略梯度，使用梯度上升的方法最大化Q估计，从而不断优化策略：

$$\nabla_{\theta} J(\pi_{\theta}) = E_{s \sim \rho^{\mu}} [\nabla_{\theta} \mu_{\theta}(s) \nabla_a Q^{\mu}(s, a) |_{a=\mu_{\theta}(s)}]$$

在确定性策略梯度的公式中， θ 是策略的参数，Agent 将使用策略 $\mu_{\theta}(s)$ 在状态 s 生成动作 a ， ρ^{μ} 表示该策略下的状态转移概率。在整个学习过程中，我们不需要真的估计策略的价值，只需要根据Critic返回的策略梯度最大化Q估计。Critic不断优化自己对 $Q(s, a)$ 的估计，Actor通过Critic的判断的梯度，求解更好的策略函数。如此往复，直到Actor收敛到最优策略的同时，Critic收敛到最准确的 $Q(s, a)$ 估计。

接下来基于这些我们介绍的DDPG模型改进的工作。

3.1 Advantage函数

借鉴DDQN^[5]的优势函数Advantage的思路，我们把critic估计的 $Q(s, a)$ 拆分成两个部分：只与状态相关的 $V(s)$ ，与状态、动作都相关的Advantage函数 $A(s, a)$ ，有 $Q(s, a) = V(s) + A(s, a)$ ，这样能够缓解critic对 Q 过高估计的问题。具体到推荐环境中，我们的策略只是对排序模型的融合参数做调整，收益主要是由状态决定的。

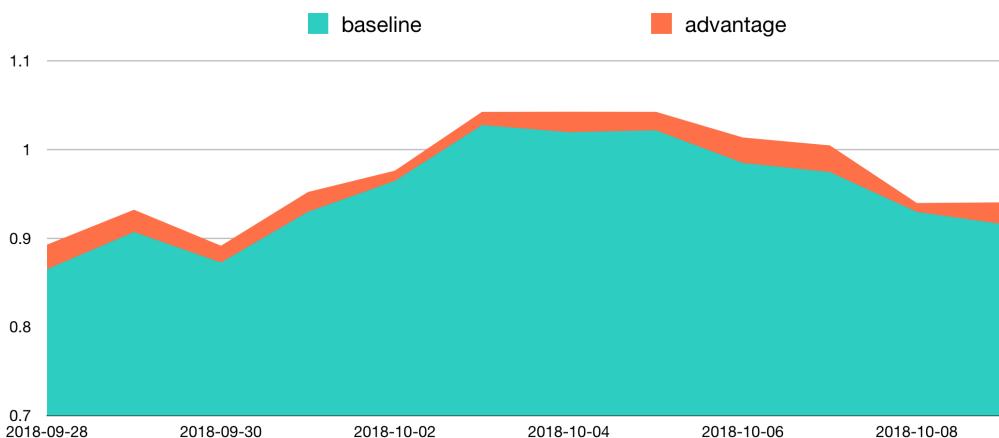


图9 实验组与基线的Q值对比

如图9所示，在实际实验中观察 $V(s)$ 和 $A(s, a)$ 均值的比值大约为 97:3，可以验证我们的这一判断。在实际训练过程中，我们先根据状态和收益训练 $V(s)$ ，再使用 $Q(s, a) - V(s)$ 的残差训练 $A(s, a)$ ，很大程度上提升了训练稳定性，并且我们可以通过残差较为直观地观测到当前策略是否优于基线。图8中 $A(s, a)$ 稳定大于 0，可以认为强化学习在自己的目标上取得了稳定的正向收益。

3.2 State权值共享

受A3C^[6]网络的启发，我们观察到DDPG的网络中Actor和Critic网络中都有State的表达，而在我们的场景中大部分参数都集中在State的部分，在十万量级，其他参数只有数千，因此我们尝试把State部分的权重做共享，这样可以减少约一半的训练参数。

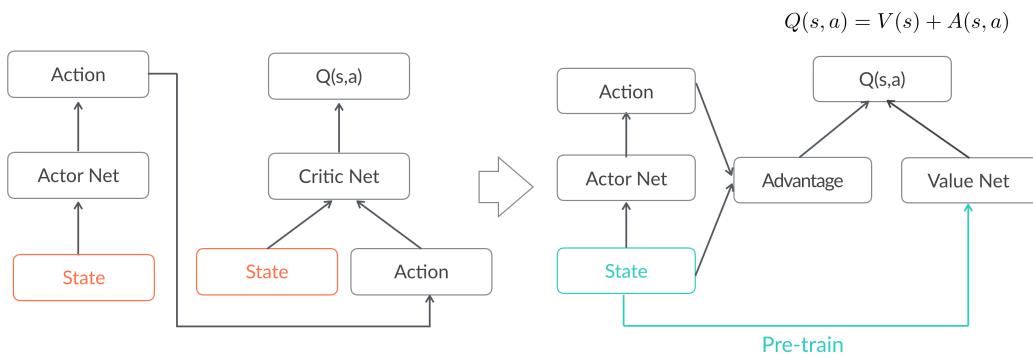


图10 使用advantage函数并做state权值共享

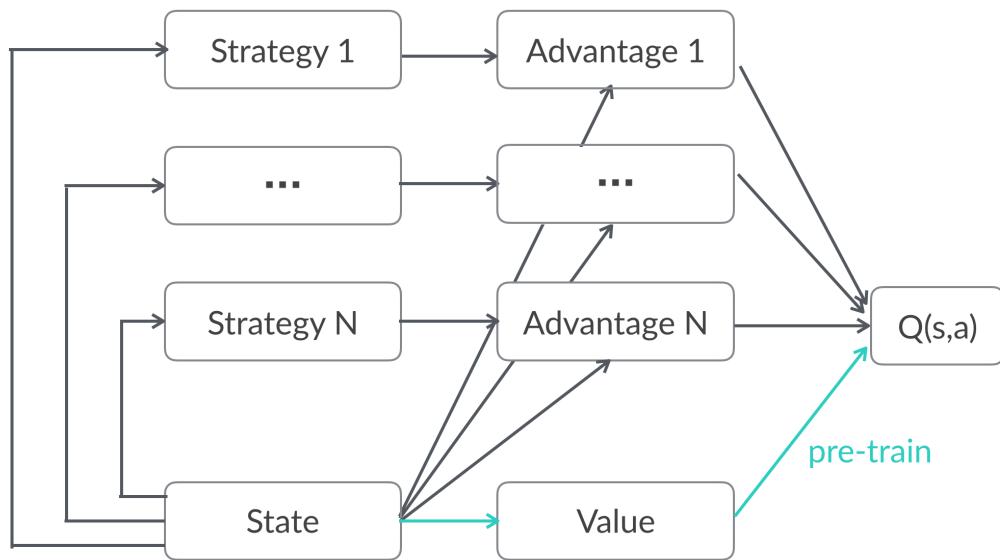
改进后的网络结构如图10所示。对于这个网络结构，我们注意到有 $V(s)$ 的分支和动作不相关，意即我们不需要具体的Action也可以学习该State下Q的期望，这就允许我们在线下使用基线策略千万级的数据量做**预训练**，线上也同时使用基线和实验流量做实时更新，从而提升训练的效果和稳定性。又因为这条更新路径包含了所有State的参数，模型的大部分参数都可以得到充分的预训练，只有Action相关的参数必须依赖Online Learning的部分，这就大幅提高了我们的实验迭代效率。原来我们需要上线后等待一周训练再观察效果，改进后上线第二天就可以开始观察效果。

3.3 On-policy

在A2C^[7]的论文里作者论述了他们的见解：同步A2C实现比异步实现的A3C表现要好。目前尚未看到任何证据证明异步引入的噪声能够提供任何性能收益，因此为了提升训练效率，我们采取了这个做法，使用同一套参数估计 Q_{t+1} 和更新 Q_t ，从而使模型参数再次减半。

3.4 扩展到多组并行策略

考虑多组强化学习实验同时在线的情况，结合A/B Test环境特点，我们把以上网络框架扩展到多Agent的情况。



$$Q(s, a) = V(s) + A(s, a)$$

图11 支持多组线上实验DDPG模型

如图11所示，线上多组实验共享State表达和 $V(s)$ 的估计，每个策略训练自己的 $A(s,a)$ 网络且能快速收敛，这样的结构一方面使训练过程更加稳定，另一方面为强化学习策略全量提供了可能性。

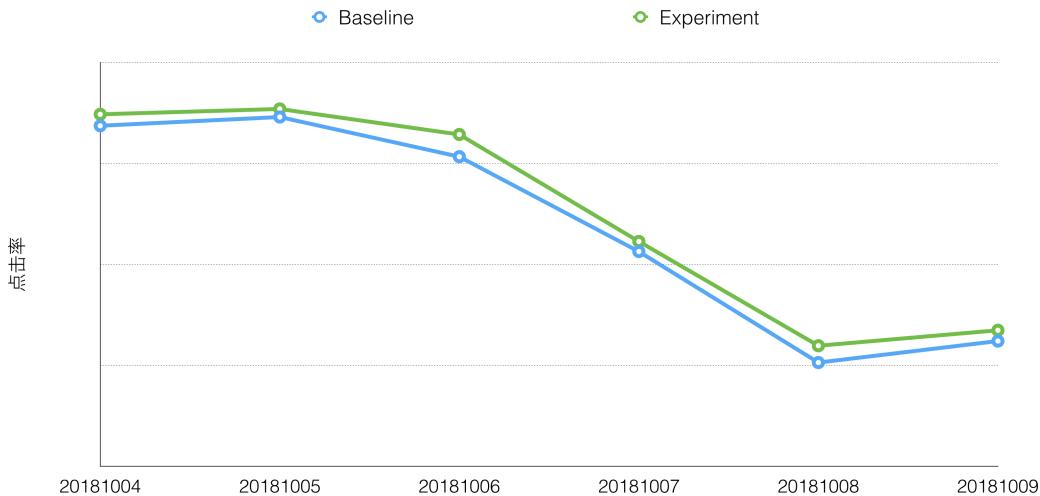


图12 点击率分天实验效果

在DDPG的改造工作中，我们使用Advantage函数获得更稳定的训练过程和策略梯度。State权值共享和On-Policy方法使我们的模型参数减少75%。Advantage函数和State权值共享结合，允许我们使用基线策略样本做数据增强，使每天的训练样本从十万量级扩展到百万量级，同时充分的预训练保证策略上线后能迅速收敛。经过这些努力，强化学习线上实验取得了稳定的正向效果，在下单率效果持平的情况下，周效果点击率相对提升0.5%，平均停留时长相对提升0.3%，浏览深度相对提升0.3%。修改过的模型与A2C的主要区别是我们仍然使用确定性策略梯度，这样我们可以少估计一个动作的分布，即随机策略方差降至0的特例。图12表明强化实习的效果是稳定的，由于“猜你喜欢”的排序模型已经是业界领先的流式DNN模型，我们认为这个提升是较为显著的。

4 基于TF的轻量级实时DRL系统

强化学习通常是在一次次试错（Trial-and-Error）中学习，实时地改进策略并获得反馈能大幅提升学习效率，尤其在连续策略中。这一点在游戏场景下很容易理解，相应地，我们也在推荐系统中构建了实时深度学习系统，让策略更新更加高效。为了支持实时更新的DRL模型和高效实验，我们针对Online Learning的需求，基于TensorFlow及TF Serving做了一些改进和优化，设计并实现了一套特征配置化的实时更新的DRL框架，在实验迭代过程中沉淀了DQN、DDQN、DDPG、A3C、A2C、PPO^[8]等模型。系统架构如图13所示：

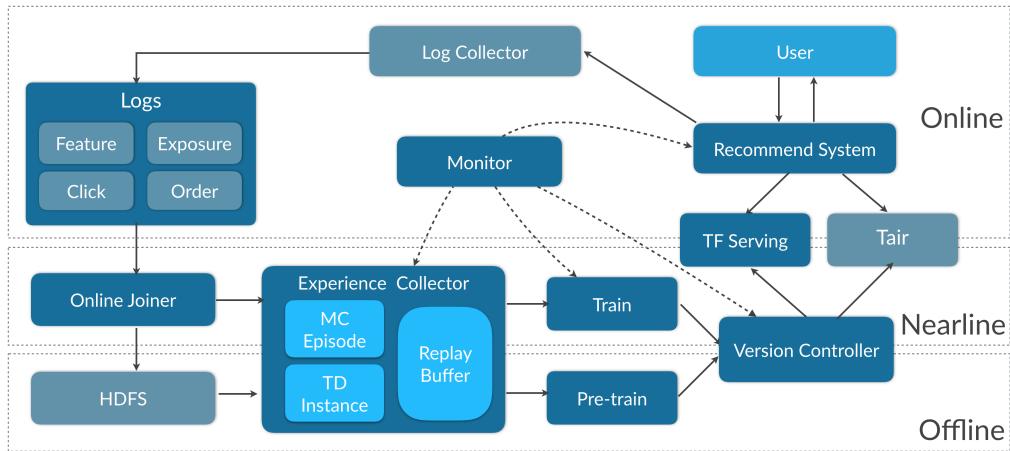


图13 实时更新的强化学习框架

训练部分工作流如下：

1. Online Joiner从Kafka中实时收集特征和用户反馈，拼接成Point-Wise粒度的Label-Feature样本，并把样本输出到Kafka和HDFS，分别支持在线和离线更新。
2. Experience Collector收集上述样本，合并为List-Wise的请求粒度，并根据请求时间戳拼接成[]列表形式的MC Episode，再经过状态转移计算拆分成形式的TD Instance，输出MC或TD格式的样本支持RL训练。
3. Trainer做输入特征的预处理，使用TensorFlow训练DRL模型。
4. Version Controller负责调度任务保证实效性和质量，并把训练完成且指标符合预期模型推送到TF Serving和Tair中，这部分只需要Actor相关的参数。Tair作为弥补TF在Online Learning短板辅助的PS，后面会具体介绍。
5. Monitor监控和记录整个训练流程中的数据量和训练指标，对不符合预期的情况发出线上告警。
6. 新模型上线前会先做离线的Pre-Train，使用基线策略的数据学习State的表达和Value net。上线后实时同时更新Actor，Advantage和Value的参数。

线上预测部分，推荐系统的Agent从Tair获取预处理参数，并将处理后的特征喂给TF Serving做前向传播，得到Action并对展现给用户的排序结果做相应的干预。

针对TensorFlow对Online Learning支持比较弱，Serving对千万级Embedding处理效率不高的问题，我们做了一些改进：

- 在线上特征的分布会随时间而改变，针对Dense特征我们自己维护增量的Z-Score算法对特征做预处理。
- Embedding特征的输入维度也经常发生变化，而TF不支持变长的Input Dimension，为此我们维护了全量的ID-Embedding映射，每次训练让模型加载当前样本集合中的高频Embedding。
- 千万级Item Embedding会大幅降低训练和预测的效率，为此我们把这部分映射在预处理中，把映射后的矩阵直接作为CNN的输入。
- 为了提升特征工程的实验效率，支持特征配置化生成模型结构。
- 此外，TF serving在更新模型一两分钟内响应时间会骤然升高，导致很多请求超时，原因有二，其一，serving的模型加载和请求共用一个线程池，导致切换模型使阻塞处理请求；其二，计算图初始化是lazy的，这样新模型后的第一次请求需要等待计算图初始化。这个问题在更新模型频Low对online learning支持比较弱的Online Learning场景影响较大，我们采用切分线程池和warm up初始化的方式解决。更具体的方案和效果可以参考美团另一篇技术博客^[9]。

5 总结和展望

强化学习是目前深度学习领域发展最快的方向之一，其与推荐系统和排序模型的结合也有更多价值等待发掘。本文介绍了强化学习在美团“猜你喜欢”排序场景落地的工作，包括根据业务场景不断调整的MDP建模，使强化学习能够取得一定的正向收益；通过改进DDPG做数据增强，提升模型的鲁棒性和实验效率，从而取得稳定的正向收益；以及基于TensorFlow的实时DRL框架，为高效并行策略迭代提供了基础。

经过一段时间的迭代优化，我们在强化学习方面也积累了一些经验，与传统的监督学习相比，强化学习的价值主要体现在：

1. 灵活的奖励塑形，能支持各种业务目标建模，包括不限于点击率、转化率、GMV、停留时长、浏览深度等，支持多目标融合，为业务目标直接负责。
2. 充满想象空间的动作设计，不需要直接的Label，而是通过网络来生成和评价策略，适合作为监督学习的补充。这点和GAN有相通之处。
3. 考虑优化长期收益对当前决策造成的影响，Agent与Environment交互频繁的场景更加能体现强化学习的价值。

同时强化学习作为机器学习的一个分支，很多机器学习的经验仍然适用于此。比如数据和特征决定效果的上限，模型和算法只是不断逼近它。对于强化学习而言特征空间主要包含在状态的建模中，我们强烈建议在状态建模上多做一些尝试，并信任模型有能力从中做出判断。再如，使用更多的训练数据降低经验风险，更少的参数降低结构风险的思路对强化学习仍然适用，因此我们认为DDPG的改进工作能够推广到不同业务的线上A/B Test场景中。此外，我们在训练过程中也遇到了强化学习对随机性敏感的问题^[10]，为此我们线上使用了多组随机种子同时训练，选择表现最好的一组参数用于实际参数更新。

在目前的方案中，我们尝试的Action是调整模型融合参数，主要考虑这是个排序问题中比较通用的场景，也适合体现强化学习的能力，而实际上对排序结果的干预能力是比较有限的。未来我们会探索不同品类、位置、价格区间等跟用户意图场景强相关属性的召回个数，调整排序模型隐层参数等方向。另外在解决学习效率低下的问题上，还将尝试Priority Sampling 提高样本利用效率，Curious Networks提升探索效率等方法。也欢迎对强化学习感兴趣的朋友们与我们联系，一起交流探索强化学习在工业界的应用与发展，同时对文章的错漏之处也欢迎大家批评指正。

参考文献

- [1] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, and Hemal Shah. Wide & deep learning for recommender systems. CoRR, 2016.
- [2] Yan, P., Zhou, X., Duan, Y. E-commerce item recommendation based on field-aware Factorization Machine. In: Proceedings of the 2015 International ACM Recommender Systems Challenge, 2015.
- [3] Mnih, Volodymyr, Kavukcuoglu, Koray, Silver, David, Rusu, Andrei A, Veness, Joel, Bellemare, Marc G, Graves, Alex, Riedmiller, Martin, Fidjeland, Andreas K, Ostrovski, Georg, et al. Humanlevel control through deep reinforcement learning. Nature, 2015.
- [4] Lillicrap, T., Hunt, J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. Continuous control with deep reinforcement learning. In International Conference, 2015 on Learning Representations, 2016.
- [5] Wang, Z., de Freitas, N., and Lanctot, M. Dueling network architectures for deep reinforcementlearning. Technical report, 2015.

- [6] Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Timothy P. Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. ICML, 2016
- [7] Y. Wu, E. Mansimov, S. Liao, R. Grosse, and J. Ba. Scalable trust-region method for deep reinforcement learning using Kronecker-factored approximation. arXiv preprint arXiv:1708.05144, 2017.
- [8] Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347, 2017
- [9] 仲达, 鸿杰, 廷稳. 基于TensorFlow Serving的深度学习在线预估. MT Bolg, 2018
- [10] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger. Deep reinforcement learning that matters. arXiv:1709.06560, 2017.

作者简介

- 段瑾, 2015年加入美团点评, 目前负责强化学习在推荐场景的落地工作。

深度学习在文本领域的应用

作者: 美团算法团队

背景

近几年以深度学习技术为核心的人工智能得到广泛的关注，无论是学术界还是工业界，它们都把深度学习作为研究应用的焦点。而深度学习技术突飞猛进的发展离不开海量数据的积累、计算能力的提升和算法模型的改进。本文主要介绍深度学习技术在文本领域的应用，文本领域大致可分为4个维度：词、句子、篇章、系统级应用。

- 词。分词方面，从最经典的前后向匹配到条件随机场（Conditional Random Field, CRF）序列标注，到现在Bi-LSTM+CRF模型，已经不需要设计特征，从字粒度就能做到最好的序列标注效果，并且可以推广到文本中序列标注问题上，比如词性标注和专门识别等。
- 句子。Parser方面，除词粒度介绍的深度学习序列标注外，还可以使用深度学习模型改善Shift–Reduce中间分类判断效果；句子生成方面，可以通过序列到序列（Seq2Seq）模型训练自动的句子生成器，可用于闲聊或者句子改写等场景。
- 篇章。情感分析方面，可以使用卷积神经网络对输入文本直接建模预测情感标签；阅读理解方面，可以设计具有记忆功能的循环神经网络来做阅读理解，这个也是近年非常热的研究问题。
- 系统级应用。信息检索方面，把深度学习技术用在文本匹配做相似度计算，可以通过BOW、卷积神经网络或循环神经网络表示再学习匹配关系（如DSSM系列），还有使用DNN做排序模型（如Google的Wide & Deep等，后面会重点介绍）；机器翻译方面，源于Seq2Seq模型到Stack–LSTM + Attention等多层LSTM网络，使得基于词的统计机器翻译模型已经被基于神经网络的翻译模型超越，并且已经应用到产品中，比如谷歌翻译、百度翻译、有道翻译等；智能交互方面，在做闲聊、对话、问答等系统时深度学习在分类、状态管理（如深度强化学习）、回复生成等环节都有很好的应用。

总之，上面这些文本领域的深度学习应用只是冰山一角，深度学习应用还有知识图谱、自动摘要、语音、图像文本生成等。总体趋势是，各文本研究和应用的方向都在尝试深度学习技术，并分别取得了进展。在文本领域，如果想跟图像、语音那样取得突破性进展还面临重重困难，如不同任务的大规模标注数据缺乏，如何建模能捕捉语言的逻辑性以及所蕴含的地域、文化特色等等。限于篇幅，本文只对美团在文本中应用较多的文本匹配、排序模型进行介绍。

基于深度学习的文本匹配

文本匹配在很多领域都有用到，尤其是信息检索相关场景，比如搜索的Query和Doc、广告中Query–Ad、搜索Suggestion中Query前缀和Query（见图1）、关键词推荐中Query和Query、文档去重时Doc和Doc等。



图1 美团Query-POI和Query Suggestion的文本匹配场景

文本匹配主要研究计算两段文本的相似度问题。相似度问题包含两层：一是两段文本如何表示可使得计算机方便处理，这需要研究不同的表示方法效果的区别；二是如何定义相似度来作为优化目标，如语义匹配相似度、点击关系相似度、用户行为相似度等，这和业务场景关系很紧密。

在解决这两个问题过程中会遇到很多难题，其中一个难题就是设计模型如何充分考虑语义。因为中文的多义词、同义词非常普遍，它们在不同的语境中表达的含义是不一样的。比如苹果多少钱一台？苹果多少钱一斤？对于前者，根据“一台”能看出它是指苹果品牌的电子设备，后者则是指水果。当然，还有很多语言现象更加困难，比如语气、语境、口语的不同表述等。

文本的表示和匹配是本节的主线，如何做到语义层面匹配就成为本节的主旨。受到整体技术的演进影响，文本的匹配技术同样有一个顺应时代的技术历程，如图2所示。

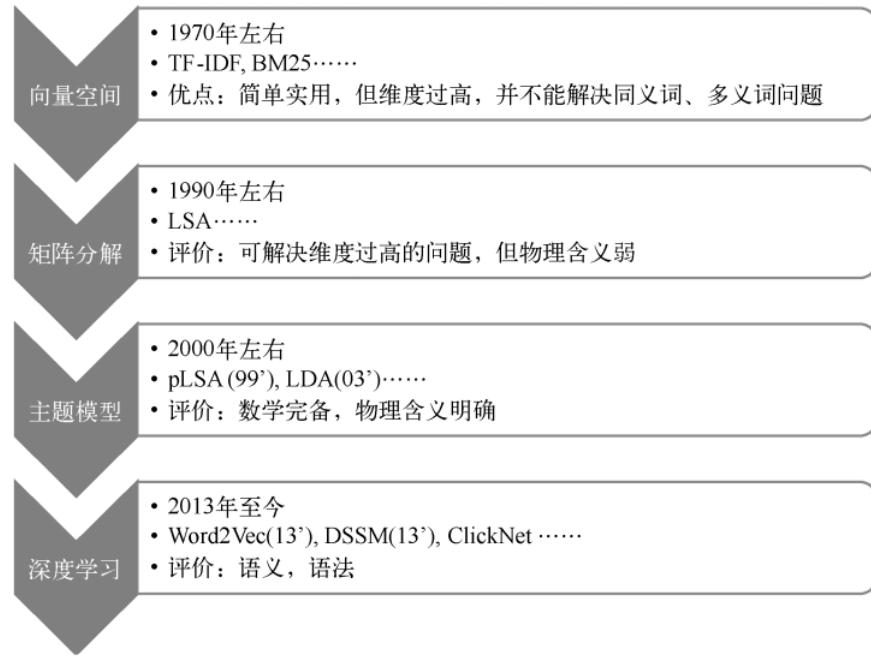


图2 语义表示匹配模型演进历程

1. 向量空间

1970年左右提出的向量空间模型，就是把文档表示词表维度的向量通过TF-IDF计算出词的权重，比如一种标准词表包含词ID、词和IDF，另一种是停用词表，如图3所示。

