

TURING

Apress®

量化

大数据时代的企业管理

METRICS

How to Improve Key Business Results

深度挖掘数据背后的价值
积极应对大数据时代的挑战
亚马逊网站全五星好评

【美】Martin Klubeck 著
吴海星 译



人民邮电出版社
POSTS & TELECOM PRESS

TURING

量化

大数据时代的企业管理

METRICS

How to Improve Key Business Results

【美】Martin Klubeck 著
吴海星 译



人民邮电出版社

北京

图书在版编目(CIP)数据

量化：大数据时代的企业管理 / (美) 克鲁贝克 (Klubeck, M.) 著；吴海星译. -- 北京：人民邮电出版社，2013.1

书名原文: Metrics: How to Improve Key Business Results

ISBN 978-7-115-29961-1

I. ①量… II. ①克… ②吴… III. ①量化—企业管理 IV. ①F272

中国版本图书馆CIP数据核字(2012)第266576号

内 容 提 要

本书是大数据时代企业进行量化管理的成功指南。首先简要介绍了量化分析工具与其他改进工具的联系，接着重点阐述了量化分析体系的构成，全面介绍了量化分析这一改善组织、提升业绩的强大工具，旨在帮助读者更加系统、科学地制定和完善自己的量化分析体系，挖掘数据背后的价值，真正将数据用于企业业绩的提升。

本书适用于所有致力于改善组织、提升业绩的数据分析项目经理、数据分析师。

量化：大数据时代的企业管理

◆ 著 [美] Martin Klubeck

译 吴海星

责任编辑 傅志红

执行编辑 张 霞

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号

邮编 100061 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

北京隆昌伟业印刷有限公司印刷

◆ 开本：700×1000 1/16

印张：17.75

字数：314千字

2013年1月第1版

印数：1-4 000册

2013年1月北京第1次印刷

著作权合同登记号 图字：01-2012-0967号

ISBN 978-7-115-29961-1

定价：45.00元

读者服务热线：(010)51095186转604 印装质量热线：(010)67129223

反盗版热线：(010)67171154

译者序

——从数据仓库到量化管理

量化管理是指以数字为基础，用数学的方法来考察和研究事物的运动状态和性能，以求对事物存在和发展的规模、程度等做出精确的数字描述和科学控制，实行标准化操作的管理模式。本书中的 Metrics 作为“以求对事物存在和发展的规模、程度等做出精确的数字描述”的工具出现，是一种为改善某种东西而对其进行详细描述的手段，用来支持决策，帮助改善组织、提升绩效。

数据仓库的概念植根于信息技术，由比尔·恩门（Bill Inmon）在 1991 年出版的《建立数据仓库》（*Building the Data Warehouse*）一书中提出。其最初的定义是一个面向主题的（Subject Oriented）、集成的（Integrated）、相对稳定的（Non-Volatile）、反映历史变化（Time Variant）的数据集合，用于支持管理决策。构建数据仓库考虑的也大多是信息技术问题，比如数据生命周期、元数据、数据质量，以及 ODS（操作数据存储）、ETL（抽取—转换—装载）。当然也有根据业务需求确定的主题域、信息、指标以及数据集市。但在构建数据仓库的知识体系中，对数据集市、主题域、信息和指标的论述，一般都局限在软件行业的需求调研方法以及数据建模，属于业务知识体系向信息技术体系的转换过渡方法。但如何提炼出一个组织的量化信息需求，确立业务知识体系，或者说如何进行描述，我在看到这本书之前，还没见过一种能把这个过程解释清楚的方法。

随着大数据概念的提出，各种 IT 技术也应运而生，比如 Hadoop 生态体系对 MapReduce 概念的开源实现，用于统计分析的 R 语言，MPP 架构的数据分析处理平台，甚至号称处理速度无与伦比的内存数据库。厂商们都在狂热地鼓吹自己的产品可以更快速、更有效地处理堆积如山的数据，从数据中淘金，把数据变成信

