

我国大数据征信的挑战及对策

李友元¹, 寇纲²

1. 西南财经大学金融创新期刊, 四川 成都 611130;
2. 西南财经大学大数据研究中心, 四川 成都 611130

摘要

大数据征信能够满足缺乏或没有信贷记录人群的金融需求, 是传统征信的有益补充。发展好大数据征信对于实现普惠金融, 促进我国社会信用体系建设有着重要的意义。阐述了美国和中国大数据征信的发展现状, 总结了大数据征信的特点, 探讨了我国大数据征信存在的不足以及面临的挑战。最后提出相应的对策及建议。

关键词

大数据技术 ; 征信体系 ; 信用评分

中图分类号 : F831.2

文献标识码 : A

doi: 10.11959/j.issn.2096-0271.2017004

Challenge and countermeasure of big data credit reporting in China

LI Youyuan¹, KOU Gang²

1. Journal of Financial Innovation, Southwestern University of Finance and Economics, Chengdu 611130, China
2. Research Center of Big Data, Southwestern University of Finance and Economics, Chengdu 611130, China

Abstract

Big data credit reporting could meet the financial needs of the people who have little or no credit records, and is a useful complement to the traditional credit. The development of big data credit reporting is of great significance to the inclusive finance and the construction of social credit system in China. The current situation of big data credit reporting in the United States and China were described and the characteristics of big data credit reporting were summarized. Also, the shortcomings and challenges of big data credit reporting in China were discussed. Finally, solutions and polices were also proposed.

Key words

big data technology, credit system, credit scoring

1 引言

近年来,随着互联网和信息行业的快速发展,数据呈爆炸式增长。根据国际数据公司(IDC)的研究,全球数据量大约每两年翻一番。大数据时代已经到来,引起了我国政府的高度重视。2015年9月,国务院出台了《促进大数据发展行动纲要》,加快大数据战略部署,深化大数据应用。我国各地方政府也陆续出台了大数据发展规划。政府的高度重视推动了我国大数据产业的快速发展。2014年中国大数据产业规模约为1 038亿元,2015年产业整体规模达到1 692亿元,增长了63%^[1]。大数据产业的高速发展,特别是大数据技术在金融产业的成功应用,使得大数据征信得到广泛的关注和重视。大数据征信从其本质上来讲,是将大数据技术应用到征信活动中,突出强调的是处理数据的数量大、刻画信用的维度广、信用状况的动态呈现、交互性等特点^[2]。

目前,我国信用信息数据库主要是中国人民银行征信系统。中国人民银行发布的《征信系统建设运行报告(2004—2014)》数据显示,截至2014年底,企业和个人征信系统分别为1 969 万户企业及其他组织、8.57亿自然人建立了信用档案;其中,有中征码的企业和其他组织1 008.7万户,有信贷记录的自然人3.5亿多人。传统信用评估模型是根据征信主体过去的信贷记录来评估信用情况。而中国人民银行征信系统有5亿多自然人没有信贷记录,未被中国人民银行征信系统覆盖的人数高达65%。如何满足这些没有征信记录的人的金融需求是亟需解决的难题,需要探索新的征信模式。这给大数据征信带来了市场机会。大数据征信摆脱了对信贷

数据的高度依赖。大数据征信采用了新的信用评估体系,人群覆盖面非常广,只要个体有注册登记、开立银行账户、纳税等活动,就可对他们在网络上留下的痕迹进行数据挖掘和分析,可作为征信体系的有效补充^[3]。因此,研究大数据征信、推动大数据征信发展,对促进我国社会信用体系建设、推动信用经济快速发展有重要的意义。

