

“科研团队量化”建模与实验报告

任务编号:

编者:李其炎、陈坤、李月

日期: 2021.8.30

摘要

科研团队承担着高校教学和科学攻关的重任,是高校建设中非常重要的环节,因此如何量化分析科研团队的科研实力,以便于团队自身和高校层面更好的改进发展是非常重要的,但现有研究中少有针对于科研团队量化分析方法的研究,故本项目以“科研团队量化分析”为落脚点,研究团队量化分析的具体思路和方法。

本项目以高校科研团队项目报告(包含若干科研团队项目基本情况,如项目名称,科研经费等)为基本分析出发点,借助于自然语言处理,网络爬虫技术划分出科研项目的学科归属,同时依赖于数学建模,数据分析等等知识挖掘评价指标,建立评价数学模型进行科研团队量化,最终获得科研团队总体排名,高校学科排名情况等,提供了进行科研团队分析的完整思路以及方法。

关键词: 科研团队, 量化分析, 数学建模, 自然语言处理

目 录

摘要.....	1
1. 问题描述.....	3
2. 问题分析.....	3
2.1. “项目学科划分”的分析.....	3
2.2. “评价指标及模型”的分析.....	4
3. 基本假设.....	5
4. 定义与符号说明.....	5
4.1. 名词解释.....	5
4.2. 符号说明.....	5
5. “项目学科划分”方向的建模与实验.....	6
5.1 承办单位特色学科获取.....	6
5.2 项目最终学科的建模与实验.....	6
5.2.1 准备工作.....	6
5.2.2 模型求解及检验.....	7
5.2.3 模型求解.....	8
5.3 方法评价与改进.....	10
5.3.1 模型优点.....	10
5.3.2 模型缺点.....	10
5.3.3 模型改进方向.....	10
6. “评价指标及模型建立”方向 TOPSIS 解法的建模与实验.....	11
6.1 学校全部科研团队实力量化评价的建模与实验.....	11
6.1.1 准备工作——建立评价指标.....	11
6.1.2 模型建立——建立 TOPSIS 模型.....	11
6.2 学校各学科综合实力量化评价的建模与实验.....	13
6.2.1 准备工作——建立评价指标.....	13
6.2.2 模型的建立与求解.....	13
6.3 各学科下科研团队实力量化评价的建模与实验.....	13
6.3.1. 准备工作——建立评价指标.....	13
6.3.2 模型的建立与求解.....	14
6.4 方法评价与改进.....	14
6.4.1 模型优点.....	14
6.4.2 模型缺点.....	14
6.4.3 模型改进方向.....	14
参考文献.....	14
附件.....	15

1. 问题描述

科研团队建设是高校建设的重要一环，承担着高校科研和学生教学的重要任务，因此如何正确地量化分析科研团队的科研实力以辅助团队和高校发展是非常重要的。本文以高校若干科研团队的相关数据（包括但不限于团队名称，负责人，科研经费，项目名称等等）为研究对象，研究分析如何从科研团队过往的历史数据中挖掘有价值信息，提取评价指标，建立评价模型，用以评价所有科研团队的总体实力、某学科下科研团队实力排名以及学校层面学科的综合实力排名。

2. 问题分析

科研团队量化分析，重点在于解决两个问题：

- (1) 借助于哪些指标来量化科研团队的科研实力
- (2) 建立怎样的评价模型来评价科研团队的科研实力。

这其中，评价指标的建立可以借助于日常经验，在现有数据中筛选出最具代表性，最具有评价价值的属性作为评价的指标，评价模型可以借助于评价体系中最常用的 TOPSIS 模型建立评价体系，完成科研团队的量化评价工作。

但欲完成目标结果，反映某学科下科研团队的实力排名以及学校层面学科的综合实力排名，仅仅筛选评价指标，建立评价模型无法完整地量化任务，更需要给出科研项目所属的学科门类，以便于统计特定学科下的具体情况以及对学科进行量化分析。如何根据现有信息完成项目的学科归属划分，是本项目重要的一环，本项目中借助于项目名称与承办单位结合完成项目学科的归属划分。

2.1. “项目学科划分”的分析

“项目学科划分”是项目进程中必不可少的环节，如何根据现有信息指标挖掘出价值，建立模型，完成学科分类。“项目学科划分”需要解决的仍旧是选取哪些属性以及如何根据属性划分学科归属两个主要问题：

- (1) 针对于项目学科划分指标选取问题：

项目名称往往是项目基本情况的概述和缩略，因此是划分学科归属的首要指标，但部分项目名称由于种种原因无法反映项目具体情况，故需要借助于承办单位辅助完成项目学科归属的划分。

- (2) 针对于如何根据属性划分学科归属：

本项目中借助于自然语言处理，网络爬虫，数学建模等等相关知识完成学科归属的划分。项目名称作为首要指标，本文中借助于 python 自动化技术将项目名称输入知网中并将匹配到的学科列表利用网络爬虫技术抓取到本地，同时为解决部分项目名称无法反映项目具体情况的问题，借助于承办单位辅助完成学科归属划分，将知网爬取到的学科列表与承办单位进行匹配，获取匹配程度列表，最终建立数学模型综合以上指标完成项目学科归属的划分。

