

基于知识图谱的行业分析

--在商业银行授信管理上的运用

温启明

第一章

知识图谱

什么是知识图谱

- 定义:
- 知识图谱. 是结构化的语义知识库, 用于以符号形式描述物理世界中的概念及其相互关系. 其基本组成单位是“实体—关系—实体”三元组, 以及实体及其相关属性—值对, 实体间通过关系相互联结, 构成网状的知识结构.
- 通过知识图谱, 可以实现Web从网页链接向概念链接转变, 支持用户按主题而不是字符串检索, 从而真正实现语义检索. 基于知识图谱的搜索引擎, 能够以图形方式向用户反馈结构化的知识, 用户不必浏览大量网页, 就可以准确定位和深度获取知识.

知识图谱的架构

- 知识图谱的架构，包括知识图谱自身的逻辑结构以及构建知识图谱所采用的技术(体系)架构。
- 其中知识图谱的逻辑结构，从逻辑上将知识图谱划分为2个层次：数据层和模式层。
- 从构建技术上分，知识图谱可划分为3个阶段：信息抽取、知识融合以及知识加工。

知识图谱的逻辑结构

- 数据层

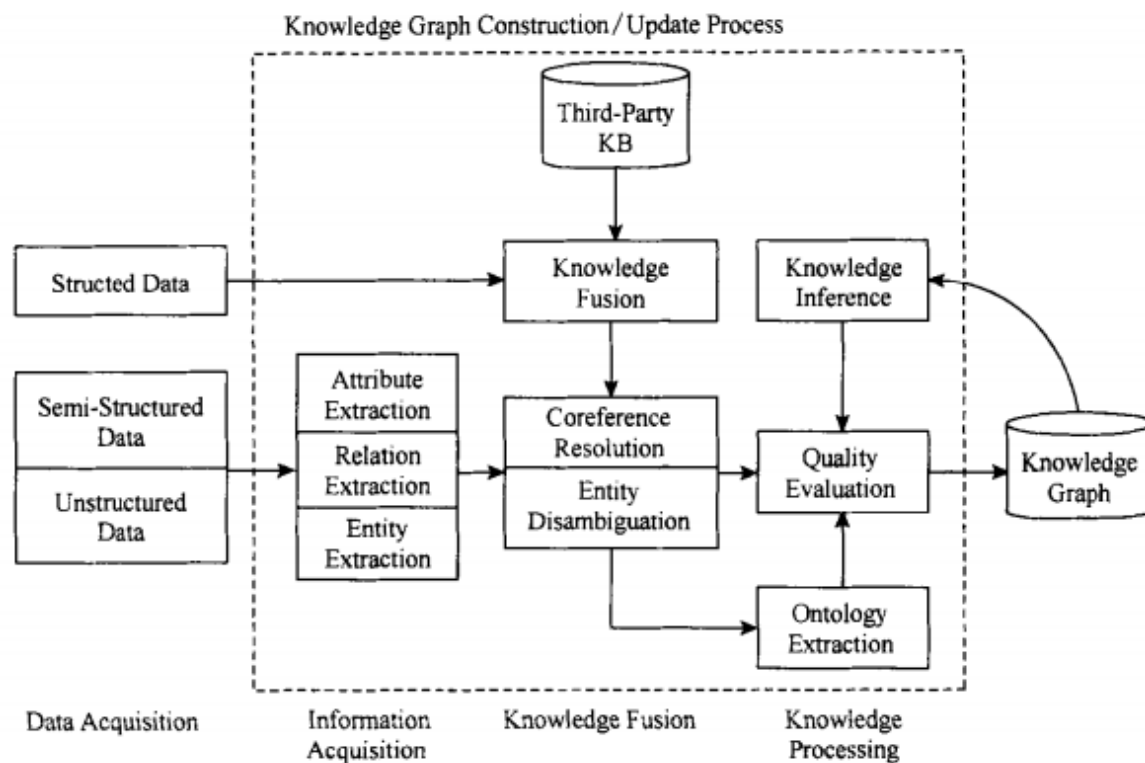
- 知识以事实(fact)为单位存储在图数据库。例如谷歌的Graphd和微软的Trinity都是典型的图数据库。如果以“实体—关系—实体”或者“实体—属性~性值”三元组作为事实的基本表达方式，则存储在图数据库中的所有数据将构成庞大的实体关系网络，形成知识的“图谱”。