2 国内外大数据征信发展现状

2.1 美国大数据征信发展现状

美国是最早将大数据应用于征信的国家,目前已经走完疯狂发展、监管成熟、合并整合3个过程,已经基本进入了由几家龙头企业主宰的成熟期阶段^[4]。美国大数据战略实施比较早,为大数据征信发展积淀了雄厚的技术基础。早在2012年3月,美国就发布了《大数据研究和发展计划》,从国家战略层面部署大数据产业。在美国政府的大力推动下,美国大数据已经形成了完整的产业布局 and 成熟的法律框架。其次,大数据征信能够满足没有信贷记录、无法进行信用评分和信用分数低的人群的金融需求,弥补了传统信用评估的不足,有强烈的市场需求。因此,美国大数据征信发展迅速。美国许多互联网公司在大数据征信领域进行了成功的探索。ZestFinance是其中的代表性企业。

ZestFinance是美国大数据征信的代表性企业,它成立于2009年,旨在利用大数据技术为没有信贷记录或者信用分数低的人群提供信贷产品和服务。ZestFinance的数据来源非常广泛,既包括传统的信贷记录等具有金融属性的结构化数据,也包括法律记录、交易信息、电

子商务、社交信息等非传统的数据和非结构化的数据。ZestFinance运用机器学习和集成学习模型建立信用评估模型,能在5 s内对近1万条原始数据、超过7万个变量进行分析,将跟信用没有直接关联的变量整合在一起,形成能够反映征信主体信用特征的相关变量,最终获得有用的信用信息,进而对征信主体进行信用评估。目前,ZestFinance开发的模型已经达到了14个,涵盖信贷审批、助学贷款、市场营销的多个方面。与传统信贷管理业务比较,ZestFinance 的处理效率提高了将近90%;风险控制方面,ZestFinance 的模型相比于传统信用评估模型性能提高了40%^[5]。

此外,美国一些资深的传统征信公司也开始发展大数据征信。益百利(Experian)公司开发出跨渠道身份识别引擎,布局投入研发社交关系数据,积极探索互联网大数据与征信的关系;艾克飞(Equifax)公司通过加大研发投入及收购行为布局大数据产品与服务;费埃哲(FICO)公司在传统FICO信用模型中引入社交媒体、电商、移动用户数据,提高了模型的用户区分度^[2]。

2.2 中国大数据征信发展现状

中国大数据征信尚处于起步阶段。我国政府高度重视社会信用体系建设和大数据发展,陆续出台了《社会信用体系建设规划纲要(2014—2020年)》《促进大数据发展行动纲要》等一系列国家政策文件,为大数据征信的发展创造了良好的政策环境。2013年,我国互联网金融蓬勃发展,特别是P2P网贷平台的快速发展,对缺失或没有信贷记录的人群进行征信的迫切需要促进了大数据征信的发展。闪银(Wecash)等互联网信用评估平台趁势而

起。闪银是由北京闪银奇异科技有限公司开发的互联网信用评估平台,成立于2013年,依托数据挖掘分析和机器学习技术实现快速精准的信用评估。利用大数据技术,闪银已经能够做到3 min完成信用评估。2015年1月5日,中国人民银行印发《关于做好个人征信业务准备工作的通知》,要求芝麻信用、腾讯征信、前海征信、鹏元征信等8家机构做好个人征信业务的准备工作。个人征信市场化由此开始,出现了一大批大数据征信服务机构。芝麻信用、前海征信、考拉征信等大数据征信机构得到快速发展。百度和京东也投资了美国ZestFinance公司,涉足大数据征信。其中以芝麻信用和前海征信比较具有代表性。

芝麻信用成立于2015年,是阿里巴巴集团旗下子公司。芝麻信用的数据来源广泛,包括电子商务数据、互联网金融交易数据、合作机构数据以及用户自己提供的信用数据。其运用大数据及云计算技术客观呈现个人的信用状况,并采用“芝麻信用分”的形式为消费者提供个人信用评分服务。芝麻信用还开通了企业信用信息查询窗口,搭建了芝麻信用商家服务平台,为商户提供信用产品和服务。目前芝麻信用已经将“芝麻信用分”应用在租车、租房、酒店住宿等生活场景;还与蚂蚁微贷旗下“花呗”“借呗”、招联金融旗下“好期贷”开展合作,提供个人贷款信用风险评估服务。

