

我国海洋经济领域研究主题的共词分析

武群芳¹, 郭鑫¹, 王继民^{1*}

(1. 北京大学 信息管理系, 北京 100871)

摘要:随着国家“一带一路”发展战略的提出,发展海洋经济成为重中之重。文章以中国知网上与“海洋经济”相关的博硕士论文题录信息为数据来源,利用共词分析、聚类分析等研究方法,对论文的关键词进行词频统计,建立高频关键词的共词矩阵;通过层次聚类将海洋经济领域的研究主题划分为海洋产业结构优化、渔业资源与可持续发展、蓝色经济区等9类研究主题;并结合战略坐标图对这些类的密度和向心度进行分析与探讨,得出研究主题地域性差异较明显、成熟研究话题与新兴研究热点并存的结论,以期对海洋经济理论与应用研究提供参考,促进我国海洋经济的持续快速发展。

关键词:海洋经济;共词分析;系统聚类;战略坐标

1 引言

海洋蕴含着巨大的发展潜力,海洋资源的开发越来越引起世界各国的重视,海洋经济成为一个国家或地区发展的重要增长点。2013年10月,习近平总书记访问东盟时提出了“21世纪海上丝绸之路”的战略构想,与“丝绸之路陆上经济带”合称为“一带一路”,成为我国未来的重要发展战略。随后,沿海各省市积极响应中央的战略决策,依据自身情况和优势,迅速出台了相关的政策和办法,发展海洋经济助力我国经济增长。

国内外学者对海洋经济进行了大量的研究,尤其是“一带一路”战略出台后,关于海洋经济的研究更是不断涌现。近年来,国内学者从全国和区域两个层面、海洋经济核算、国外发展海洋经济对我国的借鉴作用等方面进行了大量研究,既指出发展海洋经济可以促进我国经济的发展,同时也指出海洋经济发展中的负面影响,为进一步推动我国海洋经济研究提出了可行性建议^[1]。鉴于此,本文通过收集中国知网(CNKI)中与海洋经济相关的高质量博硕士论文,利用共词分析与层次聚类方法,对海洋经济领域的研究主题进行梳理,探析该领域的研究热点,以期能够对海洋经济的研究现状和研究热点进行揭示,为我国海洋经济理论与应用的深入研究提供一定的参考和借鉴,助力我国海洋经济的持续快速发展。

2 数据收集与预处理

2.1 数据收集

本次研究选取中国知网(CNKI)收录的博硕士论文作为文献信息源,因为博硕士论文普遍来说质量较高,所研究的内容也更加深入。我们以“主题”为检索项,以“海洋经济”为检索词,时间跨度限制

作者简介:武群芳(1991—),女,山东省烟台市人,硕士,主要从事文献计量、数据挖掘研究。

*通信作者:王继民,教授,主要从事数据挖掘、海洋发展指数研究。

为 2005—2015 年，检索时间为 2015 年 4 月 4 日，检索得到近 10 余年以海洋经济为研究主题的硕、博士论文共计 1 246 篇。之后，我们对这批论文题录信息的字段缺失情况进行核查，结果显示这批论文的关键词不存在缺失现象，所以共词研究的分析对象即为 1 246 篇论文的关键词字段。

2.2 词频统计与词汇预处理

我们利用自编计算机程序，对这批博硕士论文的关键词字段进行词频统计，得到其中的高频关键词。观察发现，在高频关键词中有一些较为宽泛、通用的词汇，不能有效揭示学科主题，包括“海洋”、“影响”、“发展”、“研究”等等。此外，还有一些关键词存在一义多词现象，即不同的词汇表达的是相同或相近的含义，进行共词分析时有必要将其进行合并，比如“评价指标体系”和“指标体系”、“环渤海”和“环渤海地区”，等等。

