

大数据时代腐败防治机制创新研究

■ 理学博士,四川省社会科学院党委书记、教授 李后强 ■ 工学博士,四川省社会科学院金融与财贸经济研究所副教授 李贤彬

【摘要】:腐败作为一种社会现象,随着人类社会的存在而存在。腐败与反腐败是一个问题的两个方面,随着社会的进步呈现交替发展的趋势。传统的反腐败策略主要关注事后惩处,而对事前预防关注不多。随着大数据时代的来临,腐败防治乃至腐败免疫需要创新体制机制。本文在综述国内外大数据时代反腐败相关研究文献的基础上,构建了基于国内实际的大数据腐败防治机制创新体系,运用大数据系统有指导学习、聚类分析、莫比乌斯分析、泊松分布、量子跃迁、自组织临界理论(SOC理论)、扩展限制凝聚(DLA模型)、动态免疫(DI)等创新理论体系对惩治腐败、预防腐败、腐败免疫进行了系统的分析,给出了束缚权力的大数据笼子设计方案。

一、引言

腐败就是公权私用,是全社会的毒瘤,不会仅仅存在于哪一个国家或地区,全球所有国家或地区或多或少都存在着腐败现象,承受着不同层面的腐败之痛。有些国家腐败治理有效,当前腐败的严重程度较轻,有些国家腐败现象仍大行其道。基于此全球范围内成立了众多的组织机构和多边框架协议致力于防治腐败,我国也于近期加入了国际反腐败学院(IA-CA)并在 APEC 北京会议期间通过了《北京反腐败宣言》,以此彰显中国加强反腐败的决心。腐败破坏社会公平正义,损害政府形象和公信力,阻碍经济健康发展,是必须治理的“毒瘤”。这是 2014 年 11 月 9 日闭幕的亚太经合组织第 26 届部长会议审议通过的《北京反腐败宣言》中关于腐败问题的描述,该宣言决定成立 APEC 反腐败执法合作网络,在亚太地区加大追逃追赃合作,携手打击跨境腐败行为。根据相关学者的研究估计,美国的年犯罪成本在 1 万亿美元左右,腐败的年成本占当期 GDP 的 2% 到 5%,体量达到每年 8000 亿美元到 2 万亿美元之间,9·11 以来美国用于反恐的花费已达 1 万亿美元。(1)

基于腐败是社会“毒瘤”的表述,借助唐代孙思邈《备急千金要方·诊候》对医术的描述“上医医国,中医医人,下医医病。上医听声,中医察色,下医诊脉。上医医未病之病,中医医欲病之病,下医医已病之病。”与此相对应,腐败防治的最低层级为惩治腐败,中间层级为预防腐败,最高层级为腐败免疫,对应于国家反腐败工作的三个目标:不敢腐、不能腐、不想腐。习近平总书记于十八届中共中央政治局第一次集中学习时讲话提到:“党风廉政建设,是广大干部群众始终关注的重大政治问题,委更求更加科学有效地防治腐败”。中共中央纪委也提出要切实做好反腐倡廉网络舆情收集研判、应对处置和引导工作。

党的十八届四中全会《决定》中强调“加快推进反腐败国家立法,完善惩治和预防腐败体系,形成不敢腐、不能腐、不想腐的有效机制,坚决遏制和预防腐败现象。”党的十八届四中全会《决定》学习辅导百问中强调“健全反腐败领导体制和工作机制,强化反腐败体制机制创新和制度保障”。(2)因此,反腐败工作需要新的理念、新的系统、新的技术、新的手段来惩治、防治腐败乃至实现腐败免疫。

近日,国家决定成立新的反贪局——最高人民检察院反贪污贿赂总局,以应对腐败发展的新趋势:系统化、区域化、群体化、家族化。而制约当前反腐败工作的核心问题之一就是装备与理念的落后,即是腐败治理水平的落后,因此腐败治理水平的现代化亦是国家治理水平现代化中的重要一环。

新的反贪污贿赂局将整合力量、优化职能,从有利于最高人民检察院集中精力查办大要案案,有利于强化对下业务工作的集中统一领导与指导,有利于破除制约办案工作发展的体制机制障碍出发,来成立新的反贪局。职能配置将更加科学,办案力量将进一步增强,领导班子将进一步配强。建设目标是:更强大战斗力、威慑力、公信力、灵敏高效的中国特色反腐败专门机构。