前海征信成立于2013年,是中国平安保险集团旗下全资子公司。依托中国平安保险集团,前海征信可获得大量高质量的银行、保险等金融数据;前海征信还通过合作机构获取手机通信、公安、法院、社保、学历学籍、交通违章、房产、地理位置、租房、租车、婚恋、社交网站、消费等数据;利用大数据技术深入挖掘金融和

互联网多维数据,为顾客提供信用产品和服务。前海征信针对个人信用评分开发了信用评分产品“好信度”,分数范围为300~850分,分数越高信用越好。前海征信为企业客户征信开发了“好信常贷客”,基于好信金融信息服务平台的数据共享,覆盖千万级别的客户信贷记录,能识别多头借贷行为。目前,前海征信主要为银行、P2P、小贷、消费金融等金融机构提供大数据征信产品和服务。

总结国内外的情况,大数据征信机构主要有4个特点。首先,大数据征信机构定位很明确,即满足没有信贷记录或信用评分较低的人群和小微企业的金融需求。其次,大数据征信机构掌握着核心的数据资源,而且数据来源广泛,既包含政府公共数据、金融经济数据等传统数据,也有互联网交易数据、房租等非传统数据。第三,与传统征信模式相比,大数据征信需要处理海量的数据,数据来源广泛,技术门槛很高,因此,大数据征信主要由大型互联网公司、新兴金融科技或资深传统征信机构探索发起,这些机构技术实力雄厚,具有比较成熟的大数据征信技术解决方案。第四,相比传统征信机构,大数据征信应用场景更加多样化。大数据征信机构利用大数据技术可以进行风险识别、身份认证、信贷审核、反欺诈以及个性化服务等。

与美国大数据征信机构相比,我国大数据征信机构起步晚,技术积累相对薄弱,但是发展快,在应用场景方面甚至超过了美国大数据征信机构,例如芝麻信用分的应用场景非常丰富,涵盖金融、住宿、出行等多个生活场景。但是我国大数据征信机构也存在一些不足。首先,我国大数据征信机构营销推广方式过于突破传统应用场景,可能会丧失征信的严肃性。其次,我国试点的8家大数据征信

机构大多不是独立的第三方机构,数据来源于母公司,而母公司旗下的其他子公司又涉及放贷业务,“既当裁判员又当运动员”,难以保证征信评估的独立性。第三,我国大数据征信机构数据主要是互联网平台数据以及网上抓取的其他数据,数据来源多且复杂,数据质量得不到保障,且缺少金融信贷数据作为支撑,数据采集的维度不完整,难以全面反映客户的真实信用情况,也会影响到大数据征信模型的有效性。第四,国外大数据的法律框架比较成熟,大数据征信机构归集和使用数据比较规范,能够兼顾客户的主体权益。由于我国相关法律法规不健全、监管缺失,国内大数据征信机构往往忽视了客户的主体权益。

3 我国大数据征信面临的主要挑战

大数据技术的快速发展和个人征信业务的放开,使得大数据征信服务机构蓬勃发展,成为推动我国征信领域新业态生成发展的新生力量。但是,我国大数据征信的发展面临着以下挑战。

3.1 政策和监管存在不确定性

为了规范征信体系建设,我国陆续出台了《征信业管理条例》《征信机构管理办法》《征信机构信息安全规范》《征信机构监管指引》《征信业务管理办法(草稿)》等法律法规。大数据征信作为新兴的征信模式,是传统征信的有益补充,也要受这些法律法规的约束。但是,大数据征信涉及多方面的问题,征信的内容、技术手段、数据特征和分析判断的评估方式、评估模型、主要内容与方向,甚至主要结论都会发生根本改变^[6],已经超出上述征信法律

