



# 大数据应用对 KIBS 企业 R&D 团队 知识创造影响研究

## ——以中小型软件企业为例

STUDY ON THE INFLUENCES OF BIG DATA  
APPLICATIONS ON KNOWLEDGE CREATION  
IN R&D TEAM OF KIBS ENTERPRISES  
——TAKING SMALL AND MEDIUM SIZED  
SOFTWARE ENTERPRISES AS AN EXAMPLE

黄小龙

XIAOLONG HUANG

本博士论文提交正大管理学院中国研究生院  
属正大管理学院工商管理哲学博士学位  
工商管理专业（中文体系）课程学习的一部分  
二〇一八年八月

สำนัก 书脊

大数据应用对 KIBS 企业 R&D 团队知识创造影响研究  
——以中小型软件企业为例

黄小龙  
二〇一八年





# 大数据应用对 KIBS 企业 R&D 团队 知识创造影响研究

## ——以中小型软件企业为例

STUDY ON THE INFLUENCES OF BIG DATA  
APPLICATIONS ON KNOWLEDGE CREATION  
IN R&D TEAM OF KIBS ENTERPRISES  
——TAKING SMALL AND MEDIUM SIZED  
SOFTWARE ENTERPRISES AS AN EXAMPLE

黄小龙

XIAOLONG HUANG

本博士论文提交正大管理学院中国研究生院  
属正大管理学院工商管理哲学博士学位  
工商管理专业（中文体系）课程学习的一部分

二〇一八年八月

版权归正大管理学院所有

本博士学位论文题目：大数据应用对 KIBS 企业 R&D 团队知识创造影响研究—以中小型软件企业为例，作者：黄小龙，已通过博士学位论文答辩委员会审核。本博士学位论文为正大管理学院工商管理哲学博士学位工商管理专业（中文体系）课程学习的一部分。

**博士学位论文答辩委员会签名：**

答辩委员会主席-----  
( Doctor )

主导师/答辩委员-----  
( Doctor )

副导师/答辩委员-----  
( Doctor )

答辩委员 -----  
( Doctor )

答辩委员 -----  
( Doctor )

**中国研究生院院长签名：**

-----  
(Associate Professor Doctor Ruengsak Kawtummachai)

批准日期：二〇一八年 月 日

## 摘要

论文题目： 大数据应用对 KIBS 企业 R&D 团队知识创造影响研究  
—以中小型软件企业为例

作者： 黄小龙

导师： 韩圣龙教授

学位名称： 工商管理哲学博士学位

专业名称： 工商管理专业（中文体系）

学年： 二〇一八年

近年来，服务业出现一种以网络和信息技术为背景，以智力资本为主要投入要素，服务成果具有高附加值、高赢利、专业性、知识性等特点的新型服务产业，如软件企业、网络信息业、金融业、网络专业咨询业等，这类企业称知识密集服务型企业，即 KIBS 企业。而在 KIBS 企业的知识服务产品开发过程中，R&D 团队是知识创造的主要组织形式，是企业创造价值的核心单位。软件企业属于典型的 KIBS 企业，其内部的 R&D 团队更是整个企业的创新灵魂。

大数据来临是不争的事实。大数据时代各类数据、信息、知识通过网络爆炸性地发展，传统的信息技术模式已无法满足大数据发展的步伐。因此，各行业须通过技术创新，研究新的信息技术及管理模式，来提高生产效率，增强核心竞争力。作为依赖网络和信息技术 KIBS 企业，在知识和价值生产过程中对数据的依赖程度极高，因此研究大数据应用对 KIBS 企业 R&D 团队知识创造的影响具有重要的现实意义。

目前大数据相关研究主要集中在数据处理的技术层面，而通过管理视角研究大数据尚处于起步阶段。总体上讲，现有大数据研究主要有以下问题：（1）从企业应用视角看，对大数据的技术特性、应用范围及应用价值还不了解；（2）大数据应用对企业造成的影响研究极少；（3）缺乏实证研究。

本文基于知识管理理论，从企业管理视角，通过软件企业来研究大数据应用对 KIBS 企业 R&D 团队知识创造的影响，研究主要包括四方面：（1）大数据应用的技术特性。（2）大数据应用下 R&D 团队知识创造过程发生的变化。（3）大数据应用对 R&D 团队知识创造的影响因素。（4）实证研究是否获得成功验证。

本研究通过理论和实证两方面对大数据应用、R&D 团队和知识创造三个变量

之间的维度关系进行分析研究。首先，归纳出大数据应用的技术特性，为今后的研究提供参考。其次，以案例分析法研究了大数据应用下 R&D 团队知识创造过程的变化以及其影响因素，以此来丰富知识创造理论。最后，通过统计分析方法实证了大数据应用对 R&D 团队知识创造的影响模型，丰富了管理视角下大数据应用的研究方法。本研究结论可为企业管理的大数据应用实践提供参考依据。

**关键词：**大数据 KIBS 企业 R&D 团队 知识创造 软件企业

## ABSTRACT

Title: Study on the Influences of Big Data Applications on Knowledge Creation in R&D Team of KIBS Enterprises —Taking Small and Medium Sized Software Enterprises as an Example

Author: Xiaolong Huang

Advisor: Dr. ShengLong Han

Degree: Doctor of Philosophy (Business Administration)

Major: Business Administration (Chinese Program)

Academic Year: 2018

In recent years, a new type of service industries have shown up in service industries, they use network and information technology as their main background, intellectual capital as their main inputs and high added values, high profitability, professionalism and knowledge as their outputs. This type of enterprises such as software service industry, internet industry, financial industry, internet consulting industry and so on are called Knowledge-intensive service enterprises, that is, KIBS Enterprises. In the development process of knowledge service product in KIBS enterprises, R & D team is the main organization of knowledge creation and the core unit to create enterprise value. Software enterprises are typical KIBS enterprises, their internal R & D team is the innovative soul of the enterprises.

It is an indisputable fact that big data is coming. In the era of big data, various types of data, information and knowledge explosively produced through internet, the traditional information technology model has been insufficient to meet the needs of big data development. Therefore, all industries have to develop new information technology and management model through technological innovation so as to improve productivity and enhance core competence. Because KIBS enterprises depend mainly on the information and communication technology, they are highly dependent on data in the process of knowledge and value production. Therefore, it is of great practical significance to study the influences of big data applications on knowledge creation in R&D team, KIBS enterprises.



At present, the research on big data is mainly focused on the technical level of data processing, and the study of big data from management perspective is still in its infancy at the technical level of data processing. Big data researches from the management perspective are rare. Generally speaking, the existing big data researches have the following problems: (1) From the perspective of enterprise applications, the technical characteristics, application scope and application value of big data are still not understood; (2) The influences of big data applications on enterprises are rarely studied; (3) Lack of empirical researches.

Based on the theory of knowledge management, this paper studies the influences of big data applications on knowledge creation in R & D team, KIBS enterprises through software enterprises from the perspective of enterprise management. The research mainly includes four aspects: (1) the technical features of big data applications (2) the change of the knowledge creation process in R & D team under big data applications (3) influence factors of big data applications on knowledge creation in R&D team (4) successful verification of empirical research.

This paper theoretically and empirically studies the relationship between dimensions among these three variables: big data applications, R & D team and knowledge creation. First of all, it sums up technological features of big data applications and provides references for future research. Secondly, the case study method is used to study the change and influence factors of knowledge creation process in R & D team under big data applications, so as to enrich the knowledge creation theory. Finally, the statistical analysis method is used to test and verify the influence model of big data applications on knowledge creation in R & D team, which enriches the research methodology of big data applications from management perspective. The conclusion of this study can provide reference for the implementation of big data applications in enterprise management.

**Keywords:** Big Data                      KIBS Enterprises                      R&D team  
Knowledge Creation                      Software Enterprises

# 目 录

摘 要 .....	I
ABSTRACT .....	II
目 录 .....	III
表格目录 .....	IV
图片目录 .....	V
第一章 绪论 .....	1
1.1 研究背景 .....	1
1.1.1 社会发展背景 .....	1
1.1.2 KIBS 企业和 R&D 团队重要地位及其知识创造的紧迫性 .....	2
1.2 研究现状及问题提出 .....	3
1.3 相关概念界定 .....	4
1.4 研究内容 .....	9
1.5 研究方法及技术路线 .....	9
1.5.1 研究方法 .....	9
1.5.2 技术路线 .....	10
1.6 论文研究框架 .....	11
1.7 论文创新点 .....	11
1.8 本章小结 .....	12
第二章 文献综述 .....	13
2.1 知识创造文献综述 .....	13
2.1.1 数据、信息和知识的关系 .....	13
2.1.2 知识创造相关综述 .....	14

## 目 录 (续)

2.2 KIBS 企业文献综述 .....	21
2.2.1 KIBS 企业 .....	21
2.2.2 软件企业 .....	24
2.2.3 KIBS 研究评述 .....	26
2.3 R&D 团队文献综述 .....	27
2.3.1 知识层面研究 .....	27
2.3.2 相关影响研究 .....	27
2.3.3 评价与测量研究 .....	28
2.3.4 R&D 团队研究评述 .....	28
2.4 信息技术与管理学文献综述 .....	28
2.4.1 信息技术的资源基础理论 .....	28
2.4.2 信息技术的战略管理理论 .....	29
2.4.3 信息技术对知识管理的作用 .....	29
2.4.4 文献评述 .....	31
2.5 大数据文献综述 .....	31
2.5.1 大数据发展历程 .....	31
2.5.2 大数据的特征和数据结构类型 .....	32
2.5.3 大数据技术理论 .....	33
2.5.4 大数据技术与传统信息技术 .....	37
2.5.5 大数据的影响、作用与应用 .....	37
2.5.6 大数据与管理学领域研究 .....	38
2.5.7 大数据面临的问题 .....	39
2.5.8 大数据研究评述 .....	39

## 目 录 (续)

2.6 本章小结 .....	39
第三章 软件工程及其大数据应用 .....	40
3.1 软件工程理论基础 .....	40
3.1.1 传统软件工程理论 .....	40
3.1.2 敏捷开发理论 .....	42
3.1.3 面向对象程序设计理论 .....	43
3.1.4 软件工程新技术 .....	43
3.1.5 软件开发团队 .....	45
3.2 软件工程的大数据应用 .....	46
3.3 R&D 团队软件开发的大数据应用过程 .....	47
3.4 软件 R&D 团队大数据应用与知识创造关系 .....	47
3.5 本章小结 .....	48
第四章 R&D 团队大数据应用知识创造过程案例研究 .....	49
4.1 案例 1: 保险客户行为分析系统开发项目 .....	49
4.1.1 软件公司及开发项目概述 .....	49
4.1.2 软件开发大数据应用及其知识创造过程 .....	49
4.1.3 案例中的软件研发大数据应用对知识创造影响分析 .....	51
4.2 案例 2: 软件外包测试项目案例 .....	52
4.2.1 公司及项目概述 .....	52
4.2.2 软件测试大数据应用及其知识创造过程 .....	52
4.2.3 软件测试大数据应用对知识创造影响分析 .....	52
4.3 大数据应用下的知识创造过程 .....	53
4.4 本章小结 .....	53

## 目 录 (续)

第五章 理论假设及研究模型 .....	54
5.1 主要研究变量 .....	54
5.1.1 大数据应用概念 .....	54
5.1.2 大数据应用维度 .....	54
5.1.3 个人能力因素问题 .....	55
5.1.4 团队协作能力问题 .....	56
5.1.5 R&D 团队知识创造维度 .....	56
5.2 研究假设提出 .....	56
5.2.1 大数据应用与 R&D 团队知识创造关系 .....	56
5.2.2 大数据应用与 R&D 团队知识社会化的关系 .....	57
5.2.3 大数据应用与 R&D 团队知识表出化的关系 .....	57
5.2.4 大数据应用与 R&D 团队知识联结化的关系 .....	57
5.2.5 大数据应用与 R&D 团队知识内在化的关系 .....	58
5.2.6 团队协作能力与大数据应用和 R&D 团队知识创造的关系 .....	58
5.3 研究模型提出 .....	59
5.4 本章小结 .....	59
第六章 数据研究方法及问卷设计 .....	60
6.1 数据研究方法 .....	60
6.2 调查问卷设计 .....	60
6.2.1 设计原则 .....	60
6.2.2 变量测量 .....	61
6.2.3 问卷设计 .....	65
6.3 数据采集 .....	66

## 目 录 (续)

6.3.1 小样本数据的收集 .....	67
6.3.2 大样本数据的收集 .....	67
6.3.3 大样本数据描述 .....	67
6.4 研究方法 .....	69
6.5 本章小结 .....	71
第七章 数据分析及假设检验 .....	72
7.1 小样本的预测试 .....	72
7.1.1 小样本数据描述性统计分析 .....	72
7.1.2 小样本数据信度和效度检验 .....	74
7.2 大样本数据检验 .....	84
7.2.1 描述性统计分析 .....	85
7.2.2 信度分析 .....	88
7.2.3 因子分析 .....	90
7.2.4 效度分析 .....	98
7.2.5 ICC 与 rwg 分析检验 .....	99
7.2.6 相关分析 .....	101
7.2.7 假设检验 .....	102
7.3 实证结论分析和讨论 .....	107
7.3.1 结论分析 .....	107
7.3.2 结论讨论 .....	108
7.4 本章小结 .....	108
第八章 结论与展望 .....	109
8.1 研究结论和实践启示 .....	109

## 目 录 (续)

8.2 论文主要研究贡献 .....	111
8.3 存在不足与未来展望 .....	112
参考文献 .....	116
附录 A 大数据应用对软件 R&D 团队知识创造影响调查问卷 .....	129
致 谢 .....	136
声 明 .....	137
个人简历 .....	138

## 表格目录

表 1.1	大数据典型定义 .....	7
表 2.1	知识创造过程研究 .....	19
表 2.2	知识创造影响因素 .....	21
表 2.3	大数据技术与传统信息技术的区别 .....	37
表 2.4	大数据与知识管理研究情况 .....	38
表 3.1	敏捷开发方法较之传统软件工程开发 .....	43
表 4.1	保险客户行为分析系统开发软件 R&D 团队各类人员职责 .....	49
表 6.1	技术外包性测量量表 .....	62
表 6.2	技术智能性测量量表 .....	62
表 6.3	强可视化性测量量表 .....	63
表 6.4	知识社会化测量量表 .....	63
表 6.5	知识表出化测量量表 .....	63
表 6.6	知识联结化测量量表 .....	64
表 6.7	知识内在化测量量表 .....	64
表 6.8	团队协作能力测量量表 .....	65
表 6.9	Pearson 相关系数取值范围表 .....	71
表 7.1	小样本受访者公司和个人信息统计描述表 .....	72
表 7.2	大数据应用量表信度分析表 .....	74
表 7.3	大数据应用量表 KMO 和 Bartlett 球形度检验表 .....	75
表 7.4	大数据应用量表解释总方差分析表 .....	75
表 7.5	大数据应用各维度旋转因子载荷表 .....	76
表 7.6	大数据应用各维度旋转因子载荷表 .....	76
表 7.7	大数据应用量表 KMO 和 Bartlett 球形度检验表 .....	77



## 表格目录 (续)

表 7.8	大数据应用量表解释总方差分析表 .....	77
表 7.9	大数据应用各维度旋转因子载荷表 .....	78
表 7.10	大数据应用最终量表信度分析表 .....	78
表 7.11	知识创造量表信度分析表 .....	79
表 7.12	知识创造量表 KMO 和 Bartlett 球形度检验表 .....	80
表 7.13	知识创造量表解释总方差分析表 .....	81
表 7.14	知识创造量表各维度旋转因子载荷表 .....	81
表 7.15	修正后知识创造量表各维度旋转因子载荷表 .....	82
表 7.16	再次修正知识创造量表 KMO 和 Bartlett 球形度检验表 .....	83
表 7.17	再次修正知识创造量表解释总方差分析表 .....	83
表 7.18	修正后知识创造量表各维度旋转因子载荷表 .....	83
表 7.19	大样本数据受访者公司和个人信息统计描述表 .....	85
表 7.20	大样本数据描述性统计分析图 .....	87
表 7.21	大数据应用量表信度分析表 .....	89
表 7.22	大数据应用总体量表 $\alpha$ 值检测结果图 .....	89
表 7.23	知识创造量表信度分析表 .....	89
表 7.24	大数据应用总体量表 $\alpha$ 值检测结果图 .....	90
表 7.25	大数据应用量表 KMO 和 Bartlett 球形度检验表 .....	91
表 7.26	大数据应用量表解释总方差分析表 .....	91
表 7.27	大数据应用各维度旋转因子载荷表 .....	92
表 7.28	知识创造量表 KMO 和 Bartlett 球形度检验表 .....	92
表 7.29	知识创造量表解释总方差分析表 .....	93
表 7.30	知识创造量表各维度旋转因子载荷表 .....	93

## 表格目录（续）

表 7.31	验证性因素模型拟合度表 .....	95
表 7.32	验证性因素分析结果表 .....	96
表 7.33	验证性因素模型拟合度 .....	97
表 7.34	因子分析结果 .....	98
表 7.35	各变量内部一致性 ICC(1)和 ICC(2)检验表 .....	100
表 7.36	rwg 均值和中位数表 .....	101
表 7.37	大数据应用和 R&D 团队知识创造相关检验结果表 .....	101
表 7.38	各变量维度间偏相关检验结果表 .....	102
表 7.39	大数据应用与 R&D 团队知识创造一元线性回归分析表 .....	103
表 7.40	大数据应用与 R&D 团队知识社会化多元线性回归分析表 .....	104
表 7.41	大数据应用与 R&D 团队知识表出化多元线性回归分析表 .....	104
表 7.42	大数据应用与 R&D 团队知识联结化多元线性回归分析表 .....	105
表 7.43	大数据应用与 R&D 团队知识内在化多元线性回归分析表 .....	106
表 7.44	回归模型总体情况 .....	106
表 7.46	回归系数情况表 .....	107

## 图片目录

图 1.1	IT 技术研究发展脉络	3
图 1.2	研究技术路线图	10
图 2.1	数据、信息、知识和智能关系图	13
图 2.2	Nonaka&Takeuchi 知识创造过程 SECI 模型图	15
图 2.3	引入“Ba”的 SECI 模型	16
图 2.4	组织知识创造五阶段模型及促进因素	17
图 2.5	组织的知识创造螺旋	18
图 2.6	KIBS 与客户间知识产生和传播	24
图 2.7	大数据 5V 特征	33
图 2.8	大数据平台架构	34
图 3.1	软件开发瀑布模型	41
图 3.2	软件开发一般流程	41
图 3.3	软件重用流程图	44
图 3.4	软件开发团队常见组织结构图	45
图 3.5	中小企业软件 R&D 团队应用大数据过程	47
图 3.6	软件 R&D 团队大数据应用与知识创造关系	48
图 5.1	大数据应用维度模型	55
图 5.2	大数据应用对 R&D 团队知识创造影响研究模型图	59
图 6.1	问卷覆盖区域情况图	67
图 6.2	公司应用大数据为公司业务服务的时间情况图	68
图 6.3	公司业务范围情况图	68
图 6.4	研发投入占销售收入的比例情图	68
图 6.5	受教育程度情况图	69

## 图片目录（续）

图 6.6	岗位角色情况图 .....	69
图 7.1	大数据应用关系模型图 .....	95
图 7.2	知识创造关系模型图 .....	97

# 第一章 绪论

## 1.1 研究背景

### 1.1.1 社会发展背景

信息技术（简称 IT，英文全称 Information Technology）作为近代先进生产力，它可以从生产、营销、管理、计划、决策、控制和资源优化等方面提高企业效能，进而提高企业竞争力。随着当今计算机软硬件、网络通信、智能化等技术的快速发展和广泛应用，各类数据信息发生了爆炸性增长，导致大数据时代的来临。2008 年《Nature》杂志推出专刊《Big Data》后，“大数据”正式成为各行业科学研究的大舞台。2011 年《Science》杂志出版专刊《Dealing with Data》，标志了“大数据”时代的正式到来。此后，许多国家先后将“大数据”上升为国家战略。

2012 年 3 月，美国总统奥巴马宣布启动“大数据研究与开发计划”，自此“大数据”研究上升为美国国家最高战略。2015 年 3 月十二届全国人大三次会议上，李克强总理首次在政府工作报告中提出“互联网+”。“互联网+”的深层意义是通过传统制造与互联网深度融合，完成产业升级，以此提升全社会的创新能力和生产力。而“互联网+”四大核心技术即为移动互联网、云计算、大数据、物联网。2015 年 8 月，经李克强总理签批的《关于促进大数据发展的行动纲要》指出：“要深化大数据在各行业创新应用，使开放的大数据成为促进创业创新的新动力”。

大数据快速发展早已引发产业界关注，2012 年美国通用电气公司（General Electric Company，简称 GE）发布了《工业互联网：突破智慧和机器的界限》，首次提出“工业大数据”概念，引发产业界热烈讨论。大数据在中国发展非常快，从中国贵阳大数据交易所发布的《2015 年中国大数据交易白皮书》显示，2014 年，全球大数据市场规模达到约 3607.6 亿美元，实现 53.23% 增长，比 2013 年 57.63% 的增速略有回落，但快速增长态势不变。IUD 领导决策数据分析中心在《2016 中国大数据发展报告》中提出，中国在 2020 年将成世界第一数据资源大国。

大数据正为企业创造良好的发展机遇，对企业的预测、决策、商业模式变革和价值创造等产生深远的影响。美国全球三大零售商之一的沃尔玛公司（Wal-Mart Stores）在其官网 [www.Walmart.com](http://www.Walmart.com) 上设计了自动搜索引擎，并利用语义工具进行文本分析、机器学习和同义词挖掘等，使得沃尔玛的在线购物完成率提高了约 15%。2008 年初，阿里巴巴通过对客户行为数据的挖掘和统计分析发现，客户询盘数急剧下滑。同时，欧美地区的客户对中国在线购买量也在不断下降。因此他们提前了 6 个月的时间准确预测了世界经济的总体走势，从而规避了世界金融危机。再

如海尔集团通过 **SCRM**（英文全称 **social crm**，即社会化客户关系管理）会员的大数据平台，获取了大量客户信息，然后使用“look-like”模型把用户进行分类，最后应用语义分析工具分析客户需求，从而优化用户体验。

传统信息技术经过几十年的发展和应用，尽管也出现不少由于信息技术运用不当，从而导致企业受损甚至发展失败的负面案例，但如今其仍然发展成为企业发展核心竞争力之一。但随着计算机技术、网络技术、智能化技术的不断发展和提高，数据信息量呈现爆炸性增长，传统的信息技术已经不能胜任和适应现今的各项工作任务。为此，一种全新模式的信息技术出现，即数字技术（简称 **DT**）。马云曾说过，当今人类已经从传统的 **IT** 时代进入 **DT** 时代。这样的背景下，企业管理将会遇到诸多理论和实践问题，比如基于企业视角的大数据技术特点、如何应用大数据提高企业预测和决策能力以及应用大数据对企业管理问题产生怎样的影响和变化，这些问题都需要我们去加以研究解决的。

### **1.1.2 KIBS 企业和 R&D 团队重要地位及其知识创造的紧迫性**

近年来，服务业在发达国家中所占比重越来越大，服务业内部组成结构也发生了非常大变化，出现了一种以网络和信息技术为背景，智力资本为主要投入要素，服务成果具有高附加值、高赢利、专业性、知识性等特点的新型服务产业，如软件企业、网络信息业、金融业、网络专业咨询业等，这种企业称之为 **KIBS** 企业，即知识密集型服务企业（**Knowledge-intensive Business Service**）。

知识密集型服务依托信息技术和网络技术发展迅猛，**KIBS** 的发展水平已经成为一个国家衡量其社会经济、现代化水平的重要标志。近年来，中国的 **KIBS** 同样发展快速。据统计，中国服务业对 **GDP** 的贡献从 2003 年的 7% 升至 2015 年的超过 51.1%。其主要得益于我国产业结构升级和经济转型的不断深入，这一数据的变化，也标志了中国经济发展达到了新阶段。尽管服务业对 **GDP** 的增长贡献巨大，但中国的产业结构调整仍需一个较长的过程。中国国家信息中心首席经济师范剑平表示，中国仍处在工业化中后期。这就说明中国的服务业发展仍处于初级阶段。虽然知识型服务业随着服务业整体快速发展而发展，但其所占 **GDP** 比重仅为 2015 年的 15% 左右，处于一个较低的水平。其主要原因是大多 **KIBS** 的服务产品原创性不高、技术含量低，导致产品容易模仿，知识产权保护困难，竞争优势低。中国将来要让服务业成为整个国家经济发展的主导产业，必须发挥知识型服务业作为高端服务业来支撑经济的长期发展。可见，**KIBS** 的发展问题在国家战略层面具有相当重要的地位。

**KIBS** 企业在其知识服务产品的价值创造过程主要体现了知识创造。而 **KIBS** 企业中的知识创造活动则是由不同知识产区的工作团队来开展和实施，即工作团队是 **KIBS** 企业知识创造的基本组织单元。**R&D** 团队则是企业实施知识创造最重

要、最核心的工作团队。近年来，R&D 活动得到中国国家、政府、企业等不同层面的高度重视。2014 年中国的 R&D 经费支出同比增长 12.4%，全国专利、软件著作权等的申请数量明显上升。同时，R&D 相关研究也越来越受到学界的关注，特别是管理学领域。通过在中国知网检索，R&D 相关研究论文 2012 年比 2011 年提高了约 25%，达 4257 篇。截止 2016 年，每年均比 2012 年略有增长，但基本保持在 4300 至 4500 篇之间。这些数据反映，R&D 相关研究是当今学界研究的热点。

知识是企业重要的战略资源(Grant, 1996)，确定和开拓新知识的来源可以为 企业获得持续竞争优势提供基础 (Drucher, 1998)，因此知识创造则是提高企业核 心竞争力的关键。从 KIBS 企业在中国经济、社会发展的地位及 KIBS 企业和 R&D 团队关系上看，研究 KIBS 企业及其 R&D 团队相关问题具有极其重要的实践意义。 而从知识创造在企业及国家层面的重要作用，并结合中国社会、经济的发展形势 来看，研究 KIBS 企业及其 R&D 团队的知识创造具有实践意义和紧迫性。

## 1.2 研究现状及问题提出

信息技术的研究领域中，Banker, Kauffman (2004) 和 Sidorova (2008)通过对 ERP、DSS、CAD、Internet 等研究其主题发展规律后认为：一项新兴信息技术 的研究通常有相类似的发展脉络。其发展脉络如图 1.1 所示：

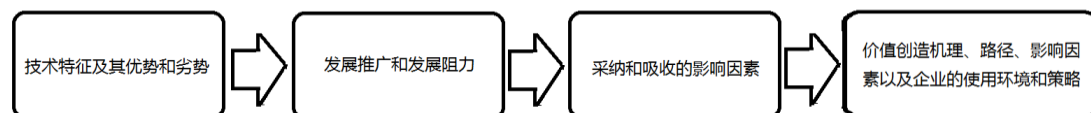


图 1.1 IT 技术研究发展脉络

尽管以上观点仍然缺乏严谨性，但它仍然可以为 我们预测新兴 IT 技术的研究 方向有明显的启示作用。

大数据是近几年来新兴的信息技术，大数据在经济管理上的研究现状，我们 以在中国关注度较高的数据库 CNKI(中国知网)、EBSCO、Google 学术等数据库， 以及国际上知识管理和信息管理领域中比较权威的期刊，如《Journal of Knowledge Management》、《Journal of Knowledge Management》、《The Journal of Information and Knowledge Management》、《Information Organization》等，来了解 情况，对近 5 年（2012—2016）大数据与经济 管理文献检索结果进行分析，发现 其研究范围主要集中在第一（技术特征及其优势和劣势）和第二阶段（发展推广 和发展阻力）。对第三阶段（采纳和吸收的影响因素）和第四阶段（价值创造机理、 路径和影响因素以及企业使用环境和策略）的研究则非常少，甚至无人涉足。

对于信息技术与知识创造的关系研究，Davenport&Prusak(1998)认为信息技术 是知识创新重要因素。Lee&Choi(2003)则进一步提出了二者的具体关系，他们认为

知识创造是知识实现价值的重要环节，各种形式的信息技术能够促进个体与组织对知识的收取、获取和交流，从而帮助他们创造新的知识。Scott(1998)通过研究认为，信息技术同时对显性知识和隐性知识的创造均有促进作用。可见，在现有的研究中，信息技术与企业知识创造关系紧密已是学界的共识。通过对近十年来关于信息技术与知识创造的研究文献查阅和分析发现，前期的研究成果大多基于传统的信息技术，如 ERP、DB、IS、Internet 等。而对新兴信息技术的研究较少，如大数据、智能化技术、移动通信、云计算等。经对已有文献的归纳总结，基于大数据应用和知识创造的研究还存在以下不足和问题。

1. 从企业应用视角看大数据的技术特性目前没有相关研究。
2. 大数据应用对知识创造有怎样的影响，目前相关研究还是空白。
3. 对大数据应用的相关研究缺乏基于统计分析的实证研究。

KIBS 企业自 20 世纪 90 年代起就成为了学界的研究热点。比如 Wood(1993)首次用知识密集等专业术语来描述 KIBS 企业典型特征。O'Farrell&Moffat(1995)定义了 KIBS。中国学者张方华和朱朝辉(2003)提出知识型企业知识创造的影响因素。魏江等(2008)则对 KIBS 企业和客户关系进行了研究，并提出 KIBS 企业和客户互动的三个阶段。而在研究 KIBS 企业知识产区问题时，团队和 R&D 团队则成为研究热点。比如 Fern(1982)就提出了团队是用来协同产生知识的一种相互讨论式的小组。Hedlund(1994)进一步对团队进行了定义，并指出团队是组织知识创造和知识转化的最基本单位。Madhavan&Grower(1998)则在研究中指出，团队活动中，R&D 活动其实就是一连串知识创造的过程。Lovelace(2007)认为，R&D 团队的知识创造活动是企业创新绩效的关键内容。但知识创新活动必须符合性能、预算、进度等预期。

从现有 KIBS 企业中 R&D 团队知识创造相关研究看，结合信息技术开展的研究非常少，而结合新兴信息技术开展研究的则更是凤毛麟角。

### 1.3 相关概念界定

#### 1. 知识

最早研究知识内涵是德国著名哲学家 Friedrich Nietzsche。他于 1882 年提出知识是人类一种基本需要，人类为了了解和改造身边的世界，就把未知的事物转变为已知事物。Drucker(1993)认为，知识是可以改变事物或人的信息，并最终成为个人、组织决策的前提、基础和能力。Nonaka 等(1995)则认为，知识是个人信念向“真实”方向验证的动态过程，他突出了个人和知识间的密切关系和动态变化。Wiig(1997)则认为，知识是一种资源，但相对其它资源来说，知识是无形资源，很难进行衡量和评价，知识可以在使用中加以处理，同时也可在处理中使用。



随着通信技术和信息技术的发展，信息被纳入到知识的研究中来。在 IS（信息系统）领域，诸多学者通过辨识数据、信息和知识三者的区别来定义知识。比如 Ackoff (1989)就提出 DIKW 等级模型，即把人类智慧划分为数据、信息、知识、智慧四个等级。同时他把数据定义为是对事实或观察的描述，是对客观事物的逻辑归纳，是用于表示客观事物的未经加工的原始素材。数据本身没有任何意义，也不反映与其它事物的关系。信息则定义为有一定关联关系和意义，能回答“who”、“what”、“where”、“when”等问题。知识定义为是对信息恰当的收集、存储、处理和应用，能回答“how”的问题。智慧是根据意识的各个层次，特别是人类道德、论理等来辨别或判断正确或者错误的过程。在管理学研究领域，信息被引入到知识的内涵中。如 Murrav 认为：“当信息被转化成一种可以获取有效行动的能力时，他就成为了知识”。Den Hertog 和 Huizenga (1997) 则把知识定义为是对信息的搜集整理，以实现某一功能或完成某项任务。Weggeman(1997)则给知识定义进行了补充:知识是个人的能力，应被看是信息、经验、技能和态度的共同产物。Ron Johnston&Rolf Blumentritt (1999) 通过信息-循环模型解释了信息和知识之间的关系和作用。他们认为，信息是通过知识来表达，信息对知识创造起到支持作用，但不能说明信息就是知识。信息要转变为知识，需要通过人的智力系统加以组织处理。Bhatt (2000)则认为，没有人的智力活动，信息或数据不能成为知识。

通过分析已有知识定义，本文定义知识为个人大脑对信息的处理，是与事实、过程、概念、解释、观点、观察和判断相关的个人的信息，知识是有意义的信息。

## 2.知识创造

学界对知识创造内涵界定主要有三种主流观点：一是“存量观”，其观点主要注重创造结果。如其代表人物 Davenport&Prusak(1998)认为，知识创造是企业通过开展一系列具有创新积极的活动，以此来增加企业的知识存量。二是“流量观”，该观点则注重创造过程。如其代表人物 Nonaka& Takeuchi 在 Polanyi(1967)提出知识可划分为隐性知识和显性知识的基础上，在 1995 年发表的“知识创造型企业”首次提出知识创造概念。他们认为，组织知识创造是指组织在企业内部或企业之间创造新显性知识和隐性知识的持续过程。这一概念的提出以及对知识创造过程的研究为知识创造理论奠定了理论基础。三是“协作观”，其重点关注知识或知识个体间的相互协作。如 Krogh(1998) 提出了知识创造是由许多不同知识背景的个体，他们通过相互协作来加快创造各种显性知识和隐性知识的过程。我国学者周小虎和陈传明(2004)也提出了，知识创造过程实质上就是对各种不同领域的知识进行相互协调与整合的过程。企业行动者的知识协调与整合能力直接决定着企业的创造活动的可能性和成功的可能性。实际上上述观点是可以相互补充。

通过剖析知识创造过程中的具体活动，Nonaka (1996) 将知识创造划分成基于旧知识的扩展型知识创造和开拓型知识创造两类。基于旧知识的扩展型知识创

造是指在现有的知识领域和知识存量下，增加知识的深度和广度，再通过对知识的提炼来发现其他相关知识。而开拓型知识创造则是指创造出新的知识领域，包括产生新的数据、新的信息、新的显性知识和隐性知识。而开拓型知识创造活动还可以细分为探索式知识创造和挖掘式知识创造。

通过对已有研究的了解和分析，并结合本论文的研究内容，我们对知识创造定义为：知识创造是组织通过信息整合产生新知识的过程，这一过程属于开拓型知识创造的活动过程。

### 3.信息技术和大数据

#### (1) 信息技术定义

1958年 Harold Leavitt&Thomas Whisler 首次提出信息技术一词。从现有的文献看，信息技术（英文全称为 Information Technology，简写 IT）主要从两大视角进行定义。一是通过信息技术具有的各种信息处理功能来定义。如 ISO 和 IEC 两个国际组织就将信息技术定义为：对不同信息进行收集、描述、表示、处理、存储、传输、安全、管理、组织等而应用的系统、工具、规范、设计和开发。达文波特也定义信息技术为获取、存贮、传递、处理、分析和使用信息的标准化技术。二是根据信息技术的组成内容进行定义。如 Harold Leavitt&Thomas Whisler(1958) 就将信息技术描述为：应用计算机处理大量信息，并运用数学方法对问题进行分析和决策，模拟高级别的思维活动。Leslie Willcocks(1996)则定义信息技术为硬件、软件、通信技术和信息处理技术。

通过对已有研究的归纳分析，并结合论文研究内容，本文对信息技术定义为：用于管理和处理数据、信息、知识所采用的各种技术总称。它包含了计算机硬软件技术、数据信息处理技术、通信技术、智能化技术、传感技术、控制技术。

#### (2) 大数据定义

随着信息技术的快速发展，各类数据发生爆炸性增长，传统的信息技术已无法满足海量数据的处理和各类服务的需要，大数据概念由此产生。目前，国际社会对大数据还没有一个统一定义。据百度统计，当今全世界共有 100 多种大数据的定义。2008 年《自然》杂志(Nature) 发表的大数据专刊中，大数据被定义为“代表着人类认知过程的进步，数据集的规模是无法在可容忍的时间内用目前的技术、方法和理论去获取、管理、处理的数据”。麦肯锡公司(McKinsey, 2011) 发布的研究报告认为，大数据是指其大小超出了典型数据库软件的采集、储存、管理和分析等能力的数据集。维克托·迈尔-舍恩伯格及肯尼斯·库克耶(2013)指出，大数据是不用随机分析法(抽样调查)这样的捷径，而采用所有数据进行分析处理。百度则给出大数据是指无法在可承受的时间范围内用常规软件工具进行捕捉、管理和处理的数据集合。“大数据”是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

表 1.1 大数据典型定义

序号	作者	年份	大数据定义
1	《Nature》杂志	2008	代表着人类认知过程的进步，数据集的规模是无法在可容忍的时间内用目前的技术、方法和理论去获取、管理、处理的数据
2	麦肯锡公司 (McKinsey)	2011	大小超出了典型数据库软件的采集、储存、管理和分析等能力的数据集。
3	维克托·迈尔-舍恩伯格及肯尼斯·库克耶	2013	大数据是不用随机分析法（抽样调查）这样的捷径，而采用所有数据进行分析处理。
4	百度		无法在可承受的时间范围内用常规软件工具进行捕捉、管理和处理的数据集合。“大数据”是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

通过对不同大数据定义的归纳分析，本文将大数据定义为：无法在可承受时间范围内用传统信息技术进行捕捉、管理和处理的数据集合。这些数据集合可以通过整理、挖掘、识别等方法变为有特定价值的信息。

#### 4.KIBS 企业和 R&D 团队

##### (1) KIBS 企业定义

KIBS 企业至今没有一个统一的定义。但鉴于知识密集型服务业在国家经济发展中的重要地位，因此学界、各类机构、组织乃至政府均有定义。学界中的定义，比如 Miles(2005)认为 KIBS 企业至少有以下三个特征：①须为私营企业或组织。②企业创造价值依靠专业知识和专家。③提供知识产品及其后续服务。Hipp 则认为，KIBS 企业是客户对象为制造或服务型的企业，并有丰富知识来源的企业。Dathe&Schmid(2000)，则认为，员工受教育水平、过程和产品创新程度均超过行业平均值的服务型企业才是 KIBS 企业。机构或组织的定义，如 OECD（世界经合组织）认为，KIBS 企业是提供金融、保险、专业科学技术等服务的私营企业。政府层面定义，如中国国务院发展研究中心定义 KIBS 为，获取、处理、分析、研究、存储各类信息，并将其转化为有用的知识，为不同用户提供其所需信息资源和信息管理服务的行业。其服务类型包括技术、咨询和电子商务等三类服务。而技术服务则具体包括信息技术、软件开发与信息资源管理、专利等服务。咨询服务则具体包括金融、保险、投资、创业、管理咨询或代理服务。电子商务服务则同时提供信息处理平台、信息咨询等服务。因此，KIBS 是一种混合知识服务。

结合论文研究内容，本文采用中国国务院发展研究中心对 KIBS 的定义。

##### (2) R&D 团队定义

团队(Workteam)作为组织创造价值的核心单位,它有着极其重要的地位。因此,团队也成为了组织理论研究领域近年来关注的热点之一。在现有团队理论研究中,不同视角有不同的定义,如 Homstrom(1987)从团队组成和目标视角定义,团队是由一组不同知识背景的代理人组成,他们独立完成任务并选择努力方向,但他们却创造出一个共同的产出。每个代理人对产出的边际贡献依赖于其他代理人的努力,不可独立观测。而 Brauner(2000)则从信息技术知识创造定义团队,他认为团队实际上可以被看作是一个信息处理的特殊单位。团队内部对信息同样进行编码、储存、处理、分析与检索,团队成员之间通过相互往来实施知识交换,并达到创造新知识和想法。

