普通高等教育经管类专业系列教材

金融大数据

谢文武　田　鹏　主　编
卢　锐　顾光同　董自光　庞　杰　副主编

清華大學出版社
北　京

内 容 简 介

本书基于大数据分析工具 Python，综合运用数据挖掘与算法，对大数据在商业银行信贷、保险营销、证券投资、宏观经济分析、金融科技、金融大数据监管等领域的应用进行介绍。本书具有三个显著特色：一是学科交叉特色明显，在金融分析中借助统计科学的大数据分析方法；二是理论与实践结合，将金融学理论与大数据分析相结合，有针对性地在具体场景中提出优化金融服务的解决方案；三是核心技能融入知识体系，将金融数据认知、收集、整理、处理、可视化应用、分析与咨询等核心技能融入课程建设与专业教学中。

本书可作为高等院校金融类相关专业的教材，也可作为金融行业从业人员的参考书。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。
版权所有，侵权必究。举报：010-62782989，beiqinquan@tup.tsinghua.edu.cn。

图书在版编目(CIP)数据

金融大数据 / 谢文武 , 田鹏主编 . -- 北京 : 清华大学出版社 , 2026. 10. -- (普通高等教育经管类专业系列教材). -- ISBN 978-7-302-72397-4

Ⅰ. F830.41

中国国家版本馆 CIP 数据核字第 20261X9Y08 号

责任编辑：刘金喜
封面设计：周晓亮
版式设计：妙思品位
责任校对：成凤进
责任印制：沈　露

出版发行：清华大学出版社
网　　址：https://www.tup.com.cn，https://www.wqxuetang.com
地　　址：北京清华大学学研大厦 A 座　　**邮　　编：**100084
社 总 机：010-83470000　　**邮　　购：**010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：天津淘质印艺科技发展有限公司
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张：**16.25　　**字　　数：**416 千字
版　　次：2026 年 10 月第 1 版　　**印　　次：**2026 年 10 月第 1 次印刷
定　　价：68.00 元

产品编号：106108-01

前言

大数据是信息化发展到一定阶段的产物。随着移动互联网、云计算、物联网、区块链、人工智能、量子计算等大数据相关技术的快速发展，人类进入“大数据时代”。大数据与土地、劳动、技术和资本等生产要素共同推动经济社会向前发展。金融大数据作为大数据的一个子集，最能体现数据作为资产的价值，其中蕴含着反映人们金融交易行为的信息。金融大数据是指金融机构(银行、保险公司、证券公司等)、厂商(信息服务提供商等)、个人和政府机构(行政、监管等部门)在支付结算、股票投资决策、成本定价、期货期权交易、债券投资、资金拆借、货币发行、票据贴现与再贴现等业务交易及相关行为中产生的信息。

“金融大数据”课程已成为“金融科技”“金融学”等金融类专业的核心课程，新课程急需配套优秀教材。本书由金融、信息技术、统计等专业骨干教师和企业业务专家合作编写，深入贯彻落实习近平总书记在中共中央政治局就实施国家大数据战略进行第二次集体学习时的重要讲话精神，弘扬和培育社会主义核心价值观，增强民族自信心和自豪感，提高学生职业道德与职业素养，落实立德树人根本任务。本书倡导中国伦理文化，用中国理论解读中国实践，用中国创新引导产业链发展，构建富有传统内涵、时代精神和中国特色的话语体系。

在与金融领域专家、教育专家和高校骨干教师多次研讨，并多次召集用人单位和金融类专业毕业生开展座谈会的基础上，编写组确立了编写的基本思路与原则：强调立德树人，突出职业精神、创新精神和实践能力培养；通过项目任务式教学，配套实训场景，将知识内容融入工作任务的实施过程中；实现知识和技能贯通一体，理论与实践相结合。

本书坚持金融高等教育理念和要求，体现技术驱动金融创新的导向，融入行业标准，按照实践项目进行编写，着力解决传统教材内容“学科化”及与社会需求脱节的问题。设计完整、综合性的教学案例，设置基于工作过程的项目化、模块化学习任务，采用问题导向、项目引领、任务驱动模式，讲练结合、育训结合、工学结合，发挥教材在推动教学方法改革、提升课堂教学质量方面的重要作用。本书以“职业化”为特色，以问题为导向，以国内大数据在金融行业的真实应用案例为载体，引入企业真实业务场景及业务流程，培养学习者发现实际问题、分析实际问题、解决实际问题的职业技能，培养金融和大数据复合的创新型高素质技术技能人才。

本书基于“00后”学生的学习特点，立足高校教育数智金融人才培养实际，充分考虑高校学生的学习能力和对理论知识的可接受程度，同时兼顾金融类本科人才培养的拓展要求，既可满足大数据背景下金融行业快速变革对人才培养的新需求，又可满足不同地域、不同类型院校的教学要求。

本书的特色体现在如下几个方面。

第一，学科交叉特色明显。在金融学和金融科技的指导下，借助统计科学的大数据分析方法进行数据分析，利用计算机语言编程解决问题。

第二，理论与实践结合。本书将金融学理论与大数据分析相结合，在数据基础上研究资产定价、风险管理、金融消费者分析、企业评估等问题，有针对性地在具体场景中提出优化金融服务的解决方案。

第三，核心技能融入知识体系。本书以金融大数据核心技能为抓手，将金融数据认知、收集、

整理、处理、可视化应用、分析与咨询等核心技能融入课程建设与专业教学，以金融数据的认知、处理、可视化展示及其在金融行业中的应用为设计主线，推动数据挖掘、Python语言及应用、量化投资策略分析、金融大数据产品设计等金融大数据分析与应用核心课程的建设。

本书共12章，各章内容如下：

第1章主要介绍大数据的起源、概念和内涵特征，包括大数据的提出、演化及大数据思维；

第2章主要介绍金融大数据的内涵，以及大数据在金融领域的应用现状；

第3章主要介绍大数据分析工具Python，包括Python基础，以及如何用Python实现金融大数据的采集、清洗、存储、统计分析和可视化；

第4章主要介绍数据挖掘与算法，包括数据挖掘基础、熵权法、分类算法、回归、时间序列和文本挖掘等；

第5章主要介绍大数据分析方法，包括数据采集、数据清洗、数据集成和可视化设计等内容；

第6章主要介绍大数据在商业银行中的应用，包括针对中小微企业信贷的数据收集、数据预处理、数据分析与挖掘、信贷投放方案决策报告等；

第7章主要介绍大数据在保险业的应用，包括数据收集、数据预处理、数据分析与挖掘、保险精准营销分析报告等；

第8章主要介绍大数据在证券投资中的应用，包括量化投资策略开发的流程和工具、数据收集、数据清洗、多因子量化交易分析和投资组合回测等；

第9章主要介绍大数据在宏观研究中的应用，包括文本数据收集、数据预处理、数据分析与挖掘、货币政策文本分析报告等；

第10章主要介绍大数据在金融科技中的应用，包括第三方支付风险分析、方案设计、技术实现和案例总结等；

第11章主要介绍金融大数据监管合规的内容，包括大数据下的金融合规问题、大数据与金融信息安全、大数据视野下的金融发展与监管创新；

第12章主要介绍金融大数据前景展望，包括区块链、人工智能、绿色金融等的应用前景。

本书深化校企合作，强调从实践案例到理论知识再到实践迁移应用，适合高校学生的认知水平。每章设有学习目标及课后思考，体现知识、技能和素质培养与学习并重的理念。

为便于教学，本书提供丰富的教学资源，包括PPT课件、教学大纲、授课教案、案例代码等，读者可通过扫描下方二维码下载。


PPT课件


教学大纲


授课教案


案例代码

本书由浙大城市学院、新道科技股份有限公司联合浙江农林大学、杭州师范大学、绍兴理工学院、宁波财经学院、台州学院、浙江树人学院、温州商学院等院校共同编写。在编写过程中，众多学者和教师倾注了大量心血，做出了重要贡献。本书得到“浙大城市学院教材建设项目”资助。在此，谨向所有参与编写工作的人员致以最诚挚的感谢！

本书由谢文武教授、田鹏高级副总裁担任主编。他们凭借深厚的学术造诣和丰富的教学、企业实践经验，全面统筹规划，精心组织协调，把控整体方向和质量，为本书的编撰奠定了坚

实基础。卢锐教授、顾光同教授、董自光副教授、庞杰副教授担任副主编，积极协助主编开展工作，他们在各自负责的章节中反复推敲，提出了许多宝贵的意见和建议，为内容的科学性、系统性和实用性贡献了智慧与力量。

谢文武教授完成第1章和第2章的编写，为本书开篇奠定了良好基调；黄关华老师承担第3章的编写工作，以严谨的态度和扎实的专业知识完成了任务；罗云峰老师编写第4章，内容条理清晰、论述精当；张馨丹老师编写第5章，展现了独特的视角和深入的分析；第6章由庞杰副教授编写，紧密结合学科前沿动态与实际应用，可以满足读者的学习与应用需求；张洁教授在第7章的编写中融入了丰富的实践经验和理论知识；董自光副教授完成了第8章，该章内容丰富、观点鲜明；战昱宁老师编写第9章，在内容编排和观点阐述上颇具匠心；李琳老师编写第10章，内容翔实、逻辑连贯；顾光同教授编写第11章，体现了其在相关领域的深入研究和独到见解；卢锐教授编写第12章，以其专业的素养和精准的表达为本书画上了圆满的句号。在编写过程中，新道科技股份有限公司的张春光、饶正奎等企业高管和顾问提供了很多宝贵意见和资料，新道科技股份有限公司的丁利明做了大量协同组织工作，在此深表感谢。

正是各位主编、副主编及参编人员和新道科技股份有限公司同仁的辛勤付出、团结协作与无私奉献，使本书得以高质量完成。他们的专业精神和对教育事业的热忱，将激励更多人在相关领域不断探索和前行。再次向他们表示衷心的感谢！

