

产业结构调整的量化的分析

徐 沈^{1,2}, 吴 涛^{1,2,3}, 李国成^{1,2}

(1. 安徽大学 智能计算与信号处理教育部重点实验室, 安徽 合肥 230039;

2. 安徽大学 数学与计算科学学院, 安徽 合肥 230039;

3. 南京大学 计算机软件新技术国家重点实验室, 江苏 南京 210093)

摘 要: 产业结构内部各大产业的协调发展对稳定和促进 GDP 增长的作用巨大, 由于资源的稀缺性, 如何将有限的社会资源合理分配以协调各大产业的发展成为十分重要的问题。文中引入粗糙集的属性重要性的概念和数据挖掘的属性相关分析法, 对安徽省和全国的产业结构指标数据作综合分析, 比较分析了三大产业 GDP 增长的影响程度。结果表明第一和第三产业对安徽省国内生产总值影响较大, 而第二产业对全国 GDP 的增长有突出重要的作用。最后提出了调整和优化安徽产业结构的一些建议。

关键词: 产业结构; 粗糙集; 属性相关分析

中图分类号: TP18; F121.3

文献标识码: A

文章编号: 1673-629X(2009)08-0178-04

Quantitative Analysis on Adjustment of Industrial Structure

XU Shen^{1,2}, WU Tao^{1,2,3}, LI Guo-cheng^{1,2}

(1. Ministry of Education Key Lab. of IC & SP, Anhui University, Hefei 230039, China;

2. School of Mathematical Sciences of Anhui University, Hefei 230039, China;

3. State Key Lab. for Novel Software Technology, Nanjing University, Nanjing 210093, China)

Abstract: The coordinated development of the major industries has played an important role on the tremendous growth of GDP, however, due to the scarcity of resources, how to allocate social resources rationally in order to participate in the coordination of various industries' development become quite an important issue. For this reason, after introducing the significance of attributes in rough sets and the property-related analysis of data mining, take Anhui for example, an analysis on the indicator for industrial data is put forward. The results show that the primary industry and service sector are more important to the GDP growth in Anhui Province and the second industry is the most important one of the three major industries to national GDP growth. At last, some recommendations for the adjustment and optimization of industrial structure in Anhui are discussed.

Key words: industrial structure; rough set; property-related analysis

0 引言

产业结构, 亦称国民经济的部门结构, 指国民经济各产业部门之间以及各产业部门的内部构成。产业结构政策作为我国产业政策的主要内容之一, 目的在于调整和优化产业结构以发展和带动整个国民经济的发展^[1,2]。调整和优化产业结构就是调整产业结构的

化比例, 使得整个国民经济也相应的按比例进行。目前的很多文献和资料仅仅从定性的角度说明了调整产业结构对 GDP 增长的理论分析, 甚至提出了一些理论假说^[3,4]; 即便是采用定量方法, 也是参照统计学和计量经济学方法作近似分析或是结合现有文献关于产业结构的变动对经济增长影响的测度方法作综合分析^[4,5]。这些定量方法都不同程度地介入了很多人为的因素, 使得分析结果缺乏客观性, 某些时候还具有片面性。粗糙集理论除数据外, 不需要任何先验信息, 客观地挖掘隐藏在数据中的信息。基于此, 文中将引入粗糙集及其属性相关分析法的相关概念, 结合安徽各产业发展的指标数据, 对改革开放 20 多年来安徽省产业结构内部各大产业的重要性给出量化结果; 并通过与全国相同产业的数据对比来综合评价安徽同全国

收稿日期: 2008-12-18; 修回日期: 2009-03-19

基金项目: 中国博士后基金面上项目(20070411028); 国家 973 计划(2004CB318108, 2007BC311003); 国家自然科学基金(60675031); 安徽省自然科学基金项目(KJ2008B093, KJ2008B452C); 安徽大学人才队伍建设项目

作者简介: 徐 沈(1983-), 男, 安徽芜湖人, 硕士研究生, 研究方向为信息粒度计算及其在经济中的应用; 吴 涛, 博士, 教授, 从事机器学习、智能计算、粒度计算及其应用的研究。