R&D(research and development,即研究与开发的简称)活动是一种探索开创性的组织活动。联合国教科文组织对 R&D 定义为增加知识总量,其包括人类社会文化方面的知识,并运用这些知识创造新应用的系统创造性工作。R&D 团队与其他团队最主要的区别是:R&D 团队是一个知识密集型的团队,成员的知识背景与水平、知识学习、知识存量、知识吸纳及知识共享等均直接决定 R&D 团队的绩效。可见,R&D 团队是团队中较为特殊的团队,其特殊性就是以知识为背景。

结合团队相关理论研究和 R&D 的内涵,本文定义 R&D 团队为:根据研发活动需要,由不同知识背景或工作领域的人员组建而成的知识型团队,通过成员间有效交往、协作、学习、激励等来完成团队共同目标,并产出有价值的知识产品。

## 5.软件企业相关概念界定及说明

### (1) 软件企业概念界定

中国信息产业部对软件企业作如下界定:软件企业即以开发、研究、经营、销售软件产品或软件服务为主的高新技术组织形态。本概念可进一步解释为:软件企业以计算机软件开发、研究、系统集成、应用服务、技术服务等为主要经营收入,同时至少有一种以上由本企业开发,并拥有知识产权的软件产品以及具备资质等级的企业。这类企业不仅仅是软件程序代码的编写,它包含了整个软件工程的研发周期以及软件产品的营销服务等。

### (2) 知识型员工、企业人员及软件研发团队的界定

知识型员工是指企业中有较高的文化素质,并具备较强的学习和创新知识的能力,能较好利用专业知识提高劳力价值的脑力劳动者。企业人员指从事软件产品开发和技术服务的知识型员工,其占企业职工总数的比重较高。软件研发团队是指企业因软件研发需要,将部分企业人员组织起来,形成一支软件研发技术队伍。该团队包含多种领域专业技术的企业人员。

### (3) 中小型软件企业的界定

本论文对中小软件企业的界定主要根据中国工业和信息化部、国家统计局、国家发展改革委、财政部在《关于印发中小企业划型标准规定的通知》对软件和

信息技术服务业所划分的标准来定义，即员工人数 $\geq 300$ 人且年销售量 $\geq 1$ 亿元的为大型企业； $300 >$ 员工人数 $\geq 100$ 人且 $1 >$ 年销售量 $\geq 1000$ 万元的为中型企业； $100 >$ 员工人数 $\geq 10$ 人且 $1000 >$ 年销售量 $\geq 50$ 万元的为中型企业； $10 >$ 员工人数且 $50 >$ 年销售量的为微型企业。因此，本文中小型软件企业界定为： $300 >$ 员工人数 $\geq 10$ 人且 $1000 >$ 年销售量 $\geq 50$ 万元。

## 1.4 研究内容

根据大数据和 KIBS 企业的发展背景,R&D 团队在 KIBS 企业中的重要地位,以及 KIBS 企业和 R&D 团队的知识密集型特性,同时结合对部分中小软件企业的调查研究和信息技术专家的预访谈,本论文确定研究内容为:通过整理归纳大数据的企业应用特性,以知识创造理论为基础,以中小型软件企业为例,分析、推导和实证检验大数据应用对 KIBS 企业中 R&D 团队实施知识创造的影响关系,最后通过中小型企软件企业案例进一步检验实证结论的正确性。

根据上述思路,本文所要研究的主要内容包括以下几个方面:

(1) 通过相关文献,对大数据技术特性、中小型软件企业应用大数据过程和大数据服务模式进行梳理。并从大数据企业应用视角,归纳出大数据技术应用特。

(2) 通过中小型软件企业的案例,分析研究大数据应用下,知识创造过程中所发生的变化,并提出相应观点。

(3) 根据对相关理论和现有研究成果进行梳理,总结归纳出大数据应用和 KIBS 企业中 R&D 团队知识创造的维度,分析大数据应用特性对 KIBS 企业中 R&D 团队知识创造的影响,提出研究假设和建立理论研究模型,最后进行实证研究。

## 1.5 研究方法及技术路线

### 1.5.1 研究方法

为了研究大数据应用与 KIBS 企业中 R&D 团队知识创造的关系,本论文采用了理论探索和实证研究相结合的研究方法,以中小型软件企业为例,对大数据企事业应用和 KIBS 企业中 R&D 团队的知识创造活动进行深入研究。具体的研究方法有:文献研究、预访谈和问卷调查、统计分析和案例分析等。

(1) 文献研究。对中国及其他国家关于大数据应用、KIBS 企业、R&D 团队、知识创造活动、中小型软件企业等理论研究成果进行梳理、归纳和比较,尽量了解与本研究相关前沿理论研究及其进展,以此提出论文的理论框架和研究假设。

(2) 预访谈和问卷调查。预访谈的目的是初步检验本研究的构想,再借鉴中国及其他国家与本研究相关的成熟量表,设计出包含本研究所有变量的调查问卷,

并发放问卷收集数据，用以检验本研究所提出的理论研究假设。

(3) 案例分析。本论文将通过案例分析来研究大数据应用对 R&D 团队知识创造过程的影响，从而发现大数据应用下，R&D 团队知识创造过程所发生的变化。

(4) 统计分析。统计分析是通过收集到的样本数据来检验论文所提出的理论研究假设。即根据论文研究需求去收集样本数据后，利用统计分析方法进行处理分析，以此检验论文研究中的各个测量量表，从而达到验证论文提出的理论研究假设。本研究统计分析工具主要采用 SPSS19.0 统计分析软件。

### 1.5.2 技术路线

本研究从分析、总结、归纳已有的相关理论研究，构建研究量变及其理论研究模型，通过问卷调查、统计分析、案例分析来检验理论研究模型的合理性和实践性。本文的技术路线如图 1.2 所示。

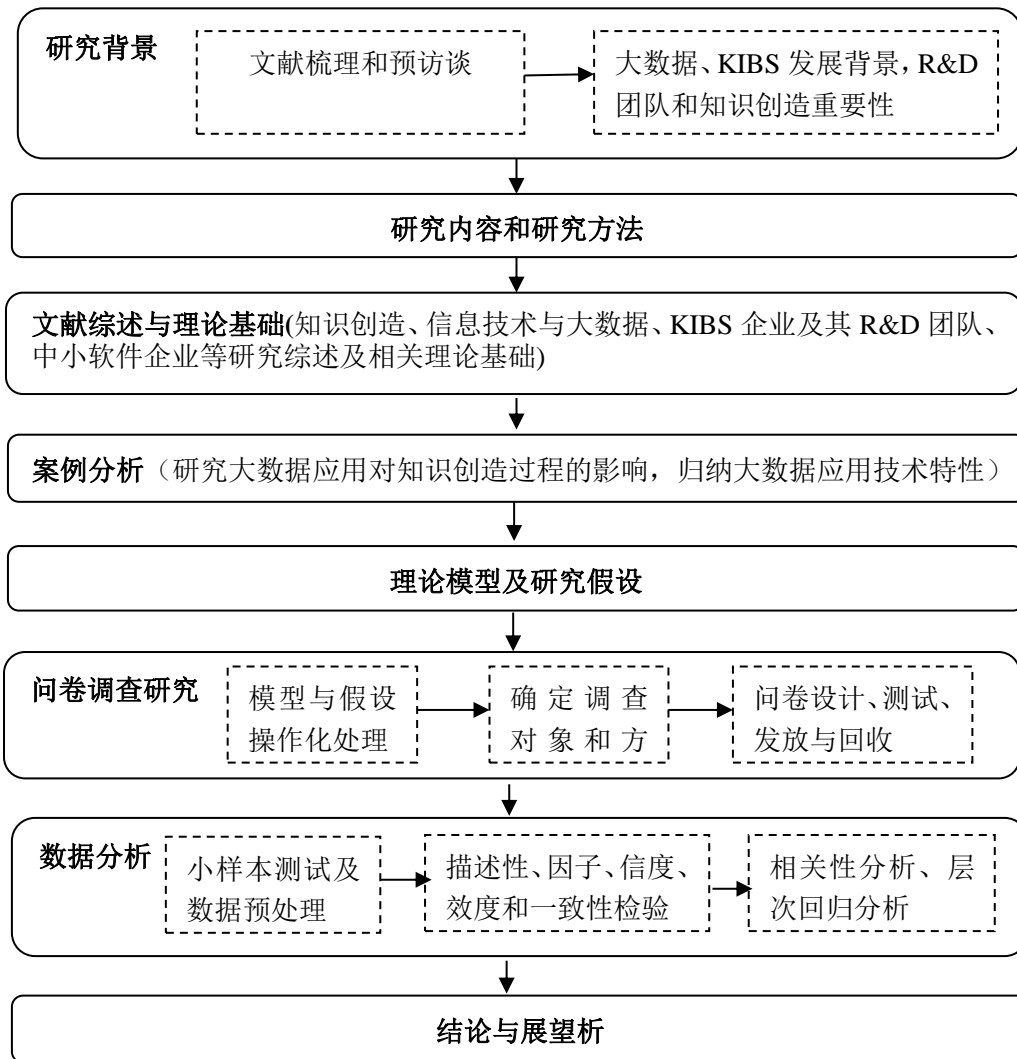


图 1.2 研究技术路线图

## 1.6 论文研究框架

根据本研究技术路线，本论文内容共有八章，每章内容大概如下：

第一章为绪论。本章主要介绍论文选题的理论研究背景和社会发展背景，提出论文主题研究的“迫切性”和意义。之后，对论文研究涉及的相关概念进行界定，同时阐述了论文的研究内容、研究方法、研究技术路线和论文研究框架，最后提出论文可能的创新点。

第二章为文献综述。对知识创造、信息技术与大数据应用、KIBS 企业及其 R&D 团队、中小型软件企业等相关研究文献进行梳理，对相关的研究成果进行分析和评述，并从中发现当前研究所存在的主要问题，同时提出本论文的研究切入点。

第三章为理论基础。通过知识创造理论、团队理论和信息技术理论等理论基础进行梳理和剖析，为本论文研究构思和框架找到作为依据和支撑的理论基础。

第四章为软件 R&D 团队大数据应用及其知识创造过程。分析了中小软件企业应用大数据面临的问题，并提出相应策略。介绍软件工程的大数据应用以及软件 R&D 团队大数据应用过程，并分析了软件 R&D 团队大数据应用和知识创造的关系。运用 2 个案例，分析软件 R&D 团队大数据应用知识创造过程发生的变化以及大数据应用对 R&D 团队知识创造产生的影响因素。

第五章为理论假设及研究模型。界定主要研究变量的概念，提出大数据与 KIBS 企业 R&D 团队知识创新之间影响关系的假设，并构建理论研究模型。

第六章为研究方法及问卷设计。首先，构建理论研究模型中变量的测量体系，通过访谈和专家意见修正问卷量表，设计出本研究完整的调查问卷。然后，选择研究样本发放和回收问卷，并收集数据样本，为检验假设的数据分析提供必须的数据。最后，对样本特征、数据分析工具和方法进行了描述说明。

第七章，数据分析与假设检验。进行小样本预测试和大样本测试，运用 SPSS21.0 统计分析软件对收集到的数据信息进行统计分析，检验本研究的假设。并对统计分析结果进一步的分析与讨论。

第八章，结论与展望。首先总结本论文的主要研究，并得出相关结论。其次与已有研究进行比较，对论文的研究结果作分析总结，并阐述本论文研究的新发现和主要贡献。最后分析本论文研究存在的不足，并提出未来的研究展望。

## 1.7 论文创新点

论文在归纳、整理、吸收中国及其他国家相关研究成果的基础上，以中国中小型软件企业为例，构建大数据应用对 KIBS 企业中 R&D 团队知识创造的影响研究模型，通过统计分析和案例分析方法，从企业应用视角研究了大数据应用与 R&D

团队知识创造之间的关系。本研究与之前的研究相比，有以下创新之处：

(1) 通过梳理大数据相关文献，从大数据的企业应用视角出发，分析并提出了大数据应用技术新特性，为大数据今后在管理学研究领域提供了新的研究视角和研究思路。

(2) 本论文通过中小型软件企业软件服务的大数据应用案例，从企业应用视角研究了大数据应用下，KIBS 企业 R&D 团队知识创造过程发生的变化，进一步丰富了 Nonaka 的 SECI 知识创造理论。

(3) 通过文献梳理和案例研究，分析了大数据应用对 R&D 团队知识创造产生的影响因素，为企业管理领域的大数据价值创造研究提供新的研究思路。同时也为企业大数据应用实践提供理论依据。

(4) 研究了大数据应用和 R&D 团队知识创造两个变量的维度，利用统计分析和案例分析方法，检验了两者关系，丰富了企业管理领域大数据研究方法。

大数据是一种新兴的信息技术，而现有的大数据相关研究文献缺乏统计分析的实证研究。随着近几年大数据技术的快速发展，应用大数据的企业越来越多，企业样本也越来越丰富。因此，本文从大数据应用和 R&D 团队知识创新两个变量的维度入手，利用数据分析方法检验了两者之间的关系，丰富了企业管理领域大数据的研究方法。

## 1.8 本章小结

本章阐述了论文的研究背景、研究现状、研究问题、相关概念界定、研究内容、研究方法、研究技术路线、论文研究框架及论文创新点。

首先，本章从中国及其他国家大数据、KIBS 企业的现实发展背景，以及 R&D 团队和知识创造的重要地位出发，结合知识创造理论、信息技术理论和团队理论，提出本论文研究的问题、内容和目的，同时分析了本论文研究的理论和实践意义。

其次，本章对论文所涉及的核心概念如知识创造、KIBS 企业、R&D 团队、信息技术、大数据、软件企业等进行了界定。

最后，本章对论文主要研究方法展开了说明，并给出了研究技术路线，论文研究框架，并对论文可能的创新点进行了说明。



## 第二章 文献综述

本章主要根据本论文所研究的内容，对中国及其他国家的文献进行回顾、梳理和述评。其主要从五个方面展开，即对知识创造、KIBS 企业和 R&D 团队、R&D 团队和知识创造、信息技术和大数据、信息技术和知识创造等方面的研究文献进行回顾、梳理和述评。本章的文献回顾和梳理将为论文最终提出理论研究模型及实证研究奠定理论基础。

### 2.1 知识创造文献综述

#### 2.1.1 数据、信息和知识的关系

知识早期被哲学家认为是“被验证过的信念”（柏拉图）和“经验的成果”（培根）。此后，管理学领域中知识则被认为是主客体相互统一、相互作用，并由人脑活动的产物。具有代表性的如 Allee (1997)提出知识具有“波粒二重性”。但随着社会的发展，时代的变迁，知识则被赋予更丰富的含义。

当今的经济管理学领域，学者在探讨知识问题前，首先将数据、信息、知识加以区分。最早区分数据、信息、知识的是 Ackoff，他于 1989 提出，根据人类发展的不同阶段，人类智力从低到高可划分四个等级，即数据、信息、知识和智慧。Nonaka 则于 1994 提出，数据、知识、信息最大区别是知识与人的信念和行为密切相关。Robert 认为数据是客观符号，信息是客观符号间的关联，知识是客观符号关联后的模式。Beckman(1999)认为数据、信息、知识有一定层次性，数据是指文本、事实、代码、图像、音频和视频等集合。信息则是有组织、结构、概括性的数据。而知识则是案例、规则、过程和模型等集合。中国学者朱祖平(2000)则认为，数据为信息的基础，信息组成部分也包括数据，数据必须经过处理和建立关系，并赋予明确意义后才能变成信息。知识则是在信息的基础上，通过推理、验证后得到的系统化规律、概念和经验。其关系如下图所示：

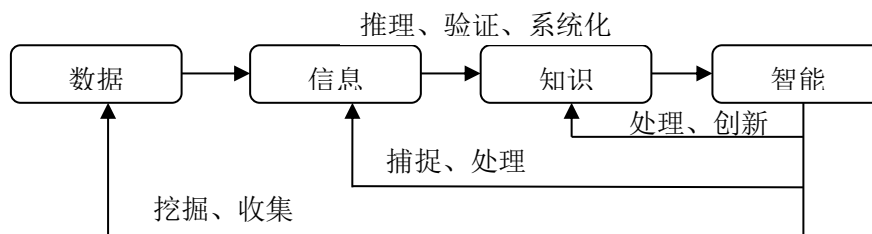


图 2.1 数据、信息、知识和智能关系图

来源于：朱相平（2000）

通过对上述研究文献的梳理和分析，本论文认为数据和信息”都是知识的组成部分，是可以和个体分离，清晰化表述的知识。

## 2.1.2 知识创造相关综述

资源基础观认为，企业是一个具有独特异质资源和能力的组合，企业的竞争优势是企业自身拥有的资源。可见，在多变、复杂的环境中企业要生存和发展，必须创造出其他企业没有的资源，以确保其竞争优势。知识不仅像劳动力、资本、土地是一种生产资源，而且是具有更重要意义的资源（Drucker, 1993），知识属于企业重要的战略资源(Grant, 1996; Teece, 1998)。Toffler(1990)、Quinn 和 Rivoli(1991)、Spender(1996)认为，知识是保证组织持久竞争优势的关键因素。Drucher（1998）则认为，确定和开拓新知识的来源为企业获得持续竞争优势提供基础。可见，知识是企业获得竞争优势的重要资源，而知识创造则是提高企业核心竞争力的关键。

### 2.1.2.1 知识创造理论

#### 2.1.2.1.1 SECI 知识创造理论模型

1966 年英国哲学家 Polanyi 首次将知识划分为显性知识（ExPlicit Knowledge）和隐性知识（Tacit Knowledge）。以此为基础，被誉为知识管理拓荒者的日本学者 Nonaka 和 Takeuchi 在他们 1995 年出版的著作《创造知识的企业-日本企业持续新的动力》中，首次系统地提出了企业知识创造模型，即 SECI 模型。Nonaka 在该著作中指出，人类的知识是隐性知识和显性知识之间的社会化相互作用而创造和扩展出来的。在知识创造过程 SECI 模型中提出了显性知识和隐性知识相互转化的四个螺旋式转换过程，即社会化(Socialization)、表出化(Externalization)、联结化(Combination)、和内在化(Internalization)。SECI 知识创造理论模型实际上是基于组织可提供帮助的假设之下的。

**社会化模式：**该模式是共享心智的模式，是隐性知识到隐性知识转化的过程，即个体间体验、分享和创造隐性知识的过程。个人只有通过观察、学习、模仿获得并创造隐性知识。这一模式通常与团队活动、组织文化有关。

**表出化模式：**这一模式是从隐性知识到显性知识的过程，即个体通过语言、文字、概念、模型等形式将零散的隐性知识明晰化。在此过程中，个体将隐性知识在团队或组织内表述为能广泛接受和理解的显性知识，知识不但可被他人共享，还能成为创造新知识的基础。这一模式通常与共同思考、概念创造有关。

**联结化模式：**这一模式是从显性知识到显性知识的过程，即将不同显性知识单元进行联结与整合，并转化成更复杂和系统的显性知识加以利用。这一模式关键步骤包括知识的获取、归纳、整合、调整和扩散。这一模式计算机数据库中的

编码化和网络化可促进知识联结。可见该模式主要来源于信息处理。

**内在化模式：**这一模式是从显性知识到隐性知识的过程，即个人通过实践，将显性知识变为隐性知识的过程。这一模式中个人的实践经历起关键作用，需要个人在实践中学习，才能改善个人心智模式，掌握技术诀窍，从而形成有价值的知识。该模式与组织学习紧密相关。

知识创造就是隐性知识和显性知识持续的、动态的互动作用的过程，这种相互作用是由四种知识转换模式间的转化来推动的，这个转化由多个转化过程交替进化，从而构成一个螺旋上升的过程，这种过程我们称之为“知识螺旋”过程。

下图为 Nonaka&Takeuchi 知识创造过程 SECI 模型图：

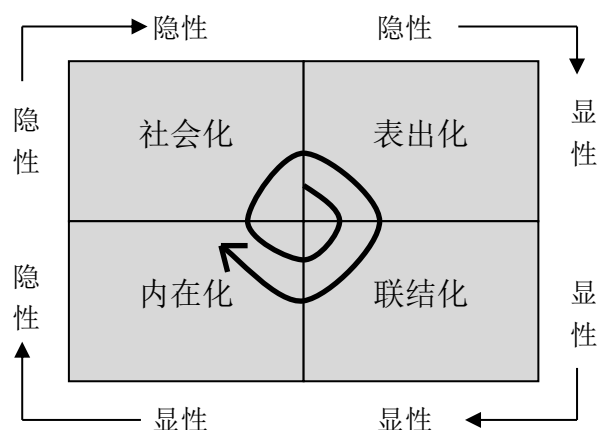


图 2.2 Nonaka&Takeuchi 知识创造过程 SECI 模型图

来源于：Nonaka&Takeuchi (1995)

### 2.1.2.1.2 知识创造场所“Ba”

1998 年 Nonaka 和 Takeuchi 又提出了 SECI 模型中知识创造背景“Ba”的概念。Nonaka&Takeuchi 认为，“Ba”是一种共享情境，是促进知识共享、知识创造和知识应用的背景。“Ba”非常关键，它为个人的知识化，进而为知识螺旋提供能量、质量和场所。

在知识创造过程中，个人不可能置身于背景之外，社会、文化、历史、技术环境等对个人都有重大影响，它是个人将信息释放为知识的场所。“Ba”是实体空间（比如办公室）、虚拟空间（比如互联网）、心智空间（比如共同理想）整合在一起形成一个整体。

Nonaka&Takeuchi 认为，“Ba”也分成四种类型，即创始“Ba”、对话“Ba”、系统“Ba”和实践“Ba”。共同化始于创始“Ba”，这个“Ba”是促进个体之间分享体验、感受、经历和心智模式的场所，它的特点是“面对面”沟通互动。表出化是通过对话“Ba”进行集体对话和思考，并运用比喻、启发、符号等将隐性知识明晰化，促成隐性知识的共享。与创始“Ba”相比，对话“Ba”更有意向和结

构化。系统“Ba”是为已经存在的显性知识提供联结化背景，比如以书面形式或者通过信息技术来传输显性知识。系统“Ba”是以群体间和事实性资料（如书面手册、模拟程序等）作为交互媒介的知识“Ba”。实践“Ba”则不同，它是以个体间和事实性资料作为交互媒介的知识“Ba”，它主要为知识内化提供背景。这背景就是通过实践中学习，来推动知识内化模式。

以下为引入“Ba”的SECI模型图：

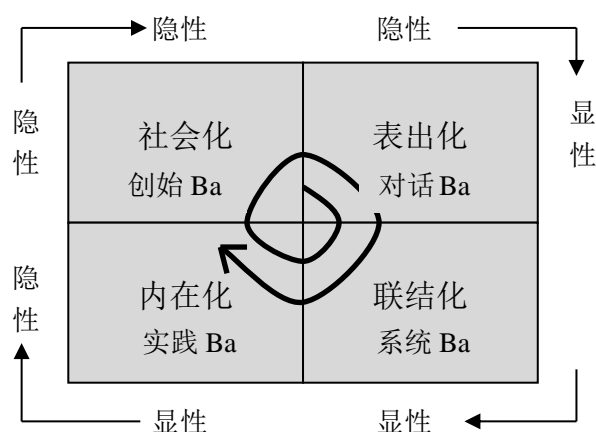


图 2.3 引入“Ba”的SECI模型

来源于：Nonaka(1998)

显然，信息技术是知识创造中一种典型、重要的“Ba”。Vaccaro 等人于 2009 年研究基于 SECI 模型，以虚拟技术为对象的组织内部知识创造过程发现，信息技术在共同化过程中起辅助作用，在表出化过程没有发现信息技术带来积极贡献，联结化过程中信息技术起到了支撑作用，内在化过程中部分信息技术可以促进个体的知识内化，同时也促进组织管理的规范化和知识传递。

由于 SECI 模型具有较强的解释力，因此随后知识创造的过程研究，大多是建立在 SECI 模型基础之上，如中国学者耿新在 SECI 模型的基础上提出了改进模型 IDE—SECI 模型。

### 2.1.2.1.3 组织知识创造新五个阶段及其促进因素

Nonaka&Takeuchi(1995)基于以上四种知识转化模式，提出了组织创造新知识的五个阶段模型，即共享隐性知识、创造概念、检验概念、建立原型和跨层次知识交流。此外，还提出了组织知识创造的促进条件，即意图(intention)、自我管理(autonomy)、波动和创造性混沌(fluctuation and Creative Chaos)、冗余(redundancy)、必要多样性(requisite variety)。以下为五阶段模型及促进因素图：

阶段一：共享隐性知识。组织不能创造知识，组织内的个体所具有的隐性知识是组织知识创造的基础。因此组织内知识共享是组织知识创造的关键。

阶段二：创造概念。通过集体对话交流隐性知识，引发集体思考，并将隐性

知识转化为明晰概念，最后用文字形式表述。

阶段三：检验概念。建立新概念后，组织需要对新概念进行检验。对企业而言，检验标准有：成本、盈利率、产品价值等。

阶段四：建立原型。将检验过的概念转化为有形、具体的模型，这个模型是将创造的新显性知识和现有的显性知识整合后产生。

阶段五：跨层次知识交流。该阶段目的是进行跨层次的知识共享，这一阶段是组织知识创造自我完善和提高的过程。它可以是组织内部，也可以是外部。

以下为组织知识创造五阶段模型及促进因素图：

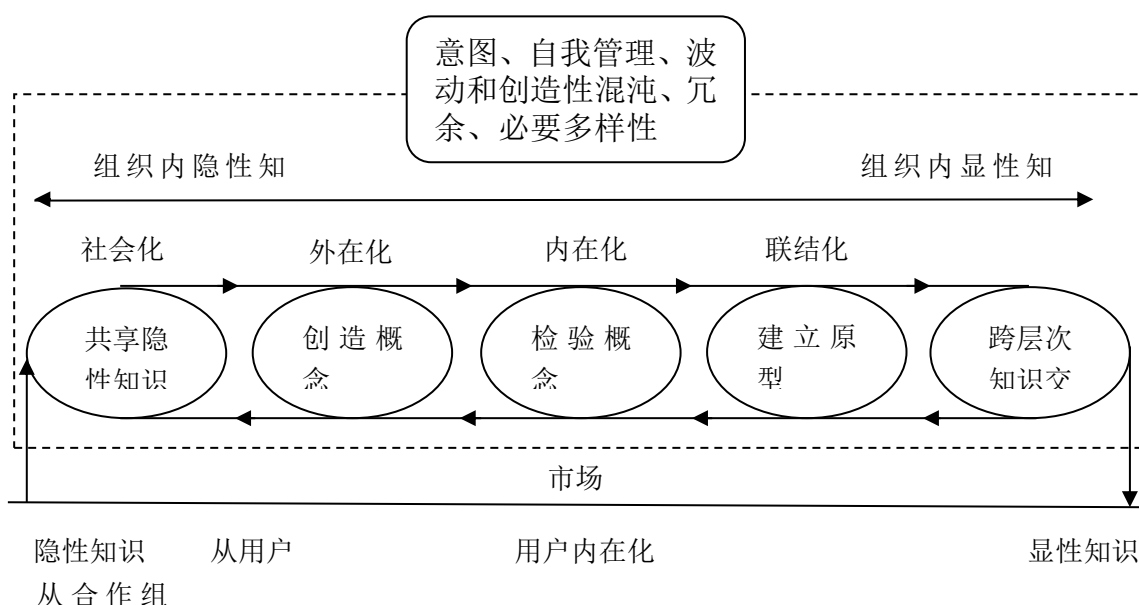


图 2.4 组织知识创造五阶段模型及促进因素

来源于：Nonaka&Takeuchi (2000)

#### 2.1.2.1.4 组织的知识创造螺旋

组织的知识创造过程可以从认识论和存在论两个维度解释。从认识论角度看，组织的知识基础是组织内部个体所具有的隐性知识，组织是通过调动个体，使个体利用自身积累的隐性知识在知识转换的四个模式中得以创造新的知识。从存在论角度看，新创造的知识不断地积累在组织内，这一知识产生的过程称为“知识螺旋”。“知识螺旋”始于个体，并随着个体间互动使范围不断扩大，乃至团队和组织的边界，并不断向前推进。以下为组织知识创造螺旋图：

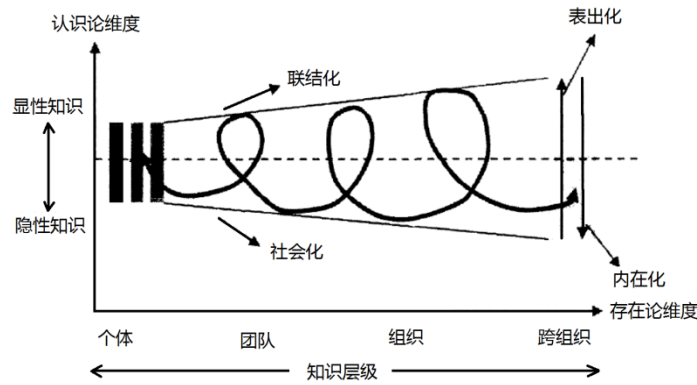


图 2.5 组织的知识创造螺旋

来源于: Nonaka&Takeuchi (2000)

### 2.1.2.2 知识创造研究

对于知识创造的研究, 管理学领域始于 20 世纪 90 年代中期。在这一时期, 学者们更多的是探索知识创造过程原理, 而对知识创造相关性研究, 如影响因素研究, 在近几年来才开始给予重视。纵观中国及其他国家对知识创造的研究, 均从两个方面展开, 即知识创造过程研究和知识创造相关性研究。为此, 本论文也将从这两方面对知识创造前期研究进行回顾、梳理和分析。

#### 2.1.2.2.1 知识创造过程研究

在 Nonaka 提出 SECI 理论后, 知识创造过程研究又相继出现了几种比较有代表性的理论, 如知识价值链理论、知识生态循环理论、社会学习周期理论等。

在知识价值链理论研究中, Holsapple&Singh(2001)首先提出了知识价值链概念, 并将知识价值链划分为知识获取、知识选择、知识生成、知识内化和知识外化五个过程。在此基础上, 中国学者芮明杰等(2004)将企业知识价值链主要环节改进为知识获取、知识融合、知识创造、知识保护和知识扩散等 6 个过程。

知识生态循环理论最具代表性的是 AAKM (美国知识管理协会) 把知识产生描述为其演化发展过程, 经历了从不可证实的知识转化为可以编码的有效的知识。中国学者吴洁&刘思峰 (2006) 基于复杂理论视角, 将热力学中的熵理论引进了知识创造过程的研究中来, 其主要对组织知识创造机理、关键、途径、机制进行了研究, 并建立了知识创造的数学模型。

社会学习周期理论具有代表性的则是 Boisot (2004)从信息学视角出发, 提出知识资产在组织体系中的扩散演化模型, 并指出知识资产的社会学习周期包括六个阶段, 即扫描、解决问题、抽象、扩散、吸收、影响。此外, Fong( 2003)针对多功能团队提出其知识创造过程包括四个阶段, 即边界交叉、知识获取、知识产生、知识整合。在此过程中, 学习则贯穿始终。

我们将知识创造过程研究归纳如下表：

表 2.1 知识创造过程研究

序号	理论	学者	观点
1	SECI 理论	Nonaka&Takeuchi (1995)、(1998)	知识创造四个螺旋式转换过程，即社会化、表出化、联结化、和内在化
2	知识价值链理论	Holsapple&Singh (2001)	知识价值链划分为知识获取、知识选择、知识生成、知识内化和知识外化五个过程
3		芮明杰 (2004)	将企业知识价值链改进为知识获取、知识融合、知识创造、知识保护和知识扩散等 6 个过程。
4	知识生态循环理论	AAKM	把知识产生描述为其演化发展过程，经历了从不可证实的知识转化为可以编码的有效的知识。
5		吴洁&刘思峰 (2006)	从复杂理论视角将熵理论引入知识创造过程研究中，主要对组织知识创造机理、关键、途径、机制进行研究，并建立知识创造数学模型。
6	社会学习周期理论	Boisot (2004)	从信息学视角出发，提出知识资产在组织体系中的扩散演化模型，并指出知识资产的社会学习周期包括六个阶段，即扫描、解决问题、抽象、扩散、吸收、影响。
7	其他	Fong (2003)	针对多功能团队提出其知识创造过程包括四个阶段，即边界交叉、知识获取、知识产生、知识整合。在此过程中，学习则贯穿始终。

来源于：本研究整理

#### 2.1.2.2.2 知识创造过程研究评述

综上所述，在知识创造过程研究中，Nonaka 的 SECI 模型最具解释力，其后学者的大多数研究均基于对 SECI 模型的改进。对 SECI 模型改进主要从两方面展开，一是引入外部知识，拓展知识转化。二是引入价值链，使知识转化变成知识创造过程的组成部分。而这些研究不足的是，没有结合具体行业甚至具体企业的业务流程进行知识创造过程研究，导致理论缺乏实践指导，操作性不强。

#### 2.1.2.2.3 知识创造的相关性研究

学者们在探索知识创造过程原理中发现，知识创造与环境密不可分。任何团体或组织的知识创造都离不开个体，但个体的知识创造也离不开团体或组织环境。Nonaka 于 1998 年提出 ‘Ba’ 的概念后，管理学界均以 ‘Ba’ 来研究组织环境对知识创造的影响作用。

Nonaka 本人于 2000 年提出了 5 种对知识创造有利的情境，即意图、自主权、波动、重复和多样性。Apostolouhe & Mentzas, Delphi Group、Shin,et al、Liebowitz

& Megbolugbe 等学者均认为知识管理中的文化和组织对知识创造具有促进作用。The Mutual Group 等则强调通过个体学习、相互交流、使用知识库学习等多种学习方式促进新知识的创造。熊德勇和金生(2004)提出了影响知识创造的 6 个因素,即战略、知识源、协同机制、知识技术、知识环境、合作愿望等。并研究了这 6 个因素在各类“Ba”中的作用和机制,使得知识创造的 SECI 模型的内部运行过程变得更清晰。韩智慧和李南(2004)认为从组织因素方面考虑,影响知识创造的因素应该包含战略制定、组织结构、组织文化、组织机制以及信息技术等五个方面内容。Smith K G 等人(2005)认为,学习在知识创造中非常重要,员工个人的知识储备和有效的学习对知识创造有促进作用。同时组织对关键人员能力的识别,并给予激励也非常重要,它能促使人与知识更好地融合,进而促进知识创造。杨德林(2005)在研究 R&D 项目组知识创造中提出,影响 R&D 项目组知识创造的因素有内部因素、管理因素、环境因素、知识库以及外部因素等。王重鸣和田茂利(2006)通过研究技术创业企业后提出,能力、性格、信息冗余和领导也对知识创造产生影响。马淑文(2007)提出,知识场为知识创造提供了知识转换、提炼和升华的环境,增强人们参与知识创造的积极性,直接影响知识创造的效率。张树娟(2008)通过研究研发团队知识创造影响因素时认为,研发团队知识创造影响因素主要包括团队人数和团队内部的互动模式。同时也提出,智力资本转换需要激励,这需要制度的保障。A.T. Arkan (2009)则把集群作为增进知识创造的“场”,并研究集群里的企业间知识交换的形成、能动性及其效率提升等影响因素,并得出了不同集群其知识创造结果有差异。V.Lee 等(2012)则认为,由于企业所处环境和所面对问题的复杂性均不同,促使企业形成有别于其他企业的学习能力,创造新知识,以保持竞争优势。G.V. Krogh 等(2012)则通过对国际化营运企业的组织结构进行研究,发现这类企业使用产品群和地理区域相互融的矩阵型组织,从而促进不在同一区域的员工之间的知识顺畅而快速地流动,从而促进组织知识创造。冯长利等(2016)采用 Fuzzy DEMATEL 方法研究供应链企业间知识创造影响因素后提出,预期的知识价值和合作意愿和倾向是知识创造的影响因素。张伟(2016)通过对资源型产业链知识创造影响因素研究分析后认为,知识传输、知识吸收、企业间知识共享均对资源型产业链的知识创造产生影响。

从知识创造相关性研究的前期成果看,其研究成果非常丰富。通过分析和参考相关文献,我们可以将与知识创造有相关性的因素归纳成五个方面,即内部因素、管理因素、环境因素、知识库以及外部因素等。本论文根据这五方面因素对以上研究文献进一步梳理,具体见下表:



表 2.2 知识创造影响因素

内部因素	管理因素	环境因素	知识库	外部因素
团队人数 团队互动模式 个体知识储备 个体学习 知识吸收 知识传输 企业内知识共享	激励机制 制度保障 对人员能力识别 协同机制	战略制定 组织结构 组织文化 组织机制 信息技术 知识环境、	知识源 知识库管理 知识技术 知识地图	预期的知识价值 合作意愿和倾向 外部信息 企业间知识共享

来源于：本研究整理

#### 2.1.2.2.4 知识创造相关性研究评述

从以上前期研究可见，目前知识创造的相关性研究大多以 Nonaka & Takeuchi 提出的“场”理论作为基础进行研究的。知识创造相关性研究前期成果虽然丰富，但仍然存在以下不足：（1）知识创造的环境影响因素没有结合具体企业或业务流程进行深入分析研究；（2）面对信息技术的快速发展，未对所出现的意义重大并影响深远的新技术可能产生的影响进行深入研究。

## 2.2 KIBS 企业文献综述

### 2.2.1 KIBS 企业

20 世纪 70 年代, Browning&Singelman(1975)就提出了生产性服务业这一概念。20 世纪 80 年代, 西方发达国家服务行业逐步取代制造业成为经济发展动力, Wood 在 1993 年调研这一时期英国服务业的发展情况时, 首次提出使用知识密集这一专业术语描述 KIBS 企业的主要特征。尽管 80 年代服务行业已经得到快速发展, 但并未得到学界的重视。随着信息技术的快速发展, 知识资源的迅猛扩张, 以知识资源作为生存依赖的知识密集型服务业从传统服务业中分离出来, 并在行业中显示了极强的生命力, 直至 20 世纪 90 年代, 知识密集型服务业才受学界的重视。

#### 2.2.1.1 KIBS 的研究现状

通过搜索相关文献, 并对其梳理和归纳, 将 KIBS 的主要研究分为五个方面, 即知识层面、创新层面、区域作用。

##### (1) 知识层面

KIBS 研究初期主要研究其知识属性、特征、作用和分类的研究。如 Wood (1993)、Moffat (1995)、Antonelli (1999)等均强调 KIBS 的知识性, 知识是 KIBS

的重要特征，同时也是第一要素。Miles 等(1995)对从知识角度对 KIBS 进行了分类，把 KIBS 分为专业型(称 P-KIBS，如咨询、法律、会计等)和技术型(称 T-KIBS，如软件、工程、研发等)。

此后的研究则包括 KIBS 企业与顾客间、知识扩散、知识管理与知识保护。Nonaka&Takeuchi(1995)、Hertog(2000)把知识创造 SECI 模型用于对 KIBS 的企业与顾客互动过程的研究。Larsen(2001)提出了 KIBS 的知识扩散观点，以此为基础 Muller&Zenker (2001)提出知识的创造与扩散的假设，Wood (2002)等则实证了这一假设。Amara&Landry (2008)提出了 KIBS 企业知识产权保护的机制。Dai&Song (2009)则提出了 KIBS 的知识管理二维结构。

### (2) 创新层面

Wood 于 1993 年指出 KIBS 能促进其客户进行创新。此后，Moffat(1995) 和 Hill(1997)提出 KIBS 企业与客户可相互促进创新。Miles(1995)、Hertog(2000)、高波(2008)等认为 KIBS 在整个创新系统内具有传递、源头和推动作用。张玉珍(2006)提出 KIBS 在知识创新系统中具有知识获取、整合、扩散和再生产等功能。赵炎&王晨(2009)对 KIBS 集群化作了分析梳理，认为 KIBS 在集群创新过程中起到化合、修正、搭桥作用。吕民乐&金妍(2016)则实证了国家层面知识密集型服务业对中国制造业创新活动产生积极作用，而行业层面上则有差异。

Miles(2005)、Wood(2005)通过研究后指出，对 KIBS 的研究应该从其创新功能与作用，上升到国家层面的创新体系。此后，中国学者也结合中国实际情况做了许多相关研究工作，如刘顺忠(2004)从宏观角度分析后提出，密集型服务业有利于国家的创新体系，同时结合中国实际情况，指出中国 KIBS 快速发展的成功因素和阻碍因素。此后，魏江等(2005)通过比较中国和欧盟的 KIBS 在国家创新系统中的作用，提出了中国的 KIBS 发展所存在的问题。

### (3) 区域作用

对于 KIBS 的区域作用，Wood(1993)就认为 KIBS 存在地理局限性，同时指出企业的选址非常重要。此后 Moffat 等(1995)认为，KIBS 的空间距离可由信息技术解决，KIBS 的区位不对其创新活动产生影响。Muller&Zenker (2001)提出，选址和区域创新系统间的距离对 KIBS 发展有影响。此后魏江(2004)、Wood(2005)实证证明，区域和国家创新系统对 KIBS 企业创新能力有显著影响。

## 2.2.1.2 KIBS 的分类与特征

### (1) KIBS 的分类

Miles 等(1995)基于知识的角度把 KIBS 企业进行了分类，一种为专业型 KIBS (P-KIBS)，这类企业为传统的专业服务行业，与技术相关性比较弱，如咨询、法律、金融等，另一种则为技术型的 KIBS(T-KIBS)，这类企业主要从事与信息通讯技术

相关的业务，与技术有较强的相关性，如软件、工程、研发等。但有些提供高知识服务，也拥有高素质员工和应用新技术的组织是否属于知识密集型服务业，在学界内没有形成统一认识，如教育机构、医疗机构、农业推广等。

中国学者结合本国的实际情况也对 KIBS 进行分类，如金雪军(2002)从知识角度对 KIBS 的种类进行了划分；魏江&陶颜(2007)则从服务生产方式角度，同时结合中国国民经济行业分类和国际标准产业分类，将 KIBS 划分成 4 大类，4 大类下又包含 14 小类。

## (2) KIBS 的特征

KIBS 的特征在学界内仍然有诸多争议，但从归纳文献来看，学者们观点一致的有两点，一是对专业知识要求高；二是以提供知识产品和服务。目前对 KIBS 的特征研究，在学界内得到广泛认可的是 Miles 于 2005 年提出的观点，他认为 KIBS 企业至少有三个特征：①须为私营企业或组织。②企业创造价值依靠专业知识和专家。③提供知识产品及其后续服务。

### 2.2.1.3 KIBS 企业的服务模式

传统的服务的服务模式具有标准化和定制化的特点，KIBS 继承了这两种服务模式。但由于服务标准化程度具有差异性，因此 Tether(2001)根据这一差异将 KIBS 的服务模式分成了四类，即完全标准化、高度标准化、完全定制化和高度定制化。在这一基础上，中国学者孙阿楠等人(2006)根据具体的软件行业特点，把软件服务公司的服务模式划分为技术推动型、客户导向型和互动型三种软件服务模式。总而言之，KIBS 的服务模式目前呈现出多元化发展，但服务模式的选择必须依靠市场和客户的需求。

### 2.2.1.4 KIBS 企业与知识创造

Strambach(2001)描述了 KIBS 企业与其客户间新知识获取、知识重组和知识传播（知识转移）的过程。其中知识获取与知识转移是 KIBS 企业与客户间的交互过程，而知识重组则发生在企业内部，即将交互过程中获得的知识与现有的知识库重新组合，创造出新的知识。如下图所示：

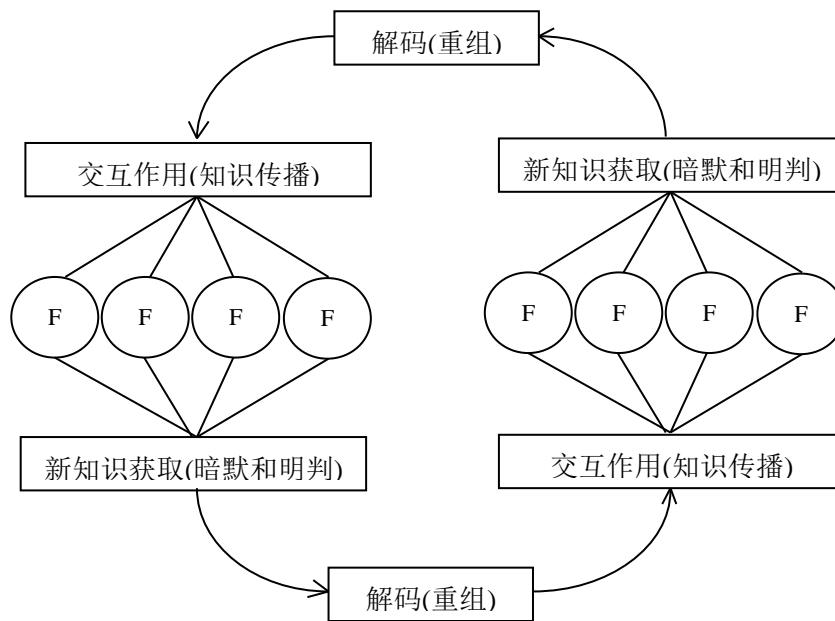


图 2.6 KIBS 与客户间知识产生和传播

来源于：Strambach(2001).