词ID	词	词IDF值	停用词表
0	金百万	3.0	的
1	价格	1.5	怎么
2	酒店	2.1	如何
...
N	丽江	4.1	是

图3 向量空间模型示例

对文本“丽江的酒店价格”分词去除停用词后，得到丽江、酒店、价格，词出现次数是1，查表IDF得到这句文本的表示：[0, 1.5, 2.1, 0, 0, ..., 0, 4.1]。其中权重使用的是TFxIDF，TF是Term在文本里的频次，IDF是逆文档频次，两者定义有很多形式，如图4所示。这里使用第二种定义。

TF变量权重		IDF变量权重	
加权机制	TF权重	加权机制	IDF权重
二元	$0, 1$	一元	1
原始数量	$f_{t,d}$	倒排文档频率	$\log \frac{N}{n_t} = -\log \frac{n_t}{N}$
Term频率	$f_{t,d} / \sum_{t' \in d} f_{t',d}$	倒排文档频率平滑	$\log \left(1 + \frac{N}{n_t} \right)$
log规范化	$1 + \log(f_{t,d})$	最大倒排文档频率	$\log \left(\frac{\max_{\{t' \in d\}} n_{t'}}{1 + n_t} \right)$
0.5浮点数规范化	$0.5 + 0.5 \cdot \frac{f_{t,d}}{\max_{\{t' \in d\}} f_{t',d}}$	概率倒排文档频率	$\log \frac{N - n_t}{n_t}$
K值浮点数规范化	$K + (1 - K) \frac{f_{t,d}}{\max_{\{t' \in d\}} f_{t',d}}$		

图4 TF和IDF的常用定义

向量空间模型用高维稀疏向量来表示文档，简单明了。对应维度使用TF-IDF计算，从信息论角度包含了词和文档的点互信息熵，以及文档的信息编码长度。文档有了向量表示，那么如何计算相似度？度量的公式有Jaccard、Cosine、Euclidean distance、BM25等，其中BM25是衡量文档匹配相似度非常经典的方法，公式如下：

$$\text{score}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}} \right)}, \quad \text{IDF}(q_i) = \log \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5}$$

其中， f 是TF， $b = 0.75$ ， $k_1 \in [1.2, 2.0]$ 。

向量空间模型虽然不能包含同义词、多义词的信息，且维度随着词典增大变得很大，但因为它简单明了，效果不错，目前仍然是各检索系统必备的特征。

2. 矩阵分解

向量空间模型的高维度对语义信息刻画不好，文档集合会表示成高维稀疏大矩阵。1990年左右，有人研究通过矩阵分解的方法，把高维稀疏矩阵分解成两个狭长小矩阵，而这两个低维矩阵包含了语义信息，这个过程即潜在语义分析。

假设有N篇文档，共有V个词，用TF-IDF的向量空间表示一个 $N \times V$ 的稀疏矩阵X，词和文档的语义分析就在这个共现矩阵上操作。这个共现矩阵通过奇异值分解，能变成三个矩阵，狭长矩阵U的维度是 $N \times K$ ，V的维度是 $K \times V$ ，中间是对角矩阵 $K \times K$ 维，如图5所示。

经过分解后，每个文档由K维向量表示，($K \ll V$)，代表了潜在语义信息，可以看成是文档需要表达的语义空间表示。V矩阵代表词在潜空间上的分布都是通过共现矩阵分解得到的。

潜在语义分析能对文档或者词做低维度语义表示，在做匹配时其性能较高（比如文档有效词数大于K），它包含语义信息，对于语义相同的一些文档较准确。但是，潜在语义分析对多义词语义的建模不好，并且K维语义向量完全基于数学分解得到，物理含义不明确。因此，在2000年左右，为解决上述问题，主题模型出现了。

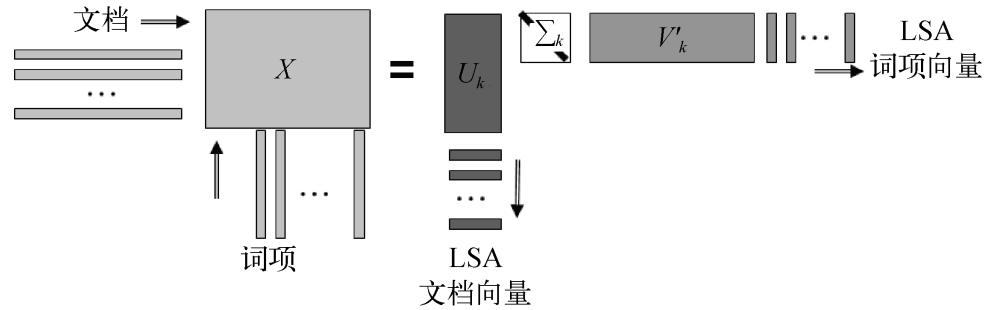


图5 潜在语义分析对词文档共现矩阵的分解示意

3. 主题模型

2000~2015年，以概率图模型为基础的主题模型掀起了一股热潮，那么究竟这种模型有什么吸引大家的优势呢？

pLSA (Probabilistic Latent Semantic Analysis)

pLSA在潜在语义分析之上引入了主题概念。它是一种语义含义，对文档的主题建模不再是矩阵分解，而是概率分布（比如多项式分布），这样就能解决多义词的分布问题，并且主题是有明确含义的。但这种分析的基础仍然是文档和词的共现频率，分析的目标是建立词/文档与这些潜在主题的关系，而这种潜在主题进而成为语义关联的一种桥梁。这种转变过渡可如图6所示。

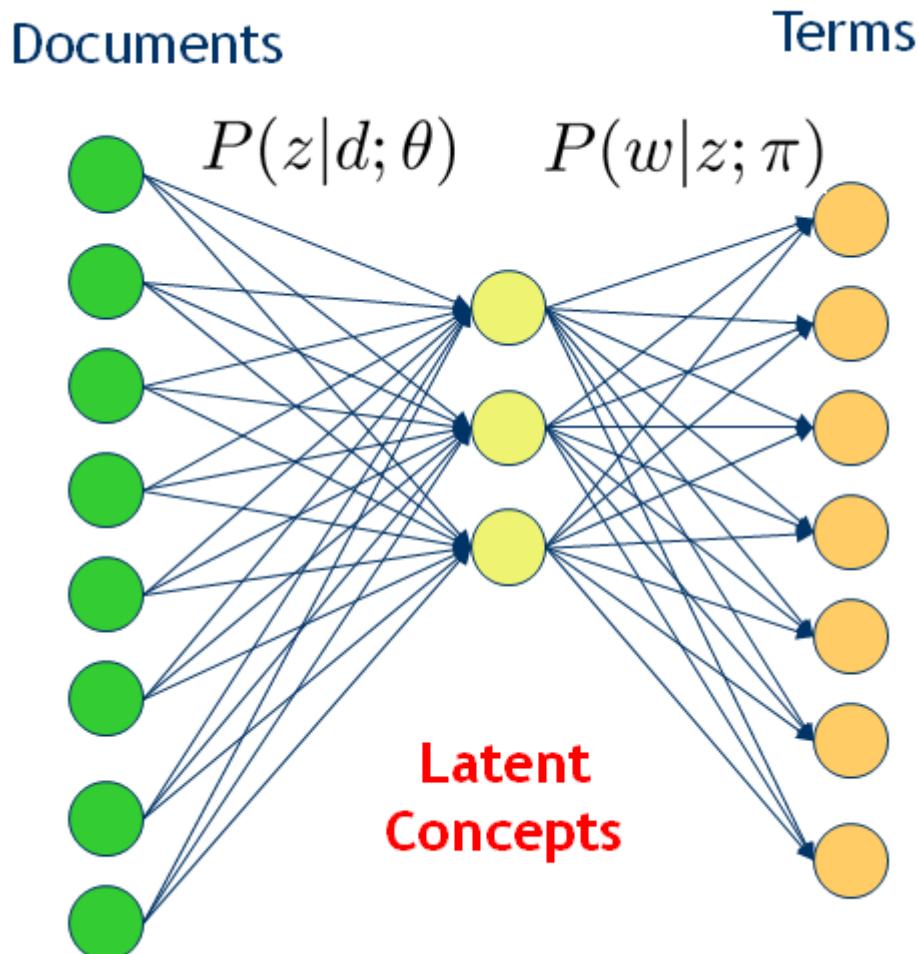


图6 从词文档共现到引入潜在主题的概率模型

假设每篇文章都由若干主题构成，每个主题的概率是 $p(z|d)$ ，在给定主题的条件下，每个词都以一定的概率 $p(w|z)$ 产生。这样，文档和词的共现可以用一种产生式的方式来描述：

$$p(d, w) = p(d) \sum_{k=1}^{k=T} p(z_k|d) p(w|z_k)$$

其概率图模型如图7所示：

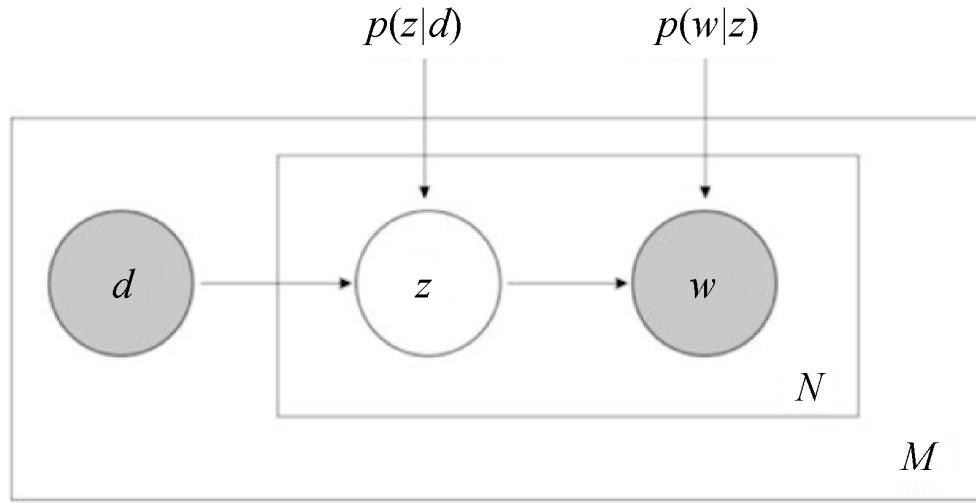


图7 pLSA概率图模型

可以通过EM算法把 $p(z|d)$ 和 $p(w|z)$ 作为参数来学习，具体算法参考Thomas Hofmann的pLSA论文。需要学习的 $p(z|d)$ 参数数目是主题数和文档数乘的关系， $p(w|z)$ 是词表数乘主题数的关系，参数空间很大，容易过拟合。因而我们引入多项式分布的共轭分布来做贝叶斯建模，即LDA使用的方法。

LDA (Latent Dirichlet Allocation)

如果说pLSA是频度学派代表，那LDA就是贝叶斯学派代表。LDA通过引入Dirichlet分布作为多项式共轭先验，在数学上完整解释了一个文档生成过程，其概率图模型如图8所示。

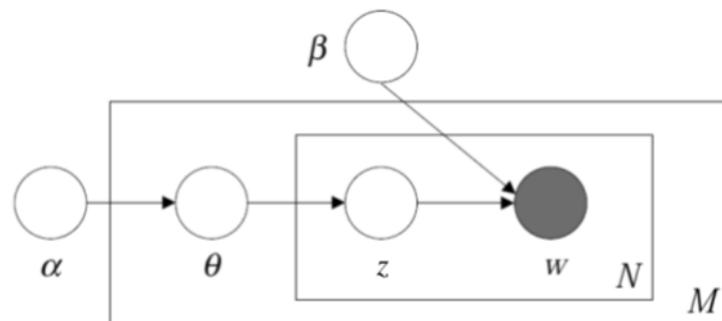


图8 LDA概率图模型

和pLSA概率图模型不太一样，LDA概率图模型引入了两个随机变量 α 和 β ，它们就是控制参数分布的分布，即文档-主题符合多项式分布。这个多项式分布的产生受Dirichlet先验分布控制，可以使用变分期望最大化（Variational EM）和吉布斯采样（Gibbs Sampling）来推导参数，这里不展开叙述。

总体来讲，主题模型引入了“Topic”这个有物理含义的概念，并且模型通过共现信息能学到同义、多义、语义相关等信息。得到的主题概率分布作为表示，变得更加合理有意义。有了文档的表示，在匹配时，我们不仅可以使用之前的度量方式，还可以引入KL等度量分布的公式，这在文本匹配领域应用很多。当然，主题模型会存在一些问题，比如对短文本推断效果不好、训练参数多速度慢、引入随机过程建模避免主题数目人工设定不合理问题等。随着研究进一步发展，这些问题基本都有较好解决，比如针对训练速度慢的问题，从LDA到SparseLDA、AliasLDA，再到LightLDA、WarpLDA等，采样速度从 $O(K)$ 降低 $O(1)$ 到。

4. 深度学习

2013年，Tomas Mikolov发表了Word2Vec相关的论文，提出的两个模型CBOW（Continuous Bag of Words，连续词袋）和Skip-Gram能极快地训练出词嵌入，并且能对词向量加减运算，这得到了广泛关注。在这项工作之前，神经网络模型经历了很长的演进历程。这里先介绍2003年Yoshua Bengio使用神经网络做语言模型的工作，Word2Vec也是众多改进之一。

神经网络语言模型

在2003年，Yoshua Bengio使用神经网络来训练语言模型比N-Gram的效果好很多，网络结构如图9所示。输入是N-Gram的词，预测下一个词。前n个词通过词向量矩阵Matrix C（维度： $n*emb_size$ ）查出该词的稠密向量 $C(w(t-1)), C(w(t-2))$ ；再分别连接到隐含层（Hidden Layer）做非线性变换；再和输出层连接做Softmax预测下一个词的概率；训练时根据最外层误差反向传播以调节网络权重。可以看出，该模型的训练复杂度为 $O(nxemb_size + nxemb_size*hidden_size + hidden_size*output_size)$ ，其中n为5~10， emb_size 为64~1024， $hidden_size$ 为64~1023， $output_size$ 是词表大小，比如为 10^7 。因为Softmax在概率归一化时，需要所有词的值，所以复杂度主要体现在最后一层。从此以后，提出了很多优化算法，比如Hierarchical Softmax、噪声对比估计（Noise Contrastive Estimation）等。

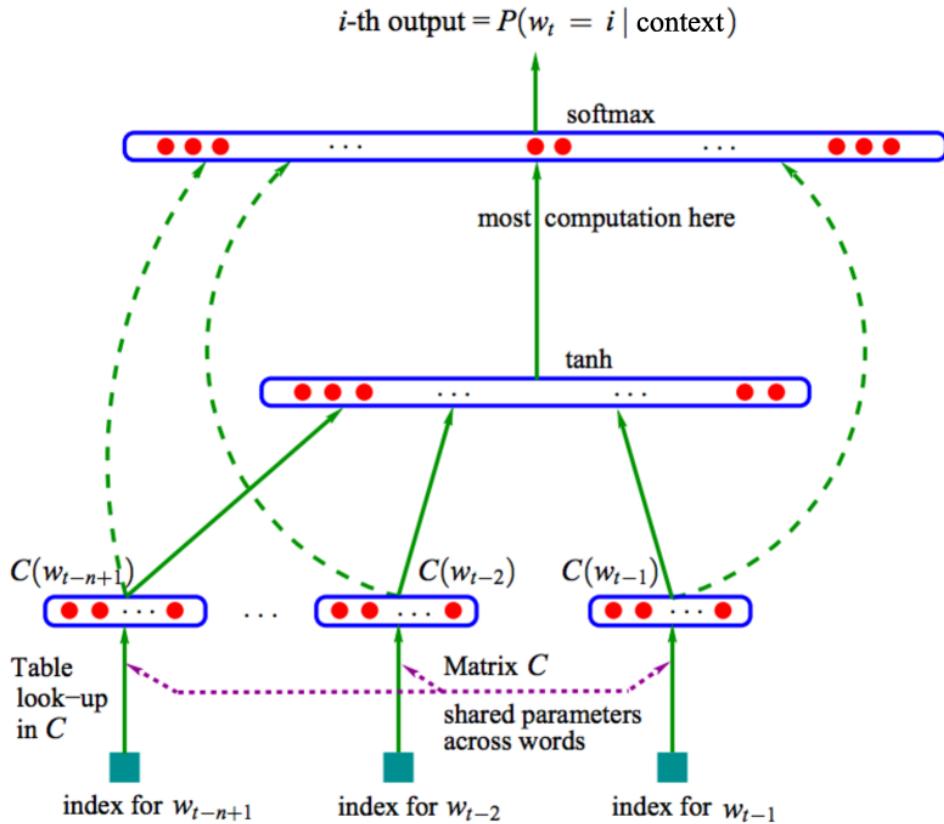


图9 神经网络语言模型结构图

Word2Vec

Word2Vec的网络结构有CBOW和Skip-Gram两种，如图10所示。相比NNLM，Word2Vec减少了隐含层，只有投影层。输出层是树状的Softmax，对每个词做哈夫曼编码，预测词时只需要对路径上的0、1编码做预测，从而复杂度从 $O(V)$ 降低到 $O(\log(V))$ 。

以CBOW为例，算法流程如下：

(1) 上下文词（窗口大小是Win）的词向量对应维相加映射到投影层。(2) 投影层经Sigmoid变换后预测当前词的编码路径（哈夫曼树）。(3) 使用交叉熵损失函数（Cross Entropy Loss）反向传播，更新Embedding层参数和中间层参数。(4) 训练使用反向传播机制，优化方法使用SGD。

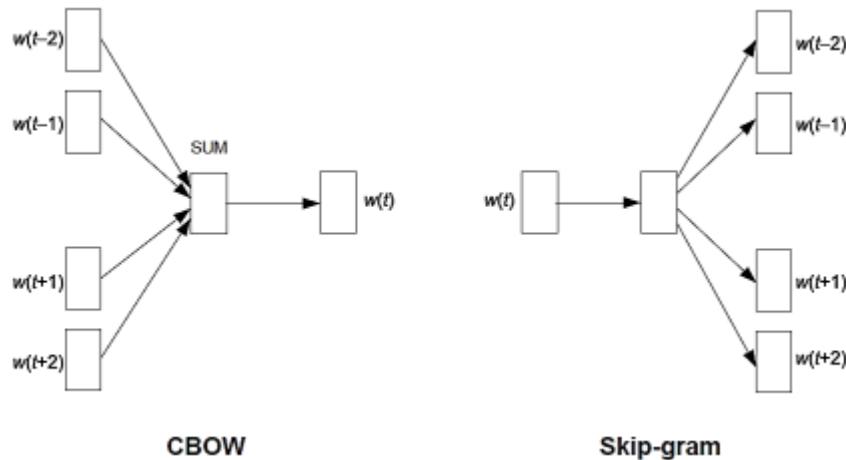


图10 Word2Vec模型示意图

从该算法流程可以看出，最外层的预测复杂度大幅降低，隐含层也去掉，这使得计算速度极大提高。该算法可得到词的Dense的Word Embedding，这是一个非常好的表示，可以用来计算文本的匹配度。但由于该模型的学习目标是预测词发生概率，即语言模型，所以从海量语料中学习到的是词的通用语义信息，无法直接应用于定制业务的匹配场景。能否根据业务场景对语义表示和匹配同时建模，以提升匹配效果呢？DSSM系列工作就充分考虑了表示和匹配。

DSSM系列

这类方法可以把表示和学习融合起来建模，比较有代表性的是微软的相关工作。下面将介绍DSSM系列内容。

(1) DSSM模型框架

DSSM网络结构如图11所示：

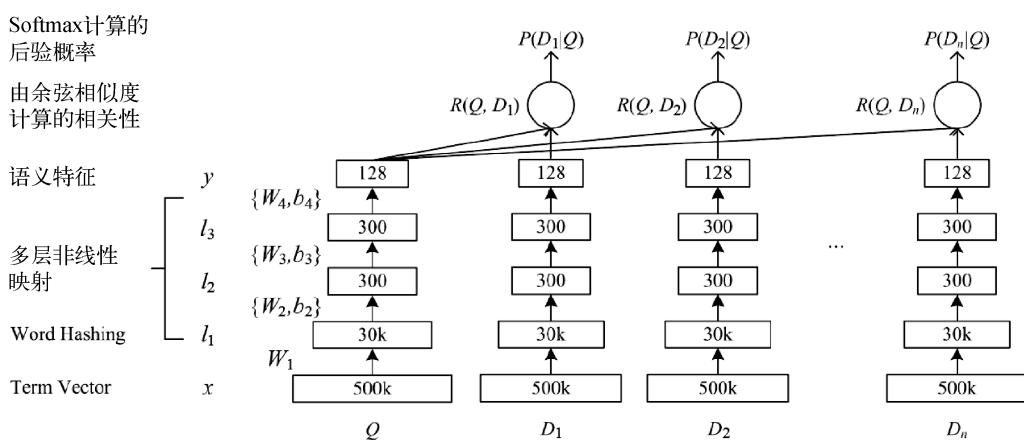


图11 DSSM网络结构图

使用搜索的点击数据训练语义模型，输入查询Query(Q)和展现点击的Doc(D)列表，先对 Q 和 D 做语义表示，再通过 Q - D 的Cosine计算相似度，通过Softmax来区分点击与否。其中，语义表示先使用词散列对词表做降维（比如英文字母的Ngram），经过几层全连接和非线性变化后得到128维的 Q 和 D 的表示。从实验结论可知，NDCG指标提升还是很明显的，如图12所示。

#	Models	NDCG@1	NDCG@3	NDCG@10
1	TF-IDF	0.319	0.382	0.462
2	BM25	0.308	0.373	0.455
3	WTM	0.332	0.400	0.478
4	LSA	0.298	0.372	0.455
5	PLSA	0.295	0.371	0.456
6	DAE	0.310	0.377	0.459
7	BLTM-PR	0.337	0.403	0.480
8	DPM	0.329	0.401	0.479
9	DNN	0.342	0.410	0.486
10	L-WH linear	0.357	0.422	0.495
11	L-WH non-linear	0.357	0.421	0.494
12	L-WH DNN	0.362	0.425	0.498

图12 DSSM和其他模型在Bing搜索数据上效果对比

**(2) CLSM **

在DSSM基础上，CLSM增加了1维卷积和池化操作来获取句子全局信息，如图13所示。通过引入卷积操作，可以充分考虑窗口内上下文的影响，从而保证词在不同语境下的个性化语义。

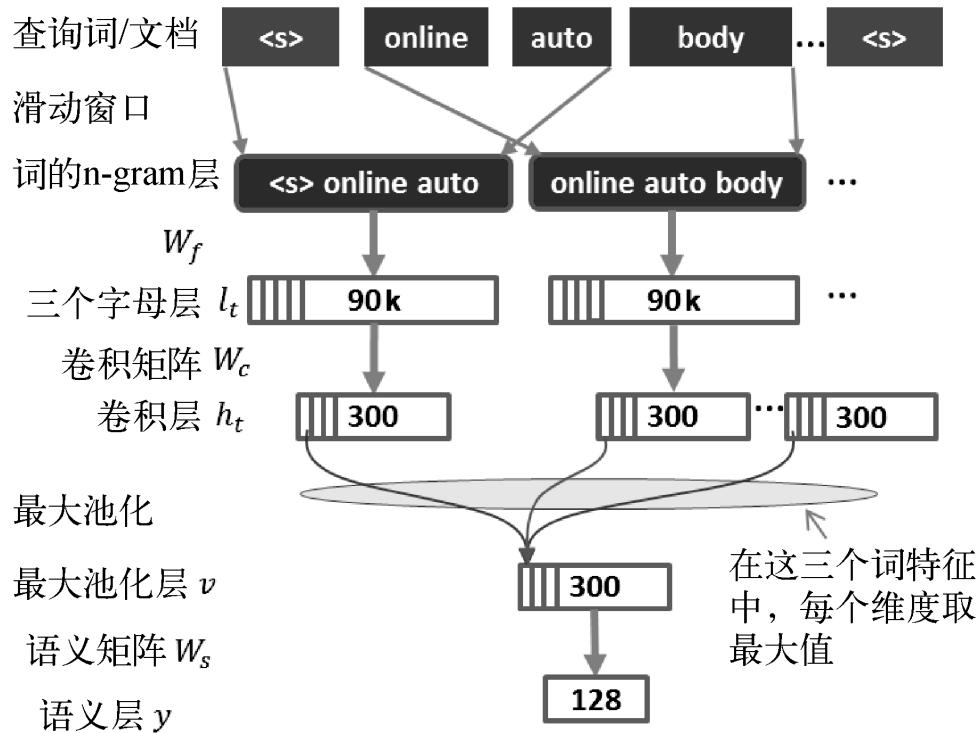


图13 CLSM网络结构图

对应效果如图14所示：

#	Models	NDCG@1	NDCG@3	NDCG@10
1	BM25	0.305	0.328	0.388
2	ULM	0.304	0.327	0.385
3	PLSA ($T=100$)	0.305	0.335 ^a	0.402 ^a
4	PLSA ($T=500$)	0.308	0.337 ^a	0.402 ^a
5	LDA ($T=100$)	0.308	0.339 ^a	0.403 ^a
6	LDA ($T=500$)	0.310 ^a	0.339 ^a	0.405 ^a
7	BLTM	0.316 ^a	0.344 ^a	0.410 ^a
8	MRF	0.315 ^a	0.341 ^a	0.409 ^a
9	LCE	0.312 ^a	0.337 ^a	0.407 ^a
10	WTM	0.315 ^a	0.342 ^a	0.411 ^a
11	PTM (maxlen = 3)	0.319 ^a	0.347 ^a	0.413 ^a
12	DSSM ($J = 4$)	0.320 ^a	0.355 ^{aB}	0.431 ^{aB}
13	DSSM ($J = 50$)	0.327 ^{aB}	0.363 ^{aB}	0.438 ^{aB}
14	CLSM ($J = 4$)	0.342 ^{aBγ}	0.374 ^{aBγ}	0.447 ^{aBγ}
15	CLSM ($J = 50$)	0.348^{aBγ}	0.379^{aBγ}	0.449^{aBγ}

