

大数据 人力资源管理

中国人事科学研究院 组织编写

● 王通讯 主编



中国人事出版社

大数据人力资源管理

中国人事科学研究院 组织编写
王通讯 主编



中国人事出版社

图书在版编目(CIP)数据

大数据人力资源管理/王通讯主编;中国人事科学研究院组织编写. —北京:中国人事出版社, 2016

ISBN 978-7-5129-1096-6

I. ①大… II. ①王… ②中… III. ①人力资源管理-研究 IV. ①F240

中国版本图书馆 CIP 数据核字(2016)第 100114 号

中国人事出版社出版发行

(北京市惠新东街 1 号 邮政编码: 100029)

*

北京市艺辉印刷有限公司印刷装订 新华书店经销

787 毫米×1092 毫米 16 开本 23.25 印张 406 千字

2016 年 5 月第 1 版 2016 年 5 月第 1 次印刷

定价: 38.00 元

读者服务部电话: (010) 64929211/64921644/84626437

营销部电话: (010) 64961894

出版社网址: <http://www.class.com.cn>

版权专有 侵权必究

如有印装差错,请与本社联系调换:(010) 50948191

我社将与版权执法机关配合,大力打击盗印、销售和使用盗版图书活动,敬请广大读者协助举报,经查实将给予举报者奖励。

举报电话: (010) 64954652

编委会

主 编 王通讯

副主编 周建华 张我伟 卢金海 徐鹏举 唐魁玉

编委会成员（按姓氏笔画排序）

王通讯	王 青	王必聪	王潇雨	田英杰	卢金海
尧 川	孙 红	吕晓庆	吴 艳	辛旭涛	刘 漪
刘 冬	刘 奇	刘志铭	李天宝	李竞伟	张宇泉
张宏僖	张我伟	张玉洲	和 云	周建华	周韬宇
赵 明	徐 起	徐鹏举	胡万明	郝 丽	唐魁玉
唐儒星	袁 哲	高健明	高烁琳	常春阳	韩丽霞
曾 辉	廖祖荣	颜世富	樊宝瑞	薛力君	

前言：人力资源管理，要让数据发声

马克思说，“一种科学，只有在成功地运用数学时，才算达到了真正完善的地步”（参见保尔·拉法格《忆马克思》）。大数据时代的到来，为人事科学进一步完善提供了有利条件。自从2013年我国大数据元年开始，我们就致力于大数据的研究，力求运用大数据的技术促进人事科学包括人事管理和人力资源管理的转型升级，并促使其进一步科学化、个性化、精准化和智能化。

习近平总书记在第二届世界互联网大会开幕式发表主旨演讲时，进一步强调，“十三五”时期，中国将大力实施网络强国战略、国家大数据战略。国务院颁布的《关于印发促进大数据发展行动纲要的通知》（以下简称《通知》）中指出，在当前，大数据已经成为推动经济转型发展的新动力，成为重塑国家竞争优势的新机遇、提升政府治理能力的新途径。《通知》还要求我们建立用数据说话、用数据决策、用数据管理、用数据创新的新的管理机制。2016年3月，中共中央印发的《关于深化人才发展体制机制改革的意见》中指出：“充分运用云计算和大数据等技术，为用人主体和人才提供高效便捷服务。”在新形势之下，我们唯一的选择就是要紧跟形势，顺势而为，尽快使人力资源工作插上大数据的翅膀，实现跨越式发展。毫无疑问，这是一项艰巨而光荣的任务。我们认为，大数据带给我们的将是一场深刻的革命：人力资源管理之发展，总体上将从“经验加感觉型”走向“数据加事实型”。

回顾新中国成立后人力资源管理的发展历程，人事管理在我国经历了相当长的一个历史阶段。从20世纪80年代之后，才有人力资源管理学说的兴起。可是，理论归理论，实践归实践。时至今日，不少单位还是沿用着老旧的一套。从21世纪开始，“信息世纪”“信息时

代”的说法风行起来。按照最时髦的说法，现今社会已经进入信息社会的大数据时代。但是，社会实际实行的一套做法，仍然承载着人事管理的厚重印记，距离现代人力资源管理，尚有很大差距。就在我们踟蹰而行、难以实施变革的时候，大数据飓风迎面吹来，呼啸而至，甚至使人来不及思考，这到底是怎么一回事？

