

相关研究

《选股因子系列研究（四十六）——日内分时成交中的玄机》2019.04.16

《量化研究新思维（十六）——他山之石：防御性因子择时》2019.04.14

《ESG 与社会责任投资系列研究（二）——ESG 评级在因子组合构建中的应用》2019.04.14

分析师:冯佳睿

Tel:(021)23219732

Email:fengjr@htsec.com

证书:S0850512080006

分析师:余浩淼

Tel:(021)23219883

Email:yhm9591@htsec.com

证书:S0850516050004

联系人:梁镇

Tel:(021)23219449

Email:lz11936@htsec.com

联系人:张振岗

Tel:(021)23154386

Email:zzg11641@htsec.com

金融科技（Fintech）和数据挖掘研究（二） ——知识图谱的构建与应用

投资要点:

- **知识图谱的介绍。**知识图谱的概念最早于 2012 年由谷歌提出，是一种直观显示各个实体关联信息的有效数据结构，旨在描述真实世界中存在的各种实体或概念及其关系。它脱胎于“语义网络”，随着网络和计算机技术的发展而逐渐演变。目前，知识图谱一词已经成为知识应用领域的一个通用词汇，不仅仅局限于谷歌搜索引擎的知识库及服务功能，而成为了人工智能领域炙手可热的概念。
- **知识图谱的构建与应用技术。**构建知识图谱的本质是一个从数据采集，数据建模到数据存储的过程。（1）通过结构化数据库、爬虫等方式获取知识图谱原始数据；（2）对数据进行预处理。其中，文本数据往往需要借助自然语言处理技术；（3）提取数据中的实体和关系。其中，不同来源的知识需要进行知识融合；（4）构建三元组；（5）存储三元组信息。可以采用关系数据库和 Neo4j 等图数据库；（6）使用知识图谱。
- **知识图谱在金融投资领域的应用。**通过知识图谱，我们可以进行产业链的挖掘，建立更为全面深入的事件驱动投资策略，构建知识图谱类因子，改善现有分析框架等。
- **产业链挖掘。**通过对比产业链的产品和服务网络中的“实体”与上市公司主营业务产品网络中的“实体”，并将相同的“实体”进行链接，即可确认不同公司在产业链中的具体位置。此外，知识图谱也可以根据不同的关系得到不同的网络信息，例如，股权网络挖掘、供应链挖掘、热点主题挖掘等。
- **事件驱动策略。**以业绩预警事件在应收账款网络中的传播为例，除了直接发生业绩预警事件的上市公司外，这些公司的主要应收账款债权人也受到相同方向的冲击，但影响幅度会有所减弱且具有一定的滞后性。类似地，其他事件也可能存在这种传递效应，例如，大股东重大负面信息（例如，偷税漏税）在股权网络中的传播、主营业务产品产量大幅减少在供应链网络中的传播等。
- **知识图谱类因子。**可以采用供应链、股权等数据构建图网络相关因子，包括：下游企业的业绩、上游企业的基本面、关联公司的数量和分散度、企业在图网络中的重要性、中心性等因子。由于这类因子包含关联结构，能够提供额外的信息，因而与传统因子的相关性通常较低。
- **对现有分析框架的改进。**以行业基本面量化为例，知识图谱标准化了不同品种/行业的分析和研究框架，由此带来的改进主要体现在以下三点。（1）可以对知识图谱中的任一节点进行预测和分析。例如，要研究铝的库存情况，只需选取与库存有关联关系的所有节点作为初始变量集；（2）将知识图谱中的结构信息带入量化模型中。例如，采用主成分分析降维，可以对每个包含关系下的分项变化汇总为一个主成分等。（3）可以在传统量化模型的基础上叠加行业观点。
- **风险提示。**数据挖掘是从历史先验数据获取经验模型的方法，存在模型失效可能。

目 录

1. 知识图谱的介绍.....	5
1.1 什么是知识图谱?	5
1.2 知识图谱的发展历史.....	6
1.2.1 知识图谱的前世	6
1.2.2 知识图谱的今生	7
2. 知识图谱的构建与应用技术.....	9
2.1 知识图谱的原始数据	10
2.2 知识图谱数据的预处理	10
2.3 实体、关系的提取	11
2.4 三元组的构建.....	11
2.5 存储三元组信息	12
2.6 知识图谱的使用	12
3. 知识图谱在金融投资领域的应用	12
3.1 产业链挖掘	12
3.2 事件驱动策略.....	14
3.3 构建知识图谱类因子	16
3.4 对现有分析框架的改进	16
4. 总结与讨论	21
5. 风险提示	21

图目录

图 1	知识图谱三元组的简单案例	6
图 2	谷歌搜索泰姬陵案例	7
图 3	谷歌搜索居里夫人案例	8
图 4	谷歌搜索《辛普森一家》导演案例	8
图 5	百度知识图谱的问答功能	9
图 6	知识图谱的构建过程	10
图 7	文本分析流程	11
图 8	RDF 存储数据结构	12
图 9	棉花领域产品关系示例	13
图 10	棉花领域概念关系示例	13
图 11	棉花产业链上游示意图	14
图 12	业绩预告事件的平均累计超额收益 (T 日-T+20 日)	15
图 13	业绩预告公司主要债权人的平均累计超额收益 (T 日-T+20 日)	15
图 14	业绩预增事件对事件公司与关联公司影响对比 (T 日-T+20 日)	16
图 15	资源品价格预测的研究框架	17
图 16	铜的基本面研究框架	17
图 17	铜的基本面研究的参考变量	18
图 18	资源品价格预测流程	19

表目录

表 1	铜价格的影响因素.....	18
表 2	资源品价格因子模型的预测结果.....	19
表 3	不同模型的胜率对比.....	20

1. 知识图谱的介绍

1.1 什么是知识图谱？

知识图谱的概念最早由谷歌提出，本质上是谷歌在 2012 年推出的一个项目，即“Knowledge Graph”。该项目将搜索结果进行知识系统化，以建立更智能的搜索引擎，旨在让用户能够更快、更便捷地发现新的信息和知识。随后，知识图谱的概念在学术界和业界普及并推广。目前，在智能搜索、智能问答以及个性化推荐领域等发挥着重要的作用。

知识图谱旨在描述真实世界中存在的各种实体或概念及其关系，因而可以简单地把它看作是一张巨大的网络。其中，网络的节点表示实体或概念，而节点之间的边则由属性或关系构成。

三元组是知识图谱最常用的表现方式，在图像网络中的基本形式为“节点 1-边-节点 2”。节点包括实体、概念、属性值等，边包括关系、属性等。节点与节点之间通过边进行连接，一个标准的三元组包括“实体 1-关系-实体 2”和“实体-属性-属性值”等。

• 节点的类型

实体：指客观独立的个体或者某种事物，可以是某一个人、某一个国家，也可以是某一种商品等等。世间万物都可以视作一种实体，如：巴菲特，中国，iphone 等。实体是知识图谱中最基础的元素，不同的实体间通过某种关系进行连接。

概念：具有相同特性的实体组成的集合，如投资大师、国家、手机等。也指事物的类别和种类、某些抽象概念，如人工智能、地理等。在某些时候，概念和实体的界限并不明显。

属性值：指实体某种属性的值，如：中国（实体）的面积（属性）为 963 万平方公里（属性值）等。

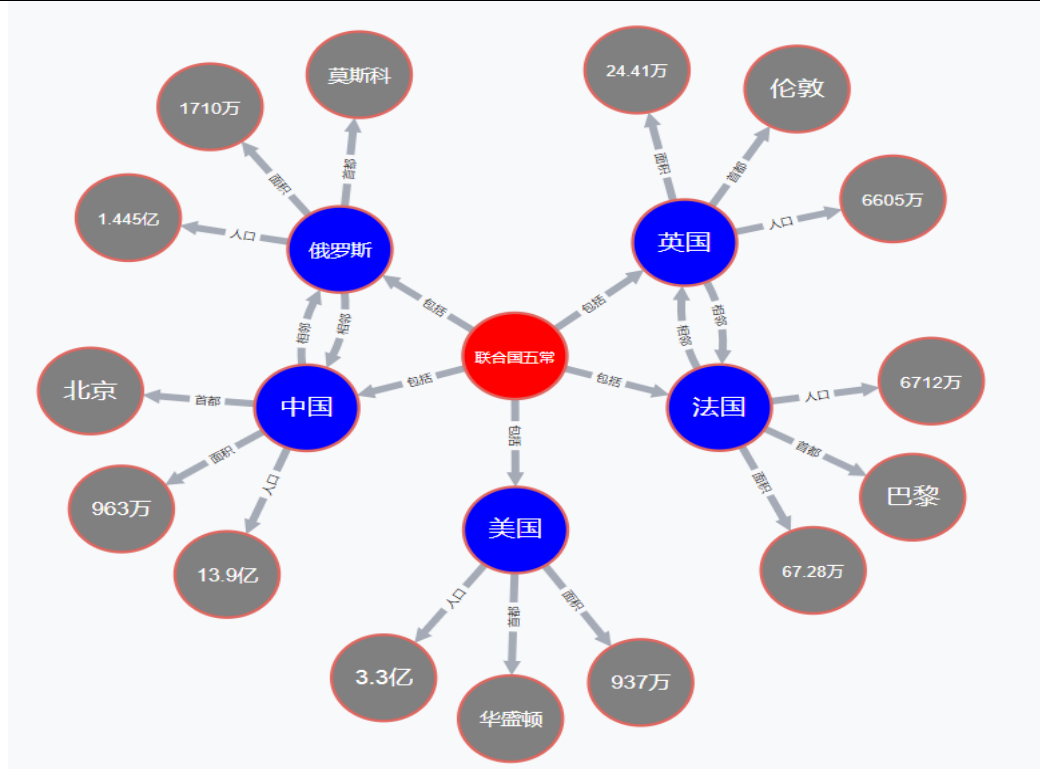
• 边的类型

关系：描述了实体及概念之间的联系，也是两个节点之间的边。比如，中国→首都→北京，中国和北京分别是两个实体，首都则是描述中国和北京的关系。

属性：也可以视作是一种关系的类型，主要指实体具有的一些特征和性质，也是实体指向属性值的边，比如生日、面积、人口数量等。比如：中国-人口-14 亿，人口便是中国的一个属性，其属性值便是 14 亿。