息，用信息构成知识。虽然决策支持、数据挖掘和数据分析喊了很多年，但直到最近，这些高级系统仍然停留在电信、金融和零售业（电子商务）等几个领域。虽然越来越多的管理者看到了 BI（BA）的力量，但对于如何掌握这种力量，还是丈二和尚摸不着头脑。甚至在 BI 得到成功应用的领域，也有点儿摸着石头过河的感觉。项目团队经常承受着很大的压力，不知道怎么找出客户自己也不知道的需求。

我曾参与过一个 BI 项目。这个项目的缘起，是因为公司在另一个行业 BI 项目上的成功。客户算是慕名而来，按他们的理解，我们的经验是可以套用到他们那个行业上的。你会在本书中看到，这是一个错误，而且是大家都会犯的错误。作者还专门为此编了一个悲剧版王子的小故事。灰姑娘留下了一双鞋，可王子没能找到红颜，所以冲冠一怒，颁布了一条触怒天下的法令，最终导致王国被颠覆。虽然我们没有王子那么悲惨，却也在跟客户磨合的过程中承受了很多痛苦。最初，我们向客户介绍各种主流的 BI 生态系产品，介绍各种数据分析算法，提供了整套的 BI 项目实施文档模板，确定了一部分主题域和指标体系，甚至还根据客户提供的测试数据训练了一个神经网络模型用于客户分群。但我们很长时间也没能给出客户想要的描述。客户和我们都很痛苦。因为从一开始，我们就没有确定根本问题，像作者说的，发动引擎开始了一段不知道要去哪儿的旅行。由于我们对新行业缺乏了解，也没有答案纲要的指导，不知道怎么帮客户确定根本问题，甚至也不知道让客户找出自己的根本问题。我们虽然知道怎么按客户已有的数据结构搭建数据仓库模型，知道怎么实现 ETL，知道怎么保证数据质量，知道怎么做数据分析，知道怎么展现分析结果，但我们不知道怎么向客户做一个完整的描述。所以在那个项目中，我们消耗了很多时间来寻找向客户描述的方法。好在我遇到了这本书，它帮助了我，所以我希望它也能帮助曾经和我一样痛苦挣扎过的人。

“量化分析”的含义要比统计分析丰富得多，但在某些方面又不如统计分析。这本书中没有关联分析、回归分析，更别提决策树和神经网络了。也没有维度和雪花模型，没有 SQL 语句，没有 R 语言。但我希望所有数据分析师或准备做数据分析的人都能读一读这本书，如果把数据分析、数据挖掘和数据仓库建模当作术，那么这本书中讲的内容就是道，是内功心法。虽然它不能帮你解决掉 SQL 中的某

个 bug，不能告诉你怎么把数据画成一幅南丁格尔玫瑰图，但它可以帮你架起数据仓库到量化管理的桥梁，让你的技术能为业务服务。

感谢图灵给我翻译本书的机会，感谢傅志红老师在我翻译初期的悉心指导和耐心解释，感谢编辑朱巍、楼伟珊和张霞对本书认真负责的审校和中肯的修改建议，使得本书最终不至于语句不通，惨不忍睹。他们的认真和敬业令人钦佩！

还要感谢我的同事们，他们和我一起经历了项目的煎熬，一起和客户交流。他们教会了我很多数据仓库方面的知识，在我的翻译工作中也提供了很多帮助和支持。最后感谢我的母亲和妻子。她们承担了家里的大部分工作，如果没有她们的理解和支持，我不可能每天抽出那么多时间来完成本书的翻译工作。

谨以此书献给那些致力于提升组织、优化流程、完善自我的人们。
把时间用在追逐梦想上吧，不要浪费时间去追逐数据。

献给我的老父亲，我真的非常想念您。我写作本书最大的希望就是
让您为我感到自豪。

致 谢

据我所知，这部分内容的目的是让那些帮助过我的人知道，我没忘记他们对本书的贡献，所以这部分让我感觉压力山大。

我不想忘掉任何提供过帮助的朋友或同事，他们在本书写作过程中帮忙评审，提出批评意见，以及编辑建议。特别是：唐·帕吉特，丹尼塔·李斯，利亚·郎，凯斯·麦金托什，马林·斯坦尼克，还有我的兄弟欧文。谢谢你们一接到通知就抽时间阅读，谢谢你们温暖的鼓励，谢谢你们陪着我。

