

可视化知识图谱之研究

——以台湾智能运输系统学术领域为例

陶冶中[▲] 简睿志 刘佑辰

(淡江大学运输管理学系 台湾 新北 25137)

摘要 随着科技日新月异,除全球智慧运输(intelligent transportation system, ITS)的发展趋势成蓬勃发展外,近年来对于大量数据下所隐藏的信息亦是越趋重视,因此,对于何种数据透过何种方式的呈现与其经挖掘后而呈现之意义,则为一大研究重点。对于 ITS 系统下九大领域间,透过知识图谱之方式,建构各领域之关联,以文献与作者上的聚焦探讨其彼此之发展历程、现况、趋势,再者深入于共同作者、关键词、参考文献、技术层面,挖掘各领域于主题上之演变、热门研究议题、前瞻研究领域,主要目的为于 ITS 系统各领域上进行整体的趋势观测与挖掘其前瞻研究及其核心价值所在,然后透过各院校的产出,可发现台湾于 20 世纪后 ITS 论文产出量有明显的上升趋势,代表着 ITS 的研究已逐渐朝一成熟稳健阶段迈进,综观台湾各大专院校于 ITS 所投入之研究产出量,以台湾交通大学 233 篇之产出为首,而各大专院校于 ITS 所投入之研究领域,则以先进交通管理服务(advanced traffic management services, ATMS)为主,对此,透过台湾投入经费于 ITS 研究之年期是可相互呼应的,而如何进一步将台湾各大专院校所投入之 ITS 研究进行其演进历程、分布特性以及群聚知识群等样貌进行挖掘,对于洞察及发现未来发展面向是一重要的关键所在。

关键词 可视化;智慧运输系统;知识图谱;先进交通管理服务

中图分类号:G203 **文献标志码**:A **doi**:10.3963/j.issn.1674-4861.2015.01.022

A Case Study of the Visualization of Intelligent Transportation System of Academic Sector in Taiwan

TAO ZhiZhong JIAN Ruizhi LIU Youchen

(Department of Transportation Management, Tamkang University, New Taipei City 25137, Taiwan, China)

Abstract: With the rapid progress of technology, not only the progress of Intelligent Transportation System is full of vitality, but the hidden information of big data is getting important. Therefore, the meaning of what data and how they are presented become hot issues. For the Nine Areas in ITS, mapping knowledge domain through construction of the connection in each area, through the literature and the authors' discussion focused to investigate the development process, status and trend, and more deeper in the co-authors, key words, reference and the technology level, for mining the change in each area, hot study topic, prospective research areas, the main goal is to observe the tendency in ITS and mining prospective research areas and find core values. It is found out that most of the ITS knowledge outputs focus on ATMS and Chiao Tung university performs the best in Taiwan.

Key words: visualization; intelligent transportation systems; mapping knowledge domain; ATMS

0 引言

当 1 个国家或都市的运输系统硬件建设达到某种程度后,亦或是政府缺乏资金提供庞大的运

输建设时,为继续提高运输容量,就必须考虑如何利用电子通信信息等高科技与管理技术将既有运输设施加值、使之更具智慧,以提供更有效率与安全的服务,也就是智能运输系统(intelligent

收稿日期:2014-05-30 修回日期:2014-11-30

▲第一作者(通讯作者)简介:陶冶中(1961~),博士,副教授,研究方向:智慧运输系统,轨道运输、永续运输。E-mail:cctao@mail.tku.edu.tw

transportation system, ITS), 因此整体来说, 运输系统可说是配合国家经济发展与运输需求成长之指标; 若从学术面论之, 其学术研究所产出之文献便可视为一具体之知识结晶, 若汇整 ITS 体系下各领域之研究主题, 即可判断各领域学术面之发展现况。

随着科技日新月异, 如今我们已身处于海量数据(big data)的时代, 过去学术上之纸本文献亦已迈入全面电子与数字化时代, 如此充足便利的电子文献资源不仅大幅度降低知识扩展的空间障碍, 同时也造就了电子化数据库的形成, 因此, 如何在大量的数据下将其信息含意有效的挖掘出来俨然已成为一重要课题。