如下图所示，我们将一些简单的三元组案例做了图形化的展示，例如，联合国五常（概念）-包括（关系）-中国（实体），中国（实体）-相邻（关系）-俄罗斯（实体），中国（实体）-人口（属性）-13.9 亿（属性值）。

图1 知识图谱三元组的简单案例



资料来源：海通证券研究所整理

1.2 知识图谱的发展历史

虽然知识图谱的概念最近几年才逐渐被大众所知，但知识图谱本身的技术已经有了深厚的基础，而其发展更是可以追溯到上世纪五六十年代。以 2012 年知识图谱概念的提出为节点，我们可以将知识图谱的发展划分成两个阶段。

1.2.1 知识图谱的前世

1956 年，在达特茅斯学院的一次会议上，“人工智能”一词在讨论中被正式提出，这也标志着人工智能学科的诞生。在后期的发展中，人工智能的研究逐渐分成三大分支流派：符号主义、连接主义和行为主义。

符号主义（Symbolism）认为，人类的认知都是基于符号的，思维只是在符号表示上的一种运算，而机器要像人一样思考才能获得智能。因此，符号主义致力于用计算机的符号操作来模拟人的认知过程。连接主义（Connectionism）从神经生理学出发，认为人的智能是人脑高层活动的结果，因此人工智能可以通过模拟大量神经元的信号传输方式来模拟大脑的智力。行为主义（Actionism）则认为人工智能源于控制论，因此相关的研究集中于模拟人在控制过程中的智能行为和作用。在人工智能前期的研究中，符号主义一直占据主导地位，逐渐发展出了启发式算法、专家系统、知识工程理论与技术等。知识图谱最初的形态“语义网络（Semantic Network）”，就是在此期间发展起来的一种用图来表示知识的结构化方式。

语义网络由 J. R. Quillian 在 1968 年提出，其原理是用相互连接的节点和边来表示知识，节点表示实体和概念，边表示节点之间的关系。语义网络本质上也可以看作是一种用于储存知识的数据结构，即基于图的数据结构。从形式上来看，语义网络用自然的语言来描述概念实体之间的关系。其中，大部分描述由用户自行定义，往往缺乏一定的标准，因此来自多个数据源的数据难以融合。

早期的语义网络的不足之处在于缺乏逻辑，并不具备推理能力。因此，从 20 世纪 7、80 年代开始，逻辑语义的表示和推理开始成为研究的重点。在这一时期，语义网络与推理学的结合形成了描述逻辑（Description Logics）。从此，基于对象的知识被形式化，

因而具备了很强的表达能力和可判定性。

万维网的出现给语义网络提供了一个天然的应用场景，让知识从封闭走向开放。我们可以通过关联来产生更多的知识，而非仅依赖于固定的内部定义。万维网最初的内容只是人类可读，计算机无法理解 and 处理。为了解决这一难题，1998 年，万维网之父 Tim Berners-Lee 提出了语义网（Semantic Web），初衷是让机器也同人类一样可以很好地获取并使用知识。简单来说，语义网就是为了使网络上的数据变得机器可读，而提出的一个通用框架。

2006 年，Berners-Lee 提出链接数据（Linked Data）的概念。网络用户可以上传各种图结构的数据，即 W3C（万维网联盟）标准的 RDF（Resource Description Framework，资源描述框架），让数据之间建立链接，从而形成链接数据。与此同时，大规模维基百科类的知识资源的出现和信息提取方法的进步，催生了一系列高质量的链接数据知识库。例如，Freebase、DBpedia 和 Yago，它们都为谷歌知识图谱项目的成功打下了基础。

1.2.2 知识图谱的今生

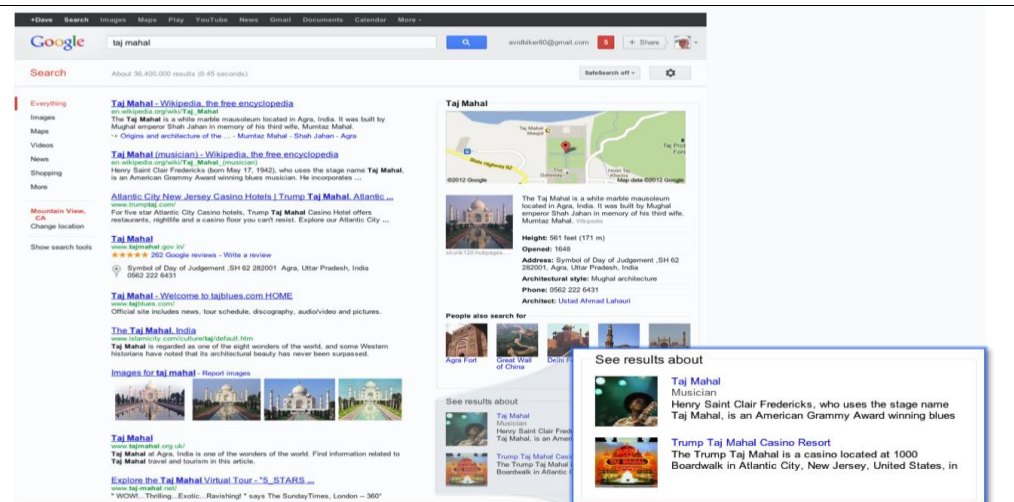
谷歌的知识图谱项目最早源于一个叫“Metaweb”的小公司。Metaweb 是开发 Freebase 的母公司，在 2010 年被谷歌收购。该公司专注于将来自多个源头的关于同一个实体或概念的描述信息融合起来，并探索这些实体的属性（例如，年龄和身高）以及彼此之间的联系，比如“身高 226cm 的 NBA 球星”和“姚明”两个描述的融合。Metaweb 存储了 1200 万个参照点（reference points），谷歌在这个基础上将规模扩张至 5.7 亿，并在这些参照点之间建立了 350 亿个连接，最终推出了“知识图谱”。

2012 年 5 月 16 日，谷歌在其官博上发表了一篇博文：Introducing the Knowledge Graph: things, not strings。在这篇文章中，谷歌的搜索引擎负责人 Amit Singhal 正式向外界介绍了知识图谱在谷歌搜索引擎中的应用。知识图谱帮助谷歌的搜索引擎进一步提升了智能化程度，实现了从 strings 到 things 的飞跃，具体表现在以下三个方面：

1. 找到正确的事物

通常，词语只有在特定的语境之下才能表达出精确的意思，而单独去搜索某个词语经常会面临一词多义的情况。例如：“小米”可以是一种农作物，也可以是科技公司、手机；“火箭”可以是实现航天飞行的运载工具，也可以是一个篮球队的名称。再如，泰姬陵（Taj Mahal）既可以指一个遗迹，也可以指一个音乐家。若我们在谷歌中搜索“泰姬陵”，则会得到如下图所示的搜索结果。在右侧，谷歌会给出泰姬陵的地图及相关介绍等。同时，也会在下方将同名的条目列出来。由此可见，借助庞大的知识图谱，谷歌可以理解实体和实体含义间的细微差别，并且将所有可能的结果归纳分组，让检索更加智能。

图2 谷歌搜索泰姬陵案例

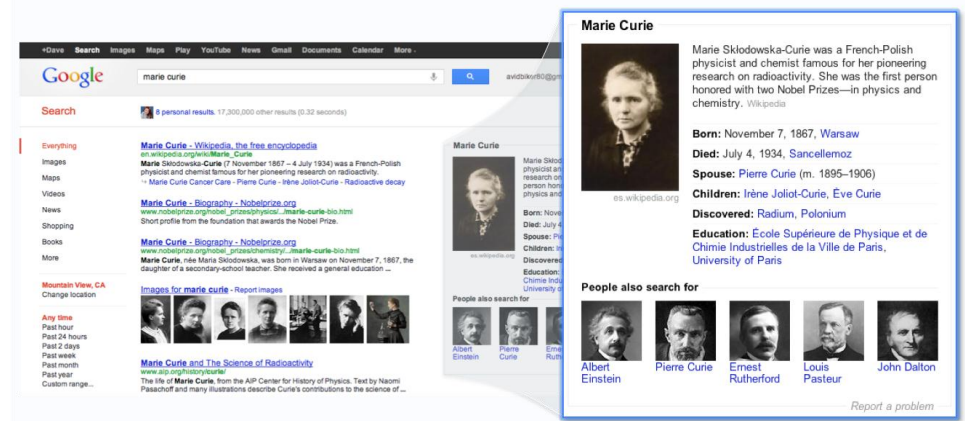


资料来源：Introducing the Knowledge Graph: things, not strings，海通证券研究所

2. 得到最优的摘要

当我们需要搜索某个事物时，借助知识图谱，谷歌可以将它的相关信息汇总，并且有条理地展示出来。但是，一个人或者一个事物的属性通常会很多，搜索者也有不同的关注点，知识图谱不可能将全部搜索结果展示给用户。因此，谷歌会通过特定的算法展示用户最关心的内容。如下图所示，当我们搜索居里夫人的时候，谷歌会将她的简介、生辰、家庭、学术成就等用户最感兴趣的信息进行汇总展示。

图3 谷歌搜索居里夫人案例



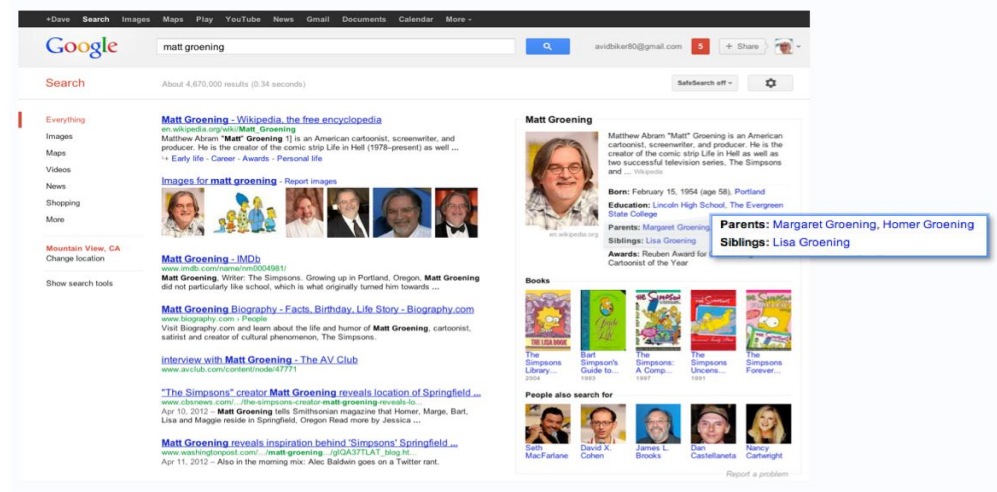
资料来源：Introducing the Knowledge Graph: things, not strings，海通证券研究所