近几年在社会科学及实务领域广受关注的大数据就为腐败治理提供了新的理念、系统、技术与手段。基于大数据的强大功能,在腐败防治的三个层次均能利用大数据相关原理与系统加以应对,惩治腐败需要基于对腐败问题的识别,可以运用大数据有指导学习、聚类分析、结果与表象的反演理论来实现;预防腐败基于对腐败路径的分析,可以运用大数据的预测功能、相关性分析功能加以实现,分析腐败行为的泊松分布、腐败过程的级别跃迁;腐败免疫基于体制机制设计,可以运用大数据遗传算法加以基因改造和前置干预来实现免疫,以自组织临界理论(SOC)突破集体腐败案件,以扩散限制凝聚模型(DLA)剖析腐败集团形成机理。以资源、权力、腐败三角系统演化理论为基础设计约束权力的大数据笼子。

二、文献综述

(一)反腐败与大数据应用的国外文献

2005 年联合国毒品和犯罪问题办公室发布了反腐败实践的调查与展望手册,该手册给出了腐败的定义、腐败检测等国际合作框架。(3)

2012 年意大利全国研究理事会的认知科学与技术研究所的 Giulia Bonelli 等人在 ERCIM-NEWS 发表的文章中提到了大数据用于犯罪与腐败调查,(4) 该文提到欧洲项目 FuturICT 推出的《犯罪与腐败观察》将利用大数据理论来面对挑战并运用新的方法来分析与理解犯罪与腐败的新现象。

2013 年经合组织、联合国毒品和犯罪问



题办公室、世界银行联合发布了《企业反腐败的道德与合规手册》,(5)《企业反腐败的道德与合规手册——实践指南》两个文件,(6)对国家反腐败法律框架、企业腐败风险评估、企业反腐败道德与合规体系的建构、相关案例实践等内容进行了十分深入的阐述。

美国伊利诺斯大学芝加哥分校政治科学系 Thomas J. Gradel 等于 2009 年到 2013 年间连续发布了 7 篇关于伊利诺斯州反腐败的调查报告,(7)(8)(9)(10)(11)(12)(13) 分别从消除腐败、腐败深度、库克郡的腐败、赞助及在芝加哥的政府机构任用亲信和犯罪行为、芝加哥和伊利诺斯的腐败领导、绿草与移植:在郊区的腐败、芝加哥警察局犯罪、腐败与掩饰等多个角度对伊利诺斯州的腐败问题进行了深入的调查。

2013 年 UN ECLAC 的 Martin Hilbert 博士发布了题为“Big Data for Development: From information to Knowledge Societies”的报告,(14) 该报告从区域发展的视角分析了大数据发展的前提条件、机遇与挑战,指出大数据带来的几个影响,其一是社会发展从“信息时代”向“知识时代”进化,其二是大数据产生的“数字鸿沟”致使区域发展不平衡,其三是需谨记大数据并非“万应灵药”,其中该文给出了政府数据公开程度、人均国民总收入与腐败感知指数之间的相关关系。

2014 年纽约大学法学院信息法协会的 Ira S. Rubinstein 教授发表的论文“Voter Privacy in the Age of Big Data”中研究了美国政府选举中政治捐赠信息的披露与防治腐败的关系,(15)分析了信息披露与免责申明手段在预防腐败中的作用。

(二)大数据与腐败防治国内文献

改革开放后有关腐败问题的研究始于 1984 年,1991 年之前处于腐败问题研究的萌芽期,1992 年到 2001 年是腐败问题研究的发展期,2002 年到目前是腐败问题研究的成熟期。2013 年胡洪彬对中国改革开放后三十年的腐败与反腐败研究运用文献计量学的视角进行了综述,(16) 该文指出国内学者对腐败与反腐败进行了持续的研究,为国家反腐倡廉建设提供了一定的智力支持,但还存在诸多不足:研究层次较低、研究内容范围较窄;比较研究与跨学科研究有待提升;缺乏科学扎实的社会调研及实证分析等。

党的十八以后,腐败与反腐败问题的学术研究逐渐走向深入,相关的研究文献逐渐面世。2013 年到 2014 年初的文献主要关注网络反腐,2014 年以来大数据在国家治理与反腐败等层面的应用逐渐展开。

2013 年李桂锋等就网络反腐热进行了理性的思考,(17) 分析了网络反腐在党风廉政建设有效监督中的非理性现状,提出了规范网络反腐的对策。2013 年胡丘陵的研究指出减少现金流通是获取大数据的最佳路径,(18) 货币数据化后将使得公共权力中的腐败概率降低。2013 年陈潭等就网络反腐的限度及其优化进行了研究,(19)分析了网络反腐的特点、网络反腐的限度、网络反腐的优化。

