

· 学科与课程建设 ·

“双一流”建设背景下我国医药类高校 ESI 潜力学科识别及预测方法实践

王雅棋¹, 丁佐奇^{1,2*}, 郝海平³¹中国药科大学理学院; ²《中国天然药物》编辑部; ³中国药科大学药学院 (南京 210009)

摘要 在我国全面开展“双一流”建设的背景下,充分利用 ESI 和 InCites 数据库对我国高校学科进行定量评估和发展预测,相关结果对推动学科建设具有重要指导意义。对我国 28 所医药类院校进行 ESI 潜力学科的识别,通过 InCites 数据库模拟检索 ESI 总被引频次和筛选最优拟合回归曲线,对潜力学科入围 ESI 排名的时间进行预测。28 所高校中 17 所高校共被识别出 24 个潜力学科,其中 23 个潜力学科预计在 2027 年 3 月之前陆续进入 ESI 排名。研究得到的潜力学科预测入围时间以及预测方法,可为各高校把握学科国际地位以及进行学科规划方案调整提供参考,有利于促进我国高校更多学科更快地入围 ESI 排名,以推动我国“双一流”建设的进程。

关键词 学科建设;潜力学科;ESI;InCites;拟合回归;时间预测

Practice of Identification and Prediction of ESI-covered Potential Disciplines in Chinese Medical Universities under the Background of “Double First-class”

Wang Yaqi¹, Ding Zuoqi^{1,2*}, Hao Haiping³¹School of Sciences; ²Editorial Department of Chinese Journal of Natural Medicines; ³School of Pharmacy, China Pharmaceutical University, Nanjing 210009, China)

Abstract: Under the background of China's “double first-class” construction, the ESI and InCites databases are fully utilized to conduct quantitative evaluation and development prediction of disciplines in universities, relevant results of which have important guiding significance for facilitating discipline construction. This study is aimed at the identification of ESI-covered potential disciplines in 28 medical universities in China, and the prediction of time for potential disciplines to be ESI ranked through simulation retrieval of ESI-covered cites in InCites and screening of optimal fitting regression curve. A total of 24 potential disciplines have been identified in 17 out of the 28 universities, 23 of which are expected to enter the ESI ranking successively before March 2027. The results and methods of identification and prediction of potential disciplines in this study can provide reference for universities to grasp the international status of disciplines and adjust the discipline planning scheme, which is conducive to the acceleration of more disciplines in China's universities being listed in ESI ranking, so as to promote China's “double first-class” construction.

Key words: discipline construction; potential disciplines; ESI; InCites; fitting regression; prediction of time

自 2015 年 11 月 5 日国务院印发《统筹推进世界一流大学和一流学科建设总体方案》(简称《总体方案》)以来,一流大学和一流学科建设成为我国当前的政策热点。同时,“双一流”开始取代“985 工

收稿日期: 2019-11-26 *通信作者: 丁佐奇, E-mail: zqding1028@163.com

基金项目: 2019 年江苏省研究生教育教学改革重大课题, 编号 JGZD19_010; 中国药科大学“双一流”建设研究生教育教学改革研究项目, 编号 JGYB201904

程”“211 工程”高水平大学建设的“单一流”,学科的重要性进一步得到凸显^[1]。学科不仅仅是一流大学建设的基础,更是一流大学其特色和优势的呈现。因此近年来,针对国内高校学科建设的评估备受重视,教育部官方最新的第四轮学科评估结果于 2017 年 12 月公布后受到热议,与 2017 年 9 月公布的首轮“双一流”学科名单存在一些出入。

目前,较为普遍的学科评估方法为利用权威的引文数据库对学科的科研产出进行评价。基本科学指标数据库(Essential Science Indicators,简称 ESI)已成为国际上权威的衡量科学研究绩效、跟踪科学发展趋势的基本分析评价工具,在国内外被广泛用于科学计量和学科评价。在我国“双一流”建设背景下,越来越多高校开始借助 ESI 数据库进行学科发展趋势的追踪和评价,ESI 数据库无疑已成为我国促进“双一流”建设的权威工具。ESI 仅统计进入全球排名前 1% 的学科数据,入围 ESI 学科排名的高校学科标志着其具有一定的国际影响力,因此,及时追踪 ESI 数据库学科排名情况可了解学科在全球所有高校和机构中所处的位置,各高校在确保已入围学科稳步增长的同时,发掘有望入围的潜力学科,把握入围 ESI 的差距,从而合理制定学科规划方案,促进各潜力学科发展。由于 ESI 数据库根据各高校和机构的学科在一定时间内所发表论文的总被引频次这一指标进行排名,所以利用各种统计方法对潜力学科入围 ESI 排名进行预测逐渐成为研究热点。笔者通过知网以“ESI 潜力学科”作为关键词进行文献检索,发现关于我国高校 ESI 潜力学科的分析研究最早可追溯至 2014 年 6 月发表的一篇文章,甚至早于《总体方案》的发布,即夏琬钧等学者通过 InCites 和 ESI 对西南交通大学自身优势学科及有发展潜力学科等方面进行统计分析^[2],这在一个方面说明 ESI 数据库对推动我国高校学科发展的作用早已存在。尤其在 2015 年“双一流”学科建设在我国拉开序幕后,这一研究课题受到越来越多学者的关注,各种各样的预测和统计方法层出不穷。