当然，必须感谢我的技术评审拉斯·齐思曼。尽管我和拉斯经常就量化分析及其使用发生争执，但当编辑要求我推荐一名技术评审时，我马上就想到了拉斯。他的诚实和激情，他渊博的知识，他的职业素养，都让他脱颖而出。非常感谢，拉斯，感谢你辛勤的工作，你提出的那些建议，还有你诚实的评价。

我要特别感谢迈克尔·朗索恩。不仅因为你是我最可靠的审阅者，还因为你一贯的鼓励和支持，给我提供了开始和坚持这项工作的动力。我非常感激你的帮助、指导以及温和而坚定的鼓励。谢谢！没有你，我真的不可能完成这项工作。

我也要感谢杰夫·奥尔森、金伯利·伯顿和安妮·贝克。尽管你们是 Apress 编辑团队中的一员，而且“只是在做本职工作”，但你们对我的帮助绝不限于此，非常感谢你们的帮助。你们诚实，一致，公正。你们专注于为 Apress 推出最好的作品，并且对作为作者的我表现出了真诚的关心。等哪天有机会，我要请你们喝酒。

俗话说，最后但同样重要的是，我要感谢我的家人。特别是我的妻子克里斯蒂娜。这段时间，你成功地假装对我的进度甚至是工作感兴趣，给我鼓励和支持。谢谢你的努力，亲爱的，我爱你。我的女儿安丽萨，谢谢你提供的童话故事，并且让我把你的作品放到书里。我期待着有更多的人把你的名字印在外套上。我爱你。

最终，特别感谢你，不管是谁，读这本书。你很特别，肯定是，因为除非觉得能找到自己的名字，没人会看这部分内容。所以你肯定是那种会把书从头到尾都看遍的极少数人之一，这也意味着你可能是会让这部分内容派上用场的读者，也可能是那种愿意分享自己的想法与好恶那类人。所以，希望我能得到你的反馈意见，先谢了。我希望这本书能在你进入量化分析体系开发的危险水域时提供导航。

注意：尽管我努力给你提供真实案例（真实的情景、交互、经历、指标和数据），但我必须保护数据提供者的隐私。所以，如有必要，我会把单个事例混合到一起。对于数据也是如此。接下来你会看到，我完全相信数据归数据提供者所有，所以我要保护出现在书中的这些实体和数据。

技术评审简介

拉斯·齐思曼是一位资深 IT 专家和顾问，系统开发生命周期各个阶段的经验都很丰富。他职业生涯的大部分时间都致力于为业务问题或机遇提供 IT 解决方案。他曾经担任过很多行业的 IT 经理和从业人员，包括银行/金融业、制造业、建筑业、零售业、医药业、电信业和医疗保健行业。齐思曼先生最近几年参与了几个 IT 和医疗组织内的业务绩效评估和管理项目，涉及了业务战略、平衡计分卡、量化分析、关键绩效指标和业务分析系统的应用。

齐思曼先生很乐意担任本书及相关量化概念的高级技术评审，并且期待本书的出版，希望本书对渴望提升、追求成熟和最佳绩效的个人、团体和组织能有所帮助。

前 言

——量化分析：基础

在所有能用于组织改善的工具中，我最关注的是量化，因为对它的呼声最高，误解也最深，而且人们对它感到恐惧，可它却非常有用。从宏观上定义量化分析，需要从以下六个方面来了解。

- 量化分析的定义
- 量化分析的原因
- 何时进行量化分析
- 量化分析的归宿
- 量化分析的所有者
- 如何进行量化分析

量化分析的定义

量化的定义方式有很多。在本书的第 1 章，我给出了一种定义和使用量化的通用语言。对于我来说，量化是一种为改善某种东西而讲述一个完整故事的手段。尽管有时候是为了优化一个业务流程，但通常都是为了改善组织。归根结底，任何改善都应该可以归结为对组织有帮助的改善。

