

知识图谱方兴未艾，消费金融应用道阻且长



麻袋研究院 | 央网
05-05

消费金融 知识图谱

本文共3955字，预计阅读时间1分34秒

作为人工智能的重要分支，知识图谱不仅可以发现数据之间的关系，而且能利用关系进行推理，从而赋予数据更多价值。正因为有着这样强大的魔力，知识图谱被应用到了搜索、地图、个性化推荐等各个领域。

而消费金融，作为一个对数据具有强依赖性的领域，对知识图谱的应用有更切实的需求，已经发展出包括反欺诈、失联客户管理、精准营销、智能搜索和可视化、问答交互在内的多种应用方式。

麻袋研究院认为，虽然应用方式多种多样，但是由于数据量少且非标准化以及建模较难的原因，导致知识图谱在消费金融领域的应用还处于初级阶段，尚未形成大规模应用——这也使知识图谱技术，成为金融科技公司未来技术竞争的关键领域。

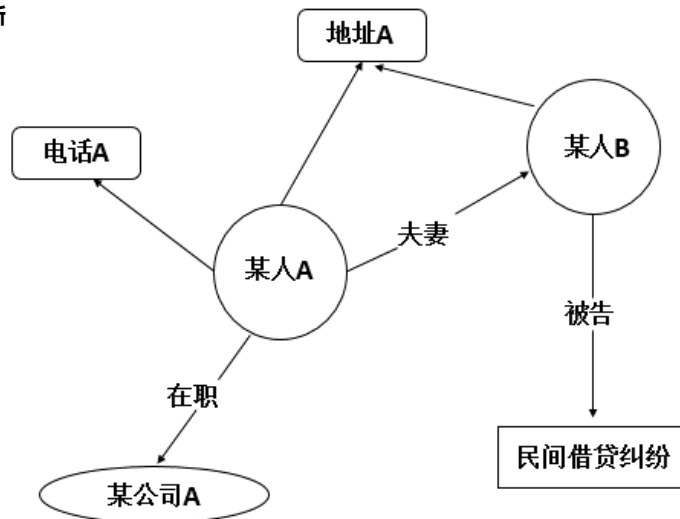
一、知识图谱概述

知识图谱本质上是描述客观世界中实体及其之间的关系的一种语义网络，由节点和边组成。在知识图谱中，每个节点表示现实世界中存在的“实体”，每条边为实体与实体之间的“关系”。

语义类型中最顶层抽象类是实体。其中，实体可分为概念实体和物理实体。概念实体简单理解就是没有具体实体的对象，如法律法规，医学知识等等。物理实体指的就是实实在在存在的实体，如杯子，汽车轮胎等。

三元组是知识图谱最通用的表现形式。所谓三元组，就是由一个表示主语的节点，一条表示谓语的边和一个表示宾语的节点组成一条记录。当围绕着一个主语有很多用三元组表示的关系呈现时，就构建了知识图谱。

图1 知识图谱示意图



资料来源：麻袋研究院

知识图谱与其他信息源不同之处是它可以由计算机直接访问，而无需任何的人工介入。由于计算机只理解正规的语言，所以知识图谱中包含的信息通常被称为结构化数据，而非结构化数据（如纯文本）只能够被人类理解。

从技术层面设计上来说，知识图谱必须考虑两个基本的权衡：一是知识图谱应该具有足够的描述性，以便复杂的知识能够被机器编码；二是这种描述性应该足够的简单，以便计算机能够快速的处理它们。

从领域上来说，知识图谱可分为通用知识图谱和特定领域知识图谱。通用知识图谱包含了大量的现实世界中的常识性知识，强调的是广度；特定领域知识图谱面向特定的垂直领域，强调的是深度。在实际中，知识图谱的应用趋势逐渐从通用领域走向特定领域。不同行业都在关注适合自身的知识图谱技术的应用方式，但目前仍未形成大规模应用。