- 模式层

- 是知识图谱的核心。在模式层存储的是经过提炼的知识，通常采用本体库来管理知识图谱的模式层，借助本体库对公理、规则和约束条件的支持能力来规范实体、关系以及实体的类型和属性等对象之间的联系。本体库在知识图谱中的地位相当于知识库的模具，拥有本体库的知识库冗余知识较少。

知识图谱的技术架构

右图给出了知识图谱技术的整体架构，其中虚线框内的部分为知识图谱的构建过程，同时也是知识图谱更新的过程。如右图所示，知识图谱的构建过程是从原始数据出发，采用一系列自动或半自动的技术手段，从原始数据中提取出知识要素(即事实)，并将其存入知识库的数据层和模式层的过程。这是一个迭代更新的过程，根据知识获取的逻辑，每一轮迭代包含3个阶段：信息抽取、知识融合以及知识加工。



构建技术-信息抽取

- 实体抽取

- 实体抽取，也称为命名实体识别(named entity recognition, NER)，是指从文本数据集中自动识别出命名实体。实体抽取的质量(准确率和召回率)对后续的知识获取效率和质量影响极大，因此是信息抽取中最为基础和关键的部分。

- 关系抽取

- 文本语料经过实体抽取，得到的是一系列离散的命名实体，为了得到语义信息，还需要从相关语料中提取出实体之间的关联关系，通过关系将实体(概念)联系起来，才能够形成网状的知识结构。研究关系抽取技术的目的，就是解决如何从文本语料中抽取实体间的关系这一基本问题。

- 属性抽取

- 属性抽取的目标是从不同信息源中采集特定实体的属性信息。例如针对某个公众人物，可以从网络公开信息中得到其昵称、生日、国籍、教育背景等信息。属性抽取技术能够从多种数据来源中汇集这些信息，实现对实体属性的完整勾画。

构建技术-知识融合

- 实体链接

- 实体链接的一般流程是：1)从文本中通过实体抽取得到实体指称项；2)进行实体消歧和共指消解，判断知识库中的同名实体与之是否代表不同的含义以及知识库中是否存在其他命名实体与之表示相同的含义；3)在确认知识库中对应的正确实体对象之后，将该实体指称项链接到知识库中对应实体。

- 知识合并

- 构建知识图谱时，可以从第三方知识库产品或已有结构化数据获取知识输入。例如，关联开放数据项目(linked open data)会定期发布其经过积累和整理的语义知识数据，其中既包括前文介绍过的通用知识库DBpedia和YAG()，也包括面向特定领域的知识库产品，如MusicBrainz和DrugBank等。

构建技术-知识加工

- 本体构建

- 数据驱动的自动化本体构建过程包含3个阶段：实体并列关系相似度计算、实体上下位关系抽取以及本体的生成』”。1)实体并列关系相似度是用于考察任意给定的2个实体在多大程度上属于同一概念分类的指标测度，相似度越高，表明这2个实体越有可能属于同一语义类别。所谓并列关系，是相对于纵向的概念隶属关系而言的。例如“中国”和“美国”作为国家名称的实体，具有较高的并列关系相似度；而“美国”和“手机”这2个实体，属于同一语义类别的可能性较低，因此具有较低的并列关系相似度。2)实体上下位关系抽取是用于确定概念之间的隶属(IsA)关系，这种关系也称为上下位关系，例如，词组(导弹，武器)构成上下位关系，其中的“导弹”为下位词，“武器”为上位词。3)本体生成阶段的主要任务是对各层次得到的概念进行聚类，并对其进行语义类的标定(为该类中的实体指定1个或多个公共上位词)。