限于作者水平，本书难免存在疏漏和不妥之处，敬请广大读者指正。

服务邮箱：476371891@qq.com，wkservice@vip.163.com。

编者

2026年3月

目录

第 1 章 大数据概论 ····· 1

1.1 大数据的起源 ····· 1

1.2 大数据的概念 ····· 2

1.3 大数据的内涵特征 ····· 2

1.3.1 特征定义 ····· 2

1.3.2 价值探讨 ····· 3

1.3.3 未来展望 ····· 4

课后思考 ····· 5

第 2 章 大数据在金融领域的应用 ····· 6

2.1 金融大数据概述 ····· 6

2.1.1 金融大数据的内涵 ····· 6

2.1.2 金融大数据的特征 ····· 7

2.2 金融科技发展中的大数据应用 ····· 8

2.2.1 金融大数据的发展趋势 ····· 8

2.2.2 金融大数据应用面临的挑战 ····· 9

2.3 大数据在银行业的应用 ····· 10

2.3.1 大数据对商业银行的影响 ····· 10

2.3.2 大数据在银行业的主要应用场景 ··· 12

2.4 大数据在证券行业的应用 ····· 13

2.4.1 大数据对证券业的影响 ····· 13

2.4.2 大数据对证券业的作用 ····· 14

2.4.3 大数据在证券业中的具体应用 ····· 15

2.5 大数据在保险业的应用 ····· 16

2.5.1 大数据对保险行业的影响 ····· 16

2.5.2 大数据在保险行业中的具体应用 ··· 18

课后思考 ····· 19

第 3 章 大数据分析工具 Python ····· 20

3.1 Python 简介 ····· 20

3.1.1 Python 的主要特点 ····· 20

3.1.2 Python 的应用领域 ····· 21

3.2 Python 基础知识 ····· 21

3.2.1 Python 的下载与安装 ····· 21

3.2.2 注释、行和缩进与变量 ····· 25

3.2.3 数据类型 ····· 26

3.2.4 运算符 ····· 27

3.2.5 Python 语句 ····· 31

3.2.6 Python 函数 ····· 33

3.3 Python 的模块、包、库 ····· 34

3.3.1 Python 中的模块 ····· 34

3.3.2 Python 中的包 ····· 35

3.3.3 Python 中的库 ····· 37

3.4 Python 数据处理步骤 ····· 38

3.5 Python 数据可视化 ····· 39

3.6 Python 网络爬虫基础 ····· 48

3.6.1 爬虫的定义 ····· 48

3.6.2 Python 爬虫架构 ····· 48

3.6.3 urllib2 实现下载网页的三种方式 ··· 49

3.6.4 第三方库 Beautiful Soup 的安装与使用 ····· 50

课后思考 ····· 51

第 4 章 数据挖掘与算法 ····· 52

4.1 数据挖掘基础 ····· 52

4.1.1 数据挖掘的概念与内涵 ····· 52

4.1.2 数据挖掘的基本流程 ····· 53

4.2 熵权法 ····· 54

4.2.1 熵权法的定义与内涵 ····· 54

4.2.2 熵权法在金融大数据分析中的应用思路 ····· 55

4.2.3 在 Python 环境下应用熵权法的基本原理 ····· 57

4.3 分类算法……58
4.3.1 分类算法的概念……58
4.3.2 分类算法的主要类型……59
4.3.3 分类算法在金融大数据分析中的应用方向与应用优势……62
4.4 回归……63
4.4.1 回归方法的基本概念、原理和应用场景……63
4.4.2 Python 环境下金融大数据分析中回归方法的实现方式、优势和局限性……63
4.4.3 回归方法在金融大数据分析中的深度应用和未来发展趋势……65
4.5 时间序列……67
4.5.1 时间序列的概念和在金融大数据分析中的地位……67
4.5.2 时间序列模型的基本形式……68
4.5.3 时间序列模型在 Python 环境下的应用……69
4.6 文本挖掘……71
4.6.1 文本挖掘的概念……71
4.6.2 文本信息的采集和预处理……72
4.6.3 文本挖掘算法设计……75
课后思考……76

第 5 章 大数据分析方法……77
5.1 大数据分析方法概述……77
5.1.1 大数据分析的应用……78
5.1.2 技术与工具……78
5.2 数据采集……79
5.2.1 大数据来源……80
5.2.2 数据采集技术……80
5.3 数据清洗……82
5.3.1 清洗挑战……82
5.3.2 清洗策略……83
5.3.3 数据清洗步骤指南……84
5.4 数据集成……84
5.4.1 集成技术概述……85
5.4.2 集成挑战与解决方案……85
5.4.3 ETL 过程的应用……86
5.5 可视化设计……87
课后思考……89

第 6 章 商业银行大数据应用：中小微企业信贷决策……90
6.1 大数据在商业银行风控中的应用……90
6.1.1 大数据应用于商业银行风控的原理……90
6.1.2 大数据在商业银行风控中应用的优势……91
6.1.3 大数据在商业银行风控中的全流程应用……92
6.2 项目背景介绍：中小微企业信贷决策……92
6.3 数据收集—— 制作指标评估清单……94
6.4 数据预处理……94
6.4.1 数据理解……94
6.4.2 利用 Python 理解数据类型……96
6.5 数据分析与挖掘……97
6.5.1 建立评价指标……97
6.5.2 运用熵权法确定权重并计算得分……104
6.6 信贷投放方案决策报告……110
6.7 本章小结……111
课后思考……112

第 7 章 保险机构大数据应用：精准营销……113
7.1 保险营销案例背景……114
7.1.1 大数据在保险营销领域的应用……114
7.1.2 大数据精准营销在保险业中的具体应用……116
7.1.3 案例介绍……117
7.2 数据收集：调查问卷设计……118
7.2.1 调查问卷设计基本原则……118

7.2.2 调查问卷设计方法与步骤 …… 119
7.2.3 数据收集案例 …… 119
7.3 数据预处理 …… 120
7.4 数据分析与挖掘 …… 121
7.4.1 用户画像分析 …… 121
7.4.2 决策树模型的应用 …… 122
7.4.3 决策树模型的 Python 代码实现 …… 123
7.5 保险精准营销分析报告 …… 127
7.5.1 背景介绍 …… 127
7.5.2 数据来源 …… 128
7.5.3 描述分析 …… 129
7.5.4 模型建立 …… 130
7.5.5 结论与展望 …… 131
课后思考 …… 131

第 8 章 证券投资大数据应用：多因子量化交易 …… 132

8.1 量化投资的基本概念 …… 132
8.1.1 量化投资的内涵与特征 …… 132
8.1.2 量化投资与传统投资方法的区别 …… 134
8.1.3 量化投资的发展历程 …… 135
8.1.4 量化策略投资环境分析 …… 137
8.2 量化投资策略开发的流程与工具 …… 138
8.2.1 量化投资策略开发的基本流程 …… 138
8.2.2 量化投资开发常用软件 …… 139
8.3 数据收集 …… 140
8.4 数据清洗 …… 141
8.4.1 任务实战 …… 142
8.4.2 实操说明 …… 142
8.5 多因子量化交易分析 …… 143
8.5.1 多因子策略介绍 …… 143
8.5.2 多因子模型的产生背景 …… 143
8.5.3 因子类型 …… 144
8.5.4 多因子选股模型的判断方法 …… 146
8.5.5 多因子选股模型的建立过程 …… 147
8.5.6 任务实战—— 多因子量化分析 …… 147
8.6 投资组合回测 …… 156
8.6.1 量化策略的两个阶段 …… 156
8.6.2 量化策略输出阶段的三种类型 …… 156
8.6.3 量化回测的实现工具 …… 158
8.6.4 BackTrader 回测流程 …… 158
8.6.5 任务实操 …… 160
课后思考 …… 167

第 9 章 宏观研究大数据应用：货币政策文本分析 …… 168

9.1 项目导入——货币政策与大数据 …… 168
9.2 货币政策相关文本数据收集 …… 169
9.2.1 确定数据源 …… 169
9.2.2 手动收集 …… 170
9.2.3 第三方数据库 …… 173
9.2.4 自动化工具 …… 173
9.2.5 文献库建设 …… 176
9.3 数据预处理 …… 177
9.3.1 文本预处理 …… 177
9.3.2 文本分析 …… 178
9.4 数据分析与挖掘 …… 181
9.4.1 内容分析 …… 181
9.4.2 时间序列分析 …… 182
9.4.3 关联规则分析 …… 184
9.4.4 文本分类 …… 185
9.4.5 预测模型 …… 186
9.5 货币政策分析报告 …… 188
9.5.1 数据可视化 …… 188
9.5.2 报告编写 …… 189
9.5.3 政策推演 …… 190
课后思考 …… 191

第 10 章 金融科技大数据应用：第三方支付风控 …… 193

10.1 案例导入：第三方支付风控背景介绍 …… 193

10.1.1 大数据与第三方支付风控原理 …… 193
10.1.2 大数据在第三方支付风控中的应用优势 …… 194
10.2 需求分析：第三方支付风险分析 …… 195
10.3 方案设计 …… 197
10.4 技术实现：风控决策系统实践操作实例 …… 198
10.4.1 任务目标及流程 …… 198
10.4.2 详细任务流程操作 …… 198
10.5 案例总结 …… 212
课后思考 …… 213

第 11 章 金融大数据监管合规 …… 214

11.1 大数据背景下的金融合规 …… 214
11.1.1 金融合规的内涵 …… 215
11.1.2 大数据背景下金融合规的特征 … 215
11.1.3 大数据背景下金融合规的挑战 … 217
11.1.4 大数据背景下金融合规的机遇 … 219
11.2 大数据与金融信息安全 …… 221
11.2.1 大数据背景下金融信息个体层面困境 …… 221
11.2.2 大数据背景下金融信息技术层面困境 …… 222
11.2.3 大数据背景下金融信息制度层面困境 …… 223
11.2.4 大数据环境下金融信息的安全防范与措施 …… 224
11.3 大数据视野下的金融发展与监管创新 …… 225
11.3.1 数字经济时代的大数据金融战略 …… 226
11.3.2 大数据金融新场景 …… 226
11.3.3 大数据背景下供应链金融发展存在的主要问题 …… 227
11.3.4 大数据视野下的金融监管困局 … 228
11.3.5 大数据视野下的金融监管建议 … 229
11.4 典型实践案例 …… 230
11.4.1 中国农业银行基于责任链的开放银行数据保护及合规实践 …… 231
11.4.2 中国工商银行信贷模型预测服务 …… 232
11.4.3 招商银行高价值客户识别模型预测服务 …… 232
课后思考 …… 233

