

领域主题动态演进分析 ——以动物资源与育种领域为例

李晓曼, 孙巍

(中国农业科学院农业信息研究所, 北京 100081)

摘要: 为了发现领域主题并研究其动态演进和发展规律, 采用文献计量和专家咨询方法, 以动物资源与育种领域为例, 从引用关系、数据流和战略坐标三种视角对主题动态演进做了分析。首先, 基于文献共被引识别领域主题, 通过共被引时间线图对领域主题进行动态演进分析; 其次, 从数据流视角研究主题随着时间的融合、分化及延续等的动态演化; 然后, 通过战略图从向心度和密度方向对主题发展演化进行分析; 最后, 采用专家咨询法对分析结果进行评判和研读。结果表明: 三个视角可以从多个方面、不同角度有效揭示主题演进情况, 且文献计量结果与专家咨询基本一致。

关键词: 主题演进; 共现分析; 可视化分析; 动物资源与育种

中图分类号: TP391

DOI: 10.3772/j.issn.1673-2286.2015.12.005

1 引言

梳理学科领域的研究热点, 揭示领域主题的发展态势, 可为科研人员及科技政策制定者提供参考与指导。在相关研究中, 共现分析以其方法的简明性和结果的可靠性成为支撑信息内容分析的重要手段和工具^[1]。通过共现现象对文献进行计量的软件有CiteSpace^[2]、VOSviewer、SciMAT、NWB、Sci²等。秦晓楠等运用CiteSpace绘制文献共被引网络图谱, 使用聚类分析将生态安全研究领域划分为5个知识群组, 并对每个群组进行系统分析^[3]; 经渊等以Web of Science数据库中的SCI、SSCI和A&HCI三个子库为数据源, 利用CiteSpace进行国家产学研协同创新研究演进分析^[4]; 马费成则以生物医学为例, 利用Pajek和CiteSpace来进行数据处理, 探讨知识演化与创新的过程^[5]; 张云等利用SciMAT对专利领域进行特定主题的动态演进分析, 分别从宏观和微观层面对文献关键词进行分析^[6]。

本文以动物资源与育种领域为例, 从引用关系视角、数据流视角和战略视角对领域主题进行动态演进分析, 并通过专家咨询的方式对分析结果进行评估和补充, 验证分析方法的可行性。

2 分析方法

2.1 基本原理

共现是一种时空相关分析方法, 应用非常广泛。科技论文中的共现是指相同或不同类型特征项共同出现的现象。共词分析和共引分析都属于比较典型的共现分析方法^[1]。

主题演化具有一定的时间特征, 每个主题都是带有时间标签的, 主题演化分析本质上是对具有时间特征的主题进行语义层面上的分析计算, 揭示主题随时间发展变化情况。基于共词分析的学科主题演化分析主要是以主题词频次或共现频次量化计算为基础, 根据主题词间关联强度的大小聚集成簇, 并绘制不同时期的学科主题网络图谱, 以观察学科主题的演化轨迹^[7]。

2.2 研究方法

共引分析可以反映学科的动态变化情况, 通过图谱中的关键节点、聚类及色彩来标识某个研究主题的演变; 数据流通过分析领域中的数据流向来寻找本领域

域热门主题时序变化情况；战略图通过对热门主题进行四个象限的归类来分析每一象限中的主题变化。因此本文选取从引用关系、数据流和战略视角对动物资源与育种领域主题的动态演进进行分析，三个视角相互补充使得分析结果更加全面地反映该领域的动态演进情况。首先，从Web of Science中遴选“动物资源与育种”领域的相关文献；第二步，根据研究需求对搜集的数据进行清洗和筛选，构建目标数据集；第三步，从引用关系视角，构建文献共被引网络，通过文献聚类可视化分析，结合网络计量分析对该领域的主题演进进行分析；第四步，从数据流视角，构建时序主题网络，通过分析不同时间段的主题间的相似度及主题的转移量对该领域的主题演进进行分析；第五步，从战略坐标视角，分别构建不同时间窗下的主题战略坐标图，通过分析每个时间窗主题所在的象限对该领域的主题演进进行分析；最后，邀请该领域专家对分析结果进行评估解读，并提出建议。