二、知识图谱在消费金融领域的应用

当前，知识图谱在消费金融领域的应用仍处于起步阶段，主要应用包括反欺诈、失联客户管理、精准营销、智能搜索和可视化和问答交互。

1. 反欺诈

反欺诈是对包含交易欺诈、网络欺诈、电话欺诈等诈骗行为进行识别的一项服务，而知识图谱的反欺诈计算是对反欺诈分析实时性的完美体现。

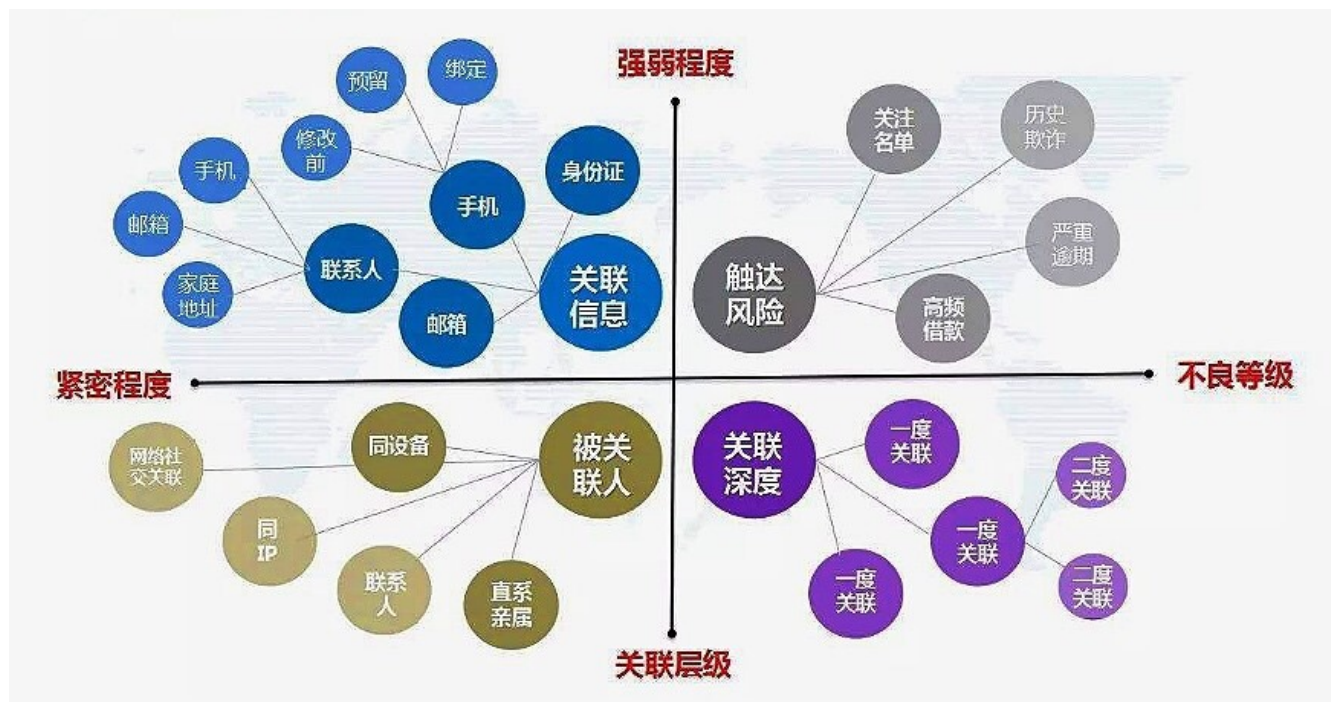
知识图谱在消费金融领域反欺诈方面的应用包括贷前审批和贷中预警。据麻袋研究院了解，目前知识图谱在消费金融领域的应用，主要集中在贷前审批管理。

在贷前的审批阶段，将借款人的基本信息、消费记录、行为记录、关系信息等整合到反欺诈知识图谱里，比如看客户一度、二度关系是否触黑，以及客户消费关联商家是否异常等。同时，利用不一致性检验，对借款人的风险进行分析和评估。比如当两个借款人填写了相同的电话号码时，这种不一致性很可能就有欺诈行为。

同时，知识图谱还可以有效甄别团体欺诈。甄别团体欺诈的方式有两种：第一，根据之前标记的黑名单客户，利用知识图谱确定与其有紧密联系的欺诈用户。根据调查，与黑名单客户有大量关联的借款用户的坏账率通常是未关联用户的2.9倍；第二，根据团体欺诈会共享部分信息的特点，利用知识图谱发现不同用户之间的信息关联性，从而识别欺诈团体。



图2 反欺诈知识图谱



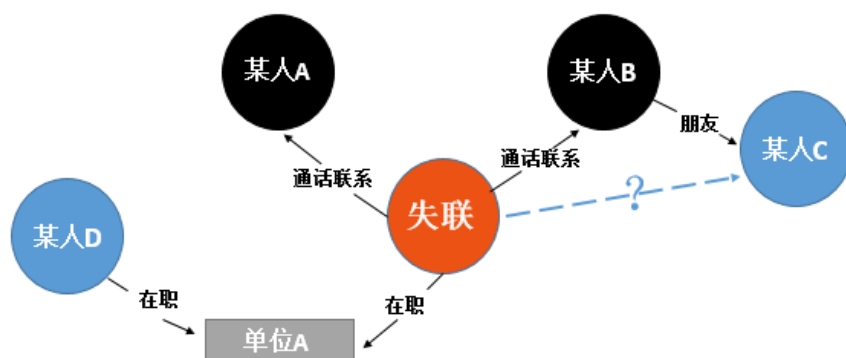
资料来源：拍拍信

以中腾信为例，其对于知识图谱的反欺诈应用，已相对成熟。贷前阶段，中腾信利用知识图谱技术对海量申请资料做快速匹配，大幅提升实时贷前反欺诈的效果。同时运行适当的实体链接分析查询，实时阻止高级的团体欺诈场景。贷中阶段，通过将实时数据纳入知识图谱模型，实现贷中的监控和额度管理。