图14 CLSM和其他模型在Bing搜索数据上效果对比

(3) LSTM-DSSM

LSTM-DSSM使用LSTM作为Q和D的表示，其他框架和DSSM一致，其网络结构图如图15所示。由于LSTM具备语义记忆功能且蕴含了语序信息，因此更适合作为句子的表示。当然也可以使用双向LSTM以及注意力模型（Attention Model）。

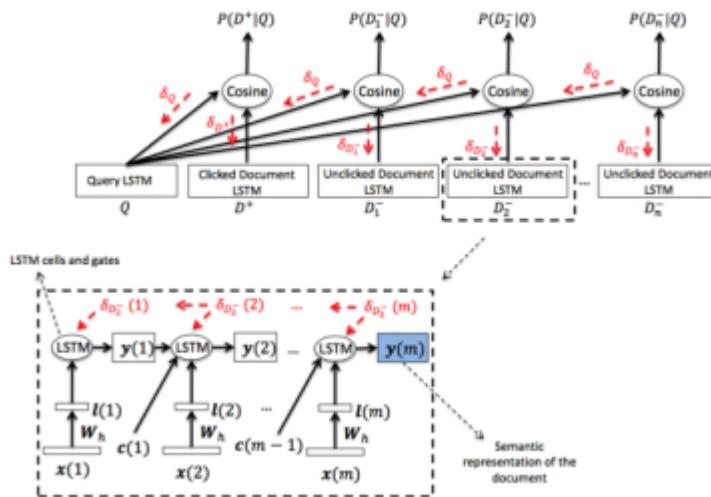


图15 LSTM-DSSM网络结构图

美团的深度学习文本匹配算法

文本的语义匹配作为自然语言处理经典的问题，可以用在搜索、推荐、广告等检索系统的召回、排序中，还可以用在文本的去重、归一、聚类、抽取等场景。语义匹配的常用技术和最新进展前文已经介绍了。

在美团这样典型的O2O应用场景下，结果的呈现除了和用户表达的语言层语义强相关之外，还和用户意图、用户状态强相关。用户意图即用户是来干什么的？比如用户在百度上搜索“关内关外”，其意图可能是想知道关内和关外代表的地理区域范围，“关内”和“关外”被作为两个词进行检索。而在美团上搜索“关内关外”，用户想找的可能是“关内关外”这家饭店，“关内关外”被作为一个词来对待。再说用户状态，一个在北京的用户和一个在武汉的用户，在百度或淘宝上搜索任何一个词条，他们得到的结果不会差太多。但是在美团这样与地理位置强相关的应用下就会完全不一样。比如在武汉搜“黄鹤楼”，用户找的可能是景点门票，而在北京搜索“黄鹤楼”，用户找的很可能是一家饭店。

如何结合语言层信息和用户意图、用户状态来做语义匹配呢？

在短文本外引入部分O2O业务场景相关特征，将其融入到设计的深度学习语义匹配框架中，通过点击/下单数据来指引语义匹配模型的优化方向，最终把训练出的点击相关性模型应用到搜索相关业务中。

针对美团场景设计的点击相似度框架ClickNet，是比较轻量级的模型，兼顾了效果和性能两方面，能很好推广到线上应用，如图16所示。

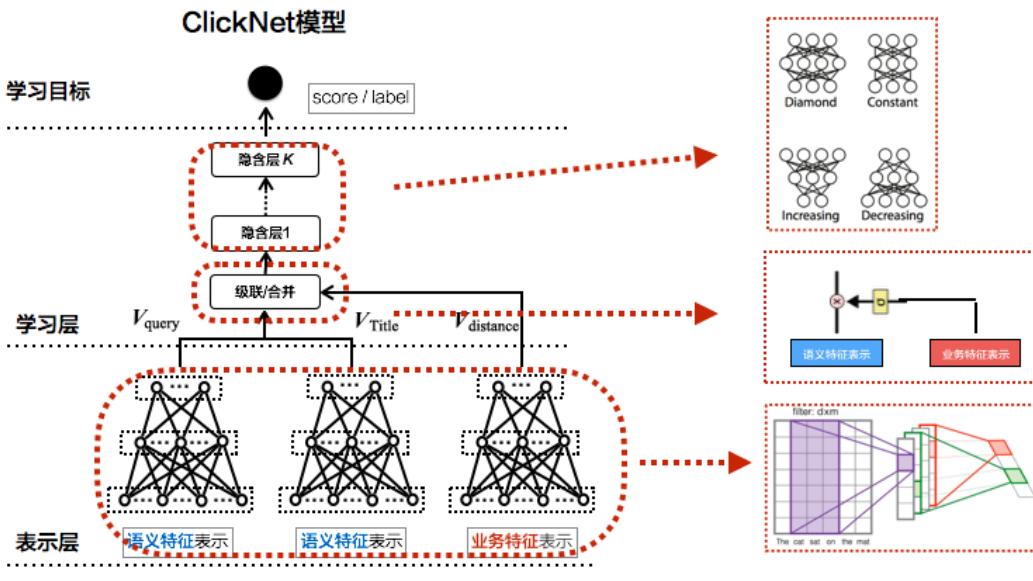


图16 ClickNet模型

- 表示层。Query和商家名分别用语义和业务特征表示，其中语义特征是核心，通过DNN/CNN/RNN/LSTM/GRU方法得到短文本的整体向量表示。另外会引入业务相关特征，比如用户或商家的相关信息、用户和商家距离、商家评价等。
- 学习层。通过多层全连接和非线性变化后，预测匹配得分，根据得分和标签来调整网络，以学习出Query和商家名的点击匹配关系。

如果ClickNet算法框架上要训练效果很好的语义模型，还需要根据场景做模型调优。首先，我们从训练语料做很多优化，比如考虑样本不均衡、样本重要度等问题。其次，在模型参数调优时考虑不同的优化算法、网络大小层次、超参数的调整等问题。

经过模型训练优化，该语义匹配模型已经在美团平台搜索、广告、酒店、旅游等召回和排序系统中上线，使访购率/收入/点击率等指标有很好的提升。

总结一下，深度学习应用在语义匹配上，需要针对业务场景设计合适的算法框架。此外，深度学习算法虽然减少了特征工程工作，但模型调优的难度会增加。因此可以将框架设计、业务语料处理、模型参数调优三方面综合起来考虑，实现一个效果和性能兼优的模型。

基于深度学习的排序模型

排序模型简介

在搜索、广告、推荐、问答等系统中，由于需要在召回的大量候选集合中选择出有限的几个用于展示，因此排序是很重要的一环。如何设计这个排序规则使得最终业务效果更好呢？这就需要复杂的排序模型。比如美团搜索系统中的排序会考虑用户历史行为、本次查询Query、商家信息等多维度信息，抽取设计出各种特征，通过海量数据的训练得到排序模型。这里只简要回顾排序模型类型和演进，重点介绍深度学习用在排序模型中的情况。

排序模型主要分类三类：Pointwise、Pairwise、Listwise，如图17所示。Pointwise对单样本做分类或者回归，即预测的得分作为排序准则，比较有代表性的模型有逻辑回归、XGBoost。Pairwise会考虑两两样本之间偏序关系，转化成单分类问题，比如比高，那这个Pair预测正，反之则负，典型的模型有

RankSVM、LambdaMART。第三类就是Listwise模型，排序整体作为优化目标，通过预测分布和真实排序分布的差距来优化模型，典型的模型如ListNet。

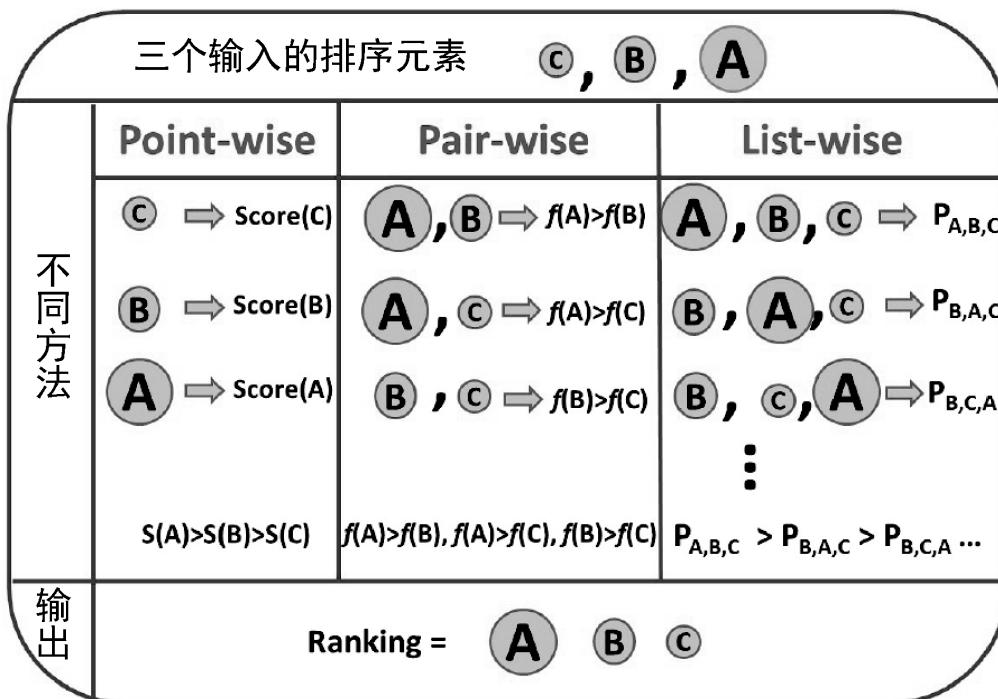


图17 排序模型分类

深度学习排序模型的演进

在排序模型的发展中，神经网络很早就被用来做排序模型，比如2005年微软研究院提出的RankNet使用神经网络做Pairwise学习；2012年谷歌介绍了用深度学习做CTR的方法；与此同时，百度开始在凤巢中用深度学习做CTR，于2013年上线。随着深度学习的普及，各大公司和研究机构都在尝试把深度学习应用在排序中，比如谷歌的Wide & Deep、YouTube的DNN推荐模型等，前面介绍的DSSM也可用来排序。下面将对RankNet、Wide & Deep、YouTube的排序模型作简单介绍。

RankNet

RankNet是Pairwise的模型，同样转化为Pointwise来处理。比如一次查询里， D_i 和 D_j 有偏序关系，前者比后者更相关，那把两者的特征作为神经网络的输入，经过一层非线性变化后，接入Loss来学习目标。如果 D_i 比 D_j 更相关，那么预测的概率为下式，其中 S_i 和 S_j 是对应Doc的得分。

$$p(D_i > D_j) = \frac{1}{1 + \exp(-\gamma(S_i - S_j))}$$

在计算得分时使用神经网络，如图18所示，每样本的输入特征作为第一层，经过非线性变换后得到打分，计算得到Pair的得分后进行反向传播更新参数，这里可以使用Mini-Batch。由于RankNet需要预测的概率

公式具有传递性，即 D_i 和 D_j 的偏序概率可以由 D_i 和 D_k 以及 D_k 和 D_j 得到，因此RankNet把计算复杂度从 $O(n^2)$ 变成了 $O(n)$ ，具体介绍可参考文献。

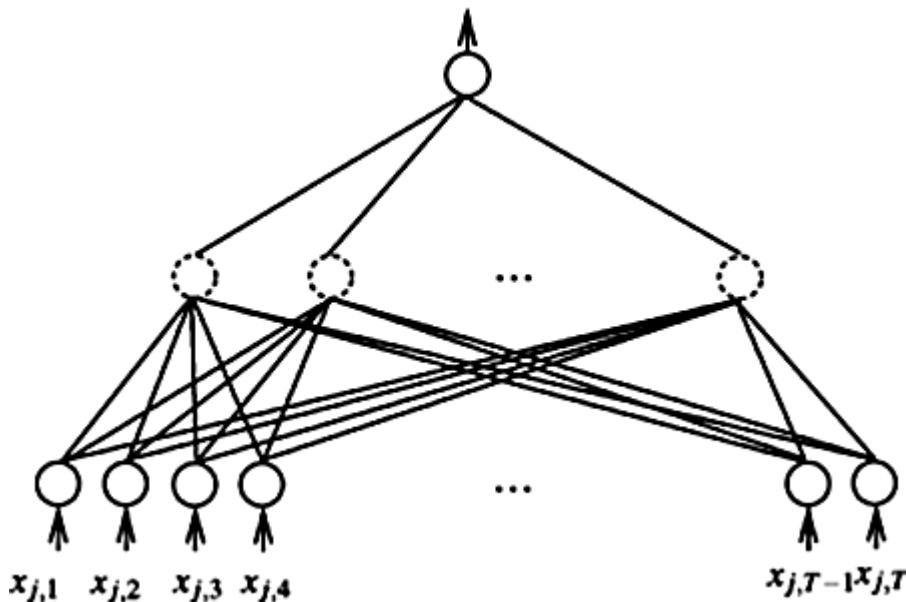


图18 RankNet示意图

当然，后续研究发现，RankNet以减少错误Pair为优化目标，对NDCG等指标（关心相关文档所在位置）衡量的效果不是太好，于是后面出现了改进模型，如LambdaRank。RankNet是典型的神经网络排序模型，但当时工业界用得多的还是简单的线性模型，如逻辑回归，线性模型通过大量的人工设计特征来提高效果，模型解释性好性能也高。当人工设计特征到一定程度就遇到了瓶颈，而深度学习能通过原始的特征学习出复杂的关系，很大程度减轻了特征工程的工作。并且GPU、FPGA等高性能辅助处理器变得普及，从而促进了深度神经网络做排序模型的广泛研究。

Wide&Deep

谷歌发表过一篇论文“Wide & Deep Learning”，其观点可以用在推荐里，比如谷歌的Apps推荐利用此观点取得很好的效果，并把模型发布在TensorFlow中。Wide & Deep整体模型结构分为Wide和Deep两部分，这两部分在最外层合并一起来学习模型，如图19所示。输入都是稀疏特征，但特征分为两种：一种适合做Deep的深度网络变化，适合时效性或者记忆性的特征，比如统计特征或者展示位置等；另一种可以直接连在最外层，适合有推广力但需要深度组合抽样的特征，比如品类、类型等。在模型优化时两部分做联合优化，Wide部分使用FTRL，而Deep使用Adagrad算法。这样，Wide和Deep对不同类型特征区分开来，充分发挥各自作用，解释性比较好。

这种思路其实可以做些扩展。比如Wide连接不是在最外层，而是在某一层，Deep的某些层也可以连接到最外层，这样能充分利用不同层抽象的Dense信息。与Wide & Deep的网络连接方式类似，如2003年NNLM和2010年RNNLM模型里的直接连接（Direct Connection），其浅层和深层的结合能很好地加速收敛，深度学习的Highway方式也是类似的。目前Wide & Deep应用较多，比如在阿里巴巴就有比较好的应用。

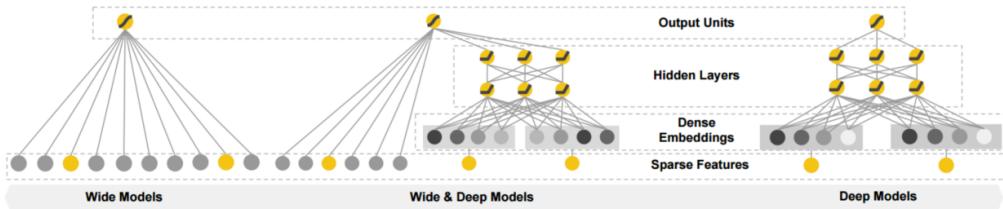


图19 谷歌的Wide&Deep模型结构

YouTube DNN排序模型

YouTube用来预测用户观看视频时长，转化为加权的逻辑回归问题。DNN排序模型和前面的工作类似，其网络结构是标准的前馈神经网络，如图20所示。DNN排序模型的特点还是在输入特征上。虽然深度学习模型对特征工程要求很低，但很多数据需要经过简单处理后才可加入模型。图20中的特征分为很多域，比如语言方面、视频方面、用户历史看过的视频ID，还有之前观看时长的统计量和归一化的值。离散值经过Embedding的处理后变成连续向量，再级联起来经过多层非线性变化后来预测最终的标签。

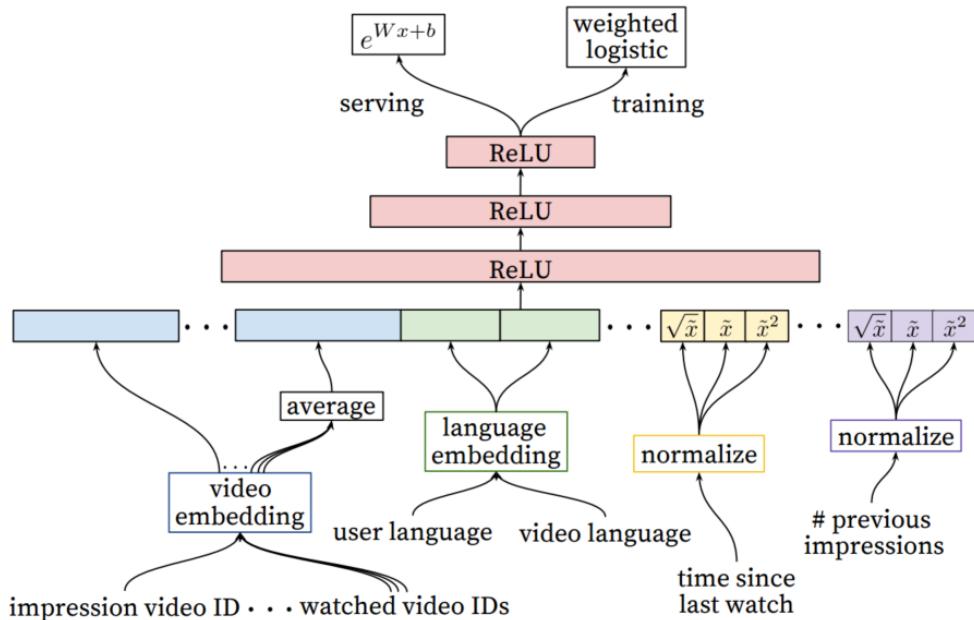


图20 YouTube的DNN排序模型结构

从上面介绍的深度学习相关排序模型可以看出，排序模型所需要的数据类型多种多样，数据的含义也各有不同，不同于图像、语音领域单一的输入形式。因此在做排序模型中，输入特征的选取和表示方式是很重要的，比如连续特征、离散特征处理、用户历史、文档特征的区分等。在美团场景中，设计排序模型需考虑业务特点，对输入特征的表示做很多尝试。

美团的深度学习排序模型尝试

在语义匹配模型中介绍了ClickNet框架，其实该框架同时也可用于排序，与语义匹配的区别主要在表示层，如图21所示。如果ClickNet用作搜索的CTR模型，那表示层的Query和Title的语义特征只是一部分，还有用户查询、用户行为、商家信息以及交叉组合特征都可以作为业务特征，并按特征的类型分不同的域。进一步讲，如果场景不包含语义匹配，模型的输入可以只有业务特征。下面简单讲解在美团用ClickNet做排序模型的尝试。

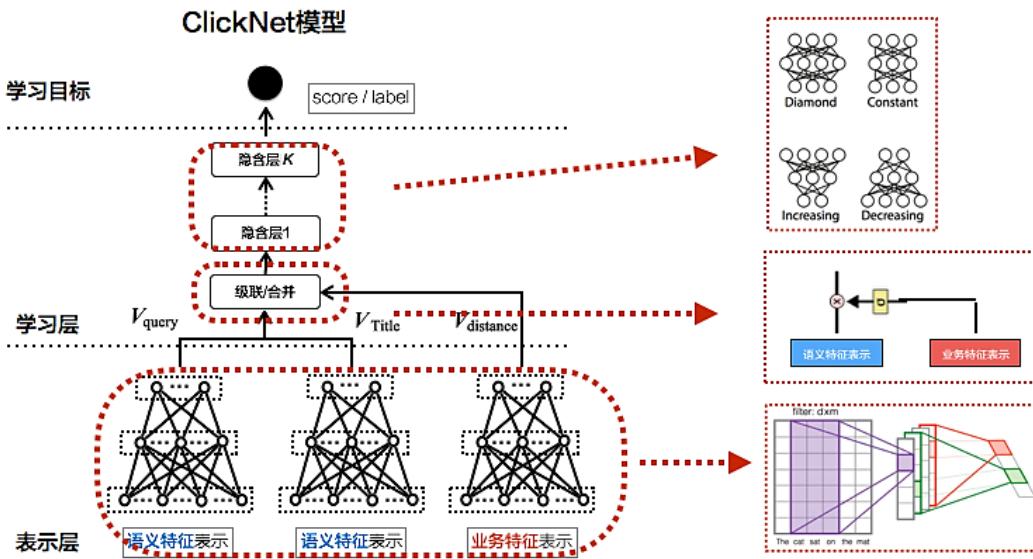


图21 ClickNet模型框架

ClickNet-v1

ClickNet设计的初衷是它作为文本的匹配模型，并作为一维语义特征加入到业务的Rank模型中以提升效果。但根据上线之后的数据分析，我们发现以语义特征表示为主、辅以部分业务特征的ClickNet在排序系统中有更好的表现。我们针对排序模型做了如下改进。

- (1) 业务特征选取。从业务方Rank已有的人工特征中，选取O2O有代表性的且没经过高级处理过的特征，比如用户位置、商家位置、用户历史信息、商家评价星级、业务的季节性等。
- (2) 特征离散化。选取的业务特征做离散化处理，比如按特征区间离散化等。
- (3) 样本处理。针对业务需要对正负例采样，对点击、下单、付费做不同的加权操作。
- (4) 信息的融合。通过引入Gate来控制语义特征和各业务特征的融合，而不仅是求和或者级联，通过样本学习出Gate的参数。

针对业务Rank的目标优化ClickNet后，效果有很大改善，但模型还是偏重于语义特征。能否直接使用ClickNet做排序模型呢？答案是可以的。只需要加重业务特征、弱化或者去除语义表示特征就可以，这样修改后的模型就是ClickNet-v2。

ClickNet-v2

ClickNet-v2以业务特征为主，替换业务Rank模型为目标，使用业务特征作为ClickNet的表示层输入，通过对各特征离散化后传入模型。和ClickNet-v1不一样的是，ClickNet-v2业务特征种类繁多，需要做深度的分析和模型设计。

比如如何考虑位置偏好问题？因为展示位置会有前后顺序，后面的展示不容易被用户看到，从而天然点击率会偏低。一种解决思路是可以把位置信息直接连接到最外层，不做特征组合处理。

再比如各业务特征通过多层非线性变化后，特征组合是否充分？一种解决思路是使用多项式非线性变换，这能很好组合多层的特征。

又比如模型组合的效果是否更好？一种解决思路是尝试FM和ClickNet的级联，或者各模型的Bagging。

此外还有模型的解释性等很多和业务场景相关的情况需要考虑。

ClickNet是基于自研的深度学习框架Ginger实现的，收敛速度和效果都非常理想。我们来看看分类任务上的一些测试，如图22所示。在Higgs数据上，基于Ginger的ClickNet比基于XGBoost的AUC提升34个千分点，使用TensorFlow实现的ClickNet比基于Ginger的AUC差3个千分点，且其速度慢于Ginger。如图23所示，ClickNet相比线性模型，其准确率都有较好的提升。

模型	数据集	训练速度	测试集AUC
XGBoost	HIGGS Train 1050w样本 Test 50w样本 28维稠密	5.7 s/epoch (32 worker*2线程, 2 server)	0.8472
Ginger		120 s/epoch (单机24线程)	0.8817
Tensorflow		300 s/epoch (单机GPU+多线程)	0.8787

图22 XGBoost和ClickNet在Higgs分类效果对比

数据名称	数据大小	特征维度	#nonzeros per instance	离散值/连续值	ginger模型结构 (lookup * hid)	准确率
train.3	#(train): 1243 #(test): 41	21	21 DENSE	连续值	16 * 8	libsvm: 87.80% ginger: 95.12%
test.3						
rcv1_train.binary	#(train): 20242 #(test): 677399	47236	74 SPARSE	连续值	128 * 64 * 32	libsvm: 95.14% ginger: 95.95%
rcv1_test.binary						
a9a (train)	#(train): 32561 #(test): 16281	123	14 SPARSE	离散值	192 * 128 * 64	libsvm: 85.07% ginger: 85.31%
a9a.t (test)						
kdda (train)	#(train): 8407752 #(test): 510302	20216830	36 SPARSE	连续值	128 * 64	liblinear: 88.53% ginger: 89.12%
kdda.t (test)						

图23 ClickNet和线性模型在分类任务对比

总结

因为深度学习的拟合能力强、对特征工程要求低，它在文本领域已经有很多应用。本章以语义匹配和排序模型为例，分别介绍了业界进展和美团场景的应用。

第一部分介绍了语义匹配经历的向量空间、潜在语义分析、主题模型、深度学习几个阶段，重点介绍了深度学习应用在语义匹配上的Embedding和DSSM系列模型，以及美团尝试的ClickNet模型。第二部分介绍了深度学习在排序模型的一些进展和美团的一些尝试。除了这两部分内容外，深度学习几乎渗透了文本的各方面，美团还有很多尝试方式，比如情感分析、对话系统、摘要生成、关键词生成等，限于篇幅不做介绍。总之，认知智能还有很长的路需要走，语言文字是人类历史的文化沉淀，涉及语义、逻辑、文化、情感等众多复杂的问题。我们相信，深度学习在文本领域很快会有很大突破。

参考文献

- [1] Thomas Hofmann. “Probabilistic Latent Semantic Analysis”. 1999.
- [2] David M.Blei, Andrew Y.Ng, Michael Jordan. “Latent Dirichlet Allocation” . 2002.
- [3] Huang, Po-Sen et al. “Learning deep structured semantic models for web search using clickthrough data” in CIKM 2013.
- [4] Shen, Yelong, He, Xiaodong, Gao, Jianfeng, et al. “A latent semantic model with convolutional–pooling structure for information retrieval” in CIKM 2014.
- [5] H. Palangi et al. “Semantic modeling with long–short–term memory for information retrieval”. 2015.

