

doi:10.3969/j.issn.1672-626x.2017.06.006

基于共词分析法的我国产业结构研究热点解析

李慧慧^{1,2}

(1. 北方民族大学 管理学院, 银川 750021; 2. 合肥工业大学 管理学院, 合肥 230009)

摘要:基于共词分析法,分析了27495篇产业结构相关论文中出现的26467个关键词,统计出63个高频关键词。从词频分析可以看出,“就业结构”“就业”“第三产业”“产业结构升级”“工业化”等关键词出现频率很高,是产业结构领域研究中的热点内容。同时,通过构建高频关键词的共词矩阵、相似矩阵、相异矩阵并运用多元统计分析发现,我国产业结构研究主要集中在第三产业、就业结构、城镇化进程这三个领域。

关键词:产业结构;共词分析;产业发展

中图分类号:F269.24;F224

文献标识码:A

文章编号:1672-626X(2017)06-0042-06

目前,我国产业结构处于重要的调整转型期,一方面,国际金融危机导致外需急剧波动,对外依存度过高的风险日益凸显;另一方面,内需也在不断萎缩,同时产能过剩、人口红利消失、生产要素成本上升等问题不断凸显。要保持经济可持续发展,就必须调整产业结构,推进产业转型升级。为此,我国学者在产业结构调整方面进行了大量研究,涌现了大批研究成果。本文拟利用共词分析法探究2012—2016年间我国产业结构研究热点,以期更加直观和科学地反映我国产业结构研究的热点领域。

一、方法简介与数据来源

共词分析来源于文献计量学的引文耦合与共被引概念^[1]。1986年,法国国家科学研究中心CNRS(Centre National de la Recherche Scientifique)的M. Callon、J.Law和A.Rip出版了第一部关于共词分析的专著——《Mapping the Dynamics of Science and Technology》^[2]。经过几十年的发展,该方法逐步在医学、人工智能等不同领域得到了广泛应用,相关领域的学者利用该方法概述了各领域的研究热点、发展趋势等。

共词分析法主要是利用文献集中词汇对或名词短语共同出现的情况来确定该文献集所代表学科中各主题之间的关系。一般认为词汇对在同一篇文献中出现的次数越多,则代表这两个主题的关系越紧密^[3]。通过这种方法可以提取所需研究领域的文献主题词,统计高频词汇,然后利用数学统计将词汇关系转化为数值或者图形。

本文从中国知网(CNKI)中的《中国期刊全文数据库》《中国硕博学位论文全文数据库》《会议论文》三个数据库中,以“产业结构”为关键词,以2012年1月—2016年12月为检索时间段,共检索到论文27495篇。通过CNKI以ENDNOTE格式提取检索到文献的题录信息。

二、关键词的确定与分析

将从CNKI中导出的题录信息导入SATI3.2文献题录信息统计分析软件,经过去重处理,共有373218条记录。然后抽取关键词进行统计,共得到关键词26467个。为了最大限度消除我们对关键词人为合并的主观影响,体现和反映作者群体对特定关键词的共识度,对关键词并未做同义词合并处

收稿日期:2017-07-06

基金项目:北方民族大学校级科研一般项目(GLXY201602)

作者简介:李慧慧(1982-),女,黑龙江牡丹江人,北方民族大学管理学院讲师,合肥工业大学管理学院博士研究生,研究方向为产业经济、人力资源管理。

理,仅仅去掉“对策”“影响”“措施”等非实质性词语,同时,因为本文以“产业结构”为检索词,所以“产业结构”在共词分析中无法发挥作用,故删除。最终得到关键词26064个。

利用SATI软件对26064个关键词进行了词频统计,选定词频高于100次的63个关键词作为本文研究的主要关键词。部分高频关键词见表1。

表1 部分高频关键词

序号	关键词	频次
1	经济增长	1205
2	经济发展	716
3	产业结构升级	472
4	低碳经济	465
5	经济发展方式	453
6	就业结构	435
7	新兴产业	423
8	区域经济	391
9	城镇化	381
10	产业结构优化	366
11	优化	339
12	可持续发展	277
13	产业转型升级	247
14	节能减排	238
15	FDI	237
16	主导产业	230
17	产能过剩	220
18	产业升级	220
19	发展	219
20	碳排放	209