3. 拓展深度和广度

谷歌通过知识图谱将大量的信息连接在一起，这种深层次的网络有时候可以让你发现一些新的事实或者新的关系。《辛普森一家》(The Simpsons) 是美国福克斯广播公司出品的一部动画情景喜剧，在美国电视史上有着非常高的地位，该剧的导演是马特·格勒宁 (Matt Groening)。

在谷歌搜索导演的名字“Matt Groening”，我们可以发现他的家庭信息。父母分别叫“Homer Groening”和“Margaret Groening”，妹妹叫“Lisa Groening”。《辛普森一家》中的辛普森夫妇也叫“Homer”和“Margaret”，女儿也叫“Lisa”，儿子“Bart”也和导演的名字“Matt”相近。因此，通过知识图谱，我们可以发现《辛普森一家》的命名灵感来自于导演自己的家庭。

图4 谷歌搜索《辛普森一家》导演案例



资料来源：Introducing the Knowledge Graph: things, not strings，海通证券研究所

本质上，谷歌并不是知识图谱这项技术的创造者，而是知识图谱这个概念的推动者。2012年，谷歌推出“知识图谱”的功能之后，知识图谱一词逐渐成为知识应用领域的一

个通用词汇。知识图谱也不仅仅局限于谷歌搜索引擎的知识库及服务功能，而是逐渐成为了人工智能领域炙手可热的概念。很多技术积累深厚的互联网公司也纷纷顺势推出自己的知识图谱，比如微软、Facebook、亚马逊、百度、阿里巴巴、美团等。

谷歌的知识图谱主要来源于 Wikipedia、Freebase 等知识库。作为全球最大的中文搜索引擎，百度的知识图谱的知识来源主要来自旗下的百科、知道和贴吧等。百度的知识图谱产品目前主要分成两大类：智能搜索和自动生成内容。百度知识图谱包含上亿的实体单位，依托知识图谱搜索引擎可以更好地理解用户搜索意图。比如，直接搜索“2020年春节还剩多少天”，百度知识图谱的问答功能会直接生成答案（如下图所示）。

图5 百度知识图谱的问答功能



资料来源：百度知识图谱，海通证券研究所

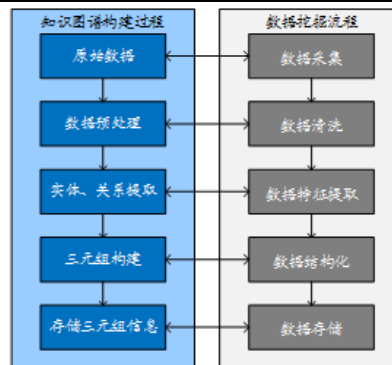
除了谷歌和百度在搜索引擎上的功能扩展之外，知识图谱在互联网领域还有着广泛的应用。比如：Facebook 在 2013 年发布社交知识图谱，主要应用于个性化推荐和智能搜索等，重点关注用户的社交偏好信息；美团点评主要关注餐饮娱乐方面，重点从美团点评的用户评价、菜品、标签等数据，构建餐饮娱乐知识图谱——“美团大脑”。

除了在传统的互联网领域，知识图谱在其他行业同样可以起到重要的作用。例如，在反欺诈场景，知识图谱有着天然的应用优势。欺诈团伙通常会用虚假的身份和信息进行诈骗，一个谎言需要无数个谎言来掩盖，而虚假的信息往往容易露出马脚。不一致性验证可以用来判断一个借款人是否具有欺诈风险。比如，借款人 A 和 B 分别填写了两家不同的公司，但是填写的公司电话却是同一个，这种情况便无法通过不一致性检验，从而形成一个风险点。

2. 知识图谱的构建与应用技术

构建知识图谱的本质是一个从数据采集，数据建模到数据存储的过程。其基本的步骤如下图所示。简单而言，知识图谱的构建过程与数据挖掘的基本流程可以一一对应，即知识图谱构建只是数据挖掘技术的一种实践。

图6 知识图谱的构建过程



资料来源：海通证券研究所整理

2.1 知识图谱的原始数据

作为一种新的组织、管理、认知客观世界的工具，知识图谱的数据来源决定了知识图谱能够描绘出怎样的一个“世界”。与互联网知识图谱不同，要构建金融领域知识图谱，原始数据不能仅仅是从金融文本当中获取到的知识。已知的公司信息，标的信息，价格数据，甚至不同证券的相关性信息都可以作为原始数据帮助构建知识图谱。

由于金融领域的“数据密集型产业”特性，目前已有非常多的数据提供商整理了金融领域中的各种数据信息。利用好这些数据，是获取知识图谱原始数据最有效的方式。

金融领域已有的数据往往会局限于和公司、可交易标的、宏观指标紧密相关的信息。然而，知识图谱的构建往往需要借助其它领域的先验知识，将金融领域的各种实体进行链接。例如，在利用知识图谱分析产业链结构时，产业上下游分别由那些产品构成等相关信息，就无法从金融数据源直接获得。需要对于该行业的先验信息进行分析，才可以获取到相关知识。

网络爬虫是获取原始数据的一种非常有效的工具。通过从特定网站上爬取所需要的内容，可以方便地获得已定位好的先验信息，这对补充构建知识图谱所需的原始数据非常有帮助。

2.2 知识图谱数据的预处理

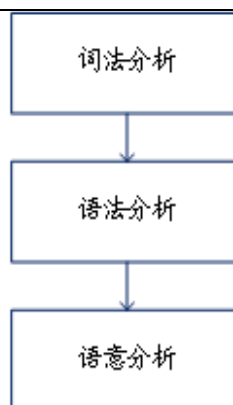
构建知识图谱所获取的原始数据有结构化与非结构化两种类型，分别对应着不同的预处理方式。

构建金融知识图谱，所需要处理的结构化数据一般包括表格化的证券基本信息，价量信息，基本面数量指标等。明晰的数据结构不仅为数据处理提供了很多便利，而且能大幅提升准确性与高效性。由此得到的知识实体和关系，也较非结构化数据更为准确。

除了传统的结构化数据本身，利用这些数据计算得到的衍生数据也是知识图谱中“实体”或者“关系”的重要来源。例如，可交易证券之间的相关系数、协整系数等。保存这些信息的时变特性，可以让知识图谱随时间演进，为更客观地描述金融市场提供可能。

非结构化数据一般包括图片、音频、视频、文本信息等。其中，文本数据是金融领域最常见的非结构化数据。而获取文本数据当中的信息，往往需要借助自然语言处理技术。具体流程如下图所示。

图7 文本分析流程



资料来源：海通证券研究所整理

对于一个句子、一段话，这种短文本而言，分析出其中的主谓宾结构，一般就可以获取知识图谱所需的“实体”与“关系”对象。而处理长文本，则需要通过词频统计，主题模型等更进一步的分析方法，判断出长文本中真正有效的“实体”对象。

2.3 实体、关系的提取

从经过处理的数据中获取有效的“实体”与“关系”是构建知识图谱的关键。在金融领域中，最核心的实体即公司、可交易证券、宏观经济指标等；最有价值的关系，往往与影响公司业绩或影响可交易证券价格的一些信息有关。然而，仅依赖金融领域的先验知识，用传统关联方法将实体进行连接，并不能发挥知识图谱的最大功效。将其他领域有用的“实体”、“关系”关联，可以有效拓展知识图谱边界，获取隐藏信息。

以产业链构建为例，若从公司财务报表和公告信息中直接挖掘，只能找出公司之间财务或股权方面的关联。而如果能提炼公司的主营产品，并利用其它产业产品之间的上下游关系，就可以将分属产业上下游不同层面的公司进行关联，获取更好的产业链结构。

建立知识图谱的另一项重要工具是知识融合，即，将不同来源的知识进行对齐、合并。在自下而上构建知识图谱的过程中，往往会在不同的知识图网中发现表示同一知识的节点。将这两个图网融合，可以避免知识孤岛，让知识链接更加稠密，提升知识的应用价值。

知识融合一般包括“实体”合并、“实体”概念合并、“实体”概念上下位关系合并等。先验信息是知识融合的关键，特别是从已知的结构化数据中得到的“实体”关联关系，可以有效帮助不同的知识图谱进行融合。

例如，将从财务信息挖掘得到的公司财务图谱与从市场价格挖掘得到的证券收益特征图谱进行融合，就可以获得财务数据、研发方向、新产品特性等信息对公司股价的影响路径，实现财务指标图谱的核心“实体”——公司与证券收益特征图谱的核心“实体”——股票之间的关联。由此可见，相较于透过复杂的文本主题分析进行“实体”关联，利用类似的结构化数据提供的信息可以大大提升知识融合的效率。

2.4 三元组的构建

RDF 是目前最为主流的“实体”-“关系”符号语义表示模型。其基本模型为有向标记图，图中每一条边对应一个三元组（主语-谓语-宾语），即前文所提到的“实体”-“关系”-“实体”。而这样一个三元组也对应了一个逻辑表达式或者关于世界的陈述。

在定义 RDF 三元组信息的同时，也可以通过扩展的数据结构保存三元组中“实体”与“关系”的属性结构。例如，“股票”实体的价格序列，线性相关“关系”的相关系数等。增加三元组的时间戳属性，更可以将知识图谱动态化，描述出不同“实体”间“关系”随时间的演进情况。

