

doi:10.3772/j.issn.2095-915x.2016.04.012

基于共词分析的微生物学植物学领域研究 热点分析

杨莲莲¹, 杨之音², 杨朝峰¹

(1. 中国科学技术信息研究所 北京 100038; 2. 吉林农业大学信息技术学院 长春 130118)

摘要: 为了解农业领域中微生物学、植物学领域的研究热点及走向, 本文采用共词分析的方法, 通过 Excel、SPSS 等软件, 对 CNKI 数据库中 2004-2014 年间 1604 篇相关文献进行统计。通过分析得出高频关键词并建立共词矩阵。进一步对高频词共词矩阵进行聚类分析和多维尺度分析, 认为植物、微生物领域研究热点主要集中在基因方面的研究进展、植物的应用分析、微生物的物理分析、生物技术及特性分析等四大主题上。其中在基因方面的研究进展属于最热门的研究领域。

关键词: 微生物学, 植物学, 共词分析, 聚类分析, 多维尺度分析

中图分类号: G35

Research on the Hotspots of Microbiology and Botany Based on the Co-Word Analysis

YANG LianLian¹, YANG ZhiYin², YANG ChaoFeng¹

(1. Institute of Scientific and Technical Information of China, Beijing 100038, China; 2. Information and Technology College, Jilin Agricultural University, Changchun 130118, China)

Abstract: In order to understand the research hotspots and trends of microbiology and botany in the agricultural field, we conducted the co-word analysis based on 1604 related literatures which are published during 2004 to 2014 from the CNKI by adopting the software, such like SPSS, Excel. We calculated the high frequency

基金项目: 本文受吉林省教育厅教育教学研究重点课题: 吉林省高等院校教育科研文献信息支撑平台构建与实施研究, 吉林农业大学科研启动基金资助项目: 区域内高校图书馆资源共享平台构建研究 (2015044) 资助。

作者简介: 杨莲莲 (1991-), 硕士研究生; 杨朝峰 (1975-), 副研究员, 研究方向: 科技政策、经济政策等; 杨之音 (1983-), 博士, 研究方向: 数字资源建设, wenjianzy@126.com。

keywords and established the co-word matrix through the analysis. Furthermore, this study applied the cluster analysis and multidimensional scaling analysis to the co-word matrix of high-frequency words. According to the analyzed results, we concluded that the research hotspots in the field of botany and microbiology mainly focused on four major themes: the research progress of gene, application analysis of plants, physical analysis of microorganism, biological technology and characteristic analysis, and the research progress of gene was the most popular topic among these research fields.

Keywords: Microbiology, botany, co-word analysis, clustering analysis, multidimensional scaling analysis

微生物学、植物学是农业领域的重要研究方向，它们包含的类别比较多，并且是现代高新生物技术的理论基础和技术基础，因此，对该领域的研究具有一定的代表性。本文通过共词分析法对该领域的研究热点进行分析，从而更好地为相关领域的研究提供理论依据。共词分析法可以发现研究对象之间的亲疏关系，挖掘隐含的或潜在的有用知识，并揭示某一学科或学科群的研究热点、结构与范式^[1]。已有研究发现，用来描述文章内容的关键词，事实上是相关科学概念可以信赖的一个指标，它们可以反映科学研究的现状^[2]。因此，本文选用关键词作为共词分析的对象。为了更加形象地展示微生物学、植物学领域的研究热点，利用聚类分析进一步计算关键词的距离，根据距离远近对高频关键词进行分类，从而发现学科的研究热点^[3]。同时对共词矩阵进行多维尺度分析，将共词矩阵中的共现词对的相似程度在低维度空间中用点与点之间的距离表示出来，以形成体现这些关键词之间相关性的共词图谱。

1 数据来源与研究方法

1.1 数据来源

本文的数据来源于 CNKI 数据库，CNKI 是

全球最大的中文知识门户网站，目前已经实现了国内 25% 的知识资源的数字化和网络化共享，基本能够满足人们的知识需求。本文选取农业科技类目下农业微生物学、植物学领域在 2004—2014 年间发表的相关文章，共检索到文献 1604 篇。将这些文献中的关键词按 500 篇一组分别导入 Excel 表格中，然后将表格中的空白行进行批量删除，并去掉表格中的空格形式。然后将所有的数据源整理到一张表上，并将个别长度大于 10 列的数据去掉，最终得到关键词的基础数据 4704 个。