团队简介

美团点评算法团队是整个美团点评技术团队的“大脑”，涵盖搜索、推荐、广告、智能调度、自然语言处理、计算机视觉、机器人以及无人驾驶等多个技术领域。帮助美团点评数亿活跃用户改善了用户体验，也帮助餐饮、酒店、结婚、丽人、亲子等200多个品类的数百万商户提升了运营效率。目前，美团点评算法团队在人工智能领域进行积极的探索和研究，不断创新与实践，致力于应用最前沿的技术，给广告大消费者带来更好的生活服务体验。

机器学习在美团配送系统的实践：用技术还原真实世界

作者: 何仁清

“

在2018 AI开发者大会（AI NEXTCon）上，美团配送AI方向负责人何仁清，分享了美团在即时配送领域中机器学习技术的最新进展，以及如何通过大数据和机器学习手段，建立对线下真实世界各种场景的感知能力，还原并预测配送过程各个细节，从而提升整体配送系统的精度。

美团“超脑”配送系统的由来

2014年，斯嘉丽·约翰逊主演的科幻片《超体》大火，影片中主人公Lucy由于无意中摄入了大量的代号为“CPH4”的神秘药物，大脑神经元获得空前的开发，获得了异乎寻常的超能力，她能够对这个世界进行全新的感知、理解和控制（比如控制无线电波），最终跨越时间和空间成为了一个超级个体。

这种对真实世界的深度感知、理解和控制，与配送AI系统对配送场景的感知、理解和配送环节控制的目标非常一致。可以说，美团要建设的AI就是配送系统的“超级大脑”。因此我们内部把配送的AI系统，简称“超脑”配送系统。



即时配送在全球快速发展

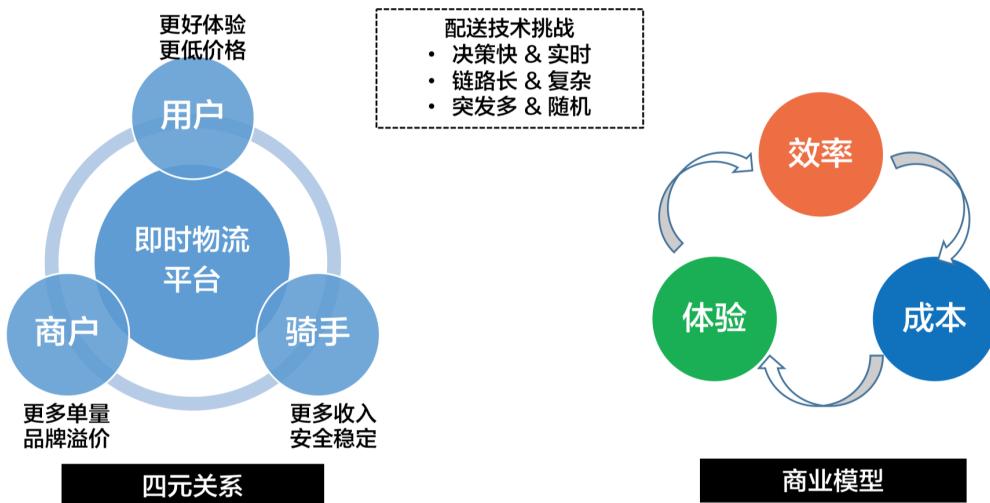
最近几年，以外卖为依托，即时配送业务在全球范围内掀起了一波快速发展的浪潮，全球各地都出现了很多创业公司，其中国外知名的包括美国的Uber Eats（全球）、英国的Deliveroo、印度的Swiggy、Zomato（分别被美团和阿里投资），印尼的go-jek等等。国内除了美团外卖、饿了么、滴滴外卖等典型代表外，而还有专注于即时配送服务创业公司，比如闪送、UU跑腿、达达、点我达等。

这种全球爆发的现象说明了两个问题：

- “懒”是人类的天性。平价、方便、快捷的服务是人类的普遍需求，尤其是在“吃”这个事情上，外卖成为了一种高频的刚需。
- 外卖的商业模式完全可行。以美团外卖为例，2018年上半年整体收入160亿，同步增长90%。根据Uber公布的数据，Uber Eats在2018第一季度占整体营业的13%。

即时配送的业务模型

即时配送，是一种配送时长1小时以内，平均配送时长约30分钟的快速配送业务。如此快速的配送时效，将传统的线上电商交易与线下物流配送（传统划分比较明确的两条业务）整合为统一整体，形成了用户、商户、骑手和平台互相交错的四元关系。



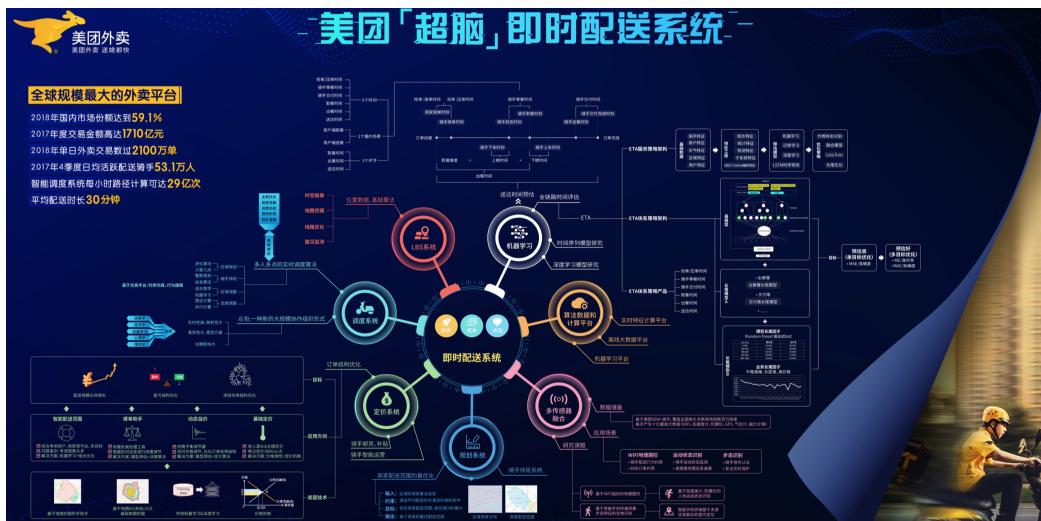
其整合力度空前紧密，几乎渗透到各个环节。以外卖搜索和排序为例，在下午时段，在用户搜索和推荐中可以看到更多的商家，因为此时运力充分，可以提供更远距离的配送服务，不仅能更好满足用户的需求，提高商家的单量，而且能够增加骑手的收入。

即时配送的核心指标是效率、成本、体验，这三者也形成了即时配送的商业模型。简单来说可以分为以下几步：

1. 首先配送效率提升
 - 让骑手在单位时间内配送更多订单，产生更多价值。
2. 然后配送成本下降
 - 更高的效率，一方面让骑手收入增加，一方面也让订单平均成本下降。
3. 然后用户体验提升
 - 低成本能够让用户（商户）以更低的价格享受更好的配送服务，从而保证更好的用户体验。
4. 进一步提升效率并形成循环
 - 更好的用户体验，让更多用户（商户）聚集过来，提升规模和密度，进一步提升配送效率。

这样，就形成了一个正向循环，不断创造更多商业价值。而技术的作用，就是加速这个正向循环。

美团“超脑”配送系统



目前互联网技术，很大部分还是针对线上产品和系统研发，整个流程可以在线上全部完成，而这也正是配送AI技术最大的不同和挑战。简单来说，类似搜索、推荐、图象和语音识别这种线上产品常用的AI技术帮助不大，因为配送必须在线下一个一个环节的进行，这就要求AI技术必须能够面对复杂的真实物理世界，必须能深度感知、正确理解与准确预测、并瞬间完成复杂决策。

为了满足这些要求，我们建设了美团“超脑”配送系统，包含以下几个方面：

- 大数据处理和计算能力
 - 算法数据和计算平台：包括实时特征计算、离线数据处理、机器学习平台等。
- 建立对世界深度感知
 - LBS系统：提供正确位置（用户/商户/骑手）以及两点之间正确的骑行导航。
 - 多传感器：提供室内定位以、精细化场景刻画、骑手运动状态识别
- 正确理解和准确预测
 - 时间预估：提供所有配送环节时间的准确预估
 - 其他预估：销量预估、运力预估等
- 完成复杂决策
 - 调度系统：多人多点实时调度系统，完成派单决策：谁来送？怎么送？
 - 定价系统：实时动态定价系统，完成定价决策：用户收多少钱？给骑手多少钱？
 - 规划系统：配送网络规划系统，完成规划决策：站点如何划分？运力如何运营？

机器学习技术挑战

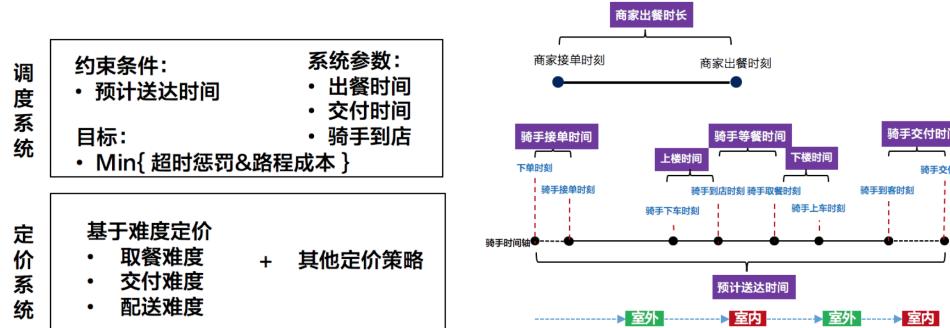
如何构建一个在真实物理世界运行的AI系统，就是我们最大的挑战。具体到机器学习方向而言，挑战包括以下几个方面：

- 精度足够高、粒度足够细
 - 时间要求：一方面是周期性变化，比如早午晚，工作假日，季节变化；一方面是分钟级的精细度，比如一个商圈单量和运力的实时变化。
 - 空间要求：一方面是不同商圈独有特性，比如CBD区域；一方面是要实现楼栋和楼层的精度，比如1楼和20楼，就是完全不同的配送难度。
 - 鲁棒性要求：处理各种不确定的能力，比如天气变化、交通变化等等。
- 线下数据质量的巨大挑战

- 大噪音：比如GPS定位漂移，尤其是在高楼附近，更不要说在室内GPS基本不可用。
- 不完备：比如商家后厨数据、堂食数据、其他平台数据，都极难获得。
- 高复杂：配送场景多样而且不稳定，随着时间、天气、路况等在不断变化。

配送系统的核心参数ETA

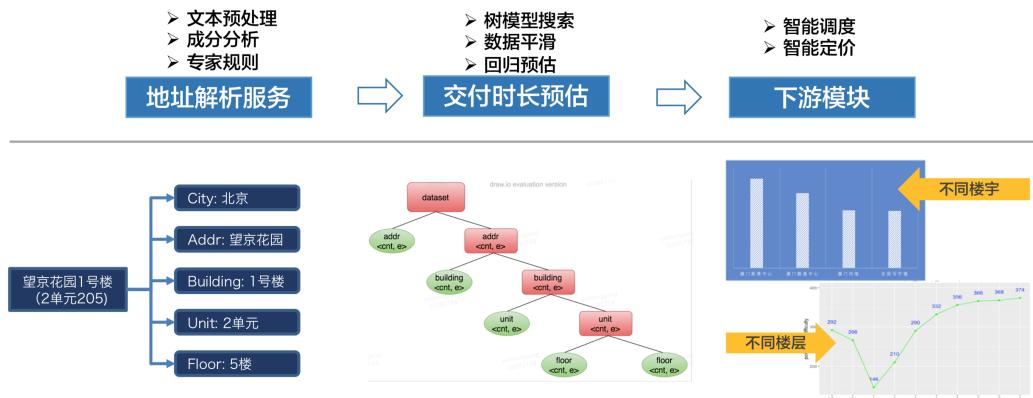
ETA (Estimated Time of Arrival, 时间送达预估) 是配送系统中非常重要参数，与用户体验、配送成本有直接关系，而且会直接影响调度系统和定价系统的最终决策。



一个订单中涉及的各种时长参数(如上图右侧所示)，可以看到有十几个关键节点，其中关键时长达到七个。这些时长涉及多方，比如骑手（接-到-取-送）、商户（出餐） 、用户（交付），要经历室内室外的场景转换，因此挑战性非常高。

通过机器学习方法，我们已经将外卖配送几乎所有环节都进行了精准预估预测。用户感知比较明显是预计送达时间，贯穿多个环节，商家列表（从配送时长角度让用户更好选择商家）、订单预览（给用户一个准确的配送时间预期）、实时状态（下单后实时反馈最新的送达时间）。当然这里面还有很多用户看不到的部分，比如商家出餐时间、骑手到店时间、交付时间等。其中交付时长，与用户关系比较大，也很有意思，下文会详细展开。

精准到楼宇和楼层的预估：交付时长



交付时长是指骑手到达用户后，将外卖交付到用户手中并离开的时间，实际是需要考虑三维空间内计算（上楼-下楼）。交付时间精准预估，有两点重要的意义，首先是客观的衡量配送难度，给骑手合理补贴；其次，考虑对骑手身上后续订单的影响，防止调度不合理，导致其他订单超时。

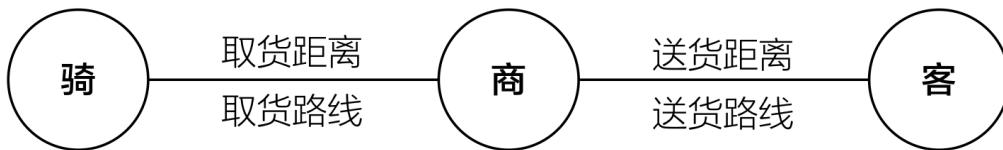
交付时长的目标是，做到楼宇和楼层的精准颗粒度，具体可以拆解为以下几步：

1. 地址的精准解析（精确到楼宇/单元/楼层）
 - 地址精度需要在5级之上（4级：街道，5级：楼宇），国内拥有这个级别精细化数据的公司屈指可数。
 - 数据的安全级别很高，我们做了很多脱敏工作，做了各种数据保护与隔离，保证用户隐私和数据安全。
 - 地址信息的多种表达方式、各种变形，需要较强的NLU技术能力。
2. 交付时长预估
 - 通过骑手轨迹进行“入客-离客”识别，并进行大量数据清洗工作。
 - 统计各个粒度的交付时长，通过树形模型实现快速搜索各个粒度的数据。
 - 因为预估精度是楼宇和楼层，数据很稀疏，很难直接进行统计，需要通过各种数据平滑和回归预估，处理数据稀疏和平滑的问题。
3. 下游业务应用
 - 给调度和定价业务，提供楼宇+楼层维度的交付时长。从上图可以看到，在不同楼宇，不同楼层交付时长的区别度还是很明显的。
 - 尤其是楼层与交付时长并不是线性相关，我们还具体调研过骑手决策行为，发现骑手会考虑等电梯的时间，低楼层骑手倾向于走楼梯，高楼层则坐电梯。

可以看到，真实世界中影响决策因素非常多，我们目前做的还不够。比如交付时长也可以进一步细化，比如准确预估骑手上楼时间、下楼时间和等待时间，这样其实能够与商家取餐环节保持一致，之所以没这么做，主要还是数据缺失，比如骑手在商家其实有两个操作数据（到店、取餐），这样能支持我们做精细化预估的，但是在用户环节只有（送达）一个操作。

举这个例子，其实是想说明，数据的完备性对我们到底有多重要。数据方面的挑战，线下业务与线上业务相比，要高出好几个等级。

配送中最重要的数据之一：地图



正确的定位 + 正确的导航

- | | |
|---|--|
| <input type="checkbox"/> 骑手定位
<input type="checkbox"/> 商家定位
<input type="checkbox"/> 用户定位 | <input type="checkbox"/> 正确距离
<input type="checkbox"/> 正确路线 |
|---|--|

配送地图的技术挑战

状态复杂

- 到店-取货-到客-交付
- 交通工具多样

混合轨迹

- 室内+室外
- 频繁切换

多点配送

- 多订单多目的地
- 骑手决策随机性

地图对配送的重要性毋庸置疑（位置和导航都不准确，配送如何进行？），前面提到的5级地址库只是其中一部分。配送地图的目标可以概括为以下两点：

- 正确的位置

- 实时部分：骑手实时位置。
- 静态部分：用户和商户准确的地址和位置。
- 正确的导航
 - 两点之间正确的距离和路线。
 - 突发情况的快速反应（封路、限行）。

如果横向对比配送、快递、打车等行业对地图的要求，其实是一件很有意思的事情，这个对于配送地图技术建设来说，是一件非常有帮助的事情。

即时配送 VS 物流快递：即时配送对地图的依赖程度明显高于物流快递

	即时配送	物流快递
地址库	5级	5级
位置	依赖度高 1) 骑手面对商家和用户位置非常多，而且不固定 2) 服务范围极大，骑手比较难熟悉整个区域	依赖度低 1) 快递员面对仓库和用户位置相对固定 2) 配送区域固定，快递员比较熟悉
导航	依赖度高 1) 路线实时规划，不确定强 2) 根据距离实时定价 3) 路线不熟悉，需要精准导航	依赖度低 1) 路线相对固定，相对确定； 2) 根据订单数量定价 3) 路线熟悉，不需要导航

即时配送 VS 出行行业：地图厂商在车载导航的优势和积累，在即时配送场景较难发挥

	即时配送	出行行业
位置	送货到家 1) 小区内部，精确到单元门 2) 需要考虑室内定位漂移的问题	路边接人 1) 门口、路口和道路旁等上车点 2) 需要考虑车辆调头、通行等问题
导航	骑行导航 1) 重点考虑天桥、隧道和小路、是否准入等特殊性 2) 配送到单元门，内部路导航要求高 3) 对实时路况要求较低	车行导航 1) 需要考虑道路等级问题、收费情况等 2) 上车点在路旁，主干路导航要求高 3) 需要较强的实时路况

从这两方面对比可以看到，在即时配送业务中，骑行地图的重要性非常之高，同时很多问题确实非常具有行业特色，通过驾车地图的技术无法很有效的解决。这样就需要建设一套即时配送业务地图的解决方案。

基于签到数据的位置校正：交付点



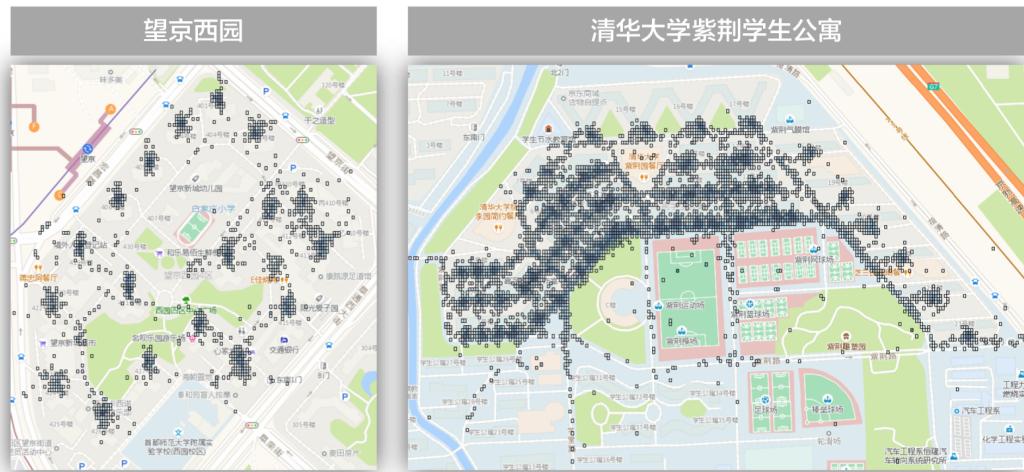
如前文所述，配送地图的方向有很多，这次我重点讲一下用户位置相关的工作“交付点挖掘”。首先看一下目前主要问题：用户位置信息有很多错误，比如：

用户选择错误

上图左，一个小区会有1期2期~N期等，用户在选择POI的时候就可能发生错误（比如1期的选了2期），两者地理位置相差非常远，很容易造成骑手去了错误的地方。这样在订单发送到配送系统的时候，我们需要做一次用户坐标纠正，引导骑手到达正确的位置。

POI数据不精细

上图右，用户本来在xx区xx栋，但是只选了xx区这个比较粗的位置信息。现实中在一个小区里面，找到一个具体xx栋楼还是非常困难的，大家可以想想自己小区中，随便说一个楼号你知道它在哪个角落吗，更别说如果是大晚上在一个你不熟悉的小区了。造成这种原因，一方面可能是用户选择不精细，还有一种可能，就是地图上没有具体楼栋的POI信息。



在实际配送中，我们都会要求骑手在完成交付后进行签到，这样就会积累大量的上报数据，对于后续进行精细化挖掘非常有帮助。大家可以先看看我们收集的原始数据（上图），虽然还是非常凌乱，但是已经能看到这其中蕴含着极高的价值，具体来说有三方面：

- 数据量大
 - 每天几千万订单，几十亿的轨迹数据。
 - 可以充分覆盖每一个小区/楼栋/单元门。
- 维度多样

- 除了骑手签到和轨迹数据，我们还有大量的用户、商户和地图数据。
- 多种数据维度可以交叉验证，有效避免数据的噪音，提高挖掘结果精度。
- 数据完备
 - 在局部（用户和商户）数据足够稠密，置信度比较高。

交付点挖掘的技术实战：挑战



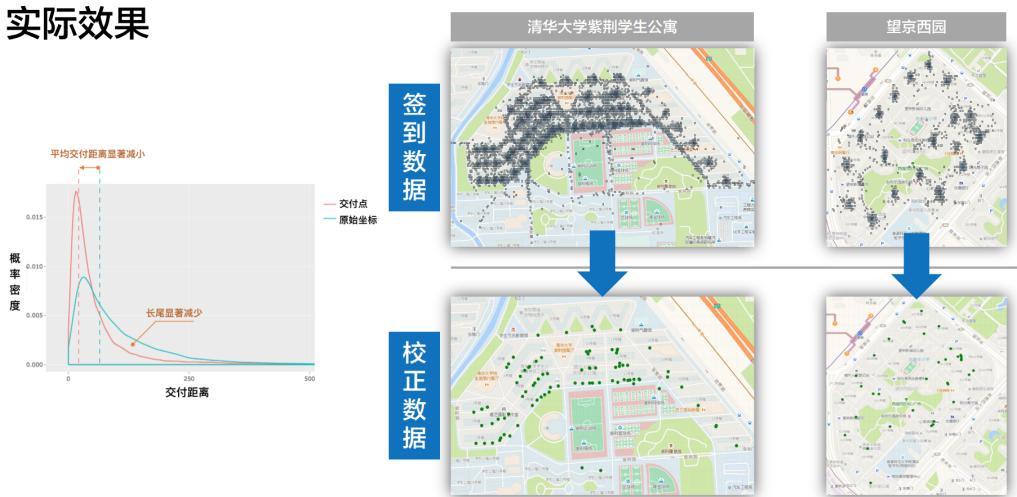
在数据挖掘实际过程中，其实并没有什么“高大上”的必杀技，无法使用流行的End2End方法，基本上还是需要对各个环节进行拆解，扎实地做好各种基础工作，基本整个挖掘过程，分为以下几个步骤：