2. 失联客户管理

在贷后管理中，知识图谱也发挥着重要作用。当借款人按时还款并且“失联”时，催收人员的工作难度加大。据专业人士告知，目前贷后管理对知识图谱的应用，主要是结合借款人授权的通话记录，挖掘出与借款人有关系的、且在相同平台借过款的新联系人，从而重新取得与借款人的联系，提高催收成功率。由于技术难度不是很大，已形成广泛应用。同时，各消费金融机构也在进一步扩充数据维度，比如工作单位等。

图3 失联客户管理知识图谱

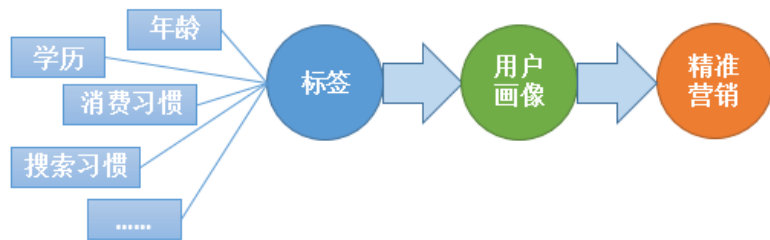


资料来源：麻袋研究院

知识图谱在精准营销中的应用有两种形式。

第一，理解用户。通过知识图谱聚合用户的基本属性，如年龄、学历、消费习惯、搜索习惯等，将这些基本属性加以分门别类，形成不同的用户标签，展现不同类别用户的身份特质及具体偏好。然后分析客户潜在需求，进行精准推送。

图4 精准营销知识图谱形式一

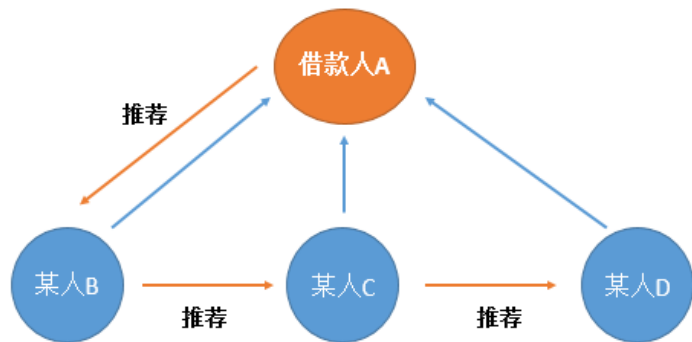


资料来源：麻袋研究院

第二，挖掘潜在客户。基于现有用户的社交网络知识图谱，根据交往方式和频次等社交行为建立关系模型，从而实现潜在用户的拓展。此外，还可以结合所有用户的标签特征，使用社区算法为全局用户进行用户细分，识别隐藏在数据深处的价值用户。

据麻袋研究院了解，目前第二种形式更为常见，主要是通过借款人之间的推荐关系，挖掘为平台带来大量资金和新客户的推荐人，分析这些推荐人的特征，从而发展新的具有类似特征的客户。

图5 精准营销知识图谱形式二



资料来源：麻袋研究院

4. 智能搜索和可视化

通过知识图谱，搜索功能可以在语义上扩展更多的搜索关键词，从而获取更全面的信息，进行风险识别和提示。比如搜索某个人的身份证号，可以返回与这个人有关的所有历史借款记录、联系人关系和其他相关的标签（如黑名单等）。然后通过图谱可视化技术，以图形网络的形式展示全方位信息，包括复杂信息和隐藏信息等。

