

美团餐饮娱乐知识图谱——美团大脑揭秘

(<https://tech.meituan.com/2018/11/22/meituan-brain-nlp-01.html>)

📅 2018年11月22日 仲远 ✎ 12907字 ⌚ 26分钟阅读

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、环境等方面的喜好，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](https://tech.meituan.com/meituan_AI_NLP.html)》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，综合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，一方面可以促进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、环境等方面的喜好，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，目前正在搜索、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

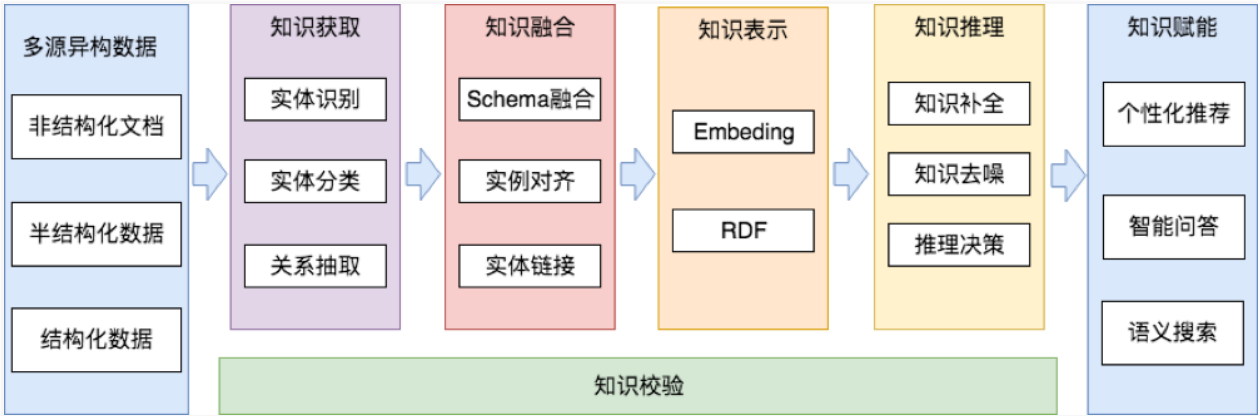


图4 知识图谱技术链