(1) 基于地址分组； (2) 数据去噪； (3) 数据聚合； (4) 置信度打分。其中主要技术挑战，主要在各种场景中保证数据挖掘质量和覆盖率，具体来说主要有三个挑战：

- 数据去噪
 - 数据噪音来源比较多样，包括GPS的漂移、骑手误操作、违规操作等各种。一方面是针对噪音原因进行特殊处理（比如一些作弊行为），另一方面要充分发挥数据密度和数据量的优势，在保证尽量去除Outlier后，依然保持可观的数据量。能够同时使用其他维度的数据进行验证，也是非常重要的，甚至可以说数据多样性和正交性，决定了我们能做事情的上限。
- 数据聚合
 - 不同区域的楼宇密度完全不一样，具有极强的Local属性，使用常规聚类方法，比较难做到参数统一，需要找到一种不过分依赖样本集合大小，以及对去噪不敏感的聚类算法。
- 重名问题
 - 这个属于POI融合的一个子问题，判断两个POI信息是否应该合并。这个在用户地址中比较常见，用户提供的地址信息一样，但实际是两个地方。这种情况下，我们的处理原则是一方面要求纠正后坐标更符合骑手签到情况，另一方面新坐标的签到数据要足够稠密。

交付点挖掘的技术实战：效果

实际效果

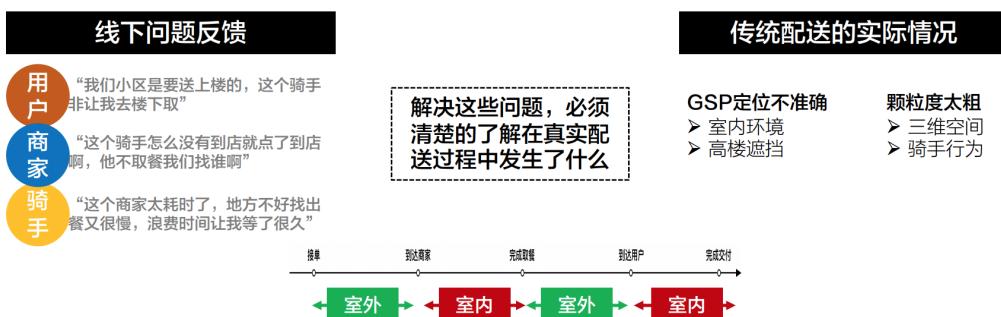


目前，我们已经上线了一版交付点，对用户位置进行主动纠正，让骑手可以更准确更快的找到用户。目前效果上看还是非常明显的。包括几个方面：

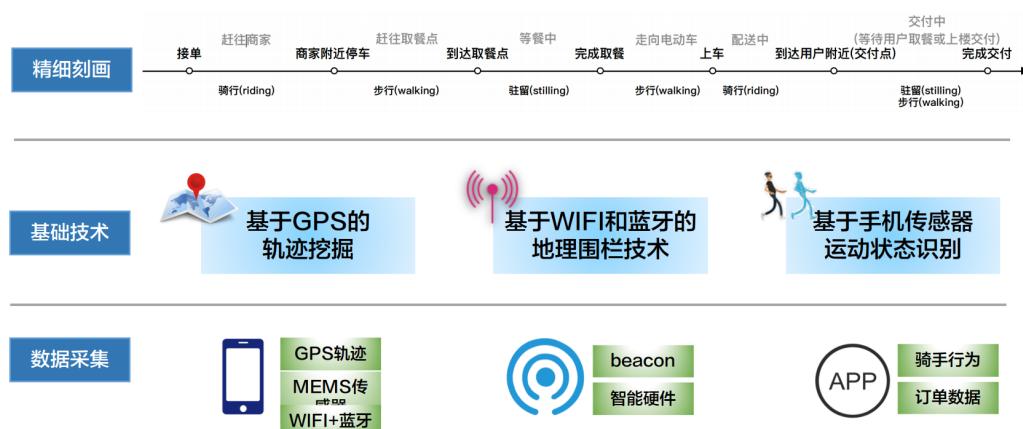
- 骑手交付距离明显降低
 - 从上图左侧部分看到，在上线前（绿色）交付距离>100M的占比很高（这个距离会导致实际位置差几栋楼，甚至不同小区），也就是用户自己选着的位置错误率比较高，导致骑手交付难度较高，对效率影响比较大。上线后（红色），交付距离明显缩短（均值左移），同时>100M的长尾比例明显下降。
- 单元门级别的高精度位置
 - 上图右侧部分看到，我们挖掘的交付点基本上能与楼宇的单元门对应。而且没有明显偏差比较大的部分。这个质量基本达到我们之前设定目标，也证明配送大数据的巨大潜力。
- 目前的问题以及后续的优化点
 - 如何提升其作为POI挖掘和发现手段的准确率？这里面有很多优化点，比如去重（交付点-位置信息的一一映射），POI信息补全和更新。
 - 如何扩大数据渠道并做到信息整合？目前主要渠道还是骑手签到和轨迹数据，这个明显有更大的想象空间，毕竟每天在全国大街小巷，有几十万骑手在进行配送，除了前面（以及后面）提到的通过手机被动采集的数据，让骑手主动采集数据，也是不错的建设思路。只不过想要做好的话，需要建立一个相对闭环数据系统，包括上报、采集、清洗、加工、监控等等。

更精细化的配送场景识别：感知

前面提到的地图技术，只能解决在室外场景的位置和导航问题。但配送在商家侧（到店、取餐）和用户侧（到客、交付）两个场景中，其实是发生在室内环境。在室内的骑手位置是在哪里、在做什么以及用户和商家在做什么，如果了解这些，就能解决很多实际问题。比如：



这个技术方向可以统称为“情景感知”，目标就是还原配送场景中（主要是室内以及GPS不准确），真实配送过程发生了什么，具体方向如下图所示：

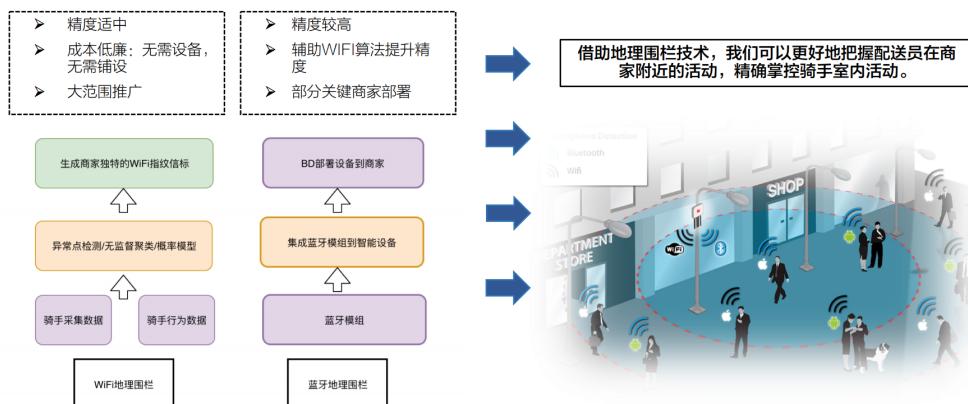


情景感知的目标就是做到场景的精细刻画（上图的上半部分），包含两个方面工作：

- 配送节点的精确刻画
 - 在ETA预估中已经展示过一些，不过之前主要还是基于骑手上报数据，这显然无法做到很高精确，必须引入更客观的数据进行描述。目前，我们选择的是WIFI和蓝牙的地理围栏技术作为主要辅助。
- 配送过程的精确刻画
 - 骑手在配送过程中经常会切换方式，比如可能某个小区不让骑电动车，那骑手必须步行，再比如骑手在商家发生长时间驻留，那应该是发生了等餐的情况（用户侧同理）。目前，我们选择使用基于传感器的运动状态识别作为主要辅助。

这些数据，大部分来自于手机，但是随着各种智能硬件的普及，比如蓝牙设备，智能电动车、智能头盔等设备的普及，我们可以收集到更多数据的数据。WiFi/蓝牙技术，以及运动状态识别的技术比较成熟，这里主要说一下概况，本文不做深入的探讨。

WiFi/蓝牙 地理围栏



运动状态识别



对于配送系统来说，比较大的挑战还是对识别精度的要求以及成本之间的平衡。我们对精度要求很高，毕竟这些识别直接影响定价、调度、判责系统，这种底层数据，精度不高带来的问题很大。

考虑成本限制，我们需要的是相对廉价和通用的解决方案，那种基于大量传感器硬件部属的技术，明显不适用我们几百万商家，几千万楼宇这种量级的要求。为此，在具体技术方面，我们选用的是WiFi指纹、蓝牙识别、运动状态识别等通用技术方案，就单个技术而言，其实学术界已经研究很充分了，而且也有很多应用（比如各种智能手环等设备）。对于我们的挑战在于要做好多种传感器数据的融合（还包括其他数据），以确保做到高识别精度。当然为了解决“Ground Truth”问题，部署一些稳定&高精度的智能硬件还是必须的，这对技术迭代优化和评估都非常有帮助。

总结

美团外卖日订单量超过2400万单，已经占有了相对领先的市场份额。美团配送也构建了全球领先的即时配送网络，以及行业领先的美团智能配送系统，智能调度系统每小时路径计算可达29亿次。如何让配送网络运行效率更高，用户体验更好，是一项非常困难的挑战，我们需要解决大量复杂的机器学习和运筹优化等问题，包括ETA预测，智能调度、地图优化、动态定价、情景感知、智能运营等多个领域。过去三年来，美团配送AI团队研发效果显著，配送时长从一小时陆续缩短到30分钟，并且还在不断提升，我们也希望通过AI技术，帮大家吃得更好，生活更好。



招聘信息

目前，即时配送业务正处于快速发展期，新的场景、新的技术问题不断涌现，团队正在迅速扩大中，急需机器学习资深专家、运筹优化技术专家、LBS算法工程师、NLP算法工程师，我们期待你的加入。扫码可查看职位详情，或者发送简历至 yewei05@meituan.com

如何基于深度学习实现图像的智能审核?

作者: 晓明

背景

美团每天有百万级的图片产生量,运营人员负责相关图片的内容审核,对涉及法律风险及不符合平台规定的图片进行删除操作。由于图片数量巨大,人工审核耗时耗力且审核能力有限。另外对于不同审核人员来讲,审核标准难以统一且实时变化。所以有必要借助机器实现智能审核。

图像智能审核一般是指利用图像处理与机器学习相关技术识别图像内容,进而甄别图像是否违规。图像智能审核旨在建立图片自动审核服务,由机器自动禁止不符合规定(负例)的图片类型,自动通过符合规定(正例)的图片类型,机器不确定的图片交由人工审核。因此,衡量智能审核系统性能的指标主要是准确率和自动化率。

通常的自动审核思路是穷举不符合规定的图片(例如水印图、涉黄图、暴恐图、明星脸、广告图等)类型,剩下的图片作为正例自动通过。这样带来的问题是新增的违规内容扩展性不足,另外必须等待所有模型构建完毕才能起到自动化过滤的作用。如果我们能主动挖掘符合规定的图片(例如正常人物图、场景一致图)进行自动通过,将正例过滤和负例过滤相结合,这样才能更快起到节省人工审核的作用。因此,我们的图像智能审核系统分为图片负例过滤模块和图片正例过滤模块,待审图片先进入负例过滤模块判断是否违禁,再进入正例过滤模块进行自动通过,剩余机器不确定的图片交由人工审核。整个技术方案如图1所示。

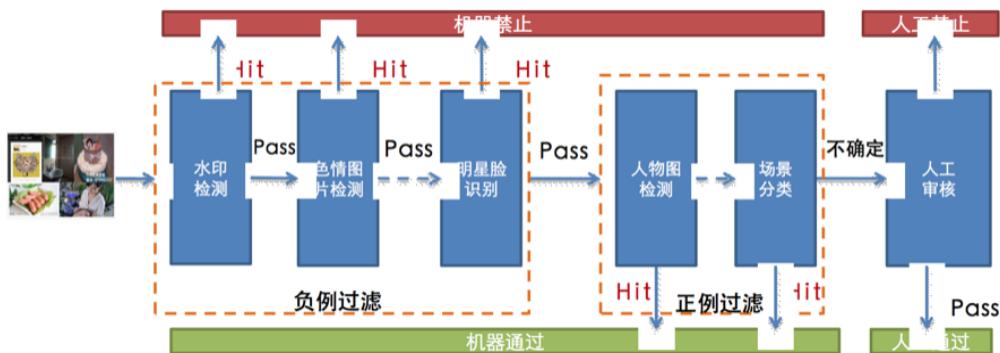


图1 图像智能审核技术方案

负例过滤和正例过滤模块中都会涉及检测、分类和识别等技术,而深度学习则是该领域的首选技术。下面将分别以水印过滤、明星脸识别、色情图片检测和场景分类来介绍深度学习在图像智能审核中的应用。

基于深度学习的水印检测

为了保护版权和支持原创内容,需要自动检测商家或用户上传的图片中是否包括违禁水印(竞对水印、其他产品的Logo)。与其他类刚体目标不同,水印具有以下特点。
 * 样式多。线下收集所涉及的主流违禁水印有20多类,每一类水印又存在多种样式。除此之外,线上存在大量未知类型的水印。
 * 主体多变。

水印在图片中位置不固定且较小，主体存在裁切变形，并且会存在多个主体交叠（多重水印），如图2所示。



图2 主体多变

- 背景复杂。由于主流水印大多采用透明或半透明方式，这使得水印中的文字标识极易受到复杂背景的干扰，如图3所示。



图3 背景复杂

传统的水印检测采用滑动窗口的方法，提取一个固定大小的图像块输入到提前训练好的鉴别模型中，得到该块的一个类别。这样遍历图片中的所有候选位置，可得到一个图片密集的类别得分图。得分高于一定阈值的块被认为是水印候选区域，通过非极大化抑制可以得到最终的结果。鉴别模型的特征可以采用文字识别领域常用的边缘方向统计特征，也可以通过CNN进行特征学习来提升对裁切、形变、复杂背景的健壮性。为了进一步改善得分的置信度，可以加入类型原型的信息，把输入图像块特征与聚类中心特征的相似度（夹角余弦）作为识别置信度。但上述方法检测效率极低，由于水印位置和大小不固定，需要在所有位置对多个尺度的图像进行判别，由此产生大量的冗余窗口。

一种思路是旨在减少滑动窗口数目的子窗口的方法。首先通过无监督/有监督学习生成一系列的候选区域，再通过一个CNN分类器来判断区域中是否包含目标以及是哪一类目标。这类方法比较有代表的是R-CNN系列。由于该类方法得到的候选框可以映射到原图分辨率，因此定位框精度足够高。

另一种解决思路时采用直接在特征图上回归的方法。我们知道，对于CNN网络的卷积层而言，输入图片大小可以不固定，但从全连接层之后就要求输入大小保持一致。因此当把任意大小的图片输入CNN直到第一个全连接层，只需要一次前向运算就可以得到所有层的特征图。然后回归的对象是待检测目标的位置信息和类别信息，它们可根据目标大小的需要在不同层次的特征图上进行回归，这类方法以Yolo、SSD为代表。该类方法的特点是在保证高检测精度的前提下实时性较好。

图4给出了上述两类框架与DPM（可变形部件模型）最佳传统方法的性能比较：

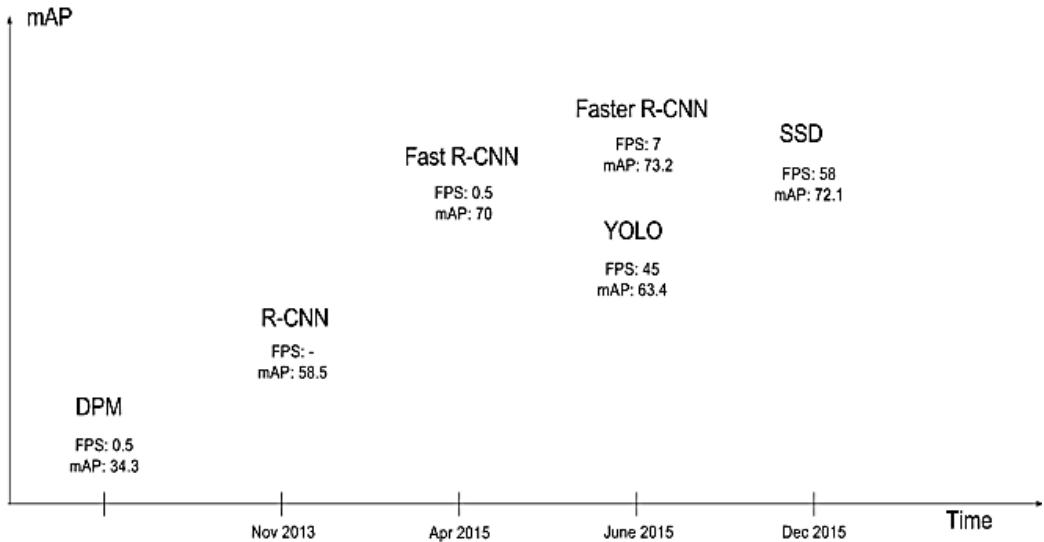


图4 基于深度学习的主流目标检测方法的性能评测

考虑到水印检测任务对定位框的精度要求不高，且需要满足每天百万量级图片的吞吐量，我们借鉴了SSD框架和Resnet网络结构。在训练数据方面，我们通过人工收集了25类共计1.5万张水印图片，并通过主体随机裁切、前背景合成等方式进行了数据增广。

基于训练得到的模型对线上数据进行了相关测试。随机选取3197张线上图片作为测试集，其中2795张图片不包含水印，包含水印的402张图片里有302张包含训练集中出现过的水印，另外的100张包含未出现在训练集中的小众水印。基于该测试集，我们评测了传统方法（人工设计特征+滑窗识别）和基于SSD框架的方法。

从图5可以看到，相比于传统方法，SSD框架无论在召回和精度上都有明显优势。进一步分析发现，深度学习方法召回了38张小众水印图片，可见CNN学习到的特征泛化能力更强。

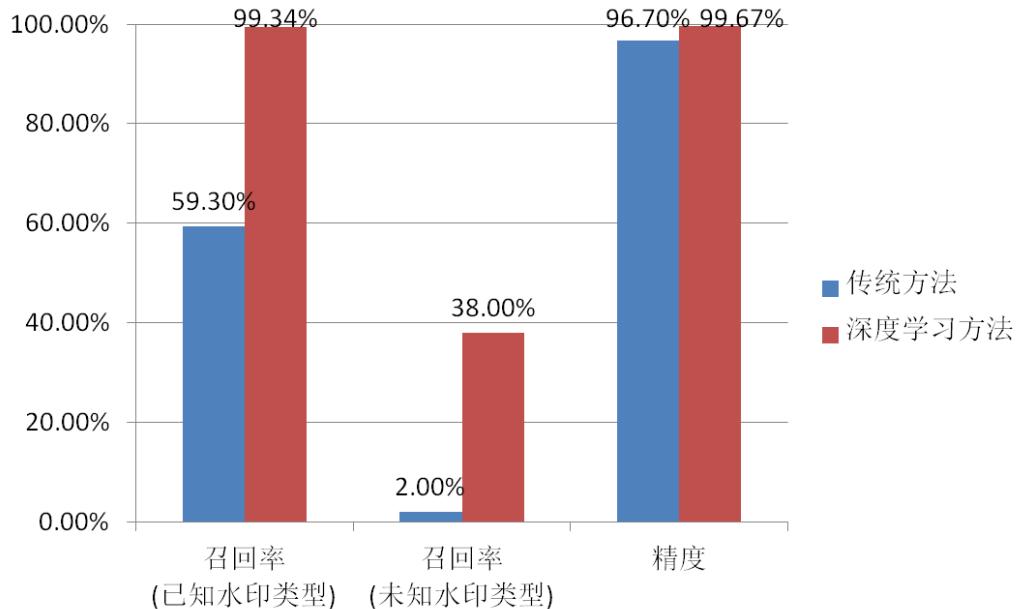


图5 水印检测性能评测

明星脸识别

为了避免侵权明星肖像权，审核场景需要鉴别用户/商家上传的图像中是否包含明星的头像。这是一类典型的人脸识别应用，具体来说是一种 $1:(N+1)$ 的人脸比对。整个人脸识别流程包含人脸检测、人脸关键点检测、人脸矫正及归一化、人脸特征提取和特征比对，如图6所示。其中深度卷积模型是待训练的识别模型，用于特征提取。下面我们将分别介绍人脸检测和人脸识别技术方案。

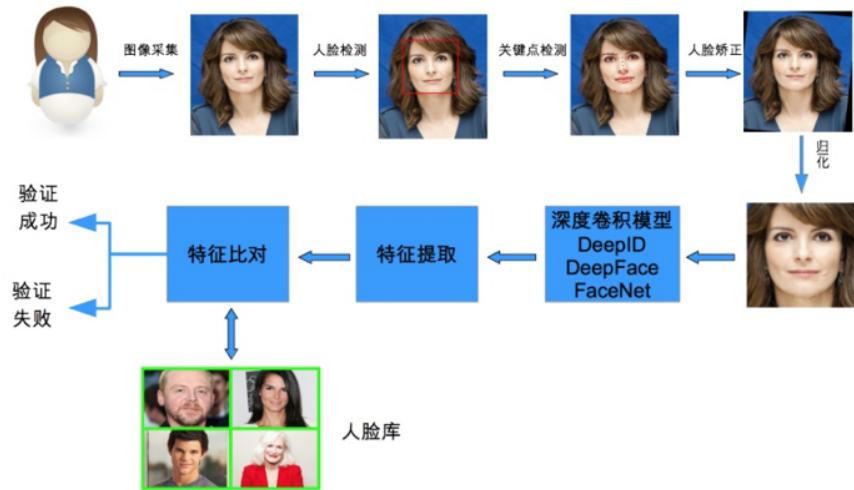


图6 明星脸识别流程

人脸检测方法可分为传统检测器和基于深度学习的检测器两类。

传统检测器主要基于V-J框架，通过设计Boosted的级连结构和人工特征实现检测。特征包括Harr特征、HOG特征和基于像素点比较的特征（Pico、NPD）等。这类检测器在约束环境下有着不错的检测效果和运行速度，但对于复杂场景（光照、表情、遮挡），人工设计的特征使检测能力会大大下降。为了提升性能，相关研究联合人脸检测和人脸关键点定位这两个任务进行联合优化（JDA），将关键点检测作为人脸检测的一个重要评价标准，但其准确率有待进一步提升。

深度学习的检测器有三种思路。第一类是沿用V-J框架，但以级联CNN网络（Cascaded CNN）替代传统特征。第二类是基于候选区域和边框回归的框架（如Faster R-CNN）。第三类是基于全卷积网络直接回归的框架（如DenseBox）。

我们采用了Faster R-CNN框架并从以下方面进行了改进：难分负例挖掘（抑制人物雕像、画像和动物头像等负例）、多层次特征融合、多尺度训练和测试、上下文信息融合，从而更好地抵抗复杂背景、类人脸、遮挡等干扰，并有效提升了对小脸、侧脸的检出率。

人脸识别主要有两种思路。一种是直接转换为图像分类任务，每一类对应一个人的多张照片，比较有代表性的方法有DeepFace、DeepID等。另一种则将识别转换为度量学习问题，通过特征学习使得来自同一个人的不同照片距离比较近、不同的人的照片距离比较远，比较有代表性的方法有DeepID2、FaceNet等。

由于任务中待识别ID是半封闭集合，我们可以融合图像分类和度量学习的思路进行模型训练。考虑到三元组损失（Triplet Loss）对负例挖掘算法的要求很高，在实际训练中收敛很慢，因此我们采用了Center Loss来最小化类内方差，同时联合Softmax Loss来最大化类间方差。为了平衡这两个损失函数，需要通过试验来选择超参数。我们采用的网络结构是Inception-v3，在实际训练中分为两个阶段：第一阶段采用

Softmax Loss+CxCenterLoss，并利用公开数据集CASIA-WebFace（共包含10 575个ID和49万人脸图片）来进行网络参数的初始化和超参数C的优选，根据试验得到的C=0.01；第二阶段采用Softmax Loss+0.01xCenter Loss，并在业务数据（5200个明星脸ID和100万人脸图片）上进行网络参数的微调。

