

基于图计算技术的车险团伙 欺诈识别研究

董静 韦云健 李丹丹 杜新凯

〔摘要〕 随着车险业务的快速发展，车险理赔中的欺诈风险问题呈明显上升趋势。财险公司因车险欺诈支出的高额费用使车险综合赔付率居高不下。本文结合阳光财产保险股份有限公司（以下简称阳光财险）车险理赔数据，基于图计算技术，将智能风险预警、网络模式匹配、人机互动检索功能融合起来落地应用，以解决车险团伙欺诈识别过程中面临的问题，防范内外部欺诈风险，挽回赔款损失。

〔关键词〕 图计算技术；车险；团伙欺诈识别

〔中图分类号〕 F840.65 〔文献标识码〕 A

一、引言

随着保险公司业务的发展，各种潜在的欺诈风险也随之增加。保险欺诈一直是保险行业面临的巨大挑战之一，高欺诈风险、高赔付率直接导致企业经营情况承压。车险作为财险公司的支柱险种，占比巨大，是保费收入的重要来

〔作者简介〕 董静、韦云健、李丹丹、杜新凯，均供职于阳光保险集团科技中心大数据和人工智能部。

源，但车险欺诈问题已成为车险发展的瓶颈，不但妨碍了车险市场的正常发展，还扰乱了社会公共秩序，加大了保险公司经营和管理风险的难度。

近年来，我国车险欺诈案件犯罪手段隐蔽，手法多样，并呈现团伙化、专业化作案趋势，共同犯罪比例明显上升。因涉及成本问题，多会出现共性关联特征的情况。跨险种、跨区域保险欺诈屡见不鲜，再加之机动车的高流动性和车险赔付标的的多样性，理赔风险管控困难，使车险成为保险欺诈的重灾区。凭借传统的欺诈识别方法已经难以高效地防范车险团伙欺诈，如何突破传统依据“人工经验”的低效率反欺诈模式，建立精准高效的保险反欺诈系统，成为保险行业亟须解决的问题之一。

本文针对车险理赔环节的团伙欺诈识别面临的痛点，通过应用人工智能、知识图谱等前沿技术，基于全方面的业务数据，搭建智能反欺诈图谱平台，从数字化、前置化、智能化三个方面升级风险管控模式，提升保险企业风险管控的效率和效果。

二、车险欺诈识别面临的问题

随着我国机动车保有量快速上升，车险保费由2000年的400余亿元增长至2019年的8180亿元，20年间增长了19倍^①，占财产险保费收入的70%以上^②。2019年初，国家发展和改革委员会出台政策促进新能源汽车、农村商用车和二手车的消费，必将推动汽车销量增长，从而带动车险发展。伴随车险业务的快速发展，车险理赔问题越发突出，其中，车险理赔中的欺诈风险问题呈现明显上升趋势。财险公司因车险欺诈而支出的高额费用使车险综合赔付率居高不下。本文结合阳光财险实际车险理赔环节的欺诈风险识别及其防范的业务流程，将车险欺诈风险识别和管控面临的问题分解为以下几个方面。

① <http://finance.sina.com.cn/wm/2020-12-16/doc-iiznezxs7241963.shtml>.

② <https://news.sina.com.cn/o/2015-12-16/doc-ifxmszek7134591.shtml>.

（一）欺诈频发，呈现三大特征

据国际保险监督官协会测算，全球每年有20%~30%的保险赔款涉嫌欺诈。欺诈手段呈现多样化、专业化、团体化等特征，共同犯罪比例明显上升。

（二）依赖人工欺诈判断难、成本高、效率低

欺诈判断庞大的工作负荷和巨大的工作压力，使欺诈判断无法逐笔勘察和审计。随着保险业务的不断扩展以及科技的进步，欺诈渗透手段不断翻新，仅依靠工作人员经验无法适应市场高速发展的要求。

（三）信息割裂，未形成风控全流程闭环管理

首先，数据共享尚未实现，“信息孤岛”仍然存在。其次，欺诈风险信息传递不及时，信息披露不全面。最后，案件核查等阶段，未实现异常案件及“黑灰名单”沉淀，缺少反哺模型优化及准确度提升的标注信息，未形成风控闭环管理，优化速度慢。