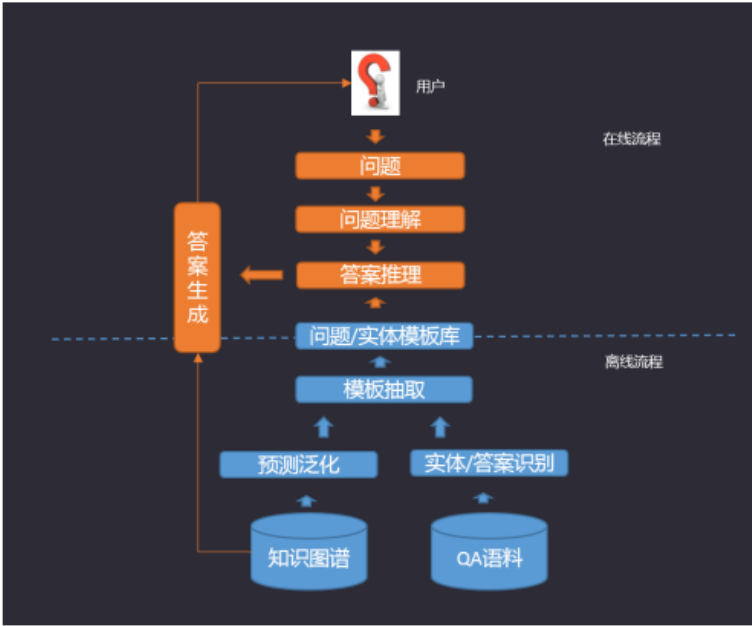
图6是宜信的智能搜索系统。该系统不仅利用公司内部积累的历史数据，还用爬虫覆盖了100多个公开网站，如人法、工商、百度、生活服务类以及十几家网贷黑名单等。通过在系统中搜索借款申请人的身份证号，就可以索引出与申请人相关的全部信息。

资料来源：宜信大数据创新中心

5. 问答交互

知识图谱在问答交互中最常用到的场景是文本客服，首先通过知识图谱创建知识库，在对用户问题进行语义理解和解析后，利用知识库查询、推理得出答案并反馈给用户。通过知识图谱，所有知识点以及连接知识点的边都被与问句关联起来，极大程度地提高了应答的关联性和准确性。

图7 知识图谱问答系统的功能流程图



资料来源：复旦大学崔万云博士《基于知识图谱的问答系统关键技术研究》

三、知识图谱应用的意义及难度

知识图谱对于消费金融领域的意义不言而喻，尤其是针对反欺诈环节。传统的反欺诈主要基于点，当出现征信数据缺失甚至空白的新客户时，识别其失信和欺诈风险的难度极大。此外，团体欺诈的盛行也给欺诈审核带来挑战。但通过知识图谱技术，可把各种信息整合成网状，比如挖掘客户与多种风险因子的关联关系，以及客户与欺诈客户、黑名单客户的关联紧密程度等等，从而更准确地判断风险高低。同时，知识图谱的应用也大幅提高了风险甄别的效率。

然而，目前将知识图谱应用于实务中的消费金融机构较少，大部分仍处于研发阶段。这主要是因为知识图谱在消费金融领域的应用存在较大的难度，主要集中在知识获取和建模方面。

1. 知识获取

在知识获取中，大多数数据都是非结构化的，比如文本信息。这些非结构化数据无法直接用于建模，所以如何把非结构化数据转化为结构化数据，再进一步整理成知识图谱所需的三元组结构是一个难点。在这个过程中，知识获取需要解决的一个核心问题是共指消解，也称为实体同义，一般要结合自然语言处理的“消歧分析”技术一起处理。比如在同一家公司工作的员工，在填写单位名称时可能对其有不同的表述方式——“阿里巴巴网络技术有限公司”，“阿里巴巴集团”和“阿里巴巴”。这种情况下，首先需要利用自然语言处理技术，把这些不同的名称指代到同一实体上。然后利用知识图谱，描述员工之间的关联关系。

2. 建模

首先是由于样本数量较少，构建有效模型的难度较大。在消费金融实务中，对于知识图谱的应用主要是垂直领域，且限定场景。但受领域和场景限制，样本数量也会受限制。此外，从不同来源大数据中抽取的知识可能存在大量的噪声和冗余，或者使用了不同的语言，从而无法建立相应的关系样本。但是随着时间的推移，样本的数量也会不断累积，逐渐形成了一些特定的关系链接和子领域。在这种情况下，构建迭代系统就显得尤为重要——通过将新的知识实时反馈给模型，从而使得模型不断地自优化。

其次，**将同一套模型应用到不同消费金融机构的难度较大**。这是因为建模过程需要结合具体的机构数据去做数据清洗。数据不同，则模型也会存在差异，所以很难建立起一套标准化模型。正是由于这个原因，目前国内为消费金融业务提供知识图谱技术支持的第三方平台数量并不多。据麻袋研究院了解，做知识图谱技术输出的方式主要是验证重要变量特征，将重要变量模型输出，从而为消费金融机构节省了挖掘部分数据特征的时间和成本。

本文系 [未央网](#) 专栏作者 [麻袋研究院](#) 发表，属作者个人观点，不代表网站观点，未经许可严禁转载，违者必究！

您使用了08分00秒阅读本文