针对上述问题，我们对这批论文的关键词进行以下的词汇预处理：

(1) 删去不具代表性，不能反映具体研究内容的关键词，本次研究删除的关键词有：海洋经济、蓝色经济、海洋、中国、影响、发展、研究、优化。

(2) 人工制定映射规则，并将规则应用到原始题录信息中，进行关键词替换。本次研究共制定了 26 条映射规则，部分映射规则如表 1 所示。

表 1 关键词映射规则（部分）

原始关键词	替换后关键词	原始关键词	替换后关键词
评价指标体系	指标体系	因子分析法	因子分析
评价体系	指标体系	微卫星标记	微卫星
对策	对策建议	浙江	浙江省
生态环境	海洋生态环境	舟山群岛	舟山
区域经济	海洋区域经济	环渤海	环渤海地区

进行词汇预处理后，需要对替换之后的关键词字段重新进行词频统计，并选取其中的高频关键词进行后续研究。在本次研究中，我们选取词频排名前 49 的关键词（词频≥7）作为研究对象，预处理之后的部分高频关键词如表 2 所示。

表 2 预处理后的高频关键词（部分）

关键词	词频	关键词	词频
海洋产业	61	因子分析	25
可持续发展	52	山东省	23
蓝色经济区	44	海洋环境	19
舟山	37	海洋区域经济	18
指标体系	36	辽宁省	17
对策建议	31	遗传多样性	16
海洋产业结构	27	发展战略	14

2.3 构造共词矩阵

在上一节中，我们选取了 49 个高频关键词，目的是为了根据关键词题录信息建立一个 49×49 的高频词共现矩阵。共现矩阵对角线上的值表示某个关键词出现的次数，非对角线上的值表示词 A 和词 B 在同一篇文章的关键词字段中共同出现的次数。本次研究的部分高频词共现矩阵如表 3 所示。

表 3 高频词共现矩阵 (部分)

高频词	海洋产业	可持续发展	蓝色经济区	舟山
海洋产业	61	2	2	2
可持续发展	2	52	1	3
蓝色经济区	2	1	44	0
舟山	2	3	0	37

2.4 构造相关矩阵和相异矩阵

在实际共词分析过程中, 关键词共现频次受各自词频大小的影响。为了能够准确揭示关键词之间的共现关系, 我们采用 Ochiai 系数将共词矩阵转换为相关矩阵。对于 w_1 和 w_2 两个关键词, 其 Ochiai 系数的计算公式为:

$$\text{Ochiai}(w_1, w_2) = w_1 \text{ 和 } w_2 \text{ 的共现次数} / \sqrt{(w_1 \text{ 出现次数} \times w_2 \text{ 出现次数})}$$

根据以上公式, 可以计算得到高频关键词的相关矩阵, 部分矩阵如表 4 所示。

表 4 高频词相关矩阵 (部分)

高频词	海洋产业	可持续发展	蓝色经济区	舟山
海洋产业	1	0.035 5	0.038 6	0.042 1
可持续发展	0.035 5	1	0.020 9	0.068 4
蓝色经济区	0.038 6	0.020 9	1	0
舟山	0.042 1	0.068 4	0	1

可以看到, 相关矩阵对角线上的值均为 1, 非对角线上的值均在 $[0, 1]$ 区间内。其中, 该数值越接近 1 表示两个关键词联系越强; 反之, 该数值越接近 0 表示两个关键词的联系越弱。使用数值 1 减去相关矩阵中的各个元素, 可以得到高频词相异矩阵, 部分矩阵如表 5 所示。

表 5 高频词相异矩阵 (部分)

高频词	海洋产业	可持续发展	蓝色经济区	舟山
海洋产业	0	0.964 5	0.961 4	0.957 9
可持续发展	0.964 5	0	0.979 1	0.931 6
蓝色经济区	0.961 4	0.979 1	0	1
舟山	0.957 9	0.931 6	1	0

相应地, 在相异矩阵中, 元素的值越接近 1 说明两个关键词距离越远, 联系越强, 数值越接近 0 说明两个关键词的距离越近联系越强。利用相异矩阵, 我们可以进行后续的聚类分析。