第 12 章 金融大数据前景展望 …… 234

12.1 金融大数据 + 区块链：应用、风险与未来趋势 …… 234
12.1.1 区块链金融的应用场景 …… 235
12.1.2 区块链金融的应用风险 …… 236
12.1.3 区块链金融安全未来趋势 …… 237
12.2 金融大数据 + 人工智能：基本服务、应用与未来趋势 …… 238
12.2.1 人工智能 + 大数据的基本发展 … 239
12.2.2 人工智能 + 大数据的具体应用 … 239
12.2.3 人工智能 + 大数据在金融领域的未来趋势 …… 240
12.3 大数据背景下的绿色金融 …… 241
12.3.1 绿色金融概述 …… 242
12.3.2 绿色金融发展的探索 …… 242
12.3.3 数字技术赋能绿色金融正成为发展趋势 …… 243
课后思考 …… 245

参考文献 …… 246

第 1 章 大数据概论

❖ 学习目标

(1) 理解大数据的起源。
(2) 理解大数据的概念。
(3) 了解大数据的核心内涵与特征。

“大数据”一词源于英文big data，其概念起源于美国。大数据最早在统计领域得到广泛应用，并在计算机通信领域引发了一场革命，随后拓展至经济、社会、科学、环境等各个领域，并成为现代国家发展战略的重要组成部分。在互联网热潮的推动下，大数据技术迅速渗透至人们生活的方方面面，吸引着人们的关注。

1.1 大数据的起源

19世纪80年代，“大数据”概念开始萌芽。1887—1890年，美国统计学家赫尔曼·霍尔瑞斯为统计1890年的人口普查数据，发明了一台电动器，将原本耗时8 年的人口普查工作缩短为1年，由此在全球范围内开启了数据处理的新纪元。1944年，卫斯理大学图书馆员弗莱蒙特·雷德预见了大数据时代的到来，他估计美国高校图书馆的藏书规模每16 年就会翻一番。1961年，德里克·普赖斯通过观察科学期刊和论文的增长规律来研究科学知识的增长，得出了“指数增长规律”，即新期刊数量的增长方式为指数型而非线性型，每15 年翻一番，每 50 年增长10倍。这些规律都表明数据知识将呈爆炸式增长，大数据时代即将来临。

在信息通信领域，数据的大规模增长与存储首先引发关注。1986年7月，哈尔·B.贝克尔在《数据通信》上发表了《用户真的能够以今天或者明天的速度吸收数据吗？》一文，预测数据记录密度将大幅增长。1997年10月，迈克尔·考克斯和大卫·埃尔斯沃思在文章《为外存模型可视化而应用控制程序请求页面调度》中较早使用了“大数据”这一术语。2001年，美国信息技术研究领域权威咨询公司Gartner 首次提出了大数据模型。同年2月，梅塔集团分析师道格·莱尼发布了题为《3D数据管理：控制数据容量、处理速度及数据种类》的研究报告，文中提及的3V 特征定义了大数据的三个维度，如今被广泛接受。此后，“大数据”这一概念在信息通信领域被普遍接受、研究和应用。

1.2 大数据的概念

*Nature*杂志于 2008 年推出的一期专刊，从互联网科技、自然与环境、网络经济和金融等多个方面介绍了海量数据带来的挑战。2008年末，一些美国著名计算机科学研究者开始认同“大数据”这一概念。业界组建起计算社区联盟(Computing Community Consortium，3C)，并发布了一份影响深远的白皮书《大数据计算：创造商务、科学和社会领域的革命性突破》。该白皮书使得大众对大数据的解读发生了显而易见的变化——从单一的数据处理机器这一角度扩展到了商业、科学、社会等各个领域，极大地丰富了“大数据”的内涵和价值，而计算社区联盟也因此被称为最早提出“大数据”概念的机构之一。2011年5月，全球知名咨询公司麦肯锡发布《大数据：创新、竞争和生产力的下一个新领域》报告，首次全方位介绍和展望大数据，具体论述了大数据的应用价值与发展前景，此后“大数据”概念逐渐被大众熟知。

究竟何为大数据？“大数据”一词可以从字面上理解为“巨大的数据量”。Manyika 等学者认为，大数据是指大小已经超出现有典型数据库获取、存储、管理和分析数据能力的数据集合。达到什么程度的数据才可以叫作大数据？目前尚未形成一个普适性的定义。一般认为，大数据的量级应该是“太字节”(Terabyte，TB，即2的40次方)。当数据规模巨大，达到某种程度时，会使数据呈现某些有价值的特性，而由于数据体量较大，这些特性无法通过传统的数据处理技术进行归纳分析，需要使用新的技术进行挖掘与分析。因此，大数据不仅指规模巨大的数据，还表示一种分析处理庞大数据的技术能力。涂子沛在其所著《大数据》一书中认为，“大数据”是指一般的软件工具难以捕捉、管理和分析的大容量数据，以“太字节”为单位。“大数据”之大，不仅在于容量大，更深层次的意义在于：人类分析和使用的数据量呈爆炸式增长，通过对海量大数据的交换、整合、挖掘和分析，可以发现新知识、创造新价值，由此带来“大知识 ”“大科技 ”“大利润”和“大发展 ”。

1.3 大数据的内涵特征

大数据作为数据，具有三个特点：一是广泛存在性，即绝大多数产品与行为均可产生数据，是否记录主要取决于技术能力与成本考量；二是非独占性，即数据可被多次使用，尤其是公开的数据可以被其他人所使用；三是多认知性，即根据使用者的不同，对同样的数据会有不同的理解和使用方式。而“大 ”为之带来的特点则是体量巨大，处理速度较快，数据类型多样，商业价值高和在线化。

1.3.1 特征定义

2001 年 2 月，梅塔集团分析师道格·莱尼发表了题为《3D 数据管理：控制数据容量、处理速度及数据种类》的研究报告，对大数据提出“3D 数据管理”的看法，即数据成长将朝三个方向发展，分别为数据即时处理的速度(velocity)、数据格式的多样化(variety)与数据量的规模(volume)，被归纳为“3V 特征 ”。之后，随着信息科技的进步，数据量的复杂程度越来越高，3V已经不足以形容新时代的大数据，因此，2012年，不仅莱尼提出调整现有的3V分析框架，而且高科技公司 IBM、国际调查机构 Gartner及IDC 等纷纷对大数据提出新的论述，在原本的速

度、多样化与规模三个特征上，增加价值性(value)和在线的(online)两个特征。

1. 数据体量巨大

一般数据库的大小在 TB 级别，而大数据的起始计量单位在PB(1PB=1024 TB) 级别，有的甚至跃升至 EB、ZB 级别，采集、存储和计算的量都非常大。百度的资料表明，其新首页导航每天需要提供的数据超过 1.5PB，这些数据如果打印出来将使用超过 5000 亿张 A4 纸。有资料证实，到目前为止，人类生产的所有印刷材料的数据量仅为 200PB。

2. 数据类型繁多

数据类型包括结构化、半结构化及非结构化数据，具体表现为图片、地理位置信息、网络日志、视频、音频等，其中个性化数据占绝大多数。多种类型的数据对数据处理能力的要求更高，已冲破了之前所限定的结构化数据的范畴。

3. 处理速度快

在数据量非常庞大的情况下，也能够做到数据的实时处理与分析，这与传统的数据挖掘技术有着本质的不同。数据体量的增大对数据的处理速度、时效性提出了更高的要求，如搜索引擎要求几分钟前的新闻能够被用户查询到，个性化推荐算法尽可能要求实时完成推荐。而大数据技术正好能满足这一需求，这也是其区别于传统数据挖掘的显著特征。

4. 价值密度低

随着互联网以及物联网的广泛应用，信息感知无处不在，大量信息的价值密度很低，即尽管数据量大，类型多，处理速度快，但真正有价值的数据却很少。以视频为例，1 小时的视频在持续不间断的监控过程中，有价值的数据可能只有一两秒。如何结合业务逻辑并通过强大的机器算法来挖掘数据价值，是大数据时代亟须解决的问题。

5. 数据是在线的

这是大数据区别于传统数据最大的特征。在互联网高速发展的背景下，数据资源不仅体量大，更重要的是表现出在线这一显著特征。数据只有在线，即数据在与产品用户或者客户产生连接的时候才有意义。如用户在使用某互联网应用时，其行为能够及时地传给数据使用方，数据使用方通过数据分析或者数据挖掘进行加工，对该应用的推送内容进行优化，将用户最想看到的内容推送给用户，就能提升用户的使用体验。

此外，业界还有人总结出大数据的其他特征，如：数据准确性(veracity)高，随着社交数据、商业交易与应用数据等新型数据源的涌现，企业越来越需要有效的信息以确保其真实性及安全性；存活性(viability)低，是指特定情况下的大数据具有很强的时效性。与传统数据服务相比，大数据服务拥有来自企业内部、外部市场及环境等不同来源的海量数据，这些数据通过传感器采集、互联网抓取等方式获取。快速发展的分布式计算及多样的数据分析模型使海量数据处理成为可能。

1.3.2 价值探讨

随着大数据时代的来临，大数据技术开始广泛应用于越来越多的领域，但只有了解大数据的价值，了解大数据究竟如何改变生活，才能更好地利用大数据。因此，需要结合时代与社会背景来具体分析大数据，理解它如何在时代变革中发挥作用。