法规体系约束的范围。而且我国政府对发展和监管大数据征信的态度也不明朗。早在2015年1月,中国人民银行就要求芝麻信用、腾讯征信、前海征信等8家民营机构做好个人征信业务的准备工作。但是截至2016年12月,8家机构仍未获中国人民银行的个人征信牌照。政府政策不明朗和监管的缺失给个人征信特别是大数据征信的发展带来很大的不确定性。

3.2 个人隐私保护和信息安全存在问题

互联网大数据深刻影响征信业的发展和变革,社会征信机构大量涌现,对征信信息数据真实性和个人信息保护提出了更高的要求,也对征信监管提出了更高要求^①。数据是征信业务的核心。大数据征信的核心也是归集和处理数据。从大数据征信的数据获取途径看,主要有自有平台累积数据、通过交易获取数据、通过技术手段获取数据、用户自己提交数据、基于综合分析获得数据等^[2]。大数据征信在获取数据过程中,信用数据来源复杂多样,且存在多重交易和多方接入的情况,隐私数据保护的边界不清晰,隐私泄露风险被迅速放大。使用信息技术获取信用信息隐蔽性强,而且大数据征信对数据的挖掘利用使得事实认定难度很大,公民维护自己合法权益面临取证难、诉讼难等问题。美国学者布鲁斯·施耐尔认为,风险防范“分为道德压力、名誉压力、制度压力以及防护机制”^[7]。大数据环境下,个人数据应用的隐私保护是一个复杂的社会问题,不仅涉及道德、法律、行业、技术等诸多领域,也涉及大量的个人、群体、企业和机构^[8]。因此,大数据征信隐私保护和信息安全问题不仅是技术问题,也是社会性问题,需要从法律、政策、技术、行业规则等方面多措并举,实行综合治理。

3.3 信用信息没有实现共享

目前,我国大数据征信机构大多拥有核心的信用数据资源,如蚂蚁征信主要掌握阿里巴巴电子商务平台数据和蚂蚁金服互联网交易数据,腾讯征信主要掌握QQ和微信社交平台数据以及微信支付等交易数据。每家互联网公司都把自己掌握的大数据作为企业的资源,它们是不会和社会分享的,比如微信里的信息腾讯就不可能和阿里巴巴共享^[9]。中国人民银行征信中心掌握着大量的金融信用信息数据,我国政府部门也积累了很多公共信用数据,但是各民营征信机构很难获取这些数据。存在信息孤岛、数据壁垒未打通、信用信息共享没有畅通渠道是我国当前大数据征信乃至整个征信行业面临的现实问题。

① <http://finance.caixin.com/20151027/100867081.html>

3.4 大数据征信模型的可靠性有待验证

由于我国社会信用体系尚不完善,信用信息共享渠道不畅通,我国大数据征信机构难以获得中国人民银行征信中心的金融信贷数据,主要采用社交网络、电子商务等与个人信用关联度较弱的互联网数据构建征信模型。互联网数据来源广、维度多,但是在数据质量和权威性方面有所欠缺。美国ZestFinance并不是完全摆脱传统征信体系,在ZestFinance进行信用评估时,传统征信数据要占到至少30%^[10]。因此,缺少传统征信数据支撑,仅仅依靠互联网数据来构建模型评估信用状况,其评估结果的准确性和科学性有待验证。其次,我国大数据征信机构掌握的核心数据资源不同,因而在构建大数据征信模型时采用的数据维度和算法都不一样,这导致同一个人在不同征信机构得到的信用评

分可能会不一样,大数据征信模型的公信力和可比性容易遭到质疑。最后,与传统征信相比,大数据征信模型要处理多达几千个甚至上万个变量,而且面临不同的应用场景可能需要开发不同的模型,这使得模型更加复杂,降低了评估模型的可解释性。