2.5 存储三元组信息

目前，主流的关系数据库都可以实现三元组信息的存储（见下图）。

图8 RDF 存储数据结构

LEFT_ENTITY_ID	RELATION_ID	RIGHT_ENTITY_ID	BICONNECTED_SIGN	S_CON_INDATE	S_CON_OUTDATE
6a731864-3d1a...	91e94e97-31...	91b00d-1d3d-45d...	0	19901219	NULL
32a8abcb-3af2...	91e94e97-31...	fdb71d2e-ca0b-493...	0	19901219	NULL
abcbe000-2b8f...	91e94e97-31...	18d14871-4a11-4bcc...	0	19901219	NULL
a8b2f334-8966...	91e94e97-31...	29001738-26d0-4f0b...	0	19901219	NULL
76f05ee0-bf77...	91e94e97-31...	9c77b282-557b-4d2b...	0	19901219	NULL
8138c312-22e8...	91e94e97-31...	9cc2ebc6-6fe7-4fd9...	0	19901219	NULL
3d9947acc-695...	91e94e97-31...	bcc80bbb-49f5-46bc...	0	19901219	NULL
e8e7b44-9ef9...	91e94e97-31...	b0d04b6d-3f23-4522...	0	19901219	NULL
72b4cbf9-eece...	91e94e97-31...	edbe893f-0c26-408e...	0	19901219	NULL
3f472ae5-085c...	91e94e97-31...	f470f291-f09d-49ed-8...	0	19901219	NULL
8dddb4f8-2f0d...	91e94e97-31...	1b25ae95-7a4b-42d9...	0	19901219	NULL

资料来源：海通证券研究所整理

如上图所示，RDF 三元组信息分别由关系数据库 LEFT_ENTITY_ID，RELATION_ID, RIGHT_ENTITY_ID 标识。其中，还保存了有效期与是否为双向的标志，提供了更丰富的信息。

在其他数据表中，通过 ENTITY_ID 或者 RELATION_ID 链接，可以查找到该“实体”或者“关系”的更多属性信息。遍历知识图谱，按条件筛选匹配“关系”，从而得到所需“实体”的关联网络是知识图谱的重要应用。这一过程往往要求“实体”或者“关系”都存在属性信息。

除了关系数据库之外，还有很多专门用于存储知识图谱三元组结构的数据库系统，如 Neo4j 等。受限于关系数据的索引结构，无论是插入、查找还是遍历，通过关系数据库方式存储的知识图谱性能都会受到影响。随着知识图谱规模的逐步扩大，利用专用数据库对知识图谱结构进行存储变得愈发重要。

2.6 知识图谱的使用

使用知识图谱首先需要确定起始的“实体”、“实体”类别，或某些“关系”，并利用它们的属性信息读取所有满足条件的三元组数据。

以已经读出的信息为起点，遍历与其连接的所有“实体”，根据预先定义的条件剔除不需要的“关系”或“实体”，从而得到第二层的 RDF 信息。

迭代上述过程，直到数据库中所有符合条件的“实体”和“关系”都已被遍历到。由此，一张全联接的知识图网便形成了，其中也包含了我们所需要了解的相应信息。

上述过程的本质是一个图的遍历过程，因而可以直接应用于产业链结构等关系图的构建。除此之外，将三元组信息向量化，通过机器学习等算法，分析其中包含的信息，也是使用知识图谱的另一种重要方式。

3. 知识图谱在金融投资领域的应用

3.1 产业链挖掘

应用知识图谱最为基础的方式，就是通过节点关系将尽可能多的信息整合在一起，并遍历某种关系下的实体集合，进而可视化地展示全景图。在金融投资领域，典型的应用场景就是产业链挖掘。

在本系列第一篇报告《数据挖掘技术框架简介》中，我们详细介绍了棉花产业链挖

掘的实例，共分为三步。

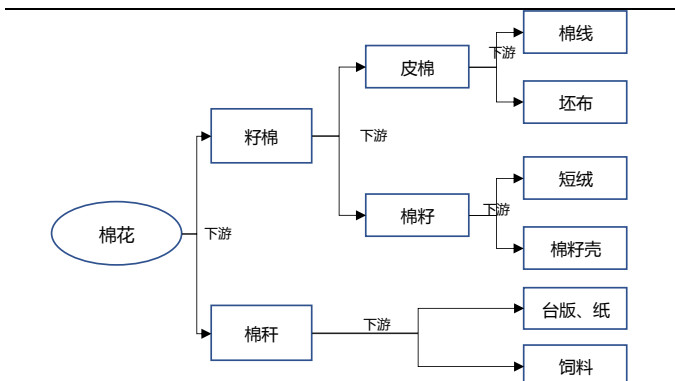
1. 构建棉花领域的产品和服务网络

首先，通过棉花相关产品和服务之间的关系，例如“棉线是皮棉的下游产品”、“棉花种植是棉花的上游服务”等，可以将棉花领域的商品和服务构建相互关系，如左下图所示。

其次，导入不同产品和服务的概念从属关系，例如“上游包括棉花种子的生产销售，农业采棉技术的服务管理等”、“下游包括棉纱、棉布等商品的生产和销售”，如右下图所示。

最终，得到具有层次信息的棉花领域的产品和服务网络。

图9 棉花领域产品关系示例



资料来源：海通证券研究所整理

图10 棉花领域概念关系示例

实体	关系	对象
棉花上游产品	包含	棉花种植
棉花上游产品	包含	棉花种子
棉花上游产业	包含	农机采棉技术与管理服务
棉花上游产品	影响	棉花供给
棉花加工	包含	染色棉花
棉花加工	包含	棉花初加工
棉花经营	包含	棉花、棉纱贸易
棉花经营	包含	棉花销售
棉花下游产业	影响	棉花需求
棉花下游产业	包含	棉纱经营
棉纱经营	包含	涤棉纱
棉纱经营	包含	改性棉纱
棉纱经营	包含	面料、床品

资料来源：海通证券研究所整理

需要注意的是，上述概念关系的建立过程中，往往需要掌握相关先验信息，如棉花的上游产业和下游产品分别包括哪些对象。这些先验信息的获取与清洗是能否构建有效知识图谱的关键所在。

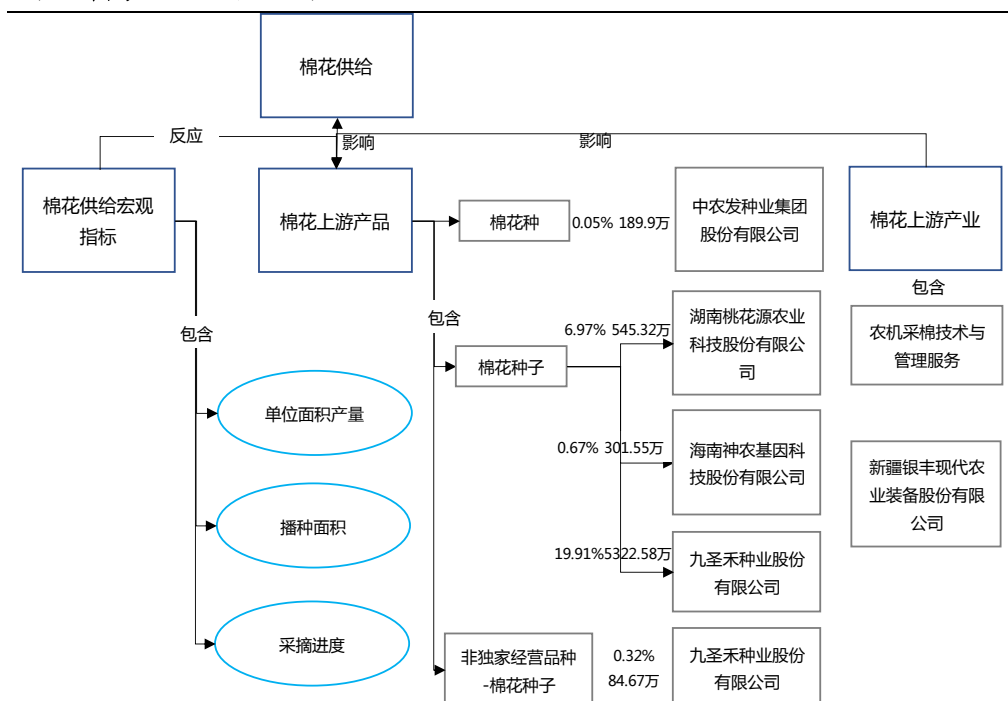
2. 导入上市公司主营业务产品信息

从结构化数据库中导入上市公司主营业务产品的相关信息，包括报告期、主营业务产品的名称、业务收入、业务收入占比等。其中，上市公司与主营产品为“实体”，主营业务为“关系”，报告期、业务收入及占比为“关系”的属性。

3. 将上市公司主营业务产品链接到棉花领域的产品和服务网络

将棉花领域的产品和服务网络中与上市公司主营业务产品网络中相同的“实体”进行链接，即可以确认不同公司在棉花产业链中的上下游位置。例如，将有棉布经营的相关主营业务的公司“实体”与棉花服务中的“棉布经营”链接，而“棉花经营”从属于棉花产业链中的“中游”这一概念，即可得到该公司处于棉花产业链的中游。据此挖掘出的部分产业链如下图所示（仅以上游为例，具体信息可参考本系列第一篇报告）。

图11 棉花产业链上游示意图



资料来源：海通证券研究所整理

通过以上方式，就可以规则化、时序化地构建棉花产业链。“规则化”是指产业链生成过程的构建摒弃了传统人工选择的方式，具有固定的生成规则，便于回溯和更新迭代。“时序化”是指产业链信息会随着时间的变化发生改变。例如，如果某个棉花种植公司主营业务变更为棉花加工，那么它在产业链中的位置就从上游变为中游。在知识图谱中，这种时间序列上的产业链变更，可以通过对实体间的关系加上时间序列属性进行维护。

除了产业链，知识图谱还可以根据不同的关系得到不同的网络信息。例如，股权网络、供应链、热点主题等。实际上，类似的应用在金融领域早有先例。典型的代表是AlphaSense，一家成立于2008年的智能搜索引擎公司。它从文件、新闻和研究报告等碎片化信息中集合所有投资信息，并进行语义分析，整合相关概念、主题、行业等，方便用户搜索、浏览和分析金融投资相关的重要数据点及整体趋势。

