

“十三五”国家重点出版物出版规划项目

图解服务的细节

051

医療ビッグデータがもたらす社会変革

大数据时代的 医疗革命

医学是一门不确定性的科学和可能性的艺术

不放过每一个数据，不轻视每一个偶然

实现医疗大数据的终极目的——尽可能延长个人的健康寿命

日本医学各领域顶尖专家展望医疗大数据应用前景&全球医疗大数据先进应用案例



推荐读物

医疗行业从业人员、医护专业

人民东方出版传媒

People's Oriental Publishing & Media



东方出版社
The Oriental Press

图解服务的细节

051

医療ビッグデータがもたらす社会変革

大数据时代的 医疗革命

〔日〕21世纪医疗论坛编
中山健夫 监修
刘波 译

人民东方出版传媒

People's Oriental Publishing & Media



东方出版社

The Oriental Press

图书在版编目 (CIP) 数据

大数据时代的医疗革命 / 21 世纪医疗论坛 编; 刘波 译.

—北京: 东方出版社, 2016.9

(服务的细节; 051)

ISBN 978-7-5060-9242-5

I. ①大… II. ①刘… ②刘… III. ①数据处理—应用—医疗保健事业—研究

IV. ①R19-39

中国版本图书馆 CIP 数据核字 (2016) 第 230914 号

IRYO BIG DATA GA MOTARASU SHAKAI HENKAKU by Takeo Nakayama / 21 Seiki Iryo Forum.

Copyright © 2014 by Takeo Nakayama / 21 Seiki Iryo Forum. All rights reserved.

Originally published in Japan by Nikkei Medical Publishing, Inc.

Simplified Chinese translation rights arranged with Nikkei Medical Publishing, Inc. / Nikkei Business Publications, Inc. through Beijing Hanhe Culture Communication Co., Ltd.

本书中文简体字版权由北京汉和文化传播有限公司代理

中文简体字版专有权属东方出版社

著作权合同登记号 图字: 01-2016-6283

服务的细节 051: 大数据时代的医疗革命

(FUWU DE XIJIE 051:DASHUJU SHIDAI DE YILIAO GEMING)

编者: 21 世纪医疗论坛

译者: 刘波

责任编辑: 吴婕 王高婷

出版: 东方出版社

发行: 人民东方出版传媒有限公司

地址: 北京市东城区东四十条 113 号

邮政编码: 100007

印刷: 北京楠萍印刷有限公司

版次: 2016 年 10 月第 1 版

印次: 2016 年 10 月第 1 次印刷

印数: 1—6000 册

开本: 880 毫米 × 1230 毫米 1/32

印张: 8.125

字数: 169 千字

书号: ISBN 978-7-5060-9242-5

定价: 38.00 元

发行电话: (010) 85924663 85924644 85924641

版权所有, 违者必究 本书观点并不代表本社立场

如有印装质量问题, 请拨打电话: (010) 85924602 85924603

前 言

“大数据”已经成为一个热词。在今天的商业世界，对大数据的驾驭度事关未来成败。

即使在医疗领域，以 2011 年为分界点，与大数据有关的医学论文发表量持续增加。从论文内容来看，涉及“公共卫生（Public health）”的篇幅最多，其次是“卫生保健（Health care）”。即便如此，当问及大数据的现状时，还没有一个人能够给出准确的答案。

大数据主要有以下三方面的特点：Volume（容量）、Velocity（迅速性）、Variety（多样性）（野村综合研究所），但迄今为止大数据还没有一个明确的定义。现在所谓的流行语（buzzword），是指“大众使用的，看似冠冕堂皇，实则概念模糊，且晦涩难懂”的词汇的典型代表。

引用维基百科的定义，所谓大数据，是指“常规数据库管

理工具等难以处理的海量数据集合”，“大数据一般分为结构化数据和非结构化数据两种，以多样性、尺度以及要求的处理速度和高度，实现容纳和应用（检索 / 共享 / 分析 / 可视化等）为目标对象”。

假设大数据符合以上描述，那么，它们已经在医疗领域存在。比如日本于 2002 年导入的 DPC 制度（Diagnosis Procedure Combination，疾病诊断分组）。DPC 制度中，2011 年之前该数据库收集的数据量为 878 万件，2012 年覆盖日本国内 1 505 家医院，普通床位超过半数——约计 48 万张床位。2011 年日本政府明令所有诊疗机构必须完成诊疗明细电子化并纳入日本国家数据库（National Database），通过该项目运作，每年收集的诊疗明细数据可达 9.5 亿件。

在构建上述国家级医疗大数据的同时，个人大数据也相继问世。由 30 亿对碱基组成的卷帙浩繁的基因组序列分析宣告完成，现阶段进行的个人基因组序列分析就是其典型代表。

另一方面，个别时候原本与医疗毫不相关的大数据甚至能够带来新的医学发现。比如：通过对全球用户检索数据分析，基本能够对流行性感冒的流行趋势做出准确的预测。对过去几十亿件检索数据与实际流感流行趋势分析的比对结果证实——检索关键词与流行性感冒的流行趋势之间存在某种相关关系。