3 研究方法

3.1 聚类分析

根据数据对象的特征属性, 聚类分析可以将数据对象集合划分为若干个不同的类, 使得同一类中的数据对象具有较大的相似性, 不同类中的数据对象具有较大的相异性^[2]。在进行共词分析时, 可以将高频词相异矩阵导入 SPSS 中进行系统聚类。本次研究将 49 个关键词聚成 9 类, 聚类结果如表 6 所示。

表 6 关键词聚类结果

类别	关键词	类别	关键词
第 1 类	海洋产业	第 4 类	舟山
	海洋产业结构		对策建议
	海洋新兴产业		海域使用权
	主导产业		无居民海岛
	产业竞争力		海洋功能区划
	辽宁省		数值模拟
	海岸带		重金属
第 2 类	可持续发展	第 5 类	海洋环境
	指标体系		治理
	因子分析		福建省
	评价		滨海旅游
	层次分析法	第 6 类	遗传多样性
	海洋渔业		微卫星
	海岛		大黄鱼
第 3 类	海洋渔业资源	第 7 类	发展战略
	聚类分析		SWOT 分析
	蓝色经济区		海洋文化产业
	山东省		影响因素
	海洋区域经济	第 8 类	海权
	产业集群		海洋文化
	海洋资源		海洋强国
	海洋生态环境		海洋战略
	协调发展	第 9 类	浙江省
	环渤海地区		海洋旅游
	海洋开发		

3.2 战略坐标图

战略坐标可以概括地呈现一个领域或亚领域的结构，它将每一个研究主题放置到直角坐标系的四个象限中，并据此描述各个主题的发展状况。在战略坐标图中，有两个较为重要的概念：密度和向心度。密度用来衡量各个类别之内的主题词内部联系的强度，反映了一个研究领域自身发展的情况；向心度则用来衡量某一类别的主题词与其他类别的主题词之间的紧密程度，反映了一个研究领域和其他领域相互影响的程度^[3]。对于研究主题的密度、向心度的计算有不同的方法，本文采用的计算公式为：

$$\text{密度} = \frac{2 \times \sum_{i,j \in K, i \neq j} E_{ij}}{n}, \quad \text{向心度} = \frac{\sum_{i \in K, j \notin K} E_{ij}}{n},$$

式中， E_{ij} 是关键词 i 和关键词 j 共现的次数， K 代表通过系统聚类得到的某一类别， n 是该类别所含关键词的数目。

利用上一节所述的聚类结果和高频词共现矩阵，结合上述计算公式，可以计算得到每个类别的密度和向心度。之后，将密度和向心度的数值进行 Z-score 规范化，根据规范化结果可以绘制“海洋经济”领域的战略坐标图，具体如图 1 所示。

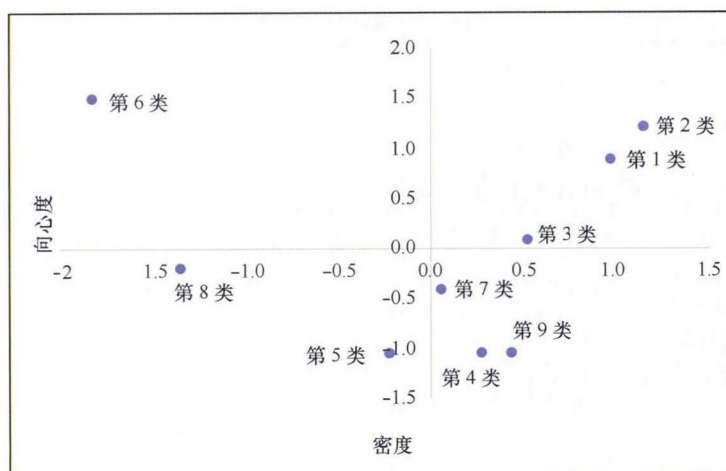


图1 海洋经济领域战略坐标图

4 结果分析

4.1 聚类分析

根据 3.1 节中的聚类结果, 共聚为 9 类, 分别是: 海洋产业结构优化、渔业资源和可持续发展、蓝色经济区、海域使用权、海洋环境治理、生物遗传多样性、海洋文化产业、海权与海洋战略、海洋旅游。下文将结合相关文献数据的具体内容对这 9 类的研究主题热点进行分析和概述。