说明：F1、F2、F3、F4 代表不同客户

KIBS 企业的知识创造同样存在 Nonaka&Takeuchi(1995)所提出的知识创造的四个过程，结合 Strambach(2001)观点，新知识获取则包括显性和隐性知识，这是社会化和表出化的过程，知识重组则是联结化的过程，而其提供知识产品是知识传播（知识转移）的过程，属于内在化的过程。

## 2.2.2 软件企业

软件企业是一种典型的 KIBS 企业，因此软件企业继承了 KIBS 所具有的特征，如知识特征等。但是软件企业另外具有一些独有的知识特征和企业特点。

### 2.2.2.1 软件企业的知识特性

软件企业是信息、知识创造和知识学习集于一体的知识型企业，与其他企业相比，它有着独特的知识型特征。其主要的知识特征是：软件企业是企业人员的智力和技术资源深度融合的知识型企业。软件企业竞争的实质是人力资源的竞争，高层次人才是软件企业最重要的竞争资本。同时，软件企业可以说是知识系统和学习系统相互渗透的组织，知识的价值和学习能力的强弱对企业人员及至整个软件企业的生存和发展都具有决定性的意义。

### 2.2.2.2 软件企业的特点

#### (1) 人力资源的知识性与专业性

软件企业的生存和发展极度依赖人力资源，所有的产品开发均依靠人和人自

身所具有的专业知识来完成，而不是借助机械或其他工具来完成。因此人力资源和智力资源是软件企业的第一要素，也是企业最重要的资源。同时，软件产品的开发并非任何人都能胜任，必须是具备相关专业知识的人才能承担。

#### (2) 产品研发过程的团队化

软件研发过程不是一两个人能完成的，必须依赖不同领域、具有不同专业知识的各类人员进行协同工作，方能最终完成整个软件工程的开发。比如一个软件开发项目，必须由一个研发团队来完成。团队中研发任务对应专业分工十分明确，如项目需求分析、系统设计、构架设计、界面设计、代码编写、系统测试等都分别有与之对应的专业人员来完成。

#### (3) 企业知识的隐含性和共享性

在软件项目的研发过程中，每个环节都有大量的显性知识和隐性知识，但其中的显性知识必须转化为员工个人的隐性知识后，才能最终完成研发任务。可见在软件企业中，产品的研发对隐性知识的依赖程度非常高，软件产品就是整个研发团队所有员工隐性知识的工作结晶。但是，团队成员的跨专业知识、经验、技巧、信息（信息不对称）等都有差异，因此企业很多核心技术、专业知识和业务信息必须共享，否则难以有效整合。

#### (4) 企业知识的生命周期性

半个世纪以来，信息技术领域的技术一直按摩尔定律来发展，即 18-24 个月都会有新产品、新技术出现，且性能翻一倍，其发展速度可见一斑。随着信息技术的快速发展，各种各样知识也层出不穷，对知识资源极度依赖的软件企业其知识的更新速度比其他企业都要快。可见，软件企业知识利用周期较之其他企业来说短很多，这就需要企业进行创新，掌握新技术和新方法，以高效、快速地获取和创造有价值的知识，并及时在企业内部共享和转移，使技术知识能够及时应用。

#### (5) 产品知识的多样化

软件产品的知识相较于其他产品来说表现形式非常多，但主要体现在显性知识。比如可以通过文字、声音、图像、视频、公式等形式表现，知识载体可以使用电子媒介来传播知识，如磁盘、光盘、U 盘、网盘等。

### 2.2.2.3 中小软件企业应用大数据面临的问题及策略

Kleinknecht(1989)分析了中小企业所面临的问题，并指出三个关键问题，即资金不足、管理能力缺乏和获得创新所需要的知识和技术难度大。而在大数据时代，中小企业所面临的问题主要有以下问题：(1)数据源获取困难。由于中小企业的资金不足，业务占行业市场份额小，不可能独自建立数据中心，只能依靠外部资源。(2)IT 基础设施不足。鉴于成本问题，中小企业不适合搭建完整的大数据处理 IT 系统（包括硬件系统、软件系统和网络）。(3)专业技术人才缺乏。人才的缺乏是中小

企业所面临的普遍问题。大数据是新兴的 IT 技术，其专业人才目前仍然无法满足市场需求。而中小企业在人才争夺上无法与大企业抗衡，因此获得大数据相关人才非常困难。(4)创新能力不足。由于中小企业所掌握的知识和技术资源不足，无法与大企业相比，使得创新能力难以提高，从而阻碍了从传统思维向大数据思维的转变。

中小软件企业应用大数据时同样面临上述问题。大数据看起来似乎只适合大企业。但从很多方面而言，大数据更适合中小企业。如中小企业灵活，服务转型快，行动更迅速；大数据应用得好，可以提高中小企业创新能力，降低企业人力资源成本等。因此，中小软件企业必须认真思考大数据的应用策略。

从现有文献看，其建议主要集中在中小企业实施大数据业务外包策略，即 IT 基础设施建设、数据的收集、存储、获取、分析、挖掘等交由大数据服务公司。但从中小软件企业实际应用效果看，大数据业务完全外包并未获得预期效果，其主要原因是：软件 R&D 活动专业性太强，大数据服务公司的分析与挖掘人员不能充分理解软件开发业务的具体需求，即分析与挖掘思想和实际业务相脱节，从而导致分析与挖掘结果和企业预期不符。因此，很多中小软件企业在将大数据业务外包的同时，也注重培养 R&D 团队的数据统计分析与数据评价能力，目的是由 R&D 团队对大数据服务公司分析与挖掘结果再次进行分析或评价数据质量和价值。也有部分中小软件企业仅要求大数据服务公司完成部分数据处理、分析和挖掘任务，将半成品数据交由 R&D 团队实施最后的数据分析与挖掘。

随着大数据的迅猛发展，大数据服务公司也不断涌现，而且其服务范围越来越宽。如美国 Precog 公司提供的大数据服务，可以从网络各种数据源采集各类数据，然后还对使用人群统计其态度、位置和其他信息，使得数据类型更加丰富，并在最后展开综合分析；Accenture 公司也提供了一站式的大数据分析和应用方案；淘宝提供的数据服务产品量子恒道和“淘师爷”，可以为淘宝卖家提供电子商务的大数据分析服务，以此来帮助卖家提高其销售量。

### 2.2.3 KIBS 研究评述

从现有对 KIBS 的研究看，研究工作主要集中在 KIBS 界定、分类、特征、知识特性、服务模式、创新作用、区域影响等方面，结合具体行业也有研究，如软件业等，而且这些研究其前期成果非常丰富。但本论文认为，目前学界对 KIBS 的研究仍存在以下不足：

(1) 对 KIBS 的知识创造研究较少，即使有也缺乏实证检验。

(2) 对于高度依赖信息技术的 KIBS 企业（特别是 T-KIBS 企业），新兴的信息技术对 KIBS 所造成的影响研究几乎没有。

## 2.3 R&D 团队文献综述

R&D 活动是高度知识密集型投资,团队越来越成为科技创新的有效组织形式。1996 年美国国家科学技术委员会就曾指出:“美国政府和企业中的 R&D 团队及其创造知识的能力,决定着下一个世纪美国在国际市场中的地位”。可见,R&D 团队在企业中的地位相当高,为此学界对 R&D 的关注度也非常高。

从目前搜索到的文献看,学界对 R&D 团队研究主要集中在三个方面,即知识层面研究、相关影响研究和评价与测量研究。

### 2.3.1 知识层面研究

诸多学者、机构认为,R&D 团队具有知识创造和知识创新的功能。Nonaka &Konno(1998)提出,R&D 团队是一种典型的知识型团队,它是企业中最理想的的学习单位和知识创造的“Ba”。《中国科学技术指标》中明确写到,R&D 活动目的是为了知识创造和知识应用而开展的系统性工作。而晏双生(2010)在区分知识创造和知识创新中提出,R&D 是知识创造和知识创新的统一。此外,很多学者还对 R&D 团队内部知识共享和知识转移进行了研究。如 Hamel(1991)通过对多个案例的研究,构建了 R&D 团队内部知识转移和知识共享的理论框架。McLure&Faraj(2005)研究后指出,R&D 团队在目标和任务具有相似性、组织激励制度、技术条件等条件下,对个体知识转移和知识共享产生积极效果。中国学者汤超颖&丁雪辰(2015)则通过知识异质性、知识要素间的相似性、企业技术创新能力所受影响等三个方面研究了企业 R&D 团队的知识基础和知识创造的关系。

### 2.3.2 相关影响研究

相关性研究主要是对 R&D 团队各类活动产生的影响与发生的关系进行研究。首先是影响因素方面的研究,如 West(1990)就指出,团队的创新氛围可以影响团队成员创新能力的发挥。West 在 1994 年实证了他自己的这一观点。中国学者在这方面做了大量的研究,如徐芳(2003)研究了 R&D 团队胜任力模型对其创新绩效的影响。杨德林&史海锋(2005)通过实证的方法,认为影响 R&D 团队知识创造的因素有内部因素、管理因素、环境因素、知识库以及外部因素等。柯江林等(2009)通过实证认为,变革型领导对 R&D 团队的创新绩效有积极的影响。张树娟&刘政(2009)研究后提出,影响 R&D 团队知识创造的因素主要有团队人数、团队领导和团队成员的互动模式。王黎莹&陈劲(2010)通过研究以团队共享心智模型为中介的研发团队创造力的影响机制后提出影响 R&D 团队创造力的四个因素,即团队任务、团队领导、团队成员、团队过程。

### 2.3.3 评价与测量研究

在对 R&D 团队各种活动的测量上, 学界做了大量研究, 因此这方面的文献也很丰富。Werner&Souder(1997)认为 R&D 团队的创新绩效较难进行测量, 原因是研发人员的投入很难观察。Kerssen&Cook(1997)则提出了适合 R&D 绩效测量的平衡记分卡模型。Loch&Tapper(2002)认为, 无论是研发的过程或结果都应列入创新绩效测量范围。Coccia(2001)认为, 团队创新绩效可以直接投资强度、自我融资难度、受培训强度、传授强度、发表论文指数、专利申请数等指标进行测量。中国学者对 R&D 团队也做了大量研究, 如路变玲(2008)等通过利用结构方程建立了 R&D 团队绩效评价模型, 同时设置了团队能力、团队互动过程、团队成果等三个评价指标。逮宇铎等人为基础研究、应用研究、技术开发研究三类 R&D 研发团队提出了的团队成员的个人绩效评价体系。

### 2.3.4 R&D 团队研究评述

从现有的文献看, 现有对 R&D 团队研究均在团队理论研究的基础上, 结合 R&D 团队的特点开展研究。其研究内容主要集中在知识层面研究、相关影响研究和评价与测量研究等三方面。前期研究成果很多, 但本论文认为仍存在以下问题:

(1) 诸多观点仍然停留在构思阶段, 没有能够开展实证研究。

(2) 针对某个行业研究 R&D 活动较多, 但研究 R&D 团队较少。

(3) 在信息技术快速发展, 许多行业生存发展高度依赖信息技术的今天, 研究新兴信息技术对 R&D 团队各种活动的影响还是空白。

## 2.4 信息技术与管理学文献综述

从信息技术在管理学研究中的相关文献看, 几乎涉及每个管理学研究领域。学者们的研究侧重于通过信息技术来研究不同的管理学理论。从相关文献看, 学界内关注度较高的有基于信息技术的资源基础理论、战略管理理论和知识管理理论。我们将从这三个方面对相关研究文献进行回顾和梳理。

### 2.4.1 信息技术的资源基础理论

运用企业资源基础理论对信息技术进行研究于 20 世纪 90 年代开始受到学者们的关注。早在 1985 年 Michael Porter 就指出, 信息技术已经渗透到经济管理的每个领域, 成为继土地、资金、劳动等生产要素之后的又一重要资源。Ross (1996) 在定义 IT 能力的同时, 把信息技术资源划分为技术资源、人力资源和关系资源三部分。Bharadwaj(2000)则对 Ross 所提出的信息技术资源进行修正, 他认为信息技术资源应该包括信息技术基础设施、人力信息技术资源和信息技术相关的无形资产。同时他认为 IT 能力是用于整合 IT 资源的能力, 其本身也是一种资源。中国学



者张嵩&黄立平(2003)从资源角度分析, IT 能力是一种可以调动、调配企业信息技术资源, 并以此获得长期竞争优势, 其包括 IT 基础设施、IT 人力资源、IT 组织文化资源和 IT 使用的无形资源。

为了更好的管理企业资源, 学者们从管理学层面结合技术层面对 ERP (Enterprise Resource Planning, 即企业资源计划) 进行研究。ERP 最早由美国 Gartner Group 公司提出, ERP 是以信息技术结合管理学理论的一种管理信息系统, 是现代管理与信息技术相结合的产物。中国学者在 ERP 上也做了大量研究, 如在管理学层面上, 仲秋雁&闵庆飞&吴力文(2004)研究了中国企业实施 ERP 过程中的成功关键因素, 并给予了实证。在技术层面上, 张巨俭&甘仞初(2003) 提出了管理信息系统发展方向重点在于 ERP 系统, 并给出了实现技术。

#### 2.4.2 信息技术的战略管理理论

由于信息技术的快速发展, 企业经营环境、经营模式及其管理范式也发生重大变革, 因此使得战略管理理论也发生深刻的变化。信息技术对战略管理理论的影响受到学界的高度关注, 纷纷研究基于信息技术的战略管理理论。

在基于信息技术的战略管理理论研究早期主要集中在战略规划上, 如 C. H. Krebel(1968)对计算机系统规划的战略问题研究。F. W. McFarlan(1971)、W.R. King (1978) 关于管理信息系统 (MIS) 的战略规划问题研究等。到了八十年代, 由于受到以波特为代表的企业战略管理结构学派的影响, 研究信息技术对企业竞争环境和竞争优势产生的影响成为学界研究的热点。如 G.L. Parson(1983)就指出, 要让企业经理清醒地认识到信息技术是企业的竞争武器, 企业必须要让信息技术资源服务于最重要的任务。Gregory. L. Parsons(1983)、F. W. McFarlan(1984)、Michael E. Porter(1985)等人则运用波特的分析方法, 对信息技术引发的竞争环境和竞争方式进行了研究。总体而言, 这一时期信息技术的企业战略管理理论研究基本以 Porter 在 1980 年、1985 年、1990 年所提出的战略管理思想为基础。八十年代之后, 信息技术的企业战略管理理论研究则更侧重于实用方面的研究。如 R. I. Benjamin and E. Levinson(1993)研究了信息技术变革下的企业管理框架。D. F. Feenv and L. R. Willcocks(1998)研究了如何通过信息技术利用核心 IS 能力。此后, 国际互联网在全球得到迅速发展, 学界则把关注度投向了基于网络的战略管理研究。如 Charles B.Sabell&Stain D. Fejeldstad(1998)研究了企业竞争优势和价值链、价值店、价值网的关系。Anders Rokstrom&Bengt Zdebel(1998)研究了企业如何在网络环境下生存的战略。Gil McWilliam (2000)研究了如何通过在线社区来建立一个强力品牌。

#### 2.4.3 信息技术对知识管理的作用

学界基于信息技术对知识管理理论的研究, 主要集中在信息技术对知识管理

的影响和作用。诺贝尔经济学奖得主 Robert Solow 对信息技术的作用提出“生产率悖论”，即企业在信息技术投入大，但收到的效果低。Sweeny(1996)也持有与 Robert Solow 一样的观点。但在 20 世纪 90 年代中后期以后，信息技术的应用对企业产生重要性突显，此时新的企业层级数据获取和新成本估算技术出现，随后的许多研究结果表明，信息技术与生产率呈正相关关系，信息技术对组织产生积极的影响。为此，基于信息技术的知识管理也就受到了学界的重视。

对相关文献的梳理和归纳，本论文以徐璞（2014）研究的信息技术在知识工作中的创新作用所提出信息技术与知识管理关系的六个方面，即知识检索、知识存储、知识共享、知识转移、知识创造以及知识应用等进行逐一回顾。

#### （1）信息技术对知识检索和存储的作用

组织知识检索和存储被称为组织记忆，是组织有效进行知识管理的重要因素。信息技术可以帮助个体或组织通过记忆要点获取大量信息，也可以帮助个体或组织将必要的信息存储到载体。因此，Chang 等人(2004)认为，良好的组织记忆是信息技术持续作用的结果。Mark&Joline 对一个存储系统进行研究，发现该系统可以辅助记忆产品信息，包括公告、服务手册、产品价格等，并帮助员工、客户快速解决问题。Fernando Olivera(2000)认为，不同信息技术对个体或组织的记忆作用是互补的。中国学者张嵩等人(2002)认为，信息技术有利于知识的获取性与重复提取性，有利于企业系统地、有条理地、规范地管理内部知识。

#### （2）信息技术对知识共享的作用

Nonaka&Takeuchi 认为，知识创造、共享和应用是企业核心竞争力的源泉。信息技术不仅能为个体或组织提供获取和存储知识的途径，还能让个体给他人分享自己的知识。但在知识共享方面，学界普遍认可信息技术解决显性知识的共享问题，但对于隐性知识的共享存在较大争议。如 Tua Haldin Herrgard(2000)提出，因为显性知识容易表达，因此可以通过很多信息技术手段，如在线培训、电子邮件等方式进行共享，而隐性知识的共享则需要个体间通过面对面沟通。但 IBM Global Services 和他的 T. J. Watson 研究中心通过做将信息技术用于隐性知识的试验认为，信息技术可以作为加强隐性知识传递的工具。

#### （3）信息技术对知识转移的作用

Leibrecht(2004)认为，计算机、网络通信、数据库等技术对组织知识管理中的知识转移过程带来巨大影响。Bolisani&Sacorso(1999)通过运用 SECI 模型研究了信息技术对组织之间进行知识转移的影响发现，不同信息技术支持不同类型的知识转移过程。如 EDI 技术支持显性知识间的转移，CD-ROM 则支持显性到隐性知识的转移，CAD 和 CAM 则支持隐性到显性知识的转移。

#### （4）信息技术对知识创造的作用

Davenport&Prusak(1998)认为信息技术是知识创造过程当中的一个重要因素。

Lee&Choi(2003)也认为,各种形式的信息技术可以帮助个体和组织收集、获取、交流知识,从而促进他们创造出新的知识。Vaccaro 等人(2009)基于 SECI 模型,以虚拟技术为对象研究组织内知识创造过程发现,信息技术在共同化过程中起辅助作用,在表出化过程没有发现信息技术带来积极贡献,联结化过程中信息技术起到了支撑作用,内在化过程中部分信息技术可以促进个体的知识内化,同时也促进组织管理的规范化和知识传递。Monsanto 公司等则强调知识地图、信息技术、人工智能、网络技术对员工知识创造的促进作用。

#### (5) 信息技术对知识应用的作用

目前信息技术对知识应用的作用研究主要集中在影响知识应用的因素和基于技术角度的知识应用两方面。为此,我们也将从这两方面进行文献回顾。

信息技术对知识应用的影响因素文献较多,如 Wesley&Daniel(1990)分析了影响个体对知识吸收能力的影响因素,认为相关知识积累、相关知识差异、相关知识重叠是关键影响因素。Fred Davis(1989)认知的有用性和无用性影响知识使用者应用知识的态度和意图。而从技术角度对知识应用研究较少,如 Sounmitra Dutta(1997)研究发现,管理决策系统自动化辅助决策过程促进了个体的知识产出。Grant(1996)认为,信息技术能够增强知识整合与应用过程。

### 2.4.4 文献评述

基于信息技术的研究几乎涉及管理学所有领域,其研究内容基本是信息技术结合管理学的各种理论和观点展开。其研究范围包括信息技术在组织中的地位和作用,信息技术对组织的影响,信息技术在组织中的应用等,前期成果很丰富,但本论文认为仍存在以下不足。

(1) 缺少基于信息技术的技术变革对管理理论进行深入研究。如信息技术有重大变革后出现的新兴技术云计算、大数据、物联网等,它不仅导致管理学研究过程上的改变,还导致思维方式发生颠覆性变化,其影响极其深远。但很少有学者在这方面展开研究。

(2) 信息技术与知识创造的研究也非常有限,即使有也是缺乏实证检验,基本是基于经验来描述信息技术对知识创造产生的作用。

## 2.5 大数据文献综述

### 2.5.1 大数据发展历程

大数据时代来临已是不争事实。目前,国家层面、企业层面、学术界等对大数据的关注度前所未有。就学术界而言,2008年《Nature》杂志推出大数据专刊,最早探讨了大数据问题,此后“大数据”成为各领域开展科学研究的热点。同年,

计算机社区联盟也发表了《大数据：创造商业、科学和社会领域的革命性突破》，其主要阐述了解决大数据关键技术所面临的问题。2011年《Science》杂志再次推出专刊《Dealing with Data》，其主要对科学研究中的大数据问题进行探讨，说明了大数据对科学研究的重要性。《Dealing with Data》专刊的推出标志了“大数据”时代的正式到来。

真正让大数据风靡全球的是全球知名咨询公司麦肯锡公司于2011年6月发布的《大数据：创新，竞争和生产力的下一个前沿》报告。该报告详尽分析了对大数据的影响、主要技术以及大数据的应用领域。2012年1月的达沃斯经济论坛中，大数据成为讨论的主题之一。美国一些知名数据管理领域学者联合发布了《Challenges and Opportunities with Big Data》白皮书，该白皮书介绍了大数据的产生、处理流程和面临的若干问题。奥巴马政府于2012年3月发布《大数据和发展倡议》。发布这一倡议的同时承诺投入2亿美元，用于启动大数据发展计划。中国也于2015年发布了《促进大数据发展行动纲要》，明确要建立标准和规范体系，推进大数据的快速发展。

### 2.5.2 大数据的特征和数据结构类型

IBM公司率先提出了大数据3V特征，即规模性（Volume）、多样性（Variety）和快速性（Velocity）。规模性指数据量非常大，至少在TB级以上。但HDS公司副总裁Hubert Yoshida则认为数据量大是相对而言的；多样性指数据不局限来源于关系数据（结构化）等传统数据类型；快速性是指快速处理动态数据的速度。

国际数据公司（IDC）则提出4V特征，即在IBM的3V特征基础上，增加价值性（Value）特征。但他们认为大数据的价值呈现稀疏特点。Bill Franks(2012)也认同这一观点，认为大数据价值利用密度低，大部分数据没有价值。同时，Bill Franks概括了大数据的共同特征是：都由机器自动生成，

Yuri Demchenko等则提出5V特征，即在4V基础上增加真实性（Veracity），真实性包括可信性、真伪性、来源和信誉、有效性和可审计性。IBM也认为，大数据必须具有真实性，真实性才能通过数据融合和数学方法来进一步提高数据质量，从而使数据具有更高的价值。Yuri Demchenko提出的5V特征如下图：

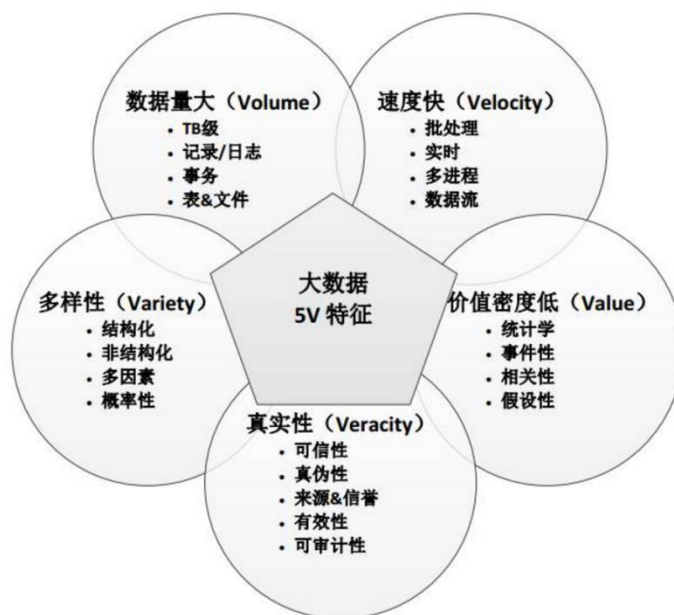


图 2.7 大数据 5V 特征

来源于：《大数据标准化白皮书 V2.0》

数据结构通常分成结构化数据、非结构化数据和半结构化数据。结构化数据即以二维表的形式存储在数据库。而方便用二维表的形式存储在数据库中的数据我们称之为非结构化数据，如文档、音频、视频、活动、会话、传感等数据。半结构化数据则介于结构化和非结构化之间的数据。

随着信息技术、通信技术、物联网技术等快速发展和应用，使得非结构化的数据大量增长，大数据应运而生。此时，传统管理结构化数据的关系型数据库远远不能满足数据变化和增长的需要。大数据的数据结构类型则包括结构化、非结构化和半结构化数据的全部数据结构类型。

### 2.5.3 大数据技术理论

大数据不是一项专门的技术，它是一种概念，是一系列技术的集合。大数据依靠如计算机软硬件、网络、云计算、数据仓库、数据挖掘、统计分析、智能识别等一系列技术构建成为一个大数据处理平台，并利用这一平台完成大数据的收集、存储、处理、挖掘、分析、智能识别和资源管理等任务。

#### (1) 大数据平台架构

目前，大数据平台从集成两三个组件发展到了完整的生态链。事实上，大数据平台的架构主要以商业为驱动，即根据不同的商业场景，选择技术措施也不同。因此，本论文所讨论的大数据平台架构是一种基础架构，不同的应用可以根据实际情况进行拓展。以下为大数据平台架构图：

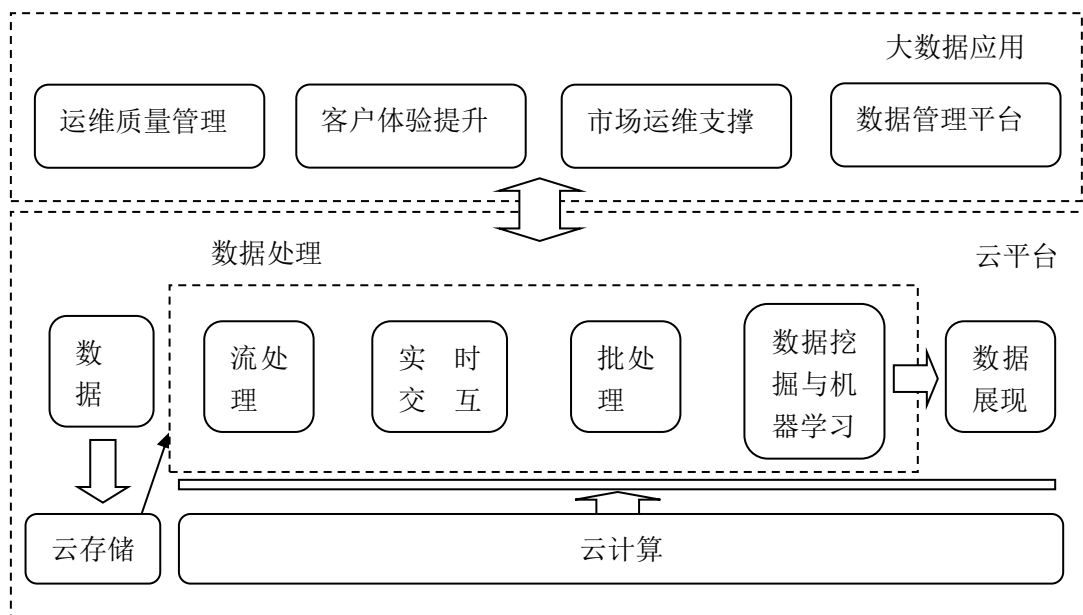


图 2.8 大数据平台架构

来源于：本论文整理

从以上大数据平台架构看，其可以分成基础设施、数据采集、数据存储、数据处理、数据展现和大数据应用等 6 部分。以下将按这 6 部分作逐一介绍。

### (2) 大数据基础设施

大数据的基础设施主要依赖云平台。云平台则包括服务器、用户计算机、云操作系统及其集成的各类软件等。云计算就是利用虚拟化技术，把网络内所有资源集成一个整体（CPU 资源池化）使用。比如用户使用计算机处理事务，通常仅由用户使用的计算机内 1 个 CPU 实施运算。但使用云计算后，就可以把网络内空闲的计算机 CPU 集成一个整体来使用，因此可能有几十个，甚至上百个或更多的 CPU 同时实施运算。可见云计算具有超强的计算能力。同样，云计算也可以把网络内所有的计算机存储资源集成一个整体（存储资源池化）使用，使得存储资源取之不尽，用之不竭。由此可见，云计算具有两个最重要的应用特性，即超强的计算能力和巨大的存储空间。

从云计算的两个重要应用特性看，大数据与云计算是绝佳配对。从大数据的量及其需要的处理速度看，处理大数据需要强大的计算能力。从大数据需要存储的量看，大数据正需要巨大的储存空间。因此，大数据极度依赖云平台，云平台是大数据必须搭配的基础设施。

### (3) 数据采集

数据量充足是大数据战略基础。大数据所有价值的挖掘都建立在数据采集的基础之上。而大数据所采集的数据不是精确而是混杂，纷繁的数据越多越好。

大数据采集的途径和方法很多，比如通过网络信息（如 Web、MIS 等）、物联网传感网（如传感器、RFID 等）、GPS 定位、摄像头（如交通摄像头）、移动终端（如手机、平板电脑等）、系统日志、公众交互平台（如 QQ、微信等）等等进行数据采集。

大数据的采集针对不同数据源采用不同的采集工具。比如针对网络信息，可以使用探针采集。将探针放置于服务器或交换机上，直接获取数据。对于 Web 的采集可以采用“网络爬虫”软件工具，对散落在互联网中的数据进行采集。而对于系统日志的采集，可以通过 Flume、Chukwa、Scribe 等工具。

#### （4）数据存储

采集到的数据需要空间存储，以保证可重复使用。大数据的存储主要依赖云平台所提供的云存储。大数据主要有结构化数据和非结构化数据两种类型，结构化数据即可以用二维表结构来逻辑表达的数据。而非结构化数据则不方便用二维表结构来逻辑表达。结构化数据存储方案主要有 SQL 结构化存储方案，而非结构化存储方案主要有 NoSQL 和 NewSQL 两种方案。

##### ①结构化数据库

结构化数据实际存储在数据库中，常用的结构化数据库有 Oracle Database、SQL Server、MySQL 等，这些数据库均执行 SQL 结构化存储方案。Oracle Database 由美国 Oracle 公司开发的一种网络分布式数据库，它是 C/S 模式（客户机/服务器模式）和 B/S 模式（浏览器/服务器模式）软件最常用的数据库之一，通常被用于大型数据库，也成为布署和管理数据库云的理想数据库系统。SQL Server 则由微软公司开发的数据库系统，也是模式 C/S 模式和 B/S 模式软件最常用的数据库系统，常用于中小型数据库。MySQL 也是 Oracle 公司旗下的产品，目前是应用在 Web 方面最好的数据库系统。它体积小，速度快，成本低，通常中小型网站的开发都使用 MySQL 做为后台数据库。

##### ②非结构化数据库

相对结构化数据库而言，非结构化数据库可以存储的数据类型更多更复杂，它能存储的数据格式包括文本、文档、图片、音频、视频、HTML、传感数据等等。非结构化数据库每个字段的记录可以由可重复和不可重复的子字段构成。它与传统的结构化数据相比，非结构化数据库突破了关系型数据库结构定义后相对固定不易改变和数据规定长度的限制。这是传统结构化数据库无法比拟的。常见遵循 NoSQL 方案标准的非结构化数据库有谷歌的 BigTable、Amazon 的 Dynamo、Apache 的 Hbase 等。

#### （5）数据处理

##### ①流处理

大数据的实时性很强，它往往不是静止的数据，而是连续不断的数据，就像

水管中的水在源源不断的流动一样。流处理就是为实时地、连续不断的数据准备的一种处理技术。它在数据不断变化的运动过程中实时捕捉分析，将可能对用户有价值的信息发送出去。流处理在许多需要进行实时处理的场景应用很多，比如股票交易需要在盘中实时监控价格的变化，一旦符合条件立即提醒交易。

典型的大数据实时处理系统有两个，一个是 Storm，另一个是 Spark Streaming。Storm 是开源的分布式实时数据处理系统，而 Spark Streaming 则是一种基于计算机内在的流数据处理系统，这种数据处理系统以迭代式和交互式为主要的数据处理模式。

### ②实时交互计算

在海量数据处理过程中，除能实时处理静态数据和流动数据，还需能实时响应用户的需求，比如用户的实时查询。传统的数据实时计算是先将数据保存到数据库管理系统，然后通过 query 和数据库管理系统进行交互，最后得到结果返回给用户。显然，这种数据实时计算方式并不适用于数据量巨大、数据类型复杂的大数据。大数据的实时计算一般采用海量并行处理 MPP 的分布式架构，即将数据的存储和处理分布到大规模的节点上，以此满足大数据的实时性要求。常用的大规模分布式处理系统如 Hadoop 的 HDFS 文件处理系统或者 NoSQL 分布式数据库。

### ③批处理

批处理即对大量、复杂的数据实施批量处理，处理时间跨度从几分钟到数小时。数据批处理发展最早，应用也最广，如电信领域针对探针数据处理的 KPI 和 KQI 计算。大数据实施批处理通常使用 Hadoop 平台。但 Hadoop 平台主要用于静态数据的批量处理，对于实时数据则力不从心。MapReduce 则是分布式和并行计算的好模型，因此对于实时性数据常使用 MapReduce 数据处理系统，但需要将 MapReduce 改造成 Pipeline 模式。

### ④数据挖掘和机器学习

数据挖掘是通过数学方法，从海量数据中获取有效的、新颖的、有潜在价值的数据和信息。而机器学习则是一种智能化技术，它模拟人的学习机制，将大量数据和信息记录在机器中，机器遇到问题时将从已记录有的数据和信息中寻找答案或解决方案。因此，机器学习到的数据信息越多，其智能化程度越高。从某种意义上讲，机器学习的科学成份更多，而数据挖掘的技术成份更多。

如今在大数据的应用中，数据挖掘和机器学习相互借鉴和渗透。机器学习使用了数据挖掘中大量的算法，如决策树、贝叶斯、聚类、关联规则、人工神经网络等算法。这些算法使得机器在学习上更主动，更有价值取向，因此也就更智能。而数据挖掘则借鉴机器学习的思维方式，研究和拓展新算法，使机器学习成为数据挖掘的重要工具，提升数据挖掘的智能化。

## (6) 数据展现



数据展现就是将计算结果以简单、直观的方式展现给用户，其目的是让用户理解和使用，为用户形成有效的统计分析和预测决策提供数据依据。这一环节其实非常重要，甚至决定大数据应用的成败。

数据展现的手段就是数据可视化，即借助图形化手段，清晰有效传达与沟通信息。数据可视化需要了解客户的价值取向，需要从美学和功能两方面同时入手，让用户理解和接受数据的同时，不觉得枯燥乏味。