本研究选取在首轮“双一流”名单中入选一流医药类学科的高校,共 21 所(北京大学、复旦大学、上海交通大学、浙江大学、郑州大学、武汉大学、华中科技大学、中山大学、暨南大学、四川大学、北京协和医学院、北京中医药大学、天津医科大学、天津中医药大学、上海中医药大学、南京中医药大学、中国药科大学、广州中医药大学、成都中医药大学、海军军医大学和空军军医大学),另外加入中国历史最悠

久的药科院校沈阳药科大学,以及 6 所江苏省有名的综合类和医药类院校(南京大学、东南大学、南京医科大学、苏州大学、徐州医科大学和南通大学),以上述 28 所院校作为研究对象,笔者利用 ESI 和 Incites 数据库通过高校未入围学科的总被引频次与 ESI 学科总被引频次阈值的比值筛选出各高校在未入围 ESI 学科中的潜力学科,并通过拟合最佳回归模型的方法对潜力学科入围 ESI 排名的时间进行预测,并以中国药科大学为例进行具体的实例解析,以期各高校把握学科国际地位以及进行学科规划方案调整提供参考。

1 潜力学科的认识

1.1 潜力值 Q 的计算

ESI 数据库中仅展示全球前 1% 研究机构的学科情况,尚未进入排名的学科的 ESI 总被引频次则是不可见的,有学者将其定义为“暗数据”范畴^[3]。通常人们认为未入围 ESI 排名的学科的潜力值即为该“暗数据”与 ESI 学科入围阈值的比值,因此首要问题是该类“暗数据”的可视化,前人研究中的解决方案大致分为两类。由于 ESI 仅对 SCI-E (Science Citation Index Expanded) 和 SSCI (Social Sciences Citation Index) 中 article 和 review 类型的论文进行统计,且被 SCI-E、SSCI 收录的每种期刊对应一个 ESI 学科,管翠中等^[4]在 Web of Science (WoS) 核心合集数据库中,按照 ESI 当前数据年限设定时间间隔、论文类型、扩展作者机构以及文献来源期刊,模拟检索学科的总被引频次;汤莹^[5]则利用 Incites 数据库根据 ESI 学科、ESI 年限和论文类型设限模拟检索各机构的学科总被引频次。以上两种模拟检索方法不可避免存在的误差在于数据库与 ESI 数据库统计来源和更新频率不完全相同,针对这一问题许多学者尝试进行精准校对,王雪^[6]以 WoS 作为模拟 ESI 统计的数据源时引入被引转换比值,并提出时间节点错位比较,减弱两个平台的统计差异;程建萍等^[7]针对 22 个 ESI 学科在 ESI 数据库和 InCites 数据库的数据偏差进行统计分析,计算出 22 个学科误差修正因子。

笔者认为,识别潜力学科的意义是从各机构自身出发,找到机构中 5 年或 10 年内有望入围 ESI 排名的学科,从而调整机构为期 5 年或 10 年的学科规划方案,以最精准的资源投入获取最大的边际效益。再退一步,除了潜力值,潜力学科的认识还取决于主观决定的潜力阈值,本身就存在一定误差,没必要对

模拟检索的结果进行精确校准。因此,本研究选取模拟检索较简单直接的 Incites 数据库,将得到的学科模拟 ESI 总被引频次记为 C_1 ,ESI 学科排名阈值记为 C_0 ,则未入围学科的潜力值 Q 为: $Q = C_1/C_0$ 。当潜力值 $Q \geq 0.8$ 时,该学科则被认为短期内有望入围 ESI 学科排名,被识别为该机构的潜力学科。