鉴于学科分类之概念, 本研究欲透过知识图谱之概念, 针对台湾 ITS 学术资源进行挖掘与分析, 最后透过绘制之方式将知识可视化, 直观的了解数据之特性。

1 知识图谱定义与内涵

知识图谱(mapping knowledge domain), 1个呈现学科、领域、专业知识和个别文章或作者等实体与其他相关的实体之邻近和相关位置的空间图, 藉由知识地图的绘制可以分析1个领域的结构、动态、社会网络和发展^[1], 主要为, 用可视化技术描述知识资源及其载体, 挖掘、分析、构建、绘制和显示知识及它们之间的相互联系, 具体来说便是将复杂的知识领域通过数据挖掘、讯息处理、知识计量和图形绘制而显示出来, 揭示知识领域的动态发展规律, 为学科研究提供切实的、有价值的参考。

而如何有效的揭露科学知识的发展至今仍是此领域持续探讨之方向, 而要描述科学知识的发展动态情形, 引文分析方法为一重要的量化方式, 可大量挖掘数据下的潜藏模式(pattern), 引文与被引文之间往往有学科内容上之关联, 透过引文群聚分析, 能了解学科间之相互关系与知识结构, 区分学科的作者族群, 分析推估学科研究间的交叉特性与发展趋势。引文分析之相关介绍如下^[2-3], 其关系见图1。

1) 引文分析法。又名引用文献分析法(citation analysis), 最早由美国学者 Kessler 基于文献引用的方法, 用于学科的结构与学科间的关联与交流之定量分析, 透过文献引用的频率分析研究可预测某一学科的影响力与重要性。

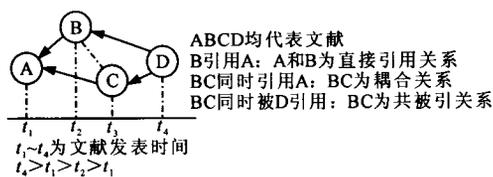


图1 引文分析形式之关联示意图

Fig. 1 Association graph of citation analysis

2) 引文耦合法。以引文分析为基础之发展, 系研究学科结构和关联之有效方法, 将共同引用相同论文的2篇或多篇文献称为耦合文献, 他们之间关系称之为引文耦合, 亦定义出耦合强度以定量地表示耦合文献间的关联, 耦合强度值即耦合文献共被引文献的篇数。

3) 共被引分析。共被引分析(co-citation)概念为 Small(1973)所提出, 指当2篇文献同时被第3篇文献所引用, 前2篇文献间的关联称之为共被引关系, 可反映出被引用文献的频率与科学知识动态变动之本质。

Derek J. deSolla Price 亦指出科学文献的引文网络中, 越频繁被引用的文章越有可能是近期所发表的文章, 而前瞻领域的研究则约由近期发表的40至50篇相关文章所构成。

知识图谱研究大致分为以下3阶段。

1) 基于共被引分析之科学知识图谱。文献共被引(document co-citation analy)为最基本的共被引关系, 展现共同被引用的文献间之结构关系进而反映学科间之关联, 透过共被引群聚分析可研究科学文献之分布现况与特征, 而文献共被引的群聚网络之变化亦能了解学科间相互关系与发展趋势。

2) 社会网络分析与科学知识图谱。社会网络分析(social network analysis)为一测量研究对象间关系之社会学研究方法, 网络中以节点代表人或组织, 连接两节点之线则代表研究对象间之关系, 主要透过可视化与数学之分析进行重要作品、作者、学科影响力和群聚分布等分析。

3) 信息视觉与科学前瞻研究知识图谱。因应网络普及、信息可视化之崛起与科学计量分析的应用, 信息可视化与科学前瞻研究知识图谱成为知识图谱兴起的第3种方式, 为科学计量学演进成知识可视化之方法。

2 知识图谱之建构

根据 Borner 等^[4], 指出知识可视化之程序依序如下。

1) 数据选取。如 ISI、Inspec 等电子数据库。

2) 分析单位。欲分析之单位,包括期刊、书籍、作者、关键词等。

3) 分析方法。分析之方法包括计量分析,如关键词分析、作者引用分析、共被引分析等,或是门坎指标的计算如量化计数。

4) 设计工具。依其相似性或是分类加以设计。相似性如数量(直接引用、共被引、共同关键词数量,以矩阵为单位)、向量(向量空间模型、潜在语意分析,以属性矩阵为单位)、相关性(Pearson R 系数);分类上如降级维度(特征向量、因素分析、主成分分析、多维尺度分析等)、群聚分析、数量(三角测量、关联定位)。

5) 结果产出。在图谱的建构软件上可看到,国外常用的讯息可视化软件主要为 Sci2 Tool, InSPIRE, SciMAT, Histcite, Pajek, Citespace, UCINET, Bibexcel, Gephi, VOSviewer, Vantage-Point, Network Workbench Too^[5], 而国内主要使用之软件为 Citespace, Histcite 和 Pajek^[6], 其中本研究预计采用 Sci2 Tool 之工具进行可视化之建构,主要原因如下。