#### 2.5.4 大数据技术与传统信息技术

大数据技术是信息技术的一次技术变革，是发展到更高层次的新型信息技术。尽管大数据技术是以传统信息技为基础发展起来，但进无论技术上还是应用上都比传统信息技术高一层次，而且有着自己独有的特点。大数据技术不同于传统信息技术主要体现在数据特征、硬件环境、数据处理技术、信息分析方法、信息管理等 5 个方面，下面将从这 5 个方面，通过文献梳理并归纳成以下表格：

表 2.3 大数据技术与传统信息技术的区别

序号	项目	传统信息技术	大数据技术
1	数量	TB 级已经满足	PB 级以上
	种类	仅使用关系型数据，结构化数据	使用非关系型数据和新型多结构数据，包括结构化、非结构化、半结构化数据。
	速度	数据产生速度慢，只能处理静态数据。	数据产生速度快，需要快速、大批量地处理动态数据。
	价值	价值密度高	价值密度低
2	硬件环境	普通的计算机或服务器即可满足	依赖云计算和云存储
3	数据处理技术	数据挖掘技术和统计分析技术	综合语义、语音、图像、视频、数据挖掘、统计等分析方法，并融合智能化技术。
4	信息分析方法	抽样分析法与因果关系分析法	全数据分析法与相关关系分析法
5	信息管理	无法精准化管理	精准化管理

来源于：本论文对宋学清&刘雨(2012)、俞立平(2016)、赵晖光&赵勇(2017)等文献归纳整理

#### 2.5.5 大数据的影响、作用与应用

维克托·迈尔(2012)在《大数据时代:生活、工作与思维的大变革》一书中指出，大数据时代会让我们生活、工作与思维发生变革。2012 年瑞士达沃经济论坛一份题为《大数据，大影响》的报告称，大数据已经成为一种经济资产，就像货币和黄金一样。麻省理工学院布伦乔尔森（2012）认为，大数据的影响就像人类发明显微镜一样，推动知识增长，促进经济繁荣，拓宽科学边界。因此可以大胆

地预见，大数据将会对人类产生深远的影响。

从现有文献看，对大数据作用的前期研究较多，如美国哈佛大学加里·金教授（2012）认为，大数据是一场革命，庞大的数据资源将会促使各领域开始量化进程。《纽约时报》（2012）指出，大数据将在商业、经济等领域的决策思路和行为发生变革。中国学者陈忻钧（2012）强调大数据的智能分析应用到电子商务。戴艳（2013）认为，利用大数据分析可以了解网络利用率，识别服务故障，改进成本效率。余来文等（2014）认为，大数据应用主要有七种，即数据存储空间出租、客户关系管理、企业经营决策指导、个性化精准营销、本地化数据集市、数据搜索、创新社会管理模式。Bernard Marr 著，宫鑫等译(2017)则指出，中小企业可以通过大数据来摸透客户心思，了解市场需求，改善经营方式，预测企业绩效等。赵眸光&赵勇(2017)则通过技术层面，详细分析大数据的应用。

### 2.5.6 大数据与管理学领域研究

随着大数据技术影响力的扩大和深化，人类的管理思想和管理实践也会发生改变，管理理论与管理实践的范式将向“数据+模型+分析”转变，从而要求揭示管理相关大数据在其中的基础性作用、影响机理以及管理范式转变的普遍规律。但从目前搜索的文献来看，研究大数据下的管理学理论还比较少，但有逐年增多的趋势。由于管理学领域涉及面太广，本文仅搜索近 5 年大数据和知识管理的研究情况，并借此了解大数据在管理学领域的研究情况，具体如下表所示。

表 2.4 大数据与知识管理研究情况

序号	数据库	2013	2014	2015	2016	2017
1	CNKI	11	34	58	79	52
2	EBSCO	48	43	46	61	41

来源于：本论文整理

说明：本表数据统计截止时间为 2017 年 8 月

通过梳理大数据在管理学领域研究发现，从创新、知识、战略、价值、决策、技术、商业、应用等层面研究较多。如 LaValle 等人(2011)研究了利用大数据分析来洞察商业价值；余来文等人(2014)研究大数据下的商业模式，并提出大数据企业以行业选择与战略定位为起点，通过盈利模式、组织能力、资源整合和资本运作来使得企业得到发展；杨善林&周开乐(2015)提出了基于管理学视角的大数据资源 6 个重要特征：复杂性、决策有用性、高速增长性、价值稀疏性、可重复开采性和功能多样性；应璇&孙济庆(2016)提出基于大数据的精细化知识服务模型。俞立平

(2016)研究了大数据下高技术企业创新路径,并最终认为大数据对高技术企业的创新具有十分重要的意义;赵眸光&赵勇(2017)从技术角度研究大数据的管理;Bernard Marr(2017)则研究了中小型企业如何应用大数据。

### 2.5.7 大数据面临的问题

大数据技术是一项新兴的信息技术,其发展到今天,技术和应用均有了一定的积累,但仍然不可避免的面临一些问题。如朱洁等(2016)认为,大数据面临的挑战是成本、实时性和安全。Bernard Marr(2017)则认为大数据的短板是大数据相关人才的缺乏(如数据科学家),而且人才成本高涨,中小型企业很难实施其大数据计划。赵眸光&赵勇(2017)则认为,大数据面临的主要问题有三方面,一是大数据管理,二是大数据技术,三是大数据工程。田溯宁(2012)认为大数据发展障碍是数据流动性和可获取性。

部分学者从技术层面提出大数据所面临的问题。如何京翔(2012)认为大数据时代应该不断更新软件分析技术。钟鸿生(2012)总结了大数据技术所面临的问题有四个,一是分析手段缺乏,导致不能充分利用数据资源。二是分析能力局限性,导致复杂问题无法解决。三是时间限制,迫使采用简单建模技术。四是运算能力不够,只能降低模型精度。

部分学者则从宏观方面提出大数据时代所面临的问题。如谢文(2012)通过研究网络公司,总结了大数据的三个挑战,一是挑战公司决策能力,二是挑战公司的数据处理能力,三是挑战公司的组织运营能力。缪其浩(2012)认为数据是资产,大数据的核心议题是如何盘活资产,使之为国家、城市、企业和个人服务。

### 2.5.8 大数据研究评述

国内外对大数据的研究主要集中在技术和应用层面。大数据研究虽处在起步阶段,但近年受到学界的高度关注,研究成果逐年递增。但存在以下问题:

(1)目前大数据研究仍处在起步和探索阶段,学术研究大多局限于宏观层面。

(2)现有大数据研究大多立足信息科学,侧重于大数据的获取、存储、处理、挖掘和信息安全等方面,而基于大数据的管理学理论研究较少。而现有大数据和管理学的研究大多基于经验性的描述,缺乏实证检验。

## 2.6 本章小结

本章节从知识创造、KIBS、R&D 团队、信息技术与管理学、大数据等五个方面出发,对中国及其他国家相关于本论文的研究文献进行了梳理、分析和归纳,通过回顾和总结前期的研究成果,发现当前管理学领域的相关研究中所存在的不足,进而为本论文提出研究问题、研究思路和研究方法上指引了方向。

## 第三章 软件工程及其大数据应用

软件开发过程就是软件工程的实施过程。因此，软件企业 R&D 团队应用大数据技术实施软件开发，实际上是软件工程的大数据应用。为了更好地理解和研究大数据应用下，R&D 团队软件开发过程中知识创造过程是如何发生变化的，我们须首先了解软件工程相关基础理论、软件工程应用大数据过程以及软件 R&D 团队大数据应用与知识创造关系。

### 3.1 软件工程理论基础

软件研发过程是知识产品的生产过程，是创造新知识的过程。因此软件工程是创造新知识的工程。软件工程理论是以工程化方法来研究开发和维护软件的理论，以此提高软件质量、实用性和降低开发成本。自 20 世纪 60 年代末期发展至今，软件工程经历了半个多世纪的发展。此间软件工程新方法、新技术不断涌现，几乎每隔 5-10 年就有一次突破性发展，极大提高了软件研发效率和软件质量。

#### 3.1.1 传统软件工程理论

软件工程理论通常以开发模型来研究软件工程的一般过程。传统软件开发模型很多，常见有瀑布模型、边做边改模型、快速原型模型、增量模型、螺旋模型等等，其中以瀑布模型最为经典。中国信息产业部和中国国家标准化管理委员会于 2007 年颁布了 GB/T8566-2007《信息技术 软件生存周期过程》，并以此来规范中国软件行业的软件开发一般过程，而这一过程就是基于瀑布模型的基础之上。软件企业通常会根据自身的实际情况，在国标版的基础上进行调整。

##### 3.1.1.1 瀑布模型简介

瀑布模型由 Royce 于 1970 年提出。该模型为最经典的软件开发模型，也是至今唯一仍被广泛使用的软件开发模型。瀑布模型是将软件生命周期划分为可行性研究、需求分析、系统设计、代码编写、软件测试、运行维护六个阶段。六个阶段顺序由顶自下，互相衔接，严格固定，不可改变。如同瀑布流水，逐级下落。

但瀑布模型在实际操作过程中往往存在诸多缺点，如（1）过于强调文档的作用，每个阶段间都会产生大量的文档，极大增加工作量。（2）由于开发过程是线性的，直到后期才能见到成果，从而增加开发风险。（3）由于开发过程的不可逆性，因此问题会积累扩大，到测试阶段才发现，会由此带来严重后果。

以下为软件开发瀑布模型图：

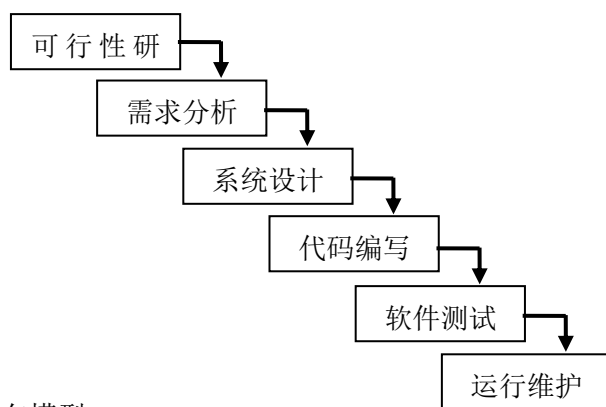


图 3.1 软件开发瀑布模型

来源于：本论文制作

### 3.1.1.2 软件开发一般过程

软件企业在接到用户订单后，需要开展一系列软件开发工作。按开发工作进度的先后划分成以下阶段：可行性分析、需求调研、需求分析、系统设计、软件开发、软件测试、试运行与验收、系统维护等，其开发流程图如下（参考 GB/T8566-2007 标准）：

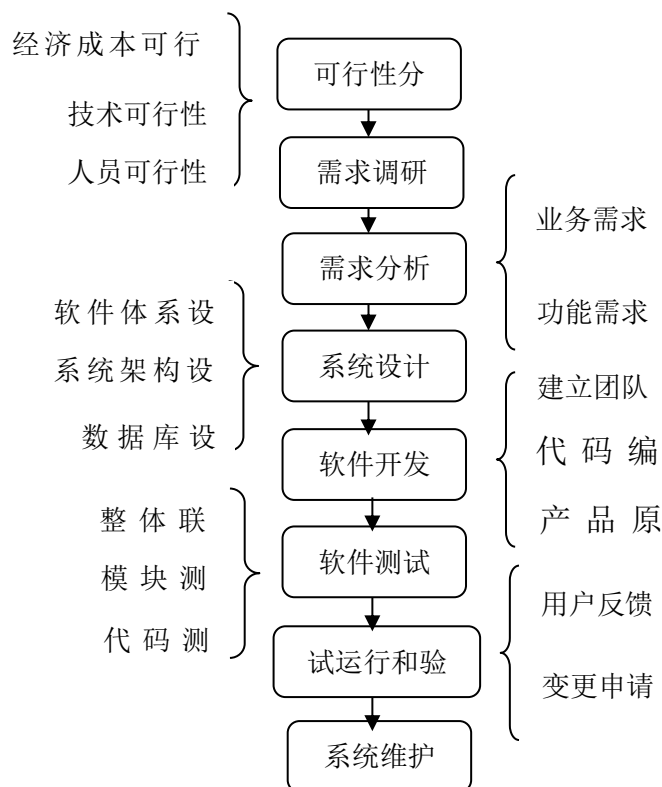


图 3.2 软件开发一般流程

来源于：本论文制作

**可行性分析：**在接到用户订单后，企业需要根据用户的初步描述，对软件开发的可行性作分析和研究。可行性分析一般从开发的经济成本、技术积累、人员配备等三个方面进行分析，并形成报告提交企业领导层决策，得到同意后进入下一个开发阶段。通常由于人员及各方面因素，中小型企业会跳过这一步骤。

**需求调研：**可行性报告通过后，企业需要派相关人员走访用户进行调研。调研过程中与用户面对面沟通，同时深入了解用户的业务流程。最后将用户对软件处理业务过程的需求形成报告。

**需求分析：**系统分析员根据需求调研结果进一步做符合软件工程规范的需求分析。需求分析首先进行业务分析，在业务分析的基础上进一步做软件的功能需求分析。通常需求调研和需求分析均由系统分析员完成。

**系统设计：**做好需求分析后即进入系统设计阶段。这一阶段主要由系统分析员根据软件需求对整个软件系统做整体规划，规划包括三方面，即体系结构设计、软件构架设计和数据库设计。

**软件开发：**在完成软件系统整体设计后，即进入具体的开发过程。在进入具体开发工作前，需要首先组建软件开发团队，再由软件开发团队成员实施代码编写，最后完成产品原型。对于许多中小型软件企业来说，由于技术人员的匮乏以及从成本上考虑，建立团队会放在首位，然后由系统分析员作为团队负责人参与到软件开发的全过程。

**软件测试：**软件测试的目的是尽量找出软件原型存在的各种 BUG。软件测试方法很多，总体而言主要进行整体联调、模块和代码等测试。

**试运行与验收：**在完成修改和测试后，即可进行软件试运行，并邀请用户对软件实施验收。在这一阶段中，用户如果发现问题或提出变更要求，则需要返回需求分析阶段，再按流程进行适当修改，直到用户满意为止。

**系统维护：**系统维护是软件投入使用后的售后环节。

### 3.1.2 敏捷开发理论

敏捷开发是根据用户需求，采用原型法、迭代法和增量法相结合的软件开发方法。敏捷开发的方法论很多，比如 Scrum 方法、微软的 MSDN 方法等。但敏捷开发一般均遵循以下原则：

(1) 根据用户需求，快速开发并交付软件原型。

(2) 欢迎用户为了竞争优势而变化需求，并根据用户需求不断修改软件，并经常发布新版本软件。

(3) 软件项目须有核心人员。团队内外均保持面对面的沟通方式，开发人员甚至与业务人员共同工作。

(4) 团队需要持续关注技术和设计。

敏捷开发方法较之传统软件工程开发更注重规避开发风险、开发人员个人因素以及开发者与用户之间的互动交流。以下为两种开发方法的区别：

表 3.1 敏捷开发方法较之传统软件工程开发

序号	传统软件开发	敏捷开发
1	首先做需求分析	把需求分析分散到整个开发过程
2	流程和工具	个人和交流
3	完备的文档	可用软件
4	为合同谈判	与客户合作
5	执行原定计划	根据需求变化而变化
6	软件完成后再发布	软件待续增量发布

来源于：邹欣（2015）

### 3.1.3 面向对象程序设计理论

面向过程程序设计与面向对象程序设计主要区别在于：面向过程程序设计的实现软件功能时需要编写所有程序代码。而面向对象程序设计的设计思想则不同，它事先将具有一定通用功能的程序代码定义成类、函数、控件或组件等（这些类、函数、控件或组件等即为对象），并将其保存以备复用。在程序设计时只需将这些类、函数、控件或组件等像积木一样搭建起来即可。可见，面向对象程序设计编写的代码量比面向过程少得多，编程效率也因此高很多。

但不管是面向过程还是面向对象的软件开发，它们开发的一般过程基本相同，即均参照 GB/T8566-2007 标准。但由于其设计思想不同，因此每个分析设计阶段所使用的工具和方法也不同。比如面向过程的程序设计使用面向过程的程序设计语言，而面向对象程序设计则使用面向对象的程序设计语言。再如功能需求分析阶段，在面向过程的开发中常使用功能模块图进行分析研究，而面向对象的开发则使用用例图来研究分析。面向对象图形建模和分析工具通常有思维导图、实体关系图、数据流图、程序流程图、UML 等。

总的来说，面向对象的软件开发有以下优点：

- (1) 编程思维与人类习惯思维一致。
- (2) 通用代码不需人为编写，而是使用成熟的代码，软件稳定性好。
- (3) 很多代码或对象可重复使用，节省人力成本。
- (4) 软件开发周期短。
- (5) 软件便于测试、修改和维护。

### 3.1.4 软件工程新技术

#### 3.1.4.1 软件重用技术

软件重用是指在不同软件开发过程中多次使用相同或相近的软件元素的过程。

软件重用技术提高了软件开发效率，降低开发成本，缩短开发周期，提高软件产品的质量。软件重用包括代码和软部件的重用、规格说明和设计过程的重用以及软件开发方法的重用。软件重用的流程如下：

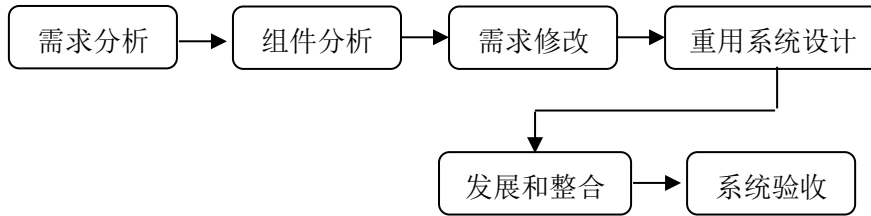


图 3.3 软件重用流程图

来源于：郑炜&吴潇雪（2016）

软件重用方法主要有开发方法的重用和软件工程的重用。开发方法的重用主要有面向对象重用方法、基于可视化重用方法、基于程序设计的重用方法。而软件工程重用主要包括应用系统重用、组件重用和对象功能重用。

### 3.1.4.2 程序代码自动生成技术

由于软件重用技术的快速发展，程序代码自动生成技术悄然流行了起来，而且发展快速，应用广泛，从嵌入式代码到 UML 图的生成均发挥重要作用。程序代码自动生成技术的出现甚至威胁到了程序员的工作岗位。

目前，程序代码自动生成方法主要有下面几种：（1）用框架自动生成基础代码。这一方法主要使用编程工具中的类库实现软件框架自动生成，如 .net 类库。（2）用脚本语言实现自动代码生成。这一方法通常使用 shell、perl 或 python 等脚本语言来自动生成代码。（3）用模版实现自动代码生成。这种方法主要用于 Web 开发上，这种代码生成包括访问数据库的数据库名称、类型、数据表以及各类数据处理操作均可自动生成。（4）用大数据云计算实现自动代码生成。大数据和云计算给代码自动生成提供了一个更加强大的支撑平台，在大数据和云计算下，程序代码自动生成将无所不能。（5）用人工智能自动生成代码。这一方法将通过机器学习、智能神经网络等技术，使得基于大数据云计算环境下的程序代码自动生成技术提高到更高层次，使程序代码实现完全自动化生成成为可能。

### 3.1.4.3 基于大数据云计算的软件工程智能化技术

随着软件系统规模越来越庞大和复杂，传统的软件工程方法解决软件问题也变得越来越困难。在大数据和云计算快速发展的背景下，基于大数据云计算新技术的软件工程智能化技术由此产生。这项新技术及应用发展很快，很多软件工程专业应用领域已经发展出许多新技术。比如庞大软件系统结构设计和项目管理，可以



应用基于智能搜索算法的软件工程技术，即把传统软件工程问题转化成基于搜索的优化问题，再应用智能搜索算法在大数据云计算平台中作问题解决。常用智能搜索算法包括遗传算法、爬山算法、模拟退火算法、蚁群算法等。

### 3.1.5 软件开发团队

#### 3.1.5.1 软件开发团队的组成

软件开发团队属于跨功能型团队(Cross-functional team)，其成员由不同工作领域的人员组成，是典型的 R&D 团队。但软件开发团队的队员一般以软件技术人员为主（如系统分析员、架构师、程序员、软件测试员、数据库工程师等），其他人员为辅（如项目管理员、视觉设计员、媒体制作员等）。软件开发团队的人员组成一般视软件项目实际开发的功能点来确定，但通常由 1-8 个软件技术人员、1 名团队负责人和 1 名项目经理组成。但如果项目需要，团队人员可多达 200 人以上。

#### 3.1.5.2 软件开发团队的结构模式

根据不同项目的人员需求，软件开发团队的结构模式形式多样，如窝蜂模式、主治医师模式、明星模式、社区模式、业余剧团模式、秘密团队模式、特工团队模式、交响乐队模式、爵士乐模式、功能团队模式、官僚模式等。大型软件公司常采用功能团队模式和官僚模式，但官僚模式容易变成“老板驱动”的开发流程。而中小型软件企业团队、敏捷开发团队和 TSP 团队（软件过程团队）最常见的结构模式如以下的团队组织结构图：

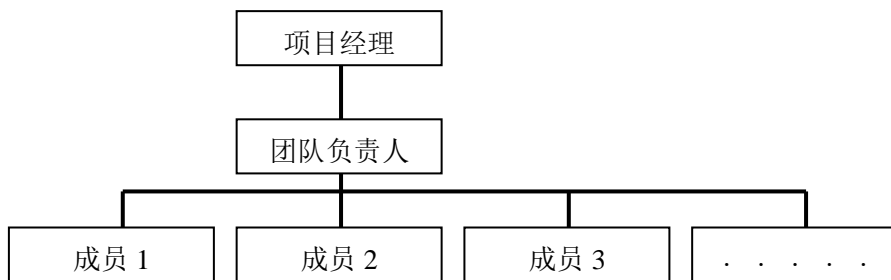


图 3.4 软件开发团队常见组织结构图

来源于：本论文整理

#### 3.1.5.3 软件开发团队的工作模式

软件开发团队的工作模式主要有集中开发模式和分布式开发模式。集中开发模式是指团队人员集中在一起办公的开发模式，分布式开发模式是指通过网络来实现不同区域的人员进行办公。前者对工作时间的要求很高，而后者则在工作时间上相对灵活。但从相关研究表明，集中开发模式比分布式开发模式工作效率更

高，集中开发模式比分布式开发模式的生产效率要高出 5% 左右。

#### 3.1.5.4 软件开发团队建设和管理

软件开发团队的建设与管理直接影响到整个团队的功能状态和研发效率，直接关系到软件项目 R&D 工作的成败以及团队目标能否实现。从归纳相关文献看，软件开发团队的建设与管理主要集中在四个方面，即建设高效团队、团队协作与沟通、团队学习和团队激励。

### 3.2 软件工程的大数据应用

传统的软件工程理论，一般以正向工程开始，工程完成后实施软件维护，如果 R&D 活动需要，再进行逆向工程或再工程等。而在大数据时代的软件工程则以逆向工程开始。由于软件重用技术大规模发展与应用后，软件资源大量积累，且大多为开源资源，而这些资源却散落在网络或各类软件企业、软件 R&D 团队、研究机构、从事软件技术的个体等等之上。因此大数据时代下的软件工程更注重数据挖掘和集约化生产。

软件工程的大数据应用最常见思想是软件工程微工程，即运用数学微积分的思想，对整个软件项目按层级进行“先微分后积分”。如软件体系和架构设计，先将其分割成可量化并且可重用的若干部分，我们把这些部分称为微资源。由于重用技术的快速发展，网络上已经积累大量的软件体系和架构的成功微资源，因此再通过数据处理、分析、挖掘、智能搜索、智能匹配等方法找到最适合本项目软件体系和架构的微资源。最后通过智能优化算法，把这些微资源像积木一样重新组合成完整的软件体系或架构。我们把最适合项目的方法、工具、资源等称为最佳实践。这些微资源的属性、特征等内容经过量化后，将通过机器学习记录在 IT 平台，以便再次使用时可以根据量化数据进行快速、准确、智能地匹配。由此可见，在大数据时代的软件工程中，是否能找到可量化且可复制的细粒度项目的最佳实践十分重要。软件工程的大数据应用，不仅体现在上述的软件体系和架构设计上，还可以应用在需求分析、数据库设计、程序编写、软件测试、版本控制等软件开发过程中的每一个阶段上。

软件工程大数据应用的核心思想是：以用户需求为导向，以重用成熟可靠资源为方法，以大数据平台实现软件开发资源的智能匹配。软件工程的大数据应用，不仅可以快速交付质量可靠的软件原型，还可以根据用户需求快速修改、测试、维护软件，这是传统软件开发所不具备的。

### 3.3 R&D 团队软件开发的大数据应用过程

软件 R&D 团队整个业务流程均以软件工程作为理论依据开展。软件 R&D 团队应用大数据为业务服务之前必须得到所有的项目业务问题，并将业务问题提交团队讨论与分析。团队每位成员必须对自己所承担的项目任务以及根据用户需求进行认真学习和思考后，根据自己的知识领域提出相应项目任务的解决方案。如存在未能解决问题则交由团队讨论解决。团队统一整理好方案后提交大数据服务公司开展大数据应用服务。大数据服务公司给 R&D 团队提交数据结果，R&D 团队对数据质量进行测评，同时进行合理性和价值分析。如在应用合理范围内，则进行快速整合并得到软件原型，并通过网络发布软件原型。再由大数据服务公司收集、分析用户对软件的评价、意见、建议及态度等，由 R&D 团队根据用户意向快速调整和修改方案和软件，直到用户满意度达到预期效果为止。以下为中小企业软件 R&D 团队应用大数据过程图：

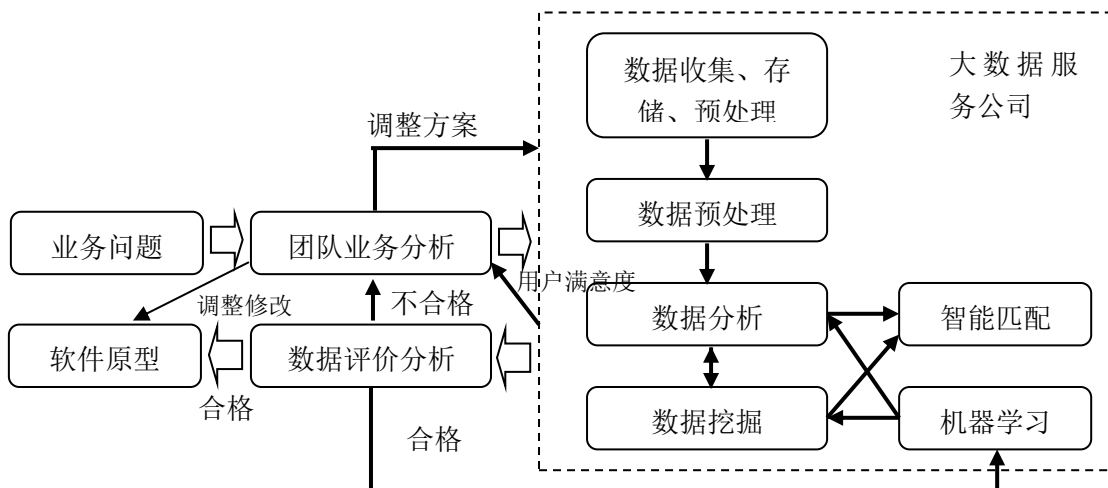


图 3.5 中小企业软件 R&D 团队应用大数据过程

来源于：本论文整理

### 3.4 软件 R&D 团队大数据应用与知识创造关系

软件 R&D 团队的软件开发过程就是知识创造的过程。而软件开发应用大数据其实是在已有资源基础上创造出新的知识。为了分析中小企业软件 R&D 团队应用大数据与知识创造的关系，我们将图 4.2.2 的过程简化为：业务问题、团队业务分析、大数据应用、数据评价和软件开发等五个阶段。从 Nonaka 和 Takeuchi 提出的知识创造 SECI 模型四个转化过程看，业务问题为社会化阶段，团队业务分析为表出化阶段，大数据应用和数据评价为联结化阶段，而软件开发则为内在化阶段，

其情境“Ba”为大数据平台。其关系如下图所示：

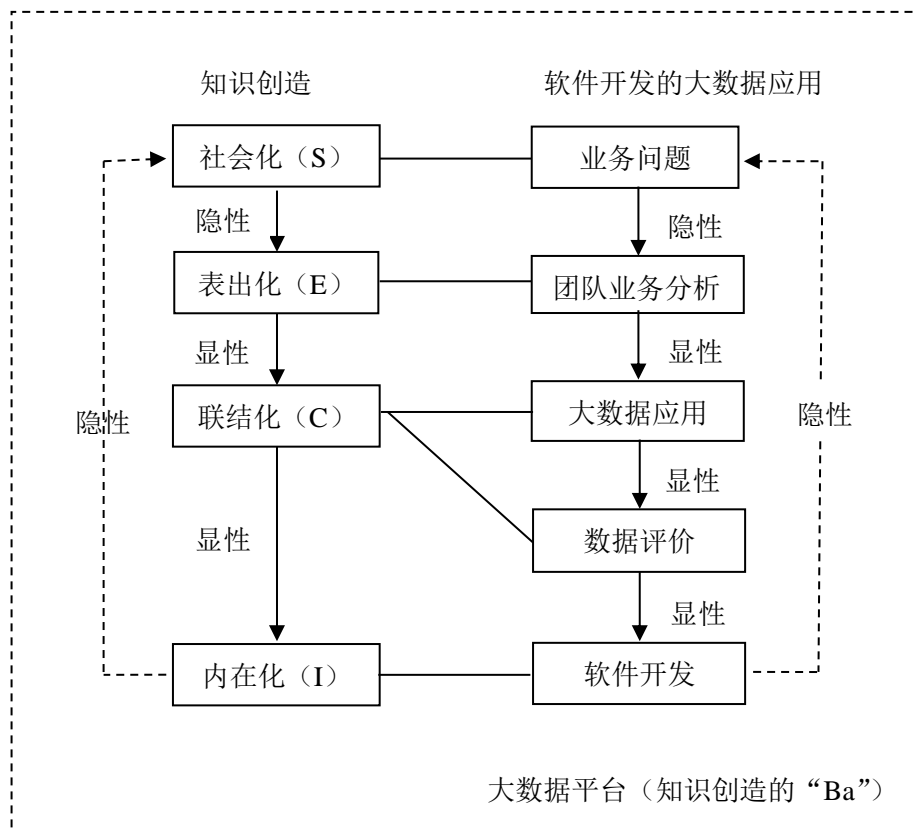


图 3.6 软件 R&D 团队大数据应用与知识创造关系

来源于：本论文整理

### 3.5 本章小结

本章节运用软件工程理论，研究了软件工程的大数据应用过程、软件企业 R&D 团队软件开发应用大数据的过程，软件 R&D 团队大数据应用与知识创造关系等，为后续的研究提供理论基础。

## 第四章 R&D 团队大数据应用知识创造过程案例研究

### 4.1 案例 1：保险客户行为分析系统开发项目

#### 4.1.1 软件公司及开发项目概述

深圳 H 软件服务公司是 H 保险公司于 2005 年投资成立的软件公司，注册资金 1000 万元，公司员工 156 人，其主要业务范围包括：H 保险公司系统内部业务、承接外包软件业务、技术咨询业务等。

保险客户行为分析系统是 H 保险公司为了动态了解保险客户投保行为而委托深圳 H 软件服务公司开发的一套分布式软件系统。该软件功能点达 1000 多个，可承受 20000 个网络业务站点同时连接，软件交付期限 12 个月。

#### 4.1.2 软件开发大数据应用及其知识创造过程

##### 阶段一：软件 R&D 团队组建

根据 H 保险公司初步提供的业务需求，结合员工的业务经验、技术背景、项目成本等方面考虑，快速组建一个 13 人的软件 R&D 团队。其中包括项目负责人 1 人，团队负责人 1 人，系统设计师 1 人，程序员 3 人，数据工程师 1 人，需求调研人员 6 人。R&D 团队各类人员职责如下表：

表 4.1 保险客户行为分析系统开发软件 R&D 团队各类人员职责

序号	人员类型	人员职责	备注
1	项目负责人	负责团队与团队外部的沟通与协调，保证软件 R&D 团队得到良好的研发环境。	
2	团队负责人	负责协调团队内部的各项工作。统一指挥每一个开发阶段的工作，保证每阶段工作按质、按量、按时完成。	
3	系统设计师	根据调研结果做用户需求分析，并对整个软件体系、架构进行设计，包括系统数据库设计。	
4	程序员	负责软件功能模块的整合，补充代码的编写，功能模块间冲突解决、模块及系统测试等。	
5	数据工程师	对软件的客户行为分析设计算法，包括数据处理、数据挖掘、智能分析、机器学习等算法。	
6	需求调研人员	深入到 H 保险公司系统内的分公司，甚至业务站点了解业务流程和业务情况。	属于团队临时人员

来源于：本论文整理

## 阶段二：软件需求分析

项目负责人与 H 保险公司沟通，让需求调研人员深入到分公司甚至业务站点做业务需求调研。需求调研人员由团队负责人统一分派任务。需求调研人员必须深入分公司或具体业务站点，通过观察、面对面交流等方法对业务流程、业务情况进行了解，并做详细记录。记录方式包括文档、图像、声音、视频等。调研完成后由团队负责统一收集整理，并交系统设计师开展软件需求分析工作。

## 阶段三：软件系统设计

系统设计师根据调研材料，按专业和行业规范，对整个软件系统做出具体的需求分析，并形成报告文档。系统设计师再根据需求报告对整个软件系统做体系结构、软件架构、功能模块及结构、数据库的设计按可重用和可量化原则进行分块，然后提取每个设计块的特征，量化并描述每个设计块特征，并形成每个设计的相关问题。

将这些设计块的量化特征和问题提交大数据服务公司，并对大数据服务公司采集的数据源提出相关要求或建议。大数据服务公司进行大数据处理，搜索挖掘已有的、成熟的、可重用的设计方案，结合智能优化算法和机器学习，最后给出每个设计块的最优设计方案。系统设计师对大数据服务公司提出的设计方案进行评价，如果符合预期，则系统设计师同意按其方案设计，或者再进行适当的修改和优化。如果不符合预期，则再次调整设计块特征指标和相关问题，并再次交由大数据服务公司分析挖掘数据，直到符合预期为止。

## 阶段四：软件原型的开发

数据工程师根据上述设计需求，提出软件中相关数据处理和分析算法的相关要求和问题，并提交程序员。程序员对整个软件系统的功能模块提出具体的编程要求、程序块量化特征和形成相关问题，再将程序块量化特征和相关问题提交大数据服务公司进行数据采集、存储、处理、挖掘和智能分析识别，最后给出最符合程序块量化特征及问题的程序代码。程序员对大数据服务公司提交的程序代码做评价，符合预期则进一步调整、修改、补充编写、整合和测试，并最终得到软件原型。不符合则重新调整程序块量化特征和问题，并再次将由大数据服务公司进行数据挖掘和智能分析识别，直至符合预期。

## 阶段五：发布软件原型和用户试用

完成软件原型的开发后，即在网络上发布软件原型，并要求用户试用。同时给予网络平台让用户反馈试用结果。

## 阶段六：调整修改软件原型和软件交付

根据用户反馈意见，对软件原型反复进行调整和修改，直到网络平台反馈的满意度超过 80% 后，即可向用户提出验收和交付。

本软件开发项目最后产生的效果是：

(1) 交付时间从 12 个月缩短到 7 个月，研发效率几乎提高 50%。

(2) 程序块代码基本不用测试，程序测试重点放在软件的整体测试上。软件质量得到大幅提高。

(3) 由于程序员数量较之以往项目少很多，设计、代码编写、软件测试等环节的工作量也较之以往少很多，因此开发成本也大幅降低。

(4) 可以根据用户导向快速修改和调整软件功能，修补软件 BUG。

#### 4.1.3 案例中的软件开发大数据应用对知识创造影响分析

在访谈过程中每位受访者对以上案例的兴奋之情均溢于言表。可以说，以上案例属于软件开发大数据应用的一个成功案例。但在成功的同时，也发现了一些大数据应用过程中影响 R&D 团队研发的因素。现分析如下：

(1) 技术外包因素。企业将大数据处理业务外包给大数据服务公司，大数据服务公司是按需按量进行收费的。为了降低开发成本，部分设计块或功能模块并没有开展大数据应用服务，因此一定程度上影响了研发效率和软件质量。

(2) 个人能力因素。案例中大数据应用过程出现由于个人的专业能力或者数学能力等原因，使得在设计分块、程序分块、特征提取及量化、问题拟定和结果评价等环节上不尽人意，并导致连续的挖掘失败，从而增加了开发成本。

(3) 智能化因素。智能化指的是通过智能算法，在网络上搜索、挖掘、分析、识别和匹配，从而自动完成相关工作。本案例曾存在智能算法适用性和机器学习深度不够的问题，导致部分特征和问题因此无法识别和匹配，最后所提交的结果不符合预期，从而影响软件研发效率和软件质量。

(4) 可视化因素。大数据可视化指通过各种表达方式（如文本、图形、声音、视频等）将结果动态地表现出来，以辅助人们更为直观、高效地洞悉大数据背后的信息、知识与智慧。可见大数据可视化可以影响大数据应用。而在本案例大数据应用过程中也曾出现因可视化问题导致对结果不理解，甚至误解的现象。

大数据可视化与传统 IT 可视化不同，传统 IT 可视化表示静态信息，是静态的可视化。大数据则是动态表示数据，是动态的可视化。而且大数据可视化还需要表示出数据的关联性和规律性。因此大数据可视化较之传统 IT 可视化更加复杂，要求更高，是一种强可视化。

(5) 团队协作能力因素。从软件工程理论和实际项目开发上看，无论是传统方式的软件开发或者是大数据应用下的软件开发，软件 R&D 团队均为软件项目开发的核心单位。因此，R&D 团队内部的协作能力，可以影响软件开发的结果。

实际上在案例中，大数据应用过程曾经出现过由于团队沟通不畅，使得在设计分块、程序分块、特征提取及量化、问题拟定和结果评价等环节上不尽人意，并导致连续的挖掘失败，从而增加了开发成本的问题。

软件研发过程实际就是知识创造的过程，因此以上影响因素就是软件 R&D 团队大数据应用下知识创造的影响因素。

## 4.2 案例 2：软件外包测试项目案例

### 4.2.1 公司及项目概述

Z 软件技术服务公司为桂林高新区的一家软件企业。公司成立于 2010 年 3 月，注册资金 300 万，目前拥有员工 39 人，主要业务范围包括：软件开发服务、软件测试服务、软件外包服务等。

2016 年承接一个软件项目的外包测试服务工作。该软件为分布式物联网软件系统，有 1500 个功能点（其中 230 个功能点为硬件驱动和自动化控制），客户端网络连接点从 0 至 30000 个，为可控连接。测试服务必须按软件开发阶段全程动态跟踪和测试。

### 4.2.2 软件测试大数据应用及其知识创造过程

阶段一：组建 5 人的测试团队。由团队根据软件设计的相关文档来研究软件每个设计环节，并设计和制定其测试方案，甚至有针对性地提前开发测试小工具。

阶段二：租用大数据服务商的大数据平台实施大数据动态跟踪测试。

阶段三：在大数据服务商现有相关数据源的基础上进行数据源拓展。拓展的数据源包括公司自存资源（包括各种软件跟踪测试记录文档、日志文件、硬件报错信息等）以及公司与其他公司结成的行业联盟所提供的软件测试共享资源。

阶段四：利用流处理技术，动态跟踪软件开发过程，并通过大数据平台的挖掘分析功能，智能搜寻当前开发与数据源中有相同或相近的情况，并预测开发过程可能会出现错误，错误预测准确则产生新的测试记录。如存在错误，但不能准确预测错误，则进行人工测试，问题解决后再形成新的测试记录。

阶段五：软件初步完成开发后，同样利用上述方法对整个软件每个功能点以及网络连接压力均进行测试，并给出测试报告，让软件开发商再次修改错误和缺陷。完全通过测试后即给出测试通过报告，在软件通过交付验收后即宣告结束。