1.2 数据来源与获取

ESI 数据库对全球所有高校及科研机构近 11 年的论文数据进行统计,每 2 个月更新一次统计数据,而 Incites 数据库每月更新一次,统计截至上一个月末的所有数据。笔者于 2019 年 11 月 ESI 数据更新后记录了 22 个 ESI 学科的入围阈值 C_0 ,即排名最后一位的总被引频次,以及 28 所高校的未入围学科名单,当前 ESI 学科排名统计的论文数据的时间跨度为 2009 年 1 月 1 日至 2019 年 8 月 31 日。同时,笔者在 Incites 数据库中对 28 所高校的未入围学科一一进行模拟检索,具体步骤为:登录 Incites 数据库,进入分析模块,选择研究数据类型为“机构”,将检索时间段设为 2009 年至 2019 年,文献类型设为“article”和“review”,研究领域设为相应的 ESI 学科,记录检索结果中各高校每个未入围学科的总被引频次。

1.3 28 所高校潜力值与潜力学科

根据潜力值的计算方法,得到 28 所高校未入围学科的潜力值 Q ,将 $Q \geq 0.8$ 的学科确定为该高校的潜力学科。本研究以中国药科大学为例进行详细的解析,表 1 中可见中国药科大学未入围学科共 18 个,其中空间科学学科无发文,材料科学学科的潜力值为 $1.066 > 0.8$,被识别为潜力学科。需要说明的是,由于 ESI 数据库中论文数来源为 SCI-E 和 SSCI,被引数据来源为 SCI-E、SSCI 和 A&HCI 三个数据库,而 Incites 数据库的论文数和被引数据均来自 SCI-E、SSCI、A&HCI、CPCI-S、CPCI-SS&H、BKCI-S、BKCISSH 七个数据库,这导致学科的 Incites 总被引频次必定略大于 ESI 总被引频次,即可能出现未入围学科 Incites 总被引频次甚至高于 ESI 阈值,潜力值 Q 大于 1 的情况。如表 2 所示,28 所高校中共有 8 个学科的潜力值大于 1,另外明显可见,28 所高校中有 17 所高校共识别出 24 个潜力学科,4 所高校识别出不止 1 个潜力学科,暨南大学和海军军医大学甚至分别识别出 4 个和 3 个潜力学科,说明如果相关院校学科建设规划部署得当,有关学科其综合实力有可能短期内得到较大提高。

表 1 中国药科大学未入围学科潜力值

编号	未入围学科	InCites 总被引频次 C_1	ESI 阈值 C_0	潜力值 Q
1	材料科学	7100	6661	1.066
2	农业科学	1887	2379	0.793
3	神经科学与行为学	3399	6818	0.499
4	植物学与动物学	1121	3002	0.373
5	免疫学	1842	5253	0.351
6	分子生物学与遗传学	4140	14628	0.283
7	环境科学与生态学	1203	4448	0.270
8	工程学	601	2680	0.224
9	计算机科学	648	3715	0.174
10	微生物学	652	5670	0.115
11	综合交叉学科	338	2960	0.114
12	社会科学总论	147	1547	0.095
13	物理学	1059	21391	0.050
14	精神病学与心理学	39	4259	0.009
15	数学	40	4572	0.009
16	经济与商业	32	4486	0.007
17	地球科学	17	6520	0.003
18	空间科学	/	38987	/

注:标“/”符号的表示中国药科大学该学科未有论文被 InCites 收录

表 2 28 所高校潜力学科

院校	潜力学科	InCites 总被引频次 C_1	ESI 阈值 C_0	潜力值 Q
北京大学	空间科学	36997	38987	0.949
复旦大学	/	/	/	/
上海交通大学	地球科学	5337	6520	0.819
浙江大学	/	/	/	/
郑州大学	神经科学学与行为学	6802	6818	0.998
武汉大学	/	/	/	/
华中科技大学	植物学与动物学	3202	3002	1.067
中山大学	经济与商业	5085	4486	1.134
暨南大学	神经科学学与行为学	5743	6818	0.842
	植物学与动物学	3112	3002	1.037
	社会科学总论	1311	1547	0.847
	分子生物学与遗传学	13076	14628	0.894
四川大学	/	/	/	/
北京协和医学院	/	/	/	/
北京中医药大学	/	/	/	/
天津医科大学	免疫学	4935	5253	0.939
	化学	6810	8233	0.827
上海中医药大学	生物学与生物化学	5778	6728	0.859
南京中医药大学	化学	7234	8233	0.879
中国药科大学	材料科学	7100	6661	1.066
广州中医药大学	化学	6606	8233	0.802
海军军医大学	植物学与动物学	3002	3002	1.000
	社会科学总论	1556	1547	1.006
	材料科学	6022	6661	0.904
空军军医大学	免疫学	4737	5253	0.902
成都中医药大学	/	/	/	/
天津中医药大学	/	/	/	/
沈阳药科大学	/	/	/	/
南京大学	/	/	/	/
东南大学	分子生物学与遗传学	13541	14628	0.926
南京医科大学	材料科学	5842	6661	0.877
苏州大学	农业科学	2567	2379	1.079
	环境科学与生态学	4313	4448	0.970
徐州医科大学	药理学与毒理学	3653	3619	1.009
南通大学	/	/	/	/