构建技术-知识加工

- 知识推理

- 知识推理是指从知识库中已有的实体关系数据出发，经过计算机推理，建立实体间的新关联，从而拓展和丰富知识网络。知识推理是知识图谱构建的重要手段和关键环节，通过知识推理，能够从现有知识中发现新的知识。例如已知(乾隆，父亲，雍正)和(雍正，父亲，康熙)，可以得到(乾隆，祖父，康熙)或(康熙，孙子，乾隆)。知识推理的对象并不局限于实体间的关系，也可以是实体的属性值、本体的概念层次关系等。例如已知某实体的生日属性，可以通过推理得到该实体的年龄属性。根据本体库中的概念继承关系，也可以进行概念推理，例如已知(老虎，科，猫科)和(猫科，目，食肉目)，可以推出(老虎，目，食肉目)。

构建技术-知识加工

- 质量评估

- 质量评估也是知识库构建技术的重要组成部分。1)受现有技术水平的限制，采用开放域信息抽取技术得到的知识元素有可能存在错误(如实体识别错误、关系抽取错误等)，经过知识推理得到的知识的质量同样也是没有保障的，因此在将其加入知识库之前，需要有一个质量评估的过程；2)随着开放关联数据项目的推进，各子项目所产生的知识库产品间的质量差异也在增大。数据间的冲突日益增多。如何对其质量进行评估，对于全局知识图谱的构建起着重要的作用。引入质量评估的意义在于：可以对知识的可信度进行量化，通过舍弃置信度较低的知识，可以保障知识库的质量。

第二章

行业分析

行业分析概述

• 定义

- 所谓“行业分析”是指对行业所处的国民经济宏观环境、行业内微观环境以及用户进行的一种综合调查分析。通过行业分析能够准确地预测今后行业的发展趋势。而企业的发展潜力和发展状况与行业特征、行业所面临的生存环境以及行业所处的发展周期有着密切的联系。行业特征是直接决定商业银行信贷投资的重要因素之一，其作用是界定行业本身所处的发展阶段和行业在国民经济中的地位，进而进行行业间的横向比较，为最终选出正确的投资对象提供较为准确的行业背景。

• 现状

- 在国外，许多国际性大银行都把行业分析作为对贷款企业分析的基础，并在进行客户风险分析时将行业风险度作为主要的判断依据，赋予其较大的分析权重，如美林银行取值为30%、JP摩根是33.3%。

行业分析方法

- 行业分析方法包括定性分析和定量分析两大类。
- 定性分析大多是从行业的经济周期、生命周期、行业结构、经营绩效等某一角度对行业进行的分析，主要包括：(1)宏观环境和经济周期分析；(2)行业基本特性分析；(3)行业生命周期分析；(4)行业竞争结构分析；(5)财务分析；(6)行业演变及其驱动力分析；(7)行业关键成功因素分析；(8)产业链和价值链分析；(9)行业吸引力分析等。
- 定量分析基于一些假设条件建立模型，从而对行业进行量化分析。主要包括：(1)层次分析法；(2)主成分分析法；(3)因子分析法；(4)灰色关联度分析法；(5)模糊评价法；(6)多层次模糊综合评价法等。

行业类型分析

- **行业生命周期分析**

- 生命周期理论将一个行业的开始到消亡划分为四个阶段：幼稚期，成长期、成熟期和衰退期，通常，每个产业都要经历这样一个由成长到衰退的发展演变过程。

- **行业的经济周期分析**

- 经济周期一般经历繁荣或持续增长时期；然后进入衰退时期，在这一时期，经济增长放缓或停止增长；接着进入恢复时期，它是下一个经济繁荣期的前奏。不同的行业对经营周期的反映不同，在经济周期的每个阶段，都会有一些行业经营和效益好于另一些行业，根据行业对经济周期的不同反映，可将其划分为增长型(growth)、被动型(defensive)和循环型(cyclical)三个类型。

