

## 大数据 文献总结 (3)

MIT-Plato 团队

1. Mining Social Science Data: a Study of Voting of the Members of the Seimas of Lithuania by Using Multidimensional Scaling and Homegeneity Analysis

[Krilavicius, Tomas](#); [Morkevicius, Vaidas](#). [Intelektine Ekonomika](#) 2 (2011).

摘要: 多维的定标 (MDS) 是一种著名统计和数据采矿技术。它对许多不同的区域中的一种 exploratory 数据分析和显现是可适用的, 诸如经济学, 尤其销售, 信誉风险分析, 心理学和计算机科学。然而, 它蒙受一些严重的弊端之苦, 即它依赖于几个主观的参数: 数据编码, 相似之处措施和模拟类型的选择。我们在新奇的应用分析投票 Lithuanian 议会 (MP) 的成员的一点名的 MDS 中显示这些弊端。我们建议使用允许从逃离的一种不同的技术在社会科学数据采矿中论及问题, 一种均衡性分析。我们简要地讨论它, 在相同的数据上说明其应用并且显示其在 MDS 上的优势。在我们集中于两种方法的技术和 methodological 方面的纸中, 因此, 它能容易地被应用分析各种各样的经济数据, 诸如电讯的客户搅乱器或者销售的用户群。我们讨论所有使用的工具, 把卷呼叫选票, 相似之处措施, 划分 (或者非部门) 成为实质性的时期的编码, 还有多样化的显现技术的 MDS 和均衡性分析的解决方案的尺寸。我们比较 MP 在生产的阴谋计划中的大多数物体代表的均衡性分析的结果的不同的显现技术: 2D 和 3D 对象阴谋计划; 范围阴谋计划, 对于每一在其中一棵最小限度的横越的树被拖拉的物体 (在我们的情况下一个派系) 的类别; 2D 和 3D 星阴谋计划, 在其中每一对象与其类别质心连接。我们为了模拟社会科学数据的研究者以推荐结束并且提出我们的关于投票的分析的未来的计划。

2. Improving image mining through Geoprocessing

[Guadagnin, R](#); [Santana, L](#); [Ferneda, E](#); [Prado, H](#). [Pattern Recognition and Image Analysis](#) 20.1 (Mar 2010): 81-85.

摘要: Geoprocessing 信息系统 (GIS) 处理构造的信息有关一些地理地方化。在

一个巨大的数据库中的那么一种使用三维的形象表示系统,在其中插入许多数据可能的大约一些种使领域感兴趣,说人口统计学的,农业,经济学,工业,如此等等。形象是能支持决策过程非常好的强大的信息来源。一个形象能被视为一套要素空间的地方化和颜色。为了解释一个形象包括引出这样要素之间的丛群和关系。本文提出 Geoprocessing 和形象采矿的一种综合支持形象为基础的的决定在诸如健康护理的几个领域中。

### 3. Distributed data mining for e-business

[Liu, Bin](#); [Cao, Shu Gui](#); [He, Wu](#). *Information Technology and Management* **12.2** (Jun 2011): 67-79.

摘要: 在以因特网为基础的电子商务环境中,大多数商业数据被分发,不同种类和私人。为了实现真实的商业智能,分发的数据的采矿大的量是必要的。通过一篇彻底的文学回顾,本文在分发的数据采矿(DDM)系统中为了电子商务识别四个主要的问题并且把现代的DDM系统分类成三个类别代表的抽样。解决这些识别的问题,本文提出名为DRHPDM(以关联为基础分层平行的数据来源分发数据采矿模型)的一新奇的DDM模型。此外,改进最后的结果的质量,数据来源被划分成为一个集中的采矿层和一个分发的采矿层,根据他们的关联。为了改进DDM系统,网络服务和多代理技术的开放,跨平台能力,和智能被采取。DRHPDM的可行性被建设一个原型系统和运用它到一个网络用法采矿情节证实。

### 4. Future for Corporate Networks

[Richardson, Ted](#). *Data Processing* **27.6** (Jul/Aug 1985): 37.