注:标“/”符号的表示该院校未识别出潜力学科

2 潜力学科入围时间的预测

在我国全面开展“双一流”建设背景下,为了进一步推动高校学科的发展和建设,在对机构进行潜力学科识别的基础上,相关研究人员还可以运用各种统计工具定量地对潜力学科入围 ESI 排名的时间进行预测,以更精确地为机构和高校提供制定学科建设规划和资源分配方案的参考。近年来,国内学者广泛将各种统计方法运用于潜力学科入围时间预测,王雪^[6]基于 ESI 统计数据的滚动周期性,引入时间序列预测模型对两所高校的潜力学科预测进行实证研究;戴莹^[8]和秦萍等^[9]利用 ESI 和 InCites 数据

库,以灰色系统理论 GM(1,1)模型的原理和方法为基础,运用 MATLAB 软件对高校潜力学科的发展趋势作出预测;汤莹^[4]通过回归分析得到总被引频次与总论文数之间的回归方程,进而通过利用时间序列预测方法预测机构未来的发文量及未来的学科总被引频次;管翠中等^[4]运用曲线拟合模型方法对高校潜力学科入围时间进行预测;陈颖瑜等^[3]联用 SWOT 分析法、文献调研法和回归分析法等,通过构建潜力学科入围所需论文篇数的线性预测公式和入围预测的指标数据模型,尝试对 ESI 潜力学科进行趋势预测。

总之,进行入围时间预测的方法主要分为两大

类:回归分析和时间序列预测。每种预测方法都有其适用条件和范围,而预测结果的准确性在很大程度上取决于假设条件的满足程度^[10]。而时间序列和回归分析的主要区别在于对数据的假设,回归分析假设每个数据都是独立的,而时间序列则是利用数据之间的相关性进行预测,即以老的样本量预测新的样本量,是一种自回归^[11]。学科的各期 ESI 总被引频次数据之间相对来说独立性高于相关性,本研究选取拟合回归曲线的方法对 28 所高校的潜力学科进行预测。

2.1 学科线、阈值线的设定与数据获取

在进行潜力学科预测时,本研究将学科线设定为自该潜力学科有论文被 InCites 数据库收录的年份起至 2018 年,每一年的年度拟 ESI 总被引频次 C_{in} ;考虑到由于 ESI 数据库和 Incites 数据库更新时间不一致以及统计来源的不同,为提高定量预测潜力学科入围时间的准确性,本文将阈值线设定为 ESI 阈值机构对应的 InCites 年度拟 ESI 总被引频次,此处的 ESI 阈值机构设定为 2019 年 11 月 ESI 数据库每个学科入围的末 10 名机构,取这些机构的年度拟 ESI 总被引频次的均值作为年度阈值,这不仅可避免选取一个末位机构时的偶然性,还能减轻末 10 名机构个体差异的影响。

数据获取主要分为两步:首先于 2019 年 11 月

笔者记录所有高校潜力学科的 ESI 排名末十位的机构名,同时在 InCites 数据库找出潜力学科首次有论文被 InCites 收录的年份,确定学科线和阈值线的起始年份,由于中国药科大学图书馆购买了 InCites 数据库自 1980 年来的数据,本研究所检索的潜力学科的 C_{in} 年份最早可至 1991 年。第二步在 Incites 数据库中检索每所高校潜力学科的年度拟 ESI 总被引频次 C_{in} ,通过设定起止年份分别为年份的前 11 年和前 1 年(比如 2010 年的 C_{in} 即为该学科在 1999 至 2009 年发表的论文的总被引频次),文献类型为“article”和“review”进行检索,获得所有高校潜力学科的学科线;以同样方式检索所有潜力学科末十位机构的 C_{in} 并取平均值,获得所有潜力学科的阈值线。