3 动物资源与育种主题演进分析

选取“动物资源与育种”领域为研究对象，为了更全面地了解本领域主题演进情况，从三个不同的视角进行主题动态演进分析。

3.1 引用关系视角的领域主题演进分析

基于文献共被引原理，采用一定的聚类方法将内容联系密切的高被引文献聚类成若干个簇，通过解读高被引文献，标记出每一个聚类的主题名称即可得到领域的核心主题。文献的发表时间以及被引时间可以看出本聚类主题的发展时间，高被引文献在发展过程中起着重要的作用。本研究利用CiteSpace绘制文献共被引网络图谱来反映领域主题的聚类情况以及发展演化情

况，过程如下：从Web of Science中遴选“动物资源与育种”领域共58,294篇文献（2000年-2014年）；根据文献被引频率的高低，选取本领域高被引文献4,682篇；以文献本身为研究对象，通过CiteSpace得到文献共被引网络图谱的时间线视图（见图1）。

图1中圆圈节点表示被引文献，节点越大表示该文献被引用的频次越大。时间线视图主要侧重于勾画聚类之间的关系，以及某个聚类中文献的历史跨度。同一聚类节点按照时间顺序排列，分布在同一条水平线上，每一个聚类代表一个主题^[2]。图1中文献共被引网络共生成了20个聚类，之间存在或多或少的主题交叉。选取节点数排名前3的聚类进行分析，分别为#0（protocol）、#1（low quality forage）、#2（feeding rapeseed），对应节点数分别为49、41、29，可以看出这三类占据了图中大部分的文献节点。图中带有“#”标识的聚类标签词来源于施引文献，是从施引文献的“标题”、“索引词”或“摘要”中提取的高频词表示。我们认为仅用某个词对聚类进行标识的准确程度相对较差，所以通过对三个类包含的高被引文献内容进行分析，得出这三个类的主题分别为：奶牛营养需要、体况与繁殖研究，动物饲料评价及瘤胃研究，动物脂质代谢及共轭亚油酸研究。通过时间线图可以得出这三个类在时间轴上的引文分布。#0（protocol）主题类中，包含的节点数最多但是节点都相对较小，说明相关研究一直在持续但是相对热度较低，是发展比较稳定的主题；#1（low quality forage）主题类节点数相对较多，所覆盖的时间区域相对较小且靠后，高中心度和高被引的文献也涵盖在此类中，说明该类发展较为迅速；#2（feeding rapeseed）主题类中，论文基本涵盖了所研究的时间区域，说明一直以来都是研究的热点。表1列出了三个聚类中出现的高被引文献以及被引频次，三个聚类的标识是软件自动标识加上文献内容分析得出的。