1.2 研究方法

本文主要采用共词分析法来探讨农业领域中微生物学、植物学领域的研究热点问题。共词分析法可以客观、形象、直接地展示相关领域的研究热点，其基础是词频分析，通过对学科领域相关文献中，高频关键词共同出现的频次规律进行分析，展现其研究结构，揭示该学科的研究现状、研究重点、发展趋势和方向^[4-7]。共词分析法立论的假设条件是：文章的关键词是文章主题内容的浓缩，两篇文章如果有两个以上的相同关键词，则认为这两篇文章在研究主题的概念、理论或方法上是相关的，内容是相似的，共词文章数量越多，表明这类关键词“距离”越近，按“距离”将一个学科内的重要关键词进行分类，从而归纳出该

学科领域的研究结构^[8]。

为了更加形象地展示农业领域研究热点，本文在共词矩阵的基础上对数据进行聚类分析和多维尺度分析。聚类分析又称为集群分析，是一种归类方法，通过聚类分析的树状图，可以看出各个关键词之间的亲疏关系，进而可以看出这些词所代表的主题结构^[9-10]。

多维尺度分析是分析研究对象的相似性或差异性的一种多元统计分析方法，一般在二维空间（或三维空间）比较容易解释，可以揭示影响研究对象相似性或差异性的未知变量，看出各个研

究热点在农业领域所处的位置。

2 基于共词分析的微生物学植物学领域研究热点分析

2.1 共词矩阵构建

2.1.1 关键词词频统计

本文对关键词基本数据按词频高低进行排序，得到高频关键词统计表，因篇幅所限，表1仅列出位于前44位的高频关键词。

表1 高频关键词列表

序号	变量名	关键词	词频(次)	序号	变量名	关键词	词频(次)
1	x1	植物	141	23	x23	16S r DNA	21
2	x2	内生真菌	90	24	x24	光合作用	21
3	x3	鉴定	74	25	x25	农作物	21
4	x4	多样性	55	26	x26	r DNA	20
5	x5	序列分析	49	27	x27	分类	20
6	x6	研究进展	47	28	x28	基因克隆	20
7	x7	分离	46	29	x29	基因组	19
8	x8	内生细菌	45	30	x30	生物学特性	19
9	x9	拟南芥	43	31	x31	物种多样性	19
10	x10	筛选	36	32	x32	植物病毒	19
11	x11	盐胁迫	34	33	x33	干旱胁迫	18
12	x12	遗传多样性	30	34	x34	基因	18
13	x13	根瘤菌	29	35	x35	种子萌发	18
14	x14	降解	29	36	x36	AM 真菌	17
15	x15	RT-PCR	27	37	x37	RAPD	17
16	x16	聚类分析	27	38	x38	黄瓜花叶病毒	17
17	x17	应用	26	39	x39	克隆	17
18	x18	原核表达	26	40	x40	微生物	17
19	x19	玉米	24	41	x41	系统发育	17
20	x20	内生菌	23	42	x42	信号转导	17
21	x21	维生素酶	23	43	x43	亚细胞定位	17
22	x22	生长	22	44	x44	影响因素	17
						总计	1352

