

# 术语引用视角下的学科交叉测度—— 以 PLOS ONE 上六个学科为例

徐庶睿<sup>1,2</sup>, 卢超<sup>1,2</sup>, 章成志<sup>1,2,3</sup>

(1. 南京理工大学信息管理系, 南京 210094; 2. 江苏省社会公共安全科技协同创新中心, 南京 210094;  
3. 江苏省数据工程与知识服务重点实验室(南京大学), 南京 210093)

**摘要** 跨学科知识转移是学科交叉形成的基础和前提, 但是目前通过引证关系挖掘出来的学科交叉仍然无法呈现其内部的交叉点。本文提出结合术语和引文内容探测学科交叉度的研究方法并进行实证分析。首先, 采集学术论文并解析引文内容; 利用术语集获取术语。然后, 统计引文内容中的术语频次和重复率。最后, 计算学科交叉度并分析学科交叉点的贡献度。实证研究表明: 不同类型的学科存在着不同的交叉水平; 相似的交叉度之下存在着不同的交叉点。结合术语和引文内容的方法能够从内容层面量化研究学科交叉问题, 反映学科交叉背后不同程度的知识交叉, 相较于基于引证关系的学科交叉研究具有一定的优势。

**关键词** 知识引用; 学科交叉测度; 跨学科知识转移; 引文内容

## Measurement of Interdisciplinary Research from the Perspective of Terminology Citation: Six Disciplines on PLOS ONE

Xu Shurui<sup>1,2</sup>, Lu Chao<sup>1,2</sup> and Zhang Chengzhi<sup>1,2,3</sup>

(1. Department of Information Management, Nanjing University of Science & Technology, Nanjing 210094;  
2. Jiangsu Collaborative Innovation Center of Social Safety Science and Technology, Nanjing 210094;  
3. Jiangsu Key Laboratory of Data Engineering and Knowledge Service (Nanjing University), Nanjing 210093)

**Abstract:** Interdisciplinary knowledge transfer is the basis and prerequisite for the formation of an interdiscipline. Current interdisciplinary research based on citation relations, however, is still unable to present an internal intersection. This paper uses terminology and citation content to explore the degrees of interdisciplinarity. First, articles containing a citation context were collected, and the terms extracted. Next, the frequency and repetition rate of the terms in the citation content were calculated. Finally, the degrees of interdisciplinarity and the contributions of the discipline intersections were calculated. The results show that different types of disciplines have different degrees of interdisciplinarity, and that different intersections exist under similar degrees. The method described in this study, which combines the terminology and citation content, can be used to quantify interdisciplinary research at the content level. It reflects the degree of knowledge intersection behind interdisciplinarity. Compared with interdisciplinary research based on citation relations, the approach has certain advantages.

**Key words:** knowledge citation; interdisciplinary measurement; interdisciplinary knowledge transfer; citation content

收稿日期: 2017-04-19; 修回日期: 2017-06-23

基金项目: 国家社会科学基金项目“面向知识创新服务的数据科学理论与方法研究”(No. 16ZAD224)。

作者简介: 徐庶睿, 女, 1992年生, 情报学专业硕士研究生, 主要研究方向为信息检索与数据挖掘; 卢超, 男, 1991年生, 管理科学与工程专业博士研究生, 主要研究方向为引文内容分析; 章成志 (ORCID: 0000-0001-8121-4796), 男, 1977年生, 博士, 教授, 博士生导师, 通讯作者, 主要研究领域包括信息组织、信息检索、数据挖掘及自然语言处理, E-mail: zhangcz@njust.edu.cn。

## 1 引言

应对诸如全球气候问题、人类基因组测序等重大问题,需要各个学科领域参与合作,交叉学科应运而生。张春美等<sup>[1]</sup>对百年来诺贝尔物理学奖、化学奖、生理学或医学奖获得者所属的研究领域做调查,发现有41.02%获奖者的研究领域属于交叉学科。交叉学科已经不断涌现并在我们人类认识世界和改造世界的过程中发挥着越来越重要的作用。

截至目前,国内外的学科交叉研究正在蓬勃发展并取得不少收获,这些研究主要以学术论文的引证关系为分析对象,集中在跨学科知识转移研究<sup>[2]</sup>和学科交叉度探测<sup>[3]</sup>两个方面。但是以引证关系确定学科分类的方法存在诸多问题。一方面,参考文献类型多样,数量众多,现有的参考文献-学科分类标准不统一,对应体系不能涵盖所有的文献类型和范围<sup>[4]</sup>;另一方面,通过引证关系挖掘出来的学科交叉仍然是浅层的,粗糙的。这样的研究无法回答诸如“当一个学科与另一个学科发生交叉时,两个学科在哪些主题下发生了交叉,在哪些主题下没有交叉?”等问题。然而,研究这个问题却有很重要的研究意义:可以使我们从微观层面了解学科交叉中知识的集成过程;为发现学科前沿话题和预测新的交叉学科创造条件。现有的研究并没有很好地回答这类问题。

在科学发展过程中,学科进行着交叉融合,学科知识进行着扩散渗透。学科知识的跨学科转移是交叉学科形成的前提和基础。Small<sup>[5]</sup>认为当研究者倾向于使用相同的词或短语来描述同一个被引用对象时,这些词或短语就成为这个被引对象的概念符号。学术界进行交流的概念符号即术语。术语具有权威性且使用广泛。基于术语角度研究学科交叉现象比基于引证关系角度更加微观具体,能揭示出知识的具体集成和扩散过程。

引文内容涵盖施引文献和被引文献的提及知识,是研究学科间和学科内部知识交流的平台<sup>[6]</sup>。首先,从论文层面看,施引文献和被引文献是通过引用关系产生联系的。其次,从内容层面看,不同于简单的引用关系,施引文献和被引文献是通过术语产生联系的。术语是施引文献和被引文献沟通的媒介。术语如果出现在施引文献的引文内容中,那么也会出现在被引文献的正文中。所以术语结合引文内容作为学科交叉测度的载体是行之有效的。

综上所述,本文对引文内容中术语进行统计分

析和学科交叉度计算,从内容分析角度量化回答学科交叉度这个问题。有别于基于引证关系的学科交叉问题研究,本文提出从术语结合引文内容的角度对学科交叉问题进行研究的方法具有极大的优势。首先,从术语结合引文内容的角度比引证关系角度更加微观具体,学科交叉度测量结果会更加准确。其次,术语是知识的基本单元<sup>[7]</sup>,术语引用的本质仍然是知识的转移,从术语角度能进一步研究学科知识的集成和扩散,探究学科知识结构的发展变化。本文选取术语为知识载体,通过统计学科论文在引文内容中引用这些术语的词量和频次,从而确定学科间的知识交叉情况。

## 2 相关研究概述

本文从学科交叉相关研究和引文内容相关研究两个角度对现有的研究工作进行述评。

### 2.1 学科交叉相关研究概述

学科交叉研究的发展前景良好。相关概念如“多学科(multidisciplinary)”、“跨学科(transdisciplinary或者cross-disciplinary)”、“交叉学科(interdisciplinary)”的定义仍众说纷纭,未达成明确共识<sup>[4]</sup>。本文仅以术语“交叉学科”(“学科交叉”与“交叉学科”的细微差别可详见文献[8])作为通用术语进行考察。