3.2 事件驱动策略

挖掘产业链等信息是知识图谱基本功能的展现。在金融投资领域，更进一步的是构建事件驱动策略。

传统的事件驱动策略主要研究事件对于发生实体的有价证券价格的影响。例如，苹果公司iPhone 5s推出时，市场评价不佳，公司股价应声下挫。这一类传统策略主要有两个特点：一是人工处理，即iPhone 5s产品市场评价与苹果公司股票价格之间的关系，主要通过人工确认；二是只研究直接发生事件的实体的股价表现。而借助知识图谱，则可以实现更为及时、深入的事件驱动策略。

首先，通过“主营业务产品”这一关系将iPhone 5s与苹果公司相关联，再通过从属关系将Twitter评论与iPhone 5s相关联。当Twitter评论出现显著的负面情绪时，就可以通过这两层关系快速链接到苹果公司股价，生成投资信号。

其次，通过“主要供应商”这一关系可以将苹果公司链接到富士康。我们发现，富士康的股价在这一事件发生后同样出现大跌，原因就是“产品市场认可度较差”这一事件在供应链网络中的传播。

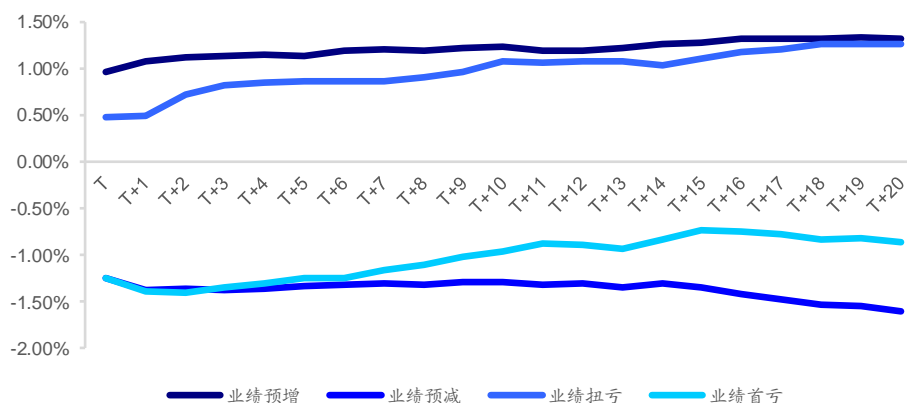
换句话说，当知识图谱中的某个节点受到冲击时（例如，负面消息），会通过图网络辐射到其他关联实体。因此，通过知识图谱可以将事件驱动策略的标的进一步扩展。同

时，如果发生事件的是某个非上市公司，甚至自然人，也可以通过股权等关系链接到相关上市公司。从而在事件驱动策略中纳入许多非金融的信息和更多层的关联结构，大大扩展了策略的适用空间。

下面，我们通过应收账款网络简单展示这类应用的效果。以传统的业绩预告事件为例，我们首先选取了影响相对显著的业绩预增、预减、扭亏、首亏四类事件。以2005/1/1-2018/12/31的业绩预告数据为回测样本，其中剔除了上市未满6个月的新股。

由于业绩变动较大的公司具有一定的行业 and 市值偏离，因此我们采用上市公司所在行业的20个相近市值公司的平均收益作为比较基准。四类事件在发生后T日至T+20日的平均累计超额收益如下图所示。

图12 业绩预告事件的平均累计超额收益（T日-T+20日）

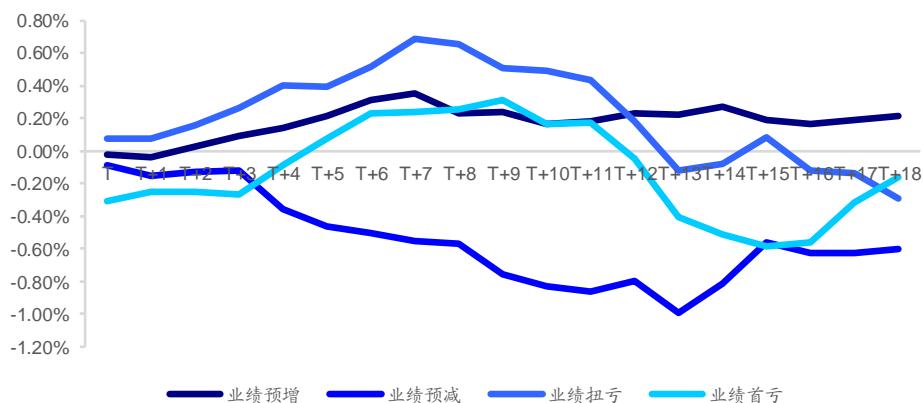


资料来源：Wind，海通证券研究所

由上图可见，业绩预增和业绩扭亏具有显著的正向效果，公告披露后的20天内存在1.3%左右的超额收益。业绩预减和业绩首亏具有显著的负向效果，公告披露后20天的超额收益在-1.0%至-1.5%左右。但是，我们也很容易发现，不论是哪一类事件，公司股价在公告当日的反应最大。之后的20个交易日内，除业绩首亏事件外，其余事件的平均累计超额收益变动并不明显。这一现象表明，在没有内幕消息的前提下，投资者实际上很难获取业绩预告等事件理论上的超额收益。

然而，知识图谱却给我们提供了一种另类的参与此类事件的方式。例如，我们可以构建上市公司之间的应收账款网络，选取披露业绩预告的公司作为关联对象，通过财报附注中的主要应收账款信息链接到主要债权人。直观上，某个公司业绩的预增或预减，都会对它的应付账款产生影响，并传导至它的债权人。因此，相关债权人在业绩预告事件发生后T日-T+20日内的股价应当也会发生同方向的变动。下图给出了这种思路下，与发生业绩预告事件的公司有债务关系的上市公司的累计超额收益。

图13 业绩预告公司主要债权人的平均累计超额收益（T日-T+20日）

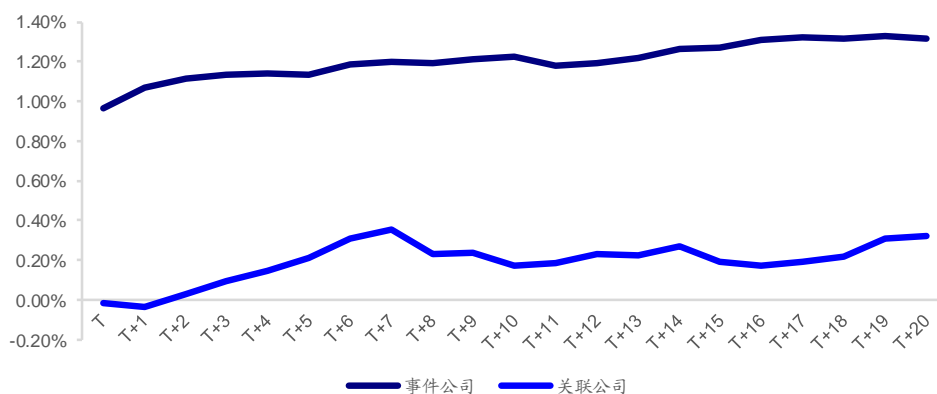


资料来源：Wind，海通证券研究所

从上图可以看到，业绩扭亏、业绩预增公司的债权人在公告披露后的 7 天内，平均超额收益单调上升，至 T+7 日的值为 0.8%、0.4% 的，此后才逐渐衰减。

我们进一步以业绩预增事件为例，对比事件公司和债权人公司在公告后的股价变化（见下图）。显然，后者的超额收益更为滞后，股价在公告当日并无明显变化，此后才逐步上涨。但是，我们也应看到，和事件公司相比，债务关联公司的平均超额收益略低。

图 14 业绩预增事件对事件公司与关联公司影响对比（T 日-T+20 日）



资料来源：Wind，海通证券研究所

其中的原因也十分直接。若 A 公司欠 B 公司大量应收账款，A 公司公告业绩预增事件，很可能会利好 B 公司应收账款的回收，因此 B 公司的股价上涨。但是，这种正面信息对 B 公司产生的影响毕竟只是间接的，其超额收益不如事件发生公司也在情理之中。

上面这个例子表明，借助知识图谱，我们完全可以拓展事件驱动策略的应用范围，从关注事件公司本身转向和它有紧密关联的公司，寻找更为确定的右侧交易机会。

事实上，我们可以根据这种思路，构建一系列基于图网络的事件驱动策略。例如，大股东重大负面事件（如，偷税漏税）在股权网络中的传播，主营业务产品大幅减产在供应链网络中的传播，等等。在海外，此类应用已经成为量化基本面投资（quantamental）的一类主流策略。一个典型的案例是一家名为 kensho 的互联网公司。它的主要产品为 warren 问答引擎，专注于寻找事件与资产之间的相关性，从而预测资产价格走势。2017 年初，该公司成功预测了特朗普当选后，美元重回涨势，以及科技股的反弹。

3.3 构建知识图谱类因子

我们还可以基于知识图谱构建另类因子，为传统的多因子模型提供额外的信息。例如，德银于 2015 年采用供应链数据构建了供应链网络相关因子，包括：下游企业的业绩、上游企业的基本面、关联公司的数量和分散度、企业在图网络中的重要性、中心性等。经过检验，这些因子都具有较为显著的超额收益。而且，由于包含了关联结构，它们与传统因子都有着很低的相关性，是传统模型的一个有益补充。

3.4 对现有分析框架的改进

知识图谱除了能提高投资者对非结构化信息的应用效率，还能改善现有的分析框架。由于知识图谱刻画了实体之间的关系，因而在预测实体的过程中，哪些是有效信息实际上已被事先确定。与此同时，预测逻辑也随之固定下来。