4 发展我国大数据征信的建议

大数据为征信发展带来机遇,基于海量、多样的数据,征信机构可以获得信用主体全方位的信息,使征信在数据来源、存储和处理方式、提供产品和服务等方面面临巨大改变^[11]。为了促进我国大数据征信健康有序发展,实现我国普惠金融,推动征信业市场化进程,促进我国社会信用体系建设,针对我国大数据征信的现状存在的主要挑战,提出以下4点建议。

4.1 加快征信立法,建立完善大数据征信的配套制度和监管机制

一是尽快出台个人信息保护法,针对大数据征信的特点,在法律法规中明确规范大数据征集中个人信息采集范围、使用授权、信息主体的权利和义务等,严格限定信用信息的采集和使用目的,防止信用信息被滥用;推进互联网实名制、数字证书等安全认证服务,加强对个人信息的保护,做好隐私保护与数据开放共享之间的平衡。二是加强信用数据信息安全体系建设,加强信用信息数据库的安全管理,防范非法入侵,泄露信用主体的信息;对大数据征信机构开展内部安全认证和评价,严厉打击非法获取、买卖信用数据的行为。三是探索符合大

数据征信特点的监管机制,实行行业准入制度,对信息保护不力的机构进行问责和处罚。四是建立完善大数据征信异议处理机制,保护信息主体的权益。五是完善信用行业协会组织,发挥行业协会自律作用,促进大数据征信行业的健康发展。

4.2 以政府为主导,建立健全信用信息共享机制

一是政府管理部门要加强顶层设计,通过征询大数据征信龙头企业意见,制定科学合理的征信大数据行业标准。二是加强政府大数据治理,加快政府公共信用数据接入中国人民银行征信系统的步伐,同时,将符合标准的大数据征信机构的信用数据融入中国人民银行征信系统,打破信息壁垒,形成畅通的信用信息共享渠道。三是探索建立大数据征信统一信用信息平台,以“市场主导,政府引导”为原则,鼓励大数据征信机构间打破信息孤岛,推动信用信息共享,最终建立能与中国人民银行征信系统存在映射关系的大数据征信统一信用信息平台。

4.3 以市场驱动为主,鼓励大数据征信机构技术创新

大数据征信对数据、资金、技术都有很高的要求。大数据征信机构只有积极进行技术创新,探索出适合中国国情的大数据征信模式和技术手段,才能更好地进行数据采集、整理和分析,开发出更符合市场需求的产品,从而更好地服务市场。政府应坚持以市场为导向,鼓励大数据征信机构采用新技术、开发新的信用产品和服务,创造良好的技术创新氛围。

4.4 加强人才培养,为大数据征信发展提供智力支持

目前,我国大数据征信人才比较紧缺。大数据征信是大数据技术在征信领域的应用,大数据征信需要既懂技术,又懂征信业务的复合型人才。政府应支持我国高等院校设置大数据征信相关专业,鼓励培养大数据征信相关的管理和技术人才;支持高校人才进行大数据征信等相关课题研究,充实大数据征信相关专家人才队伍;鼓励企业开展大数据征信人才培养。

5 结束语

大数据技术的迅速发展给征信行业带来新的发展契机。传统信用评分是根据信贷记录、公共记录等传统数据来评估个人或企业的信用状况,而无法对缺乏或无信贷记录的人群进行信用评估,导致这部分人群的金融需求无法得到满足。大数据征信能利用互联网数据等非传统金融数据对信息主体进行信用评估,是传统信用评分的有益补充。但是,大数据征信兴起时间短,所占信用市场份额很少,其信用评估的真实效果还有待检验,并不能完全替代传统征信系统。本文分析了美国和中国大数据征信的发展现状以及代表性大数据征信机构,总结了大数据征信的4个特点,结合美国大数据征信的情况,探讨了我国大数据征信存在的不足以及面临的主要挑战。最后对发展我国大数据征信提出了有针对性的政策建议。

参考文献:

[1] 贵阳大数据交易所. 2016年中国大数据交易

产业白皮书[R]. 贵阳: 贵阳大数据交易所, 2016: 18.