旧有的、我们已经习惯了的人事管理，实际上乃是工业革命的产物。它是与以物为中心的管理思想相适应的，其最典型的表现就是高耸而起的科层制。在这种组织结构之中繁忙劳作的人们，很难将自己的能动性、创造性充分发挥出来。人力资源学说强调了人力可以开发、升值的一面，这是应该充分肯定的，但是在实际操作层面，往往存在着不对称、不及时、不精确的弊端。不对称是指领导机关掌握的情况与客观事实对应不起来；不及时是指领导机关掌握的数据不是即时的，而是落后的，凭借过时的数据怎么决策？不精确是指大而化之，不能做到精准。诸如此类的弊端，严重地困扰着人力资源与社会保障工作的科学运行，也影响着人力资源管理的进一步优化。

有学者认为，所有这一切，概而观之是我们的管理还没有走出凭经验、凭印象的窠臼。换言之，人力资源管理还没有学会用数据说话，凭数据决策。这是下一步学科发展与科学决策顺利前行的不二法门、必经之路。

令人兴奋的是，大数据能够帮助人们在一定程度上做到这一点。

对于什么叫“凭数据决策”，还是给大家讲个葡萄酒定价的故事。一瓶名字叫“拉菲”的葡萄酒，动辄几千人民币，这是为什么？它值那么多钱吗？这个高的吓人的价格是通过一种怎样的机制或者规则确定下来的？为了弄清这样一个问题，我们先来看看曾经发生在美国的一件真事。

奥利是供职于美国普林斯顿大学的一位经济学教授。他多年来担任美国最顶尖经济学期刊《美国经济评论》的主编，学术水平不在话下。他一方面对经济研究非常痴心，另一方面却对葡萄酒品评饶有兴趣。而其兴趣的核心是：如何通过“数据”来评估波尔多葡萄酒的质

量，而非凭“品酒专家”的“品咂与吐掉”来对酒的等级说三道四。

他的思路是：酒是一种农产品，肯定会受到每年气候的重大影响。他采集了法国波尔多地区的气候数据，精细研究，发现如果收获季节多旱少雨，而且整个夏季平均气温较高，则该年产的葡萄就容易生产出高质量的酒。他还研究出了统计方程，做了验证。他指出，当葡萄熟透，液汁高度浓缩时，酒就是最好的；哪年夏天温度如果特别高，葡萄容易熟透，酸度就会降低；碰到炎热少雨的年份，葡萄汁也会浓缩。总之，天气越热、越干燥，那一年就容易生产出上乘的好酒。他还说，熟透的葡萄能够生产出低敏度的口感柔润的酒，液汁高浓缩的葡萄能够生产出味道醇厚的酒。

奥利的的方法是，把每年的气候数据代入自己的公式，自然就预测出任意一种酒的平均品质。照此，他曾经精确预测出 100 多个酒庄所生产的葡萄酒的品质。法国人依照他的方法，还给葡萄酒庄园排出 1 855 个等级呢。

对此，依靠“吧嗒嘴儿”而过着优裕生活的品酒师们不高兴了。这家伙不是要砸自己的饭碗吗？他们不屑一顾地说，奥利的办法“介于极端与滑稽之间”。他的那个公式也是“无法让人重视”的。因此，无论奥利在什么场合演讲、推销他这一套方法时，品酒师们无不报以一片嘘声。

一个叫帕克的作家，正在办着一份《葡萄酒爱好者》杂志，他开口大骂奥利是个“彻头彻尾的大骗子”。他讽刺奥利“就像某些影评一样，根据演员和导演来告诉你电影有多好，实际上却没有去过电影院”。他的意思很明白，应该去亲自品尝一下葡萄酒，才能够对酒的品质发表意见。但是问题是，在好几个月的时间里，人们是无法品尝到葡萄酒的。波尔多的葡萄酒，在装瓶之前需要盛放在橡木酒桶里发酵 18~24 个月，品酒专家需要在酒装在桶里 4 个月后进行第一次品尝，在这之前，葡萄酒只是发臭发酵的葡萄而已。一位著名拍卖行酒品部经理的说法正确：发酵初期的酒变化很快，不可能有人通过品尝来准确地评估酒的好坏。而奥利却能够凭借数据，得出气候与酒价的