面对这样的背景，如何精准识别打击车险欺诈团伙，防范内外部欺诈风险，为公司挽回赔款损失，对车险欺诈进行防范，是财险公司亟待解决的问题之一。

三、车险欺诈识别技术方案对比

从2015年中国保监会发布《车险反欺诈数据元》和《车险反欺诈代码集》，至2017年中国保监会起草发布《反保险欺诈应用指引第1号：车险反欺诈指引（意见稿）》，到2018年《反保险欺诈指引》正式实施和《保险标准化工作管理办法》印发，再到2019年4月开展保险公司欺诈风险管理能力自评估工作，可以看出政策层面明确要求保险公司要具备理赔风控能力，使用科技筑起反欺诈的防火墙。

风控是与具体业务强相关的技术领域，要对风险进行防控，首先要知道有哪些工具/技术可用于防控风险。

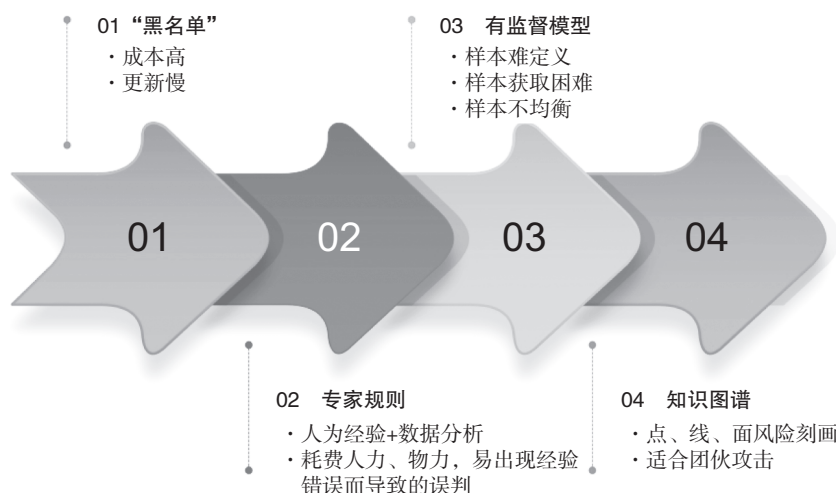


图1 欺诈识别技术区别对比

人力介入程度从高到低的欺诈识别技术依次包括“黑名单”机制、专家规则、有监督模型以及关联图谱四个层次，其中“黑名单”机制和专家规则都需要较高的人力成本，而且从行业现状来看，很难构建一份完整的业界“黑名单”，从而导致“黑名单”机制的可实施性较差，而专家规则过度依赖专家经验，存在较高的经验误判风险，主观犯错概率较大，因此，通过“黑名单”或专家规则的方式，很难构建切实可行的风控体系，而有监督模型方法通过积累的数据进行质检、稽核等多个方面的模型建设，可以在成本、理赔等环节提高欺诈识别的查准率和查全率，从而提高对未知风险预测的准确率，但难点在于样本的定义与构造，因此知识图谱这种自动化识别风险的方法吸引了越来越多的关注。

截至目前，反欺诈经历了传统反欺诈（“黑名单”、专家规则）、基于大数据反欺诈（有监督模型）以及现在的基于知识图谱（关系网络）反欺诈三次进化。其中，基于知识图谱反欺诈依赖的底层技术就是图数据库技术。相比深度学习的“黑盒子”，知识图谱的图数据结构具有更强的可解释性，应用知识图谱在风险识别能力和大规模运算方面具有突出优势，逐渐成为金融领域风控反欺诈的主要手段，借力知识图谱技术实现车险团伙欺诈识别与防范，是突破

车险反欺诈瓶颈的重要渠道之一。

阳光财险选择基于图数据库的知识图谱技术解决车险团伙反欺诈这一问题，使用图数据挖掘、图规则、图推理实现对团伙作案的识别，形成点、线、面风险的刻画，更符合车险团伙欺诈的作案特点，适合跨地区、线索分散的“专业骗保团伙”的串案识别工作，能够有效减少信息不对称，为车险反欺诈提供新途径。

四、车险团伙欺诈识别的解决方案与实践

本文使用知识图谱为整体框架，基于图计算技术，解决车险理赔开展过程中，在质检、监察、稽核等环节中面临的问题与挑战，结合阳光财险的车险理赔数据，将智能风险预警、网络模式匹配、人机互动检索功能融合落地应用，完成了智能反欺诈图谱平台的建设与实践。

