

《金融行业大数据风控研究与实践》

-段方

某世界 **100** 强企业大数据/AI 总设计师

教授 北京大学博士后

1 大数据的概念

1.1 大数据的概念和特点

1.1.1 概念

1.1.2 特点

1.1.3 数据维度的问题

1.1.4 数据的“量化”

1.1.5 大数据与人工智能

1.2 金融大数据概念和特点

1.2.1 大数据成为金融科技的核心

1.2.2 风险控制是金融业的基础

1.2.3 大数据改写了风险控制的方法

1.3 金融大数据的内容

1.3.1 客户基础数据

1.3.2 客户交易数据

1.3.3 客户理财数据

1.3.4 客户信用卡数据

1.3.5 客户服务数据等

1.4 大数据的关键点

1.4.1 数据思维的建立

1.4.2 数据的收集和整理

1.4.3 数据应用的突破点

1.4.4 技术驱动 or 应用驱动？

1.5 数据治理基础

1.5.1 “垃圾进、垃圾出”

1.5.2 数据治理的标准——DAMA

1.5.3 数据标准的定义

1.5.4 数据质量的评估标准

1.5.5 数据治理的相关应用

1.6 为什么是大数据风控？

1.6.1 自然交给数学

1.6.2 大数据更宽、更深地洞察金融

1.6.3 风控的“数据、算法、算力”

1.6.4 深度学习算法的价值

1.6.5 5G 物联网数据的价值

1.7 大数据风控的知识图谱

1.7.1 知识图谱的概念

1.7.2 知识图谱的三要素

1.7.3 知识图谱的形成

1.7.4 风控知识图谱的应用

1.8 【案例】 PayPal 大数据风控案例

2 大数据在金融领域中的应用（可选）

2.1 大数据如何颠覆金融领域？

2.1.1 从 P2P 说金融的本质

2.1.2 银行的挑战

2.1.3 股市的挑战

2.1.4 投资理财的挑战

2.1.5 人工智能切入的第一个领域？

2.2 从客户征信开始

2.2.1 客户数据的多寡

2.2.2 客户数据的“维度”

2.2.3 FICO 的缘起

2.2.4 客户征信的意义

2.2.5 【案例】国内客户征信的案例——芝麻信用

2.3 大数据在银行的应用

2.3.1 银行的数据种

类

2.3.2 银行的客户分

析 客户视图

客户欺诈分析

客户分群

2.3.3 银行的产品分析

对公产品

对私产品

其它产品

2.3.4 银行的“困惑”

比特币会颠覆什

么？纸币的“退缩

”

银行的本质被颠覆

银行业能靠“干爹”吗？

2.4 大数据在股市的应用

2.4.1 从黑石基金说起

2.4.2 股票趋势分析

2.4.3 公司风险评估

2.4.4 期货市场可以预测吗？

2.4.5 机器人炒股的“必然”

