

学者论坛

大数据行业风险管理研究

陆岷峰¹, 沈黎怡²

1.南京财经大学中国区域金融研究中心,江苏南京 210046;2.南京工业大学,江苏南京 211816

[摘要]受政策引导及市场预期的影响,大数据行业发展迅猛,但其行业风险也日渐凸显,数据分析质量风险与数据安全防护风险已成为阻碍大数据行业可持续发展的一大瓶颈。大数据行业风险主要来源于政策层面、行业层面以及大数据企业自身层面,且其风险机制复杂,各风险因素之间存在相互影响关系,正确认识风险、评价风险是促进大数据行业稳健运营的重要一环。解释结构模型对客观评价大数据行业风险机制具有积极意义,立足于风险机制的底层因素提出相关对策建议也能从根本上降低大数据行业的风险隐患,为大数据治理提供理论支撑。因此大数据风险管理应从提高个人信息安全意识、加快大数据法制建设、提高大数据企业核心竞争力等多方面入手,兼顾内外部共同发展。

[关键词]大数据;风险;解释结构模型

[中图分类号] F832.33; TP274 **[文献标识码]** A **[文章编号]** 2095-5537(2017)04-0003-10

Risk Management of Big Data Industry

LU Min-feng¹, SHEN Li-yi²

(1. A Nanjing University of Finance & Economics, Nanjing 210046; 2. Nanjing Tech University, Nanjing 211816)

[Abstract] Influenced by policy guidance and market expectation, big data industry has developed rapidly. But what followed are some prominent risks, among which data analysis quality and data security are two major bottlenecks hindering the sustainable development of the industry. These risks are at different levels: the policy level, the industry level and the enterprise level. And the risk mechanism is complex with mutual influence between different risk factors. So it is important to have a correct understanding and evaluation of the risks in order to guarantee a sound development of the industry. Interpretive structural model is helpful to objectively evaluate the risk mechanism. Countermeasures and suggestions put forward on the basis of the mechanism bottom can also help reduce the risks and provide theoretical support for the management of big data. Therefore, the management of big data should start from improving the awareness of personal information security. While taking both internal and external development into account, we also should speed up the construction of big data system and increase the core competitiveness of big data companies.

[Key words] big data; risk; Interpretive Structural Model(ISM)

[收稿日期] 2017-04-28

[作者简介] 1.陆岷峰(1962—),男,汉族,江苏省金湖人,南京财经大学中国区域金融研究中心首席研究员,教授,江苏银行总行高级会计师。研究方向:宏观经济,商业银行,互联网金融。

2.沈黎怡(1994—),女,汉族,江苏省无锡人,南京工业大学经济与管理学院硕士研究生。研究方向:供应链金融,互联网金融。

一、引言

随着信息数据的爆炸式增长,大数据概念逐渐进入大众视野,大数据行业主要通过海量数据的收集、处理等工作实现揭露规律并预测未来的目的。大数据行业兴起与发展为行业结合奠定基础,其应用范围包括营销、金融、工业、医疗、教育等诸多领域,各行各业在大数据技术的支撑下从海量数据中挖掘客户信息,寻求潜在发展空间,因此大数据技术已成为企业的核心竞争力,大数据行业发展前景广阔。据统计,2011年全球大数据市场规模仅73亿美元,随着大数据技术的逐渐成熟与市场应用推广,2015年的全球市场规模已增长至384亿美元。与此同时,国内的大数据投融资市场规模也由2011年的1.4亿美元猛增至2015年的24.5亿美元^①,大数据布局已上升至国家战略高度。国务院于2015年印发《促进大数据发展行动纲要》引导大数据行业的发展,李克强总理也在中国大数据产业峰会上充分肯定大数据的价值,提出大数据对于推进供给侧结构性改革、推动产业转型升级的重要意义。受政策引导及市场预期的影响,大数据市场热情高涨,大数据技术得到进一步的发展完善,但与此同时,行业风险也逐渐凸显,由此带来的各种乱象频发成为制约大数据行业可持续发展的一大瓶颈。大数据行业所面临的风险可细分为政策环境风险、行业环境风险与大数据企业自身风险三个层面。首先,政府层面虽对大数据行业的发展表现出高度关注,但其政策措施却处于相对滞后的状况,大数据行业的发展缺少法律政策的引导;其次,大数据行业乱象的根源在于行业整体的不协调发展,数据孤岛等阻碍行业可持续运营的问题并没有得到解决,行业协会的统筹规划作用并没有完全发挥;再次,大数据企业的内部问题,如人才、技术缺口

也对行业整体运营造成负面影响。大数据行业的价值在产业结合环节得到充分体现,其潜在风险隐患也在产业结合过程中被进一步放大,影响整体经济的运行。因此,正确认识与评估大数据行业的风险形成机制并有效降低风险已成为促进行业发展、推进产业转型升级的重中之重。