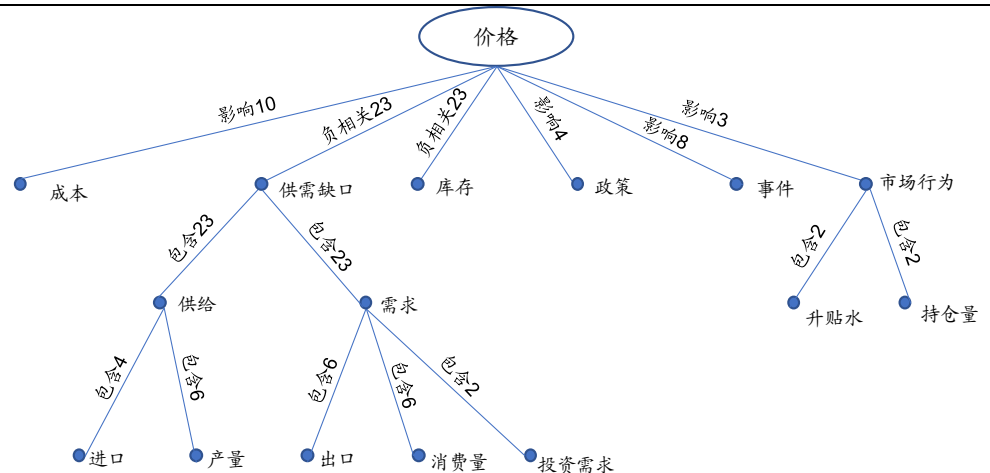
以行业基本面量化为例。我们与行业研究团队合作，开展了资源品价格预测的研究。其分析框架的构建过程可被简化为以下三步。

1. 构建资源品行业知识图谱

首先，行业研究员提供基本面研究框架和参考变量，例如，价格与供需缺口的关系、

库存所包含的指标以及各自公布的滞后期，等等。其次，我们根据以上信息，构建出资源品领域的知识图谱。最后，通过对知识图谱主要概念的提取，整合资源品行业的研究框架（如下图所示）。

图15 资源品价格预测的研究框架



资料来源：Wind，海通证券研究所

上图中每条边对应的数字代表有几个资源品存在类似的“关系”表达。考虑到“关系”数量众多，我们仅展示存在不止一个资源品有类似表达的“关系”，即所有数字都大于1。由于不同资源品研究员的关注点和输入信息有差异，因此最终代表资源品研究框架的知识图谱将包含全部“关系”。这样一来，既可以通过其他资源品的分析框架对某个资源品的现有分析框架进行补充，也可以对不同分析框架相冲突的地方进行检验和修正。

我们以有色金属——铜为例，根据上述方法，建立并完善它的基本面研究框架。具体结果如以下两图所示。

图16 铜的基本面研究框架

概念间的关系

实体	关系	对象
价格	负相关	供需缺口
供需缺口	包含	供给
供需缺口	包含	需求
库存	负相关	价格
事件	影响	价格
升贴水	影响	价格
供给	包含	产量
供给	包含	产能利用率
持仓量	影响	价格
需求	包含	下游需求
需求	包含	消费量
需求	包含	进口量
需求	包含	出口量
冶炼成本	影响	价格

概念与实体的关系

实体	关系	对象
库存	包含	总库存:LME铜
库存	包含	COMEX:铜:库存
库存	包含	库存小计:阴极铜:总计
库存	包含	上海保税区铜库存
库存	包含	LME铜:注册仓单:合计:全球
库存	包含	LME铜:注销仓单:合计:全球
价格	包含	现货结算价:LME铜
价格	包含	铜期货结算价(活跃)
价格	包含	期货收盘价(电子盘):LME3个月铜
价格	包含	长江有色市场:平均价:铜
升贴水	包含	LME铜升贴水(0-3)
升贴水	包含	平均价:铜升贴水:上海物贸
产量	包含	ICSG:全球矿山产量:当月值
产量	包含	ICSG:原生精炼铜产量:当月值
产量	包含	ICSG:再生精炼铜产量:当月值
产量	包含	产量:精炼铜(铜):矿产:当月值
产量	包含	产量:精炼铜(铜):再生:当月值
产量	包含	产量:精炼铜(铜):当月值
产量	包含	产量:铜材:当月值

资料来源：Wind，海通证券研究所

图17 铜的基本面研究的参考变量

实体的属性						
代码名称	变量名称	wind代码	频率	滞后时间	是否比率	
LMECopper_CashSettlementPrice	现货结算价:LME铜	S0029751	D	1	0	
LME3MonthsCopper_FuturesClosingPriceElectronicDisc	期货收盘价(电子盘):LME3个月铜	S0029750	D	1	0	
Copper1_ChangjiangNonferrousMarketAveragePrice	长江有色金属市场:平均价:铜	S0182161	D	0	0	
LMECopperLiters	LME铜升贴水(0-3)	S5806058	D	1	0	
AveragePrice_CopperLitreDiscountShanghaiTrade	平均价:铜升贴水:上海物贸	S5806984	D	1	0	
GlobalMineOutput_ICSG	ICSG:全球矿山产量:当月值	S0148686	M	4	0	
NativeRefinedCopperOutput_ICSG	ICSG:原生精炼铜产量:当月值	S0148689	M	4	0	
RegeneratedRefinedCopperOutput_ICSG	ICSG:再生精炼铜产量:当月值	S0148690	M	4	0	
RefinedCopperMineralsOutPut_China	产量:精炼铜(铜):矿产:当月值	S0181881	M	3	0	
RefinedCopperRenewableOutPut_China	产量:精炼铜(铜):再生:当月值	S0181883	M	3	0	
RefinedCopperOutPut_China	产量:精炼铜(铜):当月值	S0027555	M	2	0	
CopperProductOutPut_China	产量:铜材:当月值	S0027559	M	2	0	
CapacityUtilizationRatioOfRefinedCopper_ICSG	ICSG:精炼铜产能利用率:当月值	S0148693	M	4	1	
GlobalCopperProductionInMines	全球矿山铜产量	S0109369	Y	1	0	
GlobalRefinedCopperProduction	全球精炼铜产量	S5800121	Y	1	0	
GridInfrastructure	电网基本建设投资完成额:累计值	S5108453	M	2	0	
NewBuildingArea	房屋新开工面积:累计值	S0029669	M	2	0	
CarOutPut_China	产量:汽车:当月值	S0105523	M	2	0	
NewEnergyVehicleOutPut_China	产量:新能源汽车:当月值	S0243299	M	2	0	
FridgeOutPut_China	产量:冰箱:当月值	S5616347	M	2	0	
AirConditionerOutPut_China	产量:家用空调:当月值	S5616239	M	2	0	
WashingMachineOutPut_China	产量:洗衣机:当月值	S5616422	M	2	0	
GlobalConsumptionOfRefinedCopper	全球精炼铜消费量	S5800129	Y	1	0	

资料来源: Wind, 海通证券研究所

对于每个资源品,在根据图 15 建立了它的价格预测框架后,便可以提取所有与其价格相关的“实体”和“关系”,得到最终的可供进一步研究的基础数据。依然以铜为例,提取出的 RDF 数据如下表所示。

表 1 铜价格的影响因素

因子类型	包含指标	描述
库存因子	总库存:LME 铜、COMEX:铜:库存、库存小计:阴极铜:总计、上海保税区铜库存、LME 铜:注册仓单:合计:全球、LME 铜:注册仓单:合计:全球、LME 铜:注销仓单:合计:全球	主要包括各类铜的库存以及仓单的指标,为供给和需求博弈的结果
情绪因子	现货结算价:LME 铜、铜期货结算价(活跃)、期货收盘价(电子盘):LME3 个月铜、长江有色金属市场:平均价:铜、LME 铜升贴水(0-3)、平均价:铜升贴水:上海物贸	主要包括期货和现货价格以及升贴水,表征市场投资偏好
供给因子	ICSG:精炼铜产能利用率:当月值、ICSG:全球矿山产量:当月值、ICSG:原生精炼铜产量:当月值、ICSG:再生精炼铜产量:当月值、产量:精炼铜(铜):矿产:当月值、产量:精炼铜(铜):再生:当月值、产量:精炼铜(铜):当月值、产量:铜材:当月值、全球矿山铜产量、全球精炼铜产量	主要包括精炼铜以及铜矿的产量指标和产能利用率,为供需分析的一部分
需求因子	全球精炼铜消费量、进口数量:铜矿石及精矿:当月值、进口数量:废铜:当月值、进口数量:铜材:当月值、出口数量:铜材:当月值、出口数量:智利:精矿含铜量:中国、出口数量:智利:精炼铜:中国、房屋新开工面积:累计值、产量:汽车:当月值、产量:冰箱:当月值、产量:家用空调:当月值、产量:洗衣机:当月值	主要包括进出口量、消费量以及下游需求,为供需分析的一部分
成本因子	中国铜精矿现货 TC、中国铜精矿现货 RC	主要包括冶炼成本

资料来源: Wind, 海通证券研究所

2. 构建量化预测框架

为了预测资源品的价格,我们选取了三种较为典型的时间序列模型。

(1) ARIMA 模型

ARIMA 是时间序列分析中的经典模型。下文将从自动定阶的 ARIMA 模型得到的预测结果,作为其他预测模型的学术级比较基准。

(2) 序列分解模型

将预测变量分解为趋势、周期、节假日效应、随机项 4 个部分,对每个部分分别拟合一个多项式,单独预测,最终将 4 个分项的预测值汇总作为整体预测值。此类方法的典型代表是 Facebook 于 2017 年发布的开源数据预测包——“Prophet(先知)”,在预测互联网企业用户数量等领域卓有成效。下文以 Prophet 作为其他预测模型的工业级比较基准。

(3) 因子模型

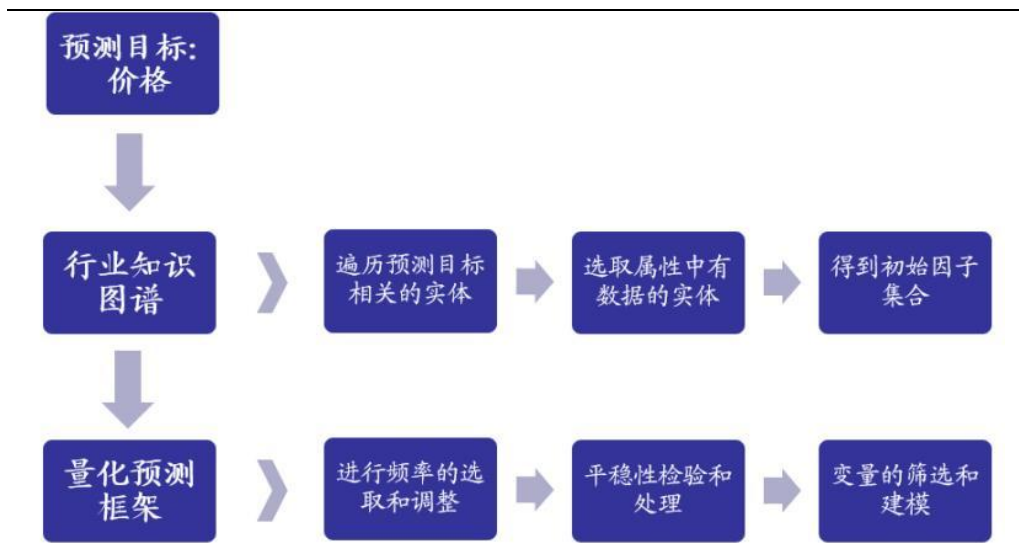
以上两个模型适用于预测变量有较强趋势性，受其他外生变量影响较小的情况。在经济预测或金融投资的应用中，包含多个变量的 VAR 模型和因子模型的应用更为广泛。

