

AI在金融科技领域的实践

高强
金融副总裁 氪信 CreditX

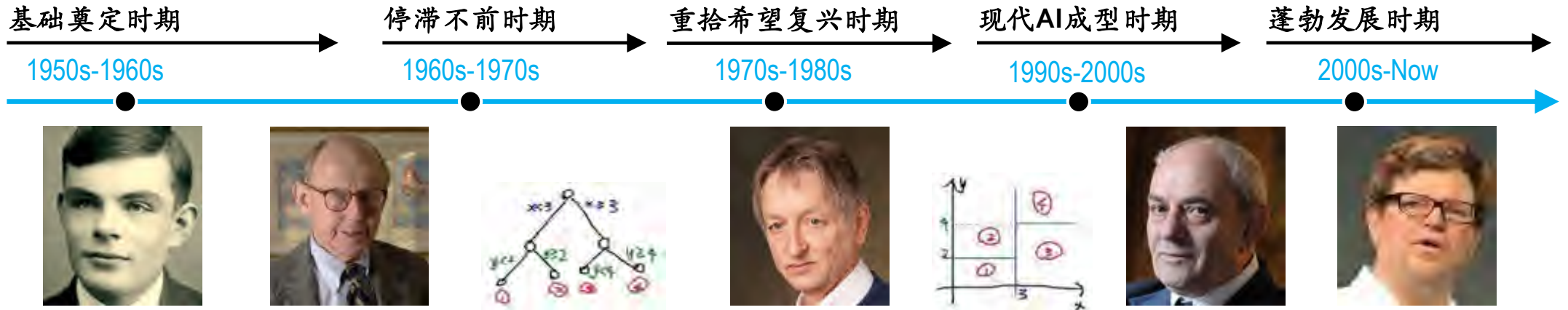
Beijing, 03/25/2017



Agenda

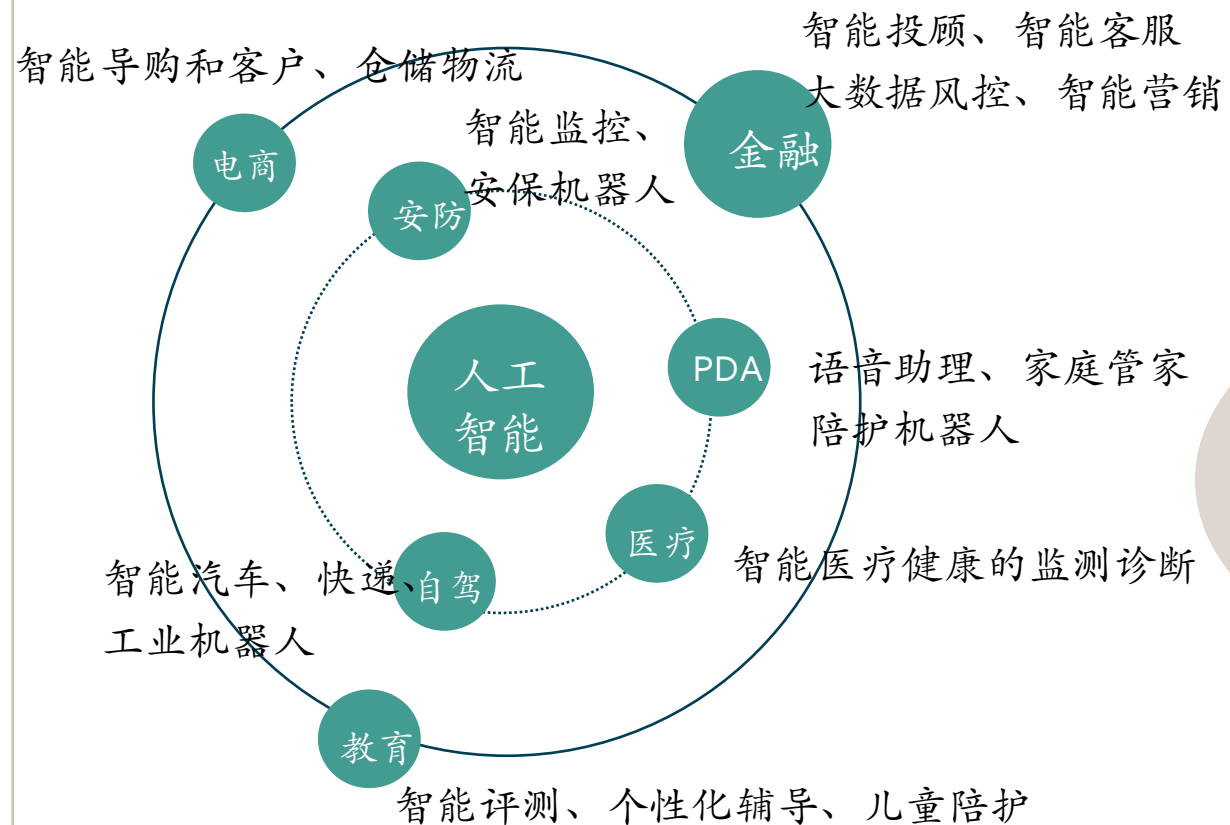
- AI+金融
- AI 驱动的金融风控解决方案
 - 大数据管理挑战 -> 知识图谱解决数据治理难题
 - 特征提取挑战 -> 深度学习挖掘弱数据的金融价值
 - 数据建模挑战 -> 集成学习框架有效整合各类风险因子
- 风控解决方案落地 -> 敏捷易集成的风控引擎系统
- 消费金融案例

人工智能的发展

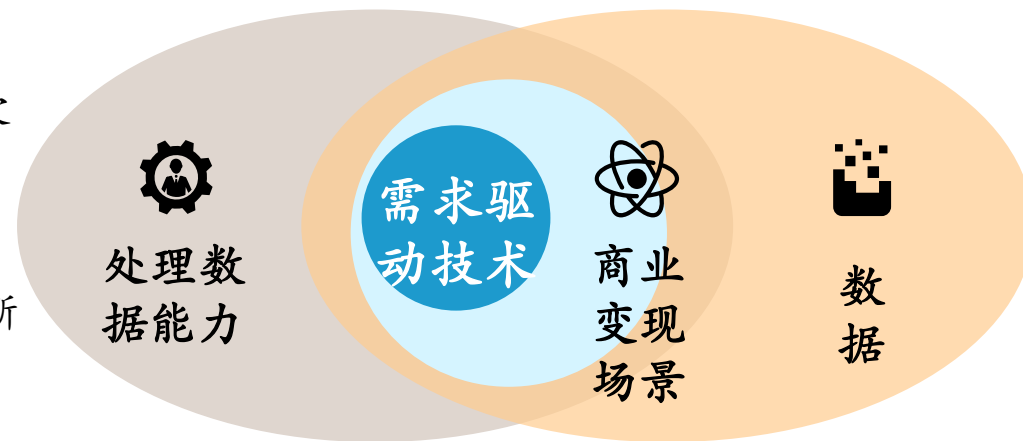


AI应用三要素

人工智能的主要应用领域



人工智能应用的三要素



金融AI的时代已经到来?