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。

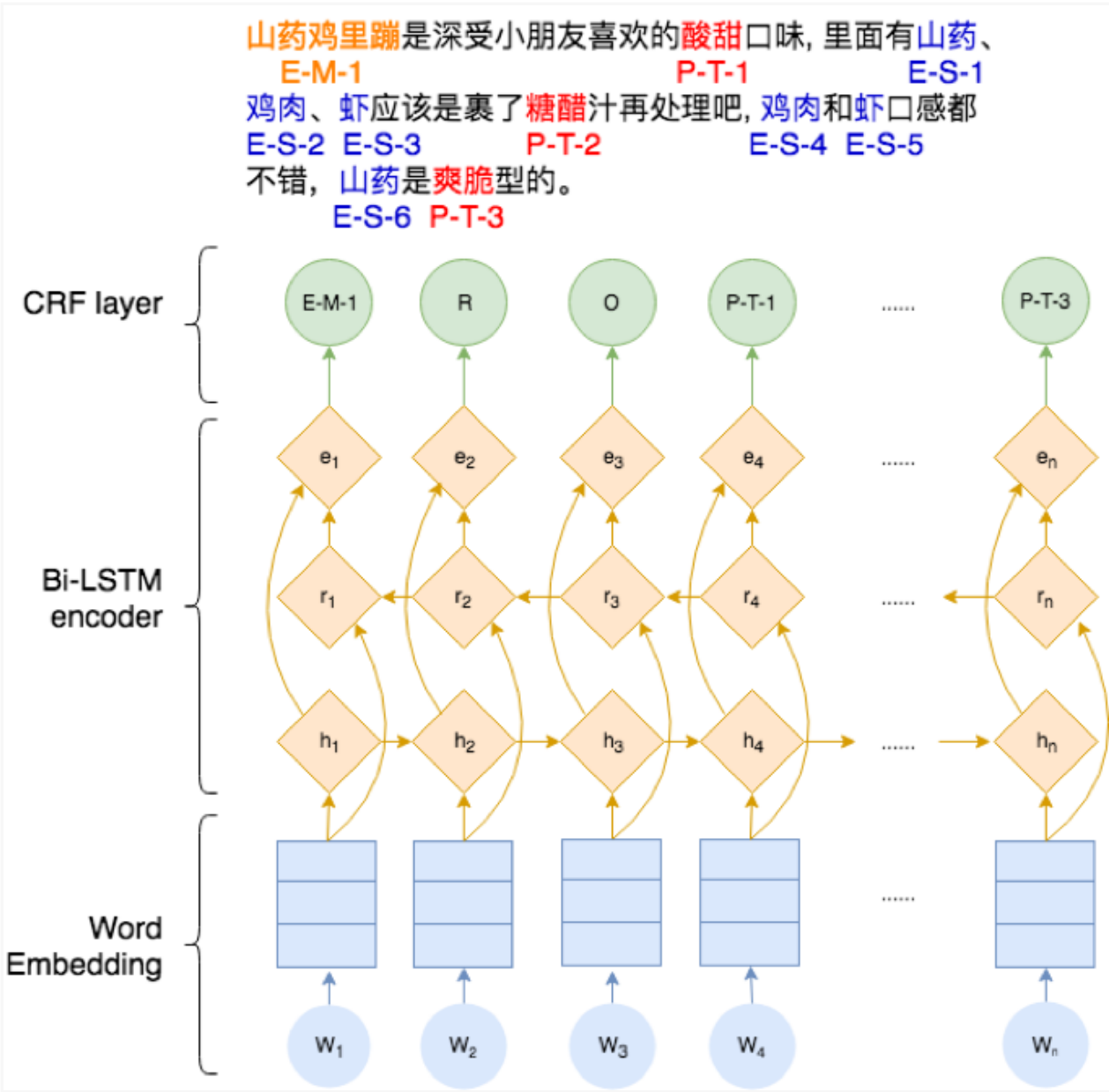


图5 BiLSTM+CRF模型

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”中抽取出来<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。

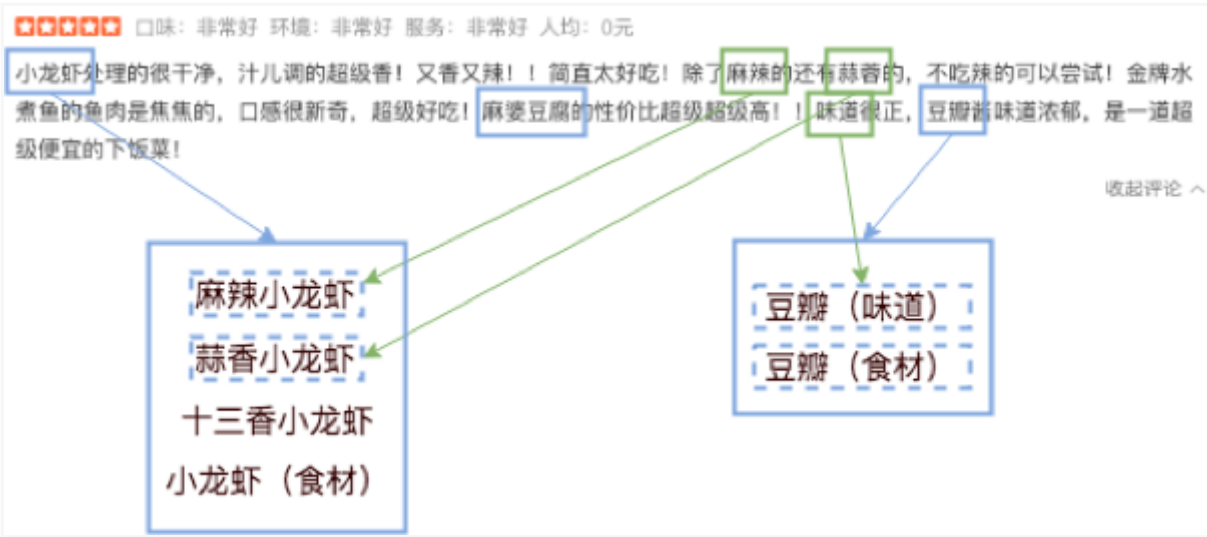
一旦图谱构建完成，如何从文本中准确匹配上图谱中相应的实体，进而延伸出相关的背景知识，则是一个实体链接问题。实体链接[7]主要依赖于实体Entity与所有Mention（文本文档中实体的目标文本）的一个多对多的映射关系表，如“小龙虾”这个Mention在图谱中实际对应的实体Entity可能是“麻辣小龙虾”的菜，也可能是“十三香小龙虾”的菜。对于从文本中识别出的Mention，利用上下文等信息，对其候选Entity进行排序，找出最可能的Entity。实体链接可以正确地定位用户所提实体，理解用户真实的表达意图，从而进一步挖掘用户行为，了解用户偏好。

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、食材等方面的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