的差距;最后,对安徽省产业结构的调整和优化方式提出一些建议。

1 粗糙集方法

粗糙集理论是一种新的处理模糊和不确定性知识的数学工具,其中心思想是在保持分类能力不变的前提下,通过知识约简导出问题的决策或分类规则^[6-8]。

定义1 一个信息表知识表达系统 S 可以表示为 $S = \langle U, R, V, f \rangle$, 这里 U 是论域,即样本全体, $R = B \cup D$ 是属性集合,子集 B 和 D 分别称为条件属性集和决策属性集, $V = \bigcup_{r \in R} V_r$ 是属性值的集合, V_r 表示属性 $r \in R$ 的属性值范围,即属性 r 的值域, $f: U \times R \rightarrow V$ 是一个信息函数,它指定中每一个对象 X 的属性值。

定义2 给定知识库 $K = (U, B)$, 对于每个子集 $X \subseteq U$ 和一个等价关系 $R \in \text{ind}(K)$, 定义两个子集: $\underline{R}X = \bigcup \{Y \in U/R \mid Y \subseteq X\}$, $\overline{R}X = \bigcup \{Y \in U/R \mid Y \cap X \neq \emptyset\}$ 。分别称它们为 X 的 R 下近似集和 R 上近似集。

定义3 设 P 和 Q 为 U 中的等价关系, Q 的 P 正域记为 $\text{pos}_P(Q)$, 即 $\text{pos}_P(Q) = \bigcup_{x \in U/Q} PX$ 。其含义是 U 中所有根据分类 U/P 的信息可以准确地划分到关系 Q 的等价类中去的对象集合。

定义4 设 Q 的 P 正域 $\text{pos}_P(Q)$ 包含了 n 个等价类, 每个等价类包含了 i 个样本, 表示为:

$\text{pos}_P(Q) = \{e_1, e_2, \dots, e_n\}; e_i, i = 1, 2, \dots, n$, 为 $\text{pos}_P(Q)$ 的某个等价类, 包含有 c_i 个样本。

因此, 定义 $\text{pos}_P(Q)$ 的阶为 $|\text{pos}_P(Q)| = \sum_i c_i^2$ 。

定义5 信息表 S 中, 条件属性列 $A \subset B$ 相对于决策属性 D 的重要度定义为:

$$\text{sgn}_A(D) = \frac{|\text{pos}_B(D)| - |\text{pos}_{B \setminus \{A\}}(D)|}{|U|}$$

2 属性相关分析方法

属性相关分析度量是计算某种度量, 用于量化属性与给定类或概念的相关性, 以帮助识别不相关或弱相关属性。这种度量包含信息增益、Gini 索引、不确定性和相关系数^[9,10]。文中采用的信息增益分析方法, 并结合经济分析的实际, 给出如下定义:

设 S 是训练样本的集合, 其中每个样本的类标号是已知的, 记为第 C_i 类。事实上, 每个样本是一个元组, 一个属性用于确定训练样本的类, 每个类都有完全相同的属性。现假定有 m 个类, S 包含 s_i 个 C_i 类样本, $i = 1, 2, \dots, m$ 。确定其中的一个类为分析的目标类,

而其他的类均为对比类, 则对于一个任意样本属于类 C_i 的可能性是 s_i/s , 其中 s 是集合 S 中对象的总数。

定义6 对于一个给定的样本分类所需的期望信息是:

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m \frac{s_i}{s} \log_2 \frac{s_i}{s}$$

具有值 $\{a_1, a_2, \dots, a_v\}$ 的条件属性列 A 可以用来将 S 划分为子集 $\{S_1, S_2, \dots, S_v\}$, 其中, S_j 包含 S 中 A 值为 a_j 的那些样本。

定义7 设 S_j 包含类 C_i 的 s_{ij} 个样本, 根据 A 的这种划分的期望信息称作的 A 熵, 计算公式是:

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + s_{2j} + \dots + s_{mj}}{s} I(s_{1j} + s_{2j} + \dots + s_{mj})$$