现有的学科交叉研究(如Wagner等<sup>[4]</sup>,Rafols等<sup>[9]</sup>以及Vugteveen等<sup>[10]</sup>)的主要研究方法可分为两类:自上而下法和自底向上法。其中,大部分研究者主要使用自上而下法。在自上而下法中,基于预先定义的学科分类目录(如Web of Science的学科分类目录)确定学科类目以及学科对应的期刊,通过对参考文献所属的学科分类进行分析从而研究学科交叉占比和学科间的关系。van Raan等<sup>[11]</sup>根据某研究所发表论文的学科领域分布和论文的施引文献的学科领域分布作为描述该研究所学科交叉性的指标。部分研究者采用自底向上的方法进行研究,即在对论文进行聚类或分类的基础上,进行相似度的多元分析,并通过可视化展示来评估类簇网络结构。Small<sup>[12]</sup>选取高共被引论文,通过多次聚类构成科学地图,对科学地图中表示学科交叉的节点进行内容分析。

当前的学科交叉研究主要集中在跨学科知识转移和学科交叉度探测研究两个方面。

#### (1) 跨学科知识转移

随着对交叉学科研究的深入,研究者逐渐意识

到交叉学科的关键是知识的集成<sup>[13-14]</sup>。美国国家科学院将交叉学科定义为“为了增进根本性认知或解决单学科范畴或研究领域无法解决的问题，由个人组成的团队研究模式，该研究模式集成了两个或两个以上学科或专业知识领域的信息、数据、技术、工具、视角、概念以及/或者理论”<sup>[14]</sup>。

Liu 等<sup>[15]</sup>在 Rafols 等<sup>[9]</sup>工作的基础上，提出使用文献计量数据进行知识集成和知识扩散分析的框架。Vugteveen 等<sup>[10]</sup>以交叉学科河流研究学为例，一方面考察了河流研究学的知识来源，另一方面考察了河流研究学的知识扩散。Liu 等<sup>[16]</sup>选取中国一级学科下权威期刊的数据构建学科知识网络，揭示学科间的知识转移结构以及从 1981 年至 2001 年的学科知识演变过程。

## (2) 学科交叉度探测

早期从文献计量领域对学科交叉进行测度最大的障碍是有效数据的缺乏<sup>[17]</sup>。在 1985 年，Porter 等<sup>[17]</sup>以期刊引用报告 (Journal Citation Reports) 作为一个合适的数据基础，使用引用学科分类 (Citations Outside Category) 为指标对 383 篇期刊的交叉学科性进行测度。实验结果证明跨领域 (工程学、生命科学、物理科学和社会科学) 的引用现象非常少。

Porter 等<sup>[18]</sup>提出 integration score 作为交叉学科测度的指标，这个指标不仅考查被引参考文献在不同学科的分布，还考察了学科间的相关性。在 Porter 工作的基础上，Rafols 等<sup>[9]</sup>引入一套新的文献计量体系，通过评价参考文献多样性来对交叉学科性进行间接测定。他们所选取的指标主要参考 Stirling<sup>[19]</sup>所提出的学科多样性分析的框架。对于学科的交叉性测度，目前越来越多的学者倾向纳入网络分析的指标进行分析。

## 2.2 引文内容相关研究概述

在文献中，引用是通过引文标记进行表示的，而围绕在引文标记附近的文本内容就是引文内容。首先，从本质上来讲，引文内容和参考文献一样，反映的是文献间的引用现象。其次，不同于参考文献，引文内容不再是抽象的、间接的代表被引文献的符号，而是实实在在代表具体概念或观点的语言<sup>[5]</sup>。结合语言学以及引文分析的基础理论，利用文本挖掘和自然语言处理技术，对引文内容从内部微观层面进行研究，从而能够探究引文内容语言层面的文本特征，丰富施引文献和参考文献之间的引用关系，

挖掘出引用背后所蕴含的动机<sup>[20]</sup>。

目前，研究者主要对引文内容进行引用标注<sup>[21-23]</sup>，情感分析<sup>[24-25]</sup>等基础分析工作，并论证了引文内容相较于引证关系，摘要等的优势<sup>[26-28]</sup>。引文内容是对被引用文献主题最好的反映<sup>[5]</sup>，在引文内容分析基础上识别的主题具有更好的主题代表性<sup>[26]</sup>，是学科间和学科内部的知识扩散分析的平台<sup>[29]</sup>。

少数研究者利用引文内容进行了学科交叉研究。主要研究成果如下：Small<sup>[12]</sup>利用学科交叉点的同被引引文内容 (co-citation contexts) 对线索词进行分析。章成志等<sup>[30]</sup>通过分析学科术语的频次，计算学科交叉度。张艺蔓等<sup>[29]</sup>融合引文内容和全文本引文分析对交叉研究领域的知识流动进行挖掘。

综上所述，目前基于引文分析方法的跨学科知识转移和学科交叉度探测还主要停留在引证关系层面，少有文献涉及文本内容层面。然而对于学科交叉研究进行分析，特别是跨学科知识的抽取，需要深入文本内容中进行知识挖掘，从而更细致地研究跨学科知识的集成和扩散。章成志等<sup>[30]</sup>通过统计学科间术语包含情况直接计算学科交叉度，不同于该研究的是，本文通过引入术语集，间接地计算学科间基于某一术语集的学科交叉度，弥补了由于术语抽取错误率较高造成学科交叉度计算存在缺陷的问题。

## 3 研究思路与实现方法

### 3.1 研究思路

如图 1(a)和图 1(b)所示，医学学科术语“vascular smooth muscle” (血管平滑肌) 出现在化学学科论文和物理学学科论文的引文内容中，我们认为这两个学科共同引用医学学科的术语。进一步地，若这两个学科中有大量学术论文的引文内容包含医学学科术语，则我们认为这两个学科存在交叉现象。

根据以上观察，本文将学科知识交叉具象化为学科间术语的引用，从引文内容角度探索学科交叉现象。如图 2 所示，借鉴 Small<sup>[31]</sup>的文献共被引思想，当学科 A 与学科 B 的引文内容中，引用  $k$  个 ( $1 \leq k \leq N$ ) 学科术语集  $TC_i, TC_{i+1}, \dots, TC_{i+k-1}$  中的术语时，我们认为学科 A 和学科 B 通过共同引用术语集 TC 的术语而产生交叉。学科 A 和学科 B 共同引用的术语越多，则学科 A 和学科 B 之间的学科交叉度越强。

During hypoxia, ROS generation from mitochondrial complex III was reported to be involved for inhibiting cellular PHD activity to stabilize HIF-1 $\alpha$  [26]. However, recent finding of decreased generation of ROS during hypoxia raised concerns regarding direct involvement of ROS in affecting PHD activity [29]. There are reports of HIF-1 $\alpha$  stabilization by inhibition of PHD activity due to mitochondrial ROS generation by Ang II in vascular smooth muscle cell [24], [30]. Transition metals like cobalt and nickel also affect PHD activity by depleting cellular ascorbate level [15]. So, ROS-mediated inhibition of PHD activity was reported as major cellular mechanism in both hypoxic and normoxic conditions. In contrast, we detected that insulin increased HIF-1 $\alpha$  transcription by a ROS sensitive Sp1 activation mechanism (Fig. 7-8)

(a) 化学论文片段<sup>①</sup>

LIG can inhibit the proliferation and cell cycle progression of vascular smooth muscle cells, associated to basic fibroblast growth factor stimulation, through the reduction of reactive oxygen species and/or the suppression of the MAPK pathway [2]. LIG also inhibits vasoconstriction induced by norepinephrine bitartrate and calcium chloride on rat abdominal aorta segments [3]. Hence, LIG is considered to be an effective agent to reduce vascular resistance; thereafter, increase blood flow and enhance microcirculation to prevent cardiovascular diseases, including atherosclerosis and hypertension [4], [5].