(1) 借镜于国外研究。美国国家科学基金会(national science foundation, NSF)、美国国立卫生研究院(the national institutes of health, NIH)、美国农业部(the US department of agriculture, USDA), 以及美国国家海洋和大气管理局(the national oceanic and atmospheric administration, NOAA)等均使用 Sci2 来进行可视化分析之研究。

(2) 可使用之数据格式较多元:包括 xml、.net、.isi、.csv、.bib、.enw、.nsf。

(3) 可抽取多种数据网络:包括定向网络、双边网络、引文网络、作者文献网络、共现网络、词共现网络、合作者网络、引文耦合网络, Sci2 在网络构建方面的功能甚至比 CiteSpace 还要更强^[9]。

3 台湾 ITS 学术产出现况

台湾交通部运输研究所于 2002 年拟订了台湾 ITS 发展的 9 大服务领域,分别为:先进交通管理服务(advanced traffic management services, ATMS)、先进用路人信息服务(advanced traveler information services, ATIS)、先进公共运输服务(advanced public transportation services, APTS)、商用车辆营运服务(commercial vehicle

operations services, CVOS)、电子收付费服务(electronic payment services, EPS)、紧急救援管理服务(emergency management services, EMS)、先进车辆控制及安全服务(advanced vehicle control and safety services, AVCSS)、弱势使用者保护服务(vulnerable individual protection services, VIPS)、信息管理服务(information management services, IMS)。

台湾院校之 ITS 学术论文产出情形,可以观察到学术论文主要投入之院校、院校学术论文产出趋势、ITS 9 大领域下之产出比例、各领域论文发展趋势以及历年产出趋势与平均产量,进而推断学术面之发展。

台湾各院校 ITS 论文产出现况如图 2,可看出各院校对于 ITS 产业之研究与投入,以台湾交通大学为首达 233 篇,其次为成功大学 153 篇,值得注意的是交通大学之硕博士论文因尚未产出 2013 年版本,因此数据为 2012 年之数据,但产出量依然惊人。

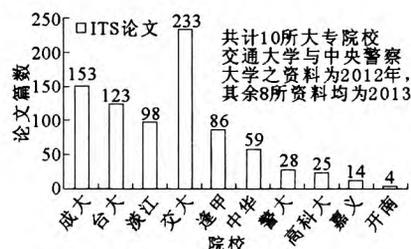


图 2 台湾各院校 ITS 硕博士论文产量

Fig. 2 The quantity of thesis of each school in Taiwan

而对于台湾 ITS 9 大领域之产出图 3,可发现台湾产出之 ITS 领域以 ATMS 为主流之研究方向,而学术研究在理论验证与实证分析时所需之交通参数数据皆须藉由健全的 ATMS 基础设施系统加以搜集,经数据搜集与处理后方可供其他领域相关数据以予更多信息服务于用路人,由此可见 ATMS 为 ITS 系统之核心研究领域,相关知识产出有利于其他领域之发展,故论文产出数量领先其他领域,符合我国当今现况;其次依序为 APTS 与 CVOS。

图 4 之研究领域与院校硕博士论文产出交叉分析有益于了解各领域分布于院校之情形。

从图 5 与图 6 ITS 论文每年产出趋势中发现,台湾 ITS 领域论文自 2000 年后便逐步呈现显著的成长趋势,台湾于 2000 年以后亦开始计划性的投入经费于 ITS 各领域,显现出台湾 ITS 知识产业已逐步迈入成熟稳健阶段。

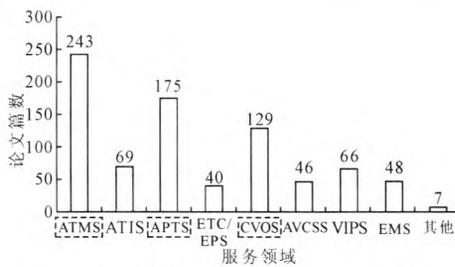


图3 台湾 ITS 领域硕博学位论文产出趋势
Fig.3 The quantity of thesis of each category in Taiwan's ITS