定义8 根据上述 A 划分而获得的信息增益为:

$$\text{Gain}(A) = I(s_1, s_2, \dots, s_m) - E(A)$$

Gain 值越大, 说明目标类和对比类的数据在属性列上的相关性越强, 数值较为接近, 因而两者差距越小, 反之亦然。

3 产业结构影响 GDP 的分析步骤

在综合分析大量原始数据的基础上提出了一种分析解决每个产业对整个国民经济的影响程度的新方法, 该方法撇开单就某个年度的经济数据分析, 综合考虑多年来的大量数据, 其分析结果较之原来的计量经济分析更为全面和客观。步骤如下:

S1. 输入原始的目标类数据和对比类数据, 分别以二维信息表的形式表示;

S2. 对两类数据采用指数化处理方法进行无量纲化处理, 使得结果数据介于 0 和 1 之间;

S3. 采用等距离划分法对标准化后的两类数据作属性的离散化处理, 文中对目标类数据和对比类数据均作 0-1 划分;

S4. 将所有属性值均相同的样本子集进行合并, 合并的结果单列为属性 count, 其值等于相同样本子集的个数, 得到目标类和对比类的简化决策信息表;

S5. 利用定义 4 计算步骤 4 的决策信息表的属性子集“第 i 产业”($i = 1, 2, 3$) 的属性重要度 $\text{sgn}_i(D)$, $\text{sgn}_i(D)$ 值越大, 表明第 i 产业对 GDP 的增长影响越大;

S6. 利用定义 6, 7 和 8 给出的属性相关分析方法计算属性子集 i (第 i 产业) 的信息增益 $\text{Gain}(i)$;

S7. 结合 $\text{sgn}_i(D)$ 和 $\text{Gain}(i)$ 的值的大小, 综合分析第 i 产业 ($i = 1, 2, 3$) 对整个 GDP 发展的影响程度, 进而得出如何从量上加以协调的结论。

4 产业结构影响 GDP 的量化分析

在经济研究和经济管理中,常用的有两大部类分类法、三次产业分类法、资源密集度分类法和国际标准产业分类法,而我国采用的是三次产业分类法,其划分标准是:产品直接取自自然界的部门称为第一产业,一般指农业;对初级产品进行再加工的部门称为第二产业,包含工业和建筑业;为生产和消费提供各种服务的部门称为第三产业,指除了第一、第二产业以外的其它行业。收集安徽(目标类)和全国(对比类)在 1978 年至 2005 年(共计 28 个年份)的国内生产总值指数(按照可比价格计算),如表 1 和表 2 所示。

表 1 国内生产总值指数(安徽)(上年=100)

年份	第一产业	第二产业	第三产业	人均国内生产总值
1978	96.89	115.47	91.79	98.78
1979	110.91	109.56	106.68	107.82
1980	93.74	112.17	118.41	100.38
1981	126.95	104.43	112.30	115.97
1982	102.61	113.83	125.89	108.03
1983	101.02	117.29	118.63	107.74
1984	116.25	125.27	121.86	119.01
1985	109.12	121.67	120.99	114.32
1986	108.01	112.74	115.07	109.94
1987	97.92	107.93	113.79	103.48
1988	96.10	115.22	107.61	103.82
1989	102.20	104.83	110.12	103.42
1990	104.43	107.04	94.88	100.17
1991	77.22	109.74	115.91	96.70
1992	116.94	118.70	113.72	115.22
1993	111.88	124.64	116.20	117.68
1994	102.98	117.01	122.03	113.58
1995	117.42	111.17	116.71	113.61
1996	107.82	114.32	114.53	112.49
1997	109.22	112.72	112.12	111.05
1998	102.60	109.70	110.82	107.19
1999	108.01	107.45	112.49	108.61
2000	101.20	109.49	111.46	107.57
2001	102.86	109.53	112.33	108.19
2002	101.02	111.60	113.03	109.16
2003	93.65	113.38	114.30	109.04
2004	108.60	114.67	114.18	112.54
2005	100.96	118.84	109.94	110.76

4.1 计算过程

现将三大产业产值指数和国内生产总值指数分别视为决策表的条件属性和结果属性,经过步骤 2、步骤 3 和步骤 4 处理,得出表 3 和表 4 所示的对于目标类和对比类的简化决策信息表。再经过步骤 5 得出安徽第 i 产业的属性重要度见表 5。