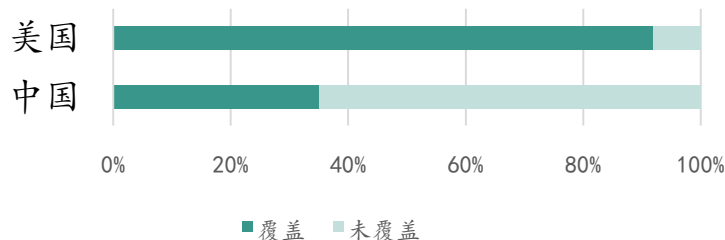
巨大的需求



市场现状

2015年中国消费信贷规模达到**19.0万亿**，同比增长23.3%。预计，未来仍将维持20%以上的增长趋势，2019年将达到**41.1万亿**

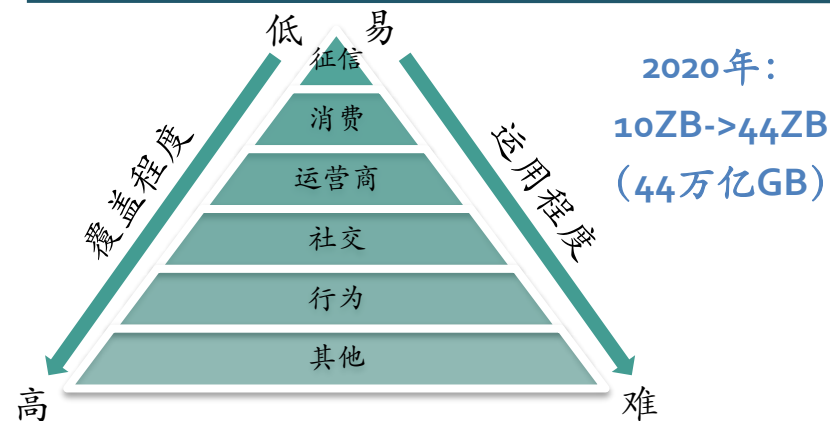
征信体系覆盖率



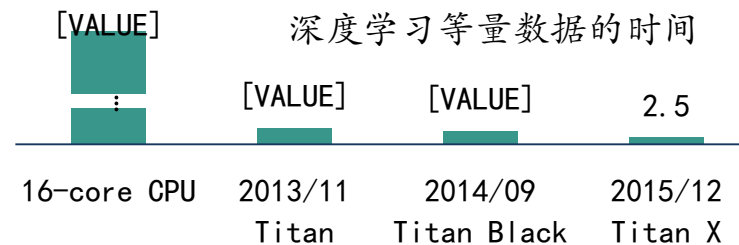
缺少信用数据群体



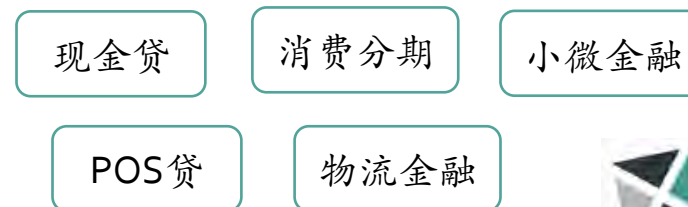
数据爆炸式增长



处理数据的能力



商业变现的场景



大数据时代金融风控之痛

风控
技术
层面

高维
困境

数据通常来自多个系统，异质异构，形式多样化；
特征工程往往会产生上千维变量，如何将其有效融合，形成 $1+1>2$ ？

价值
困境

意识到大数据的价值，但不知如何实践；