量化是改善工具。按照它们的特性，量化分析使用不同层级的信息（数据、指标、信息和其他量化指标）讲述故事。尽管我一直尽力保证叙述的完整性，但把所有内容都囊括其中几乎是不可能的。在大多数情况下，我会尽量把对改善有帮助的重要内容都包含进来。

量化为改善工作提供了很多帮助，可以让你在第一时间确定哪里出错了，工作进展如何（有没有改进以及改进是否达到了预期效果），改变之后的新环境看起来什么样。这里提到“改变”是因为改进是改变的一种，不一定是沧海变桑田。做了些新的事情，不再做某些事情，或者用不同的方式做某些事情，这些都是改变。没有改变就不可能有改进。

爱因斯坦说：“精神错乱的定义就是用同样的方式做同样的事，却期待不同的结果。”

下面是一些量化分析可以完成的事情。

□ 改善公司各个领域的“健康状况”

- 提高客户满意度
- 提高产品/服务的价值
- 提高员工满意度
- 提高业务流程的效益
- 改善战略、规划和执行力

□ 为改变提供基础

- 改变什么
- 什么时候改变
- 如何改变

□ 提供理解组织的洞察力基础

- 优势
- 问题
- 弱点
- 机会

量化分析是为以改进为目的的改变而进行的。最起码我是这么用它的。我在书中讲述了一个为改善组织整体区域而构建量化分析体系的方法，也介绍了如何为改善特定区域而构建独立的量化分析体系。我是理想主义者，对量化分析体系的信念就体现了这一点，我认为它是一个强大的改进工具。

量化分析的原因

为什么要做量化分析？为了改进。我知道这已经说过了。但是，为什么偏偏是量化分析呢？为什么不是其他那些大行其道的方法呢？（像 TQM、六西格玛、平衡计分卡等。）有趣的是，选择其中任何一种方法都会要求你使用量化分析，或者至少是量化分析的组件（数据、指标或信息）。这些改进方法首先要评测现在的状态，这是基础，这样在方法实施之后，才能看出优化流程取得了什么样的成绩。这些方法还要求你评测投入到改进工作中的时间、金钱，但没有一个能真正帮你改进你想改进的东西。

量化分析能提供一些基本帮助，所以才成为了重要的改进工具。无论要改进的是产品、服务，还是业务流程，量化分析都能提供所需信息。这种洞察力对于执行任务（满足需求或提供服务）的人很有价值，能够让他们以一种新的、通常也更全面的眼光去看待自己的工作。量化分析会帮他们找到改进方法，还能让他们看到自己能得到的好处；能够给出宝贵的反馈，让组织可以用量化分析把持续改进变成现实（而不是一个新噱头）。

量化分析也为高层管理者提供了洞察力，让团队领导可以把改进工作的成果展示给投资人，让人们注意到你的改进工作，以便得到更多的支持。

它还可以让你用客户容易理解的方式跟他们分享你的工作成果。客户能够明白事情怎么向好的方向发展。所有新产品发布时（我特别喜欢苹果公司的半年度公告）都能注意到贯穿始终的各种量化指标。

量化分析提供了洞察力，还会为你的观点的合理性提供支持。如果其他条件都旗鼓相当，数据就是决定胜负的关键。在几个部门争夺资源时，能拿出数据的会最终胜出。

何时进行量化分析

理想情况下，在组织没有成熟的时候，你不会进行任何大规模的量化分析，因为不成熟的企业没有能力组织大规模的变革活动。不过，这是针对覆盖整个企业的大规模量化分析计划而言的。如果你只是负责一个部门，只是想要实现一个

部门级别的量化分析计划，那么只需要确保你们部门能够实现变革就行了。如果你们提供一两项产品或服务，实现量化分析计划应该没问题。记住，量化分析只是一个改进工具，所以除了量化分析，你还需要实现其他改进工具。量化分析可以指明工作重点，表明工作所取得的进展，但量化分析本身不是一套解决方案。