关系，指出冬季降雨量每增加1毫米，酒价会提高0.00117美元。而且奥利能够通过对数据的分析，在葡萄刚刚收获的时候，就能预测出未来酒的品质，比品酒师亲口品尝还要提前几个月！我们可以想一想，在葡萄酒期货贸易空前活跃的今天，奥利的办法能够给葡萄酒收藏者多么巨大的帮助啊！

正因为如此，20世纪90年代美国《纽约时报》头版头条刊登出了奥利对1989年波尔多葡萄酒的预测数据，从而使更多的人了解了奥利的思想。现在，那些传统的品酒师们也开始更多地关注天气了。

伦敦克里斯蒂拍卖行酒品部主席说：“很多人认为奥利是个怪人。我也认为他在很多方面的确很怪。但是我发现，他的思想和工作会在多年后留下光辉的痕迹。他所做的努力对于打算买酒的人来说，非常有帮助。”

与奥利的葡萄酒故事有异曲同工之妙的是美国棒球界詹姆斯的故事。詹姆斯在美国办着一份名为《棒球摘要》的刊物。和奥利一样，这个人也相信数据，不大赞同通过观察球员比赛来判断球员能力的做法。而后一种做法是传统的，也是多数人认可的。

詹姆斯认为，凭着肉眼观察，不足以让人了解到评价一个球员的标准。想想吧，谁能够凭观察看到上垒率0.300的与0.275的加强击球手之间的差别呢？如果你一年观察15场比赛，那么后者就有40%可能比前者击出更多安打。谁是好的击球手，只能从他们的记录上才能看出来。詹姆斯不仅有这种认识，而且研究出了一个自己的公式：

$$\text{垒得分} = (\text{安打垒} + \text{保送上垒}) \times \text{总垒数} / (\text{打数} + \text{保送上垒})$$

可以看出，公式更加强调球员在垒的比重，尤其给那些保送上垒更频繁的球员更高的评级。这种做法让球探们非常不高兴。

球探们认为，要想找到一位优秀球员，必须开着车走上6000多公里，住上百次的汽车旅馆；经历4个多月，看完200场高中生或大学生的棒球比赛。特别是，你必须走进球场，在铝制栏杆后面第四排找到一个位子，以便能够看到别人看不到的某些细节。而且，这种机会只有一次。

这就好像那些葡萄酒品尝师只相信自己的味蕾一样，这些球探们只相信自己的眼睛。

2002年，奥克兰运动队的经理比利·比恩因为了解和认同詹姆斯的那一套找人寻才做法，决定通过数据来选择棒球球员。他比较喜欢一个叫杰里米的大学生，原因是他走动更积极。可是，旁边的球探们并不喜欢他，原因是这个小伙很胖。球探们还觉得，身材笨重的人是不可能的球队里打主力的。可是比利·比恩并不这么看。而且认为应该选上这个人。后来的故事是：杰里米比那一年进队的所有球员进步都快。4年后成为主力队员。在詹姆斯的带领启发下，美国棒球研究协会已经成立，他们还有了一套用于“棒球数据分析”的专用词汇。有兴趣的朋友可以去看这本书——《赛伯计量学》。

这个故事，又不同于前一个故事。它把我们从对葡萄酒的预测评价，拉回到对体育人才的预测评价上来了。而这正切合我们本书要讨论的主题。

这两个故事对我们人力资源管理者有什么启迪意义呢？我认为起码有这样三条：

一是大数据方法能够量化管理对象。无论是人还是其他事物，只有通过量化管理对象，才能使管理走向科学。

二是大数据的量化思路可能与我们以往的量化思路不大一致。以往的量化思路太窄，大数据能够使我们脑洞大开、猛开窍。

三是大数据管理人力资源的方法，并不排斥其他方法的使用，但是大数据是更有说服力的方法，应该努力掌握。

记得有位很有影响的历史学家说过，中国人往往对数据不够重视。历史上不少皇帝拍板决策，大多缺少数据依据。比如对旱涝灾情的估计、对前线战事的估计，多是靠眼睛、靠耳朵来估摸。如果前来汇报、上奏的大臣叫得响、哭声大，则会更加重视。这样的决策之策，是很难达到科学合理的，所以往往导致误判误事，甚至酿成大祸。摆脱长期以来历史阴影的影响，在当今仍然十分必要。