Global Big Data Exchange. White paper on China big data exchange 2016[R]. Global Big Data Exchange, 2016: 18.

[2] 孔德超. 大数据征信初探——基于个人征信视角[J]. 现代管理科学, 2016(4): 39-41.

KONG D C. A look at big data credit reporting: based on the perspective of personal credit[J]. Modern Management Science, 2016(4): 39-41.

[3] 植凤寅. 大数据征信与小微金融服务[J]. 中国金融, 2014(24): 90-93.

ZHI F Y. Big data credit reporting and microfinance[J]. China Finance, 2014(24): 90-93.

[4] 九次方大数据研究院. 美国大数据征信的应用及启示[J]. 国际融资, 2016(5): 14-17.

USFOUN BIGDATA. Enlightenment on the application of big data credit reporting in the USA[J]. International Financing, 2016(5): 14-17.

[5] 范铁光, 刘岩松. 大数据应用于信用评分模型的实践与启示[J]. 征信, 2015(2): 29-31.

FAN T G, LIU Y S. Practice of big data used in the scoring model and enlightenment[J]. Credit Reference, 2015(2): 29-31.

[6] 吴晶妹. 2015展望:网络征信发展元年[J]. 征信, 2014(12): 1-4.

WU J M. Outlook 2015: the first year for online credit reference development[J]. Credit Reference, 2014(12): 1-4.

[7] 布鲁斯·施奈尔. 我们的信任[M]. 徐小天, 译. 北京: 机械工业出版社, 2013: 78.

SCHNEIER B. Liars and outliers[M]. Translated by XU X T. Beijing: China Machine Press, 2013: 78.

[8] 袁文秀, 余恒鑫. 关于网络信息生态的若干思考[J]. 情报科学, 2005, 23(1): 144-147.

YUAN W X, YU H X. Thinking about information ecology in network space[J]. Information Science, 2005, 23(1): 144-147.

[9] 韩锋. 从大数据到区块链的迁移[J]. 金融博览, 2016(3): 16-17.

HAN F. From big data to block chain[J].

Financial View, 2016(3): 16-17.

- [10] 刘新海, 丁伟大. 大数据征信应用与启示——以美国互联网金融公司ZestFinance为例[J]. 清华金融评论, 2014(10): 93-98.

LIU X H, DING W D. Enlightenment on the application of big data credit reporting: a case study of ZestFinance[J]. Tsinghua

Financial Review, 2014(10): 93-98.

- [11] 中国人民银行征信管理局. 现代征信学[M]. 北京: 中国金融出版社, 2015: 1.

Credit Information System Bureau of the People's Bank of China. Principles of modern credit[M]. Beijing: China Financial Publishing House, 2015: 1.

作者简介



李友元 (1986-), 男, 西南财经大学金融创新期刊编辑部助理编辑, 主要研究方向为多目标决策、信用评分、互联网金融。



寇纲 (1975-), 男, 博士, 西南财经大学工商管理学院教授、博士生导师、执行院长, 长江学者特聘教授, 主要研究方向为大数据与决策、信息管理、多目标决策与优化算法、信用评分等。已出版英文专著2部, 共发表学术论文100余篇, 其中被SCI收录61篇, SSCI收录23篇。入选首届国家自然科学基金优秀青年科学基金项目、四川省百人计划青年项目、教育部新世纪优秀人才支持计划、四川省学术和技术带头人后备人选、四川省有突出贡献的优秀专家等人才计划。先后获得2013年中国管理学青年奖、第十二届四川省青年科技奖、2016年度高等学校科学研究优秀成果奖(科学技术)自然科学奖一等奖等奖项; 入选Web of Science基本科学指标(Essential Science Indicators)的全球高被引科学家(highly cited scientists)。

收稿时间: 2016-12-13