### 4.2.3 软件测试大数据应用对知识创造影响分析

根据本案例大数据应用过程所遇到的问题，现将影响大数据知识发现结果及其应用效果的因素作如下分析：

（1）数据可用性因素：由于数据源劣质数据比例过高，导致所发现的知识错误和测试预测不准确。但由于大数据应用服务的外包，企业的大数据应用往往不用考虑数据源的质量问题。因此数据可用性因素可包含在技术外包性当中。



(2) 个人数据分析能力因素。由于测试业务的开展采用租用大数据平台的方式，尽管大数据平台已经集成诸多数据挖掘、智能分析和识别等工具，但因测试情况的不确定性和复杂性，平台上的工具仍无法满足要求，因此对测试人员个人的数据分析能力要求比较高。

(3) 技术实时性因素：由于大数据动态产生大量数据，因此对网络稳定性和带宽要求较高。如果网络无法满足大数据动态处理和分析的需要，其应用效果也会大打折扣。

(4) 团队协作能力因素：从软件工程理论可知，软件测试是一种分工协作的工作，团队成员之间需要有较高的工作默契。因此，如果测试团队成员间的协作能力差，则会导致测试漏洞百出，增加人力成本和时间成本。

### 4.3 大数据应用下的知识创造过程

从以上两个案例看，大数据应用下 Nonaka&Takeuchi 提出的 SECI 知识创造理论知识创造四个转化模式并没有发生改变。但由于大数据应用改变了知识创造方式，因此导致知识创造行为发生了改变。对于知识创造行为的定义，Nonaka(1994)认为个体的知识创造行为是个体与外部环境持续互动的过程。

大数据应用下，知识转化模式中的部分知识转化行为发生了转移现象，即从一个转化模式向另一个转化模式转移。如传统软件开发方式中的程序代码编写由人工完成，属于将显性知识进行实践，并转化为自身的隐性知识的内在化过程。但从上述案例看，大数据应用下大部分程序代码并非由人工完成，而是由大数据技术自动生成，这样的过程属于显性知识转化组织或个体开展知识创造的一系列行为活动。为显性知识的联结化过程。因此，大数据应用下，原属于内在化过程的程序代码编写的知识创造行为转移到了联结化过程。

由此可得到结论：**SECI 知识创造四个转化模式中的知识转化行为不是固定在某个转化模式中，它可以随着“Ba”的变化而发生转移。**

### 4.4 本章小结

本章根据中小软件企业 R&D 团队软件开发和软件测试两个案例，分析和研究了软件 R&D 团队大数据应用知识创造过程发生的变化及其影响因素。并由此得到了本论文研究的第一个结论。同时，也获得了大数据应用下，R&D 团队知识创造过程可能的影响因素，为后续的研究提供依据。

## 第五章 理论假设及研究模型

### 5.1 主要研究变量

#### 5.1.1 大数据应用概念

Cooper R.B.&Zmud R.W. (1990) 从技术创新视角提出 IT 应用的概念, 即 IT 应用是企业信息技术框架下, 鼓励信息技术运用并发挥其作用。邝孔武&王晓敏 (1999) 认为, IT 应用指信息技术在企业的应用。王嵘冰(2015)则将云计算应用定义为, 企业采用云计算这一新兴的信息技术并通过其技术特性的应用扩散来提升企业的竞争优势。为此, 本论文也从大数据技术创新视角来定义大数据应用, 即大数据应用是指企业采用大数据技术这一新兴 IT 技术, 并通过其技术特性应用的扩散来为企业创造价值, 以此来提升和保持企业核心竞争力。而大数据应用技术特性是指大数据应用过程中其技术所具备的内在、特有的性质。

#### 5.1.2 大数据应用维度

从前面文献回顾看, 大数据的数据特性有 5V 特性, 即规模性 (Volume)、多样性 (Variety) 和快速性 (Velocity)、价值性 (Value)、真实性 (Veracity)。而对于影响大数据应用的技术特性, 学者们也提出不少观点。如任磊等 (2014) 在大数据可视分析综述中认为, 大数据的自动智能分析、动态分析和强可视化是推动大数据应用发展的基石。孙大为等 (2014) 通过分析大数据的流式计算指出, 大数据技术具有实时性、可伸缩性、系统容错性、状态一致性等特性。郭平等 (2015) 认为大数据智能计算是大数据技术新方向, 智能化将成为大数据应用新的技术特性。李建中等 (2016) 则对大数据的可用性作了论述, 认为大数据的可用性影响大数据知识发现和预测决策。从诸多文献来看, 大数据应用的技术特性才是推动企业应用大数据的根本原因。因此本论文将从大数据创新的应用特性出发, 研究大数据应用对企业知识创造的影响。

综合本文案例分析和大数据应用定义, 个人能力属于个体属性, 团队协作能力属于团队属性, 均不属于大数据应用的技术特性。技术实时性从各种大数据技术文献看, 其包含在数据处理智能化范围。根据第四章的分析, 数据可用性包含在技术外包性中, 实时性则在强可视化中得以体现。因此得到软件企业大数据应用过程中影响软件 R&D 活动知识创造的主要因素有: 技术外包因素、智能化因素、强可视化因素等三个。综合大数据应用技术特性相关文献我们将大数据应用技术特性归纳为: 技术外包性、技术智能性、强可视化性等 3 个特性维度。以下逐一

对 3 个特性维度进行分析。

### (1) 技术外包性

大数据应用和传统的 IT 应用有所不同，大数据应用过程中资源可以由“买”或“自建”转向“租”或“外包”，以此来提升企业价值创造能力。企业不用为基础设施犯愁，不用投入精力维护 IT 系统，不用提高成本招聘技术人才。这种技术模式可以扩展企业的 IT 能力，拓展企业业务，提高企业创新能力。

### (2) 技术智能性

大数据应用最核心的价值在于它能够智能分析、挖掘和发现新知识，并通过匹配重用、预测、辅助决策等方式来创造价值。大数据这一技术特性，使得企业生产效率、生产质量、价值创造能力以及技术创新能力等方面大幅提高，同时降低了企业生产成本。

### (3) 强可视化性

强可视化性是相对于传统 IT 可视化而言。由于大数据的体量巨大，类型复杂、时效性高及价值密度低，因此将数据背后蕴含的规律、知识和智慧尽可能地呈现出来是大数据应用的一个关键点。

以下为本论文的大数据应用维度模型：

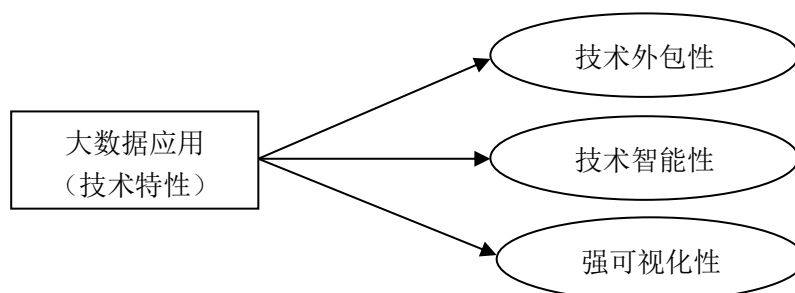


图 5.1 大数据应用维度模型

来源于：本论文整理

## 5.1.3 个人能力因素问题

个人能力影响知识创造或者影响信息技术应用下的知识创造已有诸多研究，如吴翠花(2013)和徐璞(2014)等。这些研究结果均认为个人能力确实可以个人能力影响知识创造或者影响信息技术应用下的知识创造。但在本论文的研究当中，个人能力显然不属于大数据技术的应用特性范畴，不适合做为自变量加以研究。而作为调节或中介变量加以研究，则在理论上不支持。根据罗胜强&姜嫵(2014)认为，“低层阶”变量不能调节或中介“高层阶”关系，如果在研究中发现“低层阶”变量调节或中介“高层阶”关系，这也只是属于随机误差。显然，在本研究中，个人能力属于“低层阶”变量，而大数据应用与 R&D 团队知识创造属于“高层阶”

关系，因此个人能力并不适合作为调节或中介变量在本论文中加以研究。

综上所述，个人能力不适合在本论文中加以研究。

#### 5.1.4 团队协作能力问题

团队协作能力的研究成果也很多，如 Kuhlman&arshello(1975)、Dreu&McCusker (1997)、Roch&Samuelson(1997) 在社会心理学领域的研究发现，任务结构、群体规模、成员之间的沟通、个体价值取向、任务目标等因素都可以影响群体协作行为。中国学者宋源(2009)、谢海波(2010)、赵明华(2011)等则认为，团队领导、团队构成、团队任务、团队规范、团队沟通、团队目标、团队成员个人特点、团队成员个人学习能力、团队激励等被认为是影响团队协作能力的主要因素。在本论文研究中，团队协作能力属于团队层面，但又不属于大数据应用特性，因此可以做为调节变量加以研究。

#### 5.1.5 R&D 团队知识创造维度

根据上述 KIBS 服务与知识创造关系和软件 R&D 团队大数据应用与知识创造关系可知，KIBS 企业的 R&D 团队知识创造过程与 Nonaka&Takeuchi 提出的 SECI 知识创造过程高度相符。因此，本论文将以知识的社会化、表出化、联结化和内在化作为软件 R&D 团队知识创造的 4 个维度。

### 5.2 研究假设提出

#### 5.2.1 大数据应用与 R&D 团队知识创造关系

本研究是基于信息技术理论和知识创造理论来研究大数据应用与 R&D 团队知识创造的关系。Davenport&Prusak(1998)就认为信息技术是知识创造过程当中的一个重要因素。诸多文献也指出，信息技术对知识创造有促进作用。

大数据技术是一项新兴的信息技术。大数据通过智能和动态的分析、挖掘和识别来帮助人们更高效地发现和创造新的知识。比如传统的软件开发方式程序代码编写应属于知识内在化的过程。但从案例 1 可知，大数据应用后程序代码由大数据平台依靠重用技术自动生成，最后由程序员整合成完整的软件系统。这样的开发过程使得研发效率和产品质量大幅提高。

通过案例可知，大数据应用对知识创造四个转化过程均起到促进作用。因此本研究认为，大数据应用会对 R&D 团队知识创造产生正向影响。提出以下假设：

**H0：大数据应用对 R&D 团队知识创造产生显著正向影响。**

本论文认为，大数据应用对 R&D 团队知识创造的影响包括四个维度，即知识社会化、知识表出化、知识联结化和知识内在化。

### 5.2.2 大数据应用与 R&D 团队知识社会化的关系

R&D 团队成员深入到用户的业务过程，通过观察、沟通甚至采用顶岗的方式了解用户的业务流程和业务需求，将用户的隐性知识转化为自己的隐性知识，这个过程就是 R&D 团队知识社会化的过程。

由于用户不具备软件专业化知识，无法给出解决业务问题相应的 IT 功能。因此在这一过程中，R&D 团队可以应用大数据分析相似业务及其 IT 功能的客户导向，并向客户推荐相应的业务 IT 功能，从而影响用户需求。因此本论文认为大数据应用对 R&D 团队的知识社会化具有正向影响。并提出以下假设：

**H1a: 技术外包性对 R&D 团队知识社会化产生显著正向影响。**

**H1b: 技术智能性对 R&D 团队知识社会化产生显著正向影响。**

**H1c: 强可视化性对 R&D 团队知识社会化产生显著正向影响。**

### 5.2.3 大数据应用与 R&D 团队知识表出化的关系

软件开发的需求分析阶段，是根据调研人员的记录、口述等作整理，并将业务调研结果转化成符合专业规范的文档，这一过程将需求调研获得的知识进行编码。然后，根据需求分析设计出具有创新性的软件体系结构、软件架构、功能模块等。这一过程是知识创造的过程，是隐性知识转化为显性知识的过程，也是创新性较高的过程。

传统的软件研发这一阶段完全由系统分析师人工完成。但在大数据应用下，很多需求分析与设计通过大数据相关技术，将根据蕴藏在大数据中的客户导向，新技术、新思想以及网络共享的且可重用的分析设计挖掘出来，并自动地、智能地分析和识别最优设计方案。可见大数据应用下，原由人工完成的核心工作交由大数据来完成。因此本论文认为大数据应用对 R&D 团队的知识表出化具有显著的正向影响。并提出以下假设：

**H2a: 技术外包性对 R&D 团队知识表出化产生显著正向影响。**

**H2b: 技术智能性对 R&D 团队知识表出化产生显著正向影响。**

**H2c: 强可视化性对 R&D 团队知识表出化产生显著正向影响。**

### 5.2.4 大数据应用与 R&D 团队知识联结化的关系

在完成需求分析和软件系统整体设计后，研发活动进入了程序代码编写和软件测试阶段。这一阶段首先是根据企业现有文档资料、设计资料等显性知识加以利用，其次将获取的知识转变为计算机可以接受的编码，最后将计算机编码整合成软件的各种功能模块。这一过程不仅创造了新的知识，同时也是创新性最高的过程。这一阶段是从显性知识到显性知识的过程，是通过计算机编码化和网络化进行知识联结的过程。

传统的软件研发这一阶段同样依靠程序员（或测试员）人工完成。但在大数据应用和重用技术的快速发展下，程序代码的编写、程序测试过程可以由大数据平台完成。很多成熟稳定的程序代码被挖掘出来，并自动地、智能地分析和识别最优设计方案。同时可以根据各程序块的情况，预测测试结果，使程序员高效、高质量地解决软件所存在的问题。因此本论文认为大数据应用对 R&D 团队的知识联结化具有显著的正向影响。并提出以下假设：

**H3a: 技术外包性对 R&D 团队知识联结化产生显著正向影响。**

**H3b: 技术智能性对 R&D 团队知识联结化产生显著正向影响。**

**H3c: 强可视化性对 R&D 团队知识联结化产生显著正向影响。**

### 5.2.5 大数据应用与 R&D 团队知识内在化的关系

在完成程序代码的编写和测试后，软件研发进入最后的产品产出阶段。这一阶段由程序员按照软件的整体设计，利用编写好的功能模块，像积木一样，快速整合成一个完整的软件系统，同时完成软件的整体测试。测试通过后则发布软件原型，让用户进行试用。通过用户的试用反馈，程序员转变设计思想，对相关功能进行修改，直至用户满意为止。这是程序员个人通过实践，将显性知识变为隐性知识的过程，是程序员知识内在化的过程。程序员通过在实践中学习，改善个人心智模式，掌握技术诀窍，从而形成有价值的知识的过程。

大数据应用在这一阶段中，可以对软件系统整体进行动态跟踪测试，通过机器学习技术，提前准确地预测故障点和存在的软件缺陷，以帮助程序员快速找到存在的问题并加以解决。大数据还可以进一步挖掘用户取向来帮助软件修改，使软件最终成为用户满意度极高的产品。因此本论文认为大数据应用对 R&D 团队的知识内在化具有显著的正向影响。并提出以下假设：

**H4a: 技术外包性对 R&D 团队知识内在化产生显著正向影响。**

**H4b: 技术智能性对 R&D 团队知识内在化产生显著正向影响。**

**H4c: 强可视化性对 R&D 团队知识内在化产生显著正向影响。**

### 5.2.6 团队协作能力与大数据应用和 R&D 团队知识创造的关系

团队协作不属于大数据应用技术特性。但从案例中分析中可知，团队协作因素对大数据应用下 KIBS 企业知识创造有一定影响。团队协作能力强，成员间沟通无阻，且效率高，则有利于 KIBS 企业应用大数据技术，产生新知识，开发出新的知识产品。因此，本文认为团队协作能力对大数据应用和 KIBS 企业知识创造具有正向调节作用。并提出以下假设：

**H5: 团队协作能力对大数据应用和 R&D 团队知识创造的关系具有正向调节作用。**

### 5.3 研究模型提出

经过上述的理论分析，本论文提出了大数据应用对 KIBS 企业 R&D 团队知识创造影响的相关假设。其中，大数据应用包括技术外包性、技术智能性、强可视化性等 3 个技术特性维度；R&D 团队的知识创造则以 Nonaka&Takeuchi 提出的 SECI 四个知识转化过程作为知识创造维度。本论文研究的理论模型如图 5.2 所示。

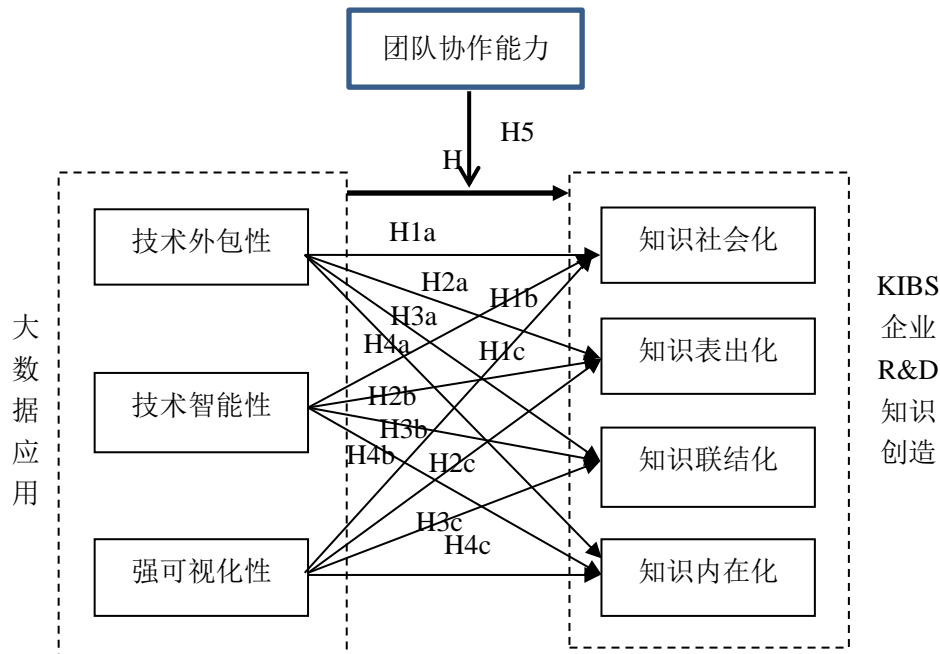


图 5.2 大数据应用对 R&D 团队知识创造影响研究模型图

来源于：本论文整理

### 5.4 本章小结

本章节首先界定了本研究的主要研究变量大数据应用的概念，其次通过知识创造 SECI 理论结合软件开发过程进行分析，并提出了本论文的研究假设，最后构建了本论文的大数据应用对 R&D 团队知识创造影响的研究模型，为接下来的实证研究提供了研究框架。

## 第六章 数据研究方法及问卷设计

### 6.1 数据研究方法

本论文的研究采用实证研究方法。实证研究范式是以实验或问卷调查的方式获取研究数据，通过统计分析方法检验所提出的理论假设。如果数据分析结果与理论假设一致，可以认为该理论假设成立，否则可拒绝该假设，最后得出研究结论。本论文的实证研究采用问卷调查方法收集数据，再运用统计分析方法对收集到的数据进行假设检验。

### 6.2 调查问卷设计

#### 6.2.1 设计原则

调查问卷是最常用的调研工具，通过调查问卷可以获得调研对象的相关数据和信息。可见，调查问卷的质量会影响到数据收集质量和效率，甚至影响到统计分析的结果。因此，调查问卷的设计必须遵循一定原则。综合李怀祖（2004）和艾尔·巴比（2005）的观点，我们提出以下调查问卷设计原则：

（1）主题明确。调查问卷内容必须围绕研究内容设置，每个题项必须精简明确，言简意赅。在确保问卷有效性和符合调研目的的前提下，问题尽量简单清楚，防止歧义，引起误解。

（2）确保问卷可答性。题项问题要通俗易懂，避免使用与受访对象不相关的专业术语，让受访者胜任答题工作。需要强调的是，为防止出于个人隐私考虑而不愿或随意答题，问卷必须采用匿名方式答题。

（3）答案设计合理。题项答案必须能覆盖所有的可能性，而且答案之间必须相互排斥，让答题者能明确其选择。

（4）结构严谨。调查问卷首先要对调研目的进行说明，让受访者理解并愿意作答。问题的排序需要从易到难，逐步引导受访者答题。

（5）问卷长度适中。问卷过长，会影响受访者的答题意愿和情绪。因此整个问卷长度要适中，尽量控制在 20 分钟内完成答题。

（6）进行预调研。问卷在正式发放前，最好能进行预调研。目的是从预调研中找出问卷存在问题，并及时调整优化问卷，提高问卷质量。

此外，为了方便受访者答题，使之方便时可在任何时间和地点答题，同时消除受访者的戒备心，本论文采用网络发放问卷的方式，同时承诺无需个人信息，数据仅用于研究，不会外泄。



## 6.2.2 变量测量

研究变量的测量均借鉴中国及其他国家已有的成熟量表。根据第五章所提出的研究假设，自变量为大数据应用，其包含技术外包性、技术智能性、强可视化性等 3 个维度。因变量为知识创造，其包含知识社会化、知识表出化、知识联结化和知识内在化 4 个维度。对于调查对象的观点，均采用 Likert 的五级标度法来衡量，答案区分为“非常不同意”、“不同意”、“不确定”、“同意”和“非常同意”，并分别给予 1、2、3、4、5 分来评分。

### 6.2.2.1 自变量的测量

从已有文献看，对新兴 IT 技术应用的测量研究较少，主要原因是缺乏基于统计分析的新兴 IT 技术应用实证研究。但近年来大数据、云计算等新兴 IT 技术基于统计分析的实证研究则有所增加。第三章对大数据和云计算的关系已有所论述，即大数据的基础设施主要依赖云平台，从云计算的两个重要应用特性看，大数据与云计算是绝佳配对。可见，大数据应用与云计算应用的技术特性是有重叠的，如技术外包性。因此，重叠部分大数据应用技术特性的测量可以借鉴云计算应用技术特性的测量。

目前，在中国对大数据应用测量最为系统的研究成果是蔡立志、武星、刘振宇于 2015 年所著的《大数据测评》，该书系统地介绍了大数据应用的测评方法和技术。为此，本文重点借鉴该书对大数据应用影响因素测评的方法。同时，本文也借鉴了王嵘冰(2015)关于云计算应用技术特性的测量。

#### (1) 技术外包性

从企业应用视角看，大数据应用的技术外包性与云计算一样包括两个目的，一是降低成本，提高成本效率。二是放下 IT 基础设施包袱，解放企业核心资源。因此二者的技术外包性是重叠的。因此，对于技术外包性的测量量表本论文参考了 Sultan(2011)、刘森(2014)和王嵘冰(2015)的研究。具体量表如下所示：

技术外包性使得对大数据服务商的服务能力、服务质量、数据质量等因素影响企业的大数据应用。Sultan(2011)认为，大数据服务商的服务能力主要指大数据分析服务提供能力。刘森(2014)和王嵘冰(2015)则认为大数据服务商的服务质量包括用户预期、用户评价和服务结果的质量等。对于数据质量，叶少波(2011)、蔡立志等(2015)认为，大数据服务商提供的数据质量体现在结果的准确性、结果与业务需求的一致性和数据动态更新。综合考虑以上研究成果，我们将通过分析服务能力、服务预期、结果评价通过率、实时服务质量、数据质量、业务需求一致性等 7 个指标测量技术外包性。

表 6.1 技术外包性测量量表

维度	测量指标	测量项	参考来源	
技术外包性	分析服务能力	大数据服务商能按公司需求提供 IT 服务	Sultan(2011)、刘森(2014)、王嵘冰(2015)、叶少波(2011)、蔡立志等(2015)	
	服务预期	大数据服务商提供业务参考符合预期率高		
	数据质量	动态更新		大数据服务商使用的数据源会持续更新
		一致性		大数据服务商使用数据源类型与业务需求相符
		准确性		大数据服务商所使用的数据得出结果准确性高
	结果评价通过率	公司对大数据服务商提供的结果评价通过率高		
信任度	相信大数据服务商能提供有价值的结果			

来源于：根据相关文献整理

### (2) 技术智能性

技术智能性主要依赖智能算法，如数据挖掘算法、网络神经算法、遗传算法等。蔡立志等（2015）提出大数据智能算法的测量应该包含三个方面指标，即预测精准率、智能化程度、自动化程度等。综合借鉴蔡立志等(2015)与刘甲学(2006)的相关研究，我们构建了以下技术智能性测量量表。和实时响应能力

表 6.2 技术智能性测量量表

维度	测量指标	测量项	参考来源
技术智能性	预测精准率	大数据服务商提供预测结果准确率高	蔡立志(2015)、刘甲学(2009)
	智能化程度	大数据服务能减少人工干预	
	自动化程度	大数据服务自动解决业务问题能力高	
	结果价值性	大数据服务经常提供新颖且有价值的结果	

来源于：根据相关文献整理

### (3) 强可视化性

蔡立志等（2015）认为大数据可视化测量包括多样性、通俗性、信任度三方面内容。多样性是指系统反应的信息足够丰富；通俗度是指可视化信息让用户通俗易懂；信任度是指用户对大数据服务的信任程度。任磊等（2014）则强调大数据可视化的实用性和扩展性。扩展性是指在显示设备无法满足大数据表示的信息量时，必须有良好的扩展性能来呈现数据。而李光达（2017）认为，可视化可以从信息量、复杂度、空间性、直观性、实时性等 5 个方面，而要实现实时性，必须有良好的基础设施（如网络基础设施等）做保障。综合上述观点，我们构建以下强可视化性测量量表：

表 6.3 强可视化性测量量表

维度	测量指标	测量项	参考来源
强可视化性	多样性	大数据服务提供的信息和呈现方式足够丰富	蔡立志(2015)、任磊(2014)、李光达(2017)
	通俗性	大数据服务提供的信息直观且通俗易懂	
	实时性	大数据服务能实时地、稳定地呈现相关信息	
	基础设施	网络稳定且信息更新速度快	
	扩展性	结果可以通过各种方式完整呈现	

来源于：根据相关文献整理

### 6.2.2.2 因变量的测量

根据研究模型因变量 R&D 团队知识创造的四个维度分别为：知识社会化、知识表出化、知识联结化和知识内在化。对于知识创造四个维度的测量研究成果很多，测量方法也非常成熟。本论文主要参考杨德林（2005）和纪慧生（2013）对以上四维度的测量。以下为本论文构建知识创造的测量量表：

表 6.4 知识社会化测量量表

维度	编号	测量项
知识社会化	1	团队加强与销售部门、客户、大数据服务商交流和信息分享
	2	注重团队内对软件开发的数据应用集体思考、经验总结和学习提高
	3	注重向团队外部学习，发现新大数据技术和市场机会，并形成创新
	4	团队内有互学互助、轻松和谐的工作环境和氛围

来源于：根据相关文献整理

表 6.5 知识表出化测量量表

维度	编号	测量项
知识表出化	1	团队内鼓励进行头脑风暴和对话来解决软件开发大数据应用问题
	2	团队内的大数据应用技巧和经验进行文字和规范整理
	3	团队内创造总结交流大数据应用的技术、经验、知识和思想的机会
	4	鼓励团队内使用比喻、类比、归纳等描述新的大数据应用知识
	5	团队对优秀员工进行宣传、经验推广和交流

续表 6.5 知识表出化测量量表

维度	编号	测量项
知识表出化	6	团队内创造思考的环境与氛围，促进大数据应用知识共享方式和工具的使用
	7	团队内建设大数据应用知识共享系统，制定知识贡献的激励机制

来源于：根据相关文献整理

表 6.6 知识联结化测量量表

维度	编号	测量项
知识联结化	1	收集团队内部和外部关于软件开发大数据应用的文献资料，通过预测等方法制定团队计划
	2	收集各类管理数据、技术资料来增加团队的知识储备
	3	团队通过应用大数据组织研发多款软件产品
	4	制定完善软件产品的服务手册及相关资料
	5	在团队内传播大数据相关的新技术、新概念和新思想
	6	运用现代化信息技术工作在团队内进行大数据技术传播
	7	鼓励员工积极查阅与大数据应用技术文献资料并提供创新建议

来源于：根据相关文献整理

表 6.7 知识内在化测量量表

维度	编号	测量项
知识内在化	1	经常通过跨部门研发项目与外部技术团队沟通大数据应用技术
	2	团队注重软件开发大数据应用相关知识和技术的收集、整理与共享
	3	团队内部通过沟通、理解来分享团队对软件开发大数据应用的愿景和价值理念
	4	团队注重整体软件开发大数据应用能力的提高，形成“传帮带”机制
	5	注重团队内部的软件开发大数据应用相关知识和业务技能的培训

来源于：根据相关文献整理

### 6.2.2.3 调节变量的测量

团队协作能力量表则参考鲍明刚(2004)和 Luo, Slotegraaf&Pan(2006)的相关研究。其主要通过沟通能力、互助能力、学习与经验分享、团队任务协作等方面测量团队协作能力。以下为本文构建团队协作能力的测量量表：

表 6.8 团队协作能力测量量表

维度	编号	测量项
团队协作能力	1	团队成员之间常互相帮助解决软件开发大数据应用上的问题
	2	团队成员之间常提出一些软件开发大数据应用有用的意见和建议
	3	团队成员之间常介绍自己在软件开发大数据应用上的经验和教训
	4	团队成员获得软件开发大数据应用新技术或知识后会与同事共同学习
	5	团队成员常在一起讨论项目开发的大数据应用相关问题

来源于：根据相关文献整理

### 6.2.3 问卷设计

本论文的研究是基于企业应用视角的大数据应用研究。目前，中国大数据应用还处在初步的发展阶段，其普及面还小，成熟度较低。况且，大数据技术是一种新兴且更高级别的信息技术，其往往是企业的核心技术，因此大部分企业不会向社会公开相关信息。因此，本论文研究无法进行大面积的深度访谈，只能通过调查问卷的方式来获取研究数据。

#### 6.2.3.1 问卷总体设计

本论文的调查问卷总体按以下设计步骤开展：

(1) 借鉴已有文献。本论文通过对大数据应用特性和 R&D 团队知识创造有关测量或评价的相关文献进行梳理，找到和本论文研究内容相符且经过检验的量表作为基础，并结合实际研究内容形成新的论文研究基础量表。

(2) 前测性访谈。形成基础量表后，邀请本领域 3 名副教授、2 名博士和 3 名博士生组成专家团，结合文献和研究内容对调查问卷相关题项进行修正。

(3) 预调查。为了使缺乏相同理解能力的受访者容易理解，消除各种疑问，使问卷调查能顺利开展。调查问卷在正式发放前，对南宁 5 家已经应用大数据的中小软件企业进行预调查。最后根据反馈意见对问卷进行再次修正。

(4) 小样本预测试。问卷经过多次修正后，还要以小样本数据进行定量检验，并根据预测试结果找出问题，提纯题项后形成正式的调查问卷。

调查问卷的结构分成四部分。第一部分为软件企业和员工个人基本情况。软件企业基本情况包括：企业性质、规模、成立年限和大数据应用时间等。员工个人基本情况包括：性别、年龄、教育背景，工作类型等。第二部分为自变量大数据应用的测量。其相关测量指标和题项参考 Sultan(2011)、刘森(2014)、王嵘冰(2015)、蔡立志(2015)、刘甲学(2006)、任磊(2014)、李光达(2017)、叶少波(2011)、李建中(2016)等。第三部分是因变量 R&D 团队知识创造的测量。其主要参考杨德林(2005)和纪慧生(2013)相关知识创造的测量量表。大数据应用技术特性从 3

个维度衡量,共 16 个题项。R&D 团队知识创造也从 4 个维度衡量,共 24 个题项。

此外,本论文对问卷题项做如下编码设计:

D 为大数据应用, D1 为技术外包性, D2 为技术智能性, D3 为强可视化性。Z 为知识创造, Z1 为知识社会化, Z2 为知识表出化, Z3 为知识联结化, Z4 为知识内在化。N 为团队协作能力。所有变量均采用 Likert 的五级标度法来衡量,答案区分为“非常不同意”、“不同意”、“不确定”、“同意”和“非常同意”,并分别给予 1、2、3、4、5 分来评分。

### 6.2.3.2 测量 R&D 团队问卷设计

由于本论文研究主要问题是大数据应用与 R&D 团队内部知识创造的关系,因此需要对 R&D 团队内部进行测量,而且研究必须结合团队层面和个人层面。因此,我们必须收集整个 R&D 团队的内部数据,并在此基础上做 R&D 团队知识创造变量的分析。

收集整个 R&D 团队的内部数据非常困难,因此我们并不要求所有的团队成员必须答题,但答题人数须超过 3 人,团队答题人数达不到 3 人的当无效问卷处理。同时,由于需要注明团队成员,这可能导致答题不再匿名。因此,问卷采用了“成员代码=团队代码+成员编号”的方式标注每一位团队成员。

## 6.3 数据采集

本研究通过对中小型软件企业进行问卷调查的方式收集数据。由于企业的大数据应用还处在起步阶段,尽管近几年大数据企业应用有快速发展的态势,但就目前而言,应用广泛程度仍不高,因此样本量较少,这增加了研究的难度。但作者本人从事多年软件创新人才培养,因此有大量学生从事软件开发工作,而且岗位大到企业董事,小到普通程序员均有分布。这为本论文的调查研究提供了保障。

考虑到信息技术的应用必须具有持续性才显现出效果,企业大数据应用效果也可能存在一定的延迟性。因此根据陈永霞&贾良定&宋继文(2006)提出的一家企业至少要经历半年的过渡期的观点,我们选择的受访企业其大数据应用都要在半年以上。同时考虑到企业所属区域的环境影响,本研究除了选取广西本地企业外,还从北京、上海、广州、深圳等经济较发达地区选取调研企业,力求选取样本尽可能地分散和随机,从而为本论文最后研究结果的适用性提供保障。

为了能获取中国不同地域的研究数据,同时方便受访者进行问卷作答和数据回收。本论文采用网络发放调查问卷的方式来收集数据,具体采用了微信上的“问卷之星”来发放调查问卷。

根据学者 Tinsley&Tinsley(1987)建议,对于影响因素分析研究,建议题项数与样本数比例为 1:5 左右。本论文问卷题项共 59 项,公司和个人信息占 12 项,论

文研究实际内容为 47 项。因此我们期望回收问卷的数量在 300 -500 份之间。

### 6.3.1 小样本数据的收集

小样本数据目的是预测试问卷的合理性和科学性，找出受访者不易理解的问题描述，删除不科学的题项，进一步修正调查问卷。小样本数据主要利用腾讯公司开发的问卷星设计电子问卷，并通过微信发放和回收预测试问卷。发放地域主要在广西和广东 12 家企业，每家软件企业应用大数据至少半年以上，为了避免由于使用同种类型的测量工具导致的同源偏差，每家企业最多发放 5 份问卷，每个部门受访者最多 2 人。小样本预试共发放预调查问卷 92 份，回收 92 份，剔除 6 份无效问卷，最后有效问卷为 86 份，问卷的有效回收率为 100%，问卷有效率 93.5%。

### 6.3.2 大样本数据的收集

大样本数据收集同样通过腾讯公司开发的问卷星，通过微信实行“滚雪球”式发放和回收问卷。问卷面向全国发放，问卷共提交了 565 份，剔除 96 份无效问卷，最后有效问卷为 469 份，问卷有效率 83%。

### 6.3.3 大样本数据描述

本论文的大样本数据覆盖并包含小样本数据，因此这里主要对大样本数据进行描述。描述从问卷覆盖区域、公司业务范围、研发投入占销售收入的比例、公司应用大数据为公司业务服务时间、教育程度、岗位角色等方面展开。

#### 1. 问卷覆盖区域情况

由于问卷发放和数据收集通过微信，采用“滚雪球”方式进行，因此问卷所覆盖的区域非常广，其包括北京、上海、广东、广西、辽宁、江苏等省、市和自治区。具体情况如下图所示：

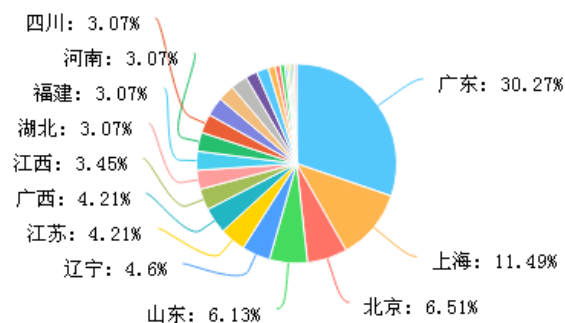


图 6.1 问卷覆盖区域情况图

#### 2. 公司应用大数据为公司业务服务的时间情况

答题者所在公司应用大数据为公司业务服务的时间情况如下图所示：

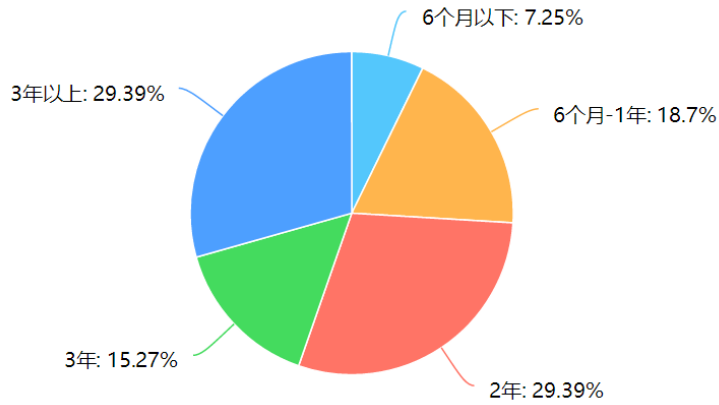


图 6.2 公司应用大数据为公司业务服务的时间情况图

### 3. 公司业务范围情况

答题者所在公司业务范围情况如下图所示：

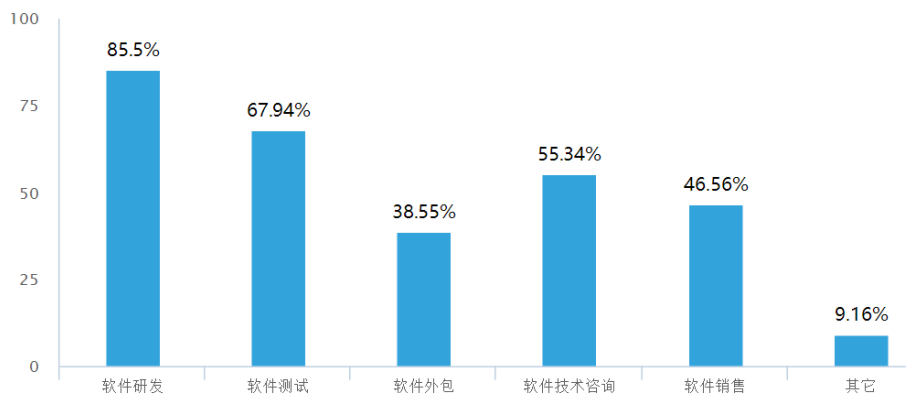


图 6.3 公司业务范围情况图

### 4. 研发投入占销售收入的比重情况

答题者所在公司研发投入占销售收入的比重情况如下图所示：

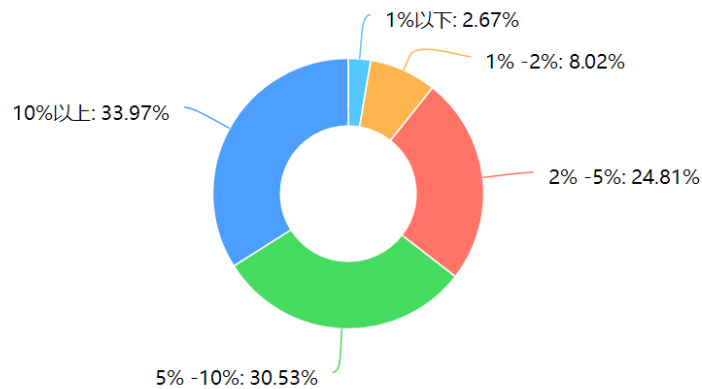


图 6.4 研发投入占销售收入的比重情况图



## 5. 受教育程度情况

答题者受教育程度情况如下图所示：

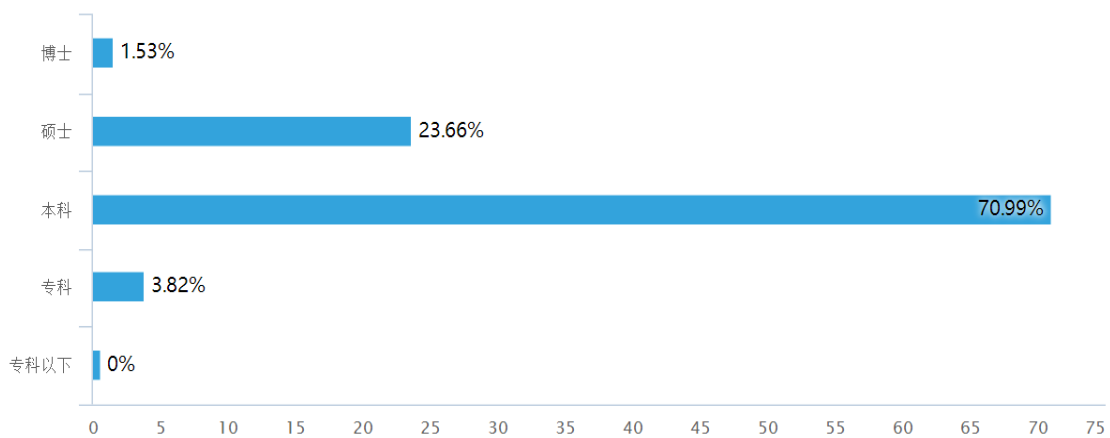


图 6.5 受教育程度情况图

## 6. 岗位角色

答题者岗位角色情况如下图所示：

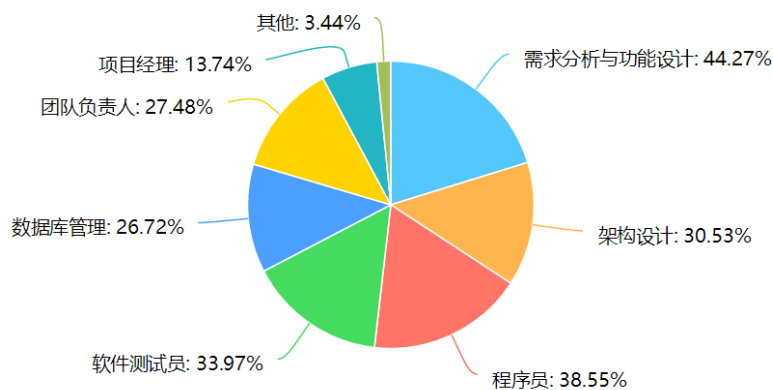


图 6.6 岗位角色情况图

从以上六个方面情况看，受访者所处地域在中国覆盖面广，公司应用大数据服务公司的时间均较长，六个月以上高达 92%。企业的业务范围均为软件企业常见业务，受访者的岗位角色也很多，受教育程度也很高，本科学历以上高达 96% 以上。大部分公司对 R&D 活动的投入也比较高。从以上情况看，本调查所获取信息来源较为全面，受访者学历普遍较高，也比较专业，具备了较好的专业知识和理解能力，使得问卷答题的准确性有一定的保障。