2.5 大数据在投资领域的应用

2.5.1 天弘基金说明了什么？

2.5.2 资本游戏里大数据的价值

2.5.3 房地产可以预测吗？

2.6 大数据在保险领域的应用

2.6.1 保险的本质——计算

2.6.2 保险的数据收集

2.6.3 保险产品的测算——精算

2.6.4 保险的数据依赖

2.7 【案例】大数据在金融领域中的应用案例

3 大数据风险控制

3.1 风险的概念及范围

3.1.1 概念（大数据+算法=风控）

3.1.2 风险的定义和范围

3.1.3 市场风险

3.1.4 信用风险

3.1.5 操作风险

3.1.6 其它风险等

3.2 大数据风控的阶段

3.2.1 贷前

3.2.2 贷中

3.2.3 贷后

3.3 大数据风控内容

3.3.1 大数据预警（对公、零售）

3.3.2 反欺诈

3.3.3 大数据征信报告

3.3.4 大数据贷前准入

3.3.5 失联修复

3.4 个人风险分析

3.4.1 客户风险分类

可疑

可

信

高

危

3.4.2 客户识别内容

你是你吗？

你是谁？

你在哪里？

3.4.3 客户生理识别

人脸识别

声纹识别

客户行为鉴权

3.4.4 客户数据维度

资产

负债

消费

收入等

3.4.5 客户标签分析

消费特征

性格特征

风险特征等

3.4.6 客户征信分析

央行的征信

FICO 的意

义 大数据征

信

3.4.7 【例】附件-芝麻信用分的案例

3.4.8 Subtopic

3.5 机构的风险控制

3.5.1 基本风险分析

3.5.2 担保链分析

3.5.3 反洗钱分析等

3.6 供应链金融风险控制

3.6.1 渗透到整个供应链条

3.6.2 丰富的数据源

3.6.3 监控的内容和方法

3.7 基于团伙风险控制

3.7.1 团伙行为特征分析

3.7.2 团伙欺诈

3.8 基于大数据的反洗钱监控

3.8.1 洗钱的行为特征

3.8.2 基本原理

3.9 【案例】附件——大数据金融风控案例

3.10 【案例】工商银行的风险控制案例

4 消费金融风控举例

4.1 概述

4.1.1 从薅羊毛说起。。。

4.1.2 消费金融涉及的范围

4.1.3 Zestfinance 模式的特点

4.2 消费金融业务模式

4.2.1 消费者

4.2.2 消费供给方

4.2.3 互联网消费金融服务平台

4.2.4 资金供给方

4.3 消费者画像

4.3.1 身份信息

4.3.2 社交数据

4.3.3 认证数据

4.3.4 消费收支数据

4.3.5 行为数据

4.3.6 互金和银行黑名单

4.3.7 设备相关数据等

4.4 全过程控制

4.4.1 贷前（准入）

身份认证

反欺诈模型和策略

风险评级

风险定价模型

4.4.2 贷中（监控）

贷中复查策略

4.4.3 贷后（追

偿） 偿还能力风险

监控 偿还意愿风

险监控 **4.5 欺**

诈风险大盘

4.5.1 整体风险趋势

4.5.2 高中低风险比例统计

4.5.3 高风险地域分析

4.5.4 高风险场景分析

4.6 【案例】附件-消费金融大数据风险控制

5 金融大数据的收集和整理

5.1 大数据的内容和范围

5.1.1 企业有哪些数据？

5.1.2 数据有哪些维度？

5.1.3 大数据的成本和价值平衡

5.2 大数据的收集

5.2.1 大数据的收集方法概述

5.2.2 能否实时收集数据？

5.2.3 数据的 ETL 过程

5.2.4 数据的管理制度

5.3 大数据的质量管控

5.3.1 什么是数据质量问题？

5.3.2 60%工作量在数据质量上？

5.3.3 数据质量管控的方法

5.3.4 数据质量能“零误差”吗？

5.4 大数据的存储技术

5.4.1 从数据仓库开始

5.4.2 HADOOP 改变了什么？

5.4.3 Excel 过时了吗？

5.4.4 如何写 SQL 语句？

5.5 大数据的安全思考

5.5.1 大数据的安全隐患

5.5.2 大数据的安全管控方法

5.5.3 是否能够永远安全？

5.6 【案例】附件-大数据的收集和整理

6 大数据的分析方法

6.1 大数据分析的价值

6.1.1 从数据中挖掘价值

6.1.2 用分析去说明“一切”

6.1.3 “四两拨千斤”的意义

6.2 大数据风控的数据指标

6.2.1 精准率

6.2.2 召回率

6.2.3 PR 曲线

6.2.4 ROC 曲线

6.2.5 提升度

6.2.6 KS 值等

6.3 大数据分析的方法

6.3.1 数据分析的形式

6.3.2 数据分析的方法

6.3.3 数据分析的难点

6.4 OLAP 分析方法

6.4.1 多维报表开始

6.4.2 多维立方体

6.4.3 【案例】 OLAP 分析工具举例

6.5 数据挖掘方法

6.5.1 什么是数据挖掘

6.5.2 数据挖掘的过程

6.5.3 数据挖掘的算法介绍

关联分析

聚类分析

分类分析等

6.6 人工智能方法

6.6.1 人工智能为什么“火”了？

6.6.2 人工智能的历史演进

6.6.3 人工智能的突破

6.6.4 人工智能算法介绍

深度学习

BP 神经网络

多层神经网络

6.6.5 ALPHAGO 改变了什么？

6.6.6 【案例】人工智能的应用案例

6.7 大数据的可视化

6.7.1 大数据为何要可视化

6.7.2 大数据的“通俗易懂”

6.7.3 大数据的可视化工具

6.8 分析报告的编写

6.8.1 如何设计报告提纲？

6.8.2 报告的故事线设置

6.8.3 报告的数据论据组织

6.8.4 报告的表现形式

6.8.5 报告的震撼性和吸引力

6.9 【案例】附件-某企业数据挖掘的应用案例汇编

7 机器学习算法基础

7.1 多维分析方法

7.1.1 OLAP 分析

7.1.2 上钻和下钻

7.1.3 用 **OLAP** 分析问题

7.2 分析算法

7.2.1 回归算法

线性回归

逻辑回归

7.2.2 决策树算法

C4.5 算

法 **CART**

算法

7.2.3 贝叶斯算法

朴素贝叶斯算法

BBN (Bayesian Belief Network) 算法

7.2.4 基于核的算法

支持向量机 **SVM** 算法

线性判别分析(**Linear Discriminate Analysis** , **LDA**)

