

基于回归分析的中国互联网发展态势指数研究

何恺,胡庆十

(安徽师范大学 经济管理学院,安徽 芜湖 241000)

摘要:近年来,我国互联网呈现迅猛的发展态势,学术界与产业界也就互联网发展陆续开展相关研究与讨论。研究以定量与定性相结合的视角,从网民规模、IPv4地址数、中国网站数量、网络国际出口带宽、网络应用状况等方面计算得出中国互联网发展态势指数并进行预测,最终从实践层面综述该指数的研究意义。

关键词:回归分析;互联网;发展;态势;指数

中图分类号:TP393 **文献标识码:**A **文章编号:**1009-3044(2013)27-6106-07

Research of Internet Development Trending Index of China

HE Kai, HU Qing-shi

(Anhui Normal University, Wuhu 241000, China)

Abstract: The Internet of China developed rapidly in recent years. Related researches and discussions have been made on the Internet development by academies and industries. The paper was written by combining quantitative and qualitative perspective. The Internet development trending index of China was calculated and forecasted using the following sub-indexes: Netizens, the number of IPv4 addresses, the number of sites, international bandwidth and network application status. Finally, implications for research were pointed out on practical level.

Key words: regression analysis; Internet; development; trend; index

1 选题的依据和意义

1.1 选题的依据和意义

近年来,中国互联网络呈现高速发展的态势。网民数、IPv4地址数、中国网站数量等互联网基础资源有了极大程度地增长。与此同时,微博、社交网站、团购等新型互联网应用不断创新与发展。研究网络发展态势,一方面,将定性化与量化相结合的研究方法应用于互联网发展态势研究,加速总体网络发展研究由定性化向量化方向的转变;另一方面,研究结论也可以为我国相关行业提供参考。

1.2 目前国内外的研究状况

经查阅相关文献,文献所载基本为对于互联网发展现象的具体阐述,尚无对于具体指标运用统计方法进行综合研究,因此,本选题方向具有创新性,且对于后续研究互联网络发展提供一定的参考。

1.3 数据来源与选用

采用CNNIC于每年1月发布的《中国互联网发展状况统计报告》(以下简称《报告》)年度统计数据。该《报告》一年两次,单数次为每年1月发布前一整年的统计报告,双数次为每年7月发布当年上半年统计报告。由于2002年以前部分数据无法获取,因而引用第11次《报告》至第31次《报告》的单数次报告数据,即2002年至2012年互联网发展状况统计数据。权重数据,运用层次分析法基本原理的数据则来源于研究者的第一手调查。

1.4 数据分析工具

研究所采用的分析工具为Excel 2003与Eviews 7。通过Excel 2003对数据进行汇总、整理及简单运算;通过Eviews 7对数据进行最小二乘回归运算。

2 统计数据基础处理:基本统计指标的确定与增长指数的构建

2.1 基本统计指标的确定

为了计算出中国互联网络发展态势指数,首先需要定义组成该指数的基本统计指标。根据历年《中国互联网络发展状况统计

收稿日期:2013-07-16

作者简介:何恺(1991-),男,安徽合肥人,研究方向为经济应用统计;胡庆十(1960-),男,安徽巢湖人,副教授,研究方向为数理统计与财务会计。

报告》，首先确定如下基本统计指标：网民规模(万人)、Ipv4地址数(万个)、域名总数(万个)、中国网站数量(万个)、中国网页数量(亿个)、网络国际出口带宽(Mbps)。

纵观这六项基本统计指标，发现域名总数，中国网站数量，中国网页数量三者存在包含关系：网站由域名和网站空间构成，通常包括主页和其他具有超链接文件的页面。即从逻辑上说，网站包含网页和域名。因此，以中国网站数量作为中国网页数量、域名总数的代表，一方面，可以有效减少基本统计指标之间的自相关；另一方面，可以有效提升基本统计指标的代表性。

经整理，基本统计指标如下所示：网民规模(万人)、Ipv4地址数(万个)、中国网站数量(万个)、网络国际出口带宽(Mbps)。

同时，为了从实际出发，突出互联网的实用性，真实反映其庞大的用户数量，决定引入网络应用状况这一指标。

网络应用状况，即通过研究《报告》中2002-2012年网民互联网应用状况，计算相应权数与年度网络应用状况(万人项)，主要反映了各项应用综合使用人数。以2012年为例，组成网络应用状况的具体应用及统计数据如下表所示：