然后,对这63个关键词进行两两共词检索,统计它们在27495篇论文中出现的次数,形成了63*

63的共词矩阵,部分矩阵如表2。

共词矩阵是一个相关矩阵,对角线上的数字是该词出现的频次,如“经济增长”共出现1205次,它与“新兴产业”同时在29篇论文中出现过。从关键词共词矩阵可以看出,产业结构的研究内容相对分散,说明学者们对产业结构的研究角度存在一定的差异性。

由于本文将运用多元统计分析方法进行相关分析,所以需将共词矩阵转化为相似矩阵和相异矩阵以满足不同的统计方法对数据的要求,于是利用SATI软件中的相似矩阵和相异矩阵功能对共词矩阵进行了转换。SATI3.2为消除多值共现矩阵中频次悬殊对统计结果造成的影响,采用Equivalence系数,将多值矩阵转化为元素值在[0,1]区间取值的相似矩阵^[4],部分相似矩阵如表3,具体公式为:

$$E_{ij} = F_{ij}^2 / F_i * F_j$$

其中, E_{ij} 为相似矩阵元素的值,对于词条 T_i 和 T_j , F_{ij} 为 T_i 和 T_j 的共现次数, F_i 为 T_i 出现的总频次, F_j 为 T_j 出现的总频次。

相似矩阵中的数字代表矩阵元素之间的相似性,两词之间的数字越大,越接近1,表示两词之间的距离越近,相似度越大;两词之间的数字越小,越接近0,表示两词之间的距离越远,相似度越小。对角线上的数据为1,表示该词自身的相关度。

由于相似矩阵中的0值过多,容易造成较大误差,因此用1与相似矩阵的数据相减的值作为相异矩阵元素的值,生成相异矩阵,部分相异矩阵如表4。相异矩阵中的数字为不相似数据,数值越大表示两词之间的距离越远,越不相似;数值越小表示两词之间的距离越近,相似度越大。

表2 部分高频关键词共词矩阵

	经济增长	经济发展	产业结构升级	低碳经济	经济发展方式	就业结构	新兴产业	区域经济	城镇化
经济增长	1205	52	18	5	23	16	29	20	16
经济发展	52	716	5	8	2	10	39	19	8
产业结构升级	18	5	472	5	0	3	1	6	10
低碳经济	5	8	5	465	3	1	1	8	1
经济发展方式	23	2	0	3	453	0	55	10	1
就业结构	16	10	3	1	0	435	0	1	5
新兴产业	29	39	1	1	55	0	423	9	0
区域经济	20	19	6	8	10	1	9	391	2
城镇化	16	8	10	1	1	5	0	2	381

表3 部分高频关键词相似矩阵

	经济增长	经济发展	产业结构升级	低碳经济	经济发展方式	就业结构	新兴产业	区域经济	城镇化
经济增长	1	0.0031	0.0006	0	0.001	0.0005	0.0016	0.0008	0.0006
经济发展	0.0031	1	0.0001	0.0002	0	0.0003	0.005	0.0013	0.0002
产业结构升级	0.0006	0.0001	1	0.0001	0	0	0	0.0002	0.0006
低碳经济	0	0.0002	0.0001	1	0	0	0	0.0004	0
经济发展方式	0.001	0	0	0	1	0	0.0158	0.0006	0
就业结构	0.0005	0.0003	0	0	0	1	0	0	0.0002
新兴产业	0.0016	0.005	0	0	0.0158	0	1	0.0005	0
区域经济	0.0008	0.0013	0.0002	0.0004	0.0006	0	0.0005	1	0
城镇化	0.0006	0.0002	0.0006	0	0	0.0002	0	0	1