该项目是全面基于知识图谱实现的反欺诈平台，项目中涉及的图论相关技术入选阳光财险2021年第二季度创新成果奖，以及2021年中小金融机构数智化转型优秀案例评选的“专家推荐TOP10优秀案例”奖，体现了阳光财险的智能化水平，增强了阳光财险的技术影响力。

智能反欺诈图谱平台的框架分为三个层级，分别是数据层、系统层、产品层。其整体框架见图2。

（一）数据层

数据层是为了解决“信息孤岛”问题，在集团层面融合各子公司的数据，实现统一管理分析。数据层收集了各个环节的业务和属性数据，包括承保数据、报案数据、理赔数据等，涵盖车辆信息、人员信息、赔案信息等，为了确保高质量的数据入图，在数据层需要完成数据清洗、过滤、整合，实体消歧等方面工作，为数据入图做充分的数据准备，进而抽取实体、关系、属性，完成车险赔案数据层关系图谱的搭建。

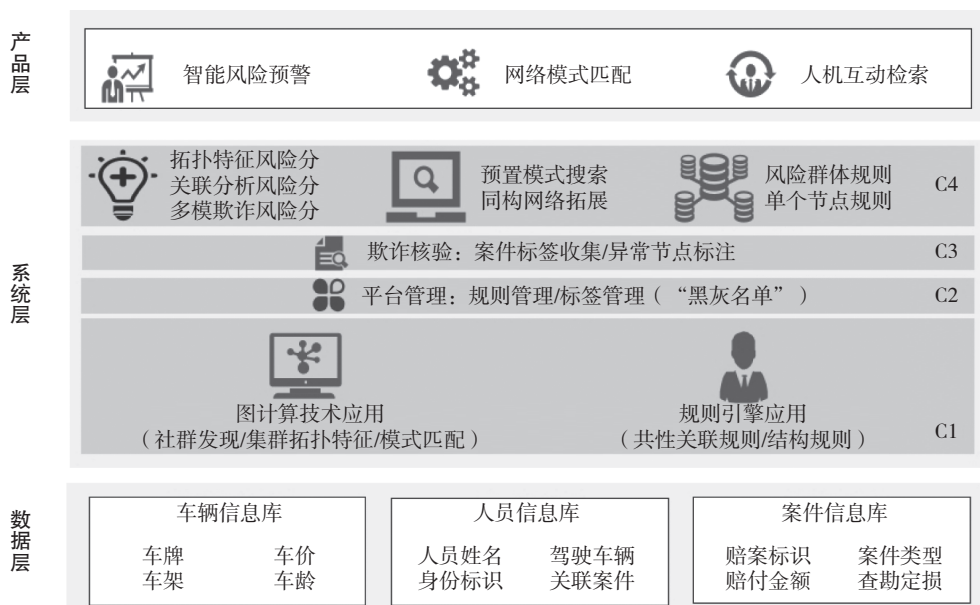


图2 智能反欺诈图谱平台框架

（资料来源：内部解决方案）

（二）系统层

系统层共分为四个细分层级，分别为图计算技术应用、平台管理、欺诈核验及模型应用。

1. 图计算技术应用

图计算技术包含与图结构相关的各类算法。目前，智能反欺诈图谱平台主要涉及四类算法，分别为计算图拓扑结构的相关算法、遍历图节点和路径的路径搜索算法、对图进行向量化表示的图嵌入算法及基于图的深度学习算法。在不同的应用场景中，可以根据不同的需求选择应用相应的算法，达成业务目标。

一方面，通过运用传统图算法、社区发现算法、图的异常检测算法等构建风险模型；另一方面，利用图卷积神经网络算法，如GraphSAGE等，挖掘图谱拓扑结构及节点自身特征和邻居特征隐含风险信息，引入图嵌入进行图表示学

