



大数据背景下精准科研信息服务

2020 年版

作者：王敏杰

时间：2020-04-06

版本：0.1



Victory won't come to us unless we go to it. — M. Moore

目录

1 前言	1
1.1 进度表	1
1.2 需要的配套	1
2 川师大数据	2
2.1 全景对比	2
2.2 学科对比	2
3 学科预测	3
3.1 统计方法	3
3.2 数学学科	3
3.3 物理学科	3
3.4 化学学科	4
3.5 工程学科	4
3.6 计算机学科	4
3.7 如果需要其他学科, 请联系作者	4
4 学院对学科的贡献	5
4.1 研究规模贡献分析	5
4.2 学术影响力贡献分析	5
5 选刊倾向与期刊推荐	6
5.1 各学科论文在各等级期刊上的分布	6
5.2 期刊推荐	6
A 统计口径	7
A.1 学科分类	7
A.2 各学科进入 ESI 的阈值	8
A.3 数据来源	8
A.4 获取方法	8
A.5 学校列表	8

第1章 前言

1.1 进度表

- 文献调研（3月底完成）
- 数据获取（4月中旬完成）
- 数学分析和模型评估（5月中旬完成）
- 可视化（6月初完成）
- 报告初稿（6月底完成）
- 研讨会（待定）
- 正式稿发布（7月初）