(b) 物理论文片段<sup>②</sup>

图1 学科论文引文内容实例片断

根据以上设想, 本文设计研究框架图如图3所示。首先, 将全文数据中的引文内容和学科信息解析出来, 本文仅将引文内容的范围限定为包含引用标记的引用句, 以及引用句的前两句和引用句的后两句, 最多共为五句话<sup>[32]</sup>。同时, 获取术语集中的术语。然后, 查找学科引文内容中的术语, 统计匹配的术语频次和重复率。接着, 根据学科交叉度公式计算学科之间的学科交叉度。最后, 进行学科交

叉点贡献分析, 探究学科交叉背后的学科知识交叉情况。

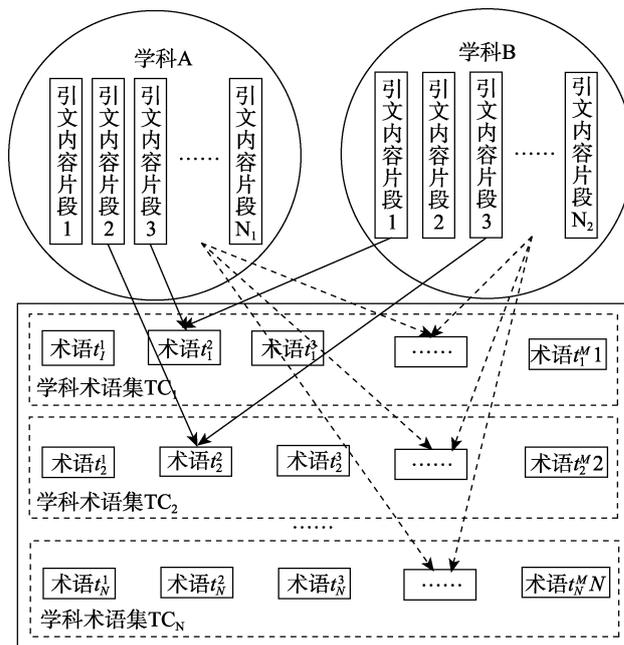


图2 术语共被引示意图

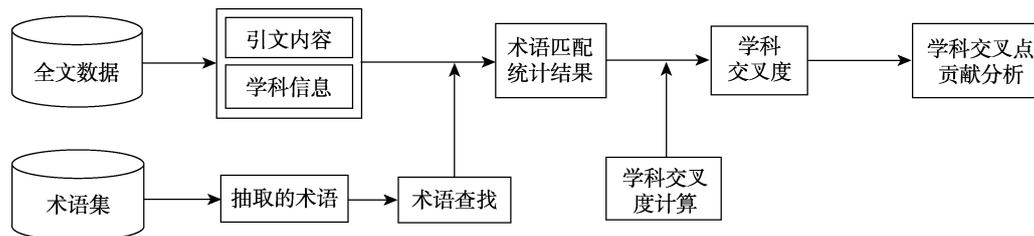


图3 术语引用视角下的学科交叉测度研究框架图

### 3.2 实现方法

#### (1) 学术论文引文内容抽取

随着信息技术的不断发展, 可解析的全文数据库逐渐得到开发, 例如, PubMed/BioMed Central、Citeseer和 arXiv 等数据库都提供可进行数据格式解析和文本内容挖掘的全文信息。利用文本挖掘和自然语言处理技术, 对全文信息中的引文内容信息进行抽取和保存。

#### (2) 术语获取

由于语言资源的不断丰富与自然语言处理技术的不断发展, 术语获取主要有两种途径: 基于机器的术语自动抽取、基于术语词典的术语获取。基于机器的术语自动抽取方法主要分为: 基于语言学的

方法、基于统计的方法以及混合方法<sup>[33]</sup>。术语词典主要包括《美国国会图书馆标题表》、《中国分类主题词表》、《医学主题词表》等。利用文本挖掘和自然语言处理技术, 可以对术语词典中的术语进行抽取和保存。

#### (3) 学科交叉度计算

学科交叉度的计算方法有很多, 如余弦相似度、杰卡德相似系数等。本文根据余弦相似度来计算学科间引用术语的相似度, 根据学科引用术语的相似度从内容分析角度量化表示为学科交叉度。任意两个学科  $D_1$  和  $D_2$  之间的相似系数  $\text{Sim}(D_1, D_2)$  指两个学科引用术语的相关程度。基于向量空间模型理论<sup>[34]</sup>, 把术语特征项  $t_1, t_2, \dots, t_n$  看成一个  $n$  维坐标系, 而

① <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0062128>, 获取日期: 2017年3月29日。

② <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0066598>, 获取日期: 2017年3月29日。

术语对应的词频  $w_1, w_2, \dots, w_n$  则为相应的权重，因此，一个学科就被表示为  $n$  维空间中的一个向量。设学科  $D_1$  和  $D_2$  表示向量空间模型中的两个向量：

$$D_1 = D_1(w_{11}, w_{12}, \dots, w_{1n})$$

$$D_2 = D_2(w_{21}, w_{22}, \dots, w_{2n})$$

于是，我们可以利用  $n$  维空间中两个向量之间的距离，即两个向量夹角的余弦值来表示学科间引用术语的相似系数。计算方法如公式 (1) 所示。

$$\text{Sim}(D_1, D_2) = \cos \theta = \frac{\sum_{k=1}^n w_{1k} \times w_{2k}}{\sqrt{\sum_{k=1}^n w_{1k}^2 \sum_{k=1}^n w_{2k}^2}} \quad (1)$$

若  $\text{Sim}(D_1, D_2) > 0$ ，则存在术语  $t_i$  使得学科  $D_1$  和  $D_2$  的术语词频  $w_{1i}$  和  $w_{2i}$  都不为零，则学科  $D_1$  和  $D_2$  共同引用了相同的术语  $t_i$ ，学科  $D_1$  和  $D_2$  在术语  $t_i$  下一定交叉，且学科交叉度的大小为学科间引用术语的相似系数  $\text{Sim}(D_1, D_2)$ 。

同理，若  $\text{Sim}(D_1, D_2) = 0$ ，则对任何一个术语  $t_i$ ，都有  $D_1$  和  $D_2$  的术语词频  $w_{1i}$  和  $w_{2i}$  至少有一个为零，则学科  $D_1$  和  $D_2$  基于所选术语是不交叉的。例如，当  $D_1 = D_1(1, 0, 0)$ ， $D_2 = D_2(0, 2, 0)$ ， $\text{Sim}(D_1, D_2) = 0$ ，学科  $D_1$  和  $D_2$  没有共同引用相同的术语，学科  $D_1$  和  $D_2$  不交叉。

(4) 学科交叉点贡献度量

利用术语对学科构建向量不仅可以计算学科间的交叉度  $\text{Sim}(D_1, D_2)$ ，而且可以根据术语交叉度公式得到任意一个分量对两学科间交叉度的贡献  $\frac{w_{1k} w_{2k}}{\sqrt{\sum_{k=1}^n w_{1k}^2 \sum_{k=1}^n w_{2k}^2}}$ 。假设学科  $D_1$  的术语向量是  $D_1 = D_1(1, 0, 0)$ ，假设学科  $D_2$  的术语向量是  $D_2 = D_2(10, 0, 0)$ ，则学科  $D_1$  和  $D_2$  的交叉度为  $10/10=1$ 。术语  $t_1$  对这两个学科交叉度的贡献为  $1/1=100\%$ ，即学科  $D_1$  和  $D_2$  之间的交叉完全由术语  $t_1$  贡献。

(5) 学科交叉下术语的范围分析

本文认为当某两个学科在某个术语集上存在术语交叉，则这两个学科必然交叉，即当  $\text{Sim}(D_1, D_2) > 0$  时，学科  $D_1, D_2$  必然交叉，这个结论已经得证。本节需要对术语集的覆盖范围进行相应的说明。如图 4 所示，当学科  $D_1, D_2$  交叉术语中的部分是术语集  $R$  的子集，则学科  $D_1, D_2$  在  $R$  下交叉是学科  $D_1, D_2$  交叉的必要而不充分条件，也就是说，此时存在学科  $D_1, D_2$  之间在术语集  $R$  以外的交叉术语。