表 2 国内生产总值指数(全国)(上年=100)

年份	第一产业	第二产业	第三产业	人均国内生产总值
1978	104.1	115.0	113.8	110.2
1979	106.1	108.2	107.9	106.1
1980	98.5	113.6	106.0	106.5
1981	107.0	101.9	110.4	103.9
1982	111.5	105.6	113.0	107.5
1983	108.3	110.4	115.2	109.3
1984	112.9	114.5	119.3	113.7
1985	101.8	118.6	118.2	111.9
1986	103.3	110.2	112.0	107.2
1987	104.7	113.7	114.4	109.8
1988	102.5	114.5	113.2	109.5
1989	103.1	103.8	105.4	102.5
1990	107.3	103.2	102.3	102.3
1991	102.4	113.9	108.9	107.7
1992	104.7	121.2	112.4	112.8
1993	104.7	119.9	112.2	112.7
1994	104.0	118.4	111.1	111.8
1995	105.0	113.9	109.8	109.7
1996	105.1	112.1	109.4	108.9
1997	103.5	110.5	110.7	108.2
1998	103.5	108.9	108.4	106.8
1999	102.8	108.1	109.3	106.7
2000	102.4	109.4	109.7	107.6
2001	102.8	108.4	110.3	107.5
2002	102.9	109.8	110.4	108.4
2003	102.5	112.7	109.5	109.3
2004	106.3	111.1	110.1	109.4
2005	105.2	111.7	110.5	109.8

数据来源:《安徽统计年鉴 1999》、《安徽统计年鉴 2006》、《中国统计年鉴 2007》。

现假设 C_1 类对应于目标类安徽, C_2 类对应于对比类全国。 C_1 类和 C_2 类均有 28 个样本, 则对给定样本所需的期望信息是:

$$I(s_1, s_2) = I(28, 28) = -\frac{28}{56} \log_2 \frac{28}{56} - \frac{28}{56} \log_2 \frac{28}{56} = 1$$

根据解析特征化所得到的如表 3 和表 4 所示的候选关系, 计算每个属性的熵, 先对属性“第一产业”的每个值计算这两个类的期望信息: 对于“第一产业” = “0”: $S_{11} = 10, S_{21} = 21, I(S_{11}, S_{21}) = 0.9073$

对于“第一产业” = “1”: $S_{12} = 18, S_{22} = 7, I(S_{12}, S_{22}) = 0.8555$

如果样本根据“第一产业”划分, 则对给定的样本进行分类所需的期望信息是:

$$E(\text{第一产业}) = \frac{31}{56} I(s_{11}, s_{21}) + \frac{25}{56} I(s_{12}, s_{22}) = 0.8842$$

因此由这样划分的信息增益是:

$$\text{Gain(第一产业)} = I(S_1, S_2) - E(\text{第一产业}) = 0.1158$$

同理可得: $\text{Gain(第二产业)} = 0.0238$, $\text{Gain(第三产业)} = 0.1752$

表 3 目标类(安徽)

序号	第一产业	第二产业	第三产业	count
1	0	1	0	2
2	1	0	0	2
3	0	0	1	6
4	1	0	1	11
5	0	1	1	2
6	1	1	1	5

表 4 对比类(安徽)

序号	第一产业	第二产业	第三产业	count
1	0	1	1	7
2	1	0	0	4
3	1	0	1	2
4	0	1	0	6
5	1	1	1	1
6	0	0	1	1
7	0	0	0	7

表 5 各产业的属性重要度 $\text{sgn}_i(D)$

$\text{sgn}_i(D)$	$\text{sgn}_1(D)$	$\text{sgn}_2(D)$	$\text{sgn}_3(D)$
安徽 (目标类)	0.429	0.238	0.333
全国 (对比类)	0.250	0.500	0.250