## 6.4 研究方法

选择合适的研究方法来检验理论假设模型非常重要。本研究应用 SPSS19.0 统

计分析软件对所收集的数据进行统计分析和假设检验。本研究所采用的统计分析方法主要有：描述性统计、信度效度检验、因子分析、相关分析和回归分析。

#### （1）描述性统计

调查问卷回收并进行数据录入后，要对受访企业、员工等基本情况进行分析，以此得到各变量的平均值、百分比、描述样本类别、特性及比例分配等情况。企业基本情况包括：企业性质、规模、成立年限和大数据应用时间等。员工基本情况包括：性别、年龄、教育背景，工作类型等。

#### （2）信度分析

为了了解问卷量表的可靠性，我们需要对问卷数据做信度分析。信度是反应测量量表在测量时的稳定性和一致性。信度是相对而言的，但信度是效度的必要条件，信度高度决定效度高度。

本研究采用 Cronbach's  $\alpha$  系数 ( $\alpha \in (0, 1)$ ) 信度检验方法对大数据应用和 R&D 团队知识创造中题项相互关联的内部一致性进行检测。利用  $\alpha$  系数对测量量表内部的一致性进行描述，如果  $\alpha$  系数值越大，表明量表测得结果越可靠。吴明隆(2003)对  $\alpha$  系数值的检验标准为， $\alpha \geq 0.8$  可靠性相当好， $0.7 \leq \alpha < 0.8$  可靠性可以接受， $0.6 \leq \alpha < 0.7$  可靠性最低接受值， $\alpha < 0.6$  量表或问卷不可信，需重新设计。

#### （3）效度分析

为了了解问卷量表的有效性，即问卷量表测量结果是否能测出本研究所要的结果，我们需要对问卷数据做效度分析。效度同样是相对而言的，它无法进行直接测量，只能进行间接测量和估量。

杜智敏(2010)提出效度分析由内容效度和结构效度组成。内容效度指研究主题在量表或问卷中的内容涵盖程度。内容效度通过本研究领域的专家依据相关理论进行逻辑分析和判断。结构效度指能测量到的理论结构和特质程度。结构效度分析通过因子分析进行检验。因子分析提取的公因子就是测量变量内容，通过某变量间的结构对问卷量表进行有效性检验，包括因子载荷和累积方差贡献率等。

#### （4）团队内部一致性分析

由于本论文需要做跨层次分析研究，即个人层面和团队层面。因此需要检验个人层面数据整合到团队层面中是否合适，这需要团队内部的一致性分析。

#### （5）相关分析

本论文就是研究自变量对因变量的影响。因此，就需要分析研究两个变量之间的相关性，即对自变量和因变量之间的相关性进行统计分析研究。本论文相关性分析采用 Pearson 积差相关系数进行检验。Pearson 相关系数取值范围如下表：

表 6.9 Pearson 相关系数取值范围表

取值范围(r)	含义
$r>0$	呈正相关
$r<0$	呈负相关
$0<r<1$	存在线性相关关系
$0<r\leq 0.3$	微弱相关
$0.3<r<0.5$	低度相关
$0.5<r<0.8$	显著相关
$0.8<r<1$	高度相关

### (6) 回归分析

相关分析只能说明变量和变量之间关系的密切程度，但并不表示两者有因果关系。比如相关性分析只能说明 A 和 B 有关系，但是 A 影响 B，还是 B 影响 A，相关性分析则无法描述。具有方向性关系的描述需用回归分析统计方法。

回归分析中的变量关系不是对称关系，回归分析中因变量是随机变量，自变量是非随机变量。回归分析可以对两个变量进行一元线性回归分析，也可以对一个因变量和多个自变量进行多元线性回归分析。回归分析是通过拟合方程的拟合优度来检验的，其指标包括标准化系数、非标准化系数、估计值标准误、决定系数  $R^2$  以及调整的相关系数  $R^2$  等。本论文研究有一个自变量对一个因变量，三个自变量对一个因变量的，因此本论文将采用一元线性回归和多元线性回归的方法进行检验和分析，基本可满足本论文研究需要。

## 6.5 本章小结

本章在理论研究模型和相关研究文献的基础上，设计了模型中变量的测量量表，再结合预访谈、专家意见形成本论文的量表。同时利用 Likert 的 5 级量表设计调查问卷。通过发放预调查问卷的方式查找问卷存在的问题并进行修正，为后面的数据分析和假设检验提供数据保障。此外，还对小样本进行预测试，为今后的实证研究奠定坚实的基础。

## 第七章 数据分析及假设检验

### 7.1 小样本的预测试

为了保证量表中各维度的准确性以及题项的合理性，除了采用如访谈等定性方法来修正以外，我们还需要采用定量的方法做进一步论证。因此，本论文在发放大样本的正式问卷之前，我们还将通过对小样本进行测试，用定量的方法检验并修正问卷。

#### 7.1.1 小样本数据描述性统计分析

我们对小样本受访者的公司和个人信息进行统计，得到以下表格：

表 7.1 小样本受访者公司和个人信息统计描述表

题项	题项分类	人数	比例
公司所在地	广东	32	66.67%
	广西	16	33.33%
公司业务范围	软件研发	46	95.8%
	软件测试	23	47.9%
	软件外包	42	87.5%
	软件技术咨询	48	100%
	软件销售	6	12.5%
	其它	0	0
公司年销售收入	100 万以下	6	12.5%
	100 - 500 万	28	58.3%
	500 - 1000 万	12	25%
	1000 万-5000 万	2	4.16%
	5000 万-1 亿	0	0
	1-10 亿	0	0
研发投入占销售收入的比例	1% 以下	0	0
	1% -2%	9	18.75%
	2% -5%	36	75%
	5% -10%	3	6.25%
	10% 以上	0	0

续表 7.1 小样本受访者公司和个人信息统计描述表

题项	题项分类	人数	比例
应用大数据为公司服务时间	6个月以下	0	0
	6个月-1年	6	12.5%
	1-2年	15	31.25%
	2-3年	23	47.9%
	3年以上	4	8.33%
性别	男	30	62.5%
	女	18	37.5%
年龄	20岁以下	0	0
	20-30岁	38	79.16%
	30-40岁	5	10.4%
	40-50岁	5	10.4%
	50岁以上	0	0
婚姻状况	已婚	29	60.4%
	未婚	19	39.6%
受教育程度	博士	0	0
	硕士	11	22.9%
	本科	37	77.08%
	专科	0	0
	专科以下	0	0
在软件研发团队内的工作	需求分析与功能设计	11	22.9%
	架构设计	16	33.33%
	程序员	32	66.67%
	软件测试员	9	18.75%
	数据库管理	9	18.75%
	团队负责人	7	14.58%
	项目经理	3	6.25%
	其他	0	0

来源于：本论文整理

从上表看，小样本问卷的受访者均来自于广东和广西，其所在公司业务范围均涵盖多方面的业务，单一业务没有。公司年销售收入主要集中在 100-500 万之间，这一范围体现了小型软件企业的营业额。从研发投入占销售收入的比例看，软件企业在研发中投入的比例比较大，2% -5%的比例范围占到 75%，这体现知识型企业对知识创造的重视程度极高。从应用大数据为公司服务时间来看，主要集中在 1-3 年，其比例高达 79.15%。可见，软件企业在近三年内对大数据的应用非常重视，并且已纷纷投入企业实践中。从年龄上看，20-30 岁有受访者占 79.16%，行业人员

平均年龄较年轻。受教育程度均为本科以上学历，受教育程度很大程度上会影响个人对知识的理解能力，受访者学历高，本身对问卷内容的明确程度和答案质量起到了一定的保障作用。

### 7.1.2 小样本数据信度和效度检验

在信度检验中，通常 CITC (Corrected Item-Total Correlation) 值大于 0.3，而且  $\alpha$  (Cronbach's  $\alpha$ ) 值在 0.6 以上，则说明该题项可以接受。如果  $\text{CITC} \leq 0.3$ ，则把该题项删除。然后再用 KMO 和 Bartlett 球形度检验来验证是否可以因子分析 ( $\text{KMO} \geq 0.7$ )，提取特征值大于 1 的因子，删除因子载荷值小于 0.5 的题项或者横跨两个因子且因子载荷值均大于 0.5 的题项。从而达到优化量表结构效度，并最终检验量表信度的目的。

#### 7.1.2.1 大数据应用信度与效度检验

##### (1) 信度检验

以下为 SPSS 19 检验出大数据应用各题项的 CITC 与  $\alpha$  值的结果：

表 7.2 大数据应用量表信度分析表

维度	题项编号	CITC	各题项 $\alpha$ 系数	量表总 $\alpha$ 系数
技术外包性	D11	0.741	0.904	0.914
	D12	0.635	0.907	
	D13	0.572	0.910	
	D14	0.616	0.908	
	D15	0.722	0.904	
	D16	0.569	0.910	
	D17	0.639	0.907	
技术智能性	D21	0.543	0.911	
	D22	0.501	0.913	
	D23	0.697	0.906	
强可视化性	D31	0.540	0.910	
	D32	0.696	0.906	
	D33	0.698	0.905	
	D34	0.669	0.906	
	D35	0.419	0.914	

来源于：本论文收集 SPSS19 检验数据后整理

从以上结果看，总体  $\alpha$  值为 0.914，可见量表总体可信度非常好。同时，所有

题项的  $\alpha$  值均大于 0.9, CITC 值均大于 0.4, 其数值均符合标准。

### (2) KMO 和 Bartlett 球形度检验

表 7.3 大数据应用量表 KMO 和 Bartlett 球形度检验表

取样足够度的 Kaiser-Meyer-Olkin 度量。		0.828
Bartlett 的球形度检验	近似卡方	437.187
	df	105
	Sig.	0.000

来源于: 本论文收集 SPSS19 检验数据后整理

从以上 KMO 和 Bartlett 检验的结果看, 大数据应用的 KMO 值为 0.828, Bartlett 球形值为 437.187, 统计量显著性水平 Sig 值为 0.000, 因此本研究大数据应用量表适合做因子分析。

### (3) 探索性因子分析

探索性因子分析常用于分析构造变量内在结构的方法, 也常用于量表的修正。其通过删除负载系数过低的变量来提高量表的信度和效度。探索性因子分析主要通过主成份分析的总方差分析, 并通过提取特征根大于 1 的公因子, 对量表结构进行检验。

根据 Hair et al (1998) 的观点, 在样本量  $\geq 50$  时, 因子载荷系数  $> 0.3$ , 则题项可认为显著的。载荷系数  $> 0.4$  时, 表示题项是重要的。载荷系数  $> 0.5$  则题项将被认为非常重要。

以下为解释总方差的检验结果:

表 7.4 大数据应用量表解释总方差分析表

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	6.949	46.326	46.326	6.949	46.326	46.326	3.446	22.971	22.971
2	1.684	11.230	57.555	1.684	11.230	57.555	3.175	21.166	44.137
3	1.089	7.257	64.813	1.089	7.257	64.813	3.101	20.675	64.813

来源于: 本论文收集 SPSS19 检验数据后整理

从以表 7.4 解释总方差的检验结果看, 三个公因子解释的总方差比例为 64.813%, 在可接受范围。其中, 解释比例最高公因子 1 为 22.971%, 公因子 2 解释比例为 21.166%, 公因子 3 为 20.675%。

再分析以下旋转因子载荷表:

表 7.5 大数据应用各维度旋转因子载荷表

因子	题项编号	因子载荷		
		1	2	3
技术外包性	D11	0.59		
	D12	0.56		
	D13	0.58		
	D14	0.62		
	D15	0.57		0.57
	D16	0.76		
技术智能性	D21			0.84
	D22			0.80
	D23			0.71
	D24			0.60
强可视化性	D31		0.89	
	D32		0.86	
	D33		0.64	
	D34		0.53	
	D35		0.54	

来源于：本论文收集 SPSS19 检验数据后整理

从上图看，D15 出现横跨 2 个因子且因子载荷值均大于 0.5，因此删除此题项，重新检验。其结果如下表：

表 7.6 大数据应用各维度旋转因子载荷表

因子	题项编号	因子载荷		
		1	2	3
技术外包性	D11		0.60	
	D12		0.56	
	D13		0.55	
	D14		0.66	
	D16		0.74	
技术智能性	D21			0.82
	D22			0.80
	D23			0.76
	D24			0.62



续表 7.6 大数据应用各维度旋转因子载荷表

因子	题项编号	因子载荷		
		1	2	3
强可视化性	D31	0.89		
	D32	0.86		
	D33	0.62		
	D34	0.55		
	D35			

来源于：本论文收集 SPSS19 检验数据后整理

从图 7.6 看，删除题项 D15 后，出现 D35 为空值（即载荷系数小于 0.5）。再删除题项 D35，重新检验。

首先检验经过两次修正后的 KMO 和 Bartlett 球形度，其结果如下：

表 7.7 大数据应用量表 KMO 和 Bartlett 球形度检验表

取样足够度的 Kaiser-Meyer-Olkin 度量。		0.834
Bartlett 的球形度检验	近似卡方	371.340
	df	78
	Sig.	0.000

来源于：本论文收集 SPSS19 检验数据后整理

由上表可见，KMO=0.834，Bartlett 球形值=371.34，统计量显著性水平 Sig=0.000，因此本大数据应用量表适合做因子分析。

其次看解释总方差，其结果如下：

表 7.8 大数据应用量表解释总方差分析表

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	6.197	47.666	47.666	6.197	47.666	47.666	3.168	24.367	24.367
2	1.586	12.196	59.862	1.586	12.196	59.862	3.152	24.248	48.615
3	1.034	7.955	67.817	1.034	7.955	67.817	2.496	19.202	67.817

来源于：本论文收集 SPSS19 检验数据后整理

从上表可知，共提取 3 个公因子，总方差累积为 67.817%，符合研究要求。最后看经过两次修正后得到的旋转因子载荷图，其结果如下：

表 7.9 大数据应用各维度旋转因子载荷表

因子	题项编号	因子载荷		
		1	2	3
技术外包性	D11		0.60	
	D12		0.53	
	D13		0.61	
	D14		0.71	
	D16		0.75	
技术智能性	D21			0.85
	D22			0.77
	D23			0.78
	D24			0.62
强可视化性	D31	0.90		
	D32	0.86		
	D33	0.59		
	D34	0.57		

来源于：本论文收集 SPSS19 检验数据后整理

从图 7.9 两次修正大数据应用旋转因子载荷图看，大数据应用量表结构得到了优化，符合研究要求。

### (3) 大数据应用最终量表信度分析

经过两次修正后的量表需要通过  $\alpha$  检验来验证最终量表的信度。以下为最终量表检验结果：

表 7.10 大数据应用最终量表信度分析表

维度	题项编号	CITC	各题项 $\alpha$ 系数	量表总 $\alpha$ 系数
技术外包性	D11	0.73	0.89	0.905
	D12	0.64	0.90	
	D13	0.56	0.90	
	D14	0.61	0.90	
	D16	0.58	0.90	

续表 7.10 大数据应用最终量表信度分析表

维度	题项编号	CITC	各题项 $\alpha$ 系数	量表总 $\alpha$ 系数
技术智能性	D21	0.62	0.90	0.905
	D22	0.54	0.90	
	D23	0.50	0.90	
	D24	0.69	0.89	
强可视化性	D31	0.55	0.90	
	D32	0.71	0.89	
	D33	0.68	0.89	
	D34	0.67	0.90	

来源于：本论文收集 SPSS19 检验数据后整理

从以上信度检验结果看，各题项的 CITC 值均大于 0.4，各题项的  $\alpha$  值均大于 0.8，量表总  $\alpha=0.905$ 。可见，大数据应用最终量表可靠性相当好，量表内部一致性也非常好，完全达到研究要求。

### 7.1.2.2 知识创造信度与效度检验

#### (1) 信度检验

以下为 SPSS 19 检验出知识创造各题项的 CITC 与  $\alpha$  值的结果：

表 7.11 知识创造量表信度分析表

维度	题项编号	CITC	各题项 $\alpha$ 系数	量表总 $\alpha$ 系数
知识社会化	Z11	0.60	0.949	0.950
	Z12	0.63	0.948	
	Z13	0.37	0.952	
	Z14	0.70	0.948	
知识表出化	Z21	0.62	0.949	
	Z22	0.72	0.947	
	Z23	0.80	0.946	
	Z24	0.54	0.949	
	Z25	0.62	0.948	
	Z26	0.77	0.947	
	Z27	0.75	0.947	

续表 7.11 知识创造量表信度分析表

维度	题项编号	CITC	各题项 $\alpha$ 系数	量表总 $\alpha$ 系数
知识联结化	Z31	0.52	0.950	0.950
	Z32	0.82	0.947	
	Z33	0.33	0.951	
	Z34	0.39	0.951	
	Z35	0.74	0.947	
	Z36	0.85	0.946	
	Z37	0.68	0.948	
知识内在化	Z41	0.62	0.948	
	Z42	0.75	0.947	
	Z43	0.69	0.948	
	Z44	0.69	0.948	
	Z45	0.79	0.946	
	Z46	0.75	0.947	

来源于：本论文收集 SPSS19 检验数据后整理

从以上结果看，量表总  $\alpha$  值为 0.950，所有题项的  $\alpha$  值均大于 0.9，CITC 值均大于 0.3，因此保留所有题项。

(2) KMO 和 Bartlett 球形度检验

从图 7.1.2.2-2 看，KMO 和 Bartlett 检验的结果为：知识创造的 KMO 值为 0.752，Bartlett 球形值为 1177.558，统计量显著性水平 Sig 值为 0.000，因此本研究知识创造量表适合做因子分析。

表 7.12 知识创造量表 KMO 和 Bartlett 球形度检验表

取样足够度的 Kaiser-Meyer-Olkin 度量。		0.752
Bartlett 的球形度检验	近似卡方	1177.558
	df	276
	Sig.	0.000

来源于：本论文收集 SPSS19 检验数据后整理

(3) 探索性因子分析

首先看解释总方差检验结果：

表 7.13 知识创造量表解释总方差分析表

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	11.827	49.277	49.277	11.827	49.277	49.277	4.202	17.508	17.508
2	2.290	9.541	58.818	2.290	9.541	58.818	4.078	16.992	34.501
3	1.708	7.118	65.936	1.708	7.118	65.936	3.649	15.205	49.705
4	1.291	5.381	71.317	1.291	5.381	71.317	3.242	13.508	63.213
5	1.107	4.610	75.927	1.107	4.610	75.927	3.051	12.714	75.927

来源于：本论文收集 SPSS19 检验数据后整理

从表 7.13 可知，共提取 5 个公因子，总方差累积为 75.927%，这与之前设定的知识创造维度不相符。再看其因子载荷的成份矩阵：

表 7.14 知识创造量表各维度旋转因子载荷表

因子	题项编号	因子载荷				
		1	2	3	4	5
知识社会化	Z11					0.586
	Z12					0.648
	Z13					0.832
	Z14	0.540				0.517
知识表出化	Z21				0.683	
	Z22		0.606			
	Z23	0.553				
	Z24			0.528		
	Z25	0.560		0.698		
	Z26	0.573		0.530		
	Z27			0.672		
知识联结化	Z31			0.878		
	Z32			0.650		
	Z33	0.580				
	Z34	0.701				
	Z35				0.519	
	Z36	0.647				
	Z37		0.570			

续表 7.14 知识创造量表各维度旋转因子载荷表

因子	题项编号	因子载荷				
		1	2	3	4	5
知识内在化	Z41				0.702	
	Z42	0.706				
	Z43		0.858			
	Z44		0.866			
	Z45	0.600				
	Z46		0.531		0.565	

从上图看，Z14、Z25、Z26、Z46 出现横跨 2 个因子且因子载荷值均大于 0.5，因此删除这 4 个题项，重新检验。其结果如下表：

表 7.15 修正后知识创造量表各维度旋转因子载荷表

因子	题项编号	因子载荷				
		1	2	3	4	5
知识社会化	Z11					0.532
	Z12					0.504
	Z13					0.799
知识表出化	Z21		0.732			
	Z22		0.520			
	Z23		0.626			
	Z24			0.628		
	Z27			0.707		
知识联结化	Z31			0.881		
	Z32			0.630		
	Z33					0.657
	Z34				0.802	
	Z35			0.533		
	Z36				0.635	
	Z37	0.547				
知识内在化	Z41		0.622			
	Z42	0.556			0.631	
	Z43	0.871				
	Z44	0.850				
	Z45	0.523			0.591	

从表 7.15 看，经过修正后的旋转因子载荷表仍然有 Z42、Z45 出现横跨 2 个因子且因子载荷值均大于 0.5，因此再删除这 2 个题项，重新检验。

首先进行 KMO 和 Bartlett 球形度检验，结果如下：

表 7.16 再次修正知识创造量表 KMO 和 Bartlett 球形度检验表

取样足够度的 Kaiser-Meyer-Olkin 度量。		0.763
Bartlett 的球形度检验	近似卡方	681.395
	df	153
	Sig.	0.000

来源于：本论文收集 SPSS19 检验数据后整理

从以上 KMO 和 Bartlett 检验的结果看，知识创造再次修正量表的 KMO 值为 0.763, Bartlett 球形值为 681.395, 统计量显著性水平 Sig 值为 0.000, 因此符合研究要求。再看解释总方差情况，检验如下表：

表 7.17 再次修正知识创造量表解释总方差分析表

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	8.460	46.999	46.999	8.460	46.999	46.999	3.619	20.104	20.104
2	1.993	11.070	58.069	1.993	11.070	58.069	3.502	19.458	39.562
3	1.274	7.080	65.149	1.274	7.080	65.149	3.347	18.593	58.155
4	1.020	5.664	70.813	1.020	5.664	70.813	2.278	12.658	70.813

来源于：本论文收集 SPSS19 检验数据后整理

从表 7.17 可知，再次修正后提取公因子数量为 4 个，符合预划分维度数，其解释总方差积累为 70.813%，符合研究要求。

其次看旋转因子载荷矩阵情况：

表 7.18 修正后知识创造量表各维度旋转因子载荷表

因子	题项编号	因子载荷			
		1	2	3	4
知识社会化	Z11			0.676	
	Z12			0.777	
	Z13			0.766	

续表 7.18 修正后知识创造量表各维度旋转因子载荷表

因子	题项编号	因子载荷			
		1	2	3	4
知识表出化	Z21	0.590			
	Z22	0.651			
	Z23	0.643			
	Z24	0.624			
	Z27	0.666			
知识联结化	Z31	0.792			
	Z32	0.542			
	Z33				0.547
	Z34				0.843
	Z35				0.655
	Z36				0.618
	Z37				0.584
知识内在化	Z41		0.514		
	Z43		0.873		
	Z44		0.865		

来源于：本论文收集 SPSS19 检验数据后整理

因子载荷的成份矩阵中所有题项的负载系数均大于 0.3，符合研究要求。而且每个题项都能归属到预期主成份当中，可见量表结构性良好。最后看最终量表总信度情况，经检验最终量表  $\alpha$  值为 0.927。

从以上检验结果看，经过两次修正的知识创造量表的可靠性相当好，量表的内部一致性也非常好，达到本论文研究要求。

### 7.1.2.3 小样本检验结论

从以上小样本检验情况看，大数据应用和知识创造量表经过 CITC 检验、探索性因子分析的修正后，其量表信度与效度均良好，量表的稳定性和内部一致性非常好，符合研究要求。

## 7.2 大样本数据检验

大样本数据是经小样本数据预测试，并修正量表后才正式发放和收集调查问卷数据。在利用大样本数据对大数据应用与知识创造间的关系进行实证研究前，我们仍然需要对大样本数据进行描述性统计分析、信度和效度分析、因子分析以



及相关分析等。

### 7.2.1 描述性统计分析

#### (1) 受访者及其公司背景描述性统计分析

实际上，在 6.3.3 节已经对大样本数据进行过受访者背景详细的描述。本节主要在 6.3.3 节的基础上分析其对本研究所存在的意义。

从问卷覆盖地域看，受访者所在公司的地域几乎覆盖全中国经济较发达的省（市、自治区），而中国知识型服务业主要集中在这些（市、自治区）。因此，问卷的结果可以体现中国软件企业的基本情况。

从公司应用大数据为公司业务服务的时间看，软件企业应用大数据的时间比较短，大多在 2 年以上。再从 2 年以上所占比例和研发投入占销售收入比例情况看，中国软件企业近年非常重视大数据的应用，大多在 2-3 年前已经应用大数据为公司服务。同时也非常重视 R&D 活动，而 R&D 活动是企业知识创造的主要活动。因此研究大数据应用对知识创造的影响有着极其重要的现实和实践意义。

从公司业务范围和职员的岗位角色看，受访者所在公司几乎覆盖软件企业所有的业务范围，其岗位角色也几乎覆盖软件企业 R&D 团队中所有角色。因此本问卷的结果可以体现中国软件企业 R&D 团队的基本情况。

从受访者受教育程度情况看，受访者的受教育程度均非常高，这也从侧面反应出知识型企业的人员构成。同时，受访者的受教育程度高，其对知识的理解能力也较高，因此受访者对问卷内容的理解程度也较高，对问卷的答案质量有一定的保障作用。

下面为大样本数据受访者公司和个人信息统计描述表：

表 7.19 大样本数据受访者公司和个人信息统计描述表

题项	题项分类	人数	比例
公司所在地	北京	17	6.51%
	上海	30	11.49%
	山东	16	6.13%
	广东	80	30.27%
	广西	11	4.21%
	其他	85	41.39%
公司业务范围	软件研发	259	98.9%
	软件测试	61	23.28%
	软件外包	211	80.53%
	软件技术咨询	262	100%

续表 7.19 大样本数据受访者公司和个人信息统计描述表

题项	题项分类	人数	比例
公司业务范围	软件销售	122	46.56%
	其它	3	1.14%
公司年销售收入	100 万以下	57	21.76%
	100 - 500 万	128	48.85%
	500 - 1000 万	71	27.1%
	1000 万-5000 万	6	2.29%
	5000 万-1 亿	0	0
	1-10 亿	0	0
研发投入占销售收入的比例	1% 以下	0	0
	1% -2%	21	8%
	2% -5%	234	89.3%
	5% -10%	7	2.67%
	10% 以上	0	0
应用大数据为公司服务时间	6 个月以下	19	7.25%
	6 个月-1 年	49	18.7%
	1-2 年	77	29.39%
	2-3 年	40	15.27%
	3 年以上	77	29.39%
性别	男	152	58.02%
	女	110	41.98%
年龄	20 岁以下	0	0
	20-30 岁	135	51.53%
	30-40 岁	115	43.89%
	40-50 岁	11	4.2%
	50 岁以上	1	0.38%
婚姻状况	已婚	69	26.34%
	未婚	193	73.66%
受教育程度	博士	4	1.53%
	硕士	62	23.66%
	本科	186	70.99%
	专科	10	3.82%
	专科以下	0	0
在软件研发团队内的工作	需求分析与功能设计	11	22.9%
	架构设计	16	33.33%
	程序员	32	66.67%
	软件测试员	9	18.75%
	数据库管理	9	18.75%
	团队负责人	7	14.58%
	项目经理	3	6.25%
	其他	0	0

来源于：本论文整理

(4) 样本数据描述性统计分析

在完成受访者及其公司背景描述性统计分析后，还需要通过最大值、最小值、均值、方差、标准差、偏度和峰度等指标，对大样本数据进行描述性统计分析，分析大样本数据的分布形态和特征，检验大样本数据是否符合研究假设的要求。

表 7.20 大样本数据描述性统计分析图

	N	均值	标准差	方差	偏度		峰度	
	统计量	统计量	统计量	统计量	统计量	标准误	统计量	标准误
D11	261.000	3.762	1.098	1.205	0.625	0.151	0.350	0.300
D12	261.000	3.805	1.185	1.404	0.664	0.151	0.622	0.300
D13	261.000	3.739	1.060	1.124	0.497	0.151	0.661	0.300
D14	261.000	3.540	1.188	1.411	0.284	0.151	1.051	0.300
D15	261.000	3.640	1.180	1.393	0.558	0.151	0.606	0.300
D21	261.000	3.632	1.260	1.587	0.732	0.151	0.491	0.300
D22	261.000	3.625	1.288	1.658	0.589	0.151	0.724	0.300
D23	261.000	3.506	1.205	1.451	0.346	0.151	0.866	0.300
D24	261.000	3.655	1.175	1.381	0.693	0.151	0.346	0.300
D31	261.000	3.716	1.188	1.412	0.698	0.151	0.335	0.300
D32	261.000	3.762	1.198	1.436	0.695	0.151	0.255	0.300
D33	261.000	3.839	1.217	1.482	0.810	0.151	0.186	0.300
D34	261.000	3.843	1.171	1.371	0.820	0.151	0.071	0.300
N11	261.000	3.613	1.259	1.584	0.695	0.151	0.530	0.300
N12	261.000	3.625	1.279	1.635	0.580	0.151	0.723	0.300
N13	261.000	3.498	1.198	1.436	0.347	0.151	0.848	0.300
N14	261.000	3.648	1.169	1.368	0.694	0.151	0.327	0.300
N15	261.000	3.774	1.166	1.360	1.075	0.151	0.547	0.300
N16	261.000	3.828	1.192	1.420	1.023	0.151	0.366	0.300
N17	261.000	3.820	1.005	1.010	0.870	0.151	0.608	0.300
Z11	261.000	3.847	1.231	1.515	1.127	0.151	0.402	0.300
Z12	261.000	3.732	1.172	1.374	0.952	0.151	0.268	0.300
Z13	261.000	3.858	1.070	1.145	0.891	0.151	0.456	0.300
Z14	261.000	3.816	1.135	1.289	0.968	0.151	0.449	0.300
Z21	261.000	3.816	1.094	1.197	0.641	0.151	0.427	0.300
Z22	261.000	3.816	1.204	1.451	0.693	0.151	0.620	0.300
Z23	261.000	3.755	1.053	1.109	0.511	0.151	0.615	0.300
Z24	261.000	3.544	1.197	1.434	0.281	0.151	1.082	0.300

续表 7.20 大样本数据描述性统计分析图

	N	均值	标准差	方差	偏度		峰度	
	统计量	统计量	统计量	统计量	统计量	标准误	统计量	标准误
Z25	261.000	3.590	1.152	1.327	0.549	0.151	0.537	0.300
Z26	261.000	3.728	1.153	1.330	0.606	0.151	0.529	0.300
Z27	261.000	3.732	1.159	1.343	0.568	0.151	0.637	0.300
Z31	261.000	3.762	1.198	1.436	0.695	0.151	0.255	0.300
Z32	261.000	3.839	1.217	1.482	0.810	0.151	0.186	0.300
Z33	261.000	3.843	1.171	1.371	0.820	0.151	0.071	0.300
Z34	261.000	3.778	1.198	1.435	0.849	0.151	0.007	0.300
Z35	261.000	3.705	1.184	1.401	0.687	0.151	0.332	0.300
Z36	261.000	3.762	1.098	1.205	0.625	0.151	0.350	0.300
Z37	261.000	3.805	1.185	1.404	0.664	0.151	0.622	0.300
Z41	261.000	3.739	1.060	1.124	0.497	0.151	0.661	0.300
Z42	261.000	3.540	1.188	1.411	0.284	0.151	1.051	0.300
Z43	261.000	3.640	1.180	1.393	0.558	0.151	0.606	0.300
Z44	261.000	3.632	1.260	1.587	0.732	0.151	0.491	0.300
Z45	261.000	3.625	1.288	1.658	0.589	0.151	0.724	0.300

来源于：本论文收集 SPSS19 检验数据后整理

从表 7.20 大样本数据描述性统计分析图看，各题项均值都在 3.498-3.858 之间，可见其分布比较均衡。各题项标准差均在 1.005-1.288 之间，说明样本数据离散度小。Klein (1998)认为，当样本数据偏度绝对值 $<3$ ，峰度绝对值 $<8$ 的时候，则可认为观测变量基本符合正态分布。而从以上统计结果看，所有题项偏度绝对值 $<1.075$ ，峰度绝对值 $<1.082$ ，远小于 Klein (1998)提出的参考值。因此可以被认为大样本数据的形态基本符合正态分布，满足本文研究假设对分析数据的基本要求。

## 7.2.2 信度分析

### 7.2.2.1 大数据应用信度检验

以下为 SPSS 19 检验出大样本数据中大数据应用各题项 CITC 与  $\alpha$  值的结果：

表 7.21 大数据应用量表信度分析表

维度	题项编号	CITC	各题项 $\alpha$ 系数	量表总 $\alpha$ 系数
技术外包性	D11	0.617	0.869	0.880
	D12	0.572	0.871	
	D13	0.606	0.869	
	D14	0.614	0.868	
	D15	0.609	0.869	
技术智能性	D21	0.554	0.872	
	D22	0.570	0.871	
	D23	0.548	0.872	
	D24	0.650	0.867	
强可视化性	D31	0.467	0.876	
	D32	0.497	0.875	
	D33	0.508	0.874	
	D34	0.498	0.875	

来源于：本论文收集 SPSS19 检验数据后整理

从以上结果看，所有题项的  $\alpha$  值大于 0.8，且所有题项 CITC 均大于 0.4，因此符合研究要求。再检验大数据应用总体量表的  $\alpha$  值，结果如下：

表 7.22 大数据应用总体量表  $\alpha$  值检测结果图

Cronbach's Alpha	项数
0.880	13

从图 7.21 可知，大数据应用总体量表的  $\alpha$  值=0.827，表明量表总体信度非常好，可靠性非常强。

### 7.2.2.2 知识创造信度检验

以下为 SPSS 19 检验出大样本数据中知识创造各题项 CITC 与  $\alpha$  值的结果：

表 7.23 知识创造量表信度分析表

维度	题项编号	CITC	各题项 $\alpha$ 系数	量表总 $\alpha$ 系数
知识表出化	Z11	0.534	0.925	0.927
	Z12	0.558	0.925	
	Z13	0.499	0.926	
	Z14	0.609	0.924	

续表 7.23 知识创造量表信度分析表

维度	题项编号	CITC	各题项 $\alpha$ 系数	量表总 $\alpha$ 系数
知识社会化	Z21	0.689	0.922	0.927
	Z22	0.555	0.925	
	Z23	0.640	0.923	
	Z24	0.618	0.924	
	Z25	0.573	0.924	
	Z26	0.535	0.925	
	Z27	0.552	0.925	
知识联结化	Z31	0.647	0.923	
	Z32	0.608	0.924	
	Z33	0.654	0.923	
	Z34	0.632	0.923	
	Z35	0.606	0.924	
	Z36	0.650	0.923	
	Z37	0.691	0.922	
知识内在化	Z41	0.508	0.925	
	Z42	0.461	0.926	
	Z43	0.470	0.926	
	Z44	0.517	0.925	
	Z45	0.465	0.926	

来源于：本论文收集 SPSS19 检验数据后整理

从表图 7.23 看，知识创造量表中各项的  $\alpha$  值均大于 0.9，可信度极好。而 CITC 的值均大于 0.4，也均在可接受范围。再从表 7.24 看，知识创造量表总体  $\alpha$  值=0.927，可信度同样很好。

表 7.24 大数据应用总体量表  $\alpha$  值检测结果图

Cronbach's Alpha	项数
0.927	23

可见，知识创造量表总体信度非常好，可靠性非常强。

### 7.2.3 因子分析

因子分析是检验量表效度的重要手段。因子分析主要有探索性因子分析和验证性因子分析。本节将采用这两种因子分析方法对大数据应用和知识创造三个量

表的大样本数据进行效度检验，从而为后面对研究假设的验证提供基础保障。需要说明的是，本论文验证性因子分析采用 AMOS21 软件进行检验。

### 7.2.3.1 探索性因子分析

本论文主要采用 KMO 和 Bartlett 球形度检验、主成份分析和最大方差对大数据应用和知识创造进行探索性因子分析。

#### (1) 数据应用探索性因子分析

大数据应用的 KMO 和 Bartlett 球形度检验结果如下图：

表 7.25 大数据应用量表 KMO 和 Bartlett 球形度检验表

取样足够度的 Kaiser-Meyer-Olkin 度量。		0.881
Bartlett 的球形度检验	近似卡方	1809.479
	df	78
	Sig.	0.000

来源于：本论文收集 SPSS19 检验数据后整理

从表 7.25 的 KMO 和 Bartlett 检验的结果看，大数据应用的 KMO 值为 0.881，Bartlett 球形值为 1809.479，统计量显著性水平 Sig 值为 0.000，因此本研究大数据应用量表适合做因子分析。

再看为解释总方差的检验结果：

表 7.26 大数据应用量表解释总方差分析表

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	5.393	41.484	41.484	5.393	41.484	41.484	3.531	27.162	27.162
2	2.060	15.846	57.331	2.060	15.846	57.331	3.042	23.399	50.561
3	1.709	13.149	70.480	1.709	13.149	70.480	2.589	19.919	70.480

来源于：本论文收集 SPSS19 检验数据后整理

从上表解释总方差的检验结果看，三个公因子解释的总方差比例为 70.480%，在可接受范围。其中，解释比例最高公因子 1 为 27.162%，公因子 2 解释比例为 23.399%，公因子 3 为 19.919%。