1. 技术变革

大数据的处理与分析正成为新兴信息技术应用融合的节点，并持续推动信息产业高速发展。移动互联网、物联网、社交网络、电子商务等是新一代信息技术的应用形态，这些应用会不断地产生即时数据，成为大数据的重要来源。云计算技术则为这些海量、多样化的大数据提供存储和运算平台，并通过分析优化，将结果反馈到应用中，使其创造出巨大的经济和社会价值。大数据价值的实现依赖新技术、新产品、新服务、新业态的产生。这在硬件与集成设备领域表现为对芯片、存储性能提出更高的要求，并催生一体化数据存储处理服务器、内存计算等市场；在软件与服务领域表现为引发了数据快速处理分析、数据挖掘技术和软件产品的发展。

2. 行业变革

大数据日益成为提高企业核心竞争力的关键因素，不同行业的企业决策正在由“业务驱动”转向“数据驱动 ”。对大数据的分析可以帮助企业为消费者提供更加快速和个性化的服务；可以为商家制定精准营销的策略提供决策支持；在公共事业领域，大数据在促进经济发展、维护社会稳定等方面起着重要作用。各行各业将在大数据技术的指导下重新定义行业的未来，这将引发全行业的变革。

3. 思维变革

在大数据时代，科学研究方法将发生重大改变。抽样调查不再是社会科学研究中普遍采取的方法，而是通过实时监测研究对象在互联网上产生的海量行为数据进行挖掘分析，揭示出规律性的内容，提出研究结论和对策。由于采集、存储、分析数据能力的提高，大数据时代下我们可以收集全体数据而非随机样本。当我们掌握了海量数据时，精确性就不那么重要了，因为我们足以分析和判断事物的发展趋势。同时，我们不再关注数据之间的因果关系，而是仅仅从数据中发现相关关系，让数据自己“发声”。使用相关关系分析法能够更快、更准确地处理数据之间的关系，而且不易受偏见的影响，提高分析决策的效率。探求数据价值取决于把握数据的人，关键是人的数据思维，与其说是大数据创造了价值，不如说是大数据思维触发了新的价值增长。从哲学意义上说，大数据的价值来自“大成智慧”。每个数据来源都有一定的片面性和局限性，只有整合各类原始数据，才能体现事物的全貌。事物的本质与规律隐藏在各种原始数据之中。不同的数据能对同一个问题提供不同角度的互补信息，可以帮助人们更深入地理解相关问题。因此汇集尽量多种来源的数据是大数据分析的关键。

数据科学是数学(统计、代数、拓扑等)、计算机科学、基础科学和各种应用科学融合的结果。能不能从大数据中得到“智慧”，关键在于对多种数据源的集成和融合。发展大数据的目的就是获得协同融合的“无缝智慧”，单靠一种数据源会导致片面性。数据的开放共享是决定大数据成败的重要前提。大数据研究与应用要改变过去各部门和各学科相互分割、独立发展的传统思路，转变为依靠不同部门、不同学科的协作。

1.3.3 未来展望

大数据的未来应用前景是非常光明的。虽然目前无法准确预测大数据最终会将人类社会带往哪种形态，但只要发展的脚步还在继续，因大数据而产生的变革浪潮将会波及地球的每个角落。未来大数据应用中一个难以绕开的问题就是用户隐私问题，如被央视曝光过的罗维邓白氏和分众无线涉及侵犯用户隐私。到目前为止，中国还没有出台专门的法律法规来定义用户隐私，

需要利用其他相关法律法规来解释有关问题。但随着民众个人隐私保护意识的日益增强，在进行大数据分析时必须遵循合法合规地获取、分析及应用数据的原则。2012 年 3 月，Gartner发表了一份题为*Information Security is Becoming a Big Data Analytics Problem*的报告，表明信息安全问题正在变成一个大数据分析问题，大规模的安全数据需要被有效地关联、分析和挖掘，并预测未来将出现安全分析平台，以及部分企业在未来 5 年将设立一个新的岗位——“安全分析师”或“安全数据分析师”。

对于大数据安全分析而言，关键不在于数据本身，而在于对数据的分析方法。大数据安全分析可以借鉴通用的数据分析技术与方法，但是当具体应用到网络安全领域时，还必须考虑到安全数据自身独有的特点及安全分析的最终目标，只有这样，大数据安全分析的应用才更有价值。例如，在进行异常行为分析、恶意代码分析或 APT(高级持续性威胁)攻击分析时，应首先构建分析模型，然后再考虑如何利用大数据分析技术(如并行计算、实时计算、分布式计算)来实现该分析模型。

课 后 思 考

(1) 大数据的演变历程是怎样的？
(2) 大数据的核心内涵是什么？
(3) 大数据的特征是什么？
(4) 大数据的未来应用前景如何？

第2章 大数据在金融领域的应用

❖ 学习目标

(1) 理解金融大数据的概念。

(2) 了解金融大数据的核心内涵与特征。

20 世纪以来，信息技术在金融业中大量且广泛地使用，从而累积了体量庞大的数据和信息。金融机构中存储着海量的数据，促使金融机构须要考虑如何将这些数据转换为可以创造实际价值的成果，为企业创造尽可能多的利润。然而，这些数据并不是为了分析目的而专门生产的，仅仅是随商业活动产生的，尽管数量庞大，但难于直接产生价值，因此需要借助大数据挖掘技术进行深度挖掘，使之成为有价值的信息。随着数据收集能力逐步提高，金融企业将产生时间连续、动态变化的面板数据，其中不仅包括用户的交易数据，也包括用户的行为数据。简单的数据搜集和归并对金融企业来说不足以有效利用这些数据，只有对其进行深度挖掘，才可以发现其中的隐性信息并利用其为客户提供更加优质的金融产品和服务。如何对多源数据进行快速高效的处理？如何应对互联网金融产生的碎片化数据，快速响应需求引发的风险问题？如何充分利用数据分析、挖掘来获取更大的经济效益？这些是金融企业进行大数据分析时面临的关键问题。

2.1 金融大数据概述

金融大数据是指金融行业中产生、收集、存储、处理和利用的海量数据，包括客户数据、交易数据、市场数据及风险数据等。这些数据具有多样性(来源于不同渠道和类型)、实时性(反映最新的市场变化和客户需求)、价值性(包含丰富的信息和情报)等特点。

2.1.1 金融大数据的内涵

金融大数据可以理解为大数据中蕴含的、反映人们金融交易行为互动信息的基本数据集合，这是一种基于“信息来源于大数据”认知而得出的理论界定。金融大数据的内涵与金融大数据的构成之间存在关联。前者会在一定程度上影响后者，这主要体现在大数据分类构成及其权重变化可能导致金融运行中出现机遇、风险或危机；金融大数据的内涵并不等价于金融大数据的构成，因为前者在一定程度上和范围内会受到政府宏观调控政策及制度安排的影响，从而使人

们难以仅依据大数据构成进行决策。这进一步涉及金融大数据的外延，以及人们根据金融大数据进行决策会不会出现偏差等讨论。不过，我们在一般理论层面上讨论金融大数据的内涵时，把焦点放在其构成上，应该说抓住了问题分析的症结。

金融大数据的内涵具有极大量、多维度和完备性等特征，人们根据金融大数据进行决策时，需要有应对这些特征的新科技手段。目前，云平台是采集和存储极大量、完备性大数据的基础设施；集约化云计算是加工和处理此类数据的主要技术手段；而机器学习、物联网、区块链等人工智能技术，则是对多维度大数据进行甄别、判断和预测的主要分析工具。人类运用新科技手段对金融大数据的挖掘、搜集、整合、分类、加工和处理，存在效用函数的评估问题。该效用函数取得最大值的关键是人们不仅要能加工和处理历史数据，而且要能加工和处理现期数据和未来数据，并且能够从历史数据、现期数据和未来数据中获得准确信息。金融大数据的内涵既可以从静态状态理解，也可以从动态过程解释。显然，经济学家分析现期数据和未来数据是对金融大数据内涵的动态研究，它是我们讨论金融大数据内涵的分析基点。

2.1.2 金融大数据的特征

虽然大数据已经成为全社会热议的话题，但到目前为止，“大数据”尚无公认的统一定义。认识大数据，要把握“资源、技术、应用”三个层次。大数据是具有体量大、结构多样、时效性强等特征的数据；处理大数据需采用新型计算架构和智能算法等新技术；大数据的应用强调以新理念辅助决策、发现新的知识，更强调在线闭环的业务流程优化。因此，大数据不仅“大”，而且“新”，是新资源、新工具和新应用的综合体。数据资产成为现代金融机构区别于传统金融机构的重要生产要素之一。对数据资产的管理、运用和挖掘，已成为现代银行业加快创新、提升管理能力等业务中最重要的工作。

金融大数据行业是以数据为核心，通过大数据技术和应用，为金融机构和客户提供智能化、精准化、高效化的服务和解决方案的产业。尚普咨询集团数据显示，金融大数据行业主要包括以下几个方面：一是金融大数据基础设施，指为支撑金融大数据产生和运行所需的硬件设备(如服务器、存储设备等)、软件平台(如数据库管理系统、分布式计算框架等)和网络通信(如云计算服务、5G网络等)等；二是金融大数据技术，指用于对金融大数据进行采集、清洗、整合、分析和挖掘等操作的技术方法和工具，如人工智能(AI)、机器学习(ML)、深度学习(DL)、自然语言处理(NLP)、计算机视觉(CV)、区块链(BC)等；三是金融大数据应用，指利用金融大数据技术为不同领域和场景提供具体功能和价值的应用产品和服务，如风险管理(RM)、信用评估(CA)、精准营销(PM)、智能投顾(RA)、反欺诈(AF)、智慧监管(SR)等；四是金融大数据生态，指由各类参与者构成的协作与竞争关系网络，包括政府部门(如央行、银保监会等)、监管机构(如征信中心、支付清算机构等)、传统金融机构(如银行、证券公司等)、互联网金融平台(如支付宝、微信支付等)、科技企业(如百度、阿里巴巴等)、创新型企业(如同盾科技、趣链科技等)以及个人与企业客户等。