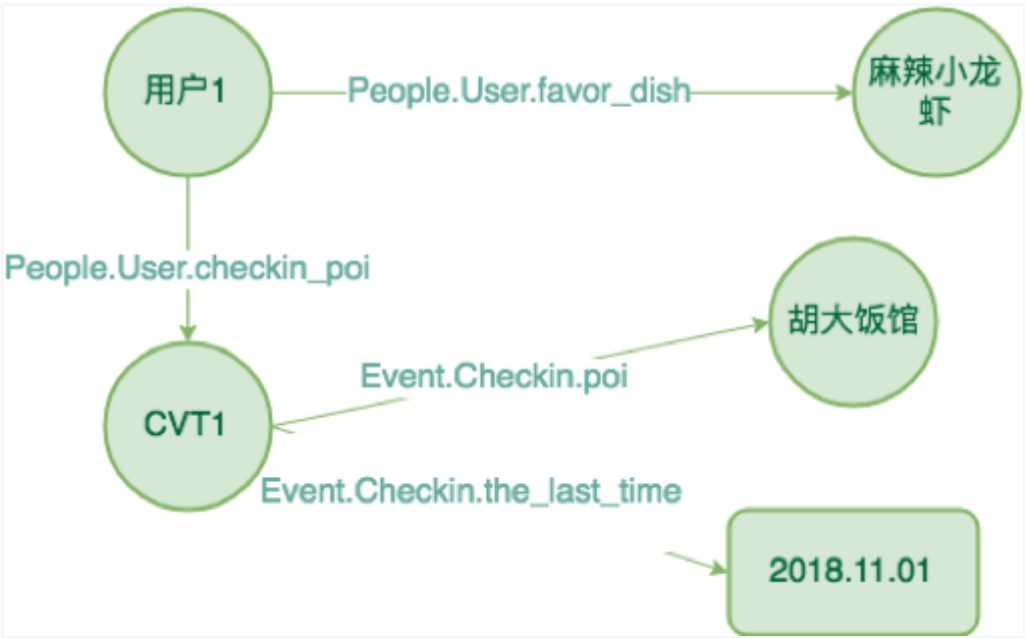
知识表示是对知识数据的一种



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], 将会在之后的系列文章里进一步讲解。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。

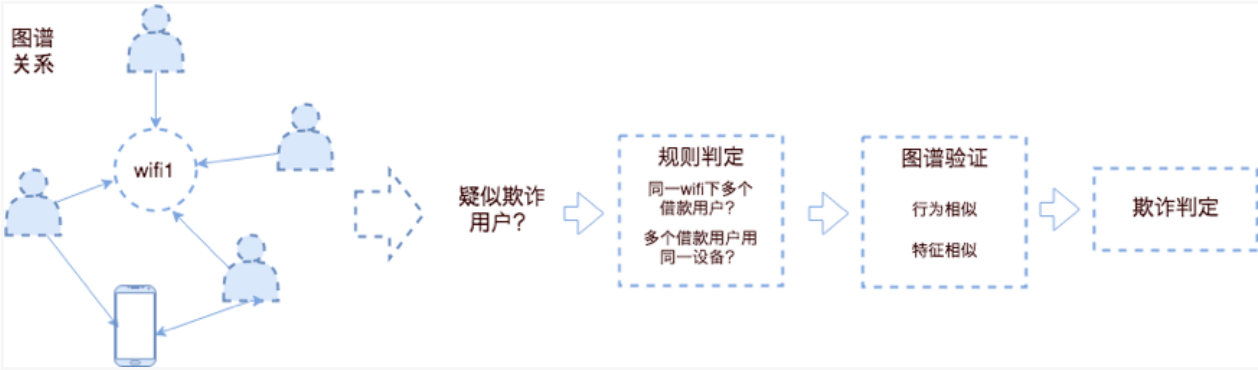


图8 知识推理在金融场景应用

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，而不只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。