大数据不仅在微观决策层面能够发挥重要作用，在宏观决策方面

照样可以。1993年，美国劳工部的专家拉里·卡茨提出一个建议：可否把政府提供就业岗位技能培训的资金拿出一部分，用在提供关于怎样申请新工作和面试上。拉里提出：多花些钱给失业者做一些就业指导，就可以缩短申请失业保险的时间。也就是说，把钱花在一个新的培训计划上，可以为国家节省20亿美元。试验证明，得到就业指导的失业者比得到技能培训指导的失业者找到新工作的时间缩短了一个星期。最好的案例是在得到指导力度最大的明尼苏达州，该项目的效果最好，找到新工作的时间缩短了4周。而且找到的工作与未参与者后来找到的工资相当。拉里采用的大数据方法主要是回归分析和随机试验。

在具有明显说服力的事实面前，原来持反对意见的国会领导人也失去了反驳机会。政府最终认为，在这位真正的专家帮助下，社会失业保险金支出下降了，再就业加快又带来了税收的增加。政府在就业指导项目上每投入1美元，就能节省大约2美元开支。像这样的事情，就凸显了社会科学作为科学的力量。

大数据作为信息社会的一项新技术，其诞生之意义我们尚未能够充分认识。大数据发展对于整个人类社会所产生的影响我们尚未充分估计。随着历史的前进，包括人力资源管理者在内的所有人士对这场变革意义的认识，还会不断加深。相信再过若干年，再回首这场颠覆性变革，人们将会欣然一笑：

两岸猿声啼不住，轻舟已过万重山！



2016年5月

目 录

上篇：大数据人力资源管理概论

第一章 大数据及其应用功能	(1)
第一节 大数据及其特点	(1)
第二节 大数据的出现	(3)
第三节 大数据的价值取决于什么	(6)
第四节 大数据在管理领域能够干什么	(8)
第五节 大数据思维	(12)
第二章 大数据与人力资源管理的关系	(15)
第一节 新世纪人力资源管理面临新形势	(15)
第二节 基于大数据的人力资源管理	(18)
第三节 加快大数据行动，关键是要做起来	(32)
第三章 基于大数据的人力资源规划	(36)
第一节 人力资源规划的含义与方法	(36)
第二节 人力资源规划的常见缺陷与问题	(42)
第三节 大数据改进人力资源规划	(49)
第四章 基于大数据的人力资源招聘	(53)
第一节 人力资源招聘的含义与方法	(54)
第二节 人力资源招聘的问题与短板	(61)
第三节 大数据改进人力资源招聘	(64)
第五章 基于大数据的人力资源使用	(69)
第一节 人力资源使用的界定与宗旨	(70)
第二节 人力资源使用中的问题	(75)

第三节	大数据改进人力资源使用	(79)
第六章	基于大数据的人力资源考核	(83)
第一节	人力资源考核的含义与功能	(83)
第二节	人力资源考核的问题讨论	(89)
第三节	大数据改进人力资源考核	(93)
第七章	基于大数据的人力资源薪酬	(98)
第一节	人力资源薪酬的含义与功能	(98)
第二节	人力资源薪酬的问题所在	(102)
第三节	大数据改进人力资源薪酬	(105)
第八章	基于大数据的人力资源培训	(112)
第一节	人力资源培训的含义与功能	(112)
第二节	人力资源培训的问题所在	(117)
第三节	大数据改进人力资源培训	(121)
第九章	基于大数据的人才评价	(127)
第一节	人才评价的定义与功能	(127)
第二节	人才评价的问题分析	(132)
第三节	大数据如何升级人才评价	(137)
第十章	基于大数据的人力资源管理创新	(145)
第一节	人力资源管理概述	(145)
第二节	人力资源管理的问题所在	(149)
第三节	大数据改进人力资源管理	(152)