将大数据技术应用于金融业不仅是技术发展的需要，也是金融业提高自身盈利能力的需要。在当前“以客户为中心、以市场为导向”的激烈竞争时代，各大金融机构准备“二次转型”的改革进程中，要想提高核心竞争能力、防范业务风险、提高业务分析数据的时效性及准确性，必须懂得如何利用现代管理信息系统进行综合分析，挖掘客户的潜在价值，利用有价值的信息改进服务手段，运用数据挖掘技术实现智能化的决策支持功能。金融行业历来对数据的重视程度非常高。随着移动互联网的发展，各种金融业务和服务的多样化，金融市场的整体规模不断

扩大。对于大数据带来的主要业务价值，参加调研的金融企业人员表示，大数据的价值是他们可以根据商业分析实现更加智能的业务决策，让决策的制定更加理性化和有依据。依靠有前瞻性的决策，金融企业可以实现生产过程中资源更优化的分配，能够根据市场变化迅速做出调整，提高用户体验以及资金周转率，从而获取更高利润。

2.2 金融科技发展中的大数据应用

金融科技是指利用最新的技术手段和创新模式，为金融业提供更加高效、便捷、安全和普惠的服务。金融科技已经成为金融行业的新兴力量，正在深刻地改变着传统金融业的格局。《金融标准化“十四五”发展规划》第六条明确提出金融业数字生态建设，包括稳步推进金融科技标准建设、系统完善金融数据要素标准、健全金融信息基础设施标准、强化金融网络安全标准防护、推进金融业信息化核心技术安全可控标准建设、稳妥推进法定数字货币标准研制等。大数据在金融科技中的应用为金融机构带来了巨大的机遇。通过收集、整理和分析海量的金融数据，金融机构能够更准确地评估风险、识别市场趋势并做出相应的决策。例如，通过对历史数据和实时市场数据进行深度学习和模型建立，金融机构可以利用预测模型预测股票价格、外汇波动等。此外，大数据还可以用于客户画像和个性化推荐，在产品设计和营销方面为金融机构提供更加精准的指导。目前金融领域的大多数业务都与大数据息息相关，海量数据的处理已经成为金融科技领域发展的重中之重。得益于多种大数据计算框架的应用以及流式数据处理技术的发展，大数据技术与金融领域的结合也越来越紧密。随着大数据技术的广泛普及和发展成熟，金融大数据应用已经成为行业热点趋势。

2.2.1 金融大数据的发展趋势

金融大数据作为大数据技术与金融行业深度融合的产物，正深刻改变着金融业务的运作模式和服务方式。随着金融行业数字化转型的加速和监管科技需求的增长，金融大数据行业迎来了前所未有的发展机遇。

1. 大数据应用水平正在成为金融企业竞争力的核心要素

金融的核心就是风控，风控以数据为导向。金融机构的风控水平直接影响坏账率、营收和利润。经过长期的数字化改造，金融机构拥有大量的信息系统，并通过这些系统积累了海量的数据，但是这些数据是分散在各个系统中的，不能实现集中分析。金融机构已经意识到需要有效地管理其日益重要的数据资产，正在主动思考和实践数据资产治理的方法。目前，金融机构正在加大在数据治理项目中的投入，结合大数据平台建设项目，构建企业内统一的数据池，实现数据的“穿透式”管理。大数据时代，数据治理是金融机构需要深入思考的命题，有效的数据资产管控可以使数据资产成为金融机构的核心竞争力。在国内，金融机构对大数据的认知已经从探索阶段进入到认同阶段。普华永道研究显示，83%的中国金融机构表示希望在大数据方面进行投资。金融行业对大数据的需求属于业务驱动型。它们迫切希望应用大数据技术使营销更精准、风险识别更准确、经营决策更具针对性、产品更具吸引力，从而降低企业成本，提高企业利润。随着更多金融机构基于大数据获得丰厚的回报，将进一步打消它们的顾虑，加速大

数据的普及。

2. 金融行业数据整合、共享和开放成为趋势

数据越关联越有价值，越开放越有价值。随着各国政府和企业逐渐认识到数据共享带来的社会效益和商业价值，全球已经掀起一股数据开放的热潮。大数据的发展需要所有组织和个人的共同协作，将个人私有、企业自有及政府自有的数据进行整合，把私有大数据变为公共大数据。目前，美欧等发达国家和地区的政府都在数据共享上做出了表率，开放大量的公共事业数据。中国政府也着力推动数据开放。一方面，国家带头着力推动政府数据公开。国务院《促进大数据发展行动纲要》提出：中央政府层面实现金税、金关、金财、金审、金盾、金宏、金保、金土、金农、金水、金质等信息系统通过统一平台进行数据共享和交换。另一方面，国家还通过推动建设各类大数据服务交易平台，为数据使用者提供更丰富的数据来源。国家发展改革委办公厅发布的关于请组织申报大数据领域创新能力建设专项的通知中明确提出，要建设大数据流通与交易平台，以支撑数据共享。

3. 金融数据与其他跨领域数据的融合应用不断强化

2015年以前，金融机构主要基于金融业自有信息进行分析。金融机构主要基于自身静态数据通过人工对内进行经营分析、产品设计、营销设计等；对外进行客户分析和行情分析。从2016年开始，大数据技术逐渐成熟，数据采集技术快速发展，通过图像识别、语音识别、语义理解等技术实现外部海量高价值数据收集，包括政府公开数据、企业官网数据、社交数据。金融机构得以通过客户动态数据的获取更深入地了解客户。未来，数据流通的市场会更健全。金融机构将可以方便地获取电信、电商、医疗、出行、教育等其他行业的数据，一方面会有力地促进金融数据和其他行业数据融合，使得金融机构的营销和风控模型更精准；另一方面，跨行业数据融合会催生跨行业的应用，使金融行业得以设计出更多的基于场景的金融产品，与其他行业进行更深入的融合。

4. 金融数据安全问题越来越受到重视

大数据的应用为数据安全带来新的风险。数据具有高价值、无限复制、可流动等特性，这些特性为数据安全管理带来了新的挑战。对金融机构来说，网络恶意攻击成倍增长，组织数据被窃的事件层出不穷。这对金融机构的数据安全管理能力提出了更高的要求。大数据使得金融机构内海量的高价值数据得到集中，并使数据实现高速存取。但是，如果出现信息泄露，可能一次性泄露组织内近乎全部的数据资产。数据泄露后还可能急速扩散，甚至出现更加严重的数据篡改和智能欺诈的情况。对个人来说，金融信息的泄露会暴露出大量的个人基本信息和消费信息等，大数据技术可以便捷地大批量收集这些信息并进行画像，这使得公民更容易受到欺诈，从而造成经济损失。

2.2.2 金融大数据应用面临的挑战

在数字经济时代，数据已成为金融行业的核心资产。金融大数据不仅能够提升金融机构的运营效率，还能为风险管理、客户服务和产品创新提供强有力的支持。然而，随着数据量的爆炸式增长，整个金融行业也面临着数据治理、隐私保护和价值挖掘的挑战。

1. 金融行业的数据资产管理应用水平仍待提高

金融行业的数据资产管理仍存在数据质量不足、数据获取方式单一、数据系统分散等一系

列问题：一是金融数据质量不足，主要体现为数据缺失、数据重复、数据错误和数据格式不统一等；二是金融行业数据来源相对单一，对于外部数据的引入和应用仍需加强；三是金融行业的数据标准化程度低，分散在多个数据系统中，现有的数据采集和应用分析能力难以满足当前大规模的数据分析要求，数据应用需求的响应速度仍显不足。

2. 金融大数据应用技术与业务探索仍需突破

金融机构原有的数据系统架构相对复杂，涉及的系统平台和供应商相对较多，实现大数据应用的技术改造难度较大，而且系统改造的同时必须保障业务系统的安全可靠运行。同时，金融行业的大数据分析应用模型仍处于探索阶段，成熟案例和解决方案仍相对较少，金融机构应用大数据需要投入大量的时间和成本进行调研和试错，一定程度上制约了金融机构应用大数据的积极性。而且，目前的应用实践反映出大数据分析的误判率还比较高，机器判断后的结果仍需要人工核查，资源利用效率和客户体验均有待提升。

3. 金融大数据的行业标准与安全规范仍待完善

当前，金融大数据的相关标准仍处于探索期，金融大数据缺乏统一的存储管理标准和互通共享平台，涉及金融行业大数据的安全规范还存在较多空白。相对于其他行业而言，金融大数据涉及更多的用户个人隐私，在用户数据安全和信息保护方面要求更加严格。随着大数据在多个金融行业细分领域中的价值应用不断深化，在缺乏行业统一安全标准和规范的情况下，单纯依靠金融机构自身管控，将带来较大的安全风险。

4. 金融大数据发展的顶层设计和扶持政策还需强化

在发展规划方面，金融大数据发展的顶层设计仍需强化。一方面，金融机构间的数据壁垒仍较为明显，数据应用仍是各自为战，缺乏有效的整合协同，跨领域和跨企业的数据应用相对较少；另一方面，金融行业数据应用缺乏整体性规划，当前仍存在较多分散性、临时性和应激性的数据应用，数据资产的应用价值没有得到充分发挥，业务支撑作用仍待加强，迫切需要通过行业整体性的产业规划和扶持政策明确发展重点，加强方向引导。

2.3 大数据在银行业的应用

银行业监管机构对于数据管理和监测的要求也在不断提高，2022年初，中国银保监会(现为国家金融监督管理总局)印发了《关于银行业保险业数字化转型的指导意见》，指出银行业保险业数字化转型取得明显成效，数字化经营管理体系基本建成，数据治理更加健全，科技能力大幅提升，网络安全、数据安全和风险管理水平全面提升。面对传统发展模式下的整体业绩承压，以及多层次资本市场大发展、金融领域关键性改革逐步深入，银行业亟须通过转变利差盈利模式和粗放式、同质化扩张路径，走上内涵式、综合化、差异化的高质量发展之路。

2.3.1 大数据对商业银行的影响

对于数据密集型的传统银行业，云计算的重要性不言而喻。云计算可以从服务模式、收益模式、风控模式、合作意愿、数据传输安全性、数据存储与维护能力六个维度对商业银行产生重大影响。