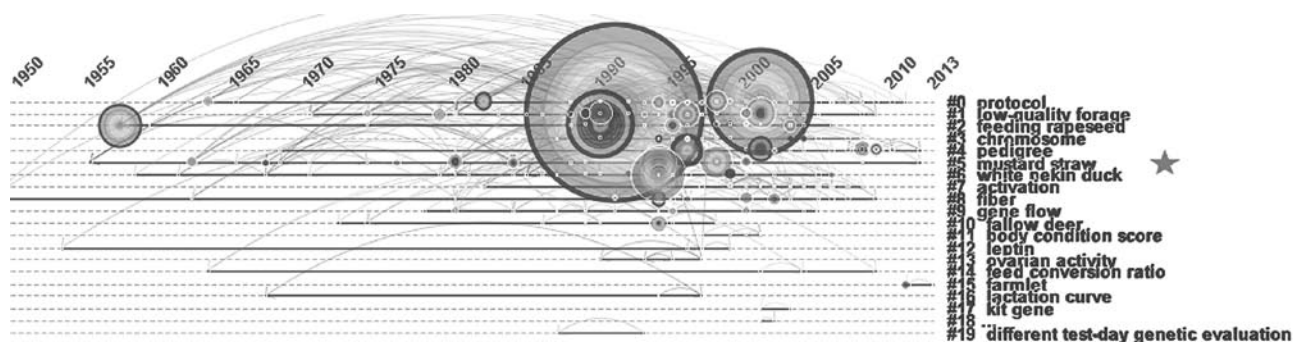


图1 “动物资源与育种”领域高被引文献聚类时间线图

表1 聚类相关文献列表

聚类号以及标注的热点主题	文献标题	被引频次
#0 (protocol) 对奶牛营养需要、体况与繁殖的研究	Nutrient Requirements of Dairy Cattle	242
	Metabolic profiles in Virginia dairy herds of different milk yield	73
	A Body Condition Scoring Chart for Holstein Dairy Cows	60
	Reproductive loss in high-producing dairy cattle: where will it end?	52
	Principal descriptors of body condition score in Holstein cows.	45
	In vitro embryo production after microinjection and ovarian dynamics following transvaginal follicular oocyte aspiration.	44
	Prediction of the Energy Value of Cow's Milk HF Tyrrell	38
#1 (low quality forage) 对动物饲料评价及瘤胃的研究	Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition.	413
	Nutrient Requirements of Dairy Cattle: Seventh Revised Edition	105
	National Research Council beef model	94
	The estimation of protein degradability in the rumen from incubation measurements weighted according to rate of passage	55
	Rapid method for determination of total fatty acid content and composition of feedstuffs and feces	49
	Automated simultaneous determination of ammonia and total amino acids in rumen fluid and in vitro media	47
	Effect of ammonia concentration on rumen microbial protein production in vitro	42
#2 (feeding rapeseed) 对动物脂质代谢及共轭亚油酸的研究	Effects of brown midrib 3 mutation in corn silage on productivity of dairy cows fed two concentrations of dietary neutral detergent fiber	42
	Official Methods of Analysis of the Association of Official Analytical Chemists	156
	A simple method for the isolation and purification of total lipides from animal tissues	135
	Nutritional regulation of milk fat synthesis.	45
	Conjugated linoleic acid is synthesized endogenously in lactating dairy cows by Delta(9)-desaturase	40
	Circadian and ultradian rhythms of peripheral cortisol concentrations in lactating dairy cows.	34
	Effect of dietary sunflower oil and pasture forage maturity on conjugated linoleic acid (CLA) content in milk fat from lactatin dairy cows	33
	Identification of the conjugated linoleic acid isomer that inhibits milk fat synthesis	32

3.2 数据流视角的领域主题演进分析

论文的关键词是科学研究内容的凝练，共词分析是通过统计关键词之间的关系来概述研究领域的微观知识结构。根据关键词共现情况对关键词进行聚类，再将时间区域划分为不同的时间段来分析关键词簇的流向可以得出持续性研究主题的发展变化情况。基于高频关键词选取本领域7,514篇文献，在SciMAT软件中设置分析单元为word，将2000-2014年的分为5个等距离的时间片（2000-2002，2003-2005，2006-2008，2009-2010，2011-2014），设置5个时间片的数据和网络的精简阈值^[8]，得到2000-2014年动物资源与育种领域

主题演进图。通过查看分析演进图中热门关键词的聚类以及类内、类间关键词随时间片的流向变化情况，来分析主题的动态发展过程。在此分析过程中，由于cattle、cow、chicken、sheep、pig等关键词属于该领域较宽泛的词，和其它微观词均关联较紧密，出现频次相对较高，会掩盖其他微观词间关联的客观呈现，因此，在SciMAT中从关键词组中删除了此类相对宽泛的关键词，以便更客观地揭示该领域主题的演化情况。

图2展示了5个时间片随着时间变化各主题的演化趋势。其中列代表时间段内分析单元的聚类主题，每一个圈代表一个主题。两个圈之间如有线相连，表示研究具有持续性，边的粗细与进化关系的权重成正比^[9]。从

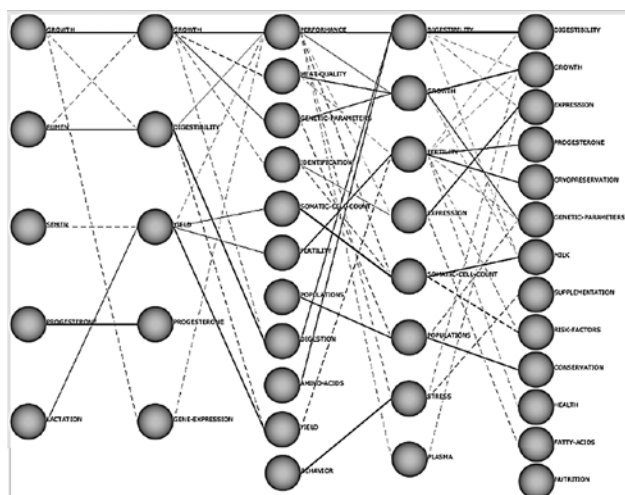


图2 “动物资源与育种”领域热点主题的时序演化趋势

图2中可以得出4条主要的主题流向,基于文献计量分析方法对各主题流信息的详细描述如下(主题是由主题类中代表性最强的词表示,其对主题内容的揭示相对片面,因此综合其中的重点关键词进行人工标注识别出每一条主题流):