尤其是面对海量非结构化、稀疏的弱金融数据，怎样才能充分挖掘数据价值？

敏捷
困境

新金融7x24小时随时在线的服务模式和客户对准实时体验的要求，以及欺诈等风险模式在博弈中快速演化；
如何动态调整模型策略应对此类风险？

机构
战略
层面

风控决策低效耗时，人工成本高，坏账率却无法有效降低，大大限制在激烈竞争格局中的业务发展规模，也无助于取得业绩的持续增长与核心竞争力的强化

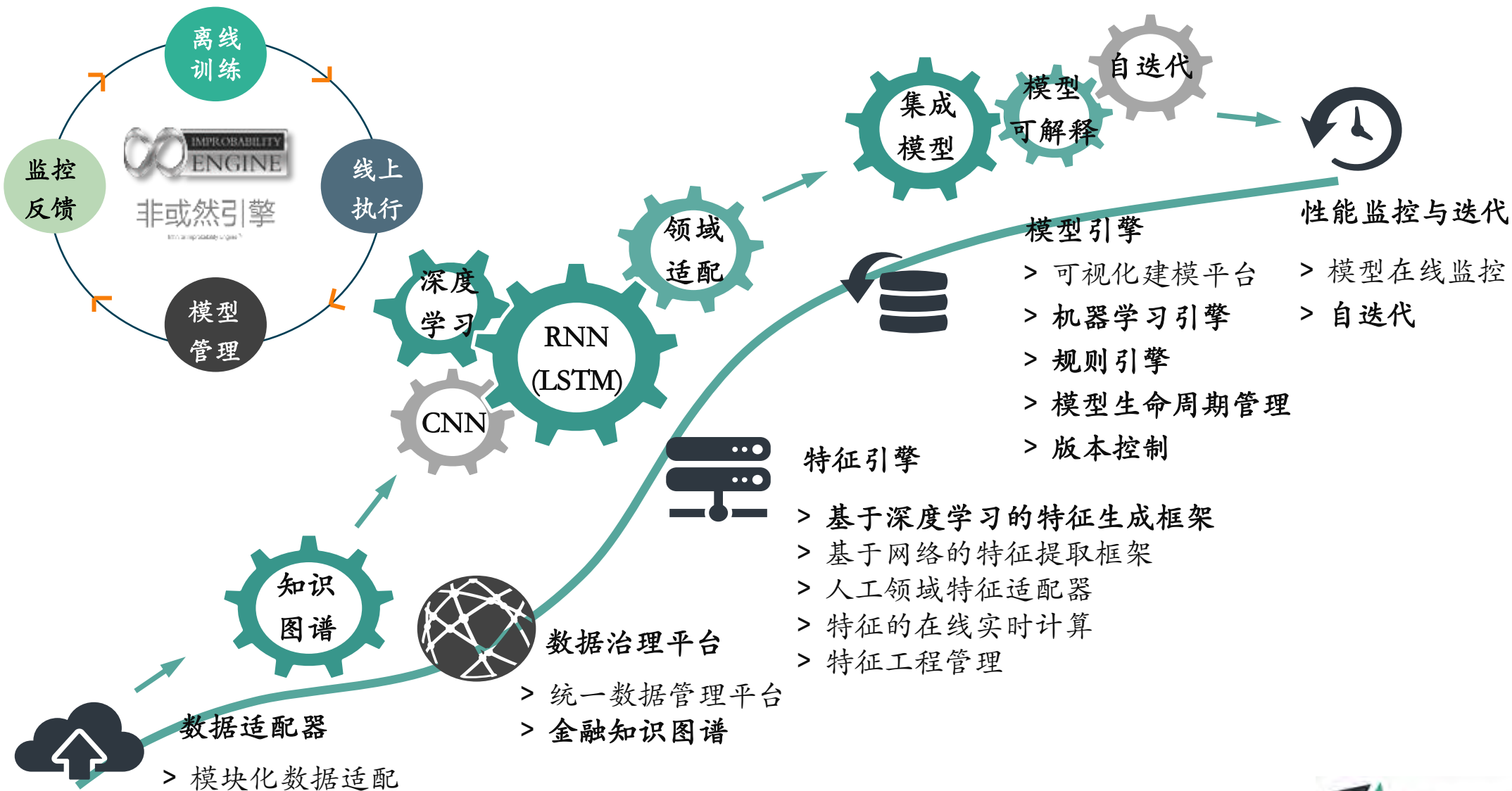
风控
业务
层面

数据
难以
整合

风险
难以
捕捉

模型
迭代
困难

AI 驱动的全流程金融风控解决方案



知识图谱：重新定义金融数据架构体系

金融风控的关键



基于纷繁复杂的各种数据，通过数据金融化和数据价值挖掘，合理建模从而进行风险程度评估

面临的挑战



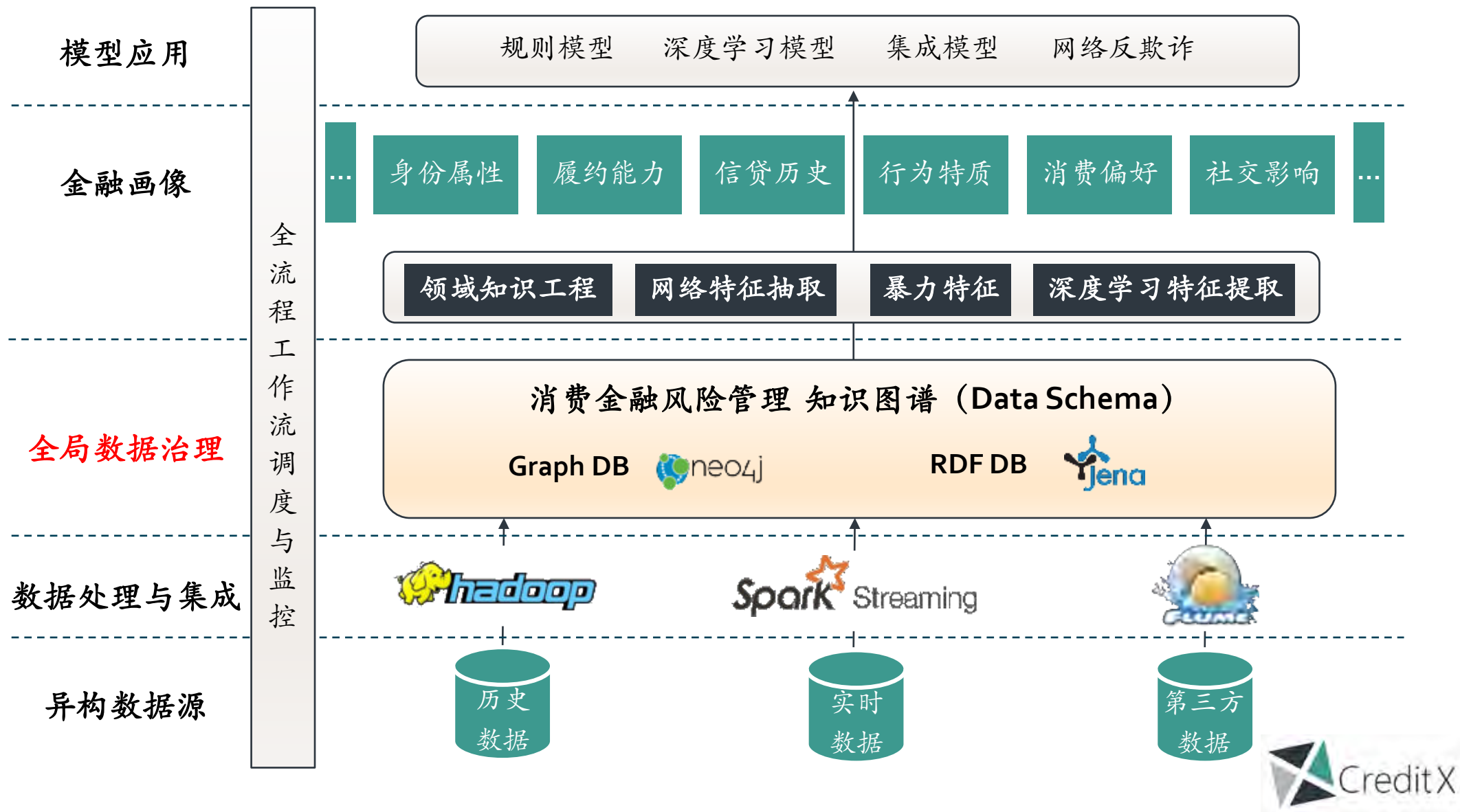
数据的种类多样性、复杂性、异质异构等特性带来数据整合、数据管理、数据使用等方面的挑战