再分析大数据应用各维度的旋转因子载荷表：

表 7.27 大数据应用各维度旋转因子载荷表

因子	题项编号	因子载荷		
		1	2	3
技术外包性	D11	0.853		
	D15	0.845		
	D14	0.798		
	D12	0.797		
	D13	0.788		
技术智能性	D21		0.862	
	D23		0.860	
	D24		0.841	
	D22		0.822	
强可视化性	D32			0.777
	D33			0.772
	D31			0.766
	D34			0.758

来源于：本论文收集 SPSS19 检验数据后整理

根据表 7.27 的数据显示，大数据应用量表所有题项的因子载荷均大于 0.7，而且每个题项均归属到预期的因子中，因此符合研究要求。

#### (2) 知识创造量表探索性因子分析

从图表 7.28 看，KMO 和 Bartlett 检验的结果为：知识创造的 KMO 值为 0.924，Bartlett 球形值为 3861.772，统计量显著性水平 Sig 值为 0.000，因此本研究知识创造量表适合做因子分析。

表 7.28 知识创造量表 KMO 和 Bartlett 球形度检验表

取样足够度的 Kaiser-Meyer-Olkin 度量。		0.924
Bartlett 的球形度检验	近似卡方	3861.772
	df	253
	Sig.	0.000

来源于：本论文收集 SPSS19 检验数据后整理

再看知识创造量表解释总方差的检验结果：



表 7.29 知识创造量表解释总方差分析表

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	9.040	39.304	39.304	9.040	39.304	39.304	4.890	21.259	21.259
2	2.485	10.803	50.107	2.485	10.803	50.107	4.606	20.024	41.284
3	2.338	10.167	60.274	2.338	10.167	60.274	3.317	14.420	55.704
4	2.044	8.886	69.160	2.044	8.886	69.160	3.095	13.456	69.160

来源于：本论文收集 SPSS19 检验数据后整理

从表 7.29 可知，共提取 4 个公因子，总方差累积为 69.16%，这与之前设定的知识创造维度相符。再看其因子载荷的成份矩阵检验结果：

表 7.30 知识创造量表各维度旋转因子载荷表

因子	题项编号	因子载荷			
		1	2	3	4
知识社会化	Z13				0.859
	Z11				0.840
	Z14				0.815
	Z12				0.808
知识表出化	Z21		0.842		
	Z25		0.784		
	Z22		0.757		
	Z23		0.756		
	Z24		0.747		
	Z27		0.743		
	Z26		0.725		
知识联结化	Z31	0.835			
	Z36	0.805			
	Z32	0.793			
	Z37	0.793			
	Z35	0.791			
	Z33	0.766			
	Z34	0.762			

续表 7.30 知识创造量表各维度旋转因子载荷表

因子	题项编号	因子载荷			
		1	2	3	4
知识内在化	Z44			0.849	
	Z42			0.766	
	Z45			0.746	
	Z41			0.738	
	Z43			0.737	

来源于：本论文收集 SPSS19 检验数据后整理

表 7.30 数据显示，因子载荷的成份矩阵中所有题项负载系数均大于 0.7，符合研究要求。而且每个题项都能归属到预期主成份当中，可见量表结构性良好。最后看最终量表总信度情况，经检验最终量表  $\alpha$  值为 0.924。因此从检验结果看，知识创造量表的可靠性相当好，量表的内部一致性也非常好，达到本论文研究要求。

### 7.2.3.2 验证性因子分析

本论文采用 AMOS23 软件对大数据应用和知识创造进行验证性因子分析，目的是通过检验大样本数据对以上两个量表的适用程度，进而检验以上两个量表的效度。

#### (1) 大数据应用验证性因子分析

大数据应用部分共有 3 个维度，共有 13 个题项，利用 AMOS23 进行验证性因素分析。先建立隐性变量和显性的关系模型，后进行检验。

第一步：建立关系模型

首先说明大数据应用三个维度对应标识及其测量题项对应标识如下：

技术外包性标识为 D1，其对应 5 个测量题项标识分别为 D11、D12、D13、D14、D15；

技术智能性标识为 D2，其 4 个测量题项标识分别为 D21、D22、D23、D24；

强可视化性标识为 D3，其 4 个测量题项标识分别为 D31、D32、D33、D34；

下图为大数据应用具体关系模型：

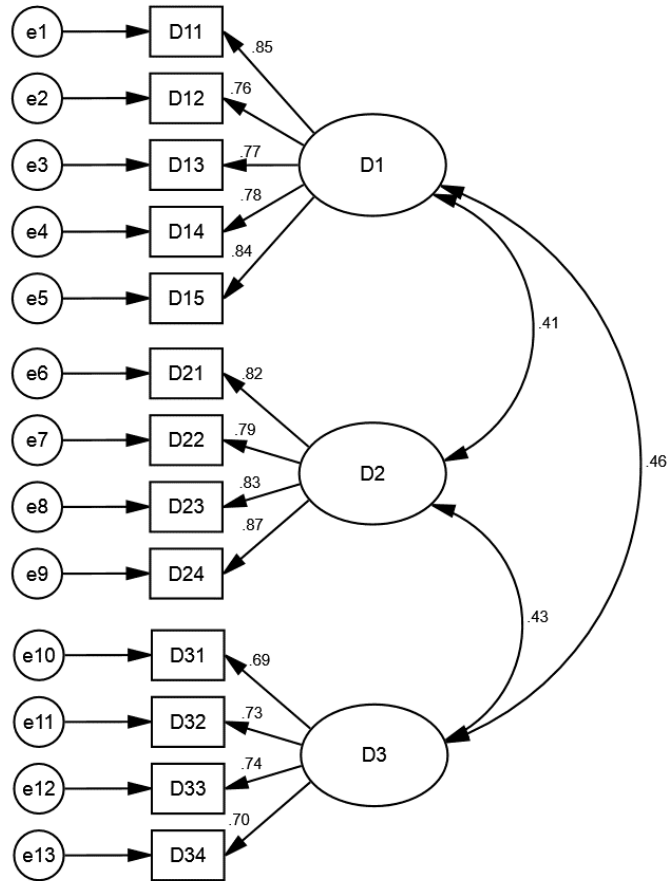


图 7.1 大数据应用关系模型图

来源于：本论文收集 AMOS23 检验数据后整理

第二步：用 AMOS23 进行检验，检验结果如下表所示：

表 7.31 验证性因素模型拟合度表

拟合指标	CMIN/DF	SRMR	RMSEA	GFI	AGFI	NFI	IFI	TLI	CFI
判断标准	<3	<0.08	<0.08	>0.9	>0.9	>0.9	>0.9	>0.9	>0.9
模型结果	1.202	0.029	0.028	0.960	0.941	0.960	0.993	0.991	0.993
结论	合格	合格	合格	合格	可接受	合格	合格	合格	合格

来源于：本论文收集 AMOS23 检验数据后整理

从上表可知 CMIN/DF 为 1.202, 小于 3 以下标准, 表明模型拟合度较好。GFI、AGFI、NFI、TLI、IFI、CFI 均达到 0.9 以上的标准, 表明模型拟合度非常好。SRMR 为 0.029, 小于 0.08, RMSEA 为 0.028, 小于 0.08, 表明模型拟合度也非常好。由此可见, 各个拟合指标均符合一般的研究标准, 可以认为这个模型具有良好的配适度。再看验证性因素分析结果：

表 7.32 验证性因素分析结果表

潜在变量	题项	非标准化负荷	S.E.	C.R.	P	标准化负荷	CR	AVE
技术外包性	D11	1				0.85	0.899	0.641
	D12	0.959	0.069	13.984	***	0.755		
	D13	0.877	0.061	14.435	***	0.772		
	D14	0.995	0.068	14.699	***	0.782		
	D15	1.06	0.065	16.278	***	0.838		
技术智能性	D21	1				0.822	0.897	0.686
	D22	0.98	0.069	14.226	***	0.788		
	D23	0.96	0.063	15.139	***	0.825		
	D24	0.992	0.061	16.27	***	0.874		
强可视化性	D31	1				0.686	0.806	0.51
	D32	1.068	0.111	9.618	***	0.727		
	D33	1.1	0.113	9.705	***	0.737		
	D34	1.01	0.108	9.391	***	0.704		

来源于：本论文收集 AMOS23 检验数据后整理

根据表 7.32 显示，所有题项的标准化因素负荷均大于 0.6，均符合大于 0.5 而小于 0.95 的标准。组成信度(CR)分别均大于 0.7，平均变异萃取量(AVE)均大于 0.5，表明各个因素均具有良好的收敛效率。

## (2) 知识创造验证性因子分析

### 第一步：建立关系模型

首先说明知识创造四个维度对应标识及其测量题项对应标识如下：

知识社会化标识为 Z1，其 4 个测量题项标识分别为 Z11、Z12、Z13、Z14；

知识表出化标识为 Z2，其 7 个测量题项标识分别为 Z21、Z22、Z23、Z24、Z25、Z26、Z27；

知识联结化标识为 Z3，其 7 个测量题项标识分别为 Z31、Z32、Z33、Z34、Z35、Z36、Z37；

知识联结化标识为 Z4，其 5 个测量题项标识分别为 Z41、Z42、Z43、Z44、Z45；

下图为知识创造具体关系模型：

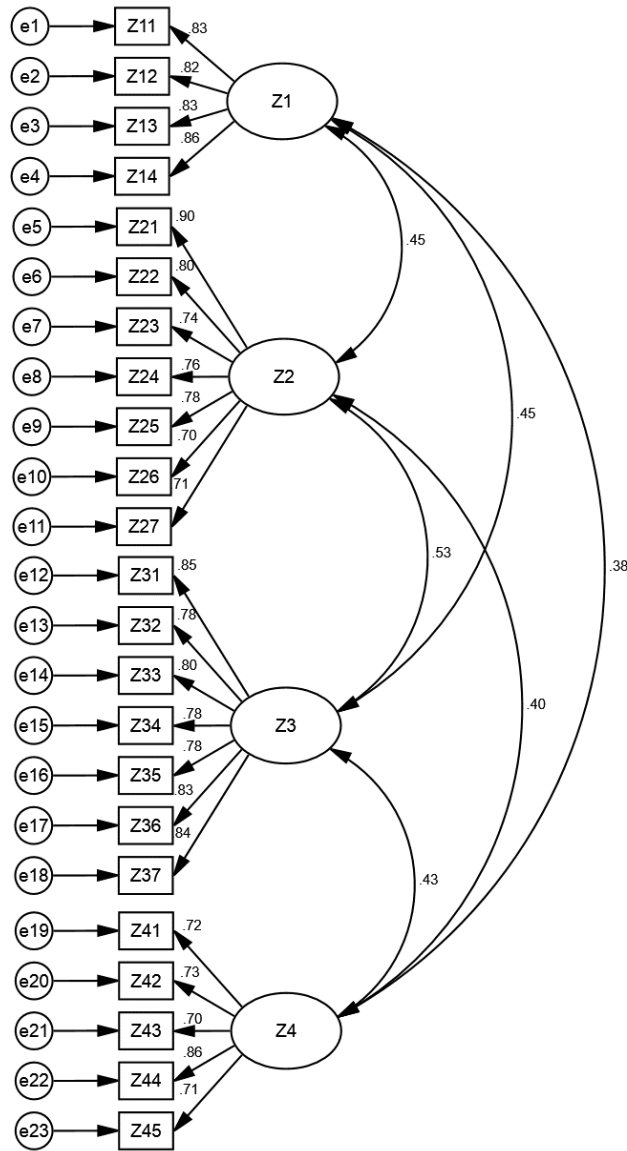


图 7.2 知识创造关系模型图

来源于：本论文收集 AMOS23 检验数据后整理

第二步：用 AMOS23 进行检验，检验结果如下表所示：

表 7.33 验证性因素模型拟合度

拟合指标	CMIN/DF	SRMR	RMSEA	GFI	AGFI	NFI	IFI	TLI	CFI
判断标准	<3	<0.08	<0.08	>0.9	>0.9	>0.9	>0.9	>0.9	>0.9
模型结果	1.232	0.037	0.030	0.920	0.902	0.931	0.986	0.984	0.986
结论	合格	合格	合格	合格	可接受	合格	合格	合格	合格

从上表可知 CMIN/DF 为 1.232，小于 3 以下标准，表明模型拟合度较好。GFI、AGFI、NFI、TLI、IFI、CFI 均达到 0.9 以上的标准，表明模型拟合度非常好。SRMR 为 0.037，小于 0.08，RMSEA 为 0.030，小于 0.08，表明模型拟合度也非常好。

由此可见，各个拟合指标均符合一般的研究标准，可以认为这个模型具有良好的配适度。再看验证性因素分析结果：

表 7.34 因子分析结果

潜在变量	题项	非标准化负荷	S.E.	C.R.	P	标准化负荷	CR	AVE
R&D 团队知识表出化	Z11	1				0.832	0.903	0.699
	Z12	1	0.065	15.315	***	0.818		
	Z13	0.949	0.061	15.595	***	0.829		
	Z14	0.966	0.058	16.519	***	0.864		
R&D 团队知识社会化	Z21	1				0.904	0.911	0.597
	Z22	0.76	0.045	16.927	***	0.797		
	Z23	0.835	0.056	14.805	***	0.738		
	Z24	0.89	0.057	15.625	***	0.762		
	Z25	0.865	0.053	16.173	***	0.778		
	Z26	0.71	0.052	13.57	***	0.699		
R&D 团队知识联结化	Z31	1				0.851	0.93	0.656
	Z32	1.014	0.066	15.345	***	0.784		
	Z33	0.9	0.057	15.726	***	0.796		
	Z34	0.998	0.066	15.11	***	0.776		
	Z35	0.97	0.063	15.357	***	0.785		
	Z36	1.024	0.061	16.734	***	0.827		
	Z37	1.051	0.061	17.327	***	0.845		
R&D 团队知识内在化	Z41	1				0.724	0.861	0.555
	Z42	1.031	0.093	11.08	***	0.735		
	Z43	0.94	0.089	10.508	***	0.696		
	Z44	1.182	0.093	12.654	***	0.855		
	Z45	0.964	0.09	10.656	***	0.706		

来源于：本论文收集 AMOS23 检验数据后整理

根据表 7.34 显示，所有题项的标准化因素负荷均大于 0.6，均符合大于 0.5 而小于 0.95 的标准。组成信度(CR)分别均大于 0.7，平均变异萃取量(AVE)均大于 0.5，表明各个因素均具有良好的收敛效率。

### 7.2.4 效度分析

杜智敏(2010)认为，效度主要分成内容效度和结构效度。为此，本论文在之前

$\alpha$  检验和因子分析检验的基础上，对大数据应用和知识创造量表进行内容效度和结构效度的分析。

#### (1) 内容效度分析

本论文在调查问卷中使用的测量题项均通过直接或间接使用中国及其他国家成熟的研究成果，并在听取专家组意见及走访企业后整理出来的。同时，大样本的调查量表是在小样本预测试后得到的，因此无论是测量问题的成熟度和因子结构的稳定度都相当高。可见，量表和调查问卷均能够体现本论文研究的内容及其概念，可以测量出大数据应用和知识创造研究过程所需要的数据，具有较好的内容效度。

#### (2) 结构效度分析

前面的  $\alpha$  检验和因子分析检验中，量表的  $\alpha$  值均大于 0.8，大数据应用的解释总方差率达 70.48%，知识创造解释总方差率达 69.16%，因子载荷系数均大于 0.5，且各题项均归属到预期的因子当中。再从验证性因子分析看，因子结构模型拟合指标均符合标准，表明了本论文模型拟合非常好。综合以上所述，两个量表各个维度和题项的因子结构清晰，因子分析结果和本论文的研究构想高度一致性，各项指标均符合要求。因此，本论文所使用的量表和调查问卷均有较好的结构效度。

### 7.2.5 ICC 与 r<sub>wg</sub> 分析检验

由于本论文是通过个人数据研究团队问题，因此需要验证个人层面数据整合到团体层面结构的有效性。验证方法主要有两种：一是 ICC 指标分析法。即利用单因素方差检验后得到 MSB 和 MSW 后，再计算 ICC (1) 和 ICC (2)，并由这两个指标来判断个人层面数据是否适合整合到团队层面数据。根据 James (1982) 给出的标准，ICC (1) >0.05，ICC (2) >0.5。二是 r<sub>wg</sub> 系数分析法。r<sub>wg</sub> 系数是指团队内部的一致性系数，即在通过多个题项测量一个变量时，团队内部成员回答是否一致。George (1990) 认为，r<sub>wg</sub> 系数 >0.7，则可视作有足够的一致性。

#### (1) ICC 分析检验

由于本量表数据要采用 SPSS19 来检验 ICC，需要重新组织数据数据。因此本论文采用先利用 SPSS19 做单因素方差检验，并且获得 MSB 和 MSW 结果后，再进行 ICC (1) 和 ICC (2) 值的计算。其计算方法如下：

$$ICC(1) = \frac{MSB - MSW}{MSB + (K - 1)MSW} \quad (\text{注：K 为团队内答题的成员人数})$$

$$CCI(2) = \frac{MSB - MSW}{MSB}$$

以下为单因素方差检验和 CCI (1) 和 CCI (2) 计算的结果：

表 7.35 各变量内部一致性 ICC(1)和 ICC(2) 检验表

	MSB	MSW	F 检验	显著性	ICC (1)	ICC (2)
技术外包性	1.032	0.388	2.535	***	0.089	0.624
技术智能性	0.996	0.362	2.434	***	0.093	0.637
强可视化性	0.939	0.336	2.565	***	0.095	0.642
R&D 团队知识表出化	0.943	0.387	2.827	***	0.078	0.590
R&D 团队知识社会化	1.065	0.397	2.536	***	0.090	0.627
R&D 团队知识联结化	0.983	0.323	2.534	***	0.107	0.671
R&D 团队知识内在化	0.998	0.343	2.533	***	0.101	0.656

来源于：本论文收集 SPSS19 检验数据后整理

从表 7.35 的数据看，ICC (1) 均大于 0.05，而 ICC (2) 也均大于 0.5，这表明各变量得到的个人层面数据适合整合到团队层面做进一步研究。

### (2) $r_{wg}$ 系数分析法

要通过 SPSS19 计算出样本中每个团队的  $r_{wg}$  值比较困难。因此我们使用 excel 辅助计算，用计算方法计算出每个团队的  $r_{wg}$  值。 $r_{wg}$  值的计算方法如下：

第一步:计算随机方差，随机方差计算公式如下：

$$\delta^2 = \frac{a^2 - 1}{12} \quad (\text{注: } \delta^2 \text{ 为随机方差, } a \text{ 为评分量表的点数})$$

由于本研究采用李特克 5 点量表法，因此  $a=5$ ，而  $\delta^2 = \frac{5 \times 5 - 1}{12} = 2$ 。

第二步:计算各团队对各题项评分值的观测方差。这一步可以用 SPSS19 中的描述性统计来辅助计算。

第三步:通过  $r_{wg}$  计算公式每个团队的  $r_{wg}$  值。计算公式如下：

$$r_{wg}(j) = \frac{J[1 - \overline{\delta_{ij}^2} / \delta^2]}{J[1 - (\overline{\delta_{ij}^2} / \delta^2)] + [\overline{\delta_{ij}^2} / \delta^2]}$$

注：J 为题目数量， $\delta^2$  为随机方差， $\overline{\delta_{ij}^2}$  为观测方差的平均值。

最后，我们求出各团队  $r_{wg}(j)$  的平均值，即  $\overline{r_{wg}(j)}$ ，用以代表每个变量的  $r_{wg}$  值，其计算结果如下表：



表 7.36  $r_{wg}$  均值和中位数表

	技术外包性	技术智能性	强可视化性	R&D 团队知识表出化	R&D 团队知识社会化	R&D 团队知识联结化	R&D 团队知识内在化
$r_{wc}$ 均值	0.837	0.866	0.889	0.924	0.891	0.887	0.936
中位数	0.925	0.928	0.919	0.917	0.936	0.906	0.911

来源于：本论文收集 SPSS19 检验数据后整理

从表 7.36 数据看， $r_{wg}$  均值均大于标准的 0.7，因此可以认为样本数据具有较好的一致性，可以整合到团队层面数据作下一步的分析检验。但在团队  $r_{wg}(j)$  值的计算结果中发现，少数团队的  $r_{wg}$  值达不到 0.7 的标准。根据刘军(2008)的提法，如果 CCI 指标满足标准，对于少数  $r_{wg}$  值达不到标准的，也可以认为数据可以整合汇聚，不必删除这些团队数据。因此，综合以上 CCI 指标的检验情况看，我们可以认为样本数据有较好的团队一致性，可将个人层面数据整合到团队层面。

## 7.2.6 相关分析

相关分析主要检验研究中各变量维度之间的相关性。吴明隆(2003)认为，当自变量相互间的相关系数大于 0.75，则有存在多重共线性问题的可能。因此其建议，在进行回归分析前，需要对自变量进行相关分析，这样可以回避多重共线性的问题的出现。因此，本论文研究将对大数据应用和知识创造两个变量的各个维度进行相关性检验。

相关性分析主要通过 Pearson 相关系数对各变量各维度间的相关性进行检验。通常标准是，Pearson 相关系数在[-1, 1]中取值，当 Pearson 相关系数>0 时，可认为存在正相关关系。当 Pearson 相关系数<0 时，则被认为存在负相关关系。其相关显著程度则以 Pearson 相关系数绝对值衡量。| Pearson 相关系数 | < 0.3 为弱相关，0.3 < | Pearson 相关系数 | < 0.5 为低度相关，0.5 < | Pearson 相关系数 | < 0.8 为显著相关，0.8 < | Pearson 相关系数 | 为高度相关。

首先，我们通过 Pearson 相关系数对大数据应用和 R&D 团队知识创造进行相关性检验，其检验结果如下：

表 7.37 大数据应用和 R&D 团队知识创造相关检验结果表

	大数据应用	R&D 团队知识创造
大数据应用	1.000	
R&D 团队知识创造	0.812*	1.000

来源于：本论文收集 SPSS19 检验数据后整理

从表 7.37 中可以看出，大数据应用与 R&D 知识创造之间的 Pearson 相关系数值为 0.812，表明两者存在着高度正相关关系。接下来看大数据应用与 R&D 知识创造各维度间的相关检验，其结果如下：

表 7.38 各变量维度间偏相关检验结果表

	技术外包性	技术智能性	强可视化性	R&D 团队知识社会化	R&D 团队知识表出化	R&D 团队知识联结化	R&D 团队知识内在化
技术外包性	1.000						
技术智能性	0.372**	1.000					
强可视化性	0.391**	0.359**	1.000				
R&D 团队知识社会化	0.412**	0.431**	0.367**	1.000			
R&D 团队知识表出化	0.490**	0.572**	0.460**	0.412**	1.000		
R&D 团队知识联结化	0.609**	0.540**	0.537**	0.409**	0.488**	1.000	
R&D 团队知识内在化	0.441**	0.408**	0.413**	0.348**	0.352**	0.398**	1.000

注：\*， $p < 0.05$ ；\*\*， $p < 0.01$

来源于：本论文收集 SPSS19 检验数据后整理

从表 7.38 检验的 r 值和 p 值结果看，各变量之间均具有显著正相关。

### 7.2.7 假设检验

相关性分析仅仅是检验两个变量间的关系紧密程度，即大数据应用和知识创造之间是否存在关联关系，但这种关系并不具备因果关系，因此就无法描述变量间的函数关系。对于确定变量间的因果关系，需要用回归分析法检验。

本论文采用的是主成分回归方法。即在做回归检验之前，先计算出每个因子的得分 ( $F_x$ ) 及其综合得分 (F)。再用  $F_x$  和 F 来进行回归分析。其计算公式如下：

(1) 因子得分计算公式

$$F_i = c_1 x_1 + c_2 x_2 + \dots + c_i x_i$$

注： $c_i$  为每个变量的主成份得分系数， $x_i$  为变量， $i=1,2,3,\dots,n$ 。

(2) 综合得分计算公式

$$Z_i = w_1 F_1 + w_2 F_2 + \dots + w_i F_i$$

注： $w_i$  为相应因子的方差贡献率， $F_i$  为因子得分， $i=1,2,3,\dots,n$ 。

根据以上公式可求出大数据应用综合得分 D，R&D 知识创造综合得分 Z，大

数据应用三个维度的因子得分分别为 D1（技术外包性）、D2（技术智能性）、D3（强可视化性），R&D 知识创造四个维度因子得分分别为 Z1（R&D 团队知识社会化）、Z2（R&D 团队知识表出化）、Z3（R&D 团队知识联结化）、Z4（R&D 团队知识内在化）。

### 7.2.7.1 大数据应用与 R&D 团队知识创造的假设检验

利用大数据应用和 R&D 团队知识创造因子综合得分进行一元线性回归分析，模型为：

$$Z = \varepsilon + bD$$

注： $\varepsilon$  为常数， $b$  为回归系数。其检验结果如下表：

表 7.39 大数据应用与 R&D 团队知识创造一元线性回归分析表

模型		非标准化回归系数	标准误差	标准化回归系数	t	R <sup>2</sup>	F	P
1	常量	2.037	0.211		9.636	0.186	59.167	0.000
	大数据应用	0.432	0.056	0.431	7.692			

来源于：本论文收集 SPSS19 检验数据后整理

从表 7.39 的分析结果看：调整 R<sup>2</sup> 系数=0.186，表明大数据应用解释了 R&D 团队 18.6% 的变化。显著性水平 P<0.01，表明拟合度比较好。F 值为 59.167，标准回归系数值为 0.431，且显著性水平 P<0.05，回归效果显著。因此可以认为大数据应用对 R&D 团队知识创造有显著的正影响，H 假设得到验证。

### 7.2.7.2 大数据应用与 R&D 团队知识社会化的假设检验

利用大数据应用三个维度，即技术外包性、技术智能性、强可视化性的因子得分和 R&D 团队知识社会化因子得分进行多元线性回归分析，其模型为：

$$Z1 = \varepsilon + b1D1 + b2D2 + b3D3$$

注： $\varepsilon$  为常数， $b1, b2, b3$  为回归系数。其检验结果如下表：

表 7.40 大数据应用与 R&D 团队知识社会化多元线性回归分析表

模型		非标准化回归系数	标准误差	标准化回归系数	t	P	VIF	R <sup>2</sup>	DW
1	常量	0.844	0.286		2.954	0.003		0.283	2.021
	技术外包性	0.270	0.067	0.241	4.043	0.000	1.273		
	技术智能性	0.281	0.059	0.280	4.763	0.000	1.238		
	强可视化性	0.196	0.067	0.172	2.909	0.004	1.259		

来源于：本论文收集 SPSS19 检验数据后整理

从上表可知 Durbin-Watson 为 2.021，在 1.5-2.5 之间，表明样本之间具有独立性，VIF 均小于 5，表明技术外包性、技术智能性、强可视化性之间不具有多重共线性，R<sup>2</sup> 为 0.283，知识社会化可被解释变异为 28.3%，技术外包性对知识社会化 ( $\beta = 0.241$ ,  $P < 0.05$ ) 具有显著正向影响，假设成立；技术智能性对知识社会化 ( $\beta = 0.280$ ,  $P < 0.05$ ) 具有显著正向影响，假设成立；强可视化性对知识社会化 ( $\beta = 0.172$ ,  $P < 0.05$ ) 具有显著正向影响。可见回归检验结果显著，假设 H1a、H1b、H1c 均成立。

### 7.2.7.3 大数据应用与 R&D 团队知识表出化的假设检验

利用大数据应用三个维度，即技术外包性、技术智能性、强可视化性的因子得分和 R&D 团队知识表出化因子得分进行多元线性回归分析，其模型为：

$$Z2 = \varepsilon + b1D1 + b2D2 + b3D3$$

注： $\varepsilon$  为常数， $b1, b2, b3$  为回归系数。其检验结果如下表：

表 7.41 大数据应用与 R&D 团队知识表出化多元线性回归分析表

模型		非标准化回归系数	标准误差	标准化回归系数	t	P	VIF	R <sup>2</sup>	DW
1	常量	0.881	0.213		4.140	0.000		0.453	1.903
	技术外包性	0.246	0.050	0.398	4.949	0.000	1.238		
	技术智能性	0.340	0.044	0.216	7.760	0.000	1.259		
	强可视化性	0.209	0.050	0.258	4.176	0.000	1.273		

来源于：本论文收集 SPSS19 检验数据后整理

由上表可以得到 Durbin-Watson 为 1.903，在 1.5-2.5 之间，表明样本之间具有独立性，VIF 均小于 5，表明技术外包性、技术智能性、强可视化性之间不具有多

重共线性， $R^2$  为 0.453，知识表出化可被解释变异为 45.3%，技术外包性对知识表出化( $\beta = 0.398, P < 0.05$ )具有显著正向影响，假设成立；技术智能性对知识表出化( $\beta = 0.216, P < 0.05$ )具有显著正向影响，假设成立；强可视化性对知识表出化( $\beta = 0.258, P < 0.05$ )具有显著正向影响。可见回归检验结果显著，假设 H2a、H2b、H2c 均成立。

#### 7.2.7.4 大数据应用与 R&D 团队知识联结化的假设检验

利用大数据应用三个维度，即技术外包性、技术智能性、强可视化性的因子得分和 R&D 团队知识联结化因子得分进行多元线性回归分析，其模型为：

$$Z3 = \varepsilon + b1D1 + b2D2 + b3D3$$

注： $\varepsilon$  为常数， $b1, b2, b3$  为回归系数。其检验结果如下表：

表 7.42 大数据应用与 R&D 团队知识联结化多元线性回归分析表

模型	非标准化回归系数	标准误差	标准化回归系数	t	P	VIF	$R^2$	DW
1 常量	0.255	0.202		1.259	0.209		0.547	2.075
技术外包性	0.390	0.047	0.391	8.245	0.000	1.273		
技术智能性	0.264	0.042	0.295	6.324	0.000	1.238		
强可视化性	0.281	0.048	0.278	5.897	0.000	1.259		

来源于：本论文收集 SPSS19 检验数据后整理

由上表可得 Durbin-Watson 为 2.075，在 1.5-2.5 之间，表明样本之间具有独立性，VIF 均小于 5，表明技术外包性、技术智能性、强可视化性之间不具有多重共线性， $R^2$  为 0.547，知识联结化可被解释变异为 54.7%，技术外包性对知识联结化( $\beta = 0.391, P < 0.05$ )具有显著正向影响，假设成立；技术智能性对知识联结化( $\beta = 0.295, P < 0.05$ )具有显著正向影响，假设成立；强可视化性对知识联结化( $\beta = 0.278, P < 0.05$ )具有显著正向影响。可见本检验回归结果显著，假设 H3a、H3b、H3c 成立。

#### 7.2.7.5 大数据应用与 R&D 团队知识内在化的假设检验

利用大数据应用三个维度，即技术外包性、技术智能性、强可视化性的因子得分和 R&D 团队知识内在化因子得分进行多元线性回归分析，其模型为：

$$Z4 = \varepsilon + b1D1 + b2D2 + b3D3$$

注： $\varepsilon$  为常数， $b1, b2, b3$  为回归系数。其检验结果如下表：

表 7.43 大数据应用与 R&D 团队知识内在化多元线性回归分析表

模型		非标准化回归系数	标准误差	标准化回归系数	t	P	VIF	R <sup>2</sup>	DW
1	常量	1.213	0.249		4.869	0.000	1.273	0.304	1.938
	技术外包性	0.266	0.058	0.268	4.565	0.000	1.238		
	技术智能性	0.201	0.051	0.226	3.912	0.000	1.259		
	强可视化性	0.229	0.059	0.227	3.893	0.000	1.273		

来源于：本论文收集 SPSS19 检验数据后整理

由上表可以得到 Durbin-Watson 为 1.938，在 1.5-2.5 之间，表明样本之间具有独立性，VIF 均小于 5，表明技术外包性、技术智能性、强可视化性之间不具有多重共线性，R<sup>2</sup> 为 0.304，知识内在化可被解释变异为 30.4%，技术外包性对知识内在化( $\beta = 0.268, P < 0.05$ )具有显著正向影响，假设成立；技术智能性对知识内在化( $\beta = 0.226, P < 0.05$ )具有显著正向影响，假设成立；强可视化性对知识内在化( $\beta = 0.227, P < 0.05$ )具有显著正向影响。可见本检验回归结果显著，假设 H4a、H4b、H4c 成立。

#### 7.2.7.6 团队协作能力对大数据应用和 KIBS 企业知识创造的调节检验

由于自变量大数据应用、调节变量团队协作能力和因变量 R&D 团队知识创造均为连续变量，因此本文采用分层回归分析检验团队协作能力的调节作用。数据去中心化后第一步检验假设 H0。第二步在第一步基础上加入调节变量得到模型 1，检验模型 1 得到其 R<sub>1</sub><sup>2</sup>。第三步在第二步基础上加入交互项（大数据应用和团队协作能力的乘积）得到模型 2，检验模型 2 得到其 R<sub>2</sub><sup>2</sup>，若 R<sub>2</sub><sup>2</sup> 显著大于 R<sub>1</sub><sup>2</sup>，且交互项的显著性水平 sig < 0.05，则表明交互项对 R&D 团队知识创造产生影响。本文调节作用检验结果如下：

表 7.44 回归模型总体情况

model	R	R Square	Adjusted R Square	Std. Error of The Estimate	R Square Change	Change Statistics				Durbin-Watson
						F Change	Df1	Df2	Sig. F Change	
1	.170 <sup>a</sup>	.032	.028	15.63671	.032	5.591	2	379	.004	1.926
2	.199 <sup>b</sup>	.041	.033	15.57002	.018	4.244	1	378	.048	

从表 7.45 看模型 1 和模型 2 的 R<sup>2</sup> 分别为 .032 和 .041，且交互项的 sig F change = 0.048，小于 0.05，证明调节效应存在。

表 7.45 回归系数情况表

model	Unstandardized Coefficients		Standardized Coefficients	t	sig
	B	Std. Error	Beta		
1 (Constant)	.483	.813		.594	.553
大数据应用	.035	.033	.061	1.055	.292
团队协作能力	.366	.157	.134	2.333	.020
2 (Constant)	1.129	.869		1.299	.195
大数据应用	.025	.033	.044	.753	.452
团队协作能力	.287	.161	.105	1.781	.076
大数据应用*团队协作能力	-.008	.004	-.112	-2.042	.043

从表 7.47 看，交互项的 sig=0.043，小于 0.05，因此，团队协作能力对大数据应用和 KIBS 企业知识创造调节作用显著。

## 7.3 实证结论分析和讨论

### 7.3.1 结论分析

从大数据应用与 R&D 团队知识创造的假设检验结果看，R<sup>2</sup> 为 0.186，显著性水平 P<0.05，回归系数值为 0.431，可见假设得到检验，大数据应用对 KIBS 企业知识创造产生显著的正向影响，这一研究结果符合本文研究预期。

又从大数据应用与 R&D 团队知识创造每个维度的假设检验结果看，假设均得到检验，即大数据应用对知识创造每个转化过程均产生显著的正向影响。这与 Vaccaro 等人(2009)认为在知识表出化过程没有发现信息技术带来积极贡献的观点并不相符。从知识联结化的假设检验结果 R<sup>2</sup> 为 0.547 看，大数据应用对知识联结化过程产生的贡献最高。以下为大数据应用与知识创造四个维度的关系分析：

#### (1) 大数据应用与知识社会化关系

尽管三个假设均成立，但从影响程度看，技术智能性对 R&D 团队知识社会化的影响程度比其他两个变量高。而强可视化性对 R&D 团队知识内在化的影响程度最低，但总体差别不大。

#### (2) 大数据应用与知识表出化关系

三个假设均成立，但从影响程度看，技术外包性对 R&D 团队知识表出化的影响程度最高。而技术智能性对 R&D 团队知识内在化的影响程度最低，但总体差别不大。

#### (3) 大数据应用与知识联结化关系

三个假设均成立，但从影响程度看，技术外包性对 R&D 团队知识联结化的影响程度比其他两个变量高。而强可视化性对 R&D 团队知识内在化的影响程度最低，但总体差别也不大。

#### (4) 大数据应用与知识内在化关系

三个假设均成立，但从影响程度看，技术外包性对 R&D 团队知识表出化的影响程度最高。而技术智能性对 R&D 团队知识内在化的影响程度最低，但总体差别不大。

### 7.3.2 结论讨论

从本论文的实证检验结论看，大数据应用对知识创造四个过程均有促进作用，因此也促进了知识型企业的知识创造，从而提高企业竞争力。大数据应用的技术特性，使得中小型企业实施大数据应用的过程当中摆脱自身 IT 能力的束缚，拓展了企业的业务能力，提高业务效率和质量，同时也简化了业务流程，减少了企业人员，从而降低企业成本。

大数据应用还能使中小型企业拓展了企业资源，使得企业内部和企业之间的信息共享更容易，沟通更流畅，从而有效降低了因地域问题而产生的数据（或信息）鸿沟，从而一定程度上缓解企业间信息不对称的问题。

总的来说，本论文的实证结果与现有管理学中信息技术相关理论的研究结论吻合，即信息技术可以促进企业的知识创造活动。

## 7.4 本章小结

本章首先采用小样本预测试的方法，在发放正式调查问卷之前再一次对问卷进行优化和修正，为后续正式调查问卷的大样本数据质量奠定了坚实基础。其次，以大样本数据对本论文提出的理论模型和研究假设进行实证检验，并获得大数据应用和 R&D 团队知识创造两个变量间影响关系的相关结论。最后，在实证结果的基础上，深入分析了大数据应用对企业知识创造的影响程度。同时也讨论了本论文的研究结论对企业所产生的影响，为企业的实践活动提供一定的指导意义。



## 第八章 结论与展望

### 8.1 研究结论和实践启示

大数据时代，企业各类生产活动的重大变化几乎都与 IT 紧密联系，IT 技术成为当代先进生产力的发展方向。IT 技术的应用可使企业的生产效率全方位提升，优化企业资源配置，从而提高企业核心竞争力。但传统的 IT 技术应用，均需要投入大量的资金和人员来建设 IT 基础设施及其软件平台，使得企业投入成本高，由此导致不少企业信息化失败。大数据概念及其处理技术的 IT 技术的出现，突出了 IT 技术创造创新的特性，使传统企业信息化方式发生了重大改变，企业不再需投入太多的资金和人员去建设信息化平台，只需将条件提供给服务商即可，让企业的大数据应用降低了门槛，在降低企业成本的同时提升了企业效率。

尽管近几年政府和企业均意识到大数据的应用能够推动企业工业化和信息化的相互融合，从而创造更高的经济效益，但目前我国企业的大数据应用没能大范围普及。企业的大数据应用会使企业发生怎样的变化，目前许多企业仍未真正了解，而学界在此领域的研究成果也甚少，这导致大部分企业对大数据应用还处于怀疑和观望状态，这也给国家大数据战略的实施带来问题。

鉴于上述原因，本论文基于知识创造理论、信息技术理论和软件工程理论，研究大数据应用与知识创造之间的作用关系。通过文献阅读，归纳了中小型企业大数据应用的 IT 特性，并在相关理论的基础上提出本论文的理论研究模型。通过案例研究，发现大数据应用下知识创造过程所发生的变化，再通过发放问卷和统计分析对理论研究模型进行实证，最后得出了大数据应用对企业知识创造产生正向影响的结论。

以下为本论文的主要研究结论：

**结论 1：大数据应用下 SECI 知识创造四个知识转化模式并未发生变化。但知识创造行为并不固定在 SECI 四个知识转化模式中的某个模式，其可随着“Ba”的变化而发生转移，即从一种转化模式转移到另一种转化模式。**

在大数据技术强大的知识创造创新能力下，尽管 Nonaka&Takeuchi 所提出的知识创造四个转化模式没有改变，但企业的知识创造行为方式却发生了变化。在大数据应用这个“Ba”下，很多知识创造行为从其他知识创造转化模式转移到了知识联结的转化模式，从而导致知识创造四个转化模式中，知识联结化模式比以往显得更为重要。因此，在大数据应用背景下，加强知识创造联结化中的知识创造行为或活动应该成为企业知识创造战略的重点。

