

所有政策,人民共建共享的伟大蓝图,实现精准医疗、健康措施的精准推荐,都离不开大量的数据分析与研究,这需要借助统计学。要让广大人民群众享有公平可及、系统连续的预防、治疗、康复、健康促进等健康服务,离不开对大量连续观察数据的分析探讨,需要运用统计学方法。要做到正确处理政府和市场关系,在基本医疗卫生服务领域政府有所为,在非基本医疗卫生服务领域市场有活力,更离不开数据的支撑,需要统计学提供科学依据才有可能达成。

随着我国卫生健康事业的发展,各种与人民生活、健康相关的数据会越来越多,需要分析研究的问题也越来越多,统计学的方法与应用越来越重要。

网络舆情分析的统计思维

朱建平 厦门大学

《统计与信息论坛》伴随着我的教学科研共同成长,30年的历程,同样验证着我国统计学科的变革与发展。“互联网+”行动的实施,驱动着时代前行,给统计学学科建设提出了严峻的挑战。时值《统计与信息论坛》创办30周年之际,就近年来在网络舆情分析应用和研究中遇到问题,谈谈自己的一点思考。

近年来,随着社会复杂程度的日益提高,网络舆情信息安全受到广泛重视,网络舆情监控及信息分析技术的研究不断深入,适宜网络舆情分析的方法不断涌现。在应用中,针对网络舆情分析的总体思路,某些步骤流程的具体实施处处体现着统计的思维和方法。

1. 网络舆情信息采集中的统计思维。传统的舆情信息汇集方法和渠道主要有文献研究、社会调查、计算机辅助电话访问调查、网络调查、舆情直报点、内参、信访工作、网络论坛等。针对网络舆情,文本数据主要建立在互联网上的Web页面和一小部分互联网应用软件系统中。此时搜集信息主要是网络定性资料收集技术,例如现有的网络爬虫或对其进行改良的网络爬虫完成信息的高效采集。这一思维将会进一步开拓或提升统计调查技术的理论与方法。

2. 网络舆情信息过滤中的统计思维。一般采用基于内容的过滤、基于网址的过滤和混合过滤等多种形式完成舆情信息预处理。在具体实施中,舆情过滤主要利用判断页面内容与主题的相关性,这是一种基于关键词的模糊匹配方法。在舆情分词方面,主要借助现有成熟的分词系统完成分词,并利用语义分析法和人工智能分词法完成分词,在此基础上提出的网络舆情热点发现分词法有效提高了网络舆情分词的精度。这一思维将进一步加深定性资料

聚类分析和判别分析的研究。

3. 网络舆情信息挖掘中的统计思维。基于Web挖掘的方法是数据挖掘在网络信息分析中的新应用,它能自动、智能地获取并发现相关舆情信息内涵和舆情热点,提高舆情处理和分析的效率和质量,实现网络舆情的智能分析和动态预警。这一思维将智能分析和统计模拟相结合,会开辟出新统计模拟的研究领域。

4. 网络舆情信息量化中的统计思维。基于语义的数据挖掘方法则主要通过分析Web文本中潜在的语义结构或借助本体、语义词典等外部语义知识,从舆情文本语义层面发现舆情规律,主要包括潜在语义分析法和基于外部语义知识的语义分析方法等。该方法能够将传统的文本分析深入到上下文的语义层面,通过量化加强对文本的语义特征提取和语义相似度计算来提高信息分析的精度。这一思维将会在完善文本数据分析技术的基础上,进一步产生新的文本数据研究的理论和方法。

网络舆情分析是一个充满机遇与挑战的研究领域,涉及多个学科领域。还应该注意,如何评判舆情分析的正确性和科学性是一个值得研究的问题。另外,民众的情感隐藏在文本语义中,现有方法大多围绕语法展开,语义层面的舆情分析技术还有待提高。

统计学在能源领域研究的应用

周四军 湖南大学

我国经济经历30多年的改革开放发展,经济又受到过金融危机的冲击,经济需要从高速增长向新的发展模式转型,即“新常态”。从原有的发展向“新常态”迈进过程中,我们将面临许多需要解决的困难和问题。习近平主席在G20杭州峰会开幕词上提出了五大发展理念——创新、协调、绿色、开放、共享,这与我国“十三五”规划纲要里的五大发展理念基本上是对应的。习主席十分看重全球治理的问题,并提出了四个重点:金融、贸易投资、能源、全球发展。其中能源利用是一个关键因素。这轮金融危机以来能源也是全球关注点,奥巴马上任以来也把能源作为美国经济的一个重点,围绕着能源问题、新能源问题,新能源与化石能源、清洁能源与化石能源的关系如何协调问题,能源引发的气候变化、温室气体排放问题等等,都是和能源相联系在一起。能源问题将是国内外经济学研究的一个长期的热点问题,统计学作为一门方法论学科,在能源领域研究的定量分析方面将发挥重要作用,包括:能源利用效