-

外部因素分析

- **技术创新因素**

- 从技术进步的表现形式上来说，可将其划分为产品品种增加型和产品质量升级型两种情况。在产品品种增加型增长模型中，新产品的引进并不会立即导致旧产品被淘汰；在产品质量升级增长模型中，技术创新是一种创造型的破坏过程（Aghion, Philippe and Howitt, Peter, 1992），上述两种模型从不同角度的侧面说明了技术创新会导致主导产业的更替，进而导致产业结构的演化。

- **政府政策的影响**

- **社会习惯的改变对社会习惯对行业发展的影响**

- **国际产业发展趋势**

需求分析

- 宏观经济环境、行业经营周期和外部因素分析，最终是要估计未来对行业产品的需求。宏观经济分析，主要是找出对行业销售有特别影响的宏观经济变量，如果能发现行业相关产品的销售和一些经济同变量的强相关关系，则是最理想的。通过按行业所处行业生命周期的位置进行分类，也可以提供预测未来需求的一个构架；而外部因素分析，可分为二类，一是稳定的外部因素，这类因素对行业的影响是比较容易预见的，另一类是易变的外部因素，这些外部因素就成为预测需求的不确定性因素，而分析外部因素对需求的影响主要判断这些不确定因素对行业产品需求的影响。通过上述分析就可以给出需求的定量预测和趋势判断。