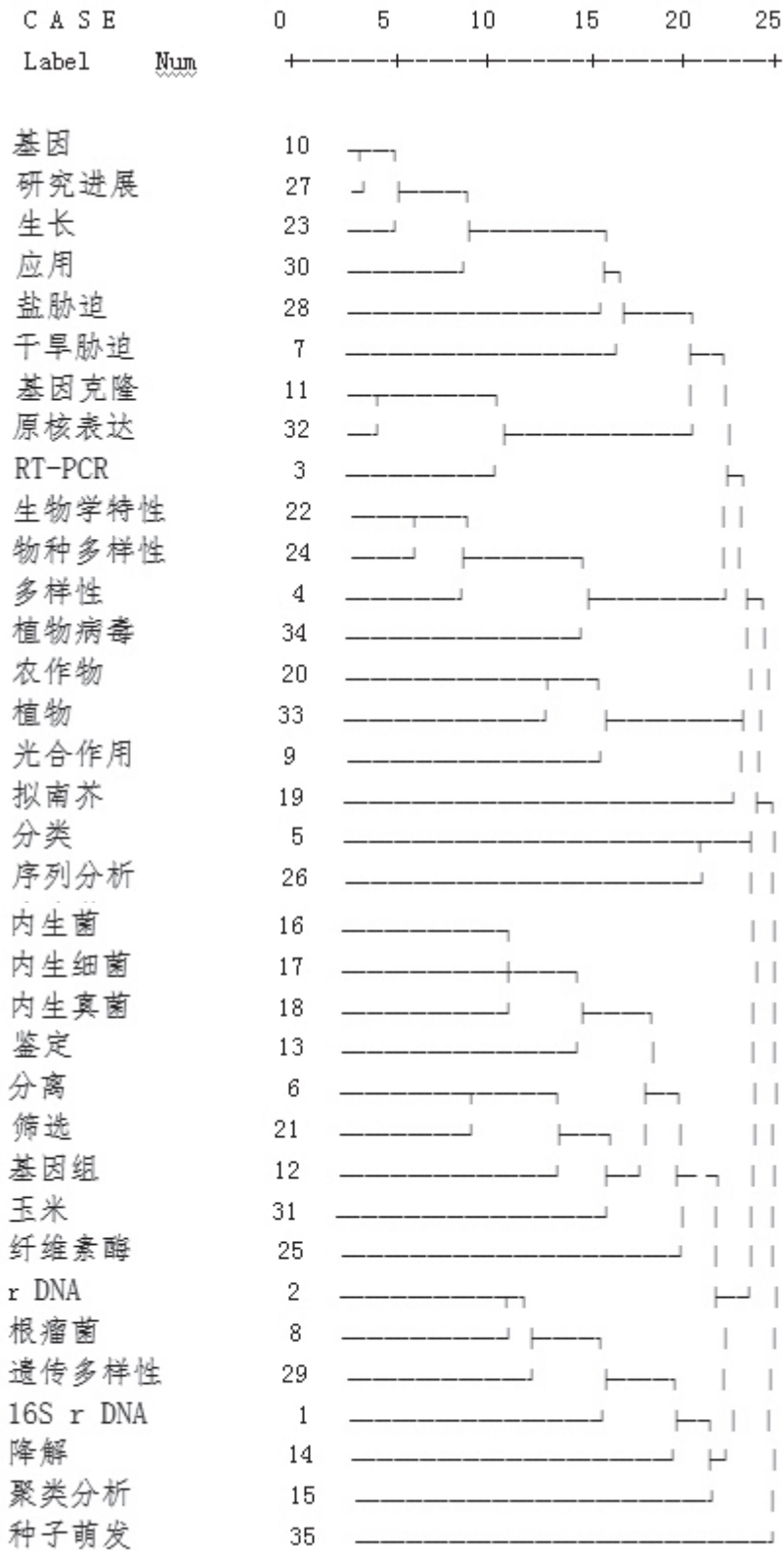


图2 高频关键词聚类树状图

利用 SPSS 软件对高频词共词矩阵进行聚类分析,得到聚类结果,如图 2 所示。通过得到的聚类树状图,可以将植物、微生物领域的研究热点分为四个大类,分别为:在基因方面的研究进展、植物的应用分析、微生物的物理分析、生物技术及特性分析。

在基因方面的研究进展。由基因、生长、研究进展、应用、盐胁迫、原核表达、干旱胁迫、基因克隆、RT-PCR、生物学特性、物种多样性、植物病毒、多样性等关键词构成。农业领域的研究从早期的多样性开始发展,进而过渡到基因克隆,逐步发展到植物病毒和原核表达方面的研究。基因是遗传的物质基础,是控制性状的基本遗传单位,可以控制生物个体的性状表现。通过对基因等生物学特性的研究,可以使微生物、植物的性状更加完善,利用转基因技术等提高农作物的产量和质量,防止病虫害,从而促进农业领域的发展。