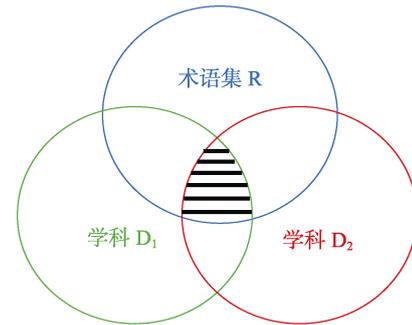


图 4 基于术语集的学科交叉度示意图

## 4 实证结果分析

### 4.1 实证数据描述

本节进行实证分析，在实证过程中对学科术语集进行简化，选取医学术语集为本文的实证术语集，选取 PLOS ONE 和《医学主题词表》为实证数据来源。

(1) 学术论文引文内容抽取

本文选取 PLOS ONE 为实证数据库。PLOS ONE (<http://journals.plos.org/plosone/>) 是目前学术界非常有影响力的开放存取期刊，载文涉及从自然科学到社会科学的 10 多种学科。该期刊对所发表论文提供结构化全文下载，易解析，是本文理想的数据来源。本文以“生物学 (Biology, BI)”、“化学 (Chemistry, CH)”、“物理学 (Physics, PH)”、“计算机科学 (Computer Science, CS)”、“数学 (Mathematics, MA)”和“心理学 (Psychology, PS)”这 6 个学科领域为样本，选取 2006 年 12 月 20 日到 2014 年 12 月 18 日这段时间内所发表的论文。为保持各学科间样本的均衡性，论文数高于 300 篇的学科均随机采样 300 篇。总共获取 1725 篇论文。

对结构化全文数据进行解析，抽取出每条参考文献的基本信息以及参考文献在全文中所对应的引用内容，从 1725 篇源文献中共获得 53869 条引文内容，如表 1 所示。其中，本文共抽取了 214187 个句子，引文内容的平均句子数约为 4 句。

表 1 学科文献和引文内容数

学科	论文数	引用内容	句子总数	平均句子数
生物学 (BI)	300	9927	40076	4.04
化学 (CH)	300	10347	41923	4.05
物理学 (PH)	282	9604	38501	4.01
计算机科学 (CS)	243	7657	29401	3.84
数学 (MA)	300	6342	24647	3.89
心理学 (PS)	300	9992	39639	3.97
总量统计	1725	53869	214187	3.98

## (2) 术语获取

利用《医学主题词表》获取医学术语。《医学主题词表》(Medical Subject Headings, 简称 MeSH, <https://www.nlm.nih.gov/mesh/>) 是美国国立医学图书馆编制的权威性主题词表, 应用广泛。许多学者将 MeSH 称为医学领域的本体<sup>[35]</sup>。MeSH 由主题词变更表、字顺表和树状结构表组成, 其中字顺表和树状结构表是 MeSH 的主要组成部分。字顺表是医学主题词表的主表。树状结构表将字顺表中互不联系的主题词通过主题词所属学科体系和逻辑关系, 加上一些必要的非主题词组成树状等级结构。

本文从 2017 版本的 MeSH 中共获取医学术语共 142968 个, 分布在 115 个主题下, 具体的主题分布情况如附表 1 所示, 其中词语量表示医学主题下的医学术语数量。

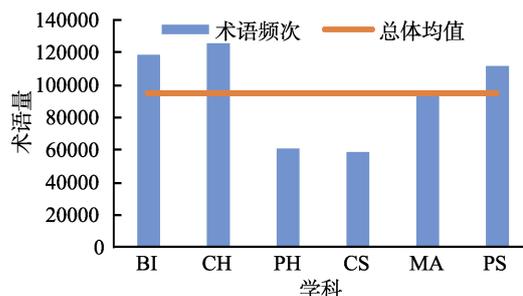
## 4.2 术语匹配统计结果分析

学科引文内容和医学术语匹配的统计结果如图 5 所示。化学学科与 MeSH 词表匹配到了最多的术语 (125815 个), 其次是生物学 (118486)、心理学 (112363) 和数学 (96272), 这几个学科匹配到的术语均高于均值 (95566)。物理学和计算机科学是与医学术语交叉最少的两个学科, 分别为 61545 和 58914。计算机学科更加偏向于工科, 因而交叉的量最少。但这类少数的交叉在 PLOS 里有一个专门的期刊刊登相关的文章, 叫做计算生物学 (Computational-Biology)。物理学和医学术语的交叉也较少。可能的原因是物理学是研究物质、能量的本质与性质, 以及它们彼此之间相互作用的科学, 与医学方面的交叉也主要集中在物理技术对医学的具体应用之中, 如核物理。

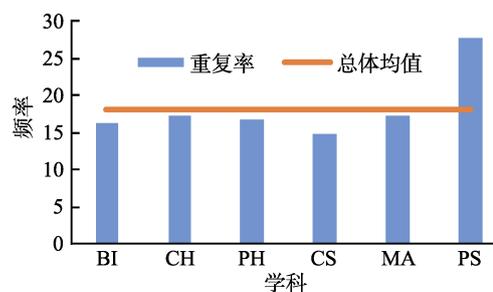
从术语的重复率来看, 心理学中术语的重复率时最高的 (27.7 次) 明显高于总体平均值 (18.0 次), 可能的原因是心理学和医学的交叉领域较为狭小, 主要集中在神经科学、临床试验等领域。与医学交叉术语频次最高的化学和生物学就表现出较低的术语重复率, 也就是说, 医学和化学、生物学等学科广泛的交叉会相对降低术语重复的几率。物理学、数学以及计算机科学这类术语频次相对较低的学科重复率持平。

## 4.3 学科交叉度比较和分析

根据公式 (1) 算得各学科之间的交叉度, 分布如图 6 所示。



(a) 术语频次



(b) 重复率

图 5 学科引文内容与医学术语匹配结果统计图

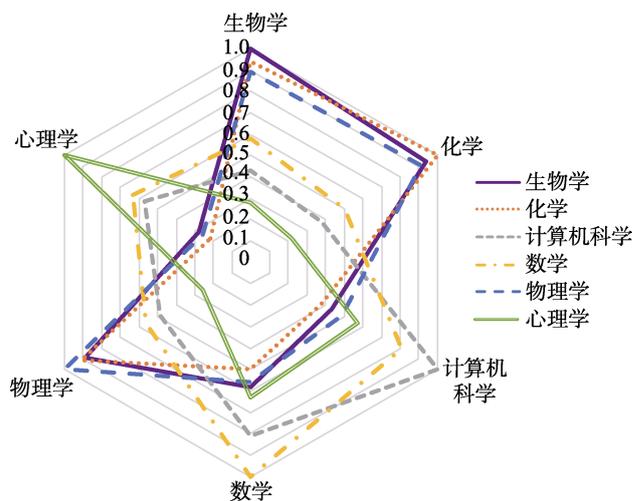


图 6 学科交叉度可视化

自然学科间, 学科交叉度普遍较高。生物学和化学的交叉度最高 (0.94), 其次是化学和物理学 (0.91), 物理学和生物学 (0.89), 计算机科学和数学 (0.81)。生物学、化学、物理学这三大学科之间的知识交叉度十分高, 这和相关研究的结果是相符的; 现实中, 如生物化学、化学物理、数学物理方法都是各个学科中的核心课程, 这也体现了学科之间的交叉度极高。

心理学和其他学科的交叉度均比较低。较与其他学科之间的交叉度, 心理学和化学最小 (0.22), 其次是心理学和物理学 (0.28)、心理学和生物学 (0.26)。主要的原因可能是心理学是一门应用性学