(1) 海洋产业结构优化。该类的主要关键词有: 海洋产业、海洋产业结构、海洋新兴产业、主导产业、产业竞争力、辽宁省等。我国海洋经济虽然得到了一定程度的发展, 但是与发达国家相比还有很大的差距, 以海洋产业结构的优化带动海洋经济的可持续发展。部分论文探讨了区域海洋主导产业的选择问题, 还需优化产业空间布局, 提升海洋产业竞争力, 形成以主导产业为核心、各产业协调发展的海洋产业体系。辽宁省委省政府提出了建设“海上辽宁”战略目标, 海洋产业的发展和实现路径成为热点话题, 研究指出辽宁省海洋产业结构变动与辽宁省经济增长的关系很密切, 实现海陆产业经济一体化, 积极培育战略性新兴产业, 加强港口建设, 发挥区域比较优势^[4], 形成具有特色的沿海经济区。

(2) 渔业资源与可持续发展。该类的关键词主要有: 海洋渔业资源、可持续发展、指标体系、因子分析、层次分析法、聚类分析等。渔业资源是海洋资源的重要组成部分, 目前我国各海域的渔业资源遭到过度捕捞, 海洋渔业资源尤其是近海渔业资源应该合理使用可持续发展; 同时, 我国还存在渔业资源性资产流失问题。大量论文探讨了实现渔业资源可持续发展的方法和途径, 提出了可行性建议; 此外还有研究可持续发展的评价问题, 目前国内外还没有比较完善的方法, 研究运用了主成分分析法、因子分析法、层次分析法和聚类分析法等多种方法建立指标评价体系, 确定指标权重, 实现评价指标体系的科学性和实用性, 可持续发展评价方法还需进一步深入研究。

(3) 蓝色经济区。该类的主要关键词有: 蓝色经济区、山东省、海洋区域经济、产业集群、海洋生态环境、协调发展、环渤海地区等。作为沿海重要省份的山东省, 于 2011 年获得了山东省半岛蓝色经济区的国家级战略发展规划, 蓝色经济区成为一个研究热点。很多论文从区域海洋经济的角度, 探讨了不同区域和城市在蓝色经济区的定位和职能, 以沿海各市为主导, 如环渤海地区, 是我国海洋经济的重要组成部分, 经济基础良好, 应该起到山东蓝色经济区桥头堡的作用, 建设中国海洋科技教育中心、沿海高科技产业带、现代化港口以及海洋文化产业、海洋旅游业等, 带动山东整体经济的发展^[5]。此外, 还有论文探讨了蓝色经济区域产业集群、海洋产业复杂网络以及海洋生态环境保护和协调发展问题。

(4) 海域使用权。该类的主要关键词有: 海域使用权、海洋功能区划、无居民海岛、对策建议、舟

山等。沿海区域拥有丰富的陆海自然资源,是经济最为发达的区域,但正面临着生态环境压力,需要规范海域使用权,合理安排海洋功能区划。有论文研究了海洋生态环境监测数据的获取和分析,能够及时了解海洋生态环境现状及动态变化趋势,对海洋突发事件进行预警预测,从而保护海洋生态环境;也有很多研究从法律的角度探讨了合理使用沿海区域的对策建议,健全法律法规,依法发展海洋经济;此外,舟山地区的无居民海岛的使用和管理也是研究的重点。

(5) 海洋环境治理。该类的主要关键词有:海洋环境、治理、福建省、滨海旅游。治理海洋环境,需要促进海洋产业结构的优化升级,该类论文主要探讨了福建省及其他沿海省份海洋环境的恶化和治理。其中,福建省在滨海旅游中具有自身的优势和特色,可有效带动海洋经济的发展,滨海旅游业将成为未来旅游经济和海洋经济的新亮点,也将有效改善当地的海洋环境。