以中国药科大学为例,表 3 中可见其潜力学科材料科学 2019 年 11 月 ESI 排名的末十位机构,由于倒数第 10 名有两位机构并列,共为 11 所机构。中国药科大学材料科学学科在 1980 年至 1990 年期间就有论文被 InCites 数据库收录,因此中国药科大学潜力学科材料科学的学科线和阈值线的时间跨度为 1991 年至 2018 年,如图 1 所示,其为该 11 所机构从 1991 年至 2018 年的 Incites 数据库年度拟 ESI 总被引频次。从图中可见,11 所机构的总被引频次随时间的变化情况各不相同,因此笔者取每年度 11 所机构的平均值作为阈值线上的数据较为合理。

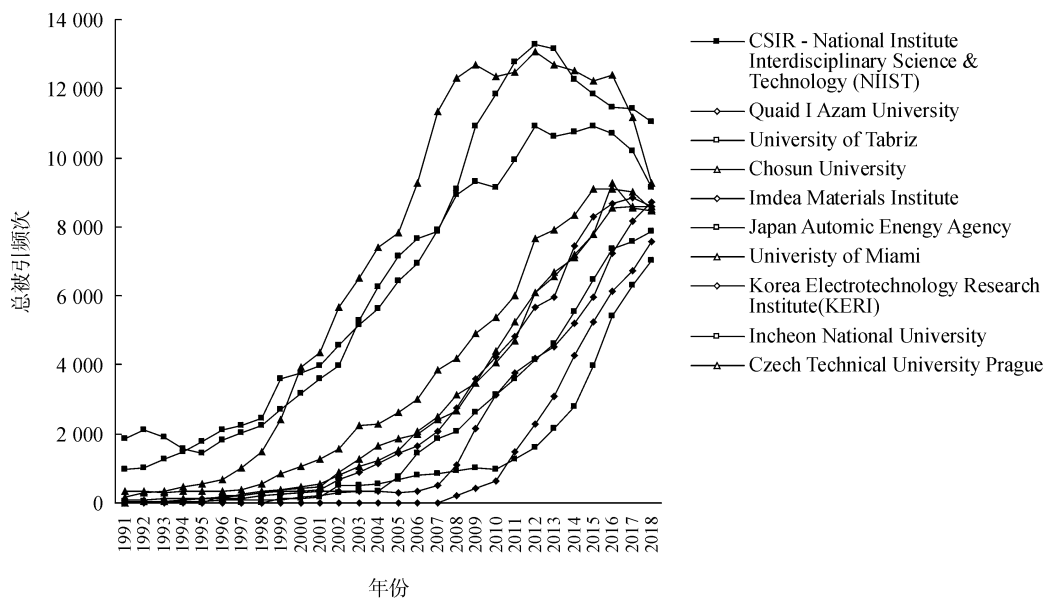


图 1 材料科学 ESI 排名末十位机构的总被引频次

此外,笔者发现中国药科大学的材料科学学科年度拟 ESI 总被引频次从 1991 年至 1999 年均为 10,而 2000 年为缺失值,自 2001 年以后从 153 开始逐渐增长。从 InCites 数据库统计的学科发文量来看,这是

由于该学科在 1980 年至 1990 年期间被 InCites 收录的论文仅一篇,且被引量维持在 10 次,而 1989 年至 1999 年间该学科无论文被 InCites 收录,2000 年重新发表一篇论文被 InCites 收录,自此开始该学科发文

量和被引量稳步增长。为了更好地拟合回归曲线,反映该学科的总被引频次增长情况,笔者将 2001 年之前的数据除去,使用 2001 年至 2018 年的材料科学学科线和阈值线的数据进行回归曲线拟合,如表 3 所示。同样的方法,本研究获取了其他 16 所高校潜力学科的学科线与阈值线。

表 3 中国药科大学材料科学学科的学科线与阈值线

年份	学科线	阈值线
2001	153	1525.300
2002	153	1929.000
2003	175	2364.400
2004	175	2676.800
2005	202	3052.000
2006	202	3513.800
2007	251	4114.100
2008	251	4308.909
2009	482	4955.000
2010	560	5386.182
2011	711	5998.636
2012	743	6816.818
2013	2208	7092.273
2014	3201	7585.727
2015	4234	8144.000
2016	5103	8746.273
2017	5669	8778.909
2018	6785	8613.909