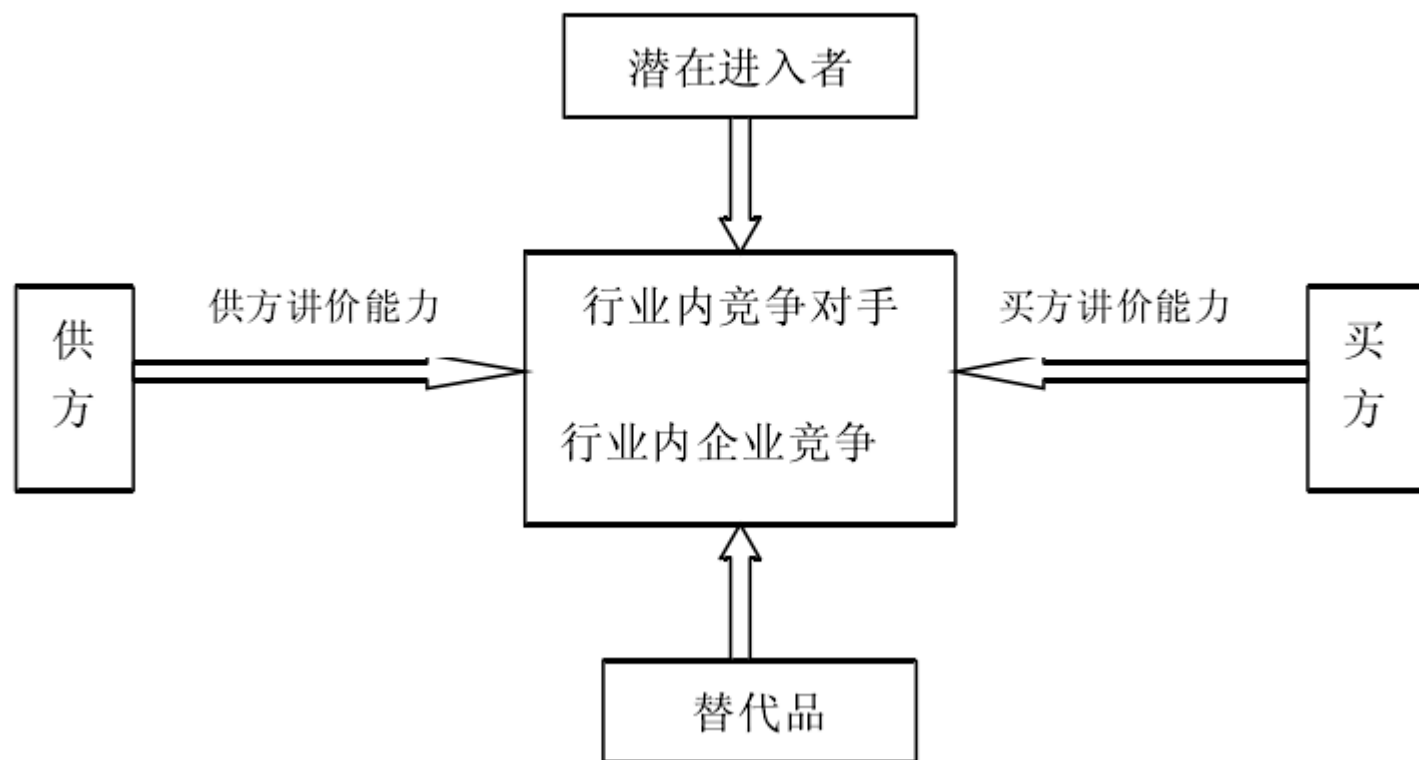
供给分析

- 供给分析主要是给出有效供给的一个基本描述。一般而言，对影响供给的因素的分析及供给的理性预测是十分困难的，这样行业分析就假定供需是平衡的，以对需求的判断来预测供给。
-