摘要: 自从通信必须因为他们的日益增长的重要性有效地被管理,公司有时选择管理通信他们自己而非把责任给一个携带者。当公众的网络提供不断高的标准的一种标准化的服务时,一种需要将总为了不满足在国际上定义的标准通信设施存在。任何网络设计应该自从扩展的率以来包括大的扩展的能力每年能是多达200%。公司要求比被公众的网络需要的连通性是更灵活的公司范围的连通性。不能由任何分组交换网络访问的私人租借的电路能提供更多安全。在社团的网络中的一个关键的因素是跳回和冗余建造成为聪明的节点,并且如果一个租

借的电路被使用，获得交替路由选择较简单的。私人租借的数字服务将是向数据在通信水平上提供综合的社团的网络的基础，声音，和其它服务。

## 5. Challenges and opportunities: Unions confront the new information technologies

[Lucore, Robert E.](#) [Journal of Labor Research](#) **23.2** (Spring 2002): 201-214.

摘要：一种新的经济和工业风景在新的信息技术的影响下面进化。新的技术总已有一种深刻地影响他们工作的工人和公共机构的安排。新的信息技术也将有一种深刻地影响今天的工人和他们的联合。本文通过讨论一些具体的工业的影响或者经济的部门概述新的信息技术密切相关的技术（因特网和之间的关系，和通过推测工人运动的潜力充分使用新的技术。

## 6. Is Big Data an Economic Big Dud?: [News Analysis]

[Glanz, James.](#) [New York Times](#), Late Edition (East Coast) [New York, N.Y] 18 Aug 2013: SR. 5.

摘要：被世界经济论坛作为“新的油”和“一个新的资产类别”命名，数据的这些巨大的装载像蒸汽机车，电栅格，钢，空调和无线电一样已经是对有改革能力的革新的 likened。

IF pencil marks on some colossal doorjamb could measure the growth of the Internet, they would probably be tracking the amount of data sloshing through the public network that spans the planet. Christened by the World Economic Forum as “the new oil” and “a new asset class,” these vast loads of data have been likened to transformative innovations like the steam locomotive, electricity grids, steel, air-conditioning and the radio.

The astounding rate of growth would make any parent proud. There were 30 billion gigabytes of video, e-mails, Web transactions and business-to-business analytics in 2005. The total is expected to reach

more than 20 times that figure in 2013, with off-the-charts increases to follow in the years ahead, according to [Cisco](#), the networking giant.

How much data is that? [Cisco](#) estimates that in 2012, some two trillion minutes of video alone traversed the Internet every month. That translates to over a million years per week of everything from video selfies and nannycams to Netflix downloads and "Battlestar Galactica" episodes.

What is sometimes referred to as the Internet's first wave -- say, from the 1990s until around 2005 -- brought completely new services like e-mail, the Web, online search and eventually broadband. For its next act, the industry has pinned its hopes, and its colossal public relations machine, on the power of **Big Data** itself to supercharge the economy.

There is just one tiny problem: the economy is, at best, in the doldrums and has stayed there during the latest surge in Web traffic. The rate of productivity growth, whose steady rise from the 1970s well into the 2000s has been credited to earlier phases in the computer and Internet revolutions, has actually fallen. The overall economic trends are complex, but an argument could be made that the slowdown began around 2005 -- just when **Big Data** began to make its appearance.

Those factors have some economists questioning whether **Big Data** will ever have the impact of the first Internet wave, let alone the industrial revolutions of past centuries. One theory holds that the **Big Data** industry is thriving more by cannibalizing existing businesses in the competition for customers than by creating fundamentally new opportunities.

In some cases, online companies like Amazon and eBay are fighting among themselves for customers. But in others -- here is where the cannibals

enter -- the companies are eating up traditional advertising, media, music and retailing businesses, said Joel Waldfogel, an economist at the [University of Minnesota](#) who has studied the phenomenon.