2.2 28 所高校潜力学科入围 ESI 时间的预测

在获取所有高校潜力学科的学科线与阈值线后,将其作为拟合回归曲线的因变量,本文把自变量设

为自获取该学科年度拟 ESI 总被引频次 C_{in} 的年数。例如,1980 年至 1990 年内某潜力学科有论文被 InCites 收录,则可获得潜力学科 1991 年的 C_{in} ,这时 1991 年的 C_{in} 对应自变量为 1,1992 年的 C_{in} 对应自变量为 2,以此类推。本研究运用 SPSS 24 软件通过分析模块的回归曲线估算功能,可对潜力学科的学科线和阈值线进行拟合回归,可供选择的回归模型有 11 种:线性模型、二次模型、三次模型、复合模型、增长模型、对数模型、S 型曲线模型、指数模型、倒数模型、幂函数模型和 Logistic 模型,笔者根据给出的模型拟合优度指标决定系数调整后的 R^2 选择最适合的拟合模型, R^2 的值越接近于 1,说明模型曲线的拟合程度越高。这里需要说明的是,在比较 R^2 之前,需要筛除掉预测值有递减趋势的模型,因为当学科线或阈值线呈递减趋势时,潜力学科入围时间的预测是没有意义的,本研究的基本假设为在可预见的短期内,潜力学科的总被引频次和 ESI 学科的阈值都是逐渐增长的。

以中国药科大学为例,表 4 列出了其潜力学科材料科学学科线与阈值线回归拟合结果中拟合优度决定系数 $R^2 > 0.5$ 的模型以及回归拟合方程,学科线模型中 R^2 最大的为三次模型,即选取三次模型对材料科学学科线进行预测;阈值线模型中 R^2 最大的仍是三次模型,但该模型在预测过程中有递减趋势被筛除,因此本研究选取 R^2 第二大的线性模型进行预测。如此,对其他所有高校的潜力学科的学科线与阈值线进行最优拟合回归模型筛选。

表 4 中国药科大学材料科学学科线与阈值线的拟合回归模型筛选

中国药科大学	回归模型	回归拟合方程	调整后的 R^2
学科线	线性	$y = -1686.464 + 360.318x$	0.725
	二次模型	$y = 1076.007 - 468.424x + 43.618x^2$	0.969
	三次模型	$y = 363.134 - 70.894x - 7.302x^2 + 1.787x^3$	0.976
	幂函数	$y = 36.595 * x^{1.458}$	0.659
	指数函数	$y = 61.448 * e^{0.256x}$	0.922
阈值线	线性	$y = 465.661 * x + 887.446$	0.989
	对数模型	$y = -499.684 + 2873.885 * \log(x)$	0.841
	二次模型	$y = 873.835 + 469.744x - 0.215x^2$	0.988
	三次模型	$y = 1571.428 + 80.736x + 49.613x^2 - 1.748x^3$	0.995
	幂模型	$y = 1186.462 * x^{0.678}$	0.962
	S 型曲线	$y = e^{8.832 - 1.972/x}$	0.661
	指数模型	$y = 1784.984e^{0.101x}$	0.951

如图 2 所示,为中国药科大学材料科学学科线与阈值线的最优拟合曲线,明显可见,两条曲线相交点,即为材料科学学科入围 ESI 排名的时间点。从图中可见,两条回归线相交于 2019 年与 2020

年之间且极接近 2020 年的位置,这说明中国药科大学材料科学学科有望于 2020 年 3 月 (ESI 数据库每年 3 月份刚好统计前 11 年的论文数据) 入围 ESI 排名。

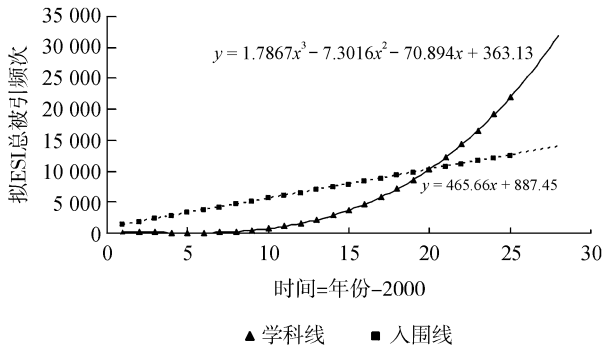


图 2 中国药科大学材料科学学科入围 ESI 时间预测

笔者用同样的方法对其他所有高校潜力学科的入围时间进行预测,结果见表 5,预计除了广州中医药大学化学学科(2034 年 3 月至 2035 年 3 月才可能入围),其他所有潜力学科在 2027 年 3 月之前都能够入围 ESI 学科排名。由于本研究用于时间预测的

建模数据截止于 2018 年,广州中医药大学的化学学科 2018 年的拟 ESI 总被引频次 C_{in} 为 3455,远低于阈值线 2018 年数据 11681.833,实际上该学科被识别为潜力学科是因为 2009 年至 2019 年 10 月 31 日的总被引频次为 6606,比上一年增长近一倍,达到了比较高的潜力值。因此,广州中医药大学化学学科 ESI 总被引频次在未来的增长速率极可能高于本研究中预测的增长速率,即在 2034 年之前就可能入围 ESI 排名。