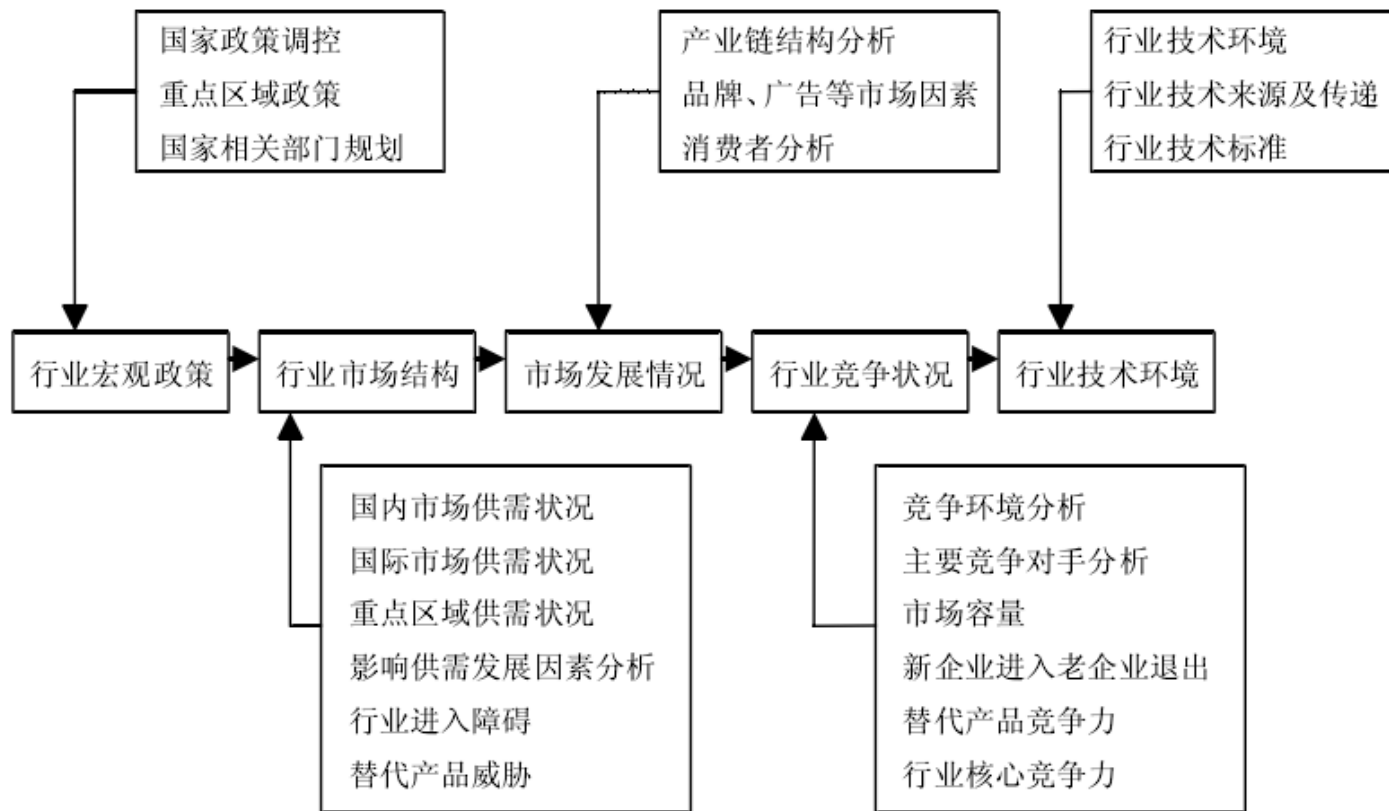
获利能力分析

- 通过供给与需求的分析和预测，就可以预测和判断行业的未来利润。一般而言，如果行业的产品供给和需求同时增长，行业的收入和利润就处于增长趋势，而如果是同时下降的则行业的收入就应处于下降趋势。一个行业要投资扩大产能或增加供给的关键是能否有获利。
-

对行业组织、行业结构进行分析



行业分析的模板



第三章

对公授信管理

中小企业信用评价体系

- 对客户进行信用等级评价的目的，一是可以根据客户信用等级状况选择适宜的授信对象；二是可以根据客户不同的信用等级，实施差别化的授信定价和采用不同的信用管理策略。由于缺乏规范的经营管理体系和健全的财务运行制度，如果以一套统一的信用评价标准来评定所有大型客户和中小企业客户的信用等级，那么显然中小企业是很难获得银行的授信支持的。亦即城市商业银行采用统一的信用评价标准，很难筛选出优良的中小企业客户，或者说会使许多潜在优质客户随同不良客户一起被淘汰掉。

中小企业信用评价指标的选择

- **1、股东评价。**鉴于中小企业的特殊性，其经济实力与其股东的经济实力一般是很难区分的，甚至更本就是混为一体的。因此，必需首先充分了解和评价企业股东的财产和经营状况。
- **2、管理者素质。**对于中小企业来说，管理者个人的素质将对企业未来的兴衰成败起关键性的作用。该项评价应包括：企业管理者的风险偏好、决策能力、创新欲望与责任心、团队建设情况等。
- **3、经营状况评价。**从较易获得并方便对比的销售收入及纳税状况指标，来评价中小企业的基本经营状况。
- **4、信用履约状况。**中小企业以往的信用记录，是决定其信用等级的重要因素。很难想象一个经常拖欠银行贷款或供应商货款的企业，会在未来的银行授信中有良好的表现。因此，必需对企业历史的各类信用履约状况进行深入的调查和评价。
- **5、经营环境评价。**相较于大企业来说，中小企业所在区域的经济环境条件、政策支持程度，对其生存和发展往往是至关重要的。
- **6、发展前景评价。**主要评价因素包括：企业所处行业排名、产品市场销售预期、产品技术含量、购销渠道的稳定性、地理位置优劣等因素。
- **7、财务状况评价。**主要通过企业的实际财务指标，来评价其债务负担能力。包括：资产运营能力、盈利能力、偿债能力和现金流量状况等。

中小企业信用评价的实施

- 1、尽职调查。城市商业银行授信人员必需对中小企业所提供的各种信息资料进行尽职调查。调查的内容应当包括两个方面，一是通过与企业主访谈、对企业进行实地调查，来核实和了解企业的各类信息；二是通过与企业利益相关者——股东、债权人、员工、供应商和顾客等以及企业所在社区有关人员和企业行为、信誉和业主个人品行等的信息。
- 2、中小企业财务报表的调整。城市商业银行应根据调查资料对企业财务报表按照其业务实际的状况进行调整，以此作为信用评级的基础。
- 3、信用评价的审查和复测。城市商业银行的授信分析与评价、授信工作尽职调查以及风险管理等工作岗位都应对中小企业客户的信用评级资料进行核实、验证，多方面强化控制、制衡与纠偏机制，以避免出现信用评级误导等问题。
- 4、进行信用评级结果的跟踪监测和复查。评级完成后，城市商业银行的评级人员要密切关注被评中小企业的情况，保持与被评价企业各方关系人的交流。