4.2 初步结论

通过对三大产业属性重要度 $\text{sgn}_i(D)$ 的计算,得出在安徽对国内生产总值影响较大的是第一和第三产业,这是因为首先安徽是一个农业大省,其 GDP 的增长主要依赖于农业的基础支撑作用;其次,新型工业化进程加快,然而安徽的第三产业却比第二产业更能影响 GDP 的增长,在保持国民经济在总体上稳步增长上起着举足轻重的作用,尤其是服务业的发展和壮大满足了社会的多层次、多元化需求,这一结论与参考文献[11]中对三大产业对经济的影响方面的判断大体一致。

从整体情况看,全国第二产业对全国 GDP 的影响却表现了突出重要的作用,其属性重要度的值明显高于其他产业。

三大产业的 Gain 值可以看出,第一产业和第三产业同全国的发展水平相关性较强,而第二产业却明显

相关性较弱,这说明安徽省的产业结构同全国的相比较具有一定的差距,表现在安徽的第一、第三产业高于全国水平,而第二产业却低于全国比重,这同参考文献[4]中对安徽产业结构的分析结果相接近。

5 对调整和优化安徽产业结构的一些建议

充分协调和把握三大产业之间的量的比例关系则有利于促进 GDP 指数快速增长,结合以上分析,尝试提出在当前形势下的安徽如何调整和优化产业结构的一些建议,具体有以下几个方面:

(1)继续保持农业农村经济的基础性作用,第一产业的属性重要度最高和我省是农业大省的现状决定了今后一段时间内,我省的第一产业仍将发挥对 GDP 增长的支撑作用。

(2)坚持以信息化带动工业化,进一步发挥二、三产业对安徽经济增长的带动作用,尤其是加快发展安徽的第三产业。

(3)由于安徽第二产业的 Gain 值在三大产业中是最小的,可见和全国工业企业的发展水平还存在着较大的差距,因此应加快本省的工业化建设步伐,至少应赶上全国的整体水平。换句话说,如果差距过大,不仅不利于工业反哺农业政策目标实现,而且也对全国范围内统一市场的形成造成不小的障碍。

(4)从全国的三大产业的属性重要度来看,第二产业对 GDP 的影响是最大的,而安徽却恰恰相反,所以安徽应积极引进发达城市工业企业的先进技术,加大对周边城市的招商引资,这样会缩小第二产业同国内平均水平的差距。

总之,安徽是经济较为落后的大省,其对产业结构的正确调整需要决策者根据全国经济形势的变化找出差距并随时作出调整。

文中引入的通过计算机领域的相关知识来对改革开放以来安徽省三大产业产值数据所作的简要分析,其优势主要在于分析方法的客观性和完整性,分析结论也符合经济发展的现实状况,在实际应用中可行。

参考文献:

[1] 方 超.安徽省产业结构与经济增长关系实证研究[J].资源开发与市场,2008,24(8):722-725.
[2] 张艺影.产业结构与经济增长相互促进作用机理分析[J].商业时代,2008(19):80-80.
[3] 刘建党.产业结构变化及其对经济增长影响的理论分析[J].北方经济:综合版,2008(13):30-32.
[4] 柯 军.安徽省产业结构升级对经济增长的贡献分析[J].安徽工业大学学报:社会科学版,2008,25(1):39-40.