①动物生长及生产性能研究: $\text{growth} \rightarrow \text{growth} \rightarrow \text{performance} \rightarrow \text{growth} \rightarrow \text{growth}$,通过对每一阶段的高频关键词进行分析可以得出:该主题流的研究以动物生长及生产性能研究为持续重点,包含对营养、饲料、消化等方面的研究。随着时间的推移,后期增加了对肉质、动物行为的关注,暗示着对动物产品质量和动物福利的重视。

②动物营养及动物消化研究: $\text{rumen} \rightarrow \text{digestibility} \rightarrow \text{digestion} \rightarrow \text{digestibility} \rightarrow \text{digestibility}$,通过对每一阶段的高频关键词进行分析可以得出:该主题流以动物营养和消化的研究为主线,涉及瘤胃消化、饲料、牛奶产量等方面研究。关键词 diet 、 amino-acids 、 digestion 、 feed 、 fermentation 、 energy 、 rumen 、 fiber 、 $\text{ruminal-fermentation}$ 等与动物营养及代谢相关度较高,并且饲料转化率高是动物遗传资源的重要内容,其他方面的研究多集中于环境效应分析。

③动物繁殖研究: $\text{lactation} \rightarrow \text{yield} \rightarrow \text{fertility} \rightarrow \text{fertility} \rightarrow \text{progesterone} + \text{cryopreservation}$,可以看出该主题流在最后一个时间片由繁殖主题分化成为两个主题黄体激素和冷冻保存。通过对每一阶段的高频关键词进行分析可以得出:该主题流的研究以动物繁殖的研究为主流,包含对哺乳期、繁殖性能、生育能力、排卵、精子、黄体激素等方面的研究,并且随着时间的推移

移体外繁殖成为动物繁殖的主要方向。

④动物基因研究: $\text{identification} \rightarrow \text{expression} \rightarrow \text{expression}$,通过对每一阶段的高频关键词进行分析可以得出:该主题流以基因鉴定、表达的研究为主流,并逐渐扩张到蛋白表达、代谢、多态性等多方面的研究。

从动物资源与育种领域主题的总体来看,随着时间的推移,各个时间片的聚类数量也在不断增加,并且主题呈现出多样化趋势,发文量显著增加。同其他主题演进一样,在该领域主题的发展过程中,也出现了主题的融合、分化、新增、消失、再生和转移现象。如在2011-2014年间,动物繁殖主题分化为卵母细胞与精子冷冻保存两个主题聚类;在2008-2011年间,氨基酸和动物消化主题聚类融合成了动物消化一个主题聚类。

3.3 战略视角的领域主题演进分析

依然选取7,514篇文献进行分析,在SciMAT中可得到2000-2014年动物资源与育种领域主题战略图。图3给出了5个时间段主题的战略坐标分布。图中节点代表聚类主题,节点的大小代表高频词的H指数;横轴表示中心度,代表主题在研究领域的重要程度;纵轴表示密度,代表主题的发展程度。第一象限中心度和密度都很高,说明这个区域的主题发展较好且较为重要,被称为引擎类的主题;第二象限密度较高,但中心度较低,说明这个区域的主题发展较好但比较孤立,这些主题往往是一些专门性、外围性的主题;第三象限中心度和密度都比较低,说明这个区域的主题是新兴的或者衰退的;第四象限中心度较高,密度较低,说明这个区域的主题是基本且稳定的^[9]。

由图3可见,对于“动物生长及生产性能”的相关研究,在2000-2008年都处于第四象限,研究密度相对较低;在2009年后处于第一象限,说明本主题的研究随着时间的推移热度逐渐上升。对于“动物营养及动物消化”的相关研究,在5个时间片中都位于第一象限,说明该主题发展势头良好,一直都是本领域的热门且重要研究主题。对于“动物繁殖”的相关研究,在2000-2002年处于第四象限,在2003-2005年处于第一象限,且密度增加,说明相关研究相对较多;在2006-2008年处于第二象限,相关研究比较孤立,与本领域其它主题联系较少,专门性程度较大;2009年左右其中心度又逐渐增大,说明从2009年开始相关主题成为本领域发展较好且较为重要的主题。对于“动物基