表4 部分高频关键词相异矩阵

	经济增长	经济发展	产业结构升级	低碳经济	经济发展方式	就业结构	新兴产业	区域经济	城镇化
经济增长	0	0.9969	0.9994	1	0.999	0.9995	0.9984	0.9992	0.9994
经济发展	0.9969	0	0.9999	0.9998	1	0.9997	0.995	0.9987	0.9998
产业结构升级	0.9994	0.9999	0	0.9999	1	1	1	0.9998	0.9994
低碳经济	1	0.9998	0.9999	0	1	1	1	0.9996	1
经济发展方式	0.999	1	1	1	0	1	0.9842	0.9994	1
就业结构	0.9995	0.9997	1	1	1	0	1	1	0.9998
新兴产业	0.9984	0.995	1	1	0.9842	1	0	0.9995	1
区域经济	0.9992	0.9987	0.9998	0.9996	0.9994	1	0.9995	0	1
城镇化	0.9994	0.9998	0.9994	1	1	0.9998	1	1	0

三、多元统计分析

多元统计分析是统计学中内容十分丰富、应用范围极为广泛的一个分支。而在共词分析中,最常用的多元统计分析方法主要有因子分析、聚类分析和多维尺度分析。本文将利用因子分析和多维尺度分析对高频关键词进行分析,进而揭示产业结构研究的现状和热点。

1. 因子分析。因子分析是利用降维的思想,通过研究众多变量之间的内部依赖关系,探求观测数据中的基本结构,并用少数几个抽象的变量来表示其基本的数据结构。其目的就是用少数几个因子去描述许多指标或者因素之间的联系,即将相互关系比较密切的几个变量归在同一类别之中,每一类变量就成为一个因子^[5]。将共词矩阵导入 SPSS21,得到表5。

从表5可以看出,共提取了5个公共因子,其累计方差贡献率为88.89%,其中第二、三、四个因子的方差贡献率较高,分别为64.08%、6.21%和10.84%,这是国内产业结构研究的三个重要领域。总体来看,我国产业结构研究呈现总体分散、相对集中的特点,这与前文统计的关键词词频结果基本一致。

从因子个数碎石图(见图1)也可以看出,从第6个公因子开始,曲线开始变平,而第四个与第五个的贡献率不高并且差别不大。根据因子具有命名解释性的特点,本文对5个公共因子分别命名(见表6)。

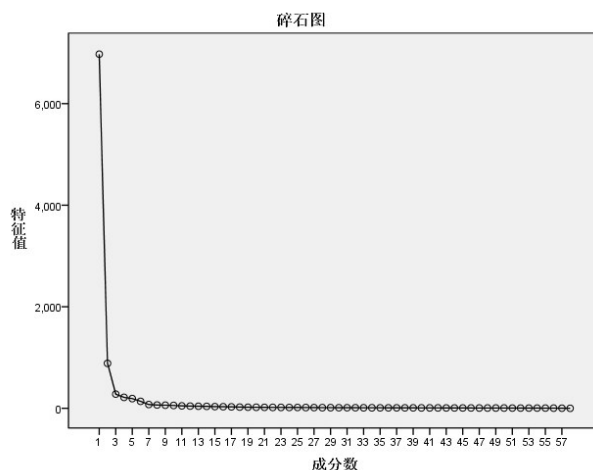


图1 碎石图

关键词的荷载量反映了关键词在其公共因子中的相关度。根据本文的实际情况,将特征值定为1。表6显示,63个关键词中共有39个参与了分类,这可能由于有些关键词尽管频次很高,但涵盖的主

表5 解释的总方差(部分)