二、文献综述

(一)大数据及大数据行业内涵研究

随着大数据的不断发展,其定义趋向多元化,在学术界尚无统一观点。麦肯锡全球研究院(2011)将大数据定义为一种解决问题的方法,即通过收集、分析海量数据,并通过实验、算法及模型等研究方法发现一定的规律或有价值的见解,并以此促进新的商业模式的形成。姜锋(2014)认为大数据产业指覆盖数据的产生、汇集、处理、应用等活动的产业,并据此提出大数据行业运营的三种商业模式,包括以大数据技术为基础的软件开发模式、以大数据平台为基础的策划、资讯模式以及以大数据应用为基础的数据流通模式。袁冰(2014)指出大数据行业包括互联网行业、信息产业等,大数据产业生态体系的扩张势不可挡。

(二)大数据行业风险研究

孟小峰等(2013)提出大数据的规模效应正导致数据存储、管理等风险的产生,并进一步对数据分析结果的有效性造成重大影响。林志刚等(2013)则从宏观角度提出大数据管理不当将强化国家间的不均衡发展程度,也使得社会阶层力量失衡。蒋洁(2014)认为大数据轮动在推动新商业模式形成的同时也引发了严重的隐私风险,且在当前外部监管措施、用户隐私认知不足的情况下,隐私风险已成为制约行业发展的一大阻碍。巴曙松等(2016)认为当前大数据风控虽取得了一定的进展,但仍然存在有效性不足等问题,其原因主要在于数

据质量缺陷、理论性障碍以及制度障碍等。

(三)大数据行业发展对策研究

在发展方向方面,吕天文(2014)认为大数据行业的发展方向主要在于提升技术及服务水平,即创新数据处理技术并拓展差异化服务模式。在风险治理方面,陈火全(2015)从大数据行业的隐私危机出发,将信誉机制引入大数据治理环节,结合认证与授权机制、数据加密机制将风险降低至可接受水平,并提出从法制层面加强信息保护,加强行业自律性。在管理策略方面,王兵等(2016)提出大数据管理的目的在于挖掘数据功能与价值,并拓展其应用范围,因此其管理策略应侧重管理框架的创新及大数据的开发和内容管理。

大数据的特性在于体量大、种类繁多、价值密度低,但数据本身没有价值,价值创造在于数据的处理与分析过程。大数据行业如何获取数据、处理数据、分析数据与传递数据价值是该行业运营发展的基础,而如何规避以上业务环节的潜在风险则成为行业可持续发展的关键。学者们对于大数据行业的发展模式、发展前景等方面作出了深入的研究,但有关大数据行业风险管理的系统分析则相对不足。因此,精准的风险认知、预防与管控能力已成为大数据行业运营发展的关键。

三、大数据行业风险及危害研究

(一)基于大数据行业风险的解释结构模型的建立

解释结构模型(ISM)的核心在于将复杂系统分解为若干子系统,在简化系统结构的基础上进行定性分析。大数据行业风险因素繁多,且各风险点之间存在相互影响关系,共同制约大数据行业的稳健运营。因此将解释结构模型纳入大数据行业风险研究是一种较为科学的研究方式,对于统筹行业发展具有重要意义。

1.风险因素的选择

陈火全(2015)主要研究了大数据治理的安全问题,提出“社会隐私意识”、“政府信息安全政策”以及大数据企业的“数据安全的防护水平”三项风险因素;巴曙松(2016)则认为数据质量问题已成为制约大数据行业发展的瓶颈,并提出大数据行业在数据使用过程中存在的问题以及外在条件对其数据获取方面的不利影响,总体归纳为大数据行业层面“数据共享程度”、政府部门“数据开放政策”、大数据企业“数据可得性”三项风险因素;杭天竹(2016)则为“技术壁垒”与“人才团队建设”两项指标提供理论支撑;“数据分析质量”指标主要来源于邱东(2014)的理论研究。徐鹏等(2013)提出政府层面的“宏观调控政策”对于大数据行业发展的影响作用,主要涉及政府对于大数据行业发展的规范指导,如产业结合、大数据应用教育等。“市场需求”、“市场竞争”与“行业门槛政策”三项指标主要源于36大数据网有关大数据行业发展现状的分析研究。因此,结合前人观点,本文将大数据行业风险因素进行分类概括,如表1所示。

表1 大数据行业风险因素

一级指标	二级指标	三级指标(风险因素)
大数据行业风险	行业风险	数据质量 n_1
		数据共享程度 n_2
		市场竞争程度 n_3
		市场需求水平 n_4
		社会隐私意识 n_5
	政策风险	行业门槛政策 n_6
		信息安全政策 n_7
		宏观调控政策 n_8
		数据开放政策 n_9
	运营风险	人才团队建设 n_{10}
		数据分析质量 n_{11}
		技术壁垒 n_{12}
		数据可得性 n_{13}
		数据安全防护 n_{14}