表1 2012年中国互联网络应用统计数据

应用	用户规模	使用率	权数
即时通信	46775	82.9%	10.24%
搜索引擎	45110	80.0%	9.88%
网络音乐	43586	77.3%	9.55%
网络视频	37183	65.9%	8.14%
网络游戏	33569	59.5%	7.35%
博客/个人空间	37299	66.1%	8.16%
微博	30861	54.7%	6.76%
电子邮件	25080	44.5%	5.50%
社交网站	27505	48.8%	6.03%
网络文学	23344	41.4%	5.11%
网络购物	24202	42.9%	5.30%
网上支付	22065	39.1%	4.83%
网上银行	22148	39.3%	4.85%
论坛/BBS	14925	26.5%	3.27%
团购	8327	14.8%	1.83%
旅行预订	11167	19.8%	2.45%
网络炒股	3423	6.1%	0.75%
总计	456569	809.6%	100.00%

其中，用户规模、使用率直接来源于《报告》，权数的计算公式如下：

$$R_i = \frac{A_i}{A}$$

其中， R_i 代表当前子应用的权数， A_i 代表当前子应用的使用率， A 代表各项子应用使用率之和。

最后计算该年网络应用状况(万人项)，计算方法为：

网络应用状况=∑(子应用用户规模 × 子应用权数)

由此算出2012年年网络应用状况为32625.7774万人项。以此类推，分别对每年一月《报告》中的子应用计算当年网络应用状况。经整理，如下图所示：

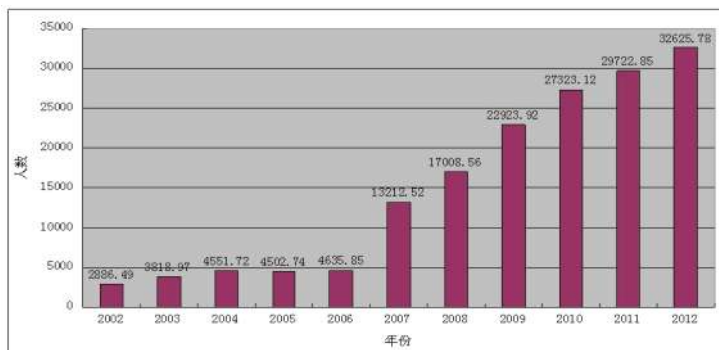


图1 2002-2012年中国互联网络应用状况条形图

最终,确定如下基本统计指标:网民规模(万人)、IPv4地址数(万个)、中国网站数量(万个)、网络国际出口带宽(Mbps)、网络应用状况(万人项)。经整理,基本数据如下表所示:

表2 2002-2012年中国互联网络基本统计指标统计数据

年份	网民规模	IPv4地址数	中国网站数量	网络国际出口带宽	网络应用状况
2002	5910.00000	2900.22000	37.00000	9380.00000	2886.48664
2003	7950.00000	4145.61000	60.00000	27216.00000	3818.96971
2004	9400.00000	5994.57280	67.00000	74429.00000	4551.72442
2005	11100.00000	7439.12960	69.00000	136106.00000	4502.74455
2006	13700.00000	9801.57440	84.00000	256696.00000	4635.84765
2007	21000.00000	13527.00000	150.00000	368927.00000	13212.52174
2008	29800.00000	18127.00000	288.00000	640287.00000	17008.56228
2009	38400.00000	23245.00000	323.00000	866367.00000	22923.91971
2010	45730.00000	27764.00000	191.00000	1098957.00000	27323.12393
2011	51310.00000	33044.00000	230.00000	1389529.00000	29722.84670
2012	56400.00000	33053.00000	268.00000	1899792.00000	32625.77742

2.2 增长指数的构建

为了反映各个基本统计指标的增长态势,同时方便后续从数量上对反映中国互联网络发展态势的各个基本统计指标进行有效整合,需要将各子指标构建为指标增长指数。

根据统计综合评价思想,进行指数设计时,考虑到互联网络发展的宏观趋势,决定采用定比计算方式来表明总体网络发展态势。计算方法如下所示:

$$P_i = \frac{x_i}{x_{2002}}$$

其中, P_i 表示该指标某年增长指数, x_i 表示该指标某年统计量, x_{2002} 表示该指标2002年(即初始统计年份)统计量。

按照以上定比计算方法分别对各个基本统计指标进行运算,数据如下表所示:

表3 2002-2012年中国互联网络基本统计指标增长指数数据

年份	网民规模	IPv4地址数	网站数量	出口带宽	网络应用状况
2002	100.00%	100.00%	100.00%	100.00%	100.00%
2003	134.52%	142.94%	162.16%	290.15%	132.31%
2004	159.05%	206.69%	181.08%	793.49%	157.69%
2005	187.82%	256.50%	186.49%	1451.02%	155.99%
2006	231.81%	337.96%	227.03%	2736.63%	160.61%
2007	355.33%	466.41%	405.41%	3933.12%	457.74%
2008	504.23%	625.02%	778.38%	6826.09%	589.25%
2009	649.75%	801.49%	872.97%	9236.32%	794.18%
2010	773.77%	957.31%	516.22%	11715.96%	946.59%
2011	868.19%	1139.36%	621.62%	14813.74%	1029.72%
2012	954.31%	1139.67%	724.32%	20253.65%	1130.29%

3 二次数据收集与整理:问卷调查与权重确定

在此前分析得到的数据的基础上,尝试建立一个全新的综合评价指数,代表中国互联网络发展水平,即中国互联网络发展态势指数。基本思想是对以上增长指数表中的数据进行加权归总。但是,权数如何才能更加科学的予以确定?决定通过进行校内

问卷调查的方式进行统计估算。

3.1 问卷的概要与具体内容

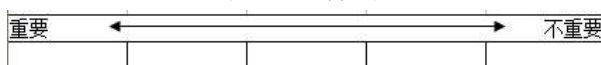
在安徽师范大学数学计算机科学学院内部发放50份调查问卷(30份以上为大样本事件),有效回收47份。在层次分析法基本原理的基础上对于方法进行简化设计,确定问卷数据分析思路,同时设计如下问卷内容(此处略去调查问卷的开头语):

表4 问卷中统计指标及其编号

统计指标	编号
网民规模	1
IPv4地址数 (如:192.168.0.10)	2
中国网站数量	3
网络国际出口带宽 (即国家的互联网主干光纤的出口带宽)	4
网络应用情况 (包括:微博、论坛、网上交易等)	5

填表说明:请凭您的经验,将以上统计指标按对互联网发展的重要性进行排序。(填写编号即可)

表5 问卷填写栏



3.2 对于问卷的统计分析

按重要性从大到小将表五中填写的统计指标从左到右进行赋权,分别为5、4、3、2、1。对收回的47份有效问卷运用Excel软件进行录入。经整理,计算得出每项基本统计指标的平均权数如下表所示(精确到小数点后四位有效数字):

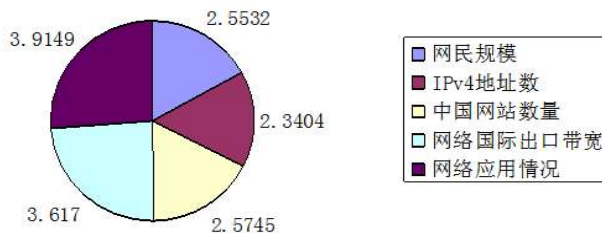


图2 统计指标平均权重

将以上数据进行整理,借鉴层次分析法基本原理,将上图中的指标两两比较,如果前者减后者为正数,则填入下表相应空格,并在相对应的后者减前者的空格中填入该正数的倒数。经计算,形成下面的分析表格:

表6 中间权数处理表

前者 \ 后者	网民规模	IPv4地址数	中国网站数量	网络国际出口带宽	网络应用情况
网民规模	1.0000	0.2128	47.0000	0.9400	0.7344
IPv4地址数	4.7000	1.0000	4.2727	0.7833	0.6351
中国网站数量	0.0213	0.2340	1.0000	0.9592	0.7460
网络国际出口带宽	1.0638	1.2766	1.0426	1.0000	3.3571
网络应用情况	1.3617	1.5745	1.3404	0.2979	1.0000

在表6的基础上,将每一纵列的项目加总,然后将纵列每项分别除以该加总值。之后,在合计栏填入前五项数据处理结果。将合计栏每项数据加总求和,其纵列每项数据除以其加总值即为该基本统计指标的最终权数。通过对数据进行处理,形成如表格7。

3.3 对于权数的自相关检验

表六中,由于n<15,使用DW检验法会使检验结果不可信,将使用图示检验法对表中数据进行检验。

通过Eviews 7,分别对各个子权数两两之间进行最小二乘回归,以前者为解释变量,后者为被解释变量,共分为四组:网民规模与IPv4地址数、网民规模与网站数量、出口带宽与网民规模、网民规模与应用情况。分别绘制四组被解释变量的残差图,可以大致看出图像变化无规律,并不呈现典型的锯齿形或循环形状变化,基本判定各子权数间不存在自相关。