"One falls, one rises -- it's pretty clear the digital kind is a substitute to the physical kind," he said. "So it would be crazy to count the whole rise in digital as a net addition to the economy."

Robert J. Gordon, a professor of economics at [Northwestern University](#), said comparing [Big Data](#) to oil was promotional nonsense. "Gasoline made from oil made possible a transportation revolution as cars replaced horses and as commercial air transportation replaced railroads," he said. "If anybody thinks that personal [data](#) are comparable to real oil and real vehicles, they don't appreciate the realities of the last century."

Other economists believe that [Big Data](#)'s economic punch is just a few years away, as engineers trained in [data](#) manipulation make their way through college and as [data](#)-driven start-ups begin hiring. And of course the recession could be masking the impact of the [data](#) revolution in ways economists don't yet grasp. Still, some suspect that in the end our current framework for understanding [Big Data](#) and "the cloud" could be a mirage.

"I think it's conceivable that the [data](#) era will be a bust for the things people expect it to be useful for," said Scott Wallsten, a senior fellow at the Technology Policy Institute and the Georgetown Center for Business and Public Policy. Some entirely new use will have to turn up for [data](#) to fulfill its economic potential, he added.

There is no disputing that a wide spectrum of businesses, from e-marketers to pharmaceutical companies, are now using huge amounts of [data](#) as part of their everyday business.

Josh Marks is the chief executive of one such company, masFlight, which helps airlines use enormous **data** sets to reduce fuel consumption and improve overall performance. Although his first mission is to help clients compete with other airlines for customers, Mr. Marks believes that efficiencies like those his company is chasing should eventually expand the global economy.

For now, though, he acknowledges that most of the raw **data** flowing across the Web has limited **economic** value: far more useful is specialized **data** in the hands of analysts with a deep understanding of specific industries. "The promises that are made around the ability to manipulate these very large **data** sets in real time are overselling what they can do today," Mr. Marks said.

Some economists argue that it is often difficult to estimate the true value of new technologies, and that **Big Data** may already be delivering benefits that are uncounted in official **economic** statistics. Cat videos and television programs on Hulu, for example, produce pleasure for Web surfers -- so shouldn't economists find a way to value such intangible activity, whether or not it moves the needle of the gross domestic product?

In addition, infrastructure investments often take years to pay off in a **big** way, said Shane Greenstein, an economist at [Northwestern University](#). He cited high-speed Internet connections laid down in the late 1990s that have driven profits only recently. But he noted that in contrast to the Internet's first wave, which created services like the Web and e-mail, the impact of the second wave -- the **Big Data** revolution -- is harder to discern above the noise of broader **economic** activity.

"It could be just time delay, or it could be that the value just isn't there," said Mr. Greenstein, who has studied the competitive success of online businesses in media, advertising and retailing.

Perhaps surprisingly, the parallel most tightly embraced by digital futurists -- the rise of the electricity grid -- is largely dismissed by those who have studied the history of the subject. The idea is that a ubiquitous Internet will make data and "cloud" computing available anywhere, like electricity through a socket.

The numerical comparisons are tantalizing. As illustrated in "The Electric City," by Harold L. Platt, the booming quantity and adoption rates of electricity flowing on the Chicago grid in the late 19th and early 20th centuries instantly bring to mind those charts showing data growth today.

Despite those similarities, Mr. Platt, a professor emeritus of history at [Loyola University Chicago](#), said it was unlikely that the revolutions unleashed in manufacturing, domestic life, transportation and high and low society by electricity could ever be matched by the data era. "I'd be hard pressed to quickly draw comparisons," he said.

But even as Mr. Platt, 68, spoke by cellphone from Chicago, fragments of today's inescapable data flood found him as he received messages from his grown children. "I have to text them or else they won't answer me back," Mr. Platt said gamely. "I'm going with the flow."