2.构建大数据行业风险因素相互关系作用矩阵

将表1所归纳风险因素视作一个完整系统,系统要素之间存在相互影响关系,解释结构模型通过邻接矩阵表示该种影响关系(见图1),矩阵中的元素 a_{ij} 表示要素 n_i 与 n_j 之间的影响关系,若要素 n_i 对 n_j 存在直接影响关系,则 $a_{ij}=1$,否则 $a_{ij}=0$ 。

(1)数据质量 n_1 是大数据企业进行大数据分析的基础,对其分析质量起到直接影响作用,即 $a_{1,11}=1$ 。

(2)在当前数据私有的背景下,大数据行业的数据共享程度 n_2 与大数据企业的数据可得性 n_{13} 成正相关,而数据共享程度则受行业竞争程度 n_3 的影响,市场竞争程度越高,大数据企业参与数据共享活动的可能性越低,即 $a_{2,13}=1, a_{32}=1$ 。

(3)目前大数据行业热情高涨的根本原因在于市场对于大数据行业应用的积极预期与政府的正面引导,因此市场对大数据行业的需求水平 n_4 是决定大数据企业发展前景的基础因素,直接影响大数据市场竞争程度 n_3 ,而政府的宏观调控政策 n_8 则是决定市场需求水平的直接因素之一,即 $a_{43}=1, a_{84}=1$ 。

(4)社会隐私意识 n_5 与政府的信息安全政策 n_7 相互影响,政府颁布信息安全政策的积极性越高,社会隐私意识越强,而社会隐私意识也反向作用于政府行为,促使其以法律政策保障大众隐私权,即 $a_{57}=1, a_{75}=1$ 。

(5)政府目前并未对大数据行业门槛作出硬性规定,在一定程度上导致行业过度竞争,引发行业乱象,因此行业门槛政策 n_6 直接影响市场竞争程度 n_3 ,较高的行业门槛能有效缓解市场过度竞争,即 $a_{63}=1$ 。

(6)当前大数据行业数据来源稀缺,而政府部门掌握的关键数据却并未开放,在一定程度上制约大数据行业的发展,就行业发展趋势而言,政府将在保障信息安全 n_7 的基础之上逐步开放数据 n_9 ,增强大数据

行业数据的可得性 n_{13} ,即 $a_{79}=1, a_{9,13}=1$ 。

(7)大数据企业的数据防护措施 n_{14} 受到内在技术条件 n_{12} 与外在政策要求 n_7 的影响。此外,大数据企业的技术条件是大数据企业运营的基础,因此也直接影响行业门槛设置。即 $a_{7,14}=1, a_{12,14}=1, a_{12,6}=1$ 。

(8)大数据企业的人才团队建设 n_{10} 代表企业的核心价值,直接影响企业的运营与数据分析质量 n_{11} 。与此同时,数据可得性 n_{13} 也决定数据分析质量,数据可得性低可能造成数据分析结果片面等负面结果,即 $a_{10,11}=1, a_{13,11}=1$ 。

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭
①	0	0	0	0	0	0	0	0	0	0	0	0	0	0
②	0	0	0	0	0	0	0	0	0	0	0	0	1	0
③	0	1	0	0	0	0	0	0	0	0	0	0	0	0
④	0	0	1	0	0	0	0	0	0	0	0	0	0	0
⑤	0	0	0	0	0	0	1	0	0	0	0	0	0	0
⑥	0	0	1	0	0	0	0	0	0	0	0	0	0	0
⑦	0	0	0	0	1	0	0	0	1	0	0	0	0	1
⑧	0	0	0	1	0	0	0	0	0	0	0	0	0	0
⑨	0	0	0	0	0	0	0	0	0	0	0	0	1	0
⑩	0	0	0	0	0	0	0	0	0	0	1	0	0	0
⑪	0	0	0	0	0	0	0	0	0	0	0	0	0	0
⑫	0	0	0	0	0	1	0	0	0	0	0	0	0	1
⑬	0	0	0	0	0	0	0	0	0	0	1	0	0	0
⑭	0	0	0	0	0	0	0	0	0	0	0	0	0	0

图1 邻接矩阵

由邻接矩阵A加上单位矩阵I得初始可达矩阵 $(A+I)$ (见图2)。

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭
①	1	0	0	0	0	0	0	0	0	0	1	0	0	0
②	0	1	0	0	0	0	0	0	0	0	0	0	1	0
③	0	1	1	0	0	0	0	0	0	0	0	0	0	0
④	0	0	1	1	0	0	0	0	0	0	0	0	0	0
⑤	0	0	0	0	1	0	1	0	0	0	0	0	0	0
⑥	0	0	1	0	0	1	0	0	0	0	0	0	0	0
⑦	0	0	0	0	1	0	1	0	1	0	0	0	0	1
⑧	0	0	0	1	0	0	1	0	0	0	0	0	0	0
⑨	0	0	0	0	0	0	0	1	0	0	0	0	1	0
⑩	0	0	0	0	0	0	0	0	0	1	1	0	0	0
⑪	0	0	0	0	0	0	0	0	0	1	0	0	0	0
⑫	0	0	0	0	0	1	0	0	0	0	0	1	0	1
⑬	0	0	0	0	0	0	0	0	0	0	1	0	1	0
⑭	0	0	0	0	0	0	0	0	0	0	0	0	0	1