2014 年谢金林基于 2004—2013 年 646 例网络反腐案例对网络反腐的发展趋势、特点与对策进行了研究。(20)2014 年潘楚雄就大数据背景下海关联理能力的现代化进行了初步研究,(21)指出应该提升海关在数据生成、节点控制、数据扩展、数据发声、数据开放等方面的能力建设。2014 年陈宇洁就网络反腐与遏制网络谣言的良性互动进行了研究。(22)2014 年刘叶婷等分析了大数据时代对政府治理的影响,指出了大数据时代政府治理面临的挑

战。(23)2014 年赵淑梅对大数据在提高党的建设科学化水平方面的作用进行了分析,(24)特别提到通过大数据,可以在相对较短的时间内发现腐败问题,提高反腐败斗争的实效性。2014 年潘芳等就基于超网络的微博反腐舆情进行了研究,(25) 分析了反腐舆情传播网络与社会网络之间的相互关系,重点研究了政府在微博反腐舆情传播中所处的作用。2014 年张亚明等就大数据作为网络反腐的新利器展开了论述,(26) 分析了大数据网络反腐的现实匹配、大数据网络反腐面临的多重挑战、大数据网络反腐的路径选择。2014 年申孟宜等就大数据时代的政府监管进行了阐述,(27) 分析了大数据的预警能力、基于大数据的网络反腐。2014 年冯燕红基于信号传递博弈模型对官员腐败的经济学分析与对策进行了研究,(28)指出在分离贝叶斯均衡范围内建立财产公示制度、加强监督、提高查处力度以及高薪养廉对策能配合模型系统地提高反腐效果。

2014 年李后强等对大数据时代社会科学研究面临的机遇与挑战进行了较全面的综述,(29)提出了研究对象的可感知性、研究工具的透视创新、研究思路的全域性、研究结果的实战性为社会科学研究的跨越式发展带来了巨大机遇,但社会科学研究领域也必须正视大数据意识缺失、社科大数据资源不足、社科数据处理能力或缺等多方面挑战,面对大数据时代的机遇与挑战,社会科学研究领域需采纳大数据行动方案加以应对:引导并倡导大数据语境、构建全域社科大数据资源、社科大数据处理能力建设。该研究成果为大数据在当前腐败防治机制创新研究领域的应用提供了理论基础。

2014 年李后强等对大数据时代社会科学研究面临的机遇与挑战进行了较全面的综述,(29)提出了研究对象的可感知性、研究工具的透视创新、研究思路的全域性、研究结果的实战性为社会科学研究的跨越式发展带来了巨大机遇,但社会科学研究领域也必须正视大数据意识缺失、社科大数据资源不足、社科数据处理能力或缺等多方面挑战,面对大数据时代的机遇与挑战,社会科学研究领域需采纳大数据行动方案加以应对:引导并倡导大数据语境、构建全域社科大数据资源、社科大数据处理能力建设。该研究成果为大数据在当前腐败防治机制创新研究领域的应用提供了理论基础。

三、腐败防治与大数据的功用

腐败是一种社会现象,广泛存在于社会生活的方方面面。由于移动互联网、社交网络、云计算、物联网、万物互联等技术的广泛应用与人类交往方式的现代化,大数据的社会透视功能日益强大,过去藏身于幕后或台下的腐败行为在大数据时代变得无处遁形。大数据技术的发展与运用正可以体现在腐败治理工作的机制创新应用之中。

腐败防治的三个境界分别是:惩治、预防、免疫。近两年备受关注的网络反腐是网民通过网络技术和数据信息对官员的腐败行为进行检举,利用社会的舆论效应引起相关部门对官员的行为进行监察和惩处,形成对执政行为的有力监督和权力的约束,使得反腐工作能够更有效、更透明,进而达到有效地预防、遏制、惩戒腐败行为的目的。总体看来,网络反腐还局限于惩治腐败的层面,预防腐败与腐败免疫单靠网络的力量还略显单薄,因此预防腐败与腐败免疫的重任就落在了大数据相关理论与技术应用的肩上了。

(一)腐败的定义与核心要素

关于腐败、反腐、腐败防治等相关术语的定义从不同侧面有不同的定义表述,本文认为美国学者格莱泽关于腐败的定义更贴合当前中国的实际。腐败:一个运转良好的政府体系丧失其功能甚至滥用其职能的朽化过程。希腊历史学家波里比阿指出,君主政体会腐败为专制政体,贵族政体会腐败为寡头政体,民主政体会腐败为暴民统治。(30)