植物的应用分析。由农作物、植物、光合作用、拟南芥、分类、序列分析等关键词构成。光合作用是植物、藻类和某些细菌在可见光的照射下,经过光反应和暗反应,利用光合色素,将二氧化碳和水转化为有机物,并释放出氧气的生化过程。序列分析是分子遗传学中一项既重要又基本的课题,在获得一个基因序列后,需要对其进行生物信息学分析,从中尽量发掘信息,从而指导进一步的实验研究。

微生物的物理分析。由内生菌、内生细菌、内生真菌、鉴定、分离、筛选、基因组、玉米、纤维素酶等关键词构成。内生菌是指那些在其生活史的一定阶段或全部阶段生活于健康植物的各种组织和器官内部的真菌或细菌,主要包括内生细菌、内生真菌、内生放线菌三大类。而纤维素酶是由多种水解酶组成的一个复杂酶系,自然界中很多真菌都能分泌纤维素酶。筛选是一种以多数物质中按预定目标就某种具有特定性质的物质

进行精选的操作过程,在生物学中则是指生物的系统、株或个体的选择。

生物技术及特性分析。由 r DNA、根瘤菌、遗传多样性、16S r DNA、降解、聚类分析、种子萌发等关键词构成。其中,种子萌发与其他关键词的距离都比较远,从树状图上看是单独的一类,说明这个类别与其他关键词所代表的主题不太一致,正处于向其他主题过渡的阶段。由于距离生物技术这一大类比较近,所以暂且将它归为这一类。遗传多样性是生物多样性的重要组成部分,对遗传多样性的认识是生物各分支学科重要的背景资料。r DNA 即核糖体 DNA,就是可以转录产生 r RNA 的基因,r DNA 的活性改变在核仁周期中发挥着重要作用。而近年来对降解方面的研究比较多,主要是因为现在环境保护问题越来越受到重视,对降解过程的研究可以为更好地消除有机制造物做出贡献。

此外,第二类中的拟南芥,以及第三类中的纤维素酶都距离所属类群较远,在所属类群的边缘,说明它们正在过渡到这一主题,具有潜在的发展趋势。另外,第一大类包含的关键词最多,说明这一研究主题相对比较集中,应该最受农业领域的学者们关注。而第二、四大类包含的关键词较少,形成的类群比较小,受关注程度相对较低。

2.3 基于共词矩阵的多维尺度分析

为了更加深入地了解微生物学、植物学领域各个主题之间的相关程度,利用 SPSS 软件对高频词的共词矩阵进行多维尺度分析。类群之间的相互距离可以反映出各个主题的相关程度,距离越大,说明主题之间越独立;距离越小,说明主题之间关系越紧密。处于边缘位置的关键词说明其研究主题比较不受关注,或者是正在向其中某一主题过渡。因此,可以通过多维尺度分析,判断学术群组在整体研究领域中的位置^[12]。结果如图 3 所示。