图2 可达矩阵

因为 $A+I \neq (A+I)^2 \neq (A+I)^3 \neq \dots \neq (A+I)^5 = (A+I)^6$,所以,可达矩阵 $M=(A+I)^5$ 的缩减矩阵为 M' (见图3)。

	①	②	③	④	⑤+⑦	⑥	⑧	⑨	⑩	⑪	⑫	⑬	⑭
①	1	0	0	0	0	0	0	0	0	1	0	0	0
②	0	1	0	0	0	0	0	0	0	1	0	1	0
③	0	1	1	0	0	0	0	0	0	1	0	1	0
④	0	1	1	1	0	0	0	0	0	1	0	1	0
⑤+⑦	0	0	0	0	1	0	0	1	0	1	0	1	1
⑥	0	1	1	0	0	1	0	0	0	1	0	1	0
⑧	0	1	1	1	0	0	1	0	0	1	0	1	0
⑨	0	0	0	0	0	0	0	1	0	1	0	1	0
⑩	0	0	0	0	0	0	0	0	1	1	0	0	0
⑪	0	0	0	0	0	0	0	0	0	1	0	0	0
⑫	0	1	1	0	0	1	0	0	0	1	1	1	1
⑬	0	0	0	0	0	0	0	0	0	1	0	1	0
⑭	0	0	0	0	0	0	0	0	0	0	0	0	1

图3 缩减矩阵

3.矩阵的分解

表2 可达集与先行集

i	R(i)	A(i)	R(i)∩A(i)
1	1,11	1	1
2	2,11,13	2,3,4,6,8,12	2
3	2,3,11,13	3,4,6,8,12	3
4	2,3,4,11,13	4,8	4
5+7	5+7,9,11,13,14	5+7	5+7
6	2,3,6,11,13	6,12	6
8	2,3,4,8,11,13	8	8
9	9,11,13	5+7,9	9
10	10,11	10	10
11	11	1,2,3,4,5,6,7,8,9,10,11,12,13	11
12	2,3,6,11,12,13,14	12	12
13	11,13	2,3,4,5+7,6,8,9,12,13	13
14	14	5+7,12,14	14

(1)区域分解

如图4所示,由 n_{11}, n_{13}, n_2, n_3 组成的子系统为矩阵A,由 n_4, n_6, n_9, n_{14} 组成的子系统为矩阵B,由 $n_1, n_5, n_7, n_8, n_{10}, n_{12}$ 组成的子系统为矩阵C。由矩阵的区域分解可得,矩阵A对矩阵B与矩阵C无影响,矩阵B对矩阵C无影响;矩阵B影响矩阵A,矩阵C影响矩阵A与B。即在大数据行业风险系统中,由市场需求水平、行业门槛政策、数据开放政策及数据安全防护风险所构成的子系统B对由数据分析质量、数据可得性、数据共享程度与市场竞争程度所构成的子系统A构成影响作用;由数据质量、社会隐私意识、信息安全政策、宏观调控政策、人

	①	③	②	③	④	⑥	⑨	⑭	①	⑤+⑦	⑧	⑩	⑫
①	1	0	0	0	0	0	0	0	0	0	0	0	0
③	1	1	0	0	0	0	0	0	0	0	0	0	0
②	1	1	1	0	0	0	0	0	0	0	0	0	0
③	1	1	1	1	0	0	0	0	0	0	0	0	0
④	1	1	1	1	1	0	0	0	0	0	0	0	0
⑥	1	1	1	1	0	1	0	0	0	0	0	0	0
⑨	1	1	0	0	0	0	1	0	0	0	0	0	0
⑭	0	0	0	0	0	0	0	1	0	0	0	0	0
①	1	0	0	0	0	0	0	0	1	0	0	0	0
⑤+⑦	1	1	0	0	0	0	1	1	0	1	0	0	0
⑧	1	1	1	1	1	0	0	0	0	0	1	0	0
⑩	1	0	0	0	0	0	0	0	0	0	0	0	1
⑫	1	1	1	1	0	1	0	1	0	0	0	0	1

图4 可达矩阵的分解

才团队建设及技术壁垒风险所构成的子系统C对子系统A与B施加影响。

(2)层级分解

根据 $R(i) \cap A(i) = R(i)$ 条件进行层级分解,由表2可知, $i=11, i=14$ 满足条件,表明 n_{11}, n_{14} 为该系统的最高层,将有关11,14的元素抽出可得表3。