如果在一些庞大的 doorjamb 上的铅笔标记能测量因特网的增长，他们将有可能追踪通过横越行星的公众的网络溅着泥水行进的数据的量。被世界经济论坛作为“新的油”和“一个新的资产类别”命名，数据的这些巨大的装载像蒸汽机车，电栅格，钢，空调和无线电一样已经是对有改革能力的革新的 likened。

增长的震惊的率将使任何父母骄傲。 有中的视频，电子邮件，网络处理和商业对商业的逻辑分析的方法的 30 十亿 gigabytes2005. 总数被期待超过 20 次到达，有了图表在 2013 年中的数字增加今后跟随，根据思科，网络巨人。

多少数据是那？ 思科在 2012 年估计那个，大约两万亿分钟的视频每月单独地横过因特网。 那每件事的每星期从视频 selfies 和 nannycams 翻译成为超过百万年成为 Netflix 下载和“Battlestar Galactica”事件。

什么是有时被称作因特网第一个波浪--从九十年代直到大约 2005 年说，--带来像电子邮件，网络，联机搜索一样完全新的服务和终于宽带。 对于其下一行动，工业已把其希望别对经济增压，和其庞大的公众关系机器，在大的数据自己的力量上。

有正好一个微小的问题： 经济并且在网络交通中的最新的波涛期间在那里停留，充其量，在忧郁中。 生产率增长的率，好好地到 2000s 中其来自七十年代的稳定的升高已在计算机和因特网革命中归因于较稍早的阶段，已实际上落下。 总体经济趋势是复杂的，但是一个争论能被做当大的数据开始出现时，放缓在 2005 年左右开始--公正。

那些因素有一些经济学家询问大的数据是否曾经将有第一因特网波浪的影响，更不用说过去几个世纪的工业的革命。 一种理论坚持大的数据工业通过在竞争中为了客户把现有的商业拆取可利用的配件比通过创造基本上新的机会更加是兴旺。

在某些情况中，像亚马逊和 eBay 一样的联机公司在他们自己中间为了客户战斗。 但是，在其他的中--这里是吃人的人进入的地方--公司吃传统的做广告，媒介，音乐并且零售商业，说 Joel Waldfogel，已学习现象的明尼苏达的的一个经济学家。

“一下降，一种升高--这是相当清楚数字种类是对物理的种类的一种替代品”，他说。“这样在中作为一张网增加计算全部升高到经济将疯狂的数字。”



罗伯特 J. Gordon，在西北的大学的经济学的一位教授，说比较大的数据加油提升的废话。“当商业的航空运输替换铁路时，汽车替换马时，由油制造的汽油使一个运输革命成为可能，以及”，他说。“如果任何人认为个人的数据与真实的油和真实的车辆相比，他们不赞赏上世纪的现实。”

其它经济学家相信当数据驾驶的开始 ups 开始雇佣时，在数据操纵中训练的工程师前进穿过学院时，大数据经济的用拳猛击正好是在几年以外，。同时，当然衰退能按照经济学家目前不抓住的方式掩蔽数据革命的影响。仍然，一些怀疑者我们的理解大的数据和“云”的当前的框架最后能是一个海市蜃楼。

“可想到的是我认为数据时代将是事情人们的一块破产期待它对有用”，Scott Wallsten 说，技术政策机构的一个高级伙伴和商业和公众的政策乔治敦中心。某种完全新的使用不得不为了数据出现完成其经济潜力，他接着说。

没有争论，从电子市场人员到药物的公司，商业的一种宽阔的频谱现在使用数据的巨大的量作为他们的日常的商业的一部分。

玩笑标记是一这样公司的特首，masFlight，这帮助航线使用巨大的数据组减少燃料消费并且改进总体表现。虽然他的第一个任务是帮助顾客为了客户与其它航线竞争，但是标记先生相信像那些一样的效率他的公司追逐应该终于扩充全球经济。