本研究结论弥补了现有知识创造研究仅侧重过程研究和相关性研究，而知识

创造行为的研究被忽视的问题，丰富了 SECI 知识创造理论。实际上，知识创造行为的研究对现实的知识创造实践活动具有指导意义。

### **结论 1 的实践启示：**

#### **(1) 调整侧重点，制定正确的企业发展战略**

从这一研究结论得知，大数据应用可对企业知识创造行为可产生影响。大数据应用下知识创造行为或活动的侧重点向知识联结化转移。因此企业需要适时调整战略侧重点，即大数据应用应成为企业知识管理战略的侧重点，以此制定出科学合理的企业知识创造战略和信息化战略，从而提高企业知识创造效率，增强企业竞争力。

#### **(2) 适时调整企业组织结构，节约企业成本的同时提高企业效率**

大数据应用可使企业许多业务模式发生改变，业务流程变得更简单，这导致企业内部或组织内部结构发生改变。因此，大数据应用下企业必须适时调整企业内部或组织内部结构，以达到节约企业成本的同时提高企业效率的目的。

### **结论 2：大数据应用对 KIBS 企业 R&D 团队的知识创造有显著正影响。**

从本文研究可知，大数据应用对 KIBS 企业 R&D 团队的知识创造有显著正影响，而且对知识创造四个转化过程均有促进作用。这一结论与 Vaccaro 等人(2009)认为在知识表出化过程没有发现信息技术带来积极贡献的观点不相符。这为信息技术对知识创造作用的进一步讨论提供新的研究佐证。在研究大数据应用对 KIBS 企业 R&D 团队的知识创造影响研究中我们得到以下实践启示：

### **结论 2 的实践启示：**

#### **(1) 大数据创造创新的 IT 特性是推动企业大数据应用的核心动力**

大数据天生为创造创新而生，其应用目的就是达到创造创新的目的，其应用过程也是一个创造创新的过程。因此大数据创造创新的 IT 特性（包括应用模式的创新）是大数据技术有别于其它 IT 技术的重要标志，也是企业实施大数据应用的动力所在。本论文通过综合文献资料，归纳总结了大数据的三个 IT 应用特性，即技术外包性、技术智能性和强可视化性。大数据技术的外包性，使得企业摆脱了 IT 能力的束缚，放下 IT 资源问题的包袱，节约企业成本，全身去发展企业核心业务和企业创新。大数据的技术智能性是大数据创造创新的核心所在，也是大数据核心价值所在。它使企业开展知识创造活动的过程得到简化，大幅提高知识创造的效率和质量，同时降低企业对人力资源的需求，也节约了企业成本。大数据强可视化性是大数据智能性的体现，大数据创造创新成果需要一种强可视化来满足用户需求和提高用户体验。

可见，大数据应用的 IT 特性是推动企业大数据应用的根本原因和发展动力。

#### **(2) 大数据应用可以提升企业效率，增强企业自身的竞争优势**

本论文实证结果表明，大数据应用对企业 R&D 团队知识创造产生显著的正向

影响，这一结果符合本论文研究的预期。大数据应用下的知识创造过程中，很多原来依赖人的知识创造行为或活动不再依赖人，而是通过机器来实现；很多工作完成时间是滞后的，但大数据应用下工作完成是实时的，创造结果的呈现也是实时的。这使得企业在大幅提高效率的同时，降低了对人力资源的需求，从而提升企业的竞争力。

**结论 3：团队协作能力对大数据应用和 KIBS 企业知识创造有调节作用。**

**结论 3 实践启示：**KIBS 企业在实施大数据应用过程当中，知识团队的协作能力会影响知识创造效果，当然也会影响知识产品的生产效率和产品质量。因此，企业在实施大数据应用的同时，必须注重企业内部知识团队的建设。

## 8.2 论文主要研究贡献

综合以上研究结果看，本论文研究的贡献主要有以下几个方面：

### （1）理论贡献方面

大数据技术是一种新兴的 IT 技术，因此其在管理学研究领域，其学术研究成果非常少，尤其关于对企业或组织的影响研究就更少。本论文研究是基于企业应用视角，研究了大数据应用对中小知识型企业 R&D 团队知识创造的影响，这丰富了管理学领域大数据的理论研究。本论文的影响研究主要包括：

①大数据应用影响知识创造过程中的行为方式。此前知识创造过程研究主要集中在知识转化模式及其场的研究。而知识创造行为没有纳入到知识创造过程研究中来。而本研究开创了这一先例。实际上，在知识创造过程研究中知识创造行为对现实中的实践活动具有指导意义。

②大数据应用对企业知识创造的影响。通过研究，实证了大数据应用对企业知识创造具有正影响作用。

③填补管理学领域大数据研究方法的空白。从已有文献看，管理学领域对大数据的研究采用案例分析为主。而本论文通过统计分析实证方法对大数据对知识创造影响进行了检验，这填补管理学领域大数据研究方法的空白。

④归纳总结了大数据应用的 IT 特征，指出了它与传统 IT 技术的差异，为今后有关大数据技术特征研究提供新的参考。

### （2）现实贡献方面

随着大数据技术的不断发展成熟，大数据应用将在企业中普及，并成为国家、企业乃至全社会信息化发展的主要动力之一。但是面对大数据这种新兴的 IT 技术，它能给国家、企业乃至全社会带来怎样的影响？本论文最后的研究结论，将为企业大数据应用决策，以及政府实施大数据战略提供相关理论依据，从而促进大数据技术进一步发展。

### 8.3 存在不足与未来展望

#### (1) 存在不足

尽管本论文研究得到了一定成果，弥补了管理研究领域大数据相关研究中的一些缺失，但由于个人能力、研究时间和研究经费等因素，本论文研究还存在以下不足：

##### ①样本数据收集

本论文研究为统计分析做了大量的前期工作，主要包括发放和回收问卷的方式获得样本数据，数据量符合研究要求，具备一定的代表性。但由于目前应用大数据的企业仍占少数，收集到的样本数据并不多，因此研究结果并不能完全客观反应大数据应用后企业的真实情况。这可能会影响到本研究成果的普适性。

##### ②中介（或控制）变量方面

本论文在分析大数据应用对企业 R&D 团队知识创造的影响过程中，并没有过多考虑其它相关影响因素。但在现实中，知识创造整个过程往往受到多种因素的影响，例如 R&D 团队应用大数据的时间等，这些中介（或控制）变量也可能会影响最后的研究结果，导致理论与实践产生差异。因此，增加控制变量的影响是今后本论文研究中需要进一步探讨和完善的。

#### (2) 未来展望

①寄希望于应用大数据来研究管理学领域中的相关问题，从而使研究摆脱对样本的依赖以及地域问题的影响，从而使结论更为准确，成果更具普适性。

②在未来的研究当中，将把其他因素的影响作用纳入到研究中来，适当增加中介变量或者控制变量，研究其对整个模型的影响，进一步完善本研究。

## 参考文献

- 艾尔·巴比(著),邱泽奇(译).(2005).社会研究方法[M].北京:北京大学出版社.
- 鲍明刚.(2004).销售人员能力差异、薪酬机制偏好及团队协作关系研究. 博士论文. 北京大学.
- Boisot.M.H 著.张群群,陈北译.(2005).知识资产—在信息经济中赢得竞争优势[M].上海世纪出版集团.
- Bernard Marr 著,宫鑫,刘姝汶,刘婷婷译.(2017).大数据专家-小企业也能用好大数据[M].中国工信出版社.
- 陈新辉.(2013).知识密集型服务企业知识创造体系研究[M].知识产权出版社.
- 蔡立志,武星,刘振宇.(2015).大数据测评[M].上海科学技术出版社.
- 陈永霞,贾良定,李超平,宋继文,张君君.(2006).变革型领导、心理授权与员工的组织承诺:中国情景下的实证研究[J].管理世界,(1):96-105.
- (美)Capers Jones 著,吴舜贤,杨传辉,韩生亮译.(2014).软件工程最佳实践[M].机械工业出版社.
- 逮宇铎,于娇,刘媛媛.(2012).三类 R&D 团队中个人绩效评价指标体系研究[J].科技进步与对策,29(19):141-144.
- 杜智敏.(2010).抽样调查与 SPSS 应用[M].北京:电子工业出版社.
- 董超,卢桂林.(2017).一本书搞懂企业大数据应用[M].化学工业出版社.
- 冯长利,赵常宁,刘丹,兰鹰.(2016).供应链企业间知识创造影响因素 Fuzzy DEMATEL 分析[J].科学学研究,5(32):734-742.
- 傅家骥.(1998).技术创新学[M].北京:清华大学出版社.

- 范群林.(2015).文化科技企业技术创新能力影响因素的跨层次研究[J].科技管理研究,20(20)133-137.
- 郭重庆.(2013).大数据时代管理学研究的战略方向[J].北京.经济与管理战略研究,27(3):3-4.
- 郭平,王可,罗阿理,薛明志.(2015).大数据分析中的计算智能研究现状与展望[J].软件学报,26(11):3010-3025.
- 郭亚军,赵路路,赵静.(2016).大数据环境下企业技术创新知识管理模式研究[J].现代情报,36(7):13-17.
- 国佳.(2012).软件企业知识管理关键因素研究.博士论文.吉林大学.
- 耿新,彭留英.(2004).企业知识的分类、分布与转化机制研究—系统化视角下对 SECI 模型的一个扩展[J].管理科学,17(4):43-48.
- 何永刚.(2007).信息系统开发过程中知识转移研究.博士论文.复旦大学.
- 何存花,姚晓民.(2014).基于决策矩阵的财务管理系统实时性评价[J].统计与决策,417(31):179-181.
- 和金生,德勇.(2004).知识管理应当研究什么?[J].科学学研究,22(1):70-75.
- 韩军,雷星晖,高琦.(2011).知识密集型服务业研究综述[J].经济论坛,(8):154-156.
- 韩智慧,李南等.(2004).知识创造的组织环境分析[J].科学学与科学技术管理,25(12):77-80.
- 胡鞍钢.(2003).建立世界最大的知识密集型社会[N].湖北日报网络版.
- 纪慧生.(2013).产品开发过程的知识创新研究[M].经济管理出版社.
- 柯江林,孙健敏,石金涛.(2009).变革型领导对 R&D 团队创新绩效的影响机制研究[J].南开评论,12(6):19-26.
- 邝孔武,王晓敏.(1999).信息系统分析与设计[M].清华大学出版社.

- 李九斤,叶雨晴,徐畅.(2015).大数据环境下会计知识转移效率提升问题[J].辽宁工程技术大学学报(社会科学版),21.8(4):378-381.
- 李朝明.(2013).基于动态能力的企业协同知识创新[M].清华大学出版社.
- 李建中,王宏志,高宏.(2016).大数据可用性的研究进展[J].软件学报,27(7):1605-1625.
- 李怀祖.(2004).管理研究方法论[M].西安:西安交通大学出版社.
- 李光达,谭章禄.(2017).基于认知科学的知识可视化过程及其影响因素研究[J].现代教育技术,27(3):64-70.
- 李民.(2013).复杂产品系统创新过程的知识创造机理及实证研究.博士论文.南京大学.
- 吕民乐,金妍.(2016).知识密集型服务业对中国制造业创新的影响—基于高技术制造业的实证分析[J].工业技术经济,(4):17-24.
- 罗胜强,姜嫵.(2014).管理学问卷调查研究方法[M].重庆大学出版社.
- 刘微微.(2016).高端装备制造业企业知识创新能力测试与评价研究[M].哈尔滨工程大学出版社.
- 刘顺忠.(2004).对创新系统中 KIBS 的功能研究[J].科学学与科学技术管理,(6):71-76.
- 刘甲学.(2006).超媒体信息空间智能导航理论与实证研究.博士论文.吉林大学.
- 刘军.(2008).整体网分析讲义-ucinet 软件实用指南[M].上海人民出版社.
- 凌云,吕王勇,张里静.(2014).基于交互验证的数据质量评估方法的研究[J].水资源与水工程学报,8(01):214-216.
- 路变玲,黄国青,闫博华.(2008).基于结构方程模型的研发团队绩效评价[J].科技管理研究,(7):315-317.
- (美)Louis Testa 著,郭稚辉译.(2011).管好团队做对事---软件企业成长手册[M].人民邮电出版社.

- 马小闲,龚国伟.(2006).信息质量评估研究[J].情报杂志,12(05):19-21.
- 马亚男.(2003).个人能力\_团队能力\_企业能力的转化扩散机制研究[J].科学学与科学技术管理,11(8)109-111.
- 马淑文.(2007).企业知识结构与知识创新关系研究[J].科技进步与对策,(12):148-151.
- 强茂山,袁尚南,温祺.(2015).工程项目团队能力的测量与评价[J].清华大学学报(自然科学版),6(55)624-632.
- 任磊,杜一,马帅,张小龙,戴国忠.(2014).大数据可视分析综述[J].软件学报,25(9):1909-1936.
- 孙大为,张广艳,郑纬民.(2014).大数据流式计算:关键技术及系统实例[J].软件学报,25(4):839-862.
- 孙以泽.(2003).数学能力的成分及其结构[J].南京晓庄学院学报,2(19):97-99.
- 孙阿楠,刘林红,夏琳.(2006).基于客户互动的软件服务创新模式研究[J].科技进步与对策,(5):88-90.
- 史海锋.(2005).R&D 项目组技术知识创造影响因素的实证研究.博士学位论文.清华大学.
- 申静.(2006).知识性服务业服务创新[M].北京图书馆出版社.
- 汤超颖,丁雪辰.(2015).创新型企业研发团队知识基础与知识创造的关系研究[J].科学学与科学技术管理,36(9):81-92.
- [美]托马斯.H.达文波特等著.(2000).信息技术的商业价值[M].中国人民大学出版社,18-19.
- 王秋彤,宁景雨.(2015).我国知识密集型服务业发展建议.合作经济与科技.检自:  
<http://www.jjykj.com/viewnews.asp?id=20463>.
- 王重鸣,田茂利.(2006).技术创业企业知识创造过程及其影响因素[J].科研管理,(6):28-31.
- 王连娟,张跃先,张翼.(2016).知识管理[M].人民邮电出版社.



- 王黎莹,陈劲.(2010).研发团队创造力的影响机制研究—以团队共享心智模型为中介[J].科学学研究,28(3):420-428.
- 王嵘冰.(2015).云计算应用对企业组织变革影响研究.博士论文.辽宁大学.
- 王苏斌等.(2003).SPSS 统计分析[M].机械工业出版社.
- 吴明隆.(2003).SPSS 统计应用实务-问卷分析与应用统计[M].北京:科学出版社.
- 吴洁,刘思峰.(2006).基于熵理论的知识创造机理研究[J].工业技术经济,25(7), 68-71.
- 吴翠花.(2013).企业知识创造与技术创新[M].西安交通大学出版社.
- 魏江,胡胜蓉,袁立宏等.(2008).知识密集型服务企业与客户互动创新机制研究:以某咨询公司为例[J].西安电子科技大学学报(社会科学版),18(3):14-22.
- 魏江.(2004).产业集群中知识密集型服务业的功能研究[J].科技进步与对策,(6):7-10.
- 温忠麟,侯杰泰,张雷.(2005).调节效应与中介效应的比较和应用[J].心理学报,37(2):268-274.
- 汪鸿昌,肖静华,谢永勤.(2013).基于企业视角的云计算研究述评与未来展望[J].外国经济与管理,6(32)13-22.
- 维克托·迈尔-舍恩伯格.(2012).大数据时代:生活、工作与思维的大变革[M]浙江人民出版社.
- 徐璞.(2014).知识工作中的创新与信息技术---基于认知视角的分析[M].上海人民出版社.
- 徐寿光著.(2003).信息技术行业 ISO9001, 2000 及 CMM 质量管理体系文件精选[M].中国标准出版社.
- 徐宗本,冯芷艳,郭迅华,曾大军,陈国青.(2014).大数据驱动的管理与决策前沿课题[J].管理世界,(11):158-163.
- 徐璞.(2012).知识工作中的创新及信息技术的作用. 博士论文.华东大学.

- 谢文.(2012).看得见的未来——谈大数据时代[EO/OL].检自：  
[http://www.china-clound.com/yunjishu/hujuzhongxin/20121106\\_16045.html](http://www.china-clound.com/yunjishu/hujuzhongxin/20121106_16045.html).
- 宋学清,刘雨.(2014).大数据:信息技术与信息管理的变革[J].情报科学,32(9):14-17.
- 许小东.(2001)关于 R&D 团队建设与管理的思考[J].科学学研究.19(2):76-81.
- 薛会娟.(2013).研发团队中的效能感与创造力的关系\_跨层次研究[J].南开管理评论.56(16)71-76.
- 熊德勇、和金生.(2004).SECI 过程与知识发酵模型[J], 研究与发展管理, 16(2): 14-19.
- 辛枫冬.(2011).网络关系对知识型服务业服务创新能力的影响研究.博士论文.天津大学.
- 杨杜.(2013).管理学研究方法(第二版)[M].东北财经大学出版社.
- 杨德林,史海锋.(2005).R&D 项目组知识创造影响因素的实证研究. 科学学与科学技术管理,(7):92-96.
- 杨善林,周开乐.(2015).大数据中的管理问题:基于大数据的资源观[J].管理科学学报, 18(5):1-8.
- 叶元龄.(2013).基于知识共享的软件企业技术创新能力研究.博士论文.北京大学.
- 叶少波.(2006).政府统计数据质量评估方法及其应用研究. 博士论文.湖南大学.
- 袁晓婷.(2010).企业 R&D 团队内部社会网络与团队知识创造关系研究.博士论文.华南理工大学.
- 姚威.(2009).产学研合作创新的知识创造过程研究.博士论文.浙江大学.
- 野中郁次郎,竹内弘高(著),李萌,高飞(译).(2006).创造知识的企业——日美企业持续创新的动力[M].北京:知识产权出版社.
- 俞立平.(2016).大数据下高技术企业创新路径研究[M].经济科学出版社.

- 余来文,陈吉乐,温著彬,封智勇.(2014).大数商业模式[M].经济管理出版社.
- 茵明杰,李鑫,任红波.(2004).高技术企业知识创新模式研究—对野中郁次郎知识创造模型的修正与扩展[J].外国经济与管理,26(5),8-12.
- 应璇,孙济庆.(2016).基于大数据的精细化知识服务模型构建[J].科研管理,37(10):153-160.
- 晏双生.(2010).知识创造与知识创新的涵义及其关系论[J].科学学研究, 8(28):1148-1152.
- 张丹宁,杜晓君.(2007).知识密集型服务企业在中小企业集群知识创造中的功能分析[J]. 东北大学学报(社会科学版),3(9):228-232.
- 张嵩,黄立平.(2003).基于资源观的企业信息技术能力分析[J]. 同济大学学报(社会科学版),(14)4:52-56.
- 张玉珍.(2006).知识密集型服务业在知识创新系统中的功能研究[J].情报杂志, (10):129-132.
- 张树娟,刘政.(2008).研发团队知识创造的影响因素及绩效评估[J].情报杂志, (1):18-20.
- 张伟.(2016).资源型产业链知识创造影响因素研究—基于贵州中部磷化工产业链的分析[J].管理学报,6(13):871-888.
- 张巨俭,甘仞初.(2003).管理信息系统的发展方向及实现技术[J].计算机应用研究, (1):8-10.
- 张文彤主编.(2004).SPSS 统计分析高级教程[M].北京:高等教育出版社,9: 213-234.
- 赵炎,王晨.(2009).知识密集型服务业的集群创新及创新系统文献综述[J].科技进步与对策,24(26):195-200.
- 赵眸光,赵勇.(2017).大数据数据管理与数据工程[M].清华大学出版社.
- 周正,辛自强.(2012).数学能力与决策的关系\_个体差异的视角[J].心理科学进展,4(20)542-551.
- 朱洁,罗华霖.(2016).大数据架构详解-从数据获取到深度学习[M].电子工业出版社.

- 邹欣.(2015).现代软件工程构建之法[M].人民邮电出版社.
- 郑炜,吴潇雪.(2016).现代软件工程 [M].西北工业大学出版社.
- 仲秋雁,闵庆飞,吴力文.(2004).中国企业 ERP 实施关键成功因素的实证研究[J].中国软科学,73-78.
- 钟鸿生.(2012).大数据的前世今生[EO/OL].检自:  
[http://blog.sina.com.cn/s/blog\\_540835df0101bh82.html](http://blog.sina.com.cn/s/blog_540835df0101bh82.html).
- Anandhi S. (2000).Bharadwaj: A Resource-based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation, MIS Quarterly, March, 24(1), 169-196.
- Ackoff, Russell. (1989).From Data to Wisdom, Journal of Applied Systems Analysis.(16): 3-9.
- Arikan A T. (2009).Interfirm knowledge exchanges and the knowledge creation capability of clusters[J].Academy of Management Review,34(4) : 658-676.
- Beckman T. (1999).The Current State of Knowledge Management, in J Liebowitz(Ed) , Knowledge Management Handbook. Boca Raton, F 1:Crc Press,1-22.
- Brauner E, Scholl W. (2000). The information processing approach as perspective for group research [J].Group Processes and Intergroup Relation, 3 (1):115-122.
- Browning.C,Singelman.J. (1975).TheEmergenceofa Service Society [M]. Springfield, 39.
- Bllatt, G.(2000) Organizing knowledge in t11e knowledge development cycle. Journal of knowledge management [J].4(1): 15-26.
- BANSAL H S,VOYER P D,MINIARD P W. (1995).Consumer he havior[M].New York:Dryden Press,21.
- Bolisani, E., Sacorso, E. (2000).Information technology management:a knowledge-based perspective.Technovation, Vol.21 No.4:473-496.

- Cooper R.B.&Zmud R.W. (1990). Information technology implementation research: a technological diffusion approach[J].*Manage Sci* , 26(2):123-39.
- Chang J.,Choi, B., Lee H. (2004).An organizational memory. For facilitating knowledge:an application to e-business architecture.*Expert System with Applications*,Vol.26 No.2,203-215.
- Cohen, W.M., Levinthal, D.A. (1990).Absorptive capacity:a new perspective on learning and innovation.*Administrative Science Quarterly*,Vol,35 No.1, 1128-1152.
- Davenport, T.H., Prusak, L. (1998).*Working Knowledge: How Organizations Manage What They Know*[M].Boston,Mass: Harvard Business School Press.
- Davenport, E.;Hall.H.(2001).New knowledge and micro-level online organization communities of practice as a development framework.*Proceedings of the Hawaii International Conference on system Sciencs*, Washington DC, IEEE Computer Society, V01.4, pp.1-10.
- Den Hertog, P.(2000).Knowledge-Intengve Business Services as Co-Producers of Innovation.*International Journal of Innovation Management*, (4):491-528.
- Davis, F.D.(1990).Perceived usefulness,perceived ease of use,and user acceptance of information.*MIS Quarterly*, Vol,13 No.3,319-340.
- Dutta, S. (1997).Strategies for implementation knowledge-base system.*IEEE Transactions on Engineering Management*, Vol,44 No.1,79-90.
- Fern E.F.,(1982).The use of focus groups for idea generation: the effect of group size,acquaintanceship, and moderator on response quantity and quality[J]. *Journal of Marketing Research*.2:1-13.
- Fong S.W. (2003).Knowledge creation in multidisciplinary projectteams: an empirical study of the processes and their dynamic interrelationships[J],*International Journal of Project Management*, 21:479-486.
- Grant, R.M.(1996).Prospering in dynamical-competitive environments:organizational capability as knowledge integration organization Science, Vol,7 No.4,375-387.

- George J.(1990).Personality,affect and behavior in groups[J].Journal of APPLIED Psychology,75:107-116.
- Hedlund G.(1994).A Model of Knowledge Management and the N-From Corporation[J].Strategy Management Journal ,15(3):73-90.
- Harold Leavitt&Thomas Whisler.(1958).Management in the 1980's.Harvard Business Review,Vol.36 No.6,pp.41-48.
- Holsapple C.W. Singh M.(2001).The Knowledge Chain Model:Activities for competitiveness [M].Expert Systems with Applications.
- HART OLIVER, BENT HOMSTROM.(1987).The theory of contract [C].Advance in Economic Theory Fifth World Congress Bewely.
- Haldin-Herrgard, T.(2000).Difficulties in diffusion of tacit knowledge in organizations,Journal of Intellectual Cpaital,Vol.1 No.4,357-365.
- Hamel G.(1991).Competition for competence and interpartner learning within international strategic alliances[J] .Strategic Management Journal, (12):83-103.
- Ikujiro Nonaka: Noboru Konno.(1998).The concept of "ba":Building a foundation for knowledge creation[J]. California Management Review.40(3):40-54.
- Krogh GV.(1998).Care in knowledge creation[J].California Management Review, 40(3):133-153.
- Krogh G V,Nonaka I, Rechsteiner L.(2012).Leadership in organizational knowledge creation: A Review and Framework[J].Journal of Management Studies, 49 (1) : 240-277.
- Kleinknecht A.(1989). Firm size and innovation[J].small Business Economics,1(3):215-222.
- Klein K.J., Lim B.C., Saltz J.L., et al., .(1989).How do they get there?An examination of the antecedents of centrality in team networks[J].Academy of Management Journal,47:952-963.
- Lee, H.,Choi, B.(2003).Knowledge management enablers,processes, and Organizations performance:an integrative view and empirical examination.Journal of management

- Information System, Vol.20 No.1, 179-228.
- Lee V, Oguntebi J. (2012). Toward learning and knowledge creation: Operationalising the Social Learning Cycle [J]. *Journal of General Management*, 37 (4) :29-53.
- Lovelace K., Shapiro D., Weingart L.R., (2001). Teams. Innovativeness and constraint adherence: Maximizing cross-functional new product a conflict communications perspective [J]. *Academy of Management Journal*, 44(4):779-793.
- Leslie Willcocks, Lester, S. (1996). The evaluation and management of information system investments: from feasibility to routine operations. In Willcocks, L. (eds), *Investing in information system: evaluation and management*. London Chapman & Hall, pp.15-36.
- Leibrecht, S., (2004). Techniques for the integration of expert knowledge into the development of environmentally sound products. *Journal of Engineering Design*, Vol.15 No.4. 353-366.
- Luo, X., Slotegraaf, R.J., and Pan, X. (2006). Cross-Functional Cooperation: The Simultaneous Role of Cooperation and Competition within Firms. *Journal of Marketing*, 70(2): 67-80.
- Loch, C.H., & Tapper. (2002). Implementing a strategy-driven performance measurement system for an applied research group. *Journal of Product Innovation Management*, 19: 185-198.
- McFarlan, F. M. (1984). "Information Technology Change -the Way you Compete". *Harvard Business Review*. Vol. 62, May-June, 98-103.
- Madhavan R., Grover R. (1998). From embedded knowledge to embodied knowledge: new product development as knowledge management [J]. *The Journal of Marketing*, 62(10):1-12.
- Miles, I., Kastrinos, N., Flanagan K., Bilderbeek, R., den Hertog, P. (1995). *Knowledge-intensive business services Users, carriers and sources of innovation*. Manchester: PR FST.
- Miles I., (2005). Knowledge intensive business services: prospects and policies [J]. *Foresight – The Journal of Future studies, Strategic Thinking and Policy*, 7(6):39-63.
- Muller, E., Zenker, A. (2001). Business services as actors of knowledge transformation: the role of KIBS in regional and national innovation systems *Research Policy*, (30): 1501-1516.

- McLure W M, Faraj S.(2005). Why should I share? examining social capital and knowledge[J]. *Management Information Systems Quarterly*, 29 (1):35-57.
- Nonaka, I.(1994). A dynamic theory of organizational knowledge creation *Organization Science*, (5):14-37.
- Nonaka L, Takeuchi H.(1995). *The knowledge-creating company*[M], Oxford: Oxford Press.
- Nonaka, I.&Konno, N.,(1998). "The Concept of Ba Building a Foundation for Knowledge Creation", *California Management Review*, 40/3, pp. 40-54.
- Nonaka L, Toyama R., Konno N.,(2000). SECI, Ba and leadership: a unified model of dynamic knowledge creation[J], *Long Range Planning*, 33:5-34.
- OECD, *Science, Technology and Industry Scoreboard 2003: Towards a Knowledge-based economy* [EB/OL]. Retrieved from <http://www.OECD.org/publications/e-book92-2003-04-1-7294/2003-4-1>.
- Olivera, F.(2000). Memory system in organizations: an empirical investigation of mechanisms for knowledge collection, storage and access. *Journal of Management studies*, Vol.37 No.6, 810-832.
- Porter, Michael E.(1985). "How Information Gives You Competitive Advantage". *Harvard Business Review*, June-August.
- Rolf Blumentritt, Ron Jolmston.(1999). Towards a strategy for knowledge management. *Technology analysis & strategic management* [J]. 11(3): 287-300.
- Ross, J. W., Beath, C. M., Goodhue, D. L.(1996). Develop Long-term Competitiveness through IT Assets, *Sloan Management Review*, 38 (1), 31-42.
- Strambach S.(2001). Innovation processes and the role of knowledge-intensive business services. Koschatzky K Kulicke M, Zenker A. *Innovation Networks — Concepts and Challenges in the European Perspectives*(M) .Physica, Heidelberg, 53-68.
- Scott, J.E..(1998). Organizational knowledge and the intranet. *Decision Support Systems*, Vol.23 No.1, pp.3-17.



- Smith K G, Collins C J, Clark K D.(2005).Existing knowledge, knowledge creation capability, and the rate of new productintroduction in high-technology firms[J].Academy of Management Journal,48( 2):346-357.
- Tether, B.S ,Hipp C. and Miles I.(2001).Standardization and Particularization in Services:Evidence from Germany [J].Research Policy,30(7):1115-1138.
- Vaccaro, A.,Veloso,F.,Brusoni, S.(2009).The impact of virtual technologies on knowledge-based processes:an empirical study.Research Policy, Vol.38 No.8,1278-1287.
- Wagner,C.(2003).Enhancing e-government in developing countries: managing knowledge through Virual communities. Electronic Journal on Information Systems in Developing Countries, Vol.14 No.4, pp.1-20.
- Wood, P. A., Bryson, J., Keeble, D.(1993).Regional patterns of small firm development in the business services:evidence from the United Kinom. Environment and Planning,(25): 677-700.
- West, M. A.(1990).The social psychology of innovation in groups, In M.A. West&J. L. Farr(Eds),Innovation and creativity at work;Psychological and organizational strategies[M].Chichester, England; Wiley, 309-333.
- Werrier, B. IVI. and Souder, W. E.(1997).Measuring R&D performance-U.S. and German Practices., Research Technology Management,40(3):28-32.
- Yoon.T,Matsushima,K. (2002).Knowledge fusion among the virtual production enterprises within the technology information infrastructure enuironment. International Engineering Management Conference, Cambridge. IEEE,Vol.1, pp.35-40.

## 附录 A 大数据应用对软件 R&D 团队知识创造影响调查问卷

### 调查问卷

#### 卷首语

尊敬的公司领导、软件团队各位朋友：

您好！这是一份关于大数据应用对企业软件 R&D 团队知识创造影响研究的学术性问卷。希望藉由您在这份问卷所提供的资料，了解大数据应用对知识创造产生的影响，以作为企业服务创造的借鉴和思考。希望您能够在百忙之中帮助我们填写这份问卷。您的意见和答案将为本研究提供非常重要的帮助。

本问卷纯属学术性研究，没有任何商业用途，请您放心并尽可能客观回答。我们承诺，我们将对您提供的所有信息严格保密。如果您对本研究结论感兴趣，我们会在研究结束之后将研究成果提供给贵方参考。烦劳之处，敬请见谅。

非常感谢您的帮助和支持！

## 填表说明

以下各部分请您对自己近年来参与的软件服务大数据应用项目进行评价，请根据当时的真实情况进行选择。

1. 题项中 1 为非常不同意，2 为不同意，3 为不确定，4 为同意，5 为非常同意，请您在相应框内打√。

2. 请根据您在一般情形下所持有的最直接的观念、感觉来回答，不需考虑太久。本答题没有对错之分，也没有标准答案。您只需客观选择即可。在答题时请您不要都打一样的分，也不要遗漏某些题项。

3. 大数据应用是指您在开展软件研发过程中，应用大数据为自己的研发工作服务。

4. 知识创造是指在软件研发过程中创造出新的且有价值的知识、技术和产品。

## 公司基本信息

(由公司相关负责人填写)

贵公司及填表人的基本信息

1. 公司所在地 \_\_\_\_\_
2. 公司存续时间 \_\_\_\_\_
3. 公司业务范围 \_\_\_\_\_  
 软件研发             软件测试             软件外包  
 软件技术咨询         软件销售             其它(请写出)
4. 公司的年销售收入  
 50 万以下         50 - 1000 万         1000 万- 1 亿         1 亿以上
5. 研发投入占销售收入的比例  
 1% 以下     1% -2%     2% -5%     5% -10%     10% 以上
6. 公司员工数量  
 10 以下     10 - 100 人     100 - 300 人     300 人以上
7. 公司应用大数据为公司业务服务的时间  
 6 个月以下     6 个月-1 年     2 年     3 年     3 年以上

## 问卷主体

如果您是自已填答问卷，请注意以下几点：

1. 答案无所谓“对”或“错”，您只要如实回答就行。
2. 请您亲自填答，不要找他人代答，也不要与他人商量。
3. 请您按照问卷中的提示打√选择答案或进行填写。
4. 请您逐题回答，不要遗漏。

### 第一部分 填表人基本情况

A1. 您的受教育程度

1.  专科以下    2.  专科    3.  本科    4.  硕士    5.  博士

A2. 您在软件研发团队内的工作

1.  需求分析与功能设计    2.  架构设计    3.  程序员    4.  软件测试员  
5.  数据库管理    6.  团队负责人    7.  项目经理    8.  其他

A3. 您所在软件研发团队代码是：\_\_\_\_\_。（注：公司名称前两字全拼+流水号），您在软件研发团队中的流水号是：\_\_\_\_\_。（由团队成员商定，保证流水号唯一即可）

### 第二部分 大数据应用

D1. 技术外包性

题项	1 非常 不同意	2 不同 意	3 不确 定	4 同 意	5 非常 同意
D11-大数据服务商能按公司需求提供 IT 服务	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D12-大数据服务商提供业务参考符合预期率高	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D13-大数据服务商使用的数据源会持续更新且与业务需求相符	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D14-公司对大数据服务商提供的结果评价通过率高	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D15-大数据服务商所使用的数据得出结果准确性高	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D16-相信大数据服务商能提供有价值的结果	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

D2. 技术智能性

题项	1 非常 不同意	2 不同 意	3 不确 定	4 同意	5 非常 同意
D21-大数据服务商提供预测结果准确率高	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D22-大数据服务能减少人工干预	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D23-大数据服务自动解决业务问题能力高	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D24-大数据服务经常提供新颖且有价值的结果	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

D3. 强可视化性

题项	1 非常 不同意	2 不同 意	3 不确 定	4 同意	5 非常 同意
D31-大数据服务商提供的信息和呈现方式足够丰富	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D32-大数据服务商提供的信息直观且通俗易懂	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D33-大数据服务能实时地、稳定地呈现相关信息	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D34-网络稳定且信息更新速度快	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D35-结果可以通过各种方式完整呈现	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### 第三部分 R&D 团队知识创造

#### Z1. 知识社会化

题项	1 非常 不同 意	2 不同 意	3 不确 定	4 同意	5 非常 同意
Z11-团队加强与销售部门、客户、大数据服务商交流和信息分享	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z12-注重团队内对软件开发的数据应用集体思考、经验总结和学习提高	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z13-注重向团队外部学习,发现新大数据技术和市场机会,并形成创新	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z14-团队内有互学互助、轻松和谐的工作环境和氛围	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

#### Z2. 知识表出化

题项	1 非常 不同 意	2 不同 意	3 不确 定	4 同意	5 非常 同意
Z21-团队内鼓励进行头脑风暴和对话来解决软件开发大数据应用问题	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z22-团队内的大数据应用技巧和经验进行文字和规范整理	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z23-团队内创造总结交流大数据应用的技术、经验、知识和思想的机会	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z24-鼓励团队内使用比喻、类比、归纳等描述新的大数据应用知识	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z25-团队对优秀员工进行宣传、经验推广和交流	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z26-团队内创造思考的环境与氛围,促进大数据应用知识共享方式和工具的使用	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z27-团队内建设大数据应用知识共享系统,制定知识贡献的激励机制	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### Z3. 知识联结化

题项	1 非常 不同意	2 不同 意	3 不确 定	4 同意	5 非常 同意
Z31-收集团队内部和外部关于软件开发大数据应用的文献资料，通过预测等方法制定团队计划	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z32-收集各类管理数据、技术资料来增加团队的知识储备	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z33-团队通过应用大数据组织研发多款软件产品	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z34-制定完善软件产品的服务手册及相关资料	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z35-在团队内传播大数据相关的新技术、新概念和新思想	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z36-运用现代化信息技术工作在团队内进行大数据技术传播	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z37-鼓励员工积极查阅与大数据应用技术文献资料并提供创新建议	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### Z4. 知识内在化

题项	1 非常 不同意	2 不同 意	3 不确 定	4 同意	5 非常 同意
Z41-经常通过跨部门研发项目与外部技术团队沟通大数据应用技术	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z42-团队注重软件开发大数据应用相关知识和技术的收集、整理与共享	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z43-团队内部通过沟通、理解来分享团队对软件开发大数据应用的愿景和价值理念	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z44-团队注重整体软件开发大数据应用能力的提高，形成“传帮带”机制	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Z45-注重团队内部的软件开发大数据应用相关知识和业务技能的培训	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



## 第四部分 团队协作能力

### N. 团队协作能力

题项	1 非常 不同 意	2 不同 意	3 不确 定	4 同意	5 非常 同意
N1-团队成员之间常互相帮助解决软件开发大数据应用上的问题	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
N2-团队成员之间常提出一些软件开发大数据应用有用的意见和建议	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
N3-团队成员之间常介绍自己在软件开发大数据应用上的经验和教训	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
N4-团队成员获得软件开发大数据应用新技术或知识后会与同事共同学习	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
N5-团队成员常在一起讨论项目开发的大数据应用相关问题	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

---

**调查到此结束，再一次感谢您的帮助！**

## 致 谢

衷心感谢导师韩圣龙教授对本人的精心指导。导师鼓励我在学术研究的路上有信心坚持前行；培养我阅读文献的习惯；指导我论文写作的整个过程中的每一个环节，深入浅出的讲解使我茅塞顿开。他的言传身教将使我终生受益。

感谢徐二明教授、罗教讲教授、吕海军教授、鲁柏翔教授、汤之敏、华国伟教授、刘春生教授、陈禀老师等热心指导与帮助，不胜感激。

感谢百色学院的领导和同事在工作学习上给予了许多支持和帮助，在此，向唐拥军书记、金长义校长、徐魁峰副书记、韦复生副校长、吴佩杰副校长等诸位领导和同事表示衷心的感谢。

感谢泰国正大管理学院洪风院长、张玉兰等老师为我们提供的服务和支持。

感谢给我指点和鼓励的同学们，他们在我的学习生活中给予诸多帮助支持。

感谢我的家人，在我紧张学习期间承担起所有家庭事务，不让我分心。家人的宽容、理解与支持才让我有时间完成学业。

黄小龙

二〇一八年八月八日

## 声 明

作者郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。尽我所知，除文中已经注明引用内容和致谢的地方外，本论文不包含其他个人或集体已经发表的研究成果，也不包含其他已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的贡献均已在独立研究报告中做了明确的说明并表示了谢意。

若有不实之处，本人愿意承担相关法律责任。

学位论文题目：大数据应用对 KIBS 企业 R&D 团队知识创造影响研究  
——以中小型软件企业为例

作 者 签 名：

日期：2018 年 8 月 8 日

## 个人简历

姓名：黄小龙

学历：

2010年7月 工程硕士学位 软件工程专业  
电子科技大学

2001年7月 理学学士学位 计算机科学教育  
广西师范学院

职业：高校教师

工作地点：百色学院，中国广西百色市中山二路21号

E-mail：741118584@qq.com

专业能力/特长：软件开发