因子模型的难点在于变量选择和参数估计。如果变量太多，就会造成“维数灾祸”，如多重共线性，从而降低估计的可靠性。变量筛选或降维的主流方法有 LASSO、PCA 和因子分析等。下文采用 LASSO+CV 的方式确定预测模型。

3. 行业知识图谱+量化预测框架

在第 1 步中，我们已从行业知识图谱中提取了每个资源品的价格“实体”，并在遍历所有的“关系”后，得到每个品种价格的影响因素（可参见上文铜的示例）。随后，将这些基础信息放入第 2 步中的 3 个预测模型。这个过程可由如下的流程图描述。

图18 资源品价格预测流程



资料来源：海通证券研究所整理

采用每个资源品截止 2018 年末的所有可得数据进行回测，参数由最近 150 个数据点确定，并按周度频率更新。因子模型的结果如下表所示。绝大部分品种的价格预测胜率（预测的涨跌方向和实际一致）均在 50% 以上。其中，动力煤、铜、天然橡胶等品种表现较优，年化收益超过 10%。不过，镍的历史数据较短，结果仅供参考。

表 2 资源品价格因子模型的预测结果

	年化收益	信息比率	最大回撤	胜率	盈亏比
镍	51.34%	2.87	7.30%	60.00%	2.34
动力煤	29.64%	1.25	17.11%	59.83%	1.41
铜	20.24%	0.87	48.96%	56.45%	1.38
原油	18.62%	0.51	83.96%	52.17%	1.20
铅	12.94%	0.75	20.28%	51.02%	1.35
天然橡胶	10.74%	0.36	59.88%	54.72%	1.11
白银	10.16%	0.61	17.33%	50.53%	1.26
PTA	8.12%	0.40	34.10%	52.28%	1.15
锌	7.69%	0.39	32.67%	50.00%	1.15
黄金	4.68%	0.32	48.63%	55.67%	1.16
焦炭	3.53%	0.10	61.12%	53.53%	1.03
铁矿石	3.49%	0.10	43.44%	51.33%	1.05
甲醇	3.38%	0.13	40.17%	51.42%	1.04
大豆	0.14%	0.01	50.52%	53.07%	1.00

锡	0.06%	0.00	7.73%	43.59%	1.00
螺纹钢	-0.20%	-0.01	69.25%	54.76%	1.00
玉米	-2.91%	-0.22	51.52%	49.02%	0.92
PVC	-3.26%	-0.19	41.15%	51.51%	0.93
棉花	-4.36%	-0.23	66.55%	50.09%	0.91
铝	-6.54%	-0.44	80.24%	50.25%	0.84
食糖	-11.89%	-0.71	71.71%	47.53%	0.71
焦煤	-16.65%	-0.43	51.92%	47.52%	0.81
沥青	-27.26%	-0.91	57.21%	49.12%	0.63

资料来源：Wind，海通证券研究所

下表进一步对比了因子模型和 ARIMA、Prophet 模型的预测效果。从胜率来看，因子模型在绝大多数品种上都表现得更好。

表 3 不同模型的胜率对比

	Auto-ARIMA	Prophet	因子模型
PTA	50.00%	48.48%	52.39%
PVC	46.88%	48.96%	52.82%
原油	48.93%	49.88%	51.46%
大豆	45.60%	49.12%	54.04%
天然橡胶	46.00%	48.31%	56.12%
棉花	48.04%	47.53%	50.43%
沥青	42.98%	46.49%	50.88%
焦炭	44.17%	54.17%	51.67%
焦煤	46.15%	50.35%	51.05%
玉米	48.86%	47.11%	50.26%
甲醇	51.17%	49.77%	51.64%
白银	52.69%	57.53%	50.54%
螺纹钢	44.80%	50.58%	55.20%
铁矿石	52.21%	54.87%	51.33%
铜	52.07%	46.59%	56.68%
铝	48.03%	48.32%	51.68%
锡	55.26%	55.26%	44.74%
镍	55.26%	36.84%	60.53%
食糖	50.69%	48.92%	47.53%
黄金	48.40%	51.60%	53.81%

资料来源：Wind，海通证券研究所

通过上述案例，我们可以看到，知识图谱标准化了不同品种的分析框架和预测过程，并对传统的方法有三点补充和完善。

（1）可以对知识图谱中的任一节点进行预测和分析。例如，要研究铝的库存情况，只需选取与库存有关联“关系”的所有节点作为初始变量集；

（2）可以将知识图谱中的结构信息带入量化模型中。例如，使用主成分分析降维，可以考虑把每个包含关系下的分项变化汇总为一个主成分等；

（3）可以在传统量化模型的基础上叠加行业观点。例如，我们回测过去 30 周的数据后发现，某个资源品的价格和库存正相关，但知识图谱中保存的行业研究员的观点为负相关。那么，有一种可能性是，库存和价格在近期受到同一个第三方因素的影响，这就促使我们进一步思考和挖掘。

在后续报告中，我们会进一步介绍基于知识图谱的 CTA 策略。

当然，除了上述应用，知识图谱在金融投资领域还有着极其广泛的空间。例如，通过对多路径的不一致性或“关系”冲突的检验和合并，进行业绩排雷等。

4. 总结与讨论

本文作为金融科技（Fintech）和数据挖掘研究系列的第二篇，主要介绍知识图谱的概念、构建方法以及在金融投资领域的应用。

知识图谱是一种直观显示各个实体关联信息的有效数据结构，其构建流程与数据挖掘的基本流程可以一一对应。在金融投资领域，通过知识图谱，我们可以进行产业链的挖掘，建立更为全面深入的事件驱动投资策略，构建知识图谱类因子，改善现有分析框架等。

本文中的应用仅为概述性的展示，我们会在后续的系列报告中详细阐述每一个案例，并根据已构建的知识图谱不断更新和扩展应用场景。

5. 风险提示

数据挖掘是从历史先验数据获取经验模型的方法，存在模型失效可能。

信息披露

分析师声明

冯佳睿 金融工程研究团队
余浩淼 金融工程研究团队

本人具有中国证券业协会授予的证券投资咨询执业资格，以勤勉的职业态度，独立、客观地出具本报告。本报告所采用的数据和信息均来自市场公开信息，本人不保证该等信息的准确性或完整性。分析逻辑基于作者的职业理解，清晰准确地反映了作者的研究观点，结论不受任何第三方的授意或影响，特此声明。

法律声明

本报告仅供海通证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

市场有风险，投资需谨慎。本报告所载的信息、材料及结论只提供特定客户作参考，不构成投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需要。客户应考虑本报告中的任何意见或建议是否符合其特定状况。在法律许可的情况下，海通证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

本报告仅向特定客户传送，未经海通证券研究所书面授权，本研究报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。如欲引用或转载本文内容，务必联络海通证券研究所并获得许可，并需注明出处为海通证券研究所，且不得对本文进行有悖原意的引用和删改。

根据中国证监会核发的经营证券业务许可，海通证券股份有限公司的经营经营范围包括证券投资咨询业务。

海通证券股份有限公司研究所

路 颖 所长
(021)23219403 luying@htsec.com

高道德 副所长
(021)63411586 gaodd@htsec.com

姜 超 副所长
(021)23212042 jc9001@htsec.com

邓 勇 副所长
(021)23219404 dengyong@htsec.com

荀玉根 副所长
(021)23219658 xyg6052@htsec.com

涂力磊 所长助理
(021)23219747 tll5535@htsec.com

宏观经济研究团队

姜 超(021)23212042 jc9001@htsec.com
于 博(021)23219820 yb9744@htsec.com
李金柳(021)23219885 lj11087@htsec.com
联系人
宋 潇(021)23154483 sx11788@htsec.com
陈 兴(021)23154504 cx12025@htsec.com

金融工程研究团队

高道德(021)63411586 gaodd@htsec.com
冯佳睿(021)23219732 fengjr@htsec.com
郑雅斌(021)23219395 zhengyb@htsec.com
罗 蕾(021)23219984 ll9773@htsec.com
沈泽承(021)23212067 szc9633@htsec.com
余浩淼(021)23219883 yhm9591@htsec.com
袁林青(021)23212230 ylq9619@htsec.com
姚 石(021)23219443 ys10481@htsec.com
吕丽颖(021)23219745 lly10892@htsec.com
周一洋(021)23219774 zyy10866@htsec.com
联系人
张振岗(021)23154386 zzg11641@htsec.com
颜 伟(021)23219914 yw10384@htsec.com
梁 镇(021)23219449 lz11936@htsec.com

金融产品研究团队

高道德(021)63411586 gaodd@htsec.com
倪韵婷(021)23219419 niyt@htsec.com
陈 瑶(021)23219645 chenyaoyao@htsec.com
唐洋运(021)23219004 tangyy@htsec.com
宋家骥(021)23212231 sjj9710@htsec.com
皮 灵(021)23154168 pl10382@htsec.com
徐燕红(021)23219326 xyh10763@htsec.com
谈 鑫(021)23219686 tx10771@htsec.com
王 毅(021)23219819 wy10876@htsec.com
蔡思圆(021)23219433 csy11033@htsec.com
联系人
谭实宏(021)23219445 tsh12355@htsec.com
庄梓恺(021)23219370 zzk11560@htsec.com
吴其右 wqy12576@htsec.com

固定收益研究团队

姜 超(021)23212042 jc9001@htsec.com
朱征星(021)23219981 zzx9770@htsec.com
周 霞(021)23219807 zx6701@htsec.com
姜珊珊(021)23154121 jps10296@htsec.com
杜 佳(021)23154149 dj11195@htsec.com
联系人
李 波(021)23154484 lb11789@htsec.com