科，而且主要研究人与动物的心理和行为活动的规律，和生物、化学这样的自然学科很难形成广泛的交叉，相反和医学这样的学科却有着紧密的联系（图 5）。

数学作为工具学科，和所有学科都有知识交叉现象的存在，且交叉度较高。数学和计算机科学的交叉度较高，为 0.81，因为计算机科学是系统性研究信息与计算的理论基础以及它们在计算机系统中如何实现与应用的实用技术的学科，数学计算理论是该学科的基础性理论之一。因此，二者学科交叉度较大。和其他学科之间，数学均能有中等偏上的交叉度水平（0.5 左右），其中心理学和数学的交叉度为 0.63，更加能够体现数学具备广泛的交叉性。

此外，值得注意的是，各个学科之间都存在一定程度的知识交叉，可能的原因是本文所选取的学科都明显是自然学科相关的学科，学科之间难免出现知识交叉。但是，由于 MeSH 主题词表存在削弱学科之间交叉度的可能性，因此这些观测的学科实际交叉度可能比学科间测度得到的学科交叉度更高。

#### 4.4 学科交叉点贡献结果分析

根据各学科的术语向量表示，我们得到了每两个学科间对学科交叉度贡献最大的 30 个术语（附表 2）。下面的分析仅以生物学和其他 5 个学科的术语交叉为例分析各学科之间的交叉术语。

总体来看，在生物和其他学科之间的术语交叉中，“cell”、“protein”和“gene”对交叉度的贡献都比较大，大致都在前 10 的位置。但在不同的学科之间，比重却各有不同。在生物学和计算机科学之间，“cell”对交叉度的贡献只有 10.31%，而在生物和化学的交叉点中，“cell”的贡献为 34.49%。而在生物和心理学的交叉术语中，“cell”、“protein”以及“gene”都不在前 30 的交叉术语中。

而且 Top30 的交叉术语也反映出不同学科自身的特色。在生物和计算机科学的交叉术语中，“method”对交叉度的贡献最高（11.3%），超过“cell”和“protein”。其他能够反映计算机科学的领域知识也在其中，如“community”、“algorithm”以及“measure”等。数学和生物的交叉术语中，体现为数学学科方面的术语有“method”、“probability”、“scale”等。这些术语都反映出数学作为工具型学科与生物的交叉主要体现在数学能够为生物等其他学科提供定量的统计分析方法。在生物与物理的交叉术语中，能够反映出物理学的术语有“tumor”、

“apoptosis”等。这些术语反映出物理学在生物医学方面的应用主要体现在：核物理相关技术对癌症的治疗以及检测等方面的运用。由于这一类的术语并不多见，因此生物和物理之间的交叉度并不如与化学的交叉度高。在生物学和心理学的交叉术语中，更多与人类、行为有关的术语大量出现，如“child”、“adult”、“patient”、“treatment”、“family”等词语。

## 5 讨论

由于六个学科的研究对象不同，导致学科的交叉程度表现出明显的区别。生物学、化学和物理学属于自然科学<sup>[36]</sup>，研究对象是客观的自然世界，生物学、化学和物理学的学科交叉度高。心理学主要研究人与动物的心理和行为活动规律<sup>[37]</sup>，心理学在与其余五个学科的交叉中自成一派。数学与自然科学甚至社会科学有很大的区别<sup>[37]</sup>，是科学不可或缺的工具<sup>[38]</sup>。数学学科的交叉行为更为基础，实验结果表现为与其他学科都有较为广泛的交叉。计算机科学自诞生起，对人类社会的进步作用巨大。计算机科学与数学相辅相成，互为基础，学科交叉程度高。

学科间的交叉现象不能仅仅通过一个或若干个指标进行测度。现有的相关研究工作中，学科间的交叉现象主要通过若干个指标进行测度，如 diversity、variety 和 balance<sup>[19]</sup>。本研究表明尽管两个学科之间的交叉度在数值上可能比较相近，但是，学科间的交叉结构却各不相同。这些微观层面的差异反映出两学科之间各自的研究领域以及这些领域在两个学科中相互渗透的情况。举例来看，数学作为工具性学科，并不是其所有的领域都会和其他学科进行互相渗透，而仅仅一部分领域，如统计分析、概率论等方面的内容在多个学科中被广泛的使用。对于生物和化学这样的学科，它们之间不仅存在着较高的学科交叉度，而且它们在多个方面都存在着交叉现象。因此，单单依靠某几个指标是无法观测出这些细粒度的学科交叉点。学科间交叉的过程变得模糊不清。

不仅如此，学科间术语交叉的测度有利于学科间交叉主题的测度。本研究的实证也反映出，学科间利用术语引用测度的交叉度可以帮助我们进行学科间主题交叉的测度。本文目前并没有对各学科间交叉的术语进行主题分析，但事实上，两学科之间的交叉是通过一次又一次的带有明确研究问题的科学研究完成的。这些研究都同时反映出交叉学科间各自具备的主题或者共有的主题。从对附表 2 的举

例分析来看,术语作为学术交流的基本单位反映出不同学科之间的交叉程度以及交叉侧重,比如,生物和心理学侧重对人为行为和心理疾病的研究,因此如“child”、“family”等相关的术语具有较高的术语贡献度。而生物和化学的交叉主要体现在分子层面的研究上,如“cell”、“protein”和“gene”。因此本研究提出的研究方法对学科之间交叉主题的研究具有重要作用。

## 6 结论

学科交叉研究一直是图书情报学科的重要研究领域,具有极重要的研究价值,体现在如跨学科知识转移、新交叉学科的产生与演化等。本文在总结前人的研究发现当前主流的借助于文献之间引证关系具有明显的弊端——难以识别学科之间具体的交叉点。为了解决这一局限,本文提出了利用统计计算学科之间共用术语的方法计算学科间交叉度的研究思路。研究结果显示该测度方法不仅能够测度学科间的交叉程度,而且能够提供学科间的交叉知识。这样不仅回答了两学科间是否交叉的问题还回答了这两个学科在哪里交叉的问题。

本研究所提出的新的学科交叉现象的研究方法,相较于传统的基于引证关系研究的学科交叉研究,具有极大的优势,主要表现在对学科交叉的具体知识交叉点的探索。从知识交叉点出发测度学科交叉度。本研究的应用前景主要包括:①基于主题词表的主题结构信息,对学科间交叉主题进行观测,对主题学科交叉度进行测量,以期对学科的交叉行为有更为微观的把握,挖掘学科交叉的类型,寻找学科交叉模式的差异,探究学科交叉动机和原因。②引入时间戳信息,如出版时间等,对学科交叉的演化过程进行更为动态化的研究分析,并分析原因。③进一步对基于术语和引文内容、基于引证关系等的交叉学科测度方法进行量化比较分析,从多个角度对交叉学科测度进行更为全面的考察。

但是,本研究也存在着较多的不足和局限性以期在下一步工作中解决。比如,本研究利用了 MeSH 测度学科间的交叉度使得学科交叉度可能小于真实值。尽管本研究已经在实现方法部分论证了利用与术语集测度学科间交叉度的可行性,但是由于 MeSH 的领域特性,学科间交叉度的观测值存在小于真实值的可能性,给本研究六学科间交叉度的测度带来一定的误差。因此,本研究的下一步工作将着重选取更加全面和广泛,包含更多学科的术语集以期获

取更加精确的学科交叉度的测度。另外本研究的实验数据规模较小,学科选取存在一定的局限性。下一步工作需要获取更多的实验数据,引入更多的学科,合并学科进行比较(如文科和理科),以期获得更有效和更一般性的结论。