习,进一步挖掘节点之间的关系。

规则引擎应用部分,通过配置基于共性关联特征和结构规则抽取车险赔案图谱中的属性特征,进而支撑产品层的应用输出。

2. 平台管理

平台管理是为平台的模型、标签、规则,包括“黑灰名单”库,提供统一的管理、维护入口,便于管理和统筹。

3. 欺诈核验

欺诈核验是在案件质检、监察、稽核等阶段,收集由业务人员标注的,或者由模型筛选出的异常案件,不仅可以填补和丰富阳光财险内部的“黑灰名单”库,同时基于图结构特征,结合更高维的有监督算法落地实现反哺模型的优化,从而提升车险团伙欺诈识别的准确度。将过往的风控策略的成果和方法进行及时的沉淀和利用,不断针对新保险欺诈风险的变化进行迭代升级,形成风控全流程的闭环管理。

4. 模型应用

模型应用是基于图计算技术的具体应用形态,针对不同业务需求及底层数据标签的积累情况,通过构建多维度模型体系,输出多维立体化的风险度评分,筛选高风险的欺诈网络,避免单一模型的“误杀”和“漏杀”;基于图搜索算法及图嵌入等图计算技术完成预置模式搜索及同构网络拓展功能,以及结合业务专家经验及图结构设置图规则,图的拓扑结构规则对新的攻击的抗性会比较高,能够泛化地识别触碰风险的特征,具有很强的鲁棒性,进而帮助质检、监察、稽核业务人员全面加强风控能力。

(三) 产品层

1. 智能风险预警

智能风险预警通过运用“算法+规则”的策略,构建了多层级、多级别的立体智能风控体系,提高了车险反欺诈风险规则的灵活性,以及提高了欺诈识别精度和效率,进而降低对专家经验的依赖程度。通过多种智能算法框架,进行联防联控,筛选高风险的欺诈网络,将其推送至业务人员的工作台供其案调使用。

一方面,智能风险预警功能使用了图结构数据及相应的图算法,可以很好地体现团伙作案的聚集性和关键人员的中心性;另一方面,基于作案成本的原因,欺诈团伙必然重复利用手中资源(如人、车、手机等)。因此,即使欺诈团伙不断在单次案件中升级作案手段,在多次作案之后他们仍会在关系网络中留下痕迹。加入图特征可使欺诈团伙画像特征更丰富,可以用来发现专家经验规则之外的新线索,提升对未知类型团伙欺诈的检测及车险反欺诈的灵活性。

2. 网络模式匹配

网络模式匹配,根据不同业务的类型特点,使用内部预置典型的高风险欺诈子图模式,对全量网络进行搜索,提取高风险网络(如闭环模式、网状模式、星型模式等)供业务人员进行调查。当然,欺诈分子的作案手段绝非一成不变,他们会根据策略拦截结果对作案方式不断进行调整,形成风险转移。为了形成攻防对抗,我们设置了同构网络扩展功能,可对指定模式的网络进行搜索。

同构网络拓展,在实际案件调查过程中,如果业务人员认为某一群体网络大概率是欺诈网络,并且平台未提出风险预警,可以选择对应的关键节点进行异常节点标注,后台模型根据关键节点特征对全量网络进行搜索匹配,并推送匹配到的相似网络。对全量的同类型的疑似欺诈网络进一步地调查和分析,实现新发高风险作案模式的提炼和识别,并将该模式添加到预置模式中,从而实现时刻更新攻防机制,对抗实时迭代更新,为企业提供全方位风险检测保障。

3. 人机互动检索

人机互动检索提供8个查询维度供业务人员在关系网络中查询,如人员姓名、身份证号、车辆车牌号、车架号等,业务人员可根据需要,从不同维度对案件网络进行检索。其中,典型使用场景是,业务人员基于智能风险预警推送的风险团伙信息,继续探查可疑骨干人员、查勘员、定损员、修理厂等相关案件及团伙;另一个场景为,业务人员从机构内部自查或者其他渠道获取可疑人员和车辆的信息,对关联的案件进行扩展和延伸。此功能更灵活,针对不同规模机构,不同地域侧重的风险点不同,业务人员可针对不同机构,自定义风险规则和阈值进行灵活组合,筛选高风险网络进行案调,实现全机构风险差异化控制。