知识图谱



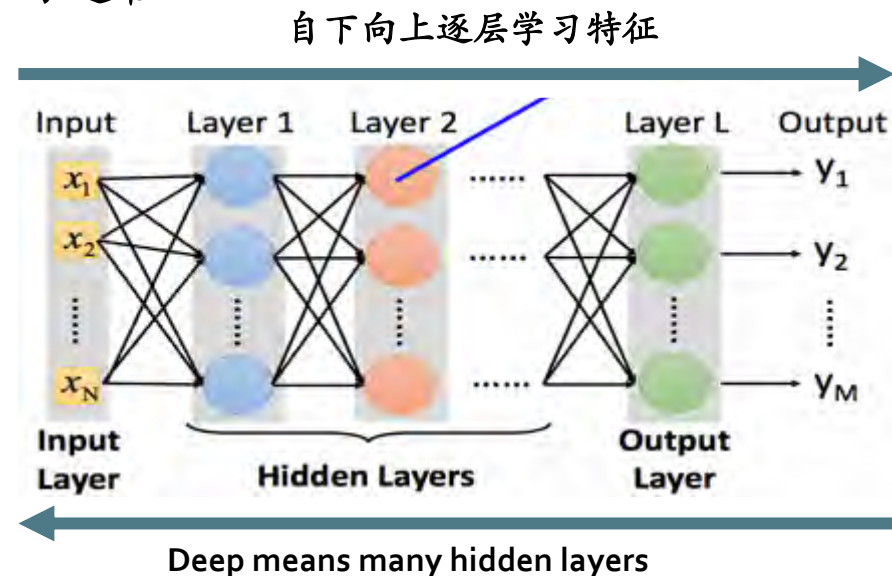
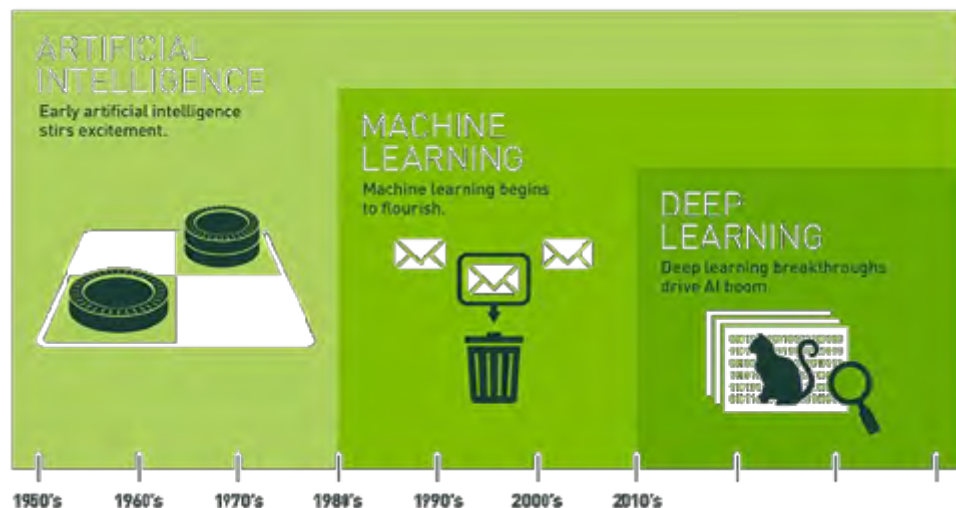
- 知识图谱本质上是语义网络，是一种基于图的数据结构，由节点和边组成。
- 在知识图谱里，每个节点表示现实世界中存在的“实体”，每条边为实体与实体之间的“关系”。知识图谱是关系的最有效的表示方式
- 知识图谱就是把所有不同种类的信息连接在一起而得到的一个关系网络，从而提供了从“关系”的角度去分析问题的能力

基于知识图谱的风险管理数据平台构建



深度学习特征提取框架：超越人工定义的深度和广度以量化金融风险

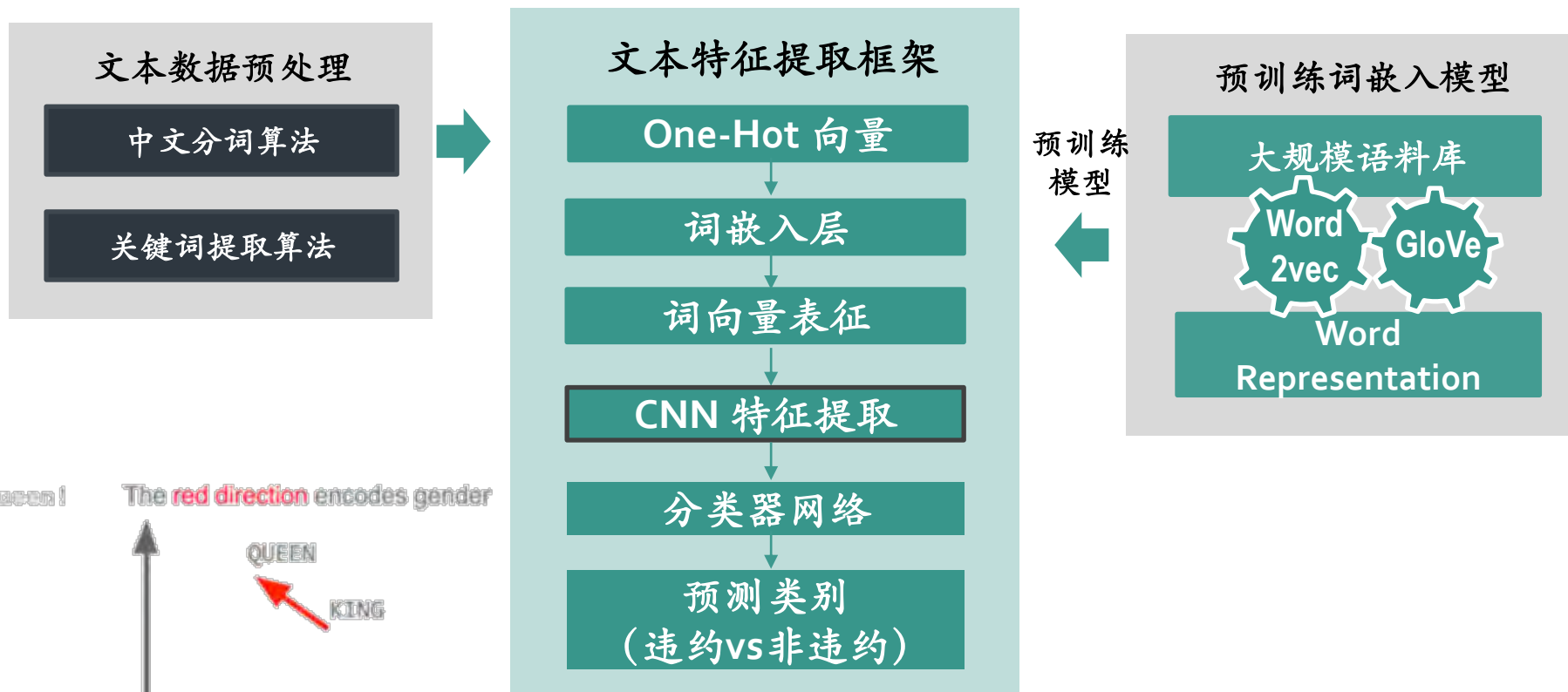
对于人工难以加工的金融非结构化数据，基于深度学习的特征生成框架自动生成特征，弥补人工定义特征的局限性；深度学习的本质是特征学习的过程



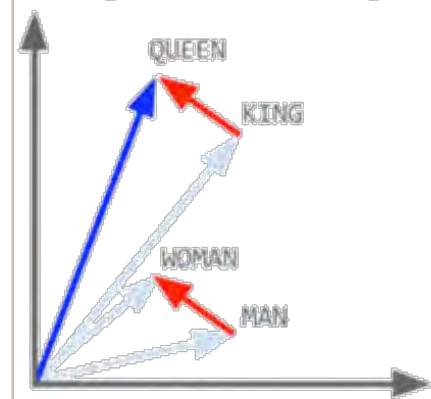
任务领域	原始输入	浅层特征 → 中层特征 → 高层特征	训练目标
语音	声强信号	频段 — 声音 — 音调 — 音素 — 单词	语音识别
图像	像素	线条 — 纹理 — 图像 — 局部 — 物体	图像识别
文本	字母	单词 — 词组 — 短语 — 句子 — 段落 — 文章	语义理解