下篇：大数据人力资源管理技术

第十一章	大数据技术概论	(160)
第一节	什么是大数据	(160)
第二节	大数据的时代	(167)
第三节	从数据到大数据	(174)

第十二章 大数据数据集成	(181)
第一节 数据集成概述	(181)
第二节 大数据集成的必要性	(189)
第三节 大数据集成的应用	(195)
第十三章 大数据清洗过滤	(202)
第一节 数据清洗概论	(202)
第二节 数据清洗方法	(206)
第三节 数据清洗工具与实战	(212)
第十四章 大数据合并压缩	(227)
第一节 数据压缩技术	(228)
第二节 数据无损压缩	(232)
第三节 数据有损压缩	(240)
第十五章 大数据数据安全	(244)
第一节 大数据安全面临的主要问题	(245)
第二节 大数据安全保障技术	(247)
第三节 数据安全技术	(251)
第四节 大数据安全趋势	(257)
第十六章 大数据知识发现	(262)
第一节 知识发现的定义与过程	(262)
第二节 知识发现的对象与分类	(267)
第三节 知识发现的方法与应用	(273)
第十七章 大数据可视呈现	(279)
第一节 数据统计与分析	(279)
第二节 数据可视化	(283)
第三节 示例展示	(293)
第十八章 大数据关联规则	(300)
第一节 关联规则的简介	(300)

第二节 关联规则挖掘	(302)
第三节 关联数据与大数据	(316)
第十九章 大数据分类聚类	(323)
第一节 概述	(323)
第二节 聚类分析算法	(328)
第三节 分类算法	(334)
第二十章 大数据应用决策	(341)
第一节 大数据应用现状	(341)
第二节 大数据应用具体案例	(346)
第三节 人力资源方面的具体应用和帮助决策	(351)
后记	(355)

上篇：大数据人力资源管理概论

第一章

大数据及其应用功能

第一节 大数据及其特点

大数据的定义

什么叫大数据？在中国最早出版的徐子沛的《大数据》一书说，大数据是那些大小已经超出了传统意义上的尺度，一般的软件工具难以捕捉、存储、管理和分析的数据。但是，多大的数据才能称得上“大”，并没有普遍适用的定义。他还说，一般认为，大数据的数量级应该是“太字节”的，也就是2的40次方的。也有专家说，实际上我们没有必要给出一个“大”的具体尺寸，因为随着技术的进步，尺寸会不断增大，而且由于研究领域的不同，“大”的界定也会不一样，没有必要统一。

说大数据是“数据集合”，很好理解；说大数据是“信息技术”，也好理解；说大数据是“服务业态”，就不大好理解了。但后者是最近国务院文件的说法。笔者认为，只有这么定义大数据，才能尽快将大数据的实际应用提到各行各业行动日程上来。

当我们讲到大数据的时候，实际上包含三层意思：一是数量很大，二是变化迅速，三是结构复杂。

大数据的重要意义不在于大，而是通过对大数据的搜集、保存、维护与共享，发现新知识，创造新价值，获得新利益，实现大发展。

关于数量大，人们会想到天文学里的描述：浩如星海。根据经济合作与发展组织（Organization for Economic Co-operation and Development, OECD）的计

算, 当今世界, 每秒发送的电子邮件总量高达 290 万件。谷歌公司每天储存数据就有 24 PB (1 PB 等于 2 的 50 次方字节), 这相当于美国国家图书馆所有纸质出版物所含数据量的上千倍。世界过去 3 年产生的数据量比以往 4 万年还要多。2010 年, 全球数据已达 1.2 亿 ZB (1 ZB 等于 2 的 70 次方字节), 而且以每年 50% 的速度增长。现在, 全球每天产生的数据足以填满 2 亿张 DVD 光盘。人们在不同领域采集到的数据之大, 已经达到前所未有的程度。