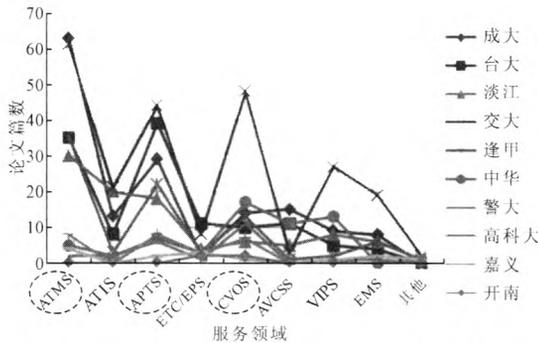


图4 ITS 研究领域与院校论文产出之交叉分析
Fig.4 The cross-analysis between ITS categories and the quantity of thesis in each school

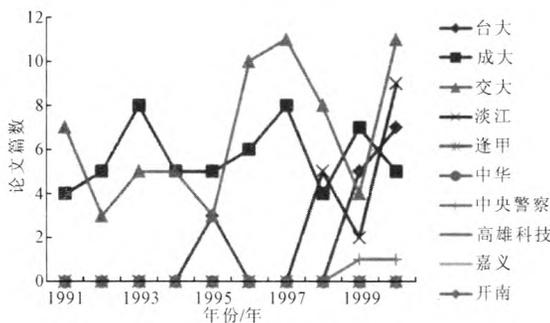


图5 ITS 草创期之硕博学位论文产出趋势
Fig.5 The trend of output of thesis in the starting period

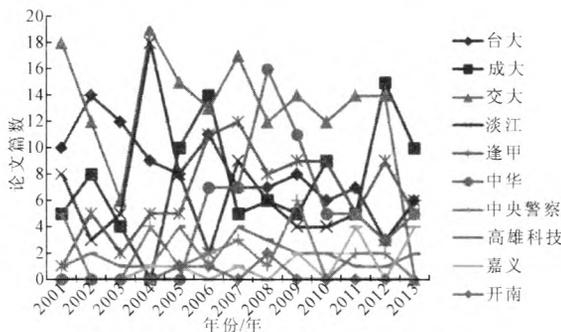


图6 ITS 奠基期之硕博学位论文产出趋势
Fig.6 The trend of output of thesis in the building period

4 主要研究方法

本研究为一实证研究,在对台湾运输相关学系 ITS 论文产出进行深入探讨,透过研究主题、

特征技术与数量分析说明各院校 ITS 领域之重要学者与其研究方向外,亦透过外国之电子数据库进行实验,利用学术群聚之状况挖掘当今前瞻研究领域与学科上之研究方向的转移,进而观察其前瞻研究之趋势是否能从台湾院校之学术论文推论之。

研究方法与内容以引文分析为基本方法,欲透过可视化软件 Sci2 进行学科知识之绘制,绘制之重点包括指导教授、关键词、参考文献、技术层面等进行图谱之绘制(群聚中心强度与各关联度),同时针对院校下同一指导教授所带领之学生的研究议题,进行分析与挖掘,试图找出硕博学位论文的产出与前瞻之研究领域之关系是否存在。

5 结束语

本研究为了解台湾 ITS 学术论文之研究趋势,欲透过知识图谱之方式呈现出其关联度,但因缺乏完整的引文数据库供知识管理使用,仍需藉由人工分类的方式进行数据之初步筛选,而若能从论文之产出内容,诸如研究方向、议题、特征技术等与当前之前瞻研究区块是能接轨的,便能确立台湾建立引文数据库之重要性。

参考文献

- [1] Reid E F, Chen Husinchun. Mapping the contemporary terrorism research domain[J]. International Journal of Human-Computer Studies, 2007,65(1):42-65.
- [2] Kessler M M. Bibliographic coupling between scientific papers[J]. American Documentation, 1963, 14 (1):10-25.
- [3] Kessler M M. Comparison of the results of bibliographic coupling and analytic subject indexing[J]. American Documentation, 1965, 16(3):223-233.
- [4] Borner K, Chen C, Boyack K W. Visualizing knowledge domains[J]. Annual Review of Information Science & Technology, 2003,37:179-255.
- [5] Cobo M J, A G López - Herrera, E Herrera - Viedma. Science mapping software tools: Review, analysis and cooperative study among tools [J]. Journal of the American Society for Information Science and Technology, 2011,62(7):1382-1402.
- [6] 邱小花,李国俊,等. 一款新的知识图谱分析软件介绍与评价[J]. 图书馆杂志, 2013,32(9):79-87. Qiu Xiaohua, Li Guojun, Xiao Ming. The introduction and evaluation of a new analysis software of knowledge mapping[J]. Library Journal, 2013, 32 (9): 79-87. (in Chinese)