很可能你已经收集了数据和指标，你可能拥有跟踪、收集数据和指标甚至能产生报表的自动化的工具，你可能要为一个年度报告提供信息，你可能已经满足了一些特定指标的需求。你所在的行业，可能已经有用了好几年（甚至几十年）的标准了。但这并不一定是我所定义的量化分析，创建这样的标准也不是本书的初衷。那些数据和指标是拿来汇报的，而不是拿来用的。它们并不是用于优化流程、改善产品或服务的。

也许是时候用这些信息维护自身利益，开发量化分析计划了。你要问的关键问题是自己有没有准备好，愿不愿意、能不能做出改变。你想改进吗？

你将在本书中学到，所谓“何时”，就是在改善某件事情之前，事先广泛收集相关信息，一般优秀的研究人员都这么做。所以，这个“何时”就是在改进工作开始之前，进行当中，以及完成之后。

量化分析的归宿

量化分析计划应该属于数据的所有者。我工作的 80% 是帮别人构建量化分析体系，通过设计、创建、实现和维护量化分析体系指导客户。在很多案例中，我也制作和发布他们的量化分析结果。但如果有人问起，我很乐意告诉他们，这些量化分析都不是我的。我只是在帮它们的所有者制作而已。我最成功的时候，就是把量化分析的维护和发布交给所有者团队的时候。然后我的角色变成了顾问，帮助他们使用和完善量化分析体系。这种转变需要时间，通常是因为缺乏技巧或资源。但把量化分析的制作、维护和发布交给它们的所有者一直都是我的最终目标。

所以，量化分析最终要交到数据所有者手里。可以挂到他们办公室的墙上，可以放在他们的共享磁盘里，也可以放在他们的网站上。那才是量化分析的归宿。但量化分析也可以出现在其他地方，比如年度报告、月例会或者公众网站上。

是否公开量化分析由所有者决定。决定在哪里发布量化分析结果时一定要慎重，要斟酌量化分析结果的使用和其他人访问该信息的必要性。组织越成熟，越适于分享量化分析结果。很多组织还没成熟到可以跟同行、客户分享量化分析结果，更别提向公众发布了。

量化分析的所有者

我每天都会处理这个“所有者”的问题。我们组织中用的很多量化分析体系都是由我创建的。我还是首席设计师、数据采集人员、分析师和结果发布人员。但就像在“归宿”那一节中提出来的，我一直寻找机会把量化分析移交给它的所有者。

所有者是谁？主要是那些交付产品、服务或执行业务流程的人，这个我在后面还会详细讲解。但量化分析的所有权可能会涉及整个组织，具体取决于如何定义所测量的项目。

这里的关键不是量化分析属于谁，而是它不属于谁。不要把一线工人排除在外，不要认为量化分析归 CEO 或高层管理人员所有。即便量化分析结果是向高层汇报的，也不意味着它们就应该只归高层所有。记住量化分析的目的。除非是 CEO 和高层管理者要优化工作流程（他们极少会直接发布产品或服务），否则你就必须把执行任务的人——那些负责让改进实际生效的人——作为所有者包括在内。如果高层管理者发现有个部门在用量化分析做改进工作，但却没跟高层分享，能有多大危害呢？有一些吧。但是请想象一下，如果有个部门发现高层管理者已经审阅了跟他们的流程、服务或产品相关的量化分析结果，但他们自己还不知道，这会激起多大的麻烦？会不会对组织造成极大伤害？我敢打赌会的。

如果高层管理者想要一个部门、单位、业务流程、服务或产品的量化分析结果，那所有涉及的人都应该是量化分析结果的分发对象。他们都应该参与到设计、创建和发布工作中来。

很简单，“所有者”就是：组织内部的所有人，以一线员工为主。

如何进行量化分析

这就是本书的写作目的。我会在书中给出一套完备的量化分析计划或特定优化领域的量化分析开发指南。之所以称为“指南”而不是“法则”，是因为法则仅指正确的方法。跟大多数事情一样，开发量化分析也不止一种正确的方法。我所提供的语言、流程和工具是基于 20 多年的经验积累得出的。这些经验中既有成功也有失败。两者都让我获益良多，我也很高兴跟你分享我学到的东西，以免你重蹈我的覆辙。而且如果你已经掉进坑里了，我的建议很简单：别再挖了。