表3 抽出11,14后的结果

i	R(i)	A(i)	R(i)∩A(i)
1	1	1	1
2	2,13	2,3,4,6,8,12	2
3	2,3,13	3,4,6,8,12	3
4	2,3,4,13	4,8	4
5+7	5+7,9,13	5+7	5+7
6	2,3,6,13	6,12	6
8	2,3,4,8,13	8	8
9	9,13	5+7,9	9
10	10	10	10
12	2,3,6,12,13,	12	12
13	13	2,3,4,5+7,6,8,9,12,13	13

由表3可得 $i=1, i=10, i=13$ 满足条件,因此 n_1, n_{10}, n_{13} 为第二层,将其抽出可得表4。

表4 抽出1,10,13后的结果

i	R(i)	A(i)	R(i)∩A(i)
2	2	2,3,4,6,8,12	2
3	2,3	3,4,6,8,12	3
4	2,3,4	4,8	4
5+7	5+7,9	5+7	5+7
6	2,3,6	6,12	6
8	2,3,4,8	8	8
9	9	5+7,9	9
12	2,3,6,12	12	12

由表4可得 $i=2, i=9$ 满足条件,因此 n_2, n_9 为第三层,将其抽出可得表5。

表5 抽出2,9后的结果

i	R(i)	A(i)	R(i)∩A(i)
3	3	3,4,6,8,12	3
4	3,4	4,8	4
5+7	5+7	5+7	5+7
6	3,6	6,12	6
8	3,4,8	8	8
12	3,6,12	12	12

由表5可得 $i=3, i=5+7$ 满足条件,因此 n_3, n_5, n_7 为第四层,将其抽出可得表6。

表6 抽出3、5+7后的结果

i	R(i)	A(i)	R(i)∩A(i)
4	4	4,8	4
6	6	6,12	6
8	4,8	8	8
12	6,12	12	12

由表6可得 $i=4, i=6$ 满足条件,因此 n_4, n_6 为第五层,将其抽出可得表7。

表7 抽出4、6后的结果

i	R(i)	A(i)	R(i)∩A(i)
8	8	8	8
12	12	12	12

由表7可得 n_8, n_{12} 为底层要素。

(二)大数据行业风险机制分析

根据以上模型分析可得出风险层级划分示意图(如图5),风险层级共6层,其中数据分析质量风险与数据安全防护风险为顶层风险,政府的宏观调控政策风险与大数据企业的技术壁垒风险为底层风险。该层级划分示意图清晰地展示了大数据行业风险的构成及其相互影响关系,其中顶层风险是大数据行业在运营过程中表现最突出、影响面最广的风险因素,而下层风险则是顶层风险的内在决定因素,下层风险因素之间的相互作用关系最终导致了顶层风险的形成,因此大数据行业的风险分析需立足于顶层风险因素,并对其进行剖析,深入研究风险体系的内在作用机制。

1. 数据分析质量的风险分析

就数据质量而言,大数据存在种类混杂

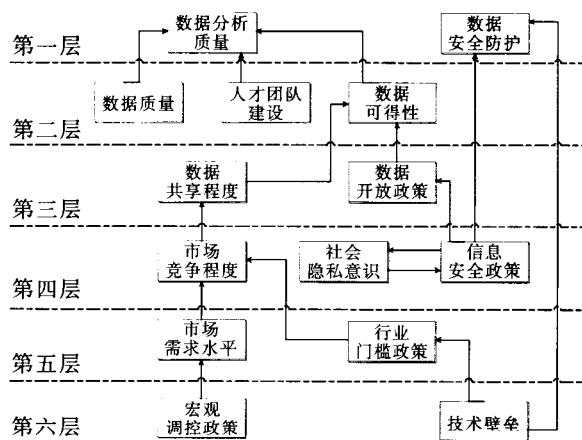


图5 各层关系示意图

且价值密度低等特性,数据失真、失效等问题正逐步瓦解大数据质量,以低质量数据为支撑的大数据分析、处理结果的可靠性不能得到保证。数据记录具有客观性,但技术及人为等因素则对其客观性造成较为明显的影响。以商业企业为例,正面的商业数据可能带来更多的商业机会,因此数据存在被篡改的风险,而大数据行业面临的数据质量风险则进一步增强,数据处理过程受到无效数据的干扰,影响数据分析结果。

就数据可得性而言,虽然大数据总量庞大且正处于逐年高速增长的态势,但依旧存在缺失。截至2016年底,央行征信系统收录的自然人规模为9.1亿,其中涵盖信贷记录的自然人仅4.3亿,4.8亿自然人的信用状况还未可知,对大数据行业的数据分析工作造成阻碍。数据资源匮乏的根源涉及多方面。首先,当前大数据行业发展模式不成熟,企业将数据资源视作客户基础,因此在大数据市场竞争激烈的背景下抗拒数据共享,最终导致数据孤岛的形成;其次,我国政府部门掌握相当部分高价值的关键数据,但出于信息保护等原因并未对公众开放,大数据企业难以通过自身途径获取该类数据。因此,提高数据可得性是大数据企业提高数据分析质量不可忽视的一环。