DNN文本类数据特征提取框架用于风险预测

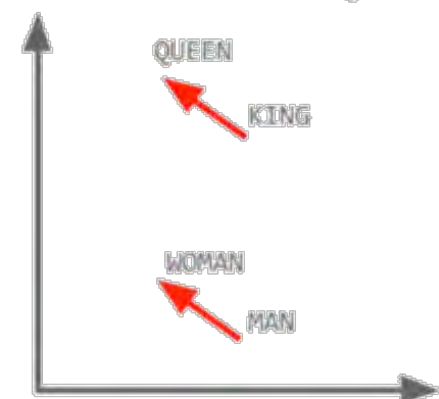
通过复杂的词向量模型将文本转化为词向量，结合卷积神经网络提取向量空间中的关系，特征抽取过程完全黑盒，自动生成抽象脱敏的特征表征



So $\text{king} + \text{man} - \text{woman} = \text{queen!}$



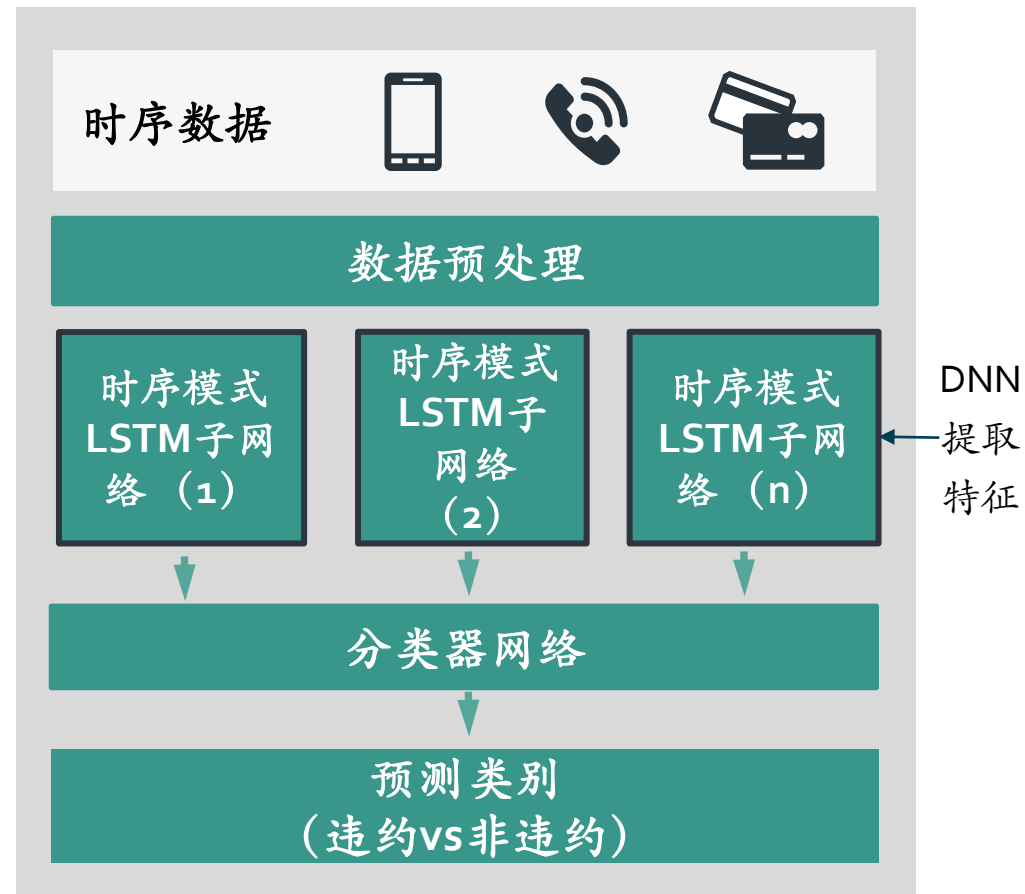
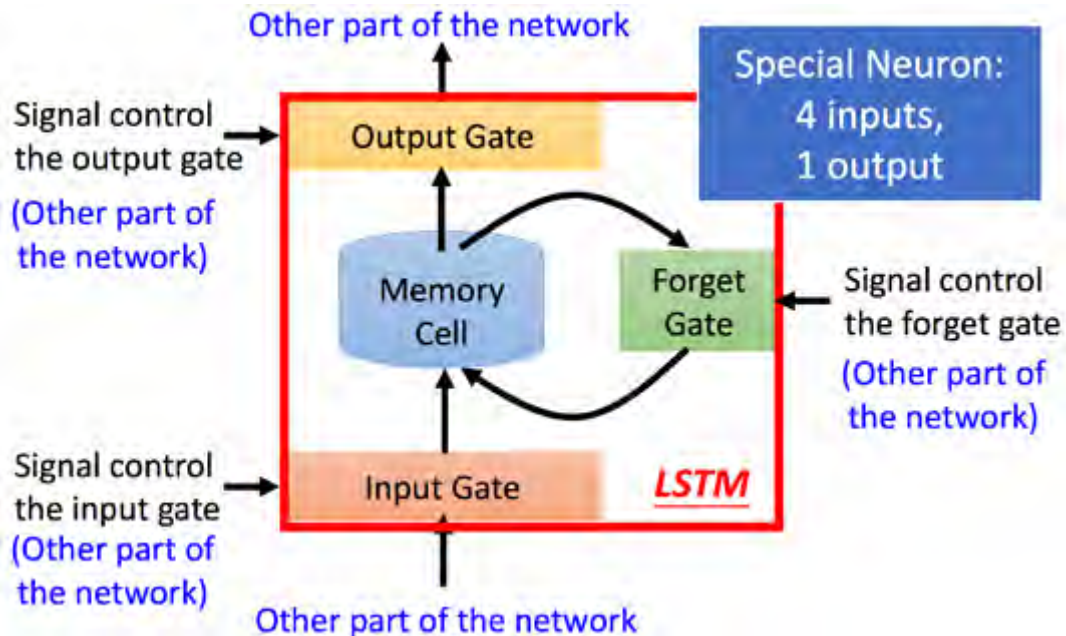
The red direction encodes gender