本书主要内容

掌握描述量化分析及其组件的统一语言是我们在本书展开讨论的基础。在编写这本书的过程中，以及在考虑所有能用来使量化分析和发布更容易的工具的时候，我想到量化分析和其他改进工具之间还有一个非常重要的差异。除了数据、指标、信息和量化分析的定义，我还要跟你分享量化分析的另一面。

首先，对我来说，量化分析本身是改进工具，同时还是其他改进工具的组成部分。即便只是用量化分析来跟踪进度或预测未来的发展趋势，量化分析也应该被看作是一种改进方法，但那还不足以让它从多如牛毛的工具中脱颖而出。我用过很多用来开发改进计划或解决组织问题的工具，全面质量管理、能力成熟度模型、精益和精细六西格玛是其中几个。每个改进方法都会用到数据和指标，六西格玛在它的整个流程中都会用到数据，不仅用来评测改进，还用来确定要改进什么，还有评测成功和目标达成的指标。甚至有把普通愿望变成 SMART 目标的指标（SMART 即 Specific、Measurable、Attainable、Realistic、Time-bound，具体、可衡量、可实现、现实性、时限性）。

很多围绕量化分析设计的工具都是用于统计分析的。这些强大的工具可以明确不同数据之间的关系、因果关系，甚至能确定数据的准确性。

但对我来说，“量化分析”的含义要比统计分析丰富得多，但在某些方面又不如统计分析。

量化分析不是统计学家的目标，也不是分析师的最爱。统计分析学院里讲的

知识要比这里涉及的广泛得多，也深入得多。我在教授量化分析时，通常都会声明我不是搞统计的，课程里面也不会讲统计学。从这些角度讲，量化分析涉及的知识有点少。

但从很多方面来看，量化分析又比统计学涉及的知识面广。它们是讲故事的一种方法，而且有很强的洞察力。当你处于抉择关头，需要选择不同的改进方向时，量化分析可以作为路标为你指明方向。

对我来说，量化分析是组织发展计划的基石，也是解决重要组织问题的工具。

正是量化分析和其他组织发展工具之间这些细微的差别，使得写这本书非常必要。统计学有专门的课程（我有个同事进入量化分析这行就是因为他是搞统计学的），各种分析工具的书也数不胜数，为这个开发的软件也很多（随便举三个例子，SPSS、MiniTab、Sigma XL），但很少有介绍或教授怎么使用量化分析的作品。在 Kaplan 和 Norton 的平衡计分卡方法和 Dean Spitzer 博士的 *Transforming Performance Measurement* 一书中，已经隐约体现出这种缺失来了。我想接过他们手中的枪，做好下一步工作——让每个人都能掌握量化分析的设计、创建和使用。

有了量化分析，你就能够利用数据、指标和信息来改善组织，并能在关键业务成果上取得成功。

我希望这本书能帮你构建量化分析体系，从而帮你改善组织。不管组织的使命有多大，量化分析都是组织改进的利器，而这本书会让量化分析尽可能简单。

目 录

第 1 章 统一语言：数据、指标、信息，天哪！	1
1.1 三头小猪的故事	1
1.2 数据	5
1.3 指标	6
1.4 信息	7
1.5 量化分析	8
1.6 根本问题	10
1.7 数据-量化分析悖论	14
1.8 量化分析体系的构成	16
1.9 总结	17
1.10 小灶	17
第 2 章 设计量化分析体系：怎么设计	20
2.1 向着根本问题，前进！	21
2.2 检查根本问题	28
2.3 构建量化分析体系	29
2.4 明确信息、指标和数据需求	32
2.5 采集指标和数据	34
2.6 如何采集数据	35
2.6.1 优先考虑自动化	36
2.6.2 置办软硬件	37
2.6.3 开展调查	37
2.6.4 用人	37
2.7 再供例子一枚	38
2.7.1 信息	39
2.7.2 指标	39
2.7.3 数据	40
2.8 回顾	41