## 参 考 文 献

- [1] 张春美,郝凤霞,闫宏秀. 学科交叉研究的神韵——百年诺贝尔自然科学奖探析[J]. 科学技术与辩证法, 2001, 18(6): 63-67.
- [2] Yan E J. Finding knowledge paths among scientific disciplines[J]. Journal of the Association for Information Science and Technology, 2014, 65(11): 2331-2347.
- [3] Zhang L, Rousseau R, Glänzel W. Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account[J]. Journal of the Association for Information Science and Technology, 2016, 67(5): 1257-1265.
- [4] Wagner C S, Roessner J D, Bobb K, et al. Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature[J]. Journal of Informetrics, 2011, 5(1): 14-26.
- [5] Small H G. Cited documents as concept symbols[J]. Social Studies of Science, 1978, 8(3): 327-340.
- [6] 马晓雷. 被引内容分析: 探究领域知识结构的新方法尝试[M]. 北京: 外语教学与研究出版社, 2011.
- [7] 冯志伟. 术语学中的概念系统与知识本体[J]. 术语标准化与信息技术, 2006(1): 9-15.
- [8] 许海云,尹春晓,郭婷,等. 学科交叉研究综述[J]. 图书情报工作, 2015, 59(5): 119-127.
- [9] Rafols I, Meyer M. Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience[J]. Scientometrics, 2010, 82(2): 263-287.
- [10] Vugteveen P, Lenders R, Van den Besselaar P. The dynamics of interdisciplinary research fields: the case of river research[J]. Scientometrics, 2014, 100(1): 73-96.
- [11] van Raan A F J, van Leeuwen T N. Assessment of the scientific basis of interdisciplinary, applied research: Application of bibliometric methods in Nutrition and Food Research[J]. Research Policy, 2002, 31(4): 611-632.
- [12] Small H. Maps of science as interdisciplinary discourse: co-citation contexts and the role of analogy[J]. Scientometrics, 2010, 83(3): 835-849.
- [13] Porter A L, Roessner J D, Cohen A S, et al. Interdisciplinary research: meaning, metrics and nurture[J]. Research Evaluation, 2006, 15(3): 187-195.
- [14] National Academy of Sciences, National Academy of Engineering, Institute of Medicine. Facilitating Interdisciplinary Re-

- search[M]. Washington DC: The National Academies Press, 2004.
- [15] Liu Y X, Rafols I, Rousseau R. A framework for knowledge integration and diffusion[J]. *Journal of Documentation*, 2012, 68(1): 31-44.
- [16] Liu C, Shan W, Yu J. Shaping the interdisciplinary knowledge network of China: a network analysis based on citation data from 1981 to 2010[J]. *Scientometrics*, 2011, 89(1): 89-106.
- [17] Porter A L, Chubin D E. An indicator of cross-disciplinary research[J]. *Scientometrics*, 1985, 8(3-4): 161-176.
- [18] Porter A L, Cohen A S, Roessner J D, et al. Measuring researcher interdisciplinarity[J]. *Scientometrics*, 2007, 72(1): 117-147.
- [19] Stirling A. A general framework for analysing diversity in science, technology and society[J]. *Journal of the Royal Society Interface*, 2007, 4(15): 707-719.
- [20] 刘盛博, 丁堃, 张春博. 引文分析的新阶段: 从引文著录分析到引用内容分析[J]. *图书情报知识*, 2015(3): 25-34.
- [21] Teufel S, Siddharthan A, Tidhar D. Automatic classification of citation function[C]// *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2006: 103-110.
- [22] 陆伟, 孟睿, 刘兴帮. 面向引用关系的引文内容标注框架研究[J]. *中国图书馆学报*, 2014, 40(6): 93-104.
- [23] 张梦莹, 卢超, 郑茹佳, 等. 用于引文内容分析的标准化数据集构建[J]. *图书馆论坛*, 2016, 36(8): 48-53.
- [24] Athar A. Sentiment analysis of citations using sentence structure-based features[C]// *Proceedings of the ACL 2011 Student Session*. Stroudsburg: Association for Computational Linguistics, 2011: 81-87.
- [25] 刘盛博, 丁堃, 张春博. 基于引用内容性质的引文评价研究[J]. *情报理论与实践*, 2015, 38(3): 77-81.
- [26] 祝清松, 冷伏海. 基于引文内容分析的高被引论文主题识别研究[J]. *中国图书馆学报*, 2014, 40(1): 39-49.
- [27] Liu S B, Chen C M. The differences between latent topics in abstracts and citation contexts of citing papers[J]. *Journal of the American Society for Information Science and Technology*, 2013, 64(3): 627-639.
- [28] 刘洋, 崔雷. 引文上下文在文献内容分析中的信息价值研究[J]. *图书情报工作*, 2014, 58(6): 101-104.
- [29] 张艺蔓, 马秀峰, 程结晶. 融合引文内容和全文本引文分析的知识流动研究[J]. *情报杂志*, 2015, 34(11): 50-54, 49.
- [30] 章成志, 徐庶睿, 卢超. 利用引文内容监测多学科交叉现象的方法与实证[J]. *图书情报工作*, 2016, 60(19): 108-115.
- [31] Small H. Co-citation in the scientific literature: A new measure of the relationship between two documents[J]. *Journal of the American Society for Information Science*, 1973, 24(4): 265-269.
- [32] Mei Q Z, Zhai C X. Generating impact-based summaries for scientific literature[C]// *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, 2008, 8: 816-824
- [33] 章成志. 多语言领域本体学习研究[M]. 南京: 南京大学出版社, 2012..
- [34] Satlon G, Wong A, Yang C S. A vector space model for automatic indexing[J]. *Communications of the ACM*, 1975, 18(11): 613-620.
- [35] 于双成, 逢大欣, 李占兵. 医学主题词表(MeSH)专指度研究[J]. *情报学报*, 1995, 14(6): 449-452.
- [36] 钱学森. 现代科学技术[N]. *人民日报*, 1977-12-09.
- [37] 胡作玄. 数学研究对象的演化[J]. *自然辩证法研究*, 1992, 8(1): 22-28
- [38] 李醒民. 知识的三大部类: 自然科学、社会科学和人文科学[J]. *学术界*, 2012(8): 5-33, 286.

附表 1 医学主题的中英文对照和主题词语量

主题序号	英文名	中文名	词语量
A01	Body Regions	身体各部分	296
A02	Musculoskeletal System	肌肉骨骼系统	929
A03	Digestive System	消化系统	285
A04	Respiratory System	呼吸系统	201
A05	Urogenital System	泌尿系统	274
A06	Endocrine System	内分泌系统	327
A07	Cardiovascular System	心血管系统	421
A08	Nervous System	神经系统	2519
A09	Sense Organs	感觉器官	427
A10	Tissues	组织	679
A11	Cells	细胞	2499
A12	Fluids and Secretions	体液和分泌物	191
A13	Animal Structures	动物结构	203
A14	Stomatognathic System	口颌系统	291

( 待续 )