腐败的三个核心要素:(1)在工资之外存在向政府官员的支付;(2)存在与这种支付相关的违反明文法律或潜在社会规范的行为;(3)上述行为给公众造成了损失或由于超工资支付引发的其他行为给公众造成了损失。腐败的度量方法:其一是向政府官员支付

的金钱数;其二是违反相关法律法规的频率;其三是腐败的社会成本。

腐败的决定因素包括腐败行为的收益与成本。腐败的收益取决于政府官员牟利的能力,成本则取决于行为暴露后的预期惩罚。政府官员牟利的手段包括:慷国家之慨,出于自身利益的考虑将政府资产转移给私人,操纵执法或法律规制。制约腐败的因素有三个:法律惩罚、职业或社会成本、精神折磨。腐败的总成本取决于各种惩罚的力度及其可能性,这种可能性是信息流、社会谴责和法律制度的函数。

(二)大数据在腐败防治领域的潜在应用

国家制定的反腐败工作的战略目标是:必须完善权力运行制约和监督体系,抓紧形成“不敢腐、不能腐、不想腐”的有效机制。坚持有腐必惩、有贪必肃,使领导干部“不敢腐”;加强体制机制创新和制度建设,强化监督管理,严肃执纪,完善激励和问责机制,使领导干部“不能腐”;通过加强理想信念教育,增强党性修养和宗旨意识,牢固树立中国特色社会主义道路自信、理论自信、制度自信,使领导干部“不想腐”。

具体的措施包括:完善惩治和预防腐败、防控廉政风险、防止利益冲突、领导干部报告个人有关事项、任职回避等方面的法律法规建设与运行实施。

目前我国的反腐败工作主要还局限于个案反腐、权力反腐、网络反腐(民间)的层级,还需要利用大数据思维与方法形成与构建制度反腐、系统反腐、国际反腐、反腐免疫的反腐败工作理想模式。

四、腐败防治机制创新前瞻

对腐败现象或腐败行为的研究需基于腐败发生、发展的全过程展开,研究方向包括:腐败行为的认定与识别,腐败严重程度的聚类,腐败表象与内核的正反面莫比乌斯分析,腐败过程的动态反演,腐败行为发生的概率分析,腐败量级的量子跃迁,集体腐败案发机制,腐败集团形成机理,腐败免疫机制设计等方面。

(一)腐败行为的识别——大数据系统有指导学习

对腐败行为的界定存在着四种基本模式:以公共权力或公共职位为中心的政治学界定,以市场为中心的经济学区界定,以法律为中心的法学界定,以社会规范为中心的社会学界定。总体来看,这四种基本模式依然属于定性的判断,随着大数据时代的来临,对于围绕在涉嫌腐败的官员周围的大数据集可以采用定量的方法加以相对科学合理并更加精确的定量界定。

对涉嫌腐败行为大数据集可以采取有指导的学习方法加以处理,有指导的学习是指:基于含有输入和输出的训练集,建立由输入变量估计输出变量的模型,并给出模型相关参数的计算算法。(31)

根据目标变量是定性的还是定量的,有指导的学习通常分为分类和回归。系统输出结果为存在腐败行为和不存在腐败行为这两种定性认定,因此腐败行为的界定属于分类的范畴。

分类是机器学习中的典型问题,是通过构成不同概念的特征进行分析,归纳出能够用于区分不同概念的判别规则,从而可以有效地应用规则将新的观测对象分配到事先制定的类别中,也称为概念学习。分类是揭示事物本质的基本途径,常见的是 0-1 学习,即目标变量只取两个值的情况。

(二)腐败严重程度的分类——大数据系统聚类分析

分析中央纪律检查委员会对腐败官员腐败严重程度的定性表述方式可以发现,违纪、严重违法、违纪违法、严重违法违纪等四种分类是最常运用的分类标准,但目前尚无对这四种分类标准的学术定义。因此对腐败官员腐败严重程度的分类即是一种无指导的学习范例,本着对历史负责、对当事人负责、对社会大众负责的科学治理目标,需要对腐败程度分类给出创新的分析方法。大数据处理技术中的聚类分析即是一种十分匹配的聚类方法。

聚类分析也称无教师学习或无指导学习,与分类学习相比,聚类的样本一般事先不做标记,需要由聚类学习算法自动确定。聚类分析是在没有训练目标的情况下将样本划分为若干簇的方法。(32)