在这里，我们希望广大医疗从业者牢记一点。那就是在现阶段，大数据能够揭示某种现象与其他现象之间存在某种相关关系。通过揭示之前我们尚不了解的相关关系，可能会为人类做出巨大的贡献。

但是，在医学世界中，即使明确某种相关关系，我们也不能盲目跟风。因为在临床一线，一个决策就能左右一个人的生命。从始至终，数据仅仅是临床决策的辅助判断材料之一。如果一味关注表面的相关关系，忽视背后的因果关系，就可能犯下无法挽回的错误。

那么，医疗从业者应该怎样面对医疗大数据？这就是贯穿本书的关键问题。

当然，即使在医疗行业，大数据也会为人们提供有价值的知识。由笔者所在的京都大学研究院医学研究科与滋贺县长滨市联合发起的共同项目——以1万名项目参与者为对象展开队列研究。该项目涵盖环境及生活方式信息、临床体检信息、基因信息、组学信息（Omics）等各种类型的医疗大数据，目的是将这些数据用于医学研究的同时，致力于构建市民健康等工作。

本书除了介绍“日本长滨市零级预防队列研究项目”的情况，作为在日本本土实施的医疗大数据项目，还将阐述与日本外科领域专业医师制度合作的手术及治疗相关NCD数据——“日

本国家临床数据库(The National Clinical Database)”以及DPC疾病诊断分组的实践应用等情况。此外,在本书第4章中,作为海外医疗大数据应用的代表,本书将围绕艾美仕日本市场研究公司(IMS JAPAN 株式会社)进行,对以英美等国为代表的海外医疗大数据应用现状展开具体阐述。关于这一章的编写工作,在此对艾美仕日本市场研究公司提供的巨大帮助表示衷心的感谢。

此外,来自日本医学界、信息学界的各位专家学者针对日本国内医疗大数据的未来展望和应用等问题提出了大量宝贵的参考意见。最后,本书最后一章重点围绕今后大数据在医疗领域应用的重点——全民个人识别编号(My Number 制度)在医疗领域应用的重要性等进行总结。

本书重点围绕大数据在医疗领域应用的意义以及医疗大数据的价值等问题进行了阐述。如果本书能够成为医疗大数据在日本应用的开端,笔者将感到荣幸之至。

京都大学研究生院医学研究科

社会健康医学系健康信息学领域教授 中山健夫

2014年5月

目 录

前言__ 001

第1章 数据、信息、知识的意义

直面“大数据”__ 003

信息来自数据分析__ 009

数据分析的方向性和风险__ 013

信息的作用是减少不确定性__ 018

正确数据分析时的注意事项__ 023

知识是信息结构化的总体__ 029

第2章 什么是医疗大数据

从茫茫数据洪流中汲取什么__ 035

呈爆炸式增长的全球数据存储量__ 039

大数据的本质是“非结构化”的多样数据__ 044

大数据的特点——4V __ 047

检索数据对流行性感冒的预测__ 051

大数据是“不问因果关系”的预测__ 054

医疗大数据真的不需要因果关系吗__ 057

全量分析改变世界的观点__ 063

大数据实现的医疗__ 066

真正有意义的医疗大数据——全民个人识别编号的应用__ 072

第3章 医疗大数据引发的思考

DPC 是反映日本医疗机构实态基准点的工具__ 079

东京医科口腔科大学研究院医疗政策信息学领域教授 伏见清秀先生

通过临床数据与其他大数据的连接，从患者角度出发治疗疾病__ 091

东京大学研究院医学系研究科医疗质量评价学讲座教授 官田裕章先生

医疗大数据与临床流行病学开拓的新的医学研究世界__ 109

京都大学研究院医学研究科药剂流行病学教授 川上浩司先生

医疗大数据 + 基因组分析的启示__ 120

京都大学研究院医学研究科附属基因医学中心主任 松田文彦先生

第4章 国外医疗大数据应用案例

- 在医疗大数据应用中落伍的日本__ 135
- 数据库构建和民间应用不断扩大的海外案例__ 139
- 欧美各国的数据库整合现状__ 144
- 国外医疗大数据对患者和社会也发挥着重要作用__ 159
- 通过应用医疗大数据，合理控制医疗费用__ 165
- 从患者视角锁定更高质量的治疗方法——美国 CER __ 169
- 利用医疗大数据培养创新型人才的必要性__ 175

第5章 有识之士眼中的医疗大数据构建及应用

- 京都大学研究院医学研究科客座教授
本庶佑先生__ 183
- 庆应大学药学部医药开发规制科学讲座教授
黑川达夫先生__ 192
- 国立信息研究所所长、东京大学教授
喜连川优先生__ 204
- IMS JAPAN 株式会社代表取缔役社长
凑方彦先生__ 212