1. 变革商业银行的服务模式

云计算可以使客户减少对传统物理营业网点和柜台的依赖性，因为网上银行、移动支付等金融服务可以由客户自主完成业务办理。站在银行的角度，在线业务办理与在线交易可以捕获到更多、更密集的客户金融信息，便于深层挖掘客户金融需求，提供个性化金融服务。此外，银行也可得以实现多元化和集约化发展。原本隶属于不同部门的业务如支付结算、现金及票据管理、供应链管理可以借助云计算强大的计算能力进行横向整合，为客户提供全方位的综合性服务。

2. 变革商业银行的收益模式

传统商业银行的盈利模式由吸储放贷到大力发展中间业务，再到如今拥抱金融科技发展新趋势，布局新型金融业务，商业银行对其收益模式的探索从未停止。以供应链金融为例，银行可以作为云计算提供方，为供应链各个节点上的企业提供云端的从采购到支付的流程处理支持，并实现供应链条上所有企业之间、企业与银行之间的信贷、支付结算等服务的对接。上述提及的金融服务，均能为银行带来除息差收益外的服务和管理收益，并且收益丰厚。

3. 变革商业银行的风控模式

云计算为银行业在风险预警和信贷监测上提供雄厚的技术支持，为重塑征信体系、建设创新型信贷评估体系开辟出一条切实可行的道路。云计算应用之前，商业银行对贷款方的信用评估还停留在比较其信用历史数据与信用历史行为，利用人工评估贷款方的违约概率，对其进行信用评级。云计算的出现使得商业银行能获取到有关借贷者的多维度数据，不局限于借贷者的借贷数据流，而是将诸如电子商务记录、社交网站信息等非结构型数据也纳入进来。这些数据蕴含了与客户信用相关的海量信息，云计算高度发达的数据处理和计算能力将在贷款风险评估方面发挥不可替代的作用。

4. 提升银行同业间的合作意愿

运营和维护数据中心的成本非常高，囊括了服务器购置成本、运营及人工成本、电力消耗、场地费用和日后扩展升级费用。大型银行具有足够的财力支持相关费用，但对于广大的中小银行来说，这些费用是一笔不菲的开支。况且，仅凭单家中小银行之力承担高端研发或将系统外包也不经济。此时，云平台将发挥使各家银行间的计算资源实现共享的作用。依托云计算技术，众多中小银行通过搭建统一的后方服务支持平台，花费最小的成本便可拥有强大的数据存储和计算能力，为后续进行高端研发和计算工作提供诸多便利。因此，各方银行可有效开展全方位、多层次合作。

5. 提升数据传输安全性

近些年，客户金融信息泄露、信用卡盗刷等问题频频发生，威胁金融消费者的财产安全和隐私安全。网上银行、移动支付等服务均通过互联网进行，这无疑增加了客户个人信息及交易信息暴露的可能，从而降低数据传输的安全性。云计算平台的出现，可在很大程度上拦截可能造成数据盗取风险的行为。如若遇到病毒侵袭程序或数据被盗取，云计算平台下的某个节点会出现异常，异常将自动发送到云端服务器，由云端专业人员做出分析并实时汇报给其他节点的客户，第一时间将风险锁定。

6. 提升数据存储与维护能力

NAS(网络附加存储)和SAN(存储区域网络)是银行传统上采用的两种主要存储技术。这两种

技术所需的基础设施庞大，管理和维护成本高昂。随着银行数据量级不断扩大，它们已越来越不能满足现行的存储需求。云计算技术可将数据存储在云端服务器集群中，如遇服务器崩溃，可通过数据复制算法及时复制到其他服务器中，有效保存海量数据。

2.3.2 大数据在银行业的主要应用场景

1. 数字化营销

按照单个客户个性化的营销方案和服务体系，金融机构依托信息化技术手段可以建立起精确的营销方案，以实现对个人客户的精准营销(precision marketing)。这种建立在精准定位基础之上的营销活动包含着对个体的关注和差异化的认同，可以最大限度地摊平企业的成本。精准营销对于每一位金融客户的兴趣、爱好、购买能力均可以做出预测和判断，根据综合化的评分向顾客推荐金融服务及产品，以保障推荐产品在其财力范围和兴趣偏好范围。

传统银行业中，认识产品、产生兴趣、付款购买三个环节是金融消费者在购买过程中必然出现的环节。在认知产品的过程中，消费者会通过网络、私人渠道进行检索，对产品信息、类别进行了解以确定其购买信息，在此过程中产生的搜索数据便可以定位消费者的收入水平、兴趣和爱好。企业可借助分布式存储和云计算深度挖掘这一系列关于该类消费者的信息，形成完整的客户关系系统(CRM系统)，从而设计出多样化的营销方案，推送给消费者，实现精准化营销。

2. 精细化管理

大宗交易数据是传统银行最为重视的业务内容，由于受制于银行较弱的数据处理能力，体量庞大、细节更多的精细化交易数据无法得到有效处理。例如，传统银行经营模式下，商业银行仅能记录每次的银行卡消费信息，无法实现实时的消费信息反馈、归集整理；在存款、贷款风险管控过程中，银行也难以对小额贷款实施有效的风险管控。一般的商业银行在记录了客户的消费和产品选择数据后，也未能充分利用这些并非为银行经营活动(如风控、催收)而生成的数据。顾客的每笔投资和消费都被记录分析之后，运用数据挖掘技术将产生信息化决策，有助于提升用户体验，银行的精细化管理水平将不断提高。

3. 低成本管理

传统银行业中的信息数据是手工化产生的，容易出现错误。特别是在信贷活动过程中，银行客户信息在记录过程中产生的错误将会给银行经营的有效性造成损失，这种传统的手工信息处理方式效率低下、准确度差、成本高。在商业银行的贷款业务中，银行需要对客户的个人信息、财务状况和抵押物品等进行尽职调查，信息获取的成本较高。但在大数据时代，商业银行对客户信息的采录过程完全自动，通过客户自填、自报，收录客户的个人信息，从而更加精确地了解到客户的实际情况，降低人工处理成本，提高工作效率。对于个体、小微层面的关注将有效解决长期以来困扰中国企业的“小微融资难”问题。

4. 集中化管理

商业银行的传统业务模式中，跨地区、跨国经营成本极高，商业银行不仅要承担开设实体机构的成本，还要承担与代理行之间产生的摩擦成本。在大数据技术的帮助下，商业银行总部机构将可以更加便捷地获取更多有价值的信息，且不局限于当地的分支机构，而是跨越了地域、时空限制。对各类多线信息的集约化管理，使得银行总部的经营权力更加集中，分支行的执行

功能得以强化；银行的管理职责与风控策略将集中于总行层面，有助于进一步提升机构的总体执行力，并减少上传下达过程中的摩擦成本。

2.4　大数据在证券行业的应用

现代证券行业具有资本密集、信息密集、智力密集和技术密集的特点，数据的产生、传播、内容及形式等都更加多样、复杂，越来越呈现出细节化、多维化、立体化的特点，对业务发展的影响也越来越大。证券行业作为世界上数据最为密集的行业之一，拥有大量的证券交易数据、客户数据、监管数据、行情数据等结构化与非结构化数据，这些数据具有体量巨大、产生速度快、类型多样、分布协同的特征，证券行业天然需要通过大数据技术来推动业务的转型升级。

2.4.1　大数据对证券业的影响

信息是证券业的命脉，归根结底，谁的用户信息量大，谁就是王者；谁掌握了数据，谁就拥有风险定价能力，谁就可以获得高额的风险收益，最终赢得竞争优势。

1. 大数据推动证券数字化战略转型

大数据技术正是证券业深入挖掘既有数据，找准市场定位，明确资源配置方向，推动业务创新的重要工具。业务转型的关键在于创新，但现阶段国内金融机构的创新往往沦为监管套利，没有能够基于挖掘客户内在需求，提供更有价值的服务。企业日常运营与大数据紧密结合，智能服务于企业，能够缩短数据分析周期，提升运营效率。基础业务重复性强，非常烦琐，大数据的运用可以促成大量全自动化结果的生成，大幅度提高基础业务的处理速度。以RSMetrics卫星情报分析公司为例，该公司采用卫星遥感技术，采集各商铺、办公用地等地的停车场流量信息，可以推断地区的客户流量信息和销售情况，由此进一步推断企业的运营状况。这种高效、准确、智能化的方式相较于传统的人工分析手段，优势不言而喻。

2. 大数据在证券行业的应用能够降低管理和运行成本，提高服务水平和利润

通过证券数据分析，金融机构能够准确地定位内部管理缺陷，制定有针对性的改进措施，实行符合自身特点的管理模式，进而降低管理运营成本。特别是大量的线上交易活动的展开，进一步降低了线下活动带来的一系列成本，也使得相关成本更加透明，方便监管。此外，证券数字化还提供了全新的沟通渠道和营销手段，使金融机构可以更好地了解客户的消费习惯和行为特征，及时、准确地把握市场营销效果，为用户提供高水平的服务。同时，用户的常用关键字也将自动录入并生成新的标签，创建出客户购买倾向模型、客户忠诚度模型、客户投资能力模型等。也就是说，标签和索引的生成是智能的，这些模型也是具有适应性的。于是，基于这种动态的数据管理系统，辅以可视化技术，证券公司可以更有针对性、更高效、更直观地向各类客户群体推送信息。

3. 大数据在证券行业的应用有助于增强风险控制能力

证券数据分析可以摒弃原来过度依靠客户提供财务报表获取信息的业务方式，转而对其资产价格、账务流水、相关业务活动等流动性数据进行动态和全程的监控分析，从而有效提升客户信息透明度。目前，通过基于大数据，整合客户的资产负债、交易支付、流动性状况、纳税和信用记录等，对客户行为进行全方位的评价，给证券数字化发展带来了巨大机遇，用户的征