中小企业授信业务定价因子

- **1、客户信用。**授信业务银行让渡的是资金或信用的使用权。银行虽然可以通过利息等方式从让渡资金或信用中得到补偿，但却面临损失本金的风险。与本金的损失相比，利息补偿仅仅是很小的一部分。防范授信业务风险的关键是准确分析客户信用，因此客户信用是授信业务定价考虑的基本因素。
- **2、资金成本。**银行向客户提供的资金来源于资金市场或社会存款。无论哪种来源，都要支付利息，同时银行还要负担人员费用、管理费用、设备费用等。这些费用共同构成资金成本。授信业务定价一般要考虑成本因素。
- **3、盈利目标。**盈利是商业银行经营的目标之一。银行往往按照业务种类的不同确定不同的盈利目标。这一目标是银行授信业务价格的重要组成部分。
- **4、市场竞争因素。**银行自身的市场策略、竞争对手的市场战略、不同授信业务的发展阶段等都会影响到授信价格的确定。
- **5、时间因素。**通常，授信业务期限越长，银行的机会成本越大，相应的授信业务价格越高。
- **6、规模因素。**银行业务存在规模效应，业务量越大，授信业务额度越大，分摊的固定成本越低，价格相应越低。
- **7、担保因素。**担保条件是影响授信能否得到足额偿还的重要因素之一。在较好的担保条件下，银行一般愿意提供较为优惠的授信价格。
- **8、银企合作的紧密程度。**一方面银行基于与企业深度交往所获得的“软”信息实施授信决策，可以有效消除信息不对称问题，降低银行授信管理的成本；另一方面企业长期“购买”和使用银行的各种产品和服务，实际上是在不断的为银行提供盈利贡献。因此，如果企业能够和银行保持长期的紧密合作关系，银行一般也会愿意进行授信价格的减让。
- **9、选择性因素。**授信业务的时间跨度一般较长，宏观环境、市场条件、客户情况都可能发生很多变化。客户可能对原有的授信业务提出不同的需求和意见。比如是否允许改变偿还贷款本息时间、提前或推迟还款、是否允许改变还款方式等。随着给予客户选择权的不同，银行会配套不同的授信价格。

授信业务定价的基本方法

- 1、成本加成定价法。即综合考虑银行实施授信的资金成本、授信相关费用、授信的风险补偿以及银行的目标收益等因素，来确定客户授信价格。这种方法忽略了银行与客户的互动合作关系和市场竞争因素，容易造成客户的逆向选择和短期行为，不利于建立稳定的银企合作关系。
- 2、优惠利率加点定价法。即将不同的客户按照其信用状况划分为不同的等级，以对给予最优良客户的价格为优惠利率，之后根据不同的客户等级确定不同的利率点数或乘数。进而对某一等级客户的定价就是优惠利率加上该等级的点数或乘以该等级的乘数。该方法确定的授信价格较为接近市场，但是，根据客户等级确定相应的加点幅度存在一定的难度。
- 3、客户盈利分析定价法。即在全面考虑客户与银行各种业务往来的成本与收益的基础上，根据银行目标利润实施授信定价，用公式表示为： $\text{来源于某客户的总收入} = \text{为该客户提供各类服务的总成本} + \text{银行目标利润}$ 。该方法充分体现了银行“以客户为中心”的经营理念，摒弃了传统“就事论事”的授信定价思维，从银行与客户的全部业务往来关系中寻求最佳的授信价格，因此该方法最有可能得出最富有市场竞争力的授信定价。