对于现在，尽管，他承认横跨网络流动的大多数未经加工的数据已限制经济价值：多的多有用在分析员的手中以对于具体的工业的深的理解被专门研究数据。“在能力周围被做在实时操纵这些十分大的数据组的诺言他们今天能把所做的卖的过多”，标记先生说。

一些经济学家争辩估计新的技术的真实的价值经常困难的，大的数据已经可能传送在官员经济统计中是没有数过的的利益。有关 Hulu 的猫视频和电视计划，例如，为了网络冲浪者生产愉快——经济学家那么不应该发现一种方式评价这样无形的活动，它是否移动国民生产总值的针吗？

此外，基础设施投资经常花费多年大规模支付，说 Shane Greenstein，西北的  
大学的一个经济学家。他引用在九十年代末放下的高速的因特网联系最近仅仅  
驱使利润。但是，与因特网第一个波浪形成对比他注意到那个，这像网络和电  
子邮件一样创造服务，第二个波浪的影响--大的数据革命--是难辩明在较广阔  
的经济活动的噪音之上。

“这能是公正的时间延迟，或者能，价值正好在那里不是”，Greenstein 先生说，  
后者已学习媒介中的联机商业的具有竞争性的成功，做广告和零售。

或许，惊奇，平行由数字 futurists 最严紧地被拥抱-- --电栅格的升高很大程  
度上被已学习主题的历史的那些人开除。想法是一普遍存在的因特网将无论何  
地像一个插座的电一样使数据和“云”使用计算机可供使用，。

数字的比较是撩弄的。正如在“电的城市”中所说明的，在晚第 19 和早第 20  
世纪在芝加哥栅格上流动的电的 Harold 通路 Platt，繁荣的数量和采纳率立即  
带来注意那些今天显示数据增长的图表。

尽管那些相似之处，Platt 先生，一位教授在 Loyola 大学芝加哥的历史，不可  
能的是说释放在通过电制造，家庭生活，运输和高和低社会中革命曾经能按数据  
时代匹配。“我将艰难迅速地拖拉比较压”，他说。

但是，68 岁，作为 Platt 先生的甚至通过来自芝加哥的电话讲话，今天的不能  
逃避的数据水灾的片段当他收到来自他成长的孩子信息时，找到他。“我不得  
不如此他们或 ck”，Platt 先生兴致勃勃地说。“我去流动。”

## 7. Big Data: A Boon to Business Intelligence

[Ohata, Michael](#); [Kumar, Arun](#). [Financial Executive](#) **28.7** (Sep 2012):  
63-64.

摘要：大的数据的诺言为了所有的竞技场中的领导人诱惑和压倒优势。对于金  
融和信息技术执行官，利用大的数据的挑战是税收的潜力，而利润提高是一种重  
要的挑战。在当前的商业环境中，用户有由他们支配的各式各样的工具，而他  
们的经验驱使，为了通信，数据，处理通过聪明的以因特网为基础应用连接的期

望社团的工人，不仅消费者和采取行动。最近的经济骚动强迫公司承认那个，甚至在好的经济条件下，增长不防止对于切实和准确数据和可靠，透明分析的过程的需要。对于数据的范例变换作为一个平台和逻辑分析的方法当一种效用握住提高一个组织的智商的潜力时，实现较更大的敏捷和提高表现。

## 8. With Big Data Comes Big Privacy Concerns

Anonymous. [Information Management 47.3](#) (May/June 2013): 7.

摘要：大数据的潜能是极大的。它造成作出和在每个领域中更棒的进步的更聪明的决定——提供了它有效地被处理而且挖掘，当然。那潜能依次在~周围保护隐私举起严重的关心。去年，世界经济的论坛处理了被来自美国，欧洲，亚洲和中东的政府官员，隐私提倡者和商务主管参加的一系列的工作室。在三个主要的区域上被集中的讨论：1. 保护和安全；2. 责任；和 3. 权利和对使用个人的数据责任。