就人才团队建设而言,大数据行业正面临严峻的人才缺口。据统计,当前全国的大数据人才约46万,预计未来3至5年内人才缺口将高达150多万^②。人才团队是大数据企业的核心竞争力,大数据人才的培养是大数据技术开发的基础,也是大数据企业实现从传统架构转向大数据时代应用程序架构的前提。数据分析过程不仅依赖于大数据技术,更需要大数据人才发挥引导作用,将大数据分析报告与现实状况相结合,提高数据分析报告的附加值,为产业结合奠定基础。

数据质量风险的成因众多,涉及行业层面、政府政策层面以及企业自身运营层面,且其危害甚广,不仅影响大数据企业的可持续经营,也对产业结合模式下其他行业的发展造成不利影响。因此,有关部门及企业应加强合作以降低数据分析质量风险。

2. 数据安全防护的风险分析

大数据行业以海量数据为支撑,但如何获取数据并使用数据已成为行业争议的焦点。个人或群体参与社会活动留下数据记录,如消费者进行购物活动、病人获取医疗服务等都将产生数据记录,该类数据被保存在购物中心、医院等机构,但数据的所有权属于个人。大数据行业从数据存储方获取数据记录,但该过程却鲜少为数据所有者获悉并批准,由此产生数据安全风险。

个人数据涉及个人隐私,大数据行业获取并使用数据的过程存在泄漏个人信息的隐患,大数据行业就如何获取数据、获取哪些数据等关键问题并未达成共识,而政府部门在规范行业运营,保护个人信息安全方面所采取的措施尚不到位,相关法律条文也存在欠缺,使得大数据企业的运营活动缺少法律引导,社会大众难以摆脱数据安全风险的威胁,所开展的维权活动也难以达到预期结果。此外,数据安全风险也来源于大数据企业的内部问题。数据资源对于大数据企业的价值较高,内部从业人员的操作风险或职业道德风险可能导致数据的泄露,而大数据企业的技术壁垒也可能导致数据保护措施的失当,为外部黑客攻击等恶性事件的发生创造了条件。

总体而言,大数据企业数据安全防护风险贯穿于其业务运营的全过程,社会公众的隐私意识与政府部门的法制建设均对其起到不可忽视的影响,但数据安全在本质上还依赖于大数据企业的自律发展及技术建设。

(三) 大数据行业风险危害研究

1. 数据安全风险导致社会大众权益受到侵害

数据安全风险是大众权益受到侵害的根源所在。大数据企业通过何种渠道获取数据、获取哪些数据等与个人信息侵害风险息息相关。大数据企业的内外部因素如操作风险及技术风险等都可能导致个人隐私泄露或被恶意利用。以2015年优购网的信息安全事件为例,在该网站购物的消费者的个人信息被泄露,导致银行钱款被盗,损失金额近百万元。大数据行业与其他行业链接的价值在于数据传递,而数据安全风险也在该种传递过程中被进一步放大。在当前互联网技术高速发展的背景下,个人在社交网站中所记录的信息以及智能手机所收录的个人资料也放大了数据安全风险,通过数据组合分析便可轻易挖掘个人信息体系,但社会大众对于该种信息暴露风险几乎无所知觉,数据安全风险已随着大数据技术的发展深入到各个行业,一旦出现数据管理不当的情况,个人权益便会受到侵害。

2. 数据分析风险导致行业多样性的退化

大数据行业的应用范围甚广,包括营销、金融、医疗等行业,大数据分析结果对于其他行业的发展起到导向作用,但大数据行业的数据分析风险及质量风险则对该种行业结合模式造成重大影响。首先,大数据行业的技术限制是导致数据分析风险的一大诱因,由于大数据行业的准入门槛较低,大数据企业的资质参差不齐,数据分析精准度不能得到保证,因此依据大数据企业数据分析结果而做出的行业发展规划或与实际情况形成偏差。其次,即便大数据行业的数据分析结果的精准度较高,将该种数据分析结果纳入行业发展规划考量也存在一定的缺陷。以精准营销为例,企业通过

客户消费记录分析得出客户消费偏好,进而为其推荐符合其偏好的产品,大数据与营销行业的结合促使营销活动更具针对性,但与此同时,行业创新性却受到抑制,消费者的潜在需求并不一定与其以往消费偏好一致。因此,过度依赖大数据技术可能造成行业多样性的退化。