So $\text{KING} - \text{MAN} + \text{WOMAN} = \text{QUEEN}$

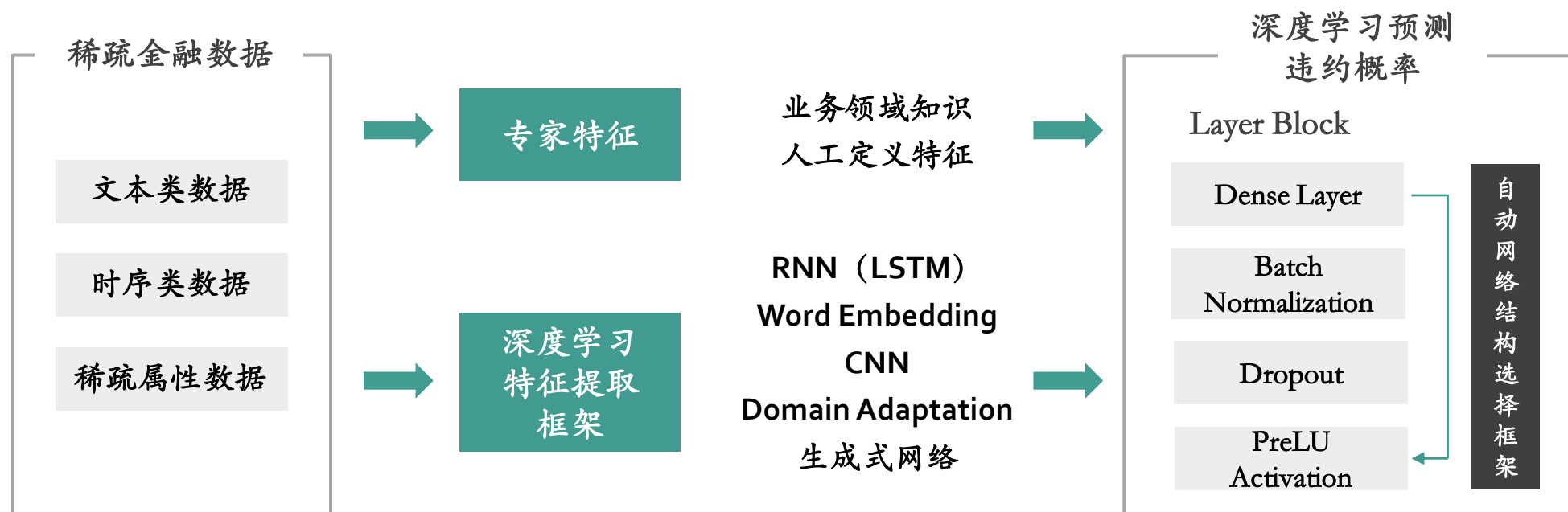
DNN时序类数据特征提取框架用于风险预测

合并学习不同周期和时序模式的循环神经网络，有效捕捉时序数据的特征隐含信息及不同时序模式下的协同影响



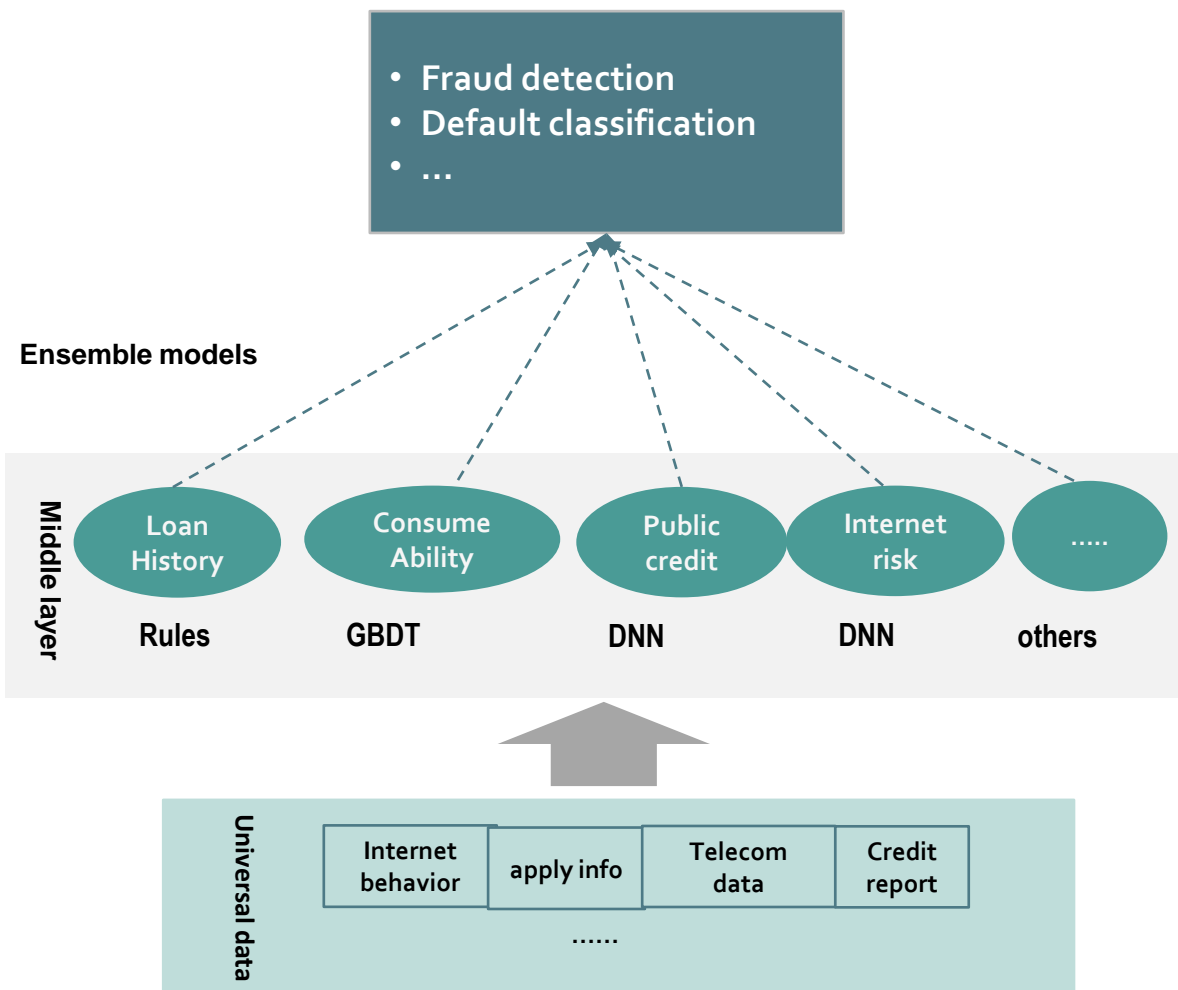
消费金融场景：融合机器和专家经验，实现全量数据价值提取

基于深度学习的特征生成框架，使用不同网络结构拟合不同的数据类型，自动从大量庞杂非结构化的数据中生成高质量的深度学习特征，并且与专家人工特征结合融入违约概率预测模型