表7 最终权数处理表

前者 \ 后者	网民规模	IPv4 地址数	网站数量	出口带宽	应用情况	合计	最终权数
网民规模	0.1227	0.0495	0.8599	0.2362	0.1135	1.3818	27.6359%
IPv4 地址数	0.5769	0.2327	0.0782	0.1968	0.0981	1.1827	23.6537%
网站数量	0.0026	0.0545	0.0183	0.2410	0.1153	0.4316	8.6320%
出口带宽	0.1306	0.2970	0.0191	0.2512	0.5187	1.2166	24.3316%
应用情况	0.1671	0.3663	0.0245	0.0748	0.1545	0.7873	15.7467%

4 中国互联网络发展态势指数

4.1 回归模型

在单独研究部分基本统计指标之后,接下来便是确定中国互联网络发展态势指数。该指数计算公式如下所示:

$P_i = 27.6359\% \times V_i + 23.6537\% \times W_i + 8.6320\% \times X_i + 24.3316\% \times Y_i + 15.7467\% \times Z_i$ 其中, P_i 为当年中国互联网络发展态势指数, V_i 为当年网民规模增长指数, W_i 为当年 Ipv4 地址增长指数, X_i 为当年中国网站数量增长指数, Y_i 为当年网络国际出口带宽增长指数, Z_i 为当年网络应用状况增长指数。

经计算,历年中国互联网络发展态势指数如下图所示:

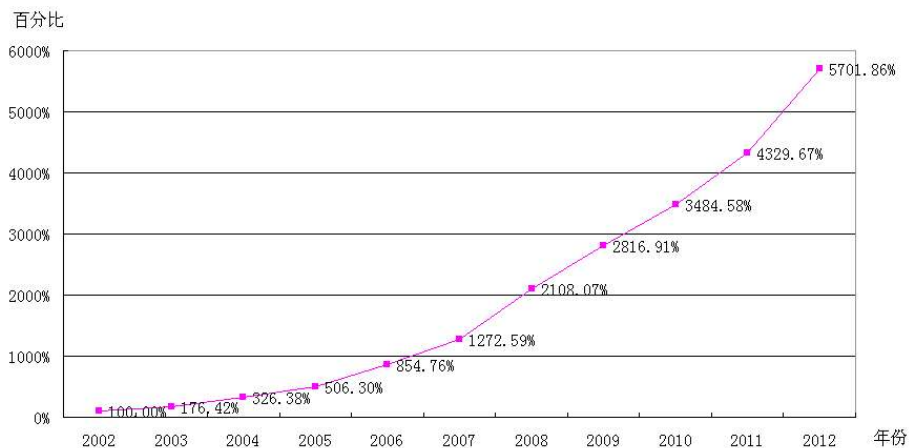


图3 中国互联网络发展态势指数折线图

接下来利用回归分析,对中国互联网络发展态势指数进行回归分析。

借助 Eviews 7 软件,先后尝试了双对数模型、半对数模型、多项式模型、迭代估计法。根据以上折线图的形态,怀疑该回归模型为二次函数,在尝试以上建模方法后,最终发现多项式模型(二次函数模型)相对简洁且回归效果最好。

设中国互联网络发展态势指数可以用序号的二次多项式近似表示,其中序号为 2002 至 2012 年的 1 至 11 个数字,即:

$$y = b_0 + b_1x + b_2x^2 + \epsilon$$

经 Eviews 7 计算,如下图所示:

Dependent Variable: INDEX
Method: Least Squares
Date: 02/13/13 Time: 21:11
Sample: 2002 2012
Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.426332	1.193886	2.032298	0.0766
NUM	-1.670129	0.457268	-3.652409	0.0065
NUM^2	0.593506	0.037114	15.99156	0.0000

R-squared	0.997360	Mean dependent var	19.70685
Adjusted R-squared	0.996700	S.D. dependent var	18.92431
S.E. of regression	1.087122	Akaike info criterion	3.231946
Sum squared resid	9.454676	Schwarz criterion	3.340463
Log likelihood	-14.77570	Hannan-Quinn criter.	3.163541
F-statistic	1511.143	Durbin-Watson stat	2.087578
Prob(F-statistic)	0.000000		