由表 5 可见,各潜力学科的学科线的最优拟合回归模型大部分(20/24)为三次模型,这说明近年来潜力学科总被引频次的增长速率越来越快。在这样的基础上,各高校若能够更进一步调整规划方案,对潜力学科加以重视和投入,在七年之内,这些潜力学科都可能成功入围 ESI 排名,成为全球排名前 1% 的学科。

表 5 17 所高校的潜力学科入围 ESI 时间预测结果

院校	潜力学科	最优拟合回归模型		预计入围 ESI 时间
		学科线	阈值线	
北京大学	空间科学	三次: $y = 6.0465x^3 - 191.08x^2 + 2225.5x - 3790.6$	二次: $y = 19.448x^2 + 1414.1x + 4584.6$	2022.03 ~ 2023.03
上海交通大学	地球科学	二次: $y = 16.231x^2 - 88.462x + 73.214$	幂: $y = 2231x^{0.4847}$	2025.03 ~ 2026.03
郑州大学	神经系统学与行为学	二次: $y = 17.311x^2 - 237.32x + 653.08$	线性: $y = 401.77x + 145.08$	2026.03 ~ 2027.03
华中科技大学	植物学与动物学	三次: $y = 0.4849x^3 - 3.4294x^2 - 18.588x + 119.55$	幂: $y = 1382.9x^{0.4338}$	2020.03 ~ 2021.03
中山大学	经济与商业	三次: $y = 0.6496x^3 - 13.537x^2 + 72.842x - 75.173$	线性: $y = 257.02x + 1039.1$	2021.03 ~ 2022.03
暨南大学	神经系统学与行为学	三次: $y = 1.266x^3 - 38.519x^2 + 320.74x - 442.72$	线性: $y = 401.77x + 145.08$	2022.03 ~ 2023.03
	植物学与动物学	二次: $y = 7.1739x^2 - 50.675x + 110.1$	幂: $y = 4.4886x^{2.0127}$	2021.03 ~ 2022.03
	社会科学总论	三次: $y = 0.2511x^3 - 7.9173x^2 + 72.597x - 133.3$	二次: $y = 2.4004x^2 + 7.4897x + 270.82$	2024.03 ~ 2025.03
天津医科大学	免疫学	三次: $y = 4.9382x^3 - 89.529x^2 + 494.65x - 618.78$	线性: $y = 803.4x + 4920.7$	2020.03 ~ 2021.03
		三次: $y = 0.7028x^3 - 15.213x^2 + 78.371x - 15.176$	三次: $y = 0.2582x^3 - 23.143x^2 + 717.99x - 1050$	2019.03 ~ 2020.03
	化学	三次: $y = 0.5184x^3 - 0.6182x^2 - 58.375x + 190.22$	线性: $y = 480.98x + 1089.3$	2024.03 ~ 2025.03
上海中医药大学	生物学与生物化学	三次: $y = 0.4445x^3 + 38.909x^2 - 291.04x + 471.35$	幂: $y = 3794.3x^{0.3395}$	2024.03 ~ 2025.03
南京中医药大学	化学	二次: $y = 40.391x^2 - 204.11x + 239.66$	幂: $y = 5536.1x^{0.3205}$	2023.03 ~ 2024.03
中国药科大学	材料科学	三次: $y = 363.134 - 70.894x - 7.302x^2 + 1.787x^3$	二次: $y = 873.835 + 469.744x - 0.215x^2$	2019.03 ~ 2020.03
广州中医药大学	化学	三次: $y = 0.069x^3 + 12.836x^2 - 112.88x + 175.32$	幂: $y = 3588x^{0.4315}$	2034.03 ~ 2035.03

(续表)