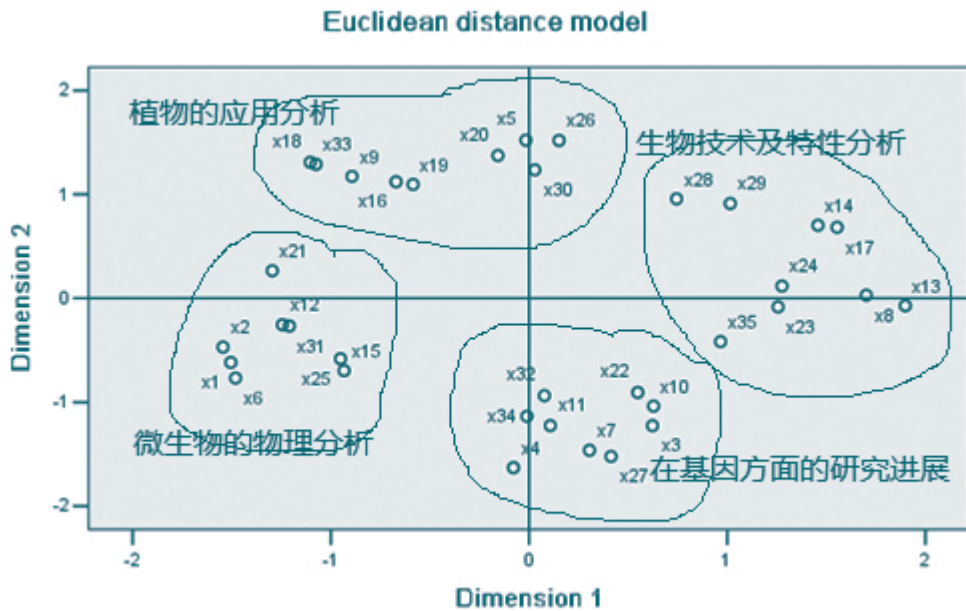


图3 多维尺度分析结果

通过图3,可以看出在基因方面的研究进展距离中心原点最近,表明这个领域是微生物学植物学领域的研究重点,属于最热门的领域。而微生物的物理分析、植物的应用分析和生物技术及特性分析这三大主题的研究距离中心原点偏远。并且,生物技术及特性分析和在基因方面的研究进展这两大类之间的关系比较紧密,而植物的应用分析和微生物的物理分析相对比较独立。

3 结论

本文以共词分析为基础,结合聚类分析以及多维尺度分析,对农业领域中微生物学、植物学领域的研究热点进行了统计分析。经过一系列的数据统计,最终得出微生物学、植物学领域的研究热点主要包括:植物和微生物的应用研究、生物技术及特性分析、在基因方面的研究进展、在基因方面的物理分析和统计分析这四大主题。其中,在基因方面的研究进展距离中心原点最近,表明这个领域是微生物学植物学领域的研究重点,

属于最热门的领域。

本文通过共词分析法挖掘出了农业领域中微生物学、植物学领域的研究热点问题,为农业领域的研究进展提供了主要线索。但是,随着农业领域结构的变化,该领域的研究热点有可能会随之改变,某些潜在领域也许会发展成研究热点,而某些热点领域也可能逐渐边缘化,进而会导致研究热点与重点的重新分布。因此,对于该领域的研究态势有待于学者们的持续关注 and 后续研究。

参考文献

- [1] 张勤, 马费成. 国内知识管理研究结构探讨——以共词分析为方法 [J]. 情报学报, 2008, 27(1):93-101.
- [2] 彭绪梅, 许振亮, 刘元芳, 等. 国外创业型大学研究热点探析: 共词可视化视角 [J]. 清华大学教育研究, 2007(6):95-100.
- [3] 刘则渊, 尹丽春. 国际科学学主题共词网络的可视化研究 [J]. 情报学报, 2006, 25(5):634-640.
- [4] Kumar S, Jan J M. Discovering Knowledge Landscapes: An Epistemic Analysis of Business and Management Field in Malaysia [J]. Procedia

-Social and Behavioral Sciences, 2012, 65:1027-1032.

[5] 蒋颖. 1995 ~ 2004 年文献计量学研究的共词分析 [J]. 情报学报, 2006, 25(4):504-512.

[6] 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.

[7] 岳洪江, 刘思峰. 国外管理学博士论文研究主题的可视化分析 [J]. 科学学与科学技术管理, 2008, 29(3):91-94.

[8] 张勤, 马费成. 国外知识管理研究范式——以共词分析为方法 [J]. 管理科学学报, 2007, 10(6):65-75.

[9] 骆方. SPSS 数据统计与分析 [M]. 北京: 清华大学出版社, 2011.

[10] 高祥宝. 数据分析与 SPSS 应用 [M]. 北京: 清华大学出版社, 2007.

[11] 王宪磊. 信息管理论 [M]. 北京: 社会科学文献出版社, 2004.

[12] 刘荣. 基于科学知识图谱的创新系统研究的历史与现状分析 [J]. 科技管理研究, 2008, 28(12):502-504.