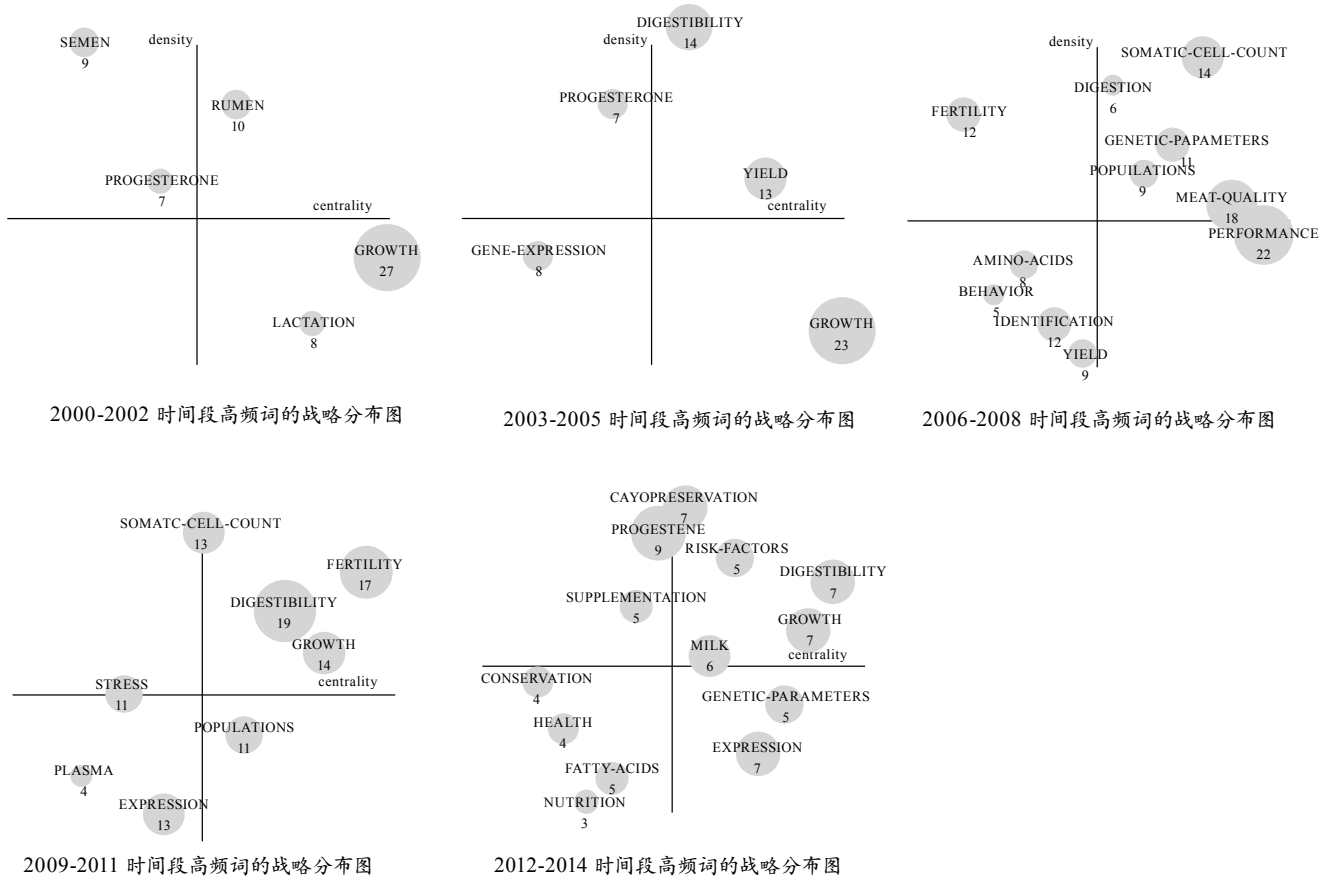


图3 2000-2014年5个时间片的高频词战略分布图

因”的相关研究,在2006-2011年间处于第三象限,属于新兴主题,中心度和密度都相对较低;在2012-2014年位于第四象限,说明该主题发展较为稳定,成为本领域的基本主题。

3.4 分析结论

通过专家对上述三个视角的分析和解读可知,利用引用关系得出三个主题在动物资源与育种领域发展态势良好,其中奶牛营养需要、体况与繁殖的研究,营养和生殖研究间存在一定的相关性,但不能反映其遗传资源特性,内容需进一步挖掘。利用共词分析得出的四个主题流基本涵盖了本领域持续性的热门研究主题,并且四条主题流的人工标注和分析是合理的,可较好地反映动物资源与育种相关研究进展,但对资源的再利用特性、可遗传特性、采用动物生长发育性状的遗传参数遗传评估、动物遗传多样性及基因多态性等方面还应再挖掘。