策略研究团队

荀玉根(021)23219658 xyg6052@htsec.com
钟 青(010)56760096 zq10540@htsec.com
高 上(021)23154132 gs10373@htsec.com
李 影(021)23154117 ly11082@htsec.com
姚 佩(021)23154184 yp11059@htsec.com
周旭辉 zxh12382@htsec.com
张向伟(021)23154141 zwx10402@htsec.com
李姝醒(021)23219401 lxx11330@htsec.com
联系人
唐一杰(021)23219406 tyj11545@htsec.com
郑子勋(021)23219733 zzx12149@htsec.com
王一潇(021)23219400 wyx12372@htsec.com

中小市值团队

张 宇(021)23219583 zy9957@htsec.com
钮宇鸣(021)23219420 ymniu@htsec.com
孔维娜(021)23219223 kongwn@htsec.com
潘莹练(021)23154122 pyl10297@htsec.com
联系人
程碧升(021)23154171 cbs10969@htsec.com
相 姜(021)23219945 xj11211@htsec.com

政策研究团队

李明亮(021)23219434 lml@htsec.com
陈久红(021)23219393 chenjiuhong@htsec.com
吴一萍(021)23219387 wuyiping@htsec.com
朱 蕾(021)23219946 zl8316@htsec.com
周洪荣(021)23219953 zhr8381@htsec.com
王 旭(021)23219396 wx5937@htsec.com

石油化工行业

邓 勇(021)23219404 dengyong@htsec.com
朱军军(021)23154143 zjj10419@htsec.com
联系人
胡 歆(021)23154505 hx11853@htsec.com
张 璇(021)23219411 zx12361@htsec.com

医药行业

余文心(0755)82780398 ywx9461@htsec.com
郑 琴(021)23219808 zq6670@htsec.com
贺文斌(010)68067998 hwb10850@htsec.com
联系人
范国钦 02123154384 fgq12116@htsec.com
梁广楷(010)56760096 lgg12371@htsec.com
吴佳桂(010)56760092 wjs11852@htsec.com

汽车行业

王 猛(021)23154017 wm10860@htsec.com
杜 威(0755)82900463 dw11213@htsec.com
联系人
曹雅倩(021)23154145 cyq12265@htsec.com

公用事业

吴 杰(021)23154113 wj10521@htsec.com
张 磊(021)23212001 zl10996@htsec.com
戴元灿(021)23154146 dyc10422@htsec.com
联系人
傅逸帆(021)23154398 fuf11758@htsec.com

批发和零售贸易行业

汪立亭(021)23219399 wanglt@htsec.com
李宏科(021)23154125 lkh11523@htsec.com
联系人
史 岳 sy11542@htsec.com
高 瑜(021)23219415 gy12362@htsec.com
谢茂莹 xmx12344@htsec.com

互联网及传媒

郝艳辉(010)58067906 hyh11052@htsec.com
孙小雯(021)23154120 sxw10268@htsec.com
毛云聪(010)58067907 myc11153@htsec.com
联系人
陈星光(021)23219104 cxg11774@htsec.com

有色金属行业

施 毅(021)23219480 sy8486@htsec.com
联系人
陈晓航(021)23154392 cxh11840@htsec.com
甘嘉尧(021)23154394 gjy11909@htsec.com

房地产行业

涂力磊(021)23219747 tll5535@htsec.com
谢 盐(021)23219436 xiey@htsec.com
杨 凡(021)23219812 yf11127@htsec.com
金 晶(021)23154128 jj10777@htsec.com

电子行业

陈 平(021)23219646 cp9808@htsec.com
尹 岑(021)23154119 yl11569@htsec.com
谢 磊(021)23212214 xl10881@htsec.com
联系人
石 坚(010)58067942 sj11855@htsec.com

煤炭行业

李 淼(010)58067998 lm10779@htsec.com
戴元灿(021)23154146 dyc10422@htsec.com
吴 杰(021)23154113 wj10521@htsec.com
联系人
王 涛(021)23219760 wt12363@htsec.com

电力设备及新能源行业

张一弛(021)23219402 zyc9637@htsec.com
房 青(021)23219692 fangq@htsec.com
曾 彪(021)23154148 zb10242@htsec.com
徐柏乔(021)23219171 x bq6583@htsec.com
联系人
陈佳彬(021)23154513 cjb11782@htsec.com

基础化工行业

刘 威(0755)82764281 lw10053@htsec.com
刘海荣(021)23154130 lhr10342@htsec.com
张翠翠(021)23214397 zcc11726@htsec.com
孙维容(021)23219431 swr12178@htsec.com
联系人
李 智(021)23219392 lz11785@htsec.com

计算机行业

郑宏达(021)23219392 zhd10834@htsec.com
杨 林(021)23154174 yl11036@htsec.com
鲁 立(021)23154138 ll11383@htsec.com
于成龙 ycl12224@htsec.com
黄竞晶(021)23154131 hjj10361@htsec.com
联系人
洪 琳(021)23154137 hl11570@htsec.com

通信行业

朱劲松(010)50949926 zjs10213@htsec.com
余伟民(010)50949926 ywm11574@htsec.com
张 弋 01050949962 zy12258@htsec.com
张峥青(021)23219383 z zq11650@htsec.com

非银行金融行业

孙 婷(010)50949926 st9998@htsec.com
何 婷(021)23219634 ht10515@htsec.com
联系人
李芳洲(021)23154127 lfz11585@htsec.com

交通运输行业

虞 楠(021)23219382 yun@htsec.com
罗月江 (010) 56760091 lyj12399@htsec.com
联系人
李 丹(021)23154401 ld11766@htsec.com

纺织服装行业

梁 希(021)23219407 lx11040@htsec.com
联系人
盛 开(021)23154510 sk11787@htsec.com
刘 溢(021)23219748 ly12337@htsec.com

建筑建材行业

冯晨阳(021)23212081 fcy10886@htsec.com
联系人
申 浩(021)23154114 sh12219@htsec.com

机械行业

余炜超(021)23219816 swc11480@htsec.com
耿 耘(021)23219814 gy10234@htsec.com
杨 震(021)23154124 yz10334@htsec.com
沈伟杰(021)23219963 swj11496@htsec.com
周 丹 zd12213@htsec.com

钢铁行业

刘彦奇(021)23219391 liuyq@htsec.com
刘 璇(0755)82900465 lx11212@htsec.com
联系人
周慧琳(021)23154399 zhl11756@htsec.com

建筑工程行业

杜市伟(0755)82945368 dsw11227@htsec.com
张欣劼 z xj12156@htsec.com
李富华(021)23154134 lf12225@htsec.com

农林牧渔行业

丁 频(021)23219405 dingpin@htsec.com
陈雪丽(021)23219164 cxl9730@htsec.com
陈 阳(021)23212041 cy10867@htsec.com
联系人
孟亚琦 myq12354@htsec.com

食品饮料行业

闻宏伟(010)58067941 whw9587@htsec.com
成 珊(021)23212207 cs9703@htsec.com
唐 宇(021)23219389 ty11049@htsec.com

军工行业

蒋 俊(021)23154170 jj11200@htsec.com
刘 磊(010)50949922 ll11322@htsec.com
张恒昭 zhx10170@htsec.com
联系人
张宇轩(021)23154172 zyx11631@htsec.com

银行业

孙 婷(010)50949926 st9998@htsec.com
解巍巍 xww12276@htsec.com
林加力(021)23214395 lj12245@htsec.com
谭敏沂(0755)82900489 tmy10908@htsec.com

社会服务行业

汪立亭(021)23219399 wanglt@htsec.com
陈扬扬(021)23219671 cyy10636@htsec.com
许樱之 xyz11630@htsec.com

家电行业

陈子仪(021)23219244 chenzy@htsec.com
李 阳(021)23154382 ly11194@htsec.com
朱默辰(021)23154383 zmc11316@htsec.com
联系人
刘 璐(021)23214390 ll11838@htsec.com

造纸轻工行业

衣楦永(021)23212208 yzy12003@htsec.com
曾 知(021)23219810 zz9612@htsec.com
赵 洋(021)23154126 zy10340@htsec.com

研究所销售团队

深广地区销售团队

蔡铁清(0755)82775962 ctq5979@htsec.com
伏财勇(0755)23607963 fcy7498@htsec.com
辜丽娟(0755)83253022 gulj@htsec.com
刘晶晶(0755)83255933 liujj4900@htsec.com
王雅清(0755)83254133 wyq10541@htsec.com
饶 伟(0755)82775282 rw10588@htsec.com
欧阳梦楚(0755)23617160 oymc11039@htsec.com
巩柏舍 gbh11537@htsec.com

上海地区销售团队

胡雪梅(021)23219385 huxm@htsec.com
朱 健(021)23219592 zhuj@htsec.com
李唯佳(021)23219384 liwj@htsec.com
黄 毓(021)23219410 huangyu@htsec.com
漆冠男(021)23219281 qgn10768@htsec.com
胡宇欣(021)23154192 hyx10493@htsec.com
黄 诚(021)23219397 hc10482@htsec.com
毛文英(021)23219373 mwy10474@htsec.com
马晓男 mxn11376@htsec.com
杨伟昕(021)23212268 yyx10310@htsec.com
张思宇 zsy11797@htsec.com
慈晓聪(021)23219989 cxc11643@htsec.com
王朝领 wcl11854@htsec.com
邵亚杰 23214650 syj12493@htsec.com
李 寅 021-23219691 ly12488@htsec.com

北京地区销售团队

殷怡琦(010)58067988 yyq9989@htsec.com
郭 楠 010-5806 7936 gn12384@htsec.com
张丽莹(010)58067931 zlx11191@htsec.com
杨羽莎(010)58067977 yys10962@htsec.com
杜 飞 df12021@htsec.com
张 杨(021)23219442 zy9937@htsec.com
何 嘉(010)58067929 hj12311@htsec.com
李 婕 lj12330@htsec.com
欧阳亚群 oyyq12331@htsec.com

海通证券股份有限公司研究所
地址：上海市黄浦区广东路 689 号海通证券大厦 9 楼
电话：(021) 23219000
传真：(021) 23219392
网址：www.htsec.com