图4 中国互联网络发展态势指数最小二乘法回归结果

得到网络发展态势指数的估计式:

$$\hat{y} = 2.426332 - 1.670129x + 0.593506x^2$$

$$t = (-3.652409) (15.99156)$$

对以上函数求导数,得到估计式为:

$$\frac{d\hat{y}}{dx} = -1.670129 + 1.187012x$$

因此,每过去一年,中国互联网络发展态势指数就增加1.187012。

4.2 模型检验

接下来,通过以下七方面对以上回归结果进行检验:

4.2.1 实际意义检验

近年来,我国互联网呈现高速发展态势。在反映互联网发展的不同方面,如网民规模、网络地址数、网站数量、网络国际出口带宽等均出现较大幅度的增长。同时,实际生活中,互联网发展速度越来越快,使我们有理由相信,随着时间的推移,中国互联网络发展态势指数呈现高速增长态势,且其增长幅度也必将越来越快。

4.2.2 R²检验

可决系数 R²=0.997360,修正可决系数 $\bar{R}^2 = 0.996700$ 。一方面,增加该多项式模型为3次与4次,则 \bar{R}^2 分别为0.996310与0.995883;减少该多项式模型为1次,则 $\bar{R}^2 = 0.903299$ 。可见,变动该估计式的次数,都会导致修正可决系数的减小。另一方面,R²与 \bar{R}^2 都十分接近于1。可见,R²检验效果很好。

4.2.3 F 检验

设显著性水平 $\alpha = 0.01$, Prob(F-statistic)=0.000000 < 0.01= α , 即有99%的把握认为解释变量联合起来对中国互联网络发展态势指数的影响是显著的。

4.2.4 截距项的显著性检验

设显著性水平 $\alpha = 0.10$, Prob=0.0766 < 0.10= α , 即有90%的把握认为截距项对中国互联网络发展态势指数的影响是显著的。

4.2.5 x 的显著性检验

设显著性水平 $\alpha = 0.01$, Prob=0.0065 < 0.01= α , 即有99%的把握认为截距项对中国互联网络发展态势指数的影响是显著的。

4.2.6 x² 的显著性检验

设显著性水平 $\alpha = 0.01$, Prob=0.0000 < 0.01= α , 即有99%的把握认为截距项对中国互联网络发展态势指数的影响是显著的。

4.2.7 标准误检验

由回归结果可以看出,3个可变系数的标准误都很小,可以看出各变量回归效果很好。

4.3 模型预测

利用 Eviews 7,对将来3年的中国互联网络发展态势指数进行预测,如下图所示:

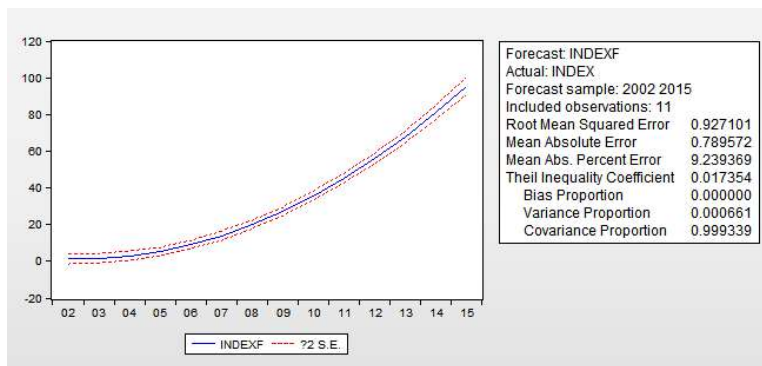


图5 中国互联网络发展态势指数及其预测图

预测结果如下表所示:

表8 2013-2015年中国互联网络发展态势指数预测表

年份	2013	2014	2015
预测值	65	77	95

从中也可以看出,对于2013-2015年指数的线性预测,基本符合2002-2012年指数的增长趋势,具有良好的线性回归效果。

5 成果应用

5.1 过去10余年中国互联网络发展历程回顾

纵观全球互联网发展,我国起步较晚。然而近十年来,我国互联网事业得到跨越式发展。据历年中国互联网络发展状况统计报告,网民规模同比增长954.31%、IPv4地址数同比增长1139.67%、中国网站数量、网络国际出口带宽、网络应用状况分别同比增长724.32%、20253.65%、1130.29%。另据崔海东《互联网建设十年回顾》一文中指出,无论是在网络建设、网络优化、网络承载技术,还是在路由设备、流量工程方法,甚至是对互联网系统性质认识上都有了巨大的飞跃与发展。这显示,经过10年高速发展,2012年我国互联网建设相比于2002年取得重大突破,互联网发展实现了历史性跨越。