1.2 需要的配套

- 吴欣益协助完成数据收集和整理（图书馆提供劳务费）

第2章 川师大数据

2.1 全景对比

横向比较 top30 所师范类高校的学科发展情况

2.2 学科对比

分面各校，高亮川师

第3章 学科预测

本章的主要工作是，计算并预测川师未来三年进入双一流学科的概率。可能一点意义也没有

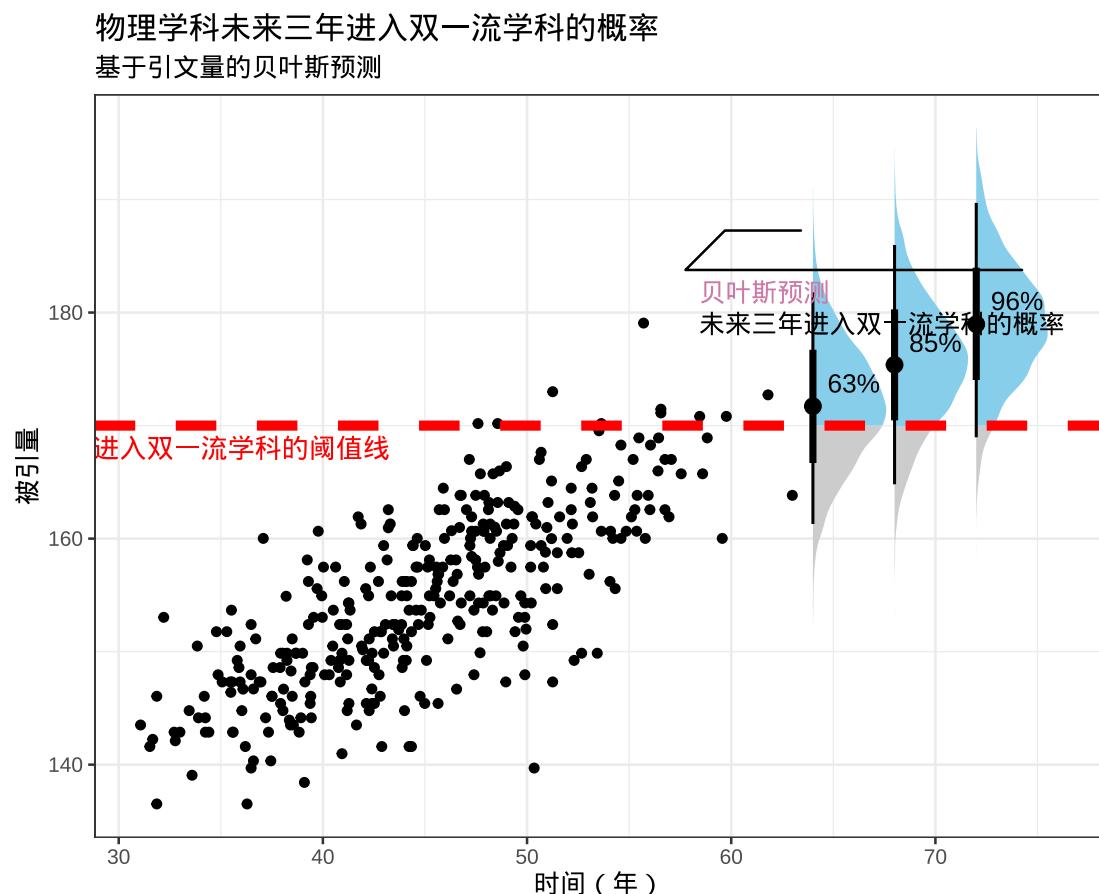
3.1 统计方法

- 贝叶斯数据分析
- 模型不好怎么办？
- log10 scale

3.2 数学学科

3.3 物理学科

```
#> # A tibble: 3 x 3
#>   weight pred_height_mean prob_above_line
#>   <dbl>          <dbl>            <dbl>
#> 1     64           172.            0.631
#> 2     68           175.            0.848
#> 3     72           179.            0.959
```



3.4 化学学科

3.5 工程学科

3.6 计算机学科

3.7 如果需要其他学科, 请联系作者

第4章 学院对学科的贡献

4.1 研究规模贡献分析

数据不准确，避免引起歧义。暂时不开展

4.2 学术影响力贡献分析

数据不准确，避免引起歧义。暂时不开展

第5章 选刊倾向与期刊推荐

5.1 各学科论文在各等级期刊上的分布

数据不准确，避免引起歧义。暂时不开展

5.2 期刊推荐

数据不准确，避免引起歧义。暂时不开展

附录 A 统计口径

A.1 学科分类

ESI22个学科

学科门类	学科
工学 (3)	计算机科学(Computer Science)
	工程科学(Engineering)
	材料科学(Materials Sciences)
生命科学 (4)	生物与生化(Biology & Biochemistry)
	环境／生态学(Environment/Ecology)
	微生物学(Microbiology)
	分子生物与遗传学(Molecular Biology & Genetics)
社会科学 (2)	一般社会科学(Social Sciences ,General)
	经济与商学(Economics & Business)
理学 (5)	化学(Chemistry)
	地球科学(Geosciences)
	数学(Mathematics)
	物理学(Physics)
	空间科学(Space Science)
农学 (2)	农业科学(Agricultural Sciences)
	植物与动物科学(Plant & Animal Science)
医学 (5)	临床医学(Clinical Medicine)
	免疫学(Immunology)
	神经科学与行为(Neuroscience & Behavior)
	药理学与毒物学(Pharmacology & Toxicology)
	精神病学／心理学(Psychology/Psychiatry)
其他 (1)	多学科(Multidisciplinary)

图 A.1: ESI 学科分类

表 A.1: 各学科进入 ESI 的阈值 (2017 年 8 月数据)

学科	阈值
农业科学	1692
生物学与生物化学	5079
化学	5981
临床医学	1855
计算机科学	2837
经济与商业	3549
工程学	1876
环境科学与生态学	3319
地球科学	4795
免疫学	3844
材料科学	3918
数学	3620
微生物学	4421
分子生物学与遗传学	10243
综合交叉学科	2087
神经系统学与行为学	4959
药理学和毒理学	2864
物理学	14029
植物学与动物学	2236
精神病学与心理学	3464
社会科学总论	1142
空间科学	27851

A.2 各学科进入 ESI 的阈值

A.3 数据来源

- 用 ESI 不用 wos
- 2010 - 2019 十年，6 个学科（数学，物理，化学，工程，计算机）
- 获取下载地址
 - 链接 1，检索学校历年发文量的
 - 链接 2，近期进入 ESI 学科的阈值

A.4 获取方法

整理的 raw-data 可以在这里找到

A.5 学校列表

- 师范类学校清单 (<https://www.dxsbb.com/news/1448.html>)
- 选取依据 (top30) (川师 25 名)

- 省份中文名英文名