7.2.5 聚类算法

K-MEANS 算法

期望最大化算法(**Expectation Maximization** , **ME**)

7.2.6 关联规则算法

Apriori 算法

7.2.7 降低维度算法

主成份分析(**Principle Component Analysis , PCA**)算法

偏最小二乘回归(**Partial Least Square Regression , PLS**)
算法

7.2.8 集成算法

随机森林算法

梯度推进机

7.3 【案例】 附件-机器学习方法在金融行业应用举例

7.3.1 社会关系网络的模型

7.3.2 逻辑回归算法建立评分卡

8 深度学习算法基础

8.1 神经网络算法的演进历史

8.1.1 传统的神经网络算法

8.1.2 神经网络算法的特点

8.1.3 神经网络算法的瓶颈点

8.2 神经网络的算法原理

8.2.1 神经元的原理

8.2.2 神经网络的原理

8.2.3 神经网络的 BP 算法

8.3 多层神经网络算法

8.3.1 单层神经网络

8.3.2 两层神经网络

8.3.3 多层神经网络——深度学习

8.4 深度学习算法

8.4.1 卷积神经网络 CNN

8.4.2 循环神经网络 RNN

8.4.3 深度神经网络 DNN

8.5 【案例】深度学习的应用案例

8.5.1 深度学习用于图像识别

8.5.2 深度学习用于语音识别

8.5.3 深度学习用于语言翻译

9 TensorFlow 等框架

9.1 人工智能的学习框架

9.1.1 开源的概念

9.1.2 学习框架的历程

9.1.3 caffe 等学习框架介绍

9.2 TensorFlow 框架内容

9.2.1 编程模型简介

9.2.2 自编码器介绍

9.2.3 多层感知机介绍

9.3 TensorFlow 开发工具

9.3.1 jupyter

9.3.2 公开的数据集

9.3.3 语音识别等案例

9.3.4 【例】某企业的 AI 平台架构介绍

9.4 TensorFlow 支持的算法

9.4.1 支持卷积神经网络算法

9.4.2 支持循环神经网络算法

9.4.3 支持深度强化学习算法

9.5 TensorFlow 的应用案例

9.5.1 Python 语言的编程

9.5.2 Python 实现 TensorFlow

9.5.3 具体的案例

9.6 【案例】附件-基于 TensorFlow 开源框架风险分析案例

10 大数据风控具体案例剖析

10.1 大数据风控的特点

10.1.1 全量风控——从抽样数据到全量数据

10.1.2 全覆盖风控——覆盖企业经营的每个毛

孔 10.1.3 风控提供数据的准确性——多维度

验证

10.2 风控大数据应用特点

10.2.1 更多的数据、更细的观察力

10.2.2 物联网数据的贷后跟踪分析

10.2.3 不同维度的数据交叉检验

10.2.4 突出数据异常点识别

10.2.5 依赖计算能力和数据范围

10.3 大数据风控的过程

10.3.1 收集收

集 10.3.2 数

据清洗 10.3.3

数 据 建 模

10.3.4 算法选

择 10.3.5 算

法使用 10.3.6

算法调优

10.4 基于风控内容的大数据分析

10.4.1 经营风

控 贷款企业的风

控

生产效益评估

生产效率评估

贷款用途评估

技术工艺风控

产品故障率的分析

工作流程评估

资源利用风控

水电等费用对比分析

人力效率评估

成本风控

成本组成分析

营销成本分析

存货资金风控

物资的库存分析

货物废品率分析

产品销售风控

广告效果量化评估

产品销售分析

10.4.2 管理风控

管理职能风控

风控各个职能部门的工作

如何 量化相关 **KPI**

如何监督 **KPI** 执行情

况 如何设置 **KPI**

10.5 风险管控

10.5.1 贷款风

险 贷款对象分析

贷款风险分析

10.5.2 营销风

险 市场风险分析

营销人员风险分析

10.5.3 财务风

险 投资风险

10.5.4 经营风险

内部控制风险

10.5.5 人员风险

人员越权异常分析

如何防止“苍蝇犯大案”？

10.5.6 新产品开发风险

市场风险分析

营销风险分析

10.5.7 信息系统风险

漏洞扫描分析

攻击测试风险

10.6 大数据精细风控的方法

10.6.1 风控数据探针

10.6.2 风控检测规则设置

10.6.3 异常监报告警

10.7 关联性风控方法

10.7.1 关联分析

10.7.2 风控数据间的关联分

析 10.8 基于风控构建“免疫

系统 ” 10.8.1 用大数据杜

绝侥幸心理 10.8.2 数据分

析管理漏洞

10.8.3 数据分析经营风险

10.9 【案例】操作违规大数据分析案例

11 总结