综上,奶牛营养需要、体况与繁殖研究,动物饲料

评价及瘤胃研究,动物脂质代谢及共轭亚油酸研究涵盖的研究范围相对较大,是整个领域持续性较长的主题;动物生长及生产性能研究、动物营养及动物消化、动物繁殖、动物基因四个主题随着时间的变化发展的程度各有不同,但都有向本领域发展较好且较为重要和稳定方向发展的趋势,说明这四个主题是该领域持续且重要的主题;其中动物基因主题是近几年才成为热门主题,属于新兴主题且发展态势良好。

4 结语与展望

本文以Web of Science为数据源,利用CiteSpace和SciMAT软件,从引用关系视角、数据流视角和战略视角对动物资源与育种领域的主题动态演进进行了分析。通过研究可知,引用关系视角的分析借助的是文献本身,研究结论都是文献层面的,文献涵盖了整个研究的内容使得分析结果较全面。共引分析可以反映学科的动态变化情况,因为论文发表之后的被引频次是在不断增加之中,因此文献的引用具有时效性,用共引分

析来映射学科结构存在时差,一些近期发表的文献由于引用频次较低很难被挖掘出来。通过主题流视角和战略视角的分析借助的是关键词,主题流视角可以挖掘出延续性的主题,战略视角可以得出主题的发展情况,两个视角配合使用得出的结果较全面,但是关键词的内容较片面。综上,对于引用关系方向,在下一步的研究中应将数据范围进行缩减,以得出更加符合本领域的研究主题。对于数据流和战略方向,在下一步的研究中应将关键词和相关文献结合起来进行分析。

参考文献

- [1] 杨立英.科技论文共现理论研究与应用[D].北京:中国科学院文献情报中心,2007.
- [2] 陈悦,陈超美.引文空间分析原理与应用:CiteSpace实用指南[M].科学出版社,2014:74-77.
- [3] 秦晓楠,卢小丽.国内生态安全研究知识图谱——基于CiteSpace的计量分析[J].生态学报,2014,34(13):3693-3703.
- [4] 经渊,郑建明.国际产学研协同创新研究演进分析[J].数字图书馆论坛,2015(6):49-53.
- [5] 马费成,陈满俊,刘向.基于科学知识图谱分析的知识演化研究——以生物医学为例[J].情报科学,2012(1):1-7,15.
- [6] 张云,华薇娜,袁顺波,等.Wos数据库中专利分析论文的主题动态演进研究[J].现代图书情报技术,2015(1):17-23.
- [7] 叶春蕾,冷伏海.基于共词分析的学科主题演化方法改进研究[J].情报理论与实践,2012(3):79-82.
- [8] SciMAT.SciMATVersion1.0UserGuide[EB/OL].[2014-03-22].<http://sci2s.ugr.es/scimat/download.html>.
- [9] Cobo M J,López-Herrera A G, Herrera-Viedma E, et al.SciMAT: A New Science Mapping Analysis Software Tool[J]. Journal of the American Society for Information Science and Technology, 2012, 63(8): 1609-1630.

作者简介

李晓曼,女,1993年生,中国农业科学院农业信息研究所硕士研究生,E-mail: sunshinexiaomanli@163.com。
孙巍,女,1978年生,中国农业科学院农业信息研究所副研究员,研究方向:农业知识组织与可视化分析,E-mail: sunwei@caas.cn。

Analysis of the Themes Dynamic Evolutions: A Case of Animal Resources and Breeding Field

LI XiaoMan, SUN Wei

(Institute of Agricultural Information, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: In order to find the themes and to determine the themes dynamic evolutions in the specific area, the paper based on the bibliometrics and expert consultation method to analyze the themes dynamic evolutions of the Animal Resource Breeding from three different perspectives. Firstly, the paper based on the references co-cited analysis to find the themes, and to analyze the themes dynamic evolutions by the co-cited timeline map. Secondly, the paper analyze the themes dynamic evolutions from the perspective of data stream. Thirdly, the paper analyze the themes dynamic evolutions through the radial and density direction of the strategic map, and finally, we let the experts of the field analysis and evaluate the research results. The results shows that: we can effectively reveal the themes evolutions from different perspectives, and the results from the bibliometrics and the expert consultation method almost unanimous.

Keywords: Theme Evolution; Co-occurrence Analysis; Visualization Analysis; Animal Resource Breeding

(收稿日期: 2015-12-08)