为了进一步提升性能，借鉴了百度采用的多模型集成策略，如图7所示。具体来说，根据人脸关键点的位置把人脸区域分割为多个区域，针对每一个区域分别训练特征模型。目前把人脸区域分割为9个区域，加上人脸整体区域，共需训练10个模型。

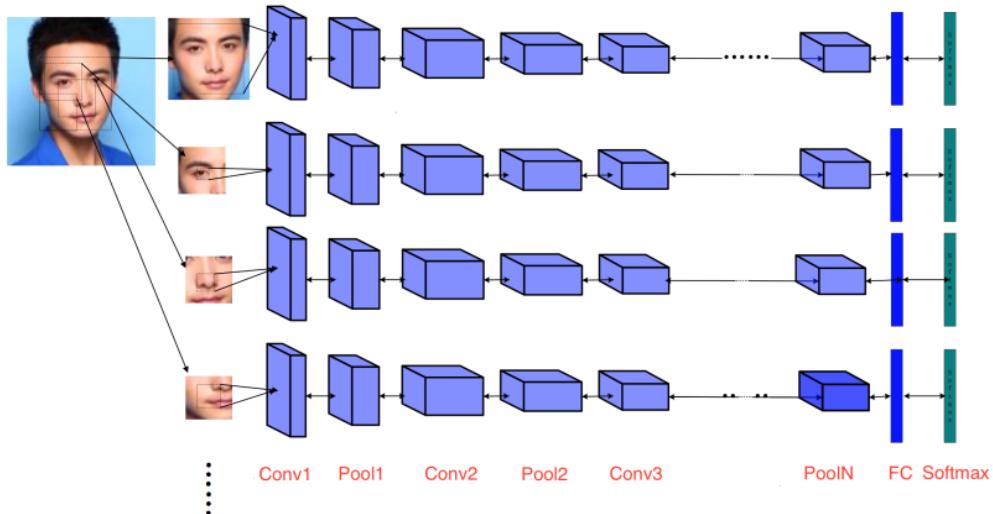


图7 基于集成学习的人脸识别

在测试阶段，对于待验证的人脸区域和候选人脸区域，分别基于图7所示的10个区域提取特征。然后对于每个区域，计算两个特征向量间的相似度（余弦距离）。最终通过相似度加权的方法判断两张人脸是否属于同一个人。表1给出了主流方法在LFW数据集上的评测结果。可以看出，美团模型在相对有限数据下获得了较高的准确率。

	训练集	网络数量	LFW准确率
DeepFace	400万	3	97.35%
DeepID	20万	25	97.45%
DeepID2+	29万	25	99.47%
FaceNet	2亿	1	99.63%
美团	150万	10	99.75%

表1 公开数据集评测结果

色情图片检测

色情图片检测是图像智能审核中重要环节。传统检测方法通过肤色、姿态等维度对图片的合规性进行鉴别。随着深度学习的进展，现有技术 [雅虎NSFW（Not Suitable for Work）模型] 直接把色情图片检测定义二分类（色情、正常）问题，通过卷积神经网络在海量数据上进行端到端训练。

对于已训练模型，不同层次学到的特征不同，有些层次学到了肤色特征，另外一些层次学习到了部位轮廓特征，还有的层次学到了姿态特征。但由于人类对色情的定义非常广泛，露点、性暗示、艺术等都可能被归为色情类，而且在不同的场景下或者面对不同的人群，色情定义标准无法统一。因此，初始学习到的

模型泛化能力有限。为了提升机器的预测准确率，需要不断加入错分样本，让机器通过增量学习到更多特征以纠正错误。除此之外，我们在以下方面进行了优化。

- 模型细化。我们的分类模型精细化了图片的色情程度：色情、性感、正常人物图、其他类。其中色情、性感、正常人物图互为难分类别，其他类为非人物的正常图片。将性感类别和正常人物图类别从色情类别中分离出来有助于增强模型对色情的判别能力。从表2中可见，相对于雅虎的NSFW模型，我们的模型在召回率方面具有明显优势。

	网络结构	准确率	召回率
NSFW	Resnet-50	81.5%	73.2%
美团	GoogLenet	81.3%	90.5%

表2 色情图片检测准确率

- 机器审核结合人工复审。在实际业务中由于涉黄检测采用预警机制，机器审核环节需要尽可能召回所有疑似图片，再结合适量的人工审核来提升准确率。因此，上层业务逻辑会根据模型预测类别和置信度将图片划分为“确定黄图”“确定非黄图”和“疑似”三部分。“疑似”部分，根据置信度由高到底进行排序，并转交人工复审。在线上业务中，“确定黄图”和“确定非黄图”部分的精度可达到99%以上，而“疑似”部分只占总图片量的3%左右，这样在保证高精度过滤的条件下可大幅节省人力。
- 支持视频内容审核。对于短视频内容的审核，我们通过提取关键帧的方式转化为对单张图片的审核，然后融合多帧的识别结果给出结论。

场景分类

作为一个贯穿吃喝玩乐各环节的互联网平台，美团的业务涉及多种垂直领域，如表3所示。有必要对运营或用户上传图片的品类进行识别，以保持与该商家的经营范围一致。此外，为了进一步改善展示效果，需要对商家相册内的图片进行归类整理，如图8所示。

品类	美食	酒店	休闲娱乐	丽人	旅游	电影	商品	生活服务	运动健身	摄影写真
占比	66.6%	6.3%	7.7%	6.6%	4.5%	3.5%	1.9%	1.1%	1.1%	0.7%

表3 美团一级品类及图片占比

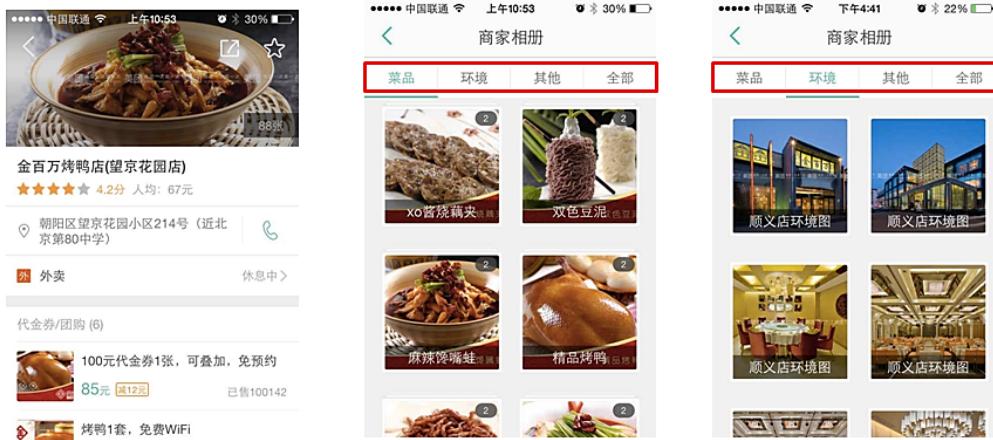


图8 商家相册图片分类

深度卷积神经网络在图像分类的相关任务上（比如ILSVRC）上已经超越人眼的识别率，但作为一种典型的监督学习方法，它对特定领域的标记样本的数量和质量的需求是突出的。我们的场景分类任务，如果完全依靠审核人员进行图片的筛选和清洗，代价较大。因此需要基于迁移学习来对模型进行微调。

迁移学习致力于通过保持和利用从一个或多个相似的任务、领域或概率分布中学习到的知识，来快速并有效地为提升目标任务的性能。模型迁移是迁移学习领域中一类常用的迁移方式，它通过学习原始域（Source Domain）模型和目标域（Target Domain）模型的共享参数来实现迁移。由于深度神经网络具有层次结构，且其隐藏层能表示抽象和不变性的特征，因此它非常适合模型迁移。

至于原始域训练的深度卷积神经网络，需要关注哪些层次的参数可以迁移以及如何迁移。不同层次的可迁移度不同，目标域与原始域中相似度较高的层次被迁移的可能性更大。具体而言，较浅的卷积层学习到的特征更通用（比如图像的色彩、边缘、基本纹理），因而也更适合迁移，较深的卷积层学习的特征更具有任务依赖性（比如图像细节），因而不适合迁移，如图9所示。



图9 深度卷积神经网络的层次结构与特征描述

模型迁移通过固定网络特定层次的参数，用目标域的数据来训练其他层次。对于我们的场景分类任务而言，首先根据分类的类别数修改网络输出层，接着固定较浅的卷积层而基于业务标注数据训练网络倒数若干层参数。如有更多的训练数据可用，还可以进一步微调整整个网络的参数以获得额外的性能提升，如图10所示。相比于直接提取图像的高层语义特征来进行监督学习，采用分阶段的参数迁移对原始域与目标域间的差异性更健壮。

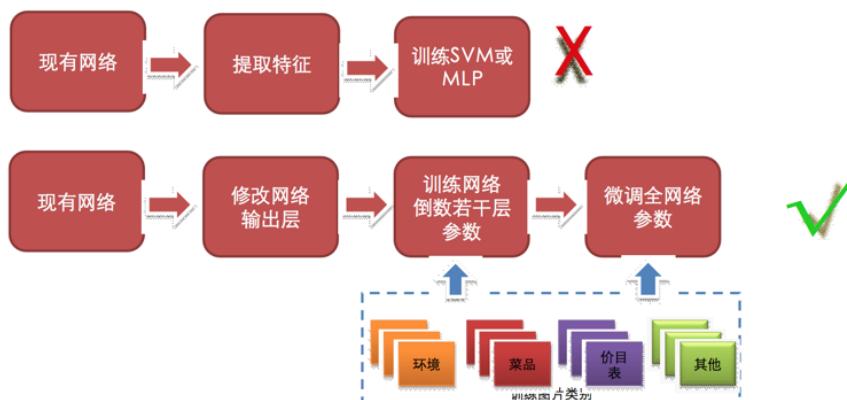


图10 基于深度卷积神经网络的模型迁移

基于上述迁移学习策略，我们在美食场景图和酒店房型图分类中进行了相关实验，基于有限（万级别图片）的标注样本实现了较高的识别准确率，测试集上的性能如表4所示。

	菜品	环境	价目表	其他	实际	查全率(%)
菜品	4773	60	8	5	4846	98.49
环境	42	3890	5	4	3941	98.70
价目表	8	4	88	5	105	83.81
其他	13	15	9	62	99	62.63
预测	4836	3969	110	76	8813/8991	
查准率(%)	98.70	98.00	80.00	81.58		平均准确率=98.02

表4 美食场景分类

如前所述，基于深度学习的图像分类与检测方法在图片智能审核中替代了传统机器学习方法，在公开模型与迁移学习的基础上，通过从海量数据中的持续学习，实现了业务场景落地。

参考文献

- [1] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod. “Robust text detection in natural images with edge-enhanced maximally stable extremal regions.” ICIP 2011.
- [2] Z Zhong,LJin,SZhang,ZFeng.“DeepText: A Unified Framework for Text Proposal Generation and Text Detection in Natural Images”. Architecture Science 2015.
- [3] Minghui Liao, Baoguang Shi, Xiang Bai, Xinggang Wang, Wenyu Liu. “TextBoxes: A Fast Text Detector with a Single Deep Neural Network”. AAAI 2017.
- [4] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn.“Towards real-time object detection with region proposal networks.” NIPS 2015.
- [5] Graves, A.; Fernandez, S.; Gomez, F.; and Schmidhuber, J. “Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks.” ICML 2006.
- [6] R Girshick,JDonahue,TDarrell,JMalik. “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.” CVPR 2014.
- [7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. “You only look once: Unified, real-time object detection”. CVPR 2016.
- [8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. Reed. “SSD: Single shot multibox detector”. ECCV 2016.
- [9] “Object detection with discriminatively trained part-based models”. TPAMI 2010.
- [10] Robust Real-time Object Detection. Paul Viola, Michael Jones. IJCV 2004.
- [11] N. Markus, M. Frljak, I. S. Pandzic, J. Ahlberg and R. Forchheimer. “Object Detection with Pixel Intensity Comparisons Organized in Decision Trees”. CoRR 2014.
- [12] Shengcai Liao, Anil K. Jain, and Stan Z. Li. “A Fast and Accurate Unconstrained Face Detector,” TPAMI 2015.
- [13] Dong Chen, ShaoQingRen, Jian Sun. “Joint Cascade Face Detection and Alignment”, ECCV 2014.
- [14] Haoxiang Li, Zhe Lin, XiaohuiShen, Jonathan Brandt, Gang Hua. “A convolutional neural network cascade for face detection”, CVPR.2015.
- [15] Lichao Huang, Yi Yang, Yafeng Deng, Yinan Yu.“DenseBox: Unifying Landmark Localization with End to End Object Detection” CVPR 2015.

- [16]Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification.CVPR 2014.
- [17]Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes.CVPR 2014.
- [18]Sun Y, Chen Y, Wang X, et al. Deep learning face representation by joint identification–verification.NIPS. 2014.
- [19]FaceNet: A Unified Embedding for Face Recognition and Clustering. CVPR 2015.
- [20]A Discriminative Feature Learning Approach for Deep Face Recognition. ECCV 2016.
- [21]Rethinking the Inception Architecture for Computer Vision. CVPR 2016.
- [22]Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. 2014.
- [23]Murray, N., Marchesotti, L., Perronnin, F. “Ava: A large-scale database for aesthetic visual analysis”. CVPR 2012.

作者简介

- 晓明，美团平台智能技术中心视觉技术负责人，曾就职于佳能研究院，三星研究院。2015年加入美团，主要致力于图像和视频相关的技术积累和业务落地，作为技术负责人主导了图像智能审核、首图优选、刷脸认证、拍照录菜等项目的上线，显著提升了用户和商家的智能化体验。

招聘信息

美团平台智能技术中心充分利用人工智能的优势来支持美团点评多个业务线，并在智能推荐、智能营销、智能经营、智能审核等多个领域都取得了很好的应用效果。长期招聘自然语言处理、计算机视觉、大规模机器学习、数据挖掘算法或工程背景的同学。欢迎有意向的同学投递简历至：

zhanghejia@meituan.com。

深度学习在OCR中的应用

作者: 美团算法团队

背景

计算机视觉是利用摄像机和电脑代替人眼，使得计算机拥有类似于人类的对目标进行检测、识别、理解、跟踪、判别决策的功能。以美团业务为例，在商家上单、团单展示、消费评价等多个环节都会涉及计算机视觉的应用，包括文字识别、图片分类、目标检测和图像质量评价等方向。本文将通过以OCR（光学字符识别）的场景来介绍深度学习在计算机视觉中的应用。

基于深度学习的OCR

文字是不可或缺的视觉信息来源。相对于图像/视频中的其他内容，文字往往包含更强的语义信息，因此对图像中的文字提取和识别具有重大意义。OCR在美团业务中主要起着两方面作用。一方面是辅助录入，比如在移动支付环节通过对银行卡卡号的拍照识别以实现自动绑卡，辅助运营录入菜单中菜品信息，在配送环节通过对商家小票的识别以实现调度核单，如图1所示。另一方面是审核校验，比如在商家资质审核环节对商家上传的身份证件、营业执照和餐饮许可证等证件照片进行信息提取和核验以确保该商家的合法性，机器过滤商家上单和用户评价环节产生的包含违禁词的图片。

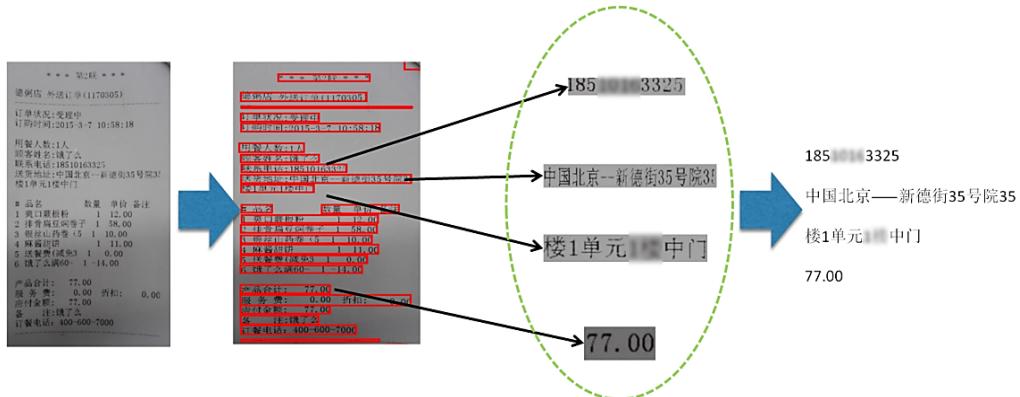


图1 图像中的文字提取和识别流程

OCR技术发展历程

传统的OCR基于图像处理（二值化、连通域分析、投影分析等）和统计机器学习（Adaboost、SVM），过去20年间在印刷体和扫描文档上取得了不错的成绩。传统的印刷体OCR解决方案整体流程如图2所示。

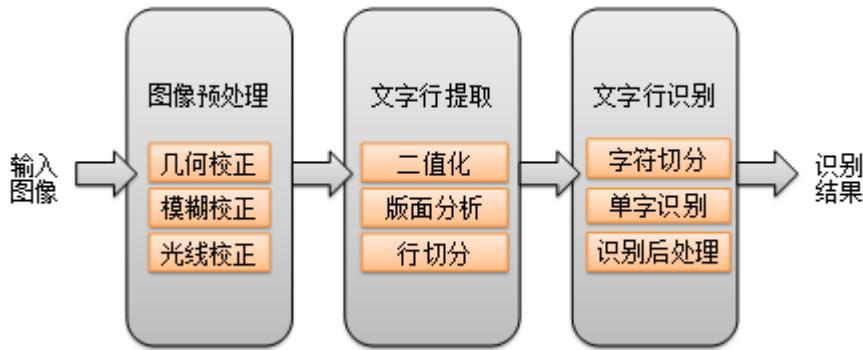


图2 传统的印刷体OCR解决方案

从输入图像到给出识别结果经历了图像预处理、文字行提取和文字行识别三个阶段。其中文字行提取的相关步骤（版面分析、行切分）会涉及大量的先验规则，而文字行识别主要基于传统的机器学习方法。随着移动设备的普及，对拍摄图像中的文字提取和识别成为主流需求，同时对场景中文字的识别需求越来越突出。因此，相比于印刷体场景，拍照文字的识别将面临以下三方面挑战：

- 成像复杂。噪声、模糊、光线变化、形变。
- 文字复杂。字体、字号、色彩、磨损、笔画宽度任意、方向任意。
- 场景复杂。版面缺失、背景干扰。

对于上述挑战，传统的OCR解决方案存在着以下不足：

- 通过版面分析（连通域分析）和行切分（投影分析）来生成文本行，要求版面结构有较强的规则性且背景可分性强（例如黑白文档图像、车牌），无法处理前背景复杂的随意文字（例如场景文字、菜单、广告文字等）。另外，二值化操作本身对图像成像条件和背景要求比较苛刻。
- 通过人工设计边缘方向特征（例如方向梯度直方图）来训练字符识别模型，在字体变化、模糊或背景干扰时，此类单一的特征的泛化能力迅速下降。
- 过度依赖于字符切分的结果，在字符扭曲、粘连、噪声干扰的情况下，切分的错误传播尤其突出。
- 尽管图像预处理模块可有效改善输入图像的质量，但多个独立的校正模块的串联必然带来误差传递。另外由于各模块优化目标独立，它们无法融合到统一的框架中进行。

为了解决上述问题，现有技术在以下三方面进行了改进。

1. 文字行提取

传统OCR（如图3所示）采取自上而下的切分式，但它只适用于版面规则背景简单的情况。该领域还有另外两类思路。

- 自底向上的生成式方法。该类方法通过连通域分析或最大稳定极值区域（MSER）等方法提取候选区域，然后通过文字/非文字的分类器进行区域筛选，对筛选后的区域进行合并生成文字行，再进行文字行级别的过滤，如图3所示。该类方法的不足是，一方面流程冗长导致的超参数过多，另一方面无法利用全局信息。

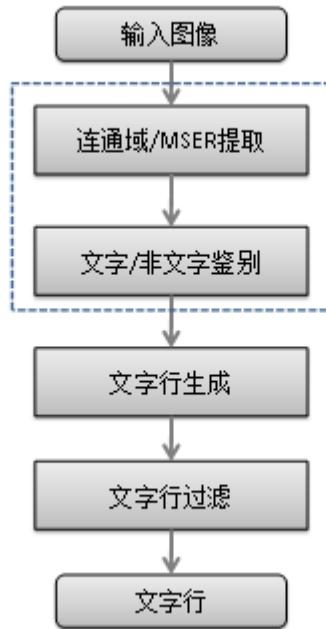


图3 基于自底向上的文字检测

- 基于滑动窗口的方法。该类方法利用通用目标检测的思路来提取文字行信息，利用训练得到的文字行/词语/字符级别的分类器来进行全图搜索。原始的基于滑动窗口方法通过训练文字/背景二分类检测器，直接对输入图像进行多尺度的窗口扫描。检测器可以是传统机器学习模型（Adaboost、Random Ferns），也可以是深度卷积神经网络。

为了提升效率，DeepText、TextBoxes等方法先提取候选区域再进行区域回归和分类，同时该类方法可进行端到端训练，但对多角度和极端宽高比的文字区域召回低。

2. 传统单字识别引擎→基于深度学习的单字识别引擎

由于单字识别引擎的训练是一个典型的图像分类问题，而卷积神经网络在描述图像的高层语义方面优势明显，所以主流方法是基于卷积神经网络的图像分类模型。实践中的关键点在于如何设计网络结构和合成训练数据。对于网络结构，我们可以借鉴手写识别领域相关网络结构，也可采用OCR领域取得出色效果的Maxout网络结构，如图4所示。对于数据合成，需考虑字体、形变、模糊、噪声、背景变化等因素。

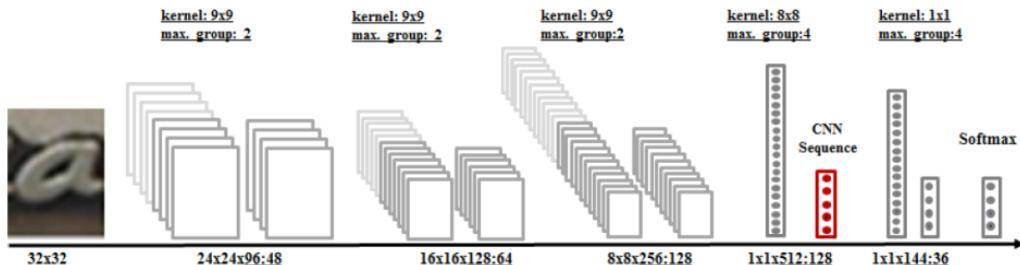


图4 基于Maxout卷积神经网络结构的单字识别模型

表1给出了卷积神经网络的特征学习和传统特征的性能比较，可以看出通过卷积神经网络学习得到的特征鉴别能力更强。

	Top1错误率（英文）	Top1错误率（中文）
HOG特征+SVM	2.4%	4.9%
CNN特征+SVM	1.8%	3.6%

表1 单字符识别引擎性能比较

3. 文字行识别流程

传统OCR将文字行识别划分为字符切分和单字符识别两个独立的步骤，尽管通过训练基于卷积神经网络的单字符识别引擎可以有效提升字符识别率，但切分对于字符粘连、模糊和形变的情况的容错性较差，而且切分错误对于识别是不可修复的。因此在该框架下，文本行识别的准确率主要受限于字符切分。假设已训练单字符识别引擎的准确率 $p=99\%$ ，字符切分准确率为 $q=95\%$ ，则对于一段长度为 L 的文字行，其识别的平均准确率为 $P=(pq)^L$ ，其中 $L=10$ 时， $P=54.1\%$ 。

由于独立优化字符切分提升空间有限，因此有相关方法试图联合优化切分和识别两个任务。现有技术主要可分为基于切分的方法（Segmentation-Based）和不依赖切分的方法（Segmentation-Free）两类方法。