(三)腐败表象与内核的正反面莫比乌斯分析

腐败官员的表象与内核完全不同,就像普通纸带一样可以分别涂上红黑两面,外表是红的,内里是黑的。而反腐倡廉的目标是让官员里外一致,即只有一个面——单侧曲面,防治腐败的目标是使拥有公权力的官员表象与内核呈现莫比乌斯带特征。

1858 年德国数学家莫比乌斯(Mobius, 1790-1868)和约翰·李斯丁发现:把一根纸条扭转 180°后,两头再粘接起来做成的纸带圈,具

有魔术般的性质。普通纸带具有两个面(即双侧曲面),一个正面、一个反面,两个面可以涂成不同的颜色;而这样的纸带只有一个面(即单侧曲面),一只小虫可以爬遍整个曲面而不必跨过它的边缘。这种纸带被称为“莫比乌斯带”,可以用参数方程式创造莫比乌斯带(图 1)。



图 1 莫比乌斯带

这个方程组可以创造一个边长为 1 半径为 1 的莫比乌斯带,所处位置为 x-y 面,中心为(0,0,0)。参数 u 在 v 从一个边移动到另一边的时候环绕整个带子。莫比乌斯带的参数方程如下:

$$x(u,v) = (1 + \frac{v}{2} \cos \frac{u}{2}) \cos(u)$$

$$y(u,v) = (1 + \frac{v}{2} \cos \frac{u}{2}) \sin(u)$$

$$z(u,v) = \frac{v}{2} \sin \frac{u}{2}$$

$$(0 \leq u < 2\pi, -1 \leq v \leq 1)$$

(四)腐败结果或表象的反演定律研究

现实社会中腐败行为并非显而易见的,类似于人类的身体疾病一样,医生能够看到的是病员身上表现出来的病理特征,医务工作者需要利用 CT、X 光等反演工具来诊断疾病的类型和严重程度。

涉嫌腐败的官员日常生活中会呈现很多异常的特征:财富的奇异增长、消费水平的显著波动、社会交往复杂度的变化等等现象,因此可以运用反演原理基于涉嫌官员的异常行为特征反推腐败行为发生的可能性与腐败的严重程度。

在物理上,反演把空间所有力方向同时反过来,因此每个坐标可用它自己的负值代替。反演是理论物理中研究微观粒子对称性的一种方法。物体在原坐标系和反演后的坐标系中各运动规律之间的关系,相当于物体和它在镜子中所成像之间的关系。时间反演即指把时间的流向倒转。

反演在数学(某些几何证明)上有很重要的作用。把(1,+∞)“放入”(0,1),可以用取倒数的方法,这是一维上的反演。二维上的反演是以一个特定的反演圆为基础来实现的:圆心 O 为反演中心,圆半径为常数 k,把点 P 反演为点 P' 就是使得 OP×OP'=k²(即 k 为 OP 和 OP' 的几何平均)。

(五)腐败行为分布——泊松分布

泊松分布是概率论中常用的一种离散型概率分布。若随机变量 X 只取非负整数值 0,1,2,…… 且其概率分布服从

$$P(X=i) = \frac{e^{-\lambda} \lambda^i}{i!}$$

则随机变量 X 的分布称为泊松分布,记作 P(λ)。这个分布是 S-D.泊松在研究二项分布的渐近公式时提出来的。泊松分布 P(λ)中只有一个参数 λ,它既是泊松分布的均值,也是泊松分布的方差。在实际事例中,当一个随机事件,例如某电话交换台收到的呼叫,来到某公共汽车站的乘客,某放射性物质发射出的粒子,显微镜下某区域中的白血球等等,以固定的平均瞬时速率 λ(或称密度)随机且独立地出现时,那么这个事件在单位时间(面积或体积)内出现的次数或个数就近似地服从泊松分布。因此泊松分布在管理科学、运筹学以及自然科学的某些问题中都占有重要的地位。

一个单位或一个行业内的腐败行为在初期呈现随机且相互独立的特征,因此在一定的时期内一个单位或一个行业的腐败行为案发现量服从泊松分布。然后随着时间的流逝,一些单位或一些行业内由于决策者之间存在千丝万缕的联系,腐败窝案层出不穷,这样腐败行为案发现量就不再服从泊松分布了。

(六)腐败过程识别——级别跃迁

通过对涉嫌腐败的官员的腐败过程分析可以发现,腐败行为呈现出一定的级别跃迁,开始是小恩小惠,进而大额受贿或贪腐,最后往往以巨额受贿贪腐等形式而引起案发。