续附表 1

主题序号	英文名	中文名	词语量
A15	Hemic and Immune Systems	血液和免疫系统	550
A16	Embryonic Structures	胚胎结构	257
A17	Integumentary System	皮肤系统	40
A18	Plant Structures	植物结构	179
A19	Fungal Structures	真菌结构	30
A20	Bacterial Structures	细菌结构	29
B01	Eukaryota	真核生物	11429
B02	Archaea	古生菌	112
B03	Bacteria	细菌	1275
B04	Viruses	病毒	2362
B05	Organism Forms	生物形态	113
C01	Bacterial Infections and Mycoses	细菌感染和真菌病	1506
C02	Virus Diseases	病毒疾病	944
C03	Parasitic Diseases	寄生虫病	538
C04	Neoplasms	肿瘤	4195
C05	Musculoskeletal Diseases	肌肉骨骼系统疾病	2450
C06	Digestive System Diseases	消化系统疾病	1372
C07	Stomatognathic Diseases	口腔颌疾病	1065
C08	Respiratory Tract Diseases	呼吸道疾病	1091
C09	Otorhinolaryngologic Diseases	耳鼻喉疾病	744
C10	Nervous System Diseases	神经系统疾病	9770
C11	Eye Diseases	眼疾病	1458
C12	Male Urogenital Diseases	男性生殖器疾病	1130
C13	Female Urogenital Diseases and Pregnancy Complications	女性生殖器疾病和妊娠并发症	1611
C14	Cardiovascular Diseases	心血管系统疾病	2372
C15	Hemic and Lymphatic Diseases	血液和淋巴系统疾病	1873
C16	Congenital, Hereditary, and Neonatal Diseases and Abnormalities	先天性、遗传性和新生儿疾病和畸形	6412
C17	Skin and Connective Tissue Diseases	皮肤和结缔组织疾病	2222
C18	Nutritional and Metabolic Diseases	营养和代谢性疾病	3010
C19	Endocrine System Diseases	内分泌系统疾病	1042
C20	Immune System Diseases	免疫系统疾病	1743
C21	Disorders of Environmental Origin	环境因素诱发疾病	12
C22	Animal Diseases	动物疾病	432
C23	Pathological Conditions, Signs and Symptoms	病理条件、体征和症状	4673
C24	Occupational Diseases	职业病	105
C25	Chemically-Induced Disorders	化学诱导的疾病	564
C26	Wounds and Injuries	创伤与损伤	1331
D01	Inorganic Chemicals	无机化合物	1059
D02	Organic Chemicals	有机化合物	12529
D03	Heterocyclic Compounds	杂环化合物	10346
D04	Polycyclic Compounds	多环碳氢化合物	3507
D05	Macromolecular Substances	大分子物质	1628
D06	Hormones, Hormone Substitutes, and Hormone Antagonists	激素、激素代用品和激素拮抗剂	1555
D08	Enzymes and Coenzymes	酶与辅酶	11368
D09	Carbohydrates	碳水化合物	1816
D10	Lipids	脂类	1590
D12	Amino Acids, Peptides, and Proteins	氨基酸、肽、蛋白质	23657
D13	Nucleic Acids, Nucleotides, and Nucleosides	核酸、核苷类和核苷酸类	1443
D20	Complex Mixtures	复杂混合物	837
D23	Biological Factors	生物因素	5601
D25	Biomedical and Dental Materials	生物医学和牙科科学	789

( 待续 )

续附表 1

主题序号	英文名	中文名	词语量
D25	Biomedical and Dental Materials	生物医学和牙科科学	789
D26	Pharmaceutical Preparations	药物制剂	423
D27	Chemical Actions and Uses	化学反应和用途	2735
E01	Diagnosis	诊断	3427
E02	Therapeutics	治疗	2822
E03	Anesthesia and Analgesia	麻醉和镇痛	130
E04	Surgical Procedures, Operative	外科操作, 手术	2415
E05	Investigative Techniques	包埋技术	5123
E06	Dentistry	牙科	684
E07	Equipment and Supplies	设备和供应	1684
F01	Behavior and Behavior Mechanisms	行为和行为机制	1982
F02	Psychological Phenomena and Processes	心理现象和过程	990
F03	Mental Disorders	精神疾病	1294
F04	Behavioral Disciplines and Activities	行为训练和活动	652
G01	Physical Phenomena	物理现象	833
G02	Chemical Phenomena	化学现象	1799
G03	Metabolism	代谢	649
G04	Cell Physiological Phenomena	细胞生理现象	704
G05	Genetic Phenomena	遗传现象	2764
G06	Microbiological Phenomena	微生物学现象	238
G07	Physiological Phenomena	生理现象	1806
G08	Reproductive and Urinary Physiological Phenomena	生殖和泌尿生理现象	559
G09	Circulatory and Respiratory Physiological Phenomena	循环和呼吸生理现象	628
G10	Digestive System and Oral Physiological Phenomena	消化系统和口腔生理现象	128
G11	Musculoskeletal and Neural Physiological Phenomena	肌肉骨骼和神经生理现象	849
G12	Immune System Phenomena	免疫系统现象	436
G13	Integumentary System Physiological Phenomena	外皮系统的生理现象	52
G14	Ocular Physiological Phenomena	眼生理现象	110
G15	Plant Physiological Phenomena	植物生理现象	97
G16	Biological Phenomena	生物学现象	650
H01	Natural Science Disciplines	自然科学学科	621
H02	Health Occupations	卫生职业	587
I01	Social Sciences	社会科学	1931
I02	Education	教育	296
I03	Human Activities	人类活动	339
J01	Technology, Industry, and Agriculture	工艺学、工业和农业	1984
J02	Food and Beverages	食物和饮料	408
J03	Non-Medical Public and Private Facilities	非医疗公共和私人设施	169
K01	Humanities	人文科学	679
L01	Information Science	情报科学	1906
M01	Persons	人群	1315
N01	Population Characteristics	人口特征	634
N02	Health Care Facilities, Manpower, and Services	卫生保健设施, 人力和服务	2049
N03	Health Care Economics and Organizations	卫生保健经济和组织	1709
N04	Health Services Administration	卫生服务行政管理	1442
N05	Health Care Quality, Access, and Evaluation	卫生保健质量, 实施, 评估	1693
N06	Environment and Public Health	环境与公共卫生	2324
V01	Publication Components	出版组件	60
V02	Publication Formats	出版格式	184
V03	Study Characteristics	研究特征	25
Z01	Geographic Locations	地理位置	742