基于切分的方法

该类方法还是保留主动切分的步骤，但引入了动态合并机制，通过识别置信度等信息来指导切分，如图5所示。

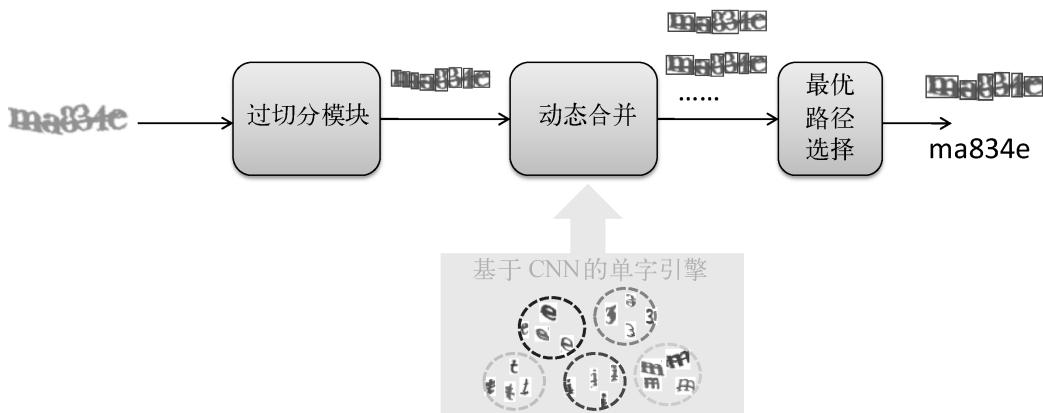


图5 基于CNN的动态切分与识别流程

过切分模块将文字行在垂直于基线方向上分割成碎片，使得其中每个碎片至多包含一个字符。通常来说，过切分模块会将字符分割为多个连续笔划。过切分可以采用基于规则或机器学习的方法。规则方法主要是直接在图像二值化的结果上进行连通域分析和投影分析来确定候补切点位置，通过调整参数可以控制粒度来使得字符尽可能被切碎。基于规则的方法实现简单，但在成像/背景复杂的条件下其效果不好。机器学习方法通过离线训练鉴别切点的二类分类器，然后基于该分类器在文字行图像上进行滑窗检测。

动态合并模块将相邻的笔划根据识别结果组合成可能的字符区域，最优组合方式即对应最佳切分路径和识别结果。直观来看，寻找最优组合方式可转换为路径搜索问题，对应有深度优先和广度优先两种搜索策略。深度优先策略在每一步选择扩展当前最优的状态，因此全局来看它是次优策略，不适合过长的文字行。广度优先策略在每一步会对当前多个状态同时进行扩展，比如在语音识别领域广泛应用的Viterbi解码和Beam Search。但考虑到性能，Beam Search通常会引入剪枝操作来控制路径长度，剪枝策略包含限制扩展的状态数（比如，每一步只扩展TopN的状态）和加入状态约束（比如，合并后字符形状）等。

由于动态合并会产生多个候选路径，所以需要设计合适的评价函数来进行路径选择。评价函数的设计主要从路径结构损失和路径识别打分两方面出发。路径结构损失主要从字符形状特征方面衡量切分路径的合理性，路径识别打分则对应于特定切分路径下的单字平均识别置信度和语言模型分。

该方案试图将字符切分和单字符识别融合在同一个框架下解决，但由于过分割是独立的步骤，因此没有从本质上实现端到端学习。

不依赖切分的方法

该类方法完全跨越了字符切分，通过滑动窗口或序列建模直接对文字行进行识别。

滑窗识别借鉴了滑动窗口检测的思路，基于离线训练的单字识别引擎，对文字行图像从左到右进行多尺度扫描，以特定窗口为中心进行识别。在路径决策上可采用贪心策略或非极大值抑制（NMS）策略来得到最终的识别路径。图6给出了滑窗识别的示意流程。可见滑窗识别存在两个问题：滑动步长的粒度过细则计算代价大，过粗则上下文信息易丢失；无论采用何种路径决策方案，它们对单字识别的置信度依赖较高。

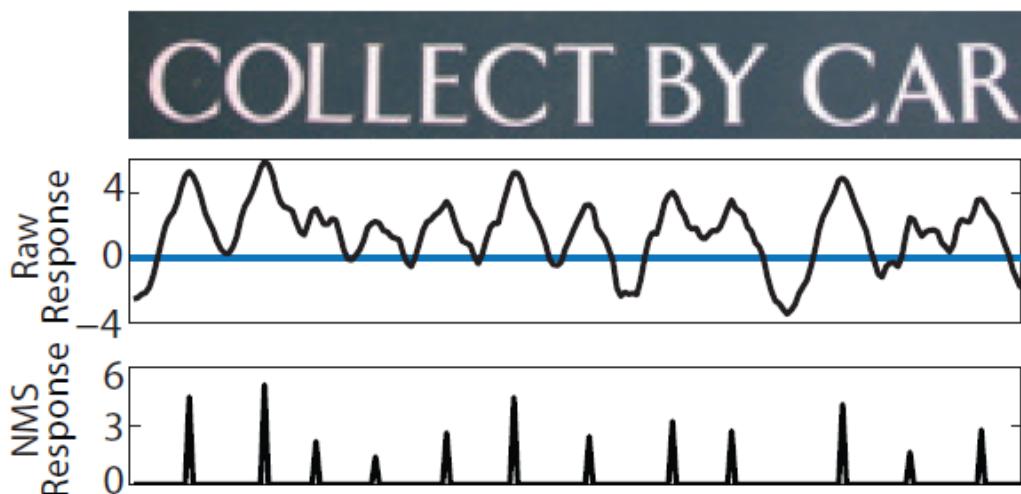


图6 基于滑动窗口的文字检测

序列学习起源于手写识别、语音识别领域，因为这类问题的共同特点是需要对时序数据进行建模。尽管文字行图像是二维的，但如果把从左到右的扫描动作类比为时序，文字行识别从本质上也可归为这类问题。通过端到端的学习，摒弃矫正/切分/字符识别等中间步骤，以此提升序列学习的效果，这已经成为当前研究的热点。

基于现有技术和美团业务涉及的OCR场景，我们在文字检测和文字行识别采用如图7所示的深度学习框架。

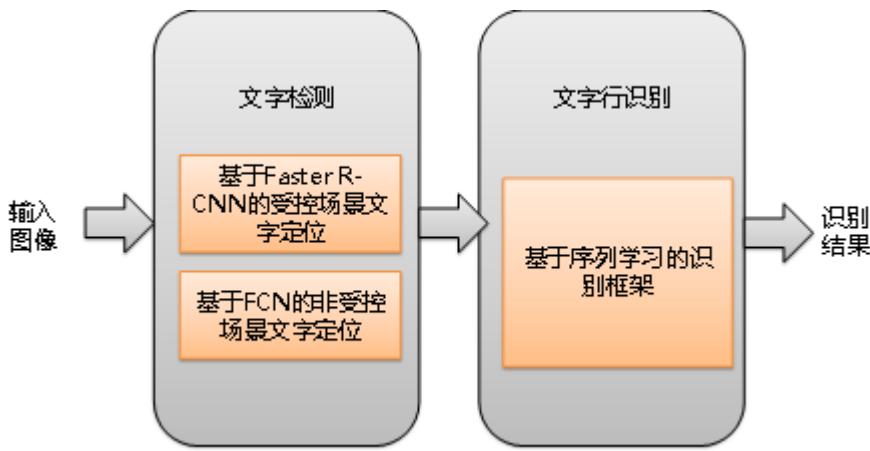


图7 基于深度学习的OCR解决方案

后面将分别介绍文字检测和文字行识别这两部分的具体方案。

基于深度学习的文字检测

对于美团的OCR场景，根据版面是否有先验信息（卡片的矩形区域、证件的关键字段标识）以及文字自身的复杂性（如水平文字、多角度），图像可划分为受控场景（如身份证、营业执照、银行卡）和非受控场景（如菜单、门头图），如图8所示。



图8 受控场景与非受控场景

考虑到这两类场景的特点不同，我们借鉴不同的检测框架。由于受控场景文字诸多约束条件可将问题简化，因此利用在通用目标检测领域广泛应用的Faster R-CNN框架进行检测。而对于非受控场景文字，由于形变和笔画宽度不一致等原因，目标轮廓不具备良好的闭合边界，我们需要借助图像语义分割来标记文字区域与背景区域。

1. 受控场景的文字检测

对于受控场景（如身份证件），我们将文字检测转换为对关键字目标（如姓名、身份证号、地址）或关键条目（如银行卡号）的检测问题。基于Faster R-CNN的关键字检测流程如图9所示。为了保证回归框的定位精度，同时提升运算速度，我们对原有框架和训练方式进行了微调。

- 考虑到关键字或关键条目的类内变化有限，网络结构只采用了3个卷积层。
- 训练过程中提高正样本的重叠率阈值。
- 根据关键字或关键条目的宽高比范围来适配RPN层Anchor的宽高比。

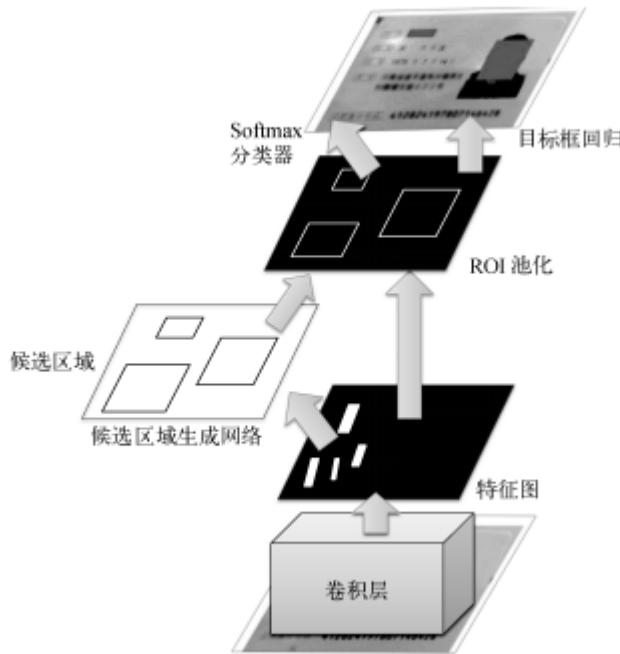


图9 基于Faster R-CNN的OCR解决方案

Faster R-CNN框架由RPN（候选区域生成网络）和RCN（区域分类网络）两个子网络组成。RPN通过监督学习的方法提取候选区域，给出的是无标签的区域和粗定位结果。RCN引入类别概念，同时进行候选区域的分类和位置回归，给出精细定位结果。训练时两个子网络通过端到端的方式联合优化。图10以银行卡卡号识别为例，给出了RPN层和RCN层的输出。



图10 基于Faster R-CNN的银行卡卡号检测

对于人手持证件场景，由于证件目标在图像中所占比例过小，直接提取微小候选目标会导致一定的定位精度损失。为了保证高召回和高定位精度，可采用由粗到精的策略进行检测。首先定位卡片所在区域位置，然后在卡片区域范围内进行关键字检测，而区域定位也可采用Faster R-CNN框架，如图11所示。

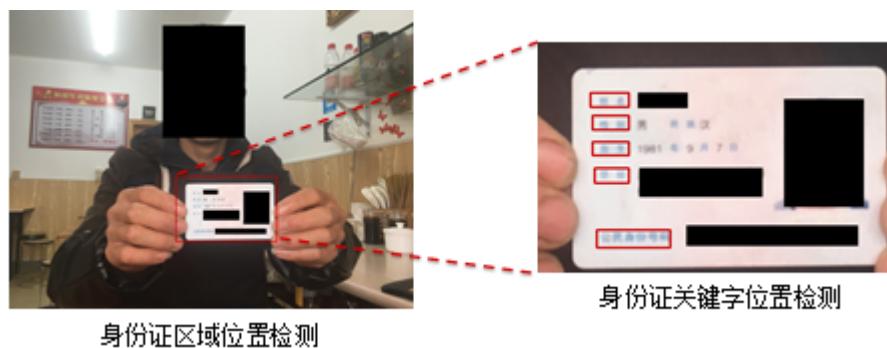


图11 由粗到精的检测策略

2. 非受控场景的文字检测

对于菜单、门头图等非受控场景，由于文字行本身的多角度且字符的笔画宽度变化大，该场景下的文字行定位任务挑战很大。由于通用目标检测方法的定位粒度是回归框级，此方法适用于刚体这类有良好闭合边界的物体。然而文字往往由一系列松散的笔画构成，尤其对于任意方向或笔画宽度的文字，仅以回归框结果作为定位结果会有较大偏差。另外刚体检测的要求相对较低，即便只定位到部分主体（如定位结果与真值的重叠率是50%），也不会对刚体识别产生重大影响，而这样的定位误差对于文字识别则很可能是致命的。

为了实现足够精细的定位，我们利用语义分割中常用的全卷积网络（FCN）来进行像素级别的文字/背景标注，整体流程如图12所示。

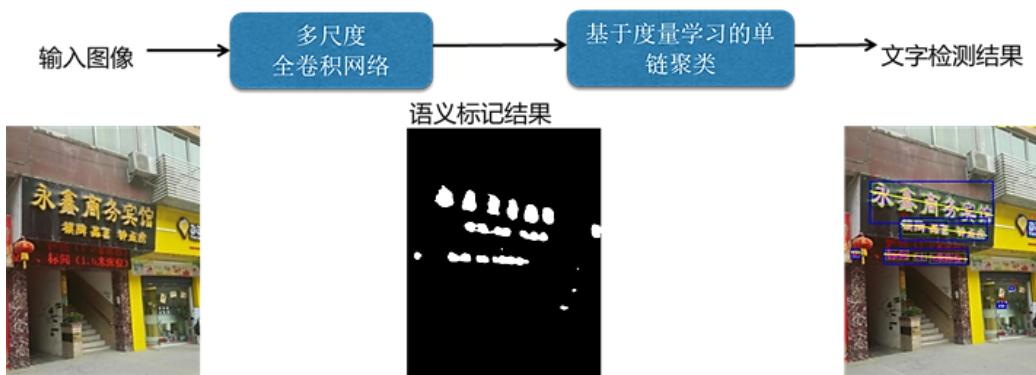


图12 基于全卷积网络的文字检测

多尺度全卷积网络通过对多个阶段的反卷积结果的融合，实现了全局特征和局部特征的联合，进而达到了由粗到精的像素级别标注，适应于任意非受控场景（门头图、菜单图片）。

基于多尺度全卷积网络得到的像素级标注，通过连通域分析技术可得到一系列连通区域（笔划信息）。但由于无法确定哪些连通域属于同一文字行，因此需要借助单链聚类技术来进行文字行提取。至于聚类涉及的距离度量，主要从连通域间的距离、形状、颜色的相似度等方面提取特征，并通过度量学习自适应地得到特征权重和阈值，如图13所示。

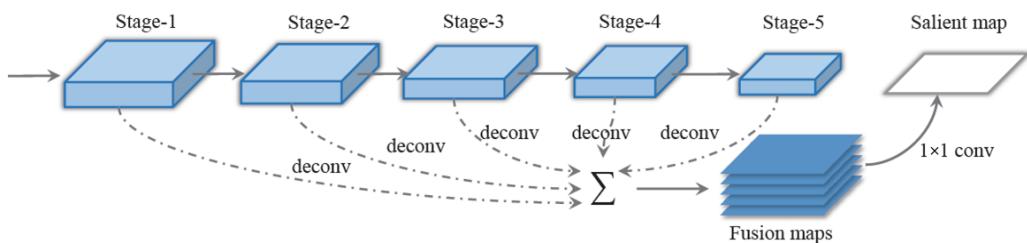


图13 基于全卷积网络的图像语义分割

图14分别给出了在菜单和门头图场景中的全卷积网络定位效果。第二列为全卷积网络的像素级标注结果，第三列为最终文字检测结果。可以看出，全卷积网络可以较好地应对复杂版面或多角度文字定位。



图14 基于FCN的文字定位结果

基于序列学习的文字识别

我们将整行文字识别问题归结为一个序列学习问题。利用基于双向长短期记忆神经网络（Bi-directional Long Short-term Memory, BLSTM）的递归神经网络作为序列学习器，来有效建模序列内部关系。为了引入更有效的输入特征，我们采用卷积神经网络模型来进行特征提取，以描述图像的高层语义。此外在损失函数的设计方面，考虑到输出序列与输入特征帧序列无法对齐，我们直接使用结构化的Loss（序列对序列的损失），另外引入了背景（Blank）类别以吸收相邻字符的混淆性。

整体网络结构分为三层：卷积层、递归层和翻译层，如图15所示。其中卷积层提取特征；递归层既学习特征序列中字符特征的先后关系，又学习字符的先后关系；翻译层实现对时间序列分类结果的解码。

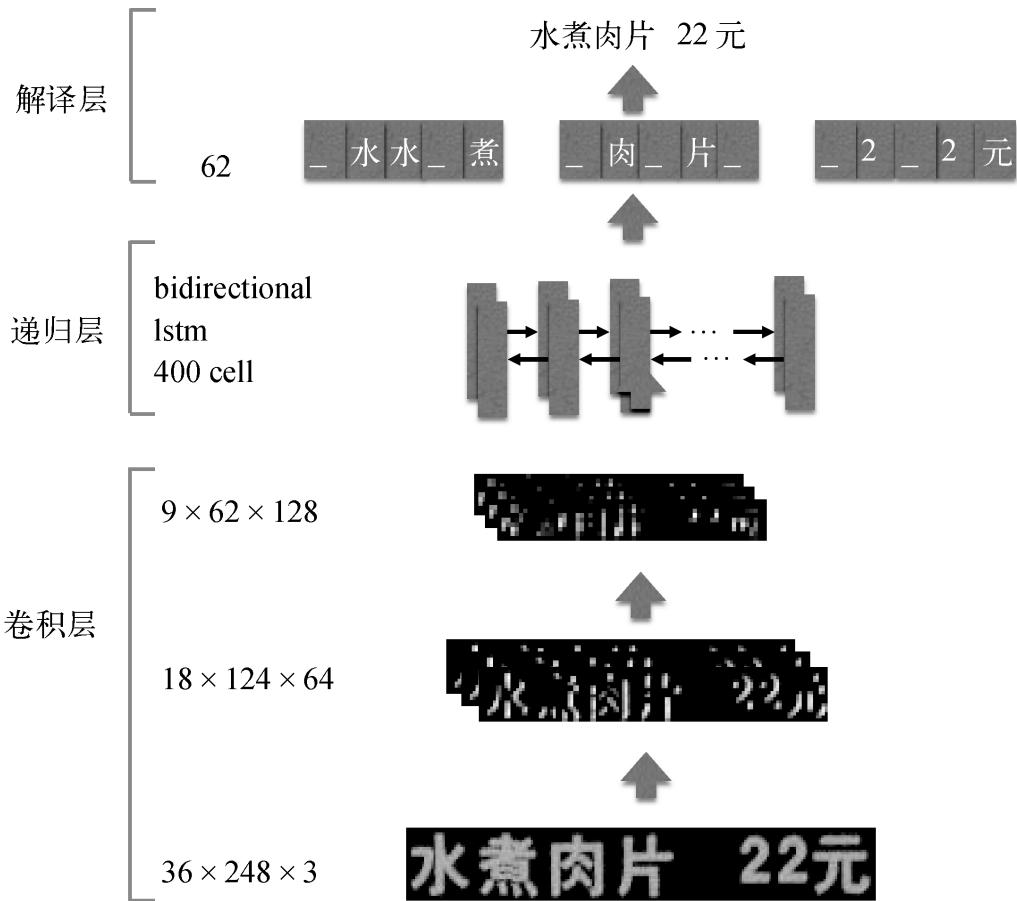


图15 基于序列学习的端到端识别框架

对于输入的固定高度 $h_0 = 36$ 的图像（宽度任意，如 $W_0 = 248$ ），我们通过CNN网络结构提取特征，得到 $9 \times 62 \times 128$ 的特征图，可将其看作一个长度为62的时间序列输入到RNN层。RNN层有400个隐藏节点，其中每个隐藏节点的输入是 9×128 维的特征，是对图像局部区域的描述。考虑到对应于某个时刻特征的图像区域，它与其前后内容都具有较强的相关性，所以我们一般采用双向RNN网络，如图16所示。

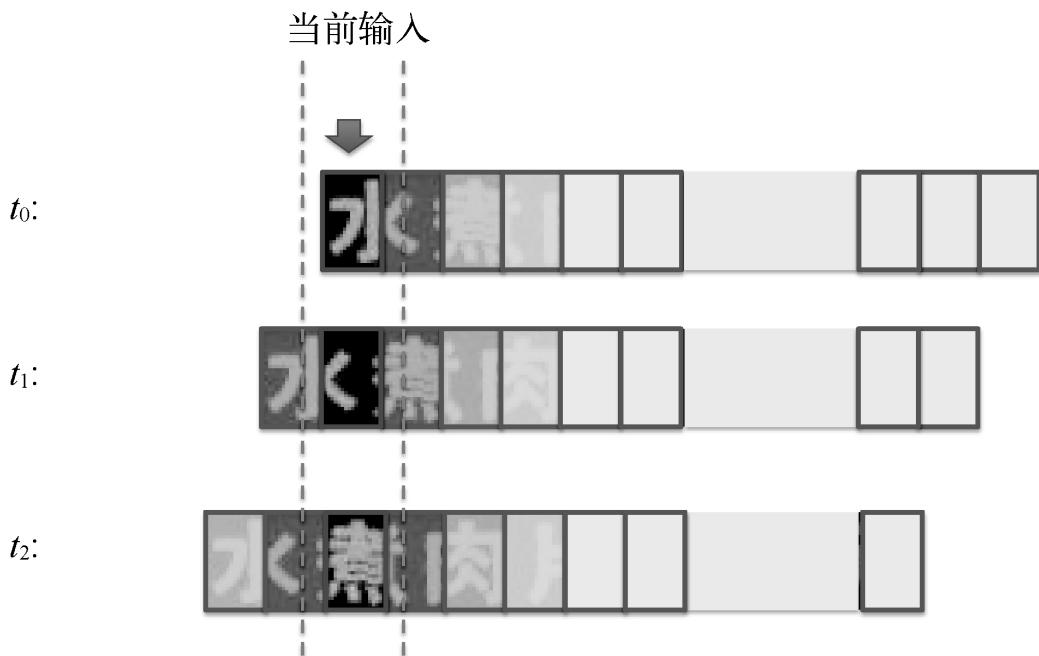


图16 双向RNN序列

双向RNN后接一个全连接层，输入为RNN层（在某个时刻）输出的特征图，输出为该位置是背景、字符表中文字的概率。全连接层后接CTC（联结主义时间分类器）作为损失函数。在训练时，根据每个时刻对应的文字、背景概率分布，得到真值字符串在图像中出现的概率 $P(\text{ground truth})$ ，将 $-\log(P(\text{ground truth}))$ 作为损失函数。在测试时，CTC可以看作一个解码器，将每一时刻的预测结果（当前时刻的最大后验概率对应的字符）联合起来，然后去掉空白和重复的模式，就形成了最终的序列预测结果，如图17所示。

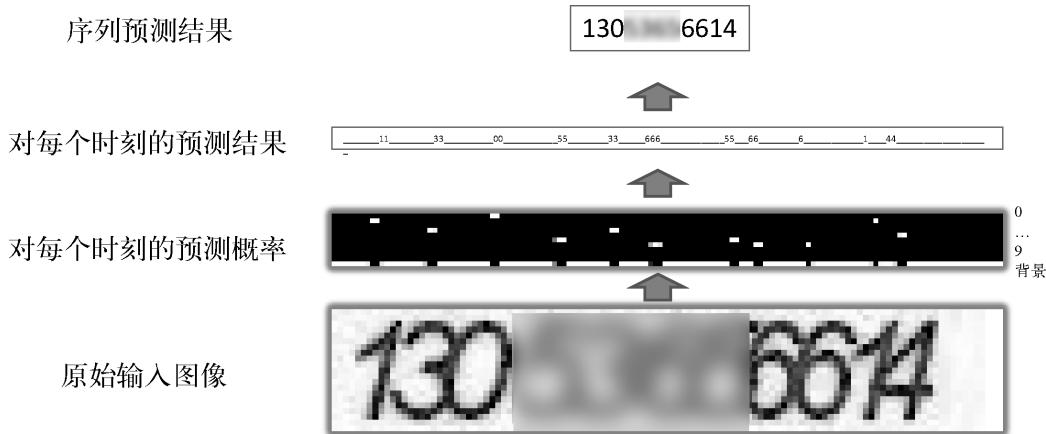


图17 CTC解码过程

从图17中也可以看出，对应输入序列中的每个字符，LSTM输出层都会产生明显的尖峰，尽管该尖峰未必对应字符的中心位置。换句话说，引入CTC机制后，我们不需要考虑每个字符出现的具体位置，只需关注整个图像序列对应的文字内容，最终实现深度学习的端到端训练与预测。

由于序列学习框架对训练样本的数量和分布要求较高，我们采用了真实样本+合成样本的方式。真实样本以美团业务来源（例如，菜单、身份证件、营业执照）为主，合成样本则考虑了字体、形变、模糊、噪声、背景等因素。

基于上述序列学习框架，我们给出了在不同场景下的文字行识别结果，如图18所示。其中前两行的图片为验证码场景，第三行为银行卡，第四行为资质证件，第五行为门头图，第六行为菜单。可以看到，识别模型对于文字形变、粘连、成像的模糊和光线变化、背景的复杂等都有较好的健壮性。



图18 文字行识别结果

基于上述试验，与传统OCR相比，我们在多种场景的文字识别上都有较大幅度的性能提升，如图19所示：

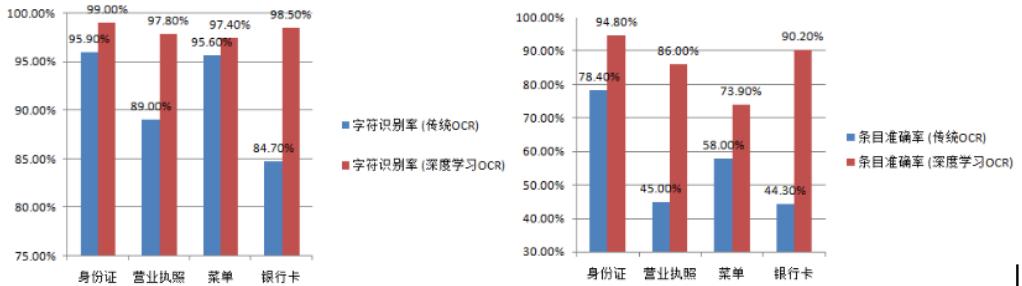


图19 传统OCR和深度学习OCR性能比较

与传统OCR相比，基于深度学习的OCR在识别率方面有了大幅上升。但对于特定的应用场景（营业执照、菜单、银行卡等），条目准确率还有待提升。一方面需要融合基于深度学习的文字检测与传统版面分析技术，以进一步提升限制场景下的检测性能。另一方面需要丰富真实训练样本和语言模型，以提升文字识别准确率。

参考文献

- [1] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod. “Robust text detection in natural images with edge-enhanced maximally stable extremal regions.” ICIP 2011.
- [2] Z. Zhong, L. Jin, S. Zhang, Z. Feng. “DeepText: A Unified Framework for Text Proposal Generation and Text Detection in Natural Images”. Architecture Science 2015.
- [3] Minghui Liao, Baoguang Shi, Xiang Bai, Xinggang Wang, Wenyu Liu. “TextBoxes: A Fast Text Detector with a Single Deep Neural Network”. AAAI 2017.
- [4] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn. “Towards real-time object detection with region proposal networks.” NIPS 2015.