3. 过度竞争阻碍大数据行业的进一步发展

在大数据行业热情高涨的背景下,大数据企业数量激增,但行业准入门槛低造成当前行业发展不规范、乱象频发的问题。大数据市场容量有限而投入过量,造成大数据行业过度竞争,各种行业报告泛滥但其准确性却不能得到保证。就当前的发展状况而言,大数据行业正处于发展的初级阶段,大数据技术尚未被完全开发,与其他行业的结合模式也相对混乱,其应用范围还比较狭窄。众所周知,大数据行业对于数据的依赖性极强,而“大数据”与“数据”概念的主要区别就在于其规模与结构的差异,因此大数据企业对于大数据的数量及质量均有较高的要求,海量且不同维度的数据资源能进一步促进大数据行业的发展完善,而单个企业在数据收集及处理方面力量有限,因此数据共享需要行业合作的支持。但在当前行业过度竞争的背景下,大数据企业持有“数据资源即客户资源”的观点,忽视了数据共享的价值,因此对于行业合作持保留态度,在一定程度上抑制了行业进步。

四、大数据行业管理策略研究

(一) 个人层面

在当前大数据技术飞速发展,大数据行业范围不断扩张的背景下,数据安全风险逐渐凸显,个人信息安全遭受严峻的挑战。信息安全风险对于个人的侵害反映在各个方面,但个人对于该项风险的警惕性却相

对较低,部分社会群体对于信息安全的认识比较到位,但其应对信息安全风险的表现也仅限于认知层面,并未将隐私侵害等恶性事件诉诸法律或作出其它抵制行为。因此,提高个人信息安全意识对于为降低数据安全风险、规范大数据行业发展至关重要。一方面,加强对公众的信息安全教育,普及数据安全风险的危害,将信息安全上升至社会意识层面。另一方面,个人在互联网进行浏览信息、购物、注册等活动时应妥善处理个人信息,切勿将个人信息泄露给不受信任的平台或机构,在遭遇隐私侵害等事件时应主动寻求法律帮助,从个人层面遏制大数据行业的不规范运营行为。

(二) 政府层面

1. 出台相应政策法规,开放政府数据,降低大数据行业数据风险

大数据行业乱象在一定程度上反映出政府监管的不到位,数据安全风险侵害个人权益的现象时有发生,但相关法律政策却相对滞后,并未在该方面给予个人足够的信息安全保护,也并未对大数据行业的数据收集及处理工作作出明确的指导,难以支撑信息安全工作的开展。因此,有关部门应尽快落实法律政策,明确大数据行业的数据收集、处理规范,主要包括明确数据使用目的、确保个人对其信息使用情况的知情权、限制个人信息的留存期限、保障数据真实性等关键问题。此外,致使大数据行业发展受限的另一问题为数据质量问题,市场开放数据紧缺,关键数据主要由政府部门掌握的现状进一步放大了数据质量问题,因此有关部门应修订信息公开条例,在控制数据安全风险的条件下最大程度开放政府数据,支撑大数据行业的进一步发展。

2. 提高准入门槛,推动产业结合,降低大数据行业过度竞争风险

大数据行业运营不规范的原因不仅在

于多重数据风险,还在于过度竞争风险。行业准入门槛低导致大数据行业鱼龙混杂,部分大数据企业资质低下,在数据收集、处理、分析等运营环节的专业性不强,不但阻碍行业的稳健发展,更于整体经济运行不利。首先,政府应明确大数据行业准入门槛,对大数据企业的资质进行审核,提高行业整体运营水平。其次,为缓解大数据行业的过度竞争风险,政府还应推动产业结合,挖掘大数据行业的潜在应用价值,创新行业发展,促进产业链的形成,利用大数据产业实现整体经济转型。具体而言,政府在引导行业发展时可促进大数据行业以科研力量为基础、以政府政策为支撑、以用户需求为导向的运营模式。

(三)行业层面

1.以人才驱动行业发展

数据本身没有价值,真正创造行业价值的是大数据技术与大数据人才,而当前大数据人才紧缺,在一定程度上也成为行业不规范运营的诱因之一。大数据行业的特殊性要求其专业人员具备较高的IT技术与数据分析技术,以实现大数据产业与其他产业之间的有机结合,为其他行业提供精准的商业决策建议,因此大数据企业应将人才战略纳入企业发展规划之中。首先,企业与高校合作培养大数据人才,共同建立校外科研基地,充分利用高校的培养模式选拔人才,包括大数据工程师、分析师、规划师等。其次,提高企业内部大数据人才待遇,并为其提供专业技能培训机会,吸收复合型人才。再次,组建大数据人才团队并形成行业协作的大数据人才建设模式。

2.以区块链技术弥补大数据市场漏洞

当前,由于大数据技术不完善或市场机制不健全等因素引发一系列风险,正对大数据行业的深入发展构成威胁,行业局限性逐渐凸显,但与此同时,日趋成熟的区块

链技术可以为大数据行业的发展难题提供解决对策。区块链技术是几大基础技术的有机结合,包括分布式技术、密码学、共识机制与时间戳,该种结合模式赋予了区块链技术“去中心化”与“去信任化”特征,为解决当前数据孤岛问题的主要突破口,同时也避免了数据所有权的错配问题,而区块链技术特有的加密及时间戳技术也使数据的真实性得到有效保证。总体而言,区块链技术能在最大程度上为大数据行业解决数据风险问题,因此大数据行业应正视区块链技术,将其与大数据技术相结合,以区块链技术为大数据技术弥补漏洞,引导行业未来发展趋势。