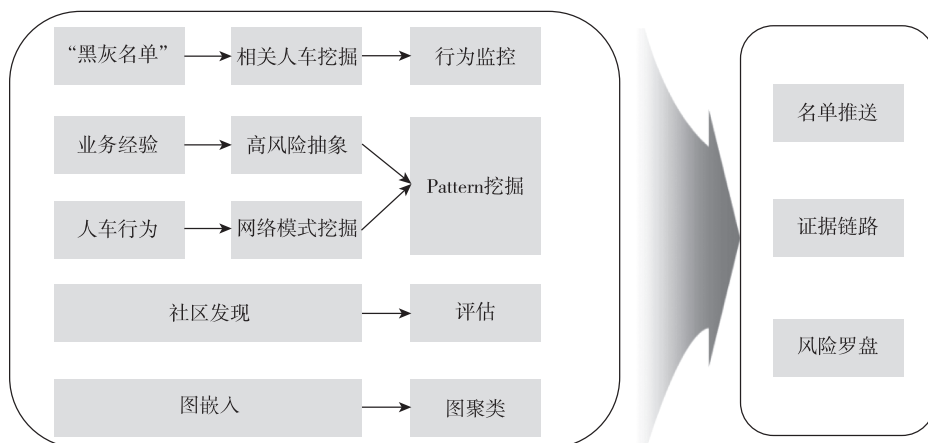


图3 智能反欺诈图谱平台功能汇总
(资料来源：内部解决方案)

综合以上三个功能，智能反欺诈图谱平台可为前端质检、监察、稽核业务人员提供高风险的欺诈名单推送、证据链路以及整个欺诈团伙的风险罗盘信息，对车险团伙的欺诈行为进行多角度、全方位的检测和识别，从而提升车险反欺诈管理的完整性和严密性。

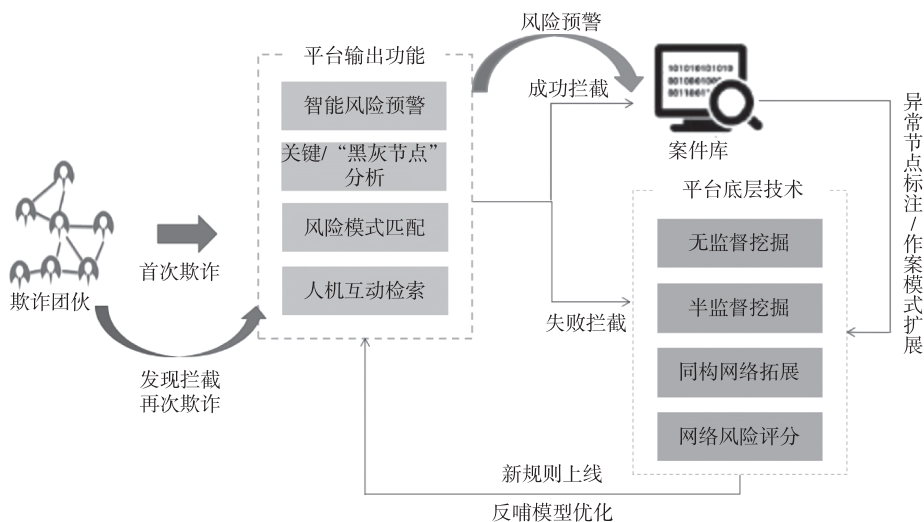


图4 智能反欺诈图谱平台流程
(资料来源：内部解决方案)

从智能反欺诈图谱平台流程来看,对于欺诈团伙多次作案之后再次作案的情况,通过平台输出的三大类风险发现机制发送风险预警,如果拦截成功,积累“黑灰名单”案件库的同时,进行策略分析、异常节点标注,从而将作案模式进行扩展实现同构网络拓展,并通过有监督学习计算网络风险评分,挖掘欺诈共性特征,应用于智能风险预警;如果拦截失败,通过无监督和半监督挖掘隐藏风险,通过分析产生新的策略并实施拦截,帮助业务人员增加新的专家规则,多种方式组合补充反欺诈策略,准确定位“黑产”团队和高危群体,降低重复对抗的操作成本。

智能反欺诈图谱平台是基于阳光财险内部数据搭建的,相较于第三方保险科技公司整体解决方案,更符合自身业务流程的特点和属性,有利于提高阳光财险在保险反欺诈中的工作效率,降低保险反欺诈付出的成本。

智能反欺诈图谱平台构建的风控体系是从规则决策到有监督再到无监督的发展关系。对于已知欺诈类型,通过业务规则就可以判断;对于未知欺诈模式,则需要使用无监督的异常检测技术,结合业务场景发现线索,挖掘团伙网络所有隐藏信息。从“黑产”动态感知、规律发现、模型迭代、业务拦截、案例学习,形成风控全流程闭环管理,提供高效准确的综合判定结果,为公司降赔增效发挥重要作用。