- [5]Graves, A.; Fernandez, S.; Gomez, F.; and Schmidhuber, J. “Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks.” ICML 2006.
- [6]R Girshick, JDonahue, TDarrell, JMalik. “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.” CVPR 2014.
- [7]J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. “You only look once: Unified, real-time object detection”. CVPR 2016.
- [8]W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. Reed. “SSD: Single shot multibox detector”. ECCV 2016.
- [9]“Object detection with discriminatively trained part-based models”. TPAMI 2010.
- [10]Robust Real-time Object Detection. Paul Viola, Michael Jones. IJCV 2004.
- [11]N. Markus, M. Frljak, I. S. Pandzic, J. Ahlberg and R. Forchheimer. “Object Detection with Pixel Intensity Comparisons Organized in Decision Trees”. CoRR 2014.
- [12]Shengcai Liao, Anil K. Jain, and Stan Z. Li. “A Fast and Accurate Unconstrained Face Detector,” TPAMI 2015.
- [13]Dong Chen, ShaoQingRen, Jian Sun. “Joint Cascade Face Detection and Alignment”, ECCV 2014.
- [14]Haoxiang Li, Zhe Lin, XiaohuiShen, Jonathan Brandt, Gang Hua. “A convolutional neural network cascade for face detection”, CVPR.2015.
- [15]Lichao Huang, Yi Yang, Yafeng Deng, Yinan Yu.“DenseBox: Unifying Landmark Localization with End to End Object Detection” CVPR 2015.
- [16]Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification.CVPR 2014.
- [17]Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes.CVPR 2014.
- [18]Sun Y, Chen Y, Wang X, et al. Deep learning face representation by joint identification–verification.NIPS. 2014.
- [19] FaceNet: A Unified Embedding for Face Recognition and Clustering. CVPR 2015.
- [20] A Discriminative Feature Learning Approach for Deep Face Recognition. ECCV 2016.
- [21] Rethinking the Inception Architecture for Computer Vision. CVPR 2016.
- [22] Alex Krizhevsky, IlyaSutskever, Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. 2014.
- [23] Murray, N., Marchesotti, L., Perronnin, F. “Ava: A large-scale database for aesthetic visual analysis”. CVPR 2012.

团队简介

美团点评算法团队是整个美团点评技术团队的“大脑”，涵盖搜索、推荐、广告、智能调度、自然语言处理、计算机视觉、机器人以及无人驾驶等多个技术领域。帮助美团点评数亿活跃用户改善了用户体验，也帮助餐饮、酒店、结婚、丽人、亲子等200多个品类的数百万商户提升了运营效率。目前，美团点评算法团队在人工智能领域进行积极的探索和研究，不断创新与实践，致力于应用最前沿的技术，给广告大消费者带来更好的生活服务体验。

AI技术在智能海报设计中的应用

作者: 晓星

背景

在视觉设计领域中，设计师们往往会因为一些简单需求付出相当多的时间，比如修改文案内容，设计简单的海报版式，针对不同机型、展位的多尺寸拓展等。这些工作需要耗费大量的时间、人力成本（5~6张/人日），但对设计师的进步成长起到的作用却非常有限。另一方面，精准营销是未来的大趋势，在大流量背景下，首页的海报资源展位需要展示“千人千面”的效果，这对海报的生产效率也提出了非常高的要求。所以，我们美团外卖技术团队尝试结合AI技术，来协助设计师避免这种低收益、高重复的任务，同时低成本、高效率、高质量地完成海报图片的生成。本文以Banner（横版海报）为例，介绍我们在海报设计与AI技术结合方面所进行的一些探索和研究。

分析

什么是Banner的设计过程？我们尝试总结了对Banner设计的理解，Banner的设计过程是一系列的具备某种特征属性的素材图层的有序叠加过程。这里的特征属性既包括颜色、形状、纹理、主题等视觉属性，也包括位置、大小、贴边等空间属性。在这个过程中，哪些环节可以被机器算法所探索呢？文献[1]研究了如何调整图像的颜色分布，使杂志封面的视觉效果更加符合人眼的视觉特性；文献[2]以此为基础，引入了基于显著性识别的图像裁剪，并使用优化方法来解决布局问题。阿里巴巴的“鹿班”系统在去年双十一当天，生成1.7亿张Banner，京东内部也在孵化玲珑和莎士比亚系统，更加智能地设计文案和Banner。

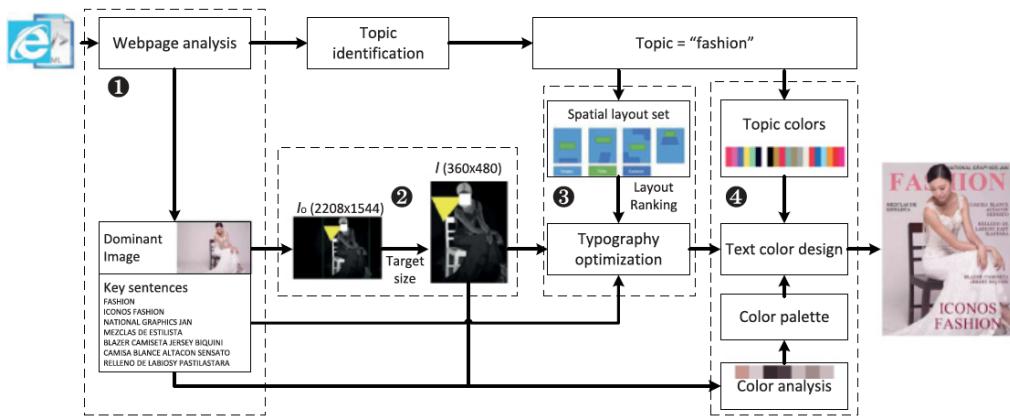


图1 封面配色&布局设计[2]

在设计领域的一些子问题上，可以用算法来挖掘出数据背后的规律（如图1所示）。那么，能否构建一个完整的学习算法和处理系统，统一解决Banner设计中所有的子问题（配色、布局、搭配、生成）呢？

技术方案

素材图层是Banner的基础元素，其本身可以被**特征化**，同时组成Banner的若干元素间的叠加顺序可以被**序列化**，因此，算法实际是在学习“**在什么时候，选择某种素材，放在哪里**”。

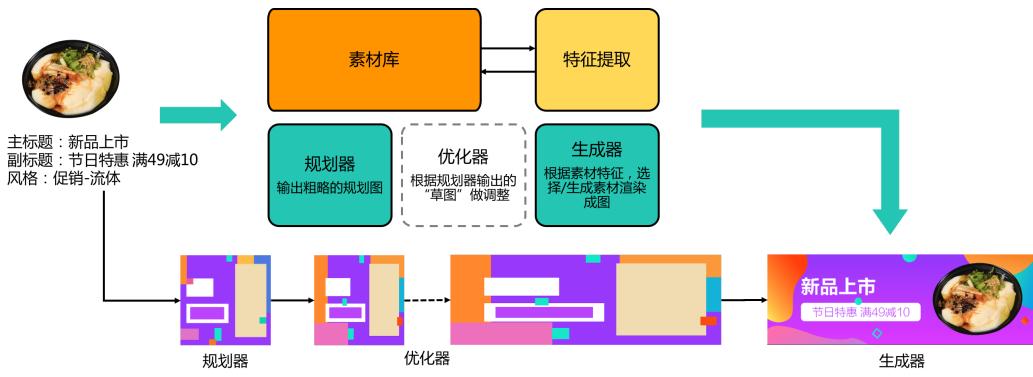


图2 流程框架

如图2所示，为了解决这个问题，我们设计规划器、优化器、生成器来共同构建海报设计的**学习与生产过程**。其中：

1. 规划器从数据中学习设计师对不同风格下的设计习惯与规律；
2. 优化器基于美学质量和设计原则，对前者的输出结果做精细化调整；
3. 最后，由生成器选取/生成素材并渲染成图；
4. 素材库作为这三个环节的基础，负责**素材管理和素材标签化**。

素材库

如何提取素材图片的特征属性，这是比较典型的分类问题。在计算机视觉领域，传统方案是提取图像的颜色、梯度等低级语义特征[3]，结合传统的分类器（LR、SVM等）来实现分类。近年来，基于深度学习的方法因为能表达更为复杂的语义特征，逐渐成为主流方法[4]。如图3所示我们提取传统的低级语义特征，以及基于CNN的高级语义特征，来共同完成素材**特征属性提取**。

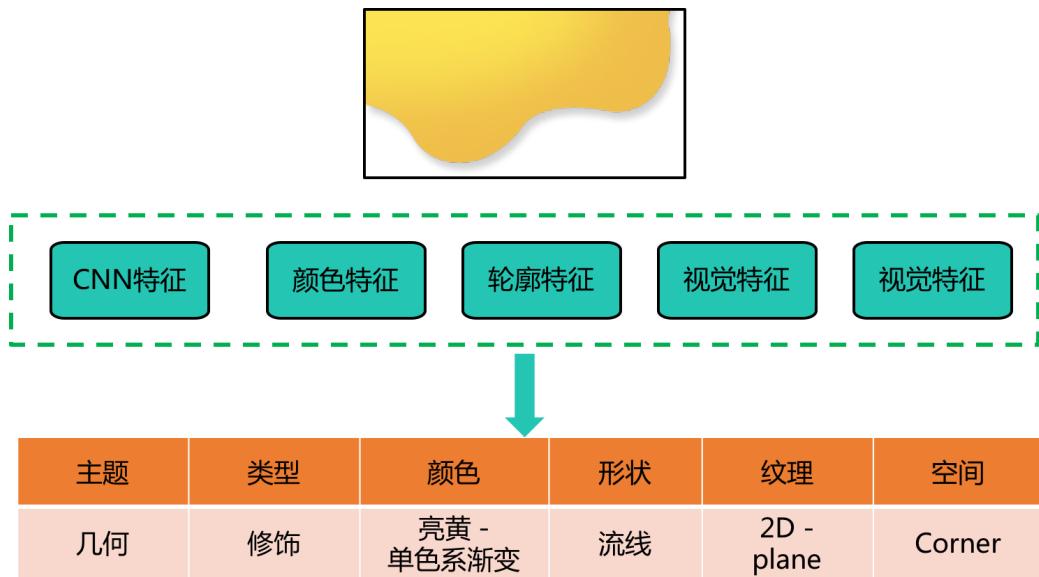


图3 素材库-特征提取

规划器

完成素材的数据化工作后，怎样学习Banner的设计过程？

作为一种生成模型，对抗生成网络（GAN）[5]在近年广为应用，其优势是可以端到端地训练图像生成模型，但在我们的应用场景下，GAN存在以下两个问题：

1. GAN的过程更像是“黑盒”的过程：输入方面，虽然Conditional-GAN之类的方法可以实现某种程度有条件地可控生成，但对于Banner设计任务来说，其输入信息（文案、目标风格、主体信息）仍然过于复杂；
2. 输出方面，GAN直接生成源数据（即图像），但非常缺乏解释性。我们需要的是更加直观、更有解释性的信息，比如素材的类型、颜色、轮廓、位置等。

在上文中有提到，Banner设计过程是**素材图层依次叠加**的过程。因此，我们可以用**序列生成模型**来拟合这个过程[6]。在建模过程中，我们把素材视作词汇（Word），海报视作句子（Sentence），词汇索引视作离散化的特征索引，素材叠加顺序就可以视为句子中的词顺序[7]。

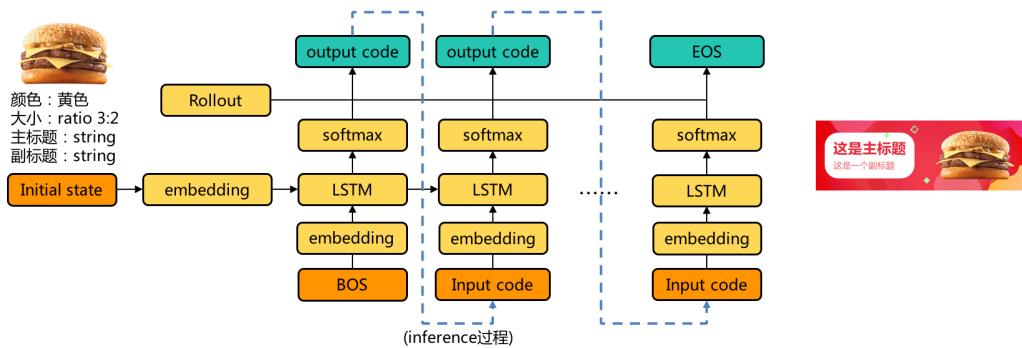


图4 规划器-序列生成

图4是我们使用的序列生成模型，输入主体信息和目标风格，输出素材特征的序列。为了增强预测过程中多条路径结果的多样性，我们在监督性地训练每个时刻的输出之外，还引入了评估整个序列合理性的Object loss。如图5所示，借鉴SeqGAN的思想，Object loss可以由判别器来提供[8]。

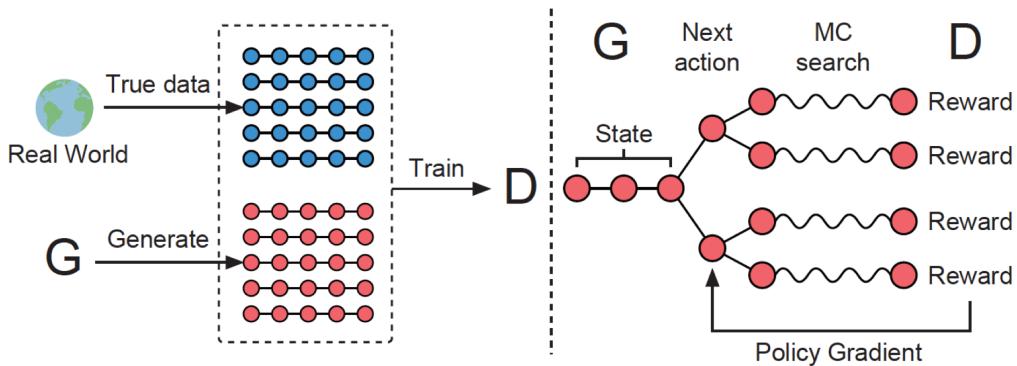


图5 SeqGAN[8]

优化器

规划器预测素材的量化特征，为了确保最终成图符合美学标准，需要一个后处理的过程（图6）。我们用优化器来解决这个问题。从本质上讲，这是一个**优化过程**。通过和设计师们的沟通，我们设计了一些基于

常规设计理念和美学标准的目标函数，动作集合包括移动、缩放、亮度调整等，结合优化方法，提升Banner的视觉效果。



图6 优化器

生成器

优化后的素材特征序列，通过生成器来渲染成图。如图7所示，对于素材库检索不到符合某种特征属性的素材的情况，我们设计了图像风格迁移的方法来实现图像特征迁移。这里的特征可以是颜色、形状、纹理等低级特征，也可以是某种语义化的视觉风格特征，对后者来说，可以将源图像的内容Content和目标图像的风格Style在某种特征空间（CNN中的某一层）里做融合，实现**风格迁移**[9,10]。



图7 素材生成

应用场景及功能拓展

“千人千面”的精准营销是未来营销策略的趋势，这对商品素材的丰富性提出了非常高的要求；从为商家赋能的角度来说，也需要为商家提供更多样的海报版式，这也要求系统具备海报风格的快速学习能力和拓展能力。对此，在常规设计风格的研究之外，我们从以下3个方面做了一些拓展研究。

主体图片加工

商品素材的丰富度与美学质量是精细化营销及海报美学质量非常重要的一环。其中最核心的要求是图像分割的能力[11,12]。以全卷积网络（FCN）为基础，如图8所示，我们采取以下几个在目标分割领域常见的技巧，来实现商品图片的目标分割：

1. Encoder–Decoder结构
2. 空洞卷积
3. 多尺度特征融合
4. Two–Stage微调网络

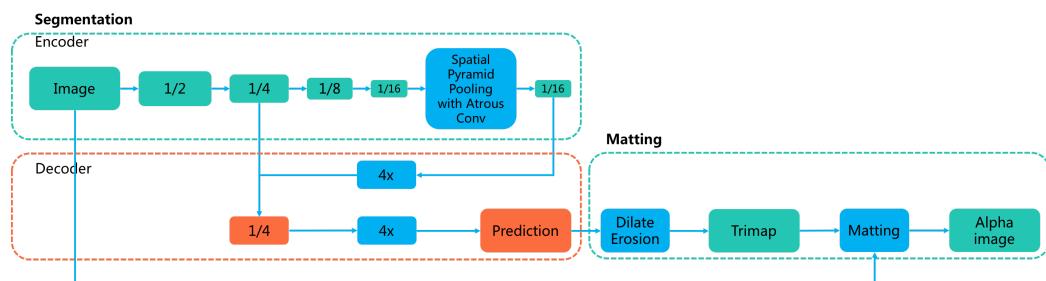


图8 图像语义分割&抠图（结构图部分参考DeepLab v3+[12]）

这种基于语义分割方法的结果，在专业设计师人工评审质量的过程中，发现主体边缘有时会出现明显的锯齿感。经过分析，我们认为有以下两个原因：

1. 语义分割模型把问题建模为一个“像素级分类过程”，每一个像素的类别都天然地被假设为“非此即彼”，大多数的 Segmentation模型都采用Cross-Entropy作为损失函数；
2. 因此，无论是从模型结构（CNN）还是从损失函数的角度来说，分割模型会更倾向于全局信息从而丢弃局部结构信息，导致只能得到边缘比较粗糙的分割结果。

为此，如图8所示，我们在图像分割的输出结果之外，结合了Image Matting方法：

1. 对分割模型的输出结果做形态学变换，生成三值Trimap，分别表示前景区、背景区、未知区；
2. 应用常规的Matting方法，比如Bayesian、Close-Form等，以原图像和Trimap图像为输入，输出前景图的Alpha通道；
3. Matting可以使前景主体的边缘更加平滑，视觉质量更高（图9）。



图9 商品主体抠图

另外，基于图像美学质量评分模型，我们会优先选择质量分高的图片作为主体素材来源。对中低分的图片，未来可以考虑借鉴Cycle-GAN[13]的思想，设计基于半监督和GAN的图像增强网络，美化后再经过图像分割产生主体素材。

海报模板拓展

上述的常规设计风格的学习属于一种广义的设计风格，需要设计师先期投入很多精力做风格划分以及数据收集、处理。为了快速适配热点场景，我们借鉴**图像检索**技术（如图10所示），提取素材图片的**CNN特征及颜色特征**，使用欧式距离来度量素材相似度，这能节省人工打标签的成本，实现基于固定模板的自动拓展及生成（海报临摹）。



图10 素材图像检索与模板拓展

多分辨率拓展

在日常工作中，设计师在设计出Banner后，往往要花费很长时间对不同展位、不同版本、不同机型做多尺寸适配（如图11所示）。能否用算法来协助人工提效？在素材已经确定，并且相互的位置关系近乎确定的条件下做多分辨率适配，这本质上是一个**优化问题**，我们在上文布局优化器的基础上，增加元素的局部相对位置与全局绝对位置的拓扑关系作为目标函数。目前，系统支持在某个固定宽高比 $\pm 30\%$ 范围内的**任意分辨率适配**，未来会进一步扩大适配范围。

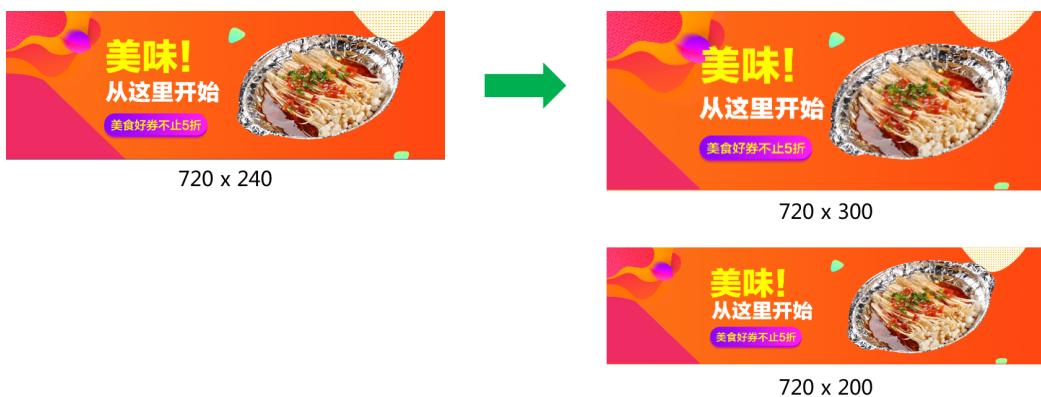


图11 多分辨率拓展

总结

目前，我们的Banner智能设计系统为钻展（外卖首页广告位）、商家店铺装修等业务提供稳定的设计能力；素材加工等子能力也在为外卖、闪购等商品图片提供技术支持。后续我们会从扩展常规设计风格、语义相关的颜色及素材挖掘、自动解析数据、构建自评估学习闭环等方面继续研究，进一步提高算法的设计能力和适用性，尽可能协助设计师提高效率，降低高重复性工作的时间和经济成本。

参考文献

- [1] A. Jahanian, J. Liu, D. Tretter, Q. Lin, E. O'Brien-Strain, S. Lee, N. Lyons, and J. P. Allebach. "Automatic Design of Colors for Magazine Covers". In Proc. IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, 2013
- [2] X. Y. Yang, T. Mei, Y. Q. Xu, Y. Rui, S. P. Li. "Automatic Generation of Visual–Textual Presentation Layout". ACM Transactions on Multimedia Computing, Communications, and Applications, 2017
- [3] David G. Lowe. "Distinctive Image Features from Scale–Invariant Keypoints". International journal of computer vision, 2004
- [4] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". NIPS, 2012
- [5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio. "Generative Adversarial Networks". NIPS, 2014
- [6] K. Kawakami. "Supervised Sequence Labelling with Recurrent Neural Networks". Studies in Computational Intelligence, 2008
- [7] T. Mikolov. "Statistical Language Models based on Neural Networks". 2012
- [8] L. Yu, W. Zhang, J. Wang, Y. Yu. "SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient". AAAI, 2017
- [9] L.A. Gatys, A.S. Ecker, M. Bethge. "Image Style Transfer Using Convolutional Neural Networks". CVPR, 2016

- [10] Y. Li, M.Y. Liu, X. Li, M.H. Yang, J. Kautz. “A Closed-form Solution to Photorealistic Image Stylization”. ECCV, 2018
- [11] J. Long, E. Shelhamer, T. Darrell. “Fully Convolutional Networks for Semantic Segmentation”. CVPR, 2015
- [12] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam. “Encoder–Decoder with Atrous Separable Convolution for Semantic Image Segmentation”. ECCV, 2018
- [13] J.Y. Zhu, T. Park, P. Isola, A. A. Efros. “Unpaired Image–to–Image Translation using Cycle–Consistent Adversarial Networks”. ICCV, 2017

作者简介

- 晓星，2017年6月加入美团，目前主要负责美团外卖图像内容挖掘、增强、生成方面的相关工作，致力于图像相关技术的积累及落地。



扫码关注技术团队
微信公众号

tech.meituan.com
美团技术博客