院校	潜力学科	最优拟合回归模型		预计入围 ESI 时间
		学科线	阈值线	
海军军医大学	植物学与动物学	三次: $y = 0.4338x^3 - 9.7914x^2 + 77.616x - 120.37$	幂: $y = 756.4x^{0.5936}$	2022.03 ~ 2023.03
	社会科学总论	三次: $y = 0.128x^3 + 1.1745x^2 - 37.914x + 92.903$	指数: $y = 298.71e^{0.0826x}$	2023.03 ~ 2024.03
	材料科学	三次: $y = 1.5481x^3 - 17.494x^2 + 41.67x + 16.87$	二次: $y = 3.22x^2 + 381.83x + 433.56$	2021.03 ~ 2022.03
空军军医大学	免疫学	三次: $y = 0.1222x^3 + 8.9571x^2 - 144.1x + 420.04$	三次: $y = 0.2582x^3 - 23.143x^2 + 717.99x - 1050$	2019.03 ~ 2020.03
东南大学	分子生物学与遗传学	三次: $y = 2.4973x^3 - 10.354x^2 - 108.23x + 359.78$	线性: $y = 803.4x + 4920.7$	2021.03 ~ 2022.03
南京医科大学	材料科学	三次: $y = 2.3513x^3 - 32.792x^2 + 146.8x - 139.07$	线性: $y = 465.66x + 887.45$	2021.03 ~ 2022.03
苏州大学	农业科学	三次: $y = -1.0395x^3 + 37.166x^2 - 99.498x + 144.69$	幂: $y = 1291.3x^{0.3563}$	2020.03 ~ 2021.03
	环境科学与生态学	三次: $y = 1.5678x^3 - 38.801x^2 + 267.97x - 399.8$	指数: $y = 775.31e^{0.0989x}$	2026.03 ~ 2027.03
徐州医科大学	药理学与毒理学	三次: $y = 0.4152x^3 - 8.7725x^2 + 59.015x - 53.178$	线性: $y = 180.36x + 742.99$	2022.03 ~ 2023.03

3 结语

本研究利用 ESI 和 InCites 数据库对我国 28 所高校进行 ESI 潜力学科识别,并用拟合回归曲线和模拟检索的方法对潜力学科入围 ESI 排名的时间进行预测,并以中国药科大学为例详细解析了识别和预测的方法和过程。任何预测模型既然是根据已有的实际资料建立的,就不可避免地存在误差。本文的预测方法也存在其他几乎所有预测方法具有的误差,即用单纯基于数学模型或统计方法进行预测,而未考虑所有动态的难以预料的变化,许多外界因素如论文激励政策的变化、学科前沿学者的流动、国际合作的加强等,都有可能使学科论文产出和绩效偏离原来的增长趋势。尽管学科在国际上的排名受影响的因素太多,但进行潜力学科的识别和预测对推动我国高校“双一流”建设仍具有较大的指导意义并有一定的促进作用。在本文修稿过程中,中国药科大学的材料科学学科已于 2020 年 1 月进入 ESI 排名,在本研究预测的 2020 年 3 月这个节点之前,证明预测准确,同时也彰显了中国药科大学近年来在材料学科相关领域的强劲发展势头,也表明药理学学科对于材料学科等交叉学科具有重要支撑作用。若每所高校能够集中资源和精力,根据学科历史上的增长趋势进行准确规划,努力实现学科自身的进步和突破,潜力学科进入全球 ESI 前 1% 指日可待。

参考文献

- [1] 马陆亭. 一流学科建设的逻辑思考[J]. 高等工程教育研究, 2017(1):62-68.
- [2] 夏琬钧,赵颖梅,刘云,等. 基于 Incites 和 ESI 的高校科研产出统计与分析——以西南交通大学为例[J]. 四川图书馆学报, 2014(3):14-16.
- [3] 陈颖瑜,邓小茹. 数据科学驱动下 ESI 潜力学科入围预测方法的实践探索[J]. 中华医学图书情报杂志, 2019,28(8):35-40.
- [4] 管翠中,范爱红,贺维平,等. 学术机构入围 ESI 前 1% 学科时间的曲线拟合预测方法研究——以清华大学为例[J]. 图书情报工作, 2016,60(22):88-93.
- [5] 汤莹. 我国经济学与商学学科分析与潜力预测——基于 ESI 和 InCites 数据分析[J]. 上海教育评估研究, 2019,8(4):66-71.
- [6] 王雪. 基于时间序列模型的高水平学科预测研究[J]. 情报杂志, 2019,38(6):45-49,117.
- [7] 程建萍,刘建辉,叶玫. 基于 ESI 的潜力学科预测模型修正和实证分析[J]. 情报科学, 2018,36(12):22-24,40.
- [8] 戴莹. 灰色预测在高校学科发展中的应用——以合肥工业大学为例[J]. 情报探索, 2018(9):13-17.
- [9] 秦萍,李雪琛,梅秀秀. 高校潜力学科发展预测研究[J]. 情报杂志, 2015,34(1):88-91.
- [10] 徐碧霞. 对三种预测方法的评价[J]. 预测, 1984(6):42-45.
- [11] Nottingham QJ, Cook DF. Local linear regression for estimating time series data[J]. Computational Statistics and Data Analysis, 2001,37(2):209-217.