集成学习模型框架：最大化AI与金融业务的结合深度

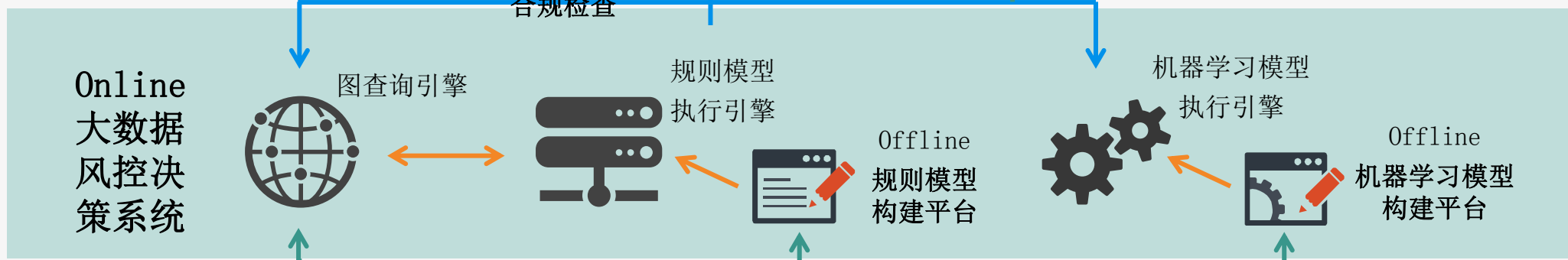
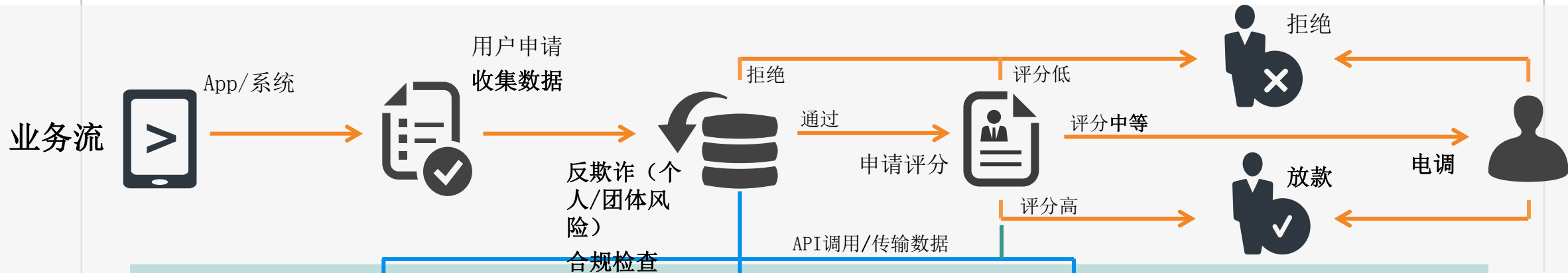
集成学习模型



为什么要使用集成学习模型？

- 不同维度/领域的数据具有不同的特点，需要使用不同的建模方法，集成学习框架可以支持不同类型模型算法作为子模型
- 单一模型预测能力有限,通过“集各家之所长”的模型效果更好
- 可以高效的对稀疏、超高维、非线性数据建模
- 具有更强的容错和抗扰动能力
- 单独领域的子模型可以快速迁移应用到新业务领域，实现快速成型和持续优化