入得当,就既可能在 $f = f_0$ 时使电压放大倍数增大,又不会因正反馈过强而产生自激振荡。

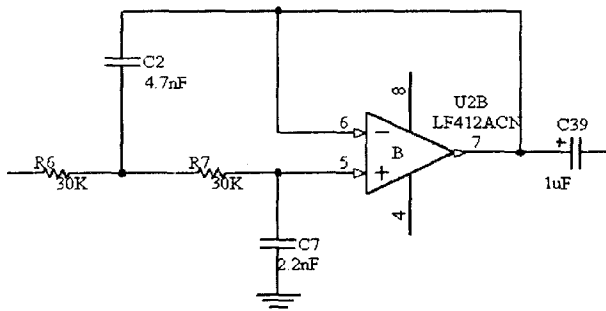


图6 低通滤波环节

其传递函数

$$A_u(s) = \frac{A_{up}(s)}{1 + [3 - A_{up}(s)]sRC + (sRC)^2} \quad (6)$$

在式(6)中,只有当 $A_{up}(s)$ 小于3时,即分母中 s 的一次项系数大于零,电路才能稳定工作,而不产生自激振荡。

2.2.3 二值化处理环节

图7所示为将滤波后得到的E、F低频信号过零比较,进行二值化处理,得到数字信号^[7]。

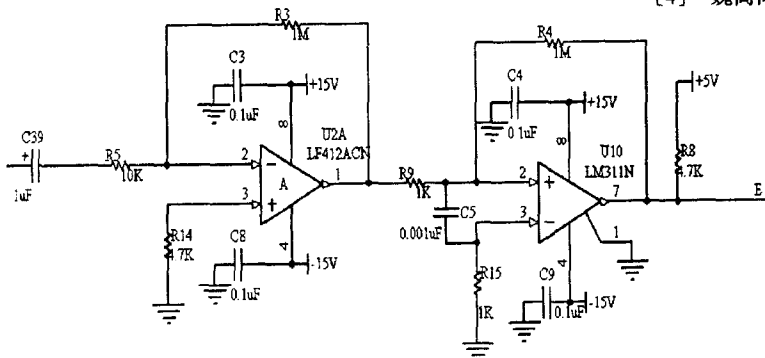


图7 二值化处理环节

2.2.4 数字相位比较器

数字相位比较器是E-F相位检测系统的最重要部分。因为FPGA可在线编程、软件仿真,具有较高的灵活性,便于修改和调试,因此在电路中采用FPGA器

件作为相位比较器。根据E-F相位信号检测原理,利用QUARTUS软件^[8]设计了相位比较方案。

3 结束语

介绍了DVD光学头的E-F相位差信号的检测原理,提出了E-F相位差信号的检测方法,并给出了实现方案。

实验结果表明,基于清华大学自主开发的数字伺服系统平台,在CD一倍速模式下伺服系统处于聚焦闭环、循迹开环的条件下,采用现场可编程门阵列(FPGA)实现的E-F相位差检测系统可以正确检测出DVD光学头的E-F相位差参数。

参考文献:

- [1] 徐端颐. 光盘存储系统设计原理[M]. 北京: 国防工业出版社, 2000.
- [2] 徐端颐. 高密度光盘数据存储[M]. 北京: 清华大学出版社, 2003.
- [3] 120mm DVD-Read-Only Disk. Standard ECMA-267 [M]. 3rd Edition. [s.l.]: [s.n.], 2001.
- [4] 魏高尚. DVD光学头伺服误差信号分析[J]. 应用激光, 2004, 24(4): 237-239.
- [5] 马建设. DVD光学头参数评价方法及测试装置的研究[R]. 北京: 清华大学精密仪器与机械学系, 2003.
- [6] 方楠. DVD光学头评价仪伺服系统的设计研究[D]. 北京: 清华大学精密仪器与机械学系, 2005.
- [7] Ma Jianshe, Fang Nan, Zhou Hui, et al. Test and evaluation of characteristic parameters in error detectors of DVD pickup[J]. Journal of Optoelectronics • Laser, 2002, 13(3): 237-239.
- [8] Altera International Limited. QUARTUS II User Manual [M]. [s.l.]: [s.n.], 2006.

(上接第181页)

- [5] 曹斌. 产业结构变动对经济增长影响的测度方法综述[J]. 兰州商学院学报, 2005, 21(5): 72-78.
- [6] 王国胤. Rough集理论与知识获取[M]. 西安: 西安交通大学出版社, 2001.
- [7] 张文修, 吴伟志. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.
- [8] Pawlak Z. Rough set theory and its applications to data analysis[J]. Cybernetics and Systems: An International Journal,

1998, 29: 661-681.

- [9] 韩家炜, 坎伯. 数据挖掘概念与技术[M]. 范明, 等译. 北京: 机械工业出版社, 2001.
- [10] Devore J, Peck R. Statistics: The Exploration and Analysis of Data[M]. New York: Duxbury Press, 1997.
- [11] 王伟, 高亮, 吴涛. 粗糙集在经济分析中的应用[J]. 计算机技术与发展, 2008, 18(4): 158-160.