(6) 生物遗传多样性。该类有3个关键词:遗传多样性、微卫星、大黄鱼,主要是生物学科的研究。大黄鱼是我国重要的海洋经济鱼类,由于过度捕捞,大黄鱼资源严重枯竭,目前我国大黄鱼面临着生长变缓等种质退化的问题,对大黄鱼养殖业造成致命打击。研究利用微卫星DNA分子标记等手段,对其繁殖行为、受精生物学、遗传多样性以及繁殖模式等方面进行了研究,旨在提高大黄鱼种质资源的改良选育。还有个别论文运用微卫星技术对我国其他海洋生物遗传多样性进行了研究。

(7) 海洋文化产业。该类的关键词有4个,分别是:海洋文化产业、发展战略、SWOT分析、影响因素。海洋文化产业是具有知识产权高附加值的产业,在海洋经济中发挥着越来越大的作用,学术界对海洋文化产业给予了诸多关注,海洋文化产业正逐步成为各国经济发展新的增长点,但是相关理论研究还很薄弱,未成为成熟的产业。本类主题的论文主要对我国文化产业的发展现状、影响因素、发展战略、潜力指数、具体发展路径等进行了探讨,很多研究运用了SWOT分析法进行分析,提出具有现实意义的对策和建议。

(8) 海权与海洋战略。该类的主要关键词有4个,分别是:海权、海洋战略、海洋强国、海洋文化。海洋作为战略屏障,对一国的安全有重要的战略意义,对海洋的争夺将成为世界政治斗争的中心议题之一。海权是国家主权的重要组成部分,该类论文探讨了海权的产生和含义的发展演变;此外,还分析了中国的海洋权益和定位,应据此制定中国自身海洋发展战略,从海洋大国向海洋强国跨越式发展,如积极发展海洋文化增强软实力等。

(9) 海洋旅游。该类只有2个关键词:浙江省和海洋旅游。海洋旅游作为海洋经济的重要组成部分,具有很大的发展潜力和空间。目前“浙江海洋经济示范区”、“浙江舟山群岛新区”已写入国家海洋发展战略规划之中,且随着近年来浙江海洋经济强省建设、旅游经济强省建设和“海上浙江”建设的推进,更为浙江海洋旅游业的快速发展提供了良好的政策环境。该类论文围绕着浙江省海洋旅游,主要探讨了旅游产业融合,创建旅游新业态、健全和创新海洋旅游管理体制等问题。

4.2 战略坐标图分析

战略坐标图(见图1)显示,第1类海洋产业结构优化、第2类渔业资源与可持续发展和第3类蓝色经济区均处于第一象限,其中第1类和第2类的密度、向心度都远高于其他类别。这说明海洋产业、海洋渔业资源和海洋可持续发展等主题在海洋经济领域的研究已经非常成熟,且与海洋经济的其他研究主题密切相关。海洋产业结构的优化是伴随着我国整体经济结构调整提出的,海洋产业结构优化既是目标也是方法途径,不管是发展海洋旅游还是海洋文化产业,都是经历了海洋产业结构的优化升级,提高我国海洋产业的竞争力;海洋渔业是海洋经济的传统领域,对其研究一直很多,随着可持续发展要求的提出,渔业资源的循环利用也成为了热门话题,因此这两类研究比较成熟完备,且居于核心的位置。第6类是唯一一个位于第二象限的主题类,该类拥有极高的密度和极低的向心度。可以看到,这一类的研究主题是利用微卫星技术研究大黄鱼的遗传多样性,主要属于生物学的研究范畴,该主题内部发展非常成熟,但处于海洋经济这一学科的边缘位置。第四象限包括第4类海域使用权、第7类海洋文化产业和第9类海洋旅游这3个类,其研究主题分别为海域使用问题、海洋文化产业的发展战略、浙江省海洋旅游,这3个类虽然向心度