五、展望

随着保险科技的发展,相关车险反欺诈的数据应用更加需要数据的及时、准确、丰富和全面,因此,数据的有效整合、标准化接入收集成为行业发展的重要要求。我国保险公司已经意识到数据中的价值,但是把需要的数据转换成可用的形式,并产生效益,在实际车险反欺诈应用中仍然面临一定的障碍与挑战。

1. 信息采集的重视程度低

保险行业尽管经过整顿和多次培训,但是从业人员仍然缺乏对反欺诈应用所需数据和出险之后总结数据的汇总上传。对于大数据车险反欺诈应用分析所需的修理厂数据、三者驾驶员信息、伤情鉴定数据的链接和收集仍显不足。

2. 已采集的数据质量不佳

车险反欺诈系统的有效应用依赖于各保险公司上传的数据质量。为了提高车险反欺诈信息数据的规范性、准确性和完备性，保监会制定了车险反欺诈数据规范，但效果并不理想，目前仍存在为了快速完成调度任务，在录入信息时没有确保信息的准确性，导致上传的字段数据存在缺失和不准确的现象。

3. 欺诈风险相关损失数据的积累不足

现阶段业内各保险公司欺诈事件的相关数据积累少，理论上对欺诈风险的认识不到位，造成缺乏针对欺诈事件全面的数据记录，欺诈风险大多以定性记录为主，缺乏对事故造成的实际与潜在损失额的分析与记录。

车险相关数据的不断规范在一定程度上有利于保险科技反欺诈的发展，以技术手段减少原先数据收集人员的负担，提高数据收集效率，加快推进各个地区的数据收集标准化和保险公司、修理厂、汽车4S店的数据收集标准化，使车险数据信息收集快速有效，将交通、银行征信等外部数据与车险行业的数据连通，拓展数据连通的维度，实现共享数据的更广泛应用。

在“智能+”科技浪潮的推送下，中国保险行业正面临前所未有的局面，一方面，中国保险深度和密度与全球平均水平相差近50%，保险行业面临市场持续增长、产品结构逐步优化、欺诈风险高等经营挑战。另一方面，在保险科技、监管机构等推动下，保险科技成为保险公司发展破局的关键，从产品、营销、承保、理赔、运营等所有环节重塑保险价值体系，彻底改变保险行业的传统风险管控模式。

未来，我们将持续结合阳光财险的实际业务场景，积极探索风控数字化、前置化、智能化的可落地模式，打磨数据产品，服务前端业务，不断深化科技创新能力，持续为数字化转型赋能。

六、附录 主流图数据库技术选型分析

在5G、物联网、人工智能等技术的推动下，企业数据呈现爆发式增长，数据间的关联复杂度也随之剧增。在处理复杂关联数据时，传统关系型数据库运算效率较低，难以进一步挖掘海量关系数据背后的价值。为了更好地利用数据

间的连接，发现隐藏的关系，企业需要选择合适的数据库技术将实体间的关系进行存储，并实现数据模型的灵活拓展。

图数据库是专门针对复杂关联深链查询优化，成为近年来崛起速度最快的数据库技术。由表示实体的节点和表示节点之间关系或连接的边组成。每个节点都有一个唯一的标识符、传出和/或传入边，以及属性或键值对，将不同种类的信息连接在一起得到一个关系网络，提供了从“关系”的角度去分析问题的能力。图存储是所有图数据库最重要的功能之一。

随着市场对图技术的认知程度和需求的上升，图数据库技术不断升温，图数据库是数据库领域近期发展最快的领域之一。据Gartner预言，图数据库将很快约占数据库领域10%的市场。DB-engines近年数据库流行趋势也显示，图数据库相比其他主流数据库受欢迎程度遥遥领先，可以明显看出图数据库越来越被大众所关注。