附表2 学科间 Top 30 交叉术语

学科交叉	交叉术语及术语贡献度 (Top 30)
BI ∩ CH	<b>cell</b> (34.49%) <b>protein</b> (21.60%) <b>gene</b> (7.60%) human(2.83%) role(2.68%) mouse(2.08%) method(1.37%) treatment(1.16%) patient(1.13%) concentration(0.95%) dna(0.85%) table(0.74%) form(0.72%) growth(0.72%) membrane(0.62%) time(0.61%) strain(0.59%) plant(0.55%) apoptosis(0.54%) tumor(0.53%) cancer(0.48%) enzyme(0.46%) disease(0.44%) population(0.41%) tissue(0.38%) temperature(0.36%) infection(0.35%) production(0.34%) genome(0.34%) mutation(0.31%)
BI ∩ CS	<b>method</b> (11.30%) <b>cell</b> (10.31%) <b>protein</b> (6.80%) time(5.97%) human(4.96%) population(3.56%) role(2.50%) gene(2.18%) form(2.01%) patient(1.99%) <b>set</b> (1.81%) <b>community</b> (1.76%) will(1.74%) table(1.69%) id(1.65%) est(1.39%) <b>measure</b> (1.34%) <b>algorithm</b> (1.26%) research(0.94%) growth(0.91%) forest(0.91%) disease(0.83%) treatment(0.81%) probability(0.72%) temperature(0.69%) scale(0.69%) concentration(0.68%) production(0.68%) behavior(0.62%) standard(0.57%)
BI ∩ MA	<b>cell</b> (18.08%) <b>method</b> (7.16%) <b>gene</b> (6.12%) <b>patient</b> (4.84%) <b>population</b> (4.36%) time(4.32%) protein(4.32%) human(3.03%) infection(2.49%) role(2.18%) treatment(2.15%) disease(1.74%) mouse(1.58%) <b>form</b> (1.53%) <b>table</b> (1.44%) will(1.12%) <b>risk</b> (1.10%) strain(1.07%) id(0.93%) measure(0.85%) est(0.80%) community(0.76%) set(0.74%) growth(0.71%) <b>probability</b> (0.64%) <b>scale</b> (0.61%) association(0.56%) animal(0.54%) mutation(0.54%) cancer(0.54%)
BI ∩ PH	cell(45.42%) <b>protein</b> (16.50%) gene(3.27%) method(2.12%) role(1.98%) human(1.78%) time(1.66%) concentration(1.46%) form(1.03%) membrane(1.03%) treatment(0.96%) mouse(0.86%) table(0.81%) growth(0.78%) patient(0.75%) strain(0.70%) dna(0.66%) tissue(0.62%) <b>tumor</b> (0.58%) population(0.57%) neuron(0.52%) id(0.42%) will(0.41%) disease(0.38%) <b>apoptosis</b> (0.37%) est(0.34%) production(0.32%) mutation(0.32%) <b>enzyme</b> (0.31%) temperature(0.30%)
BI ∩ PS	<b>patient</b> (9.98%) <b>treatment</b> (5.90%) time(4.81%) population(4.70%) role(4.31%) human(4.29%) method(4.17%) association(2.57%) risk(2.41%) <b>child</b> (2.38%) research(2.20%) measure(2.07%) scale(1.99%) who(1.98%) id(1.78%) form(1.74%) behavior(1.69%) est(1.57%) community(1.32%) table(1.28%) will(1.24%) affect(1.21%) <b>report</b> (1.14%) health(1.06%) ability(1.01%) <b>family</b> (0.97%) disease(0.94%) <b>adult</b> (0.90%) prevalence(0.86%) <b>culture</b> (0.81%)
CH ∩ CS	<b>method</b> (11.44%) cell(10.23%) protein(10.12%) human(6.09%) time(5.06%) role(2.76%) form(2.26%) patient(2.00%) table(1.98%) gene(1.94%) set(1.81%) population(1.73%) id(1.67%) temperature(1.63%) est(1.40%) <b>solution</b> (1.27%) concentration(1.22%) <b>algorithm</b> (1.14%) will(1.12%) research(1.06%) measure(0.94%) treatment(0.93%) standard(0.81%) growth(0.78%) community(0.76%) disease(0.74%) production(0.65%) <b>relative</b> (0.62%) hand(0.59%) behavior(0.59%)
CH ∩ MA	cell(18.51%) method(7.48%) protein(6.63%) gene(5.61%) patient(5.01%) human(3.83%) time(3.79%) treatment(2.55%) role(2.48%) population(2.19%) form(1.78%) table(1.75%) disease(1.60%) mouse(1.55%) infection(1.39%) cancer(1.03%) <b>risk</b> (1.00%) id(0.97%) est(0.83%) temperature(0.82%) concentration(0.81%) set(0.76%) will(0.75%) solution(0.71%) mutation(0.70%) strain(0.69%) tumor(0.68%) animal(0.66%) growth(0.63%) <b>measure</b> (0.61%)
CH ∩ PH	cell(39.41%) protein(21.48%) gene(2.54%) concentration(2.28%) human(1.91%) role(1.91%) method(1.88%) membrane(1.26%) time(1.23%) form(1.01%) treatment(0.96%) table(0.83%) mouse(0.72%) dna(0.71%) tumor(0.68%) patient(0.66%) enzyme(0.65%) tissue(0.62%) temperature(0.61%) growth(0.58%) peptide(0.50%) solution(0.50%) cancer(0.43%) inhibition(0.40%) strain(0.39%) id(0.37%) in vitro(0.36%) lipid(0.35%) mutation(0.35%) apoptosis(0.33%)
CH ∩ PS	<b>patient</b> (10.93%) treatment(7.42%) human(5.75%) role(5.18%) method(4.61%) time(4.46%) association(2.74%) research(2.69%) <b>population</b> (2.50%) risk(2.32%) form(2.14%) id(1.96%) <b>behavior</b> (1.75%) est(1.73%) who(1.69%) report(1.68%) table(1.65%) affect(1.63%) measure(1.58%) ability(1.26%) family(1.08%) cancer(1.01%) disease(0.91%) scale(0.91%) will(0.87%) cell(0.79%) relative(0.74%) adult(0.67%) life(0.67%) culture(0.65%)
CS ∩ MA	<b>method</b> (17.38%) time(10.44%) population(5.41%) measure(4.04%) algorithm(3.47%) will(2.92%) patient(2.49%) probability(2.17%) id(2.08%) set(1.97%) est(1.90%) human(1.89%) scale(1.89%) research(1.80%) community(1.69%) cell(1.56%) paper(1.51%) form(1.40%) standard(1.15%) table(1.13%) risk(1.10%) behavior(1.04%) power(0.91%) solution(0.86%) disease(0.84%) who(0.84%) cost(0.75%) role(0.65%) procedure(0.61%) literature(0.61%)
CS ∩ PH	<b>method</b> (13.21%) time(10.31%) cell(10.03%) protein(5.75%) human(2.85%) will(2.73%) <b>algorithm</b> (2.67%) form(2.40%) id(2.39%) est(2.10%) solution(1.83%) population(1.81%) set(1.66%) measure(1.63%) table(1.62%) role(1.51%) concentration(1.40%) behavior(1.19%) community(1.13%) temperature(1.00%) patient(0.99%) scale(0.91%) probability(0.81%) research(0.80%) standard(0.79%) <b>map</b> (0.67%) growth(0.63%) gene(0.62%) <b>noise</b> (0.59%) tissue(0.58%)
CS ∩ PS	research(8.31%) time(7.87%) <b>method</b> (6.85%) <b>measure</b> (6.70%) <b>scale</b> (4.17%) population(3.95%) patient(3.47%) behavior(3.38%) id(2.70%) est(2.53%) who(2.26%) will(2.19%) child(2.09%) community(1.99%) human(1.81%) risk(1.64%) form(1.07%) algorithm(1.00%) treatment(0.93%) set(0.93%) association(0.92%) role(0.87%) literature(0.85%) need(0.84%) paper(0.81%) standard(0.80%) woman(0.78%) face(0.77%) hand(0.76%) researcher(0.72%)
MA ∩ PH	cell(20.59%) method(9.81%) time(8.75%) protein(4.28%) patient(2.83%) population(2.59%) form(2.14%) will(2.06%) gene(2.04%) human(2.04%) treatment(1.78%) table(1.62%) id(1.57%) role(1.54%) est(1.42%) measure(1.20%) disease(1.17%) solution(1.16%) concentration(1.05%) scale(0.94%) algorithm(0.88%) probability(0.85%) risk(0.80%) behavior(0.79%) set(0.79%) tissue(0.76%) strain(0.69%) standard(0.64%) tumor(0.63%) mutation(0.61%)
MA ∩ PS	patient(10.21%) time(6.90%) population(5.85%) method(5.25%) measure(5.12%) risk(4.97%) scale(4.48%) who(3.47%) research(3.44%) treatment(2.99%) child(2.40%) behavior(2.32%) id(1.83%) est(1.76%) will(1.70%) health(1.41%) human(1.34%) association(1.21%) community(1.04%) form(0.99%) table(0.92%) woman(0.86%) survey(0.79%) prevalence(0.79%) disease(0.78%) table(0.71%) standard(0.67%) literature(0.61%) affect(0.60%) need(0.60%)
PH ∩ PS	time(10.12%) patient(6.07%) method(5.94%) treatment(5.08%) behavior(3.93%) scale(3.20%) role(3.17%) id(3.14%) measure(3.06%) human(3.00%) population(2.90%) est(2.90%) form(2.53%) will(2.37%) research(2.26%) risk(1.82%) affect(1.73%) association(1.62%) table(1.50%) report(1.18%) ability(1.18%) who(1.11%) community(1.04%) cell(0.86%) lead(0.72%) <b>need</b> (0.71%) <b>environment</b> (0.71%) family(0.69%) standard(0.69%) disease(0.66%)

注: BI代表生物学, CH代表化学, CS代表计算机科学, MA代表数学, PH代表物理学, PS代表心理学。

(责任编辑 车尧)