较高,但密度低于各研究主题的平均水平,这也就意味着这 3 个主题和海洋经济领域的其他研究主题联系较为紧密,处在该领域较为核心的位置;但同时这 3 个主题自身的发展还不够成熟,虽然具有一定的发展潜力但仍然有待后续的探索。第 5 类海洋环境治理、第 8 类海权与海洋战略处于战略坐标图的第三象限,其对应的研究主题分别为福建省的海洋环境治理、海权和海洋强国战略,这 2 个研究主题的密度和向心度都很低,说明类团内部成员之间联系松散,且与其他类的联系也不紧密,可能会分解演化至其他类中。

海洋强国战略和“一带一路”战略的实施和推进,学界对于海洋经济的研究势必会迎来新的发展。随着相关研究的不断深入,这种结构关系会发生一定的变化:边缘结构中的关键词可能会进入核心结构,而核心结构中的词也可能会退出,进入边缘结构^[6]。

5 结语

本文使用共词分析方法,对以“海洋经济”为主题的优质博硕士文献进行了直观、科学的分析,并进行了一些讨论与解读,具有一定的现实意义。从关键词的共词矩阵出发,将其转化为相关矩阵和相异矩阵,利用 SPSS 进行层次聚类分析,最终获得九大研究主题,分别是:海洋产业结构优化、渔业资源和可持续发展、蓝色经济区、海域使用权、海洋环境治理、生物遗传多样性、海洋文化产业、海权与海洋战略、海洋旅游;在此基础上,根据聚类结果绘制了战略坐标图,对每一个主题的成熟程度、核心程度等进行了分析。进而得出以下结论:

第一,海洋经济研究的地域性差异比较明显,往往不同地区讨论的侧重点不同。研究主题主要受当地政策引导和自然条件的影响,如山东省独有的蓝色经济区,浙江省的小黄鱼遗传多样性研究。应该考虑到各地区各省市的特点,结合当地需紧迫解决的问题,加强经济政策的针对性和引导性,^[1]鼓励学术创新;不同地区之间也应相互借鉴和学习,实现海洋区域经济共同发展。

第二,通过聚类和战略坐标分析发现,传统成熟的研究话题与新兴研究热点并存,且成熟研究话题仍然居于核心地位,如海洋产业结构优化、渔业资源和可持续发展,仍然有很多研究成果,是海洋经济的热点话题,在今后的研究中仍然需要得到关注;新的研究热点包括海洋文化产业、海洋旅游、海域使用权等,是较为具体的子话题,目前发展还不够成熟,研究的角度也不全面,在未来应该得到重视。

需要说明的是,本研究属于探索性研究,因此难免存在一定的局限性。首先数据本身还不够完善,由于缺乏该领域的标准主题词表,我们对关键词的映射规则不完全规范,论文发表也有一定的时滞性,使得关键词列表存在一定的偏差;其次如何确定高频关键词的阈值仍是有待讨论的问题,因为低阈值不利于聚类,但有助于一些隐含主题或前瞻主题的外现,而高阈值则恰好相反,容易损失信息,造成理解有误或不完整^[7],今后有待改进。

参考文献:

- [1] 张举钢,张彦彦. 我国海洋经济研究综述[J]. 港口经济, 2013 (8): 42—45.
- [2] Han J W, Kamber M, et al. 数据挖掘:概念与技术(原书第 3 版)[M]. 范明,孟小峰,译. 北京:机械工业出版社,2012.
- [3] 崔鹏,孙宝文,王天梅,等. 基于共词分析的网络虚拟社会领域热点及演进态势研究[J]. 情报杂志, 2013 (2): 41—44.
- [4] 高学文,许欣欣. 辽宁省海洋经济发展问题研究[J]. 现代经济信息, 2011 (4): 254—255.
- [5] 冯瑞. 蓝色经济区研究述评[J]. 东岳论丛, 2011, 32(5): 189—191.
- [6] 魏瑞斌,王三珊. 基于共词分析的国内 Web2.0 研究现状[J]. 情报探索, 2011(1): 1—5.
- [7] 马费成,望俊成,陈金霞,等. 我国数字信息资源研究的热点领域:共词分析透视[J]. 情报理论与实践, 2007, 30(4): 438—443.