5.2 中国互联网络未来发展态势的影响

5.2.1 对互联网产业的影响

互联网产业是以互联网为依托、以信息技术为主要支撑手段的现代服务业。随着中国互联网发展态势指数的不断攀升,中国互联网产业也呈现了不断攀升的发展趋势。根据艾瑞咨询统计的数据显示,2005年中国网络经济市场规模为146亿元;2012年,中国网络经济市场规模达3850.4亿元,同比增长54.1%;从季度来看,2012年第四季度,中国网络经济市场规模为1167.6亿元,同比增速和环比增速分别为50.5%和13.3%。三网融合、电子商务、微博等受到人们的广泛关注。可以预见,互联网产业将在未来的经济、社会生活等领域发挥出不可估量的作用。

5.2.2 对大学生生活的影响

随着中国互联网的飞速发展,互联网已经几乎走进了每一个大学生的生活。他们在互联网上冲浪,获得大量的信息。大学生作为网民规模的重要组成部分,其上网呈现高速发展态势。互联网在促进大学生学习教育上,通过多媒体与网络辅助等技术,合作学习,以掌握面对信息爆炸时代所必须的知识与技能。同时,网络的多样化与动态性也能很好地锻炼大学生的独立思考能力与实践能力。互联网的发展在提升全民特别是年轻一代的素质都起到了不可磨灭的积极作用。

5.2.3 对社会文化的影响

中国互联网的飞速发展,从数量上即中国互联网络发展态势指数的快速增长,且其增长速度越来越快,这也带动了大众社会生活的较大变化。在信息生活上,体现了全民媒体化的发展态势:论坛、个人博客、微博、SNS的兴起与发展,促进了媒体由传统媒体向自媒体的转变,使得信息传播更加私人化、平民化与自主化。在社区生活中,互联网发展态势指数的不断提高也导致了虚拟世界的出现与发展。网络不断发展所带来的松散性、对话性、广泛性的特点使得虚拟世界出现了去中心化的发展态势。主体相对平等、限制相对较少,促进互联网资源共享与决策公开的发展进程。

5.3 互联网发展瓶颈探析

在研究中我们发现,IPv4地址数已经成为制约互联网络发展的瓶颈。根据通过调查得到的五大统计指标权重中,排名第三的是IPv4地址数。纵观互联网发展历程,会发现,IPv4的数量已经远远不能满足网络发展的需求。近几年来IPv4的地址即将耗尽,其主要原因在于Internet用户使用Internet连接的移动设备,以及连接Internet的ADSL modem、有线modem的爆炸式增长,使得IPv4地址迅速趋向于枯竭,迫使不得不开发和采用一种新的Internet协议——IPv6作为替代解决方案。在应对数量枯竭危机上,IPv6可以将地球上每一粒沙子都赋予一个地址。可以有效解决互联网终端设备移动化,多样化的发展趋势。因此,加快IPv4地址向IPv6地址的过渡,成为促进互联网络健康、可持续发展的重要途径。

5.4 研究意义

研究中,我们以官方统计数据为依托,以校内问卷调查为依据,运用层次分析基本原理与统计回归方法,提出中国互联网络发展态势指数的概念,促进了中国互联网络发展的总体态势从定性化研究向量化研究的转变。该指数模型可以较有效的对未来互联网络发展趋势进行预测,具有良好的信度和效度。同时,研究成果也可以被用于政府部门以及企业等机构的决策参考。

参考文献:

- [1] CNNIC.中国互联网络发展状况统计报告[R].CNNIC,2002-2012.
- [2] 陈科.综合评价中几种确定权数的方法[J].中国卫生统计,1993,10(3):38-40.
- [3] 靳庭良.计量经济学[M].成都:西南财经大学出版社,2011,3 :66-69.
- [4] 李莉.管理定量分析 决策中常用的分析方法[M].上海:上海交通大学出版社, 2007 :184-196.
- [5] 傅德印,张旭东.EXCEL与多元统计分析 附实用计算机程序[M].北京:中国统计出版社, 2007 :83.
- [6] 崔海东.互联网建设十年回顾[J].电信工程技术与标准化,2005(9):1-5
- [7] 艾瑞咨询集团.网络经济核心数据[EB/OL].[2005-2012]. <http://www.iiresearch.com.cn/>.
- [8] 刘中望.媒介新技术:互联网与当代生活方式[J].湘潭大学学报(哲学社会科学版),2010(1):95-97.