3.以行业协会促进大数据企业协调发展

大数据行业的规范发展需要企业的稳健运营、政府政策的有力支撑以及行业协会的引导。在当前大数据行业乱象频发,而法律政策处于相对滞后的背景下,大数据行业协会应充当沟通双方的桥梁。一方面,从行业主体层面对大数据企业的运营活动进行指导。第一,规范大数据企业数据收集、处理等业务活动的开展,降低数据安全风险对社会群体及大数据行业的侵害;第二,组织会员开展培训与交流互动,促进数据共享与行业协作,降低过度竞争对行业造成的负面影响。另一方面,大数据行业协会应与有关政府部门保持信息畅通,利用自身的特殊地位为政府制定政策提供依据,并积极宣传法律法规,以提高大数据行业自律性。

注释:

①数据来源:中国产业信息网.2016年中国大数据行业发展趋势及市场规模预计[EB/OL].<http://www.chyxx.com/industry/201607/431178.html>.

②数据来源:百度百家.人才大数据报告:大数据行业将面临“全球人才荒”[EB/OL].<http://datayuan.baijia.baidu.com/article/547043>.

(下转第19页)

注释:

①资料来源:中国人民银行《2016年金融机构贷款投向统计报告》。

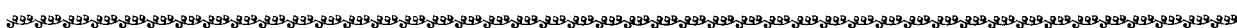
参考文献:

[1]张哲.供应链金融融资产品定价模式研究[D].天津:天津财经大学,2011.
[2]巴曙松,游春.我国小微型企业贷款保证保险相关问题研究[J].经济问题,2015(1):1-6.
[3]郭娜.政府?市场?谁更有效——中小企业融资难解决机制有效性研究[J].金融研究,2013(3):194-206.
[4]张占斌.中国经济新常态的趋势性特征及政策取向[J].国家行政学院学报,2015(1):15-20.
[5]袁成杰.经济新常态下中小企业融资效率问题研究[J].河北金融,2015(2):25-26.
[6]张伟斌,刘可.供应链金融发展能降低中小企业融资约束吗?——基于中小上市公司的实证分析[J].经济科学,

2012(3):108-118.

[7]李建军,胡风云.中国中小企业融资结构、融资成本与影子信贷市场发展[J].宏观经济研究,2013(5):7-11.
[8]徐学锋,夏建新.关于我国供应链金融创新发展的若干问题[J].上海金融,2010(3):23-26.
[9]胡跃飞,黄少卿.供应链金融:背景,创新与概念界定[J].金融研究,2009(8):194-205.
[10]王欢.供应链金融主要的融资模式探析[J].商场现代化,2010(11):11.
[11]李向文,冯茹梅.物流与供应链金融[M].北京:北京大学出版社,2012.
[12]汤曙光,任光标.银行供应链金融——中小企业信贷的理论、模式与实践[M].北京:中国财政经济出版社,2010.

责任编辑:董跃进 尚会妍



(上接第11页)

参考文献:

[1]李学龙,龚海刚.大数据系统综述[J].中国科学,2015(1):1-44.
[2]姜锋.大数据行业应用和商业模式研究[J].软件科学与工程,2014(4):20-28.
[3]袁冰.大数据行业应用现状与发展趋势分析[J].中国新通信,2014(24):75-76.
[4]蒋洁.大数据轮动的隐私风险与规制措施[J].情报科学,2014(6):18-23.
[5]巴曙松,侯畅,唐时达.大数据风控的现状、问题及优化路径[J].金融理论与实践,2016(2):23-26.
[6]孟小峰,慈祥.大数据管理:概念、技术与挑战[J].计算机研究与发展,2013(1):146-169.
[7]林志刚,彭波.大数据管理的现实匹配、多重挑战及趋势判断[J].改革,2013(8):15-23.
[8]陈火全.大数据背景下数据治理的网络安全策略[J].

宏观及研究,2015(8):76-84.

[9]王兵,崇阳,杨彬.大数据时代的大数据管理研究[J].无线互联科技,2016(8):130-131.
[10]吕天文.中国大数据行业分析与展望[J].电源世界,2014(2):4-6.
[11]杭天竹.大数据环境下的企业信息系统内部控制风险探析[J].中国管理信息化,2016(17):59-62.
[12]邱东.大数据时代对统计学的挑战[J].统计研究,2014(1):16-22.
[13]徐鹏,王以宁,刘艳华,张海.大数据视角分析学习变革——美国《通过教育数据挖掘和学习分析促进教与学》报告解读及启示[J].远程教育杂志,2013(6):11-17.
[14]刘自强.大数据的寒冬已至,谁降到下,谁成巨人[EB/OL].(2017-03-23)[2017-05-01].http://www.36dsj.com/archives/78408.

责任编辑:王春艳 尚会妍