信活动相较之前更加及时、有效、详细。此外，随着数据量的增大，加之合理且高效的筛选、整合、处理分析，基于大数据的风险控制和信用管理体系可以具备更高的准确性，也可以进行实时动态监测，以达到更优的事前、事中、事后控制效果。目前，证券行业的事前风险控制尚有欠缺，在开展互联网金融业务的同时，必须关注信用的核心地位，树立风险控制意识，通过多元化数据和模型建立传统模式与互联网金融相结合的风险控制、征信体系，以减少产品风险系数与客户风险承担能力不匹配的情况出现。另外，寻求不同金融领域和机构的合作、数据共享，将增加数据的容量和有效性。

4. 推动证券监管制度变革

传统的证券监管制度侧重于对违法行为的事后处罚，而在交易监管，特别是内幕交易监管方面，由于此类行为隐蔽性强、界限模糊，仍难以有效捕捉。这一问题的根源在于信息披露不充分及数据处理深度不足。此外，监管方式的选择也与市场效率的提升存在一定冲突。而引入大数据观念和技术，就是弥补传统证券监管制度缺陷、推动证券监管制度变革的重要拐点。将分析样本延伸至全社会、全数据，利用数据间的相互验证，提高准确度；工作重心转向数据分析和工具优化，分析侧重点趋向数据相关分析；对市场主体的活动监管从监管交易者的直接行为扩展到关联行为，进行全方位监管，提高事前预防水平。为了避免降低市场效率并保障数据的安全性，有必要对数据进行标准化处理，以及对数据使用和分析标准立法，来约束大数据的收集、使用和分析行为。

2.4.2 大数据对证券业的作用

1. 大数据可以提升证券业的个性化服务水平

证券行业作为综合类金融服务产品的提供者，在大数据的背景下，将有能力快速搜集高质量的信息，以设计出更符合客户需求的产品组合，并且可以根据客户偏好的改变及时调整。同时，由于中介服务的竞争逐渐同质化，争夺的焦点将来必然落在价格上。如果标准化同质服务不再能够给券商带来正常利润，那么券商必须转变经营思路，将通道业务转变成包含增值服务的金融服务。大数据能够通过对客户消费行为模式进行分析，提高客户转化率，开发出个性化的产品以满足不同客户的需求。越来越多的证券公司开始采用数据驱动的方法，通过一系列信息的收集、存储、管理和分析，给客户提供更好的决策，充分体现了以客户为中心的服务理念。

2. 大数据能够帮助证券公司避免客户流失

依据客户历史交易行为和流失情况创建大数据分析模型，预测客户流失的概率。比如海通证券自主开发的“基于数据挖掘算法的证券客户行为特征分析技术”主要应用于客户深度画像以及基于画像的用户流失概率预测，通过对海通证券100多万样本客户、半年交易记录的海量信息分析，建立了客户分类、客户偏好、客户流失概率等模型。该项技术的初衷是希望通过客户行为的量化分析，测算客户将来可能流失的概率。

3. 大数据在量化投资方面的应用

量化投资策略在欧美发达国家和地区的金融市场已经相对成熟，由于收益巨大，成为大数据最早应用的领域。证券业已经进入了一个大数据信息时代，证券的数据模型越来越复杂多样，数据的总量和种类都有重大的突破。大数据在处理证券数据时，通过对主力和散户的行为、轨迹分析，对主力资金和散户资金的去向追踪，对主力、散户和市场之间关系的理解，能够很好

地增加投资胜率。个人投资者将能够轻松使用大数据获得实证支持，降低交易风险，投资能力将大幅提升。大数据也让科技公司第一次有机会挑战传统的金融分析师，利用对各种数据的量化、重组和整合，提供不同的交易策略，让投资者能够科学地分析全球投资市场。

2.4.3 大数据在证券业中的具体应用

作为业务发展的重要技术支撑手段，大数据技术已经逐渐融入证券公司日常工作的各个环节，通过大数据建设总体规划，实现公司内外部数据的融合，建立统一和多维化的公司数据模型，实现数据驱动的业务模式创新，原来业务运营是靠人、靠经验，现在是以数据驱动，所有的过程和结果都可以量化，凸显数据驱动业务创新的新业态。

1. 智能管理

随着证券公司大数据中台的建设，各部门各业务数据统一接入，可以实现数据集中、口径一致，确保公司各级部门均可在保证数据隐私和安全的前提下使用数据，充分发挥数据作为证券公司重要资产的业务价值，实现公司内部数据共通共享。在业务管理上，实现产品、客户等数据的企业级视图的统一，有效促进业务的集成和协作，并为企业级分析、协同开发、交叉销售提供基础。同时按角色聚合业务数据，让不同层级的管理者及员工对业务目标始终有着全面清晰的认知，比如恒泰证券的网点管理平台和员工端超级APP，就是基于大数据技术提升中台能力，进而提升公司的管理效率和投产效率。

2. 客户营销

随着互联网深化发展，线上流量红利基本殆尽，而证券行业特有的获客成本高、用户转化节点多、转化周期长的特点却没有改变，需要利用大数据技术精准营销客户，在促进活跃用户稳步增长的同时不断提升用户的生命周期。这种建立在精准定位基础上的营销活动可以最大限度地摊平运营成本。根据客户多维度的数据描绘每个客户的画像，包括风险等级、个人属性、账户资产、操作数据、偏好信息等，通过标签和推荐模型给不同类别的客户匹配不同的资讯与产品信息，从而提升营销成功率，减少冗余信息对客户的干扰。在传统的数据运算模式下，数据集中存储、串行计算，即客户数据存储在一台服务器上，必须先计算完一个客户的数据再计算下一个客户的数据，当客户数量巨大时，存储器的存储负荷过重，运算非常浪费时间。而恒泰证券的大数据平台在做系统架构时选择采用分布式存储和并行计算的运算方式，将客户数据存储在分散的存储器中，构成一个虚拟存储系统，可以在同一时间计算客户数据，极大提升了运算效率，实现了面向千万量级客户的实时精准营销。

3. 客户安全

2020年4月，有A股投资者股票账户被盗，卖出全部持仓股，全仓买入“庄股”济民制药，次日一字跌停。新闻一出市场哗然，遭受损失的股民更是走上了维权之路。在账户安全层面，证券公司有责任和义务通过金融科技的力量，尽力保护投资者免遭“黑产”对财富安全的侵袭。恒泰证券为此推出安全微管家服务，实时拦截异常风险，极速提醒客户异常。但是黑客也会相应地升级针对措施，如大量变换IP、缓滴间歇式撞库，让传统的安全规则在现有的攻防形势下显得不太够用。前文提及的证券公司都搭建了数据埋点系统，基于大数据平台的实时数据采集和流式计算系统，收集了用户的访问时间、停留时间、页面路径、页面单击详情等数据。这些数据沉淀下来，除了应用于用户画像、精准营销、产品数据分析等业务，在安全领域也大有用

武之地。通过整合用户历史信息，形成一个多维行为画像作为基线，当用户实时行为产生变化而不可信时，即可实时提升风控等级，如通过通知客户、拦截验证甚至锁定账户的方式保障账户安全。大数据与AI技术相互融合，通过机器学习模型，挖掘聚类高风险行为特征，能达到事前防范、提前预警的效果。

4. 产品投研

大数据平台的技术与数据本身也可以直接转化为智能化的产品和服务，将用户产生的数据反哺于用户服务。例如，金融街证券即将上线的热股产品，就是一款基于用户搜索、浏览等行为数据，并结合行情交易量等信息研发的智能产品。热股产品创新性地引入物理模型，基于大数据平台实时分析计算，利用牛顿冷却定律，实时计算出当前的股市热点。在策略模型投研层面，证券公司还可以基于大数据机器学习技术，对技术面、基本面、资金面、新闻研报等多种因子进行组合，甚至引用新闻舆论等外部数据来构建产品逻辑，使用历史大数据进行训练，并且能通过新数据不断改进模型参数。同时根据每一个客户在流动性需求、风险属性、资金门槛属性的具体情况，用量化模型自动计算出适合客户的大类资产配置方案，实现金融普惠化。

2.5 大数据在保险业的应用

随着网络和信息技术的不断普及，人类产生的数据量正在呈指数级增长，而云计算的诞生更是直接把我们送进了大数据时代。“大数据”作为时下最时髦的词汇，开始向各行业渗透、辐射，颠覆很多行业特别是传统行业的管理和运营思维。在这一大背景下，大数据也触动着保险行业管理者的神经，搅动着保险行业管理者的思维；大数据在保险行业释放出的巨大价值吸引着诸多保险行业人士的兴趣和关注。探讨和学习如何借助大数据为保险行业运营管理服务也是当今该行业管理者面临的挑战。

2.5.1 大数据对保险行业的影响

大数据应用，其真正的核心在于挖掘数据中蕴藏的情报价值，而不是简单的数据计算。那么，对于保险行业来说，管理者应该如何借助大数据为保险行业的运营管理服务呢？同时，大数据应用又将如何凸显其在保险行业的情报价值呢？

1. 大数据有助于保险产品市场定位精准化

成功的品牌离不开精准的市场定位，可以说，一个成功的市场定位能够使一个企业的品牌加倍、快速成长，而基于大数据的市场数据分析和调研是企业进行品牌定位的第一步。保险行业企业要想在无硝烟的市场中分得一杯羹，需要构建大数据战略，拓宽保险行业调研数据的广度和深度，从大数据中了解保险行业市场构成、细分市场特征、消费者需求和竞争者状况等众多因素，在科学系统的信息数据收集、管理、分析的基础上，提出更好的解决问题的方案和建议，保证企业品牌市场定位独具个性化，提高企业品牌市场定位的行业接受度。

企业开拓新市场需要投入大量的人力、物力和精力，若市场定位不精准或者出现偏差，给投资商和企业自身带来的后期损失是巨大的，甚至是毁灭性的，由此可以看出市场定位对保险行业市场开拓的重要性。只有定位准确乃至精确，企业才能开发出满足市场需求的产品，从而在竞争中立于不败之地。然而，要想做到这一点，必须有足够的信息数据供保险行业研究人员