初始特征值 ^a			提取平方和载入			旋转平方和载入		
合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
6974.200	72.537	72.537	6974.200	72.537	72.537	354.853	3.691	3.691
885.868	9.214	81.750	885.868	9.214	81.750	6161.261	64.082	67.772
279.296	2.905	84.655	279.296	2.905	84.655	597.437	6.214	73.986
215.349	2.240	86.895	215.349	2.240	86.895	1042.066	10.838	84.824
191.783	1.995	88.890	191.783	1.995	88.890	390.879	4.065	88.890
137.725	1.432	90.322						
74.131	.771	91.093						
65.517	.681	91.775						
61.540	.640	92.415						
56.105	.584	92.998						
47.677	.496	93.494						
42.930	.447	93.941						
40.298	.419	94.360						
38.094	.396	94.756						
33.849	.352	95.108						

表6 因子分析确定的研究热点

因子名称	关键词	分质	因子名称	关键词	分质
三次产业	第三产业	0.964	城市化	城市化	0.882
	产业结构调整	0.392		工业化	0.749
	劳动力市场	0.163		经济结构	0.155
	第二产业	0.72		经济发展	0.163
	农村剩余劳动力	0.139		服务业	0.107
	从业人员	0.428		城市化进程	0.146
	第一产业	0.614		城市化水平	0.182
	三次产业	0.219		农民收入	0.148
	经济结构调整	0.199		就业	0.993
	劳动力就业	0.212		劳动力	0.25
产业结构	城市化进程	0.161	经济增长	劳动力转移	0.194
	就业比重	0.413		失业	0.465
	产业结构	0.944		大学生	0.418
	就业弹性	0.454		经济增长	0.965
	结构偏离度	0.662		就业增长	0.233
	劳动者	0.149		技术进步	0.181
	比较劳动生产率	0.387		劳动就业	0.196
	城市化进程	0.152		劳动生产率	0.237
	实证分析	0.253			
	结构调整	0.276			
大学生就业	0.174				
产值结构	0.277				
新疆	0.303				
产业结构偏离度	0.402				

上,实现两者之间的有机结合及良性互动,是我国经济社会发展面临的今后面临的重大课题,也是学者们今后的一个重点研究方向。

参考文献:

- [1] 储节旺,郭春侠.共词分析法的基本原理及EXCEL实现[J].情报科学,2011,(6):931-934.
- [2] Callon M, Law J, Rip A. Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World[M]. Macmillan, 1986: 89.
- [3] 贺颖,邱均平.同行评议专家遴选的科学计量方法与实证研究[J].图书情报工作,2012,(3):33-37.
- [4] 刘启元,叶鹰.文献题录信息挖掘技术方法及其软件SATI的实现——以中外图书情报学为例[J].信息资源管理学报,2012,(1):50-58.
- [5] 赵飞娟.西安市创新型科技人才素质模型构建与实证研究[D].西安:西安邮电学院,2011.
- [6] 张凌,周春雷,寇广增.基于共词分析的国内知识服务研究[J].图书情报工作,2009,(12):64-68.
- [7] 贾利军,管静娟,杨娅蓓.我国就业能力的研究热点及其解析[J].华东师范大学学报(哲学社会科学版),2013,(5):24-31.
- [8] 郭晓川.内蒙古新型城镇化、产业结构合理化协调发展研究[J].内蒙古大学学报(自然科学版),2017,(7):1-10.

(责任编辑:彭晶晶)

Research Focus on the Industrial Structure Based on Co-word Analysis

LI Hui-hui^{1,2}

(1. School of Management, BeiFang University of Nationalities, Yinchuan 750021, China; 2. School of Management, HeFei University of Technology, HeFei 230009, China)

Abstract: Based on the co-word analysis method, this paper analyzes 4067 keywords in 2732 articles related to industrial structure, and counts 58 high frequency keywords. From the frequency analysis, we can see that the key words such as "Employment Structure", "Employment", "upgrading of an industrial structure", "Urbanization" and "Industrialization" are very frequent and are hot topics in industrial structure research. At the same time, through the construction of co-word matrix, similarity matrix and dissimilarity matrix of high-frequency keywords, we find that China's industrial structure is mainly concentrated in the tertiary industry, employment structure and urbanization process by multivariate statistical analysis.

Key words: industrial structure; co-word analysis; industrial development