关于变化迅速, 可以把时下的信息洪流与公元 1439 年古登堡发明印刷机时的情况做一比较。那时欧洲的信息存量 50 年翻了一番, 今天的世界每 3 年就会翻一番。2003 年, 人类第一次破译人体基因密码, 辛辛苦苦工作了十年, 才完成 30 亿对碱基对的排序, 而现在 15 分钟即可完成。任何一个正在运行中的数据仓库, 分分秒秒都在不停地更新着它的数据, 逝者如流水, 变化总不停。

关于结构复杂, 是指大数据的来源多种多样, 数据形式多种多样。不仅有图书、报纸出版物上的, 还有银屏、视屏、电影、电视上的; 不仅有结构化的, 还有大量非结构化的; 不仅有来源于汽车导航、购物小票上的, 还有来源于各种感应器搜集到的。有人用“数据毛球”形象地描绘它, 倒也贴切。

(一) 大数据重视事物的关联性

大数据有一个重要特点, 就是“不讲为什么, 重视关联性”。如果发现了某种关联性, 就可以加以利用。凭借自有的卫星信息系统进行商品管理的沃尔玛公司, 发现在他们的卖场里, 有不少顾客在购买婴儿尿布的同时, 都要买上几罐罐装啤酒。这是为什么? 不知道。美国人讨论了好长时间也不知道。但是, 掌握这种关联性的卖场经理, 就可以告诉上架员, 要把罐装啤酒与婴儿尿布摆放在一起。这么做, 果然提升了这两种商品的销售量。

在人力资源流动方面, 国家发改委的研究人员发现了“榨菜指数”“方便面指数”, 就是它们的销售量与国内劳动力的流动、流量、流向高度重合。

(二) 大数据的价值重在挖掘

对于大数据, 不仅要搜集它, 更重要的是挖掘它。挖掘就是分析, 目的是从中寻找关系、重点、规律, 洞察其发展趋势。这就为管理者提供了莫大的帮助。凭借大数据, 管理者将可以大幅度提升各行各业管理水平、治理能力。有专家认为, 数据挖掘技术主要有关联分析、聚类分析、分类及预测。

前些年, 北京大学发生留学生食堂爆炸事件。肇事者搞了爆炸后迅速逃逸。北京市公安局利用大数据方法, 轻轻一抓, 就把他从茫茫人海中抓了出来。

(三) 大数据将颠覆诸多传统

社会科学研究常用的“抽样调查”，曾经被认为是社会文明得以建立的牢固基石。其实，它只是在技术手段受到限制的特定时期，解决特定问题的一种无奈方法。现在，已经可以收集到过去无法收集到的信息，保存与计算也不成问题，所以应该是“样本就等于全部”。而且这样做，比使用抽样调查方法得出的结论要准确得多。

有专家称，大数据将颠覆 13 个行业。互联网金融就是一个明显的案例。

由于大数据具有“数据充足”“抓取力强”“刷新及时”的特点，所以，在众多领域能够将人数据化、将岗位数据化、将资源数据化，最终能够增加产出。这样一来，自然能够引起管理者的高度重视。作为生产产品与提供服务的企业，通过记录、分析、挖掘这些数据，能够发现过去没有发现的问题与规律，从而达到提升人力资源管理效率与组织产出效率的目的。

试举一例：农业发展银行江苏分行紧跟时代潮流，积极探索人力资源管理网络化创新，通过建立网络平台，统筹网络学院，在线考试，员工四卡考核，实现了基于大数据的人力资源管理，使组织将过去“经验加感觉”的定性管理，走向了“事实加数据”的定量管理。其重大意义在于迈向了人力资源管理的智慧化。

第二节 大数据的出现

大数据是怎么出现的？

当今的世界，基本上一切都可以用数字表达，所以叫数字化世界、数字化生存。纸质上的数据只是很小的一部分。我们每天生产的电视电影、录音歌曲、手机拍摄的照片、卫星拍摄的图像，乃至个人计算机上记录的数据，加上数十亿计的感应器搜集到的信息，可以说达到了海量之大。

一、我们每个人都是数据的制造者

一个人打开电视机，走进电梯间，行驶在高速公路，以及下班路上到超市购物打出的小票单据，无不留下数据的足迹。早在很多年前，人们就开始对数据加以利用。例如，航空公司利用数据弄清楚了应该给机票如何确定价位，银