第6章 实现医疗大数据的价值最大化

医疗大数据带来的益处__ 223

整合中的医疗系统数据库__ 229

全民个人识别编号在医疗行业的应用__ 232

当务之急是培养“人类数据科学家”__ 236

第1章

数据、信息、
知识
的意义

直面“大数据”

“量变会转化为质变”。

曾屹立于国际象棋世界王座不倒的象棋大师卡斯帕罗夫（Garry Kasparov）在与 IBM 推出的超级计算机深蓝（Deep Blue）展开人机大战后，说了这样一番意味深长的话。

“至少在国际象棋的世界里，人类已经无法战胜拥有压倒性数据和计算能力的计算机！”

量变的确会转化为质变。

今天，随着大数据时代的到来，摆在我们眼前的课题也已经蜕变为量和质的问题。怎样正确驾驭这些以令人恐惧的势头持续增加的数据，又怎样将它们转化为高质量的信息？

毋庸置疑，各种数据的爆发式增加会为商业世界创造新的

机遇，也会为医学领域的学术研究带来巨大的福音。但是，数据同时也是一柄双刃剑，读取方式错误，可能造成致命的灾难性后果。

研究人员在面对数据时，应该从以下两方面的视角出发。

这就是“假设验证”和“假设生成（一说探索）”。其中，假设验证可以说是研究人员的基本态度，为验证自己提出的假设正确与否需要收集大量数据。当数据与假设不一致，还要对假设进行修改。通过反复的“假设与验证”程序，假设逐渐凝练，继而找到通向真理的道路。

那么，假设来自哪里？

首先，是研究人员的“限定范围内的观察区域”和“大脑”。其次，“先行研究（现存文献）”是迄今为止主要假设的生成场所。其样貌正逐步改变。从庞大的数据海洋中，或许在与人类前所未有的感知形式相左的过程中，某个新假设突然出现。日新月异的计算机分析技术，例如数据挖掘也许会承担假设萌芽的一个辅助角色。结果，某个研究人员提出的假设也许呈现出颠覆传统的奇葩内容。数据挖掘（Data mining），又译为资料探勘、数据采矿。它是数据库知识发现中的一个步骤。

以抢在竞争对手前面出招为制胜策略的商业世界为例，商业人惯常的思维模式是根据数据挖掘的结果生成某种假设，在

这个无论成败与否的假设阶段，首先果断出手，小试牛刀。

例如：市场营销界有一个著名的定论，即“来超市买啤酒的男性顾客很大概率也会购买婴儿尿布”。这是某连锁超市对庞大的历史销售数据进行研究后得出的结论。那么，据此类推，如果这家超市大胆尝试在饮料区的收银台附近摆放婴儿尿布会如何？其结果，即使尿布销量平平，对这家超市的销售业绩也无妨大碍。

但是，医疗的世界绝非如此。

为什么？假设大数据分析获得如下结果，证实：“人在食用某种特定食品后，会出现血压降低的情况。”当然，仅仅根据这个结果，没有哪个研究人员会得出“这种食品具有降压功效”的结论。但是，即便如此，从事人体血压波动研究的研究人员却不能轻易放过这项数据。因为从这种食品中所含的某种有效降压成分与血压之间的关系展开分析，他们可能得出某个新的假设。

对大数据进行分析，就能获得传统无法与之比拟的真知灼见，继而获得大量假设。

这其中既包括新派生的某种假设，也可能是一些“无聊或不值一提”的假设。可以说，随着大数据时代的到来，医疗从业人员最应拿出的态度，第一是“假设验证”，第二是“假设生

成”的平衡。

如前所述，假设验证是接近科学的根本。但是，凡事过犹不及，如果太钻牛角尖，就可能陷入诸如“执着于寻找对自己提出的假设有利的数据，仅仅将着眼点放在这类数据上”；或者“即使客观数据无法形成佐证，也固执地生搬硬套在自己的假设上……”这样危险的境地。

让人遗憾的是，这恐怕就是近年来在社会上掀起狂澜的“科学论文抄袭问题”之类事件的背景之一。另一方面，与假设验证相比，假设生成在某种程度上需要自由思考，即所谓的“独立假设”，这对科学家来说具有极大的吸引力，但同时必须远离个人主观或自以为是的根源陷阱，以一颗虚怀若谷的坦诚之心，倾听以事实为根据的客观数据。

接下来，对这种情况下浮现出来的某个征兆进行切实的提炼，将这个雾里看花的事物作为假设的萌芽进行耐心培育，在此基础上寻找相关数据。当假设以某种形式逐渐形成雏形，此时需要做的是从验证的视角与数据进行关联——这种假设生成看似具有无穷的吸引力，实则可能以某种完全背离传统知识积累的、无法解释的关系呈现出来。大数据时代的假设生成，其优点和注意点带有这种前所未有的鲜明特征。

本节将从医疗从业者面对大数据时的正确立场出发，按假