CreditX 智能风控落地方案

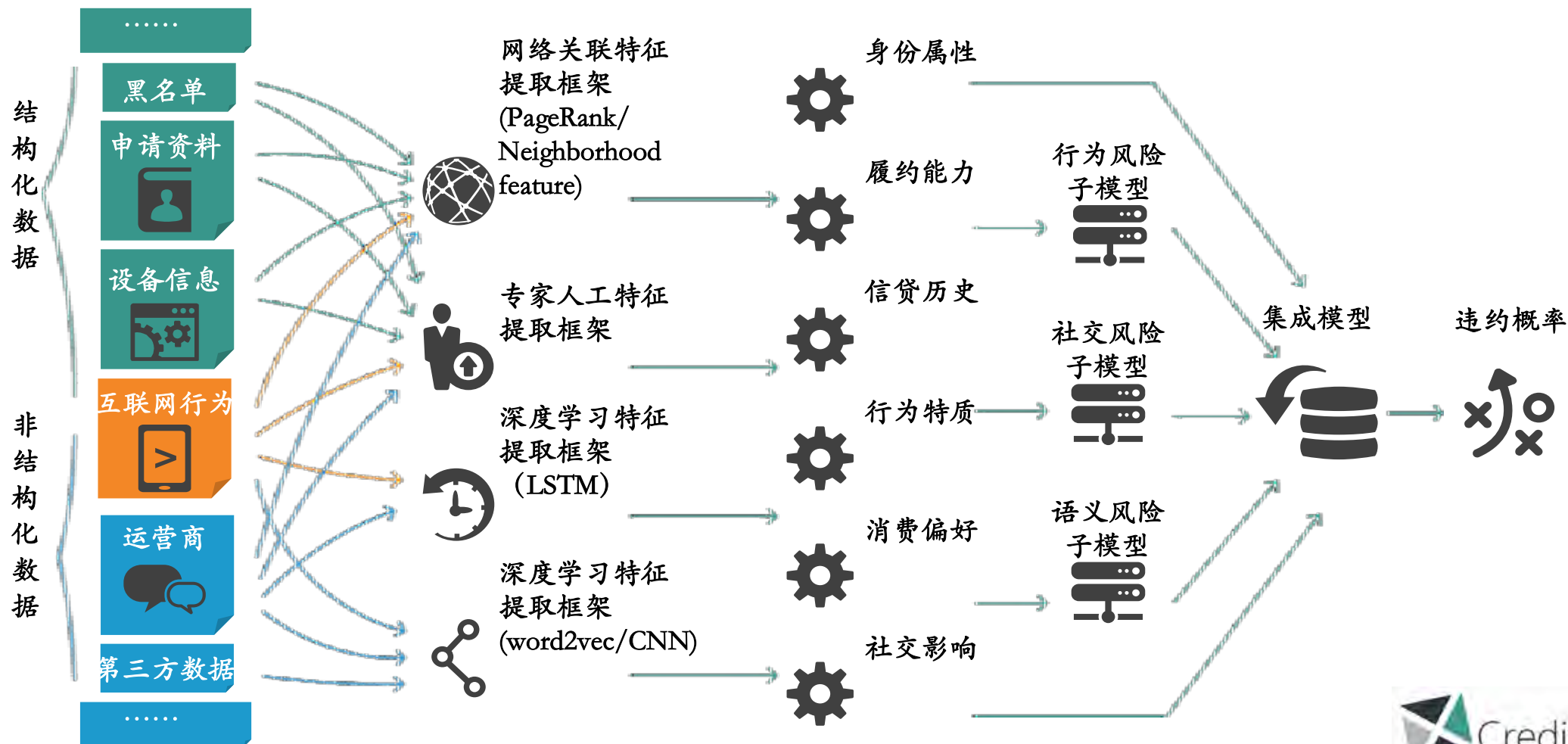


底层数据架构

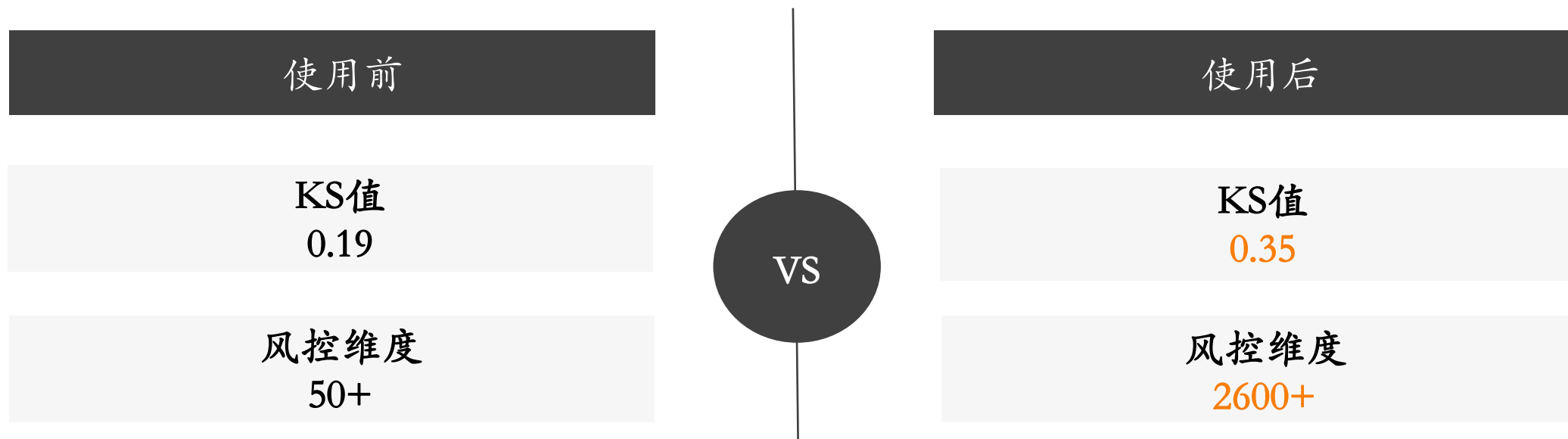


集成学习模型在消费信贷场景的应用案例

消费金融天然带有大量低饱和、稀疏、高维的数据，集成模型针对不同领域数据使用不同模型对数据进行处理和建模，产生一组“个体学习器”，并将多个学习器进行组合，泛化能力显著提升

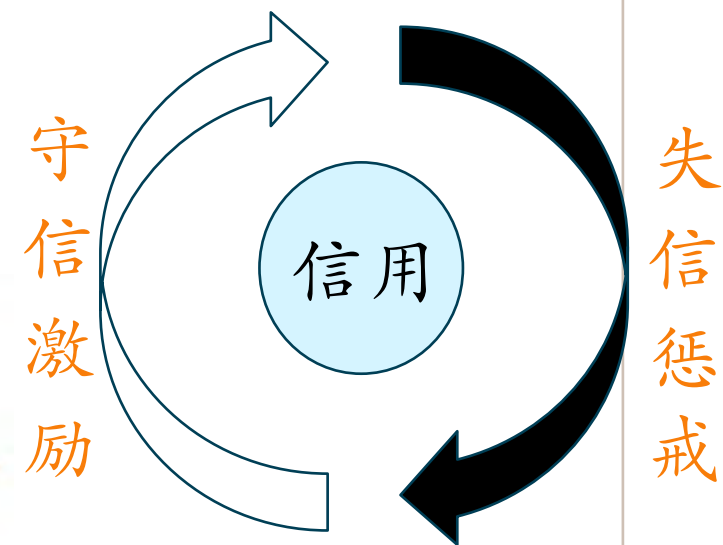
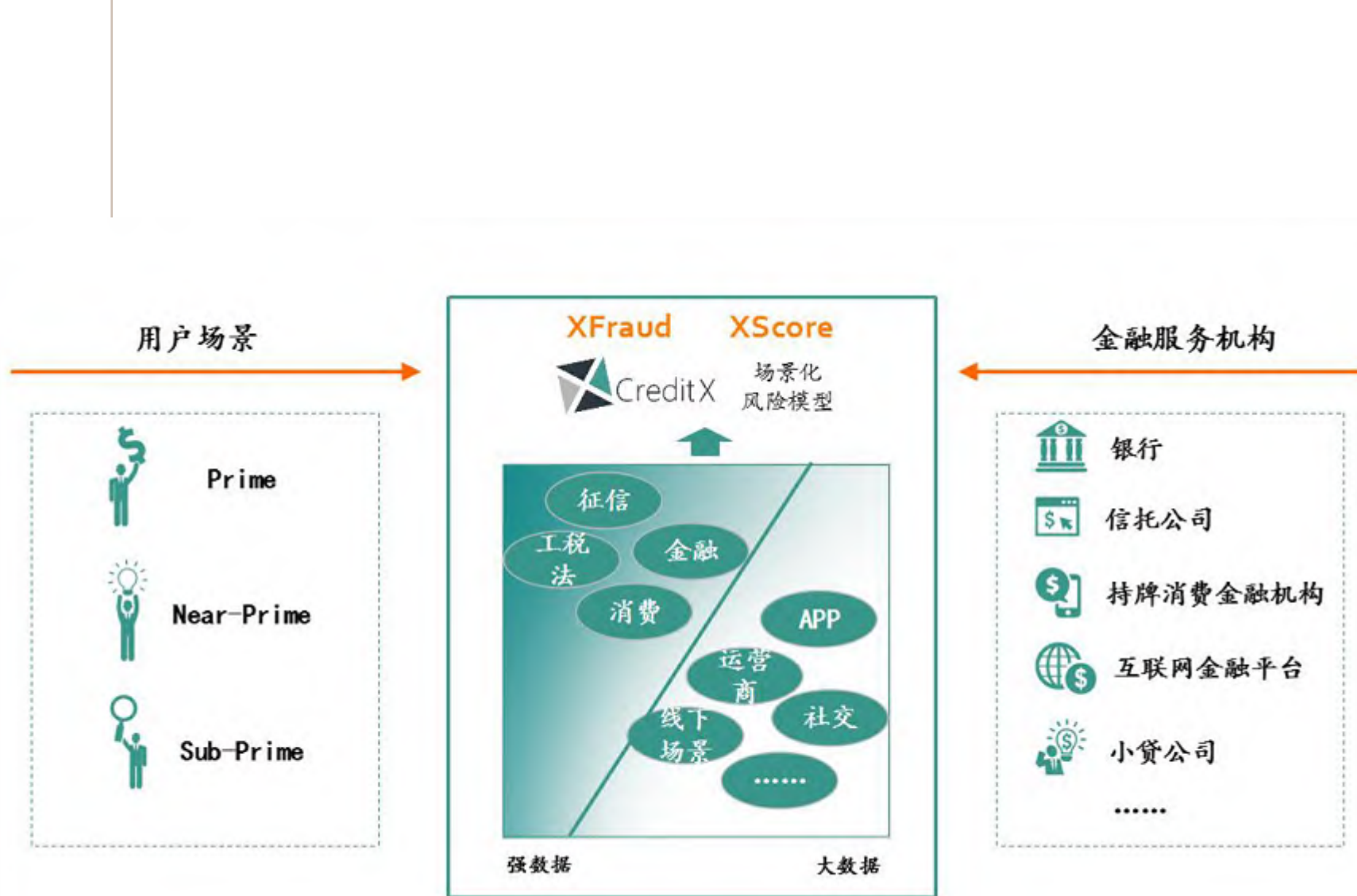


大型消费金融场景



坏账率下降 46%

大数据+AI 重塑信用价值、助力普惠金融





Thanks!

GAO QIANG
gq@creditx.com