分析和判断。在传统方式下，分析数据主要来自统计年鉴、行业管理部门数据、相关行业报告、行业专家意见及属地市场调查等，这些数据多存在样本量不足、时间滞后和准确度低等缺陷，研究人员能够获得的信息量非常有限，使准确的市场定位存在数据瓶颈。随着大数据时代的到来，借助数据挖掘和信息采集技术，不仅能给研究人员提供足够的样本量和数据信息，还能建立基于大数据的数学模型，对未来市场进行预测。当然，依靠传统的人工数据收集和统计显然难以满足大数据环境下的数据需求，这就需要依靠相关数据公司自动化数据采集工具的帮助。

2. 大数据有助于保险行业市场营销数字化

在保险行业市场营销工作中，无论是产品、渠道、价格还是顾客，每一项工作都与大数据的采集和分析息息相关，而以下两个方面又是保险行业市场营销工作中的重中之重：一是通过获取数据并加以统计分析来充分了解市场信息，掌握竞争者的商情和动态，明确产品在竞争群中所处的市场地位，以达到“知彼知己，百战不殆”的目的；二是企业通过积累和挖掘保险行业消费者档案数据，分析顾客的消费行为和价值偏好，便于更好地为消费者服务和发展忠诚顾客。以保险行业对顾客的消费行为和偏好分析为例，如果企业平时善于积累、收集和整理消费者的消费行为方面的信息数据，如消费者购买产品的花费、选择的产品渠道、偏好产品的类型、产品使用周期、购买产品的目的、消费者家庭背景、工作和生活环境、个人消费观和价值观等，建立消费者大数据库，便可通过统计和分析来掌握消费者的消费行为、兴趣偏好和产品的市场口碑现状，再据此制定有针对性的营销方案和战略，投消费者所好，其带来的营销效应是可想而知的。因此，大数据中蕴含着出奇制胜的力量，如果企业管理者善于在市场营销中加以运用，将成为保险行业在市场竞争中立于不败之地的有力武器。

3. 大数据有利于推动保险业收益管理科学化

需求预测是通过对建构的大数据进行统计与分析，采取科学的预测方法，通过建立数学模型，使企业管理者掌握和了解保险行业潜在的市场需求、未来一段时间每个细分市场的产品销售量和产品价格走势等，从而使企业能够通过价格的杠杆来调节市场的供需平衡，并针对不同的细分市场实行动态定价和差别定价。需求预测的好处在于可以提高企业管理者对保险行业市场判断的前瞻性，并在不同的市场波动周期以合适的产品和价格投放市场，获得潜在的收益。细分市场为企业预测销售量和实行差别定价提供了条件，其科学性体现在通过保险行业市场需求预测来制定和更新价格，最大化各个细分市场的收益。敏感度分析是通过需求价格弹性分析技术，对不同细分市场的价格进行优化，最大限度地挖掘市场潜在的收入。大数据时代的来临为企业收益管理工作的开展提供了更加广阔的空间。需求预测、细分市场和敏感度分析对数据需求量很大，而传统的数据分析大多采集企业自身的历史数据来进行预测和分析，容易忽视整个保险行业信息数据，因此难免使预测结果存在偏差。企业在实施收益管理过程中，如果能在自有数据的基础上依靠一些自动化信息采集软件收集更多的保险行业数据，了解更多的保险行业市场信息，将会对制定准确的收益策略、赢得更高的收益起到推进作用。

4. 大数据驱动保险行业需求开发

随着论坛、博客、微博、微信、电商平台、点评网等媒介在PC端和移动端的创新与发展，公众分享信息变得更加便捷和自由，而公众分享信息的主动性促进了“网络评论”这一新型舆论形式的发展。微博、微信、点评网及各类评论板块中数以亿计的网络评论形成了交互性大数据，其中蕴藏着巨大的保险行业需求开发价值，值得企业管理者重视。大数据并不是一个

神秘的字眼，只要保险行业企业平时善于积累和运用自动化工具收集、挖掘、统计和分析这些数据，为己所用，就会有效地提高市场竞争力和收益能力，获得良好的效益。

2.5.2 大数据在保险行业中的具体应用

大数据技术在保险业的应用具有得天独厚的优势。首先，保险业拥有大量数据。其次，保险经营的基础是大数法则，而大数据的核心是预测，是评估概率的最有力武器。大数据影响着保险行业的日常运作，在价值链的方方面面都有所体现，尤其以营销、产品、定价、服务和管理等方面为重点。

1. 客户市场细分

首先是精准营销。保险公司需要的数据来源不仅仅是保险公司的客户的数据，还应包括保险体系外的数据，如合作方数据、互联网社交数据和行为数据等。这些数据均与保险市场细分密切相关。通过对客户信息和客户行为的深度分析，能预测及了解客户需求，挖掘潜在客户，推荐适合的产品，从而实现个性鲜明的精准营销。大数据时代的营销，不应采用同样的广告或营销手段面对所有群体，而应针对细分的客户群体进行差异化宣导。对行为数据、意图数据等进行细分，可以识别用户对某一产品的关注度，从而确定更为合适的营销手段。例如，为每天手机上网时间超过5小时的客户推荐健康险和手机意外险；为喜食肉类、爱喝酒的客户推荐重大疾病保险；为使用大屏幕手机的用户推荐碎屏险；为携程、途牛网用户推荐旅游相关保险产品等。保险行业可以通过精准数据分析来寻找有保险需求的目标客户，并开展针对性营销，根据客户需求推送相关保险产品信息，避免采用信息群发的方式引发客户反感或抵触情绪。

其次是洞悉客户需求，实现产品开发个性化。保险公司在开发保险产品时，通常以市场需求为基础进行产品创新，或者对现有产品进行改良和组合，以适应市场需要、提高自身竞争力。在充分了解客户需求后，可通过积累大规模数据，基于与承保标的风险相关的已知数据预测损失情况，从而开发出针对不同客户群体的个性化产品。

2. 优化定价体系

以下针对“大数据”与“大数法则”进行分析。有些观点认为大数据颠覆了大数法则，实际上，虽然两者都是在大量数据基础上寻找规律、进行风险分析和预测，但在保险产品定价机制中的作用是完全不同的。保险是经营风险的学科，基于统计学的大数法则是精算理论的核心，我们需要依托大数法则确保行业基准纯风险损失率厘定的充足和公平。也就是说，大数法则是保险业运行管理的数理逻辑，是保险业不可动摇的理论和定价基础。大数据主要发挥保险定价的辅助作用，特别是采集客户交易、搜索等行为的网络数据进行相关性分析，找寻数据背后的匹配规律，推动保险公司细化客户群，优化精算定价模型，建立科学、有效的保险费率浮动模式和差别化定价机制。因此，大数据并没有颠覆大数法则，而是对保险费率市场化形成机制的重要优化和改进，是一种依托新技术、更加精细化的风险管理辅助工具。

3. 提高保险理赔的准确性

理赔管理对于提高保险公司的盈利能力和建立企业信誉等起着关键作用。大数据可以进一步提高保险理赔的准确性和及时性。 在大数据时代，保险公司可以通过公共信息实时获得客户的出险信息，并及时主动联系客户提供理赔等服务。不同的报案手段可以应用不同的大数据方法，比如，保险公司可以对电话报案的号码进行定位，大致了解事故发生的地点，或通过技术手段

获取用户的GPS 定位信息，以便确定报案地点。在确认事故原因方面，可以通过远程发送或上传现场图片的方式，通过图片分析挖掘技术进行事故原因分析，确认各方责任，为客户实现更为快捷、方便的赔付，给客户带来更好的赔付体验。

大数据方法同样可以应用在与第三方的合作上，比如，如果客户在高速公路上驾车出险，该客户向交警报案，保险公司就能够及时获得报案信息，在客户没有向保险公司提出索赔申请之前主动向其提供理赔服务；保险公司可以通过与4S店的合作，获取每辆汽车的维修、保养情况，查勘员在查勘时就很容易得知车辆状况，以便进行适当的、合理的赔付；当客户遭遇意外伤害事故时，客户也不必到保险公司进行索赔，医院可以将客户信息直接传送给保险公司，保险公司的预付赔款可以直接地划到客户的账户内，解客户燃眉之急。

4. 防止保险欺诈

反欺诈大数据还可以提高保险业的管理能力，防止保险欺诈等行为的发生。从本质上看，欺诈是双方信息不对称导致的，而大数据能够弱化部分信息的不对称，从而降低保险公司的赔付成本，保障保险公司的正当权益。在反欺诈方面，保险公司通常依靠一些固定的审核标准，以及理赔人员的经验积累，来判断是否存在保险欺诈。由于缺乏行业内协作机制和共享的信息平台，调查的质量主要依赖于理赔人员的个人经验和责任心，以及与公安机关的合作情况。大数据技术的战略意义不仅在于掌握大量数据信息，更重要的是，如何对这些信息进行专业化的分析和处理，从而迅速获得隐藏其后的有用信息并转化为效益。国外的保险公司已经开始在健康保险领域应用大数据技术防范保险欺诈，它们积累的丰富经验值得我们学习。

保险业发展大数据的前景广阔，但也不能忽略发展变革带来的问题及面临的挑战。首先，大数据的发展对存储的安全性要求越来越高，个人的各种隐私信息不断融入大数据的分析中，数据的隐私保护还有待进一步完善。其次，保险公司在处理风险相关信息的大数据分析能力方面相对薄弱，还需要继续加强和提高，如何更好地利用自身的数据资源，是对保险公司的一大考验。此外，越来越多的互联网企业和科技公司凭借其在大数据方面的优势进入保险行业，将使保险行业的内外部竞争进一步加剧。

课 后 思 考

(1) 金融大数据的内涵是什么？
(2) 大数据对金融行业的影响主要包括哪些方面？
(3) 大数据对银行业的影响主要包括哪些方面？
(4) 大数据对证券业的影响主要包括哪些方面？
(5) 大数据对保险业的影响主要包括哪些方面？
(6) 金融大数据在应用中面临哪些问题？