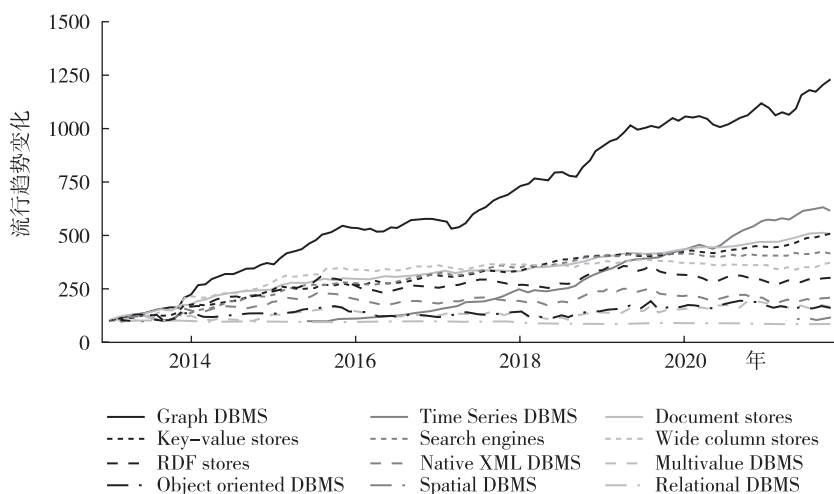


图5 不同类型数据库流行趋势

(资料来源: DB-engines.com)

目前国内外市场图数据库领域也涌现出了很多产品，在对图数据库进行技术选型时，分析DB-engines排名前15位的图数据库，主要考虑开源、成熟度、稳定性、性能、易用性、文档丰富度几个指标。

排名			DBMS	数据库模型	评分		
2021-10	2021-09	2020-10			2021-10	2021-09	2020-10
1.	1.	1.	Neo4j	Graph	57.87	+0.24	+6.53
2.	2.	2.	Microsoft Azure Cosmos DB	Multi-model	40.29	+1.76	+8.28
3.	4.	5.	Virtuoso	Multi-model	4.69	+0.27	+2.12
4.	3.	3.	ArangoDB	Multi-model	4.45	-0.34	-1.10
5.	5.	4.	OrientDB	Multi-model	4.05	-0.20	-1.43
6.	6.	8.	GraphDB	Multi-model	2.65	+0.04	+0.56
7.	7.	7.	JanusGraph	Graph	2.52	-0.07	+0.12
8.	8.	6.	Amazon Neptune	Multi-model	2.39	-0.03	-0.09
9.	9.	13.	TigerGraph	Graph	1.99	+0.08	+0.89
10.	12.	11.	Stardog	Multi-model	1.93	+0.23	+0.45
11.	10.	10.	Dgraph	Graph	1.74	-0.07	+0.06
12.	11.	9.	Fauna	Multi-model	1.57	-0.16	-0.22
13.	13.	12.	Giraph	Graph	1.28	0.00	+0.14
14.	14.	14.	AllegroGraph	Multi-model	1.25	-0.01	+0.22
15.	15.	18.	Nebula Graph	Graph	1.14	+0.04	+0.59

图6 排名前15位的图数据库

通过对比以上六个指标，最终我们选择了 Neo4j作为图存储引擎，原因如下：

（1）Neo4j社区版为开源项目，对商业应用友好，拥有对源代码的控制力，能够保证数据安全和可用性；

（2）以 Neo4j 为代表的原生图（Native Graph）数据库，以高度自由且规范的方式管理和存储数据，起步早，发展时间长，生态完整，支持多种语言，技术成熟度高；

（3）ACID完整的事务管理特性，数据库安全可靠，确保实时显示数据的合法性和准确性，是企业级应用的重要特性，运行稳定有保证；

（4）对于数据库而言，Neo4j导入数据有自带的csv格式导入工具，数据单次写入性能强，同时，Neo4j采用index-free adjacency机制，查询、遍历等性能指标优秀；

（5）Cypher是Neo4j的图形查询语言，允许用户存储和检索图数据库中的数

据。类似关系数据库中的SQL，Cypher设计借鉴了基于SQL、Python语言惯用做法，查询语言语法友好、直观、易学，学习成本低，易用性强；

（6）强大且便利的第三方Python包，有最具规模和最活跃的社区，可用成功案例及解决方案多，相关参考和学习文档丰富；

（7）Neo4j提供一套图算法库GDS，增加更多图算法，同时引入并实现了面向图学习的企业级图嵌入和图表示的机器学习模型算法过程，是全球第一个“原生图机器学习库”。