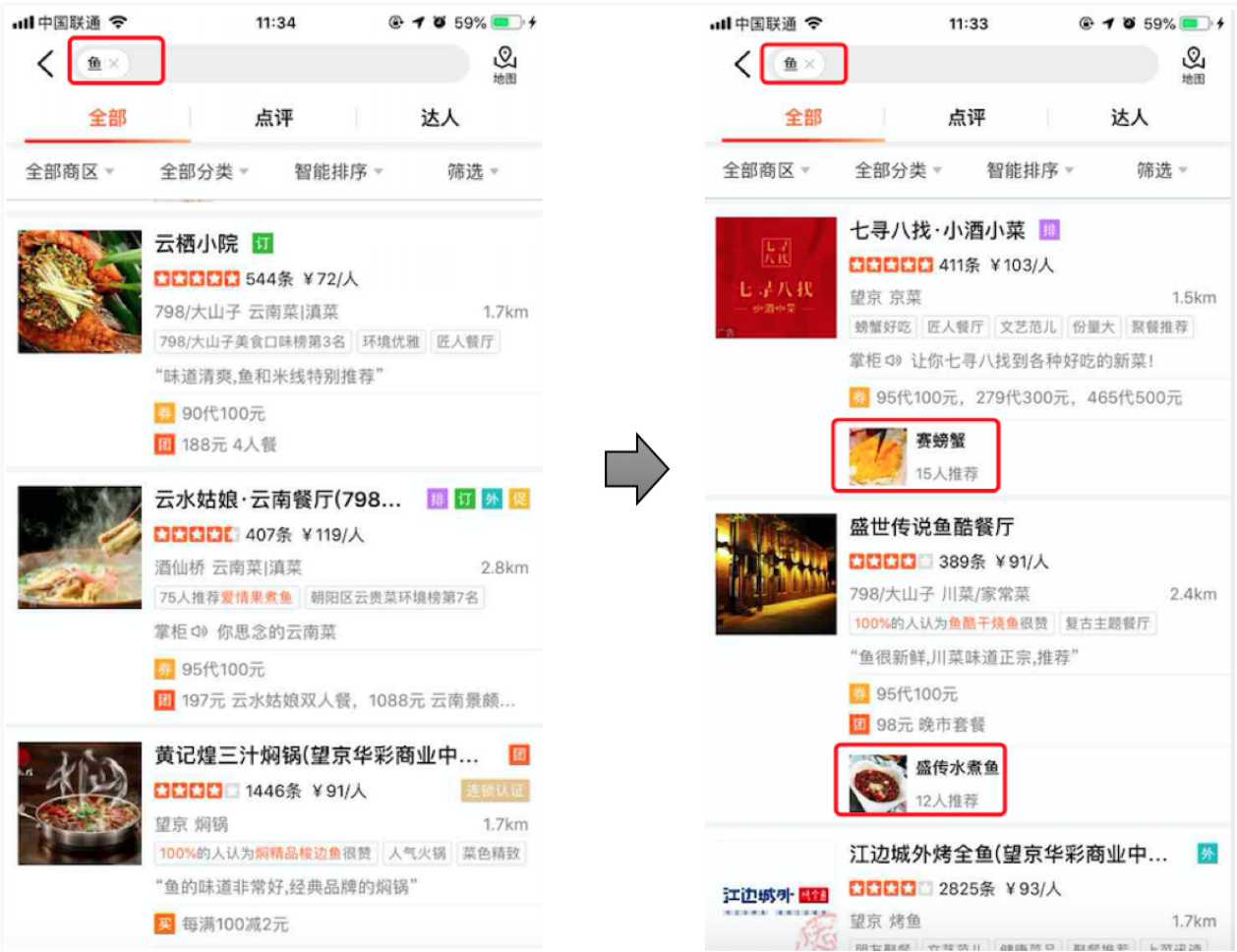
美团大脑的业务应用

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。

金融风险管理和反欺诈：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：知识图谱的应用算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 intelligence 165.1 (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。
- 王琚，博士，美团AI平台NLP中心产品和数据负责人。在此之前，王琚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 [🔗 \(https://zhaopin.meituan.com/job-detail?](https://zhaopin.meituan.com/job-detail?jobId=291801448272331465)

[jobId=291801448272331465\)](https://zhaopin.meituan.com/job-detail?jobId=291801448272331465) 、 知识图谱算法工程师/专家/研究员 [🔗](https://zhaopin.meituan.com/job-detail?jobId=291802012322333388)

[\(https://zhaopin.meituan.com/job-detail?jobId=291802012322333388\)](https://zhaopin.meituan.com/job-detail?jobId=291802012322333388)

工程岗： C++/Java研发专家/工程师 [🔗 \(https://zhaopin.meituan.com/job-detail?](https://zhaopin.meituan.com/job-detail?jobId=295725036608130037)

[jobId=295725036608130037\)](https://zhaopin.meituan.com/job-detail?jobId=295725036608130037) 、 AI平台研发工程师/专家 [🔗 \(https://zhaopin.meituan.com/job-detail?](https://zhaopin.meituan.com/job-detail?jobId=292557987555804498)

[jobId=292557987555804498\)](https://zhaopin.meituan.com/job-detail?jobId=292557987555804498)

产品岗： AI产品经理/专家 [🔗 \(https://zhaopin.meituan.com/job-detail?jobId=291803922350539431\)](https://zhaopin.meituan.com/job-detail?jobId=291803922350539431)

(NLP、数据方向)