但在根据属性划分学科归属的进程中，会衍生出诸多附加问题，如项目全称输入知网会造成匹配结果为空，故需要对项目名称进行关键词提取的操作，另外为接近特定院校的学科设置，因此对机器获得的结果采取人工筛选方式以获得更加精准的学科划分结果。

2.2. “评价指标及模型”的分析

项目处理的重点在于选取哪些指标作为评价标准以及如何科学地建立评价体系完成科研团队量化。项目的预计产出结果主要分为三个方面：

- (1) 科研团队的总实力排名
- (2) 特定学科下科研团队排名
- (3) 学校层面学科实力排名。

为使得评价体系一致，参考价值高，本文中对于三种情况的量化均采用 TOPSIS 评价模型，仅建立不同的指标完成量化分析。

一、评价指标的选取：

- (1) 针对于“科研团队总体排名”的问题：

科研经费作为已有数据中的显式数据，应作为评价体系中的首要因素，但为提高数据的可参考价值，本实验中选用年均经费用以消除项目年限对于量化分析的影响。同时，还需要挖掘现有数据中的隐式数据，项目所属学科划分之后，可以统计得到的某一科研团队负责的项目数量以及某一科研团队负责的项目所跨学科的数量，本实验中选取此三个指标作为科研团队总体排名时的指标，进而应用 TOPSIS 模型计算得到科研团队的综合评价指数，进而获得科研团队的总体排名。

- (3) 针对于“特定学科下科研团队的排名”的问题：

相比于“总体排名”问题，此时跨学科数量不能作为团队量化的指标，进而利用科研团队总体排名中获得的评价指数作为新增的指标，其他的两个指标：年均经费和项目数量不做改动，进而应用 TOPSIS 模型获得特定学科下科研团队的综合评价指标。

- (4) 针对于“学校层面的学科实力排名”的问题：

相比于以上两问题，此时项目数量更改为某一学科下项目的总数量，同时为保证学科评价的公平性，将年均经费改为学科下所有项目总经费作为评价学科实力的一个指标，同时对本学科下所有团队的综合评价指数做平均处理，获得评价的第三个指标，进而应用 TOPSIS 模型获得学校层面的学科的总体排名。

二、评价模型的建立：

在评价指标明确确立的基础上，选取何种模型量化科研团队实力是决定项目参考价值的另外一重点指标，在对比分析现有评价模型的基础下，根据项目的实际情况，选用 TOPSIS 模型作为科研团队评价的量化模型：

TOPSIS 模型的基本思路在于在现有评价指标中，分别选取出最优解和最劣解，通过计算每一样本与最优解、最劣解之间的距离，并通过公式借助于两个距离计算出综合评价指数，进而借用此作为科研团队排名的依据，获得科研团队的总体排名。同时为保证评价体系的一致性，针对于“某一学科下的科研团队排名”和“学校层面的学科排名”在本实验中均采用 TOPSIS 模型进行量化。

3. 基本假设

- (1) 假设科研团队所有项目基本情况均已包含在项目文件中，不考虑科研团队存在除项目文件包含外的任何项目结果。
- (2) 假设项目文件中的所有数据均真实可靠，具备一定的参考价值

4. 定义与符号说明

4.1. 名词解释

- (1) 效益型指标 (Profitability indicators) 对评价指标具有增长方向的指标。
- (2) 正理想解、负理想解 (Positive and negative ideal solution) 在各个备选方案中某属性值里最优与最劣的解。
- (3) 排列指标值 (index value of ranking) 经模型计算，所得各科研团队的综合评价指数，以此为依据可初步判断各对象的综合实力大小。

4.2. 符号说明

符号	说明
$A = (a_{ij})_{m \times n}$	TOPSIS 模型原始评价指标数据决策矩阵
$B = (b_{ij})_{m \times n}$	TOPSIS 模型指标数据规范矩阵
$C = (c_{ij})_{m \times n}$	TOPSIS 模型指标数据加权矩阵
$w = [w_1, w_2, \dots, w_n]^T$	TOPSIS 模型权重向量
C^*	TOPSIS 模型正理想解
C^0	TOPSIS 模型负理想解
s_i^*	i 到正理想解的距离
s_i^0	i 到负理想解的距离
f_i^*	TOPSIS 模型获得的综合评价指数

5. “项目学科划分”方向的建模与实验

5.1 承办单位特色学科获取

项目所属学科的划分决定于知网预处理匹配出的学科以及承办单位对匹配出的学科是否开设或研究。即在知网所匹配出学科（小于等于 10）之中，逐一进行该学科与承办单位的匹配，以此来获取项目最终对应的学科。在这一过程中，承办单位所对应学科的获取显得尤为重要。

结合实际情况，本项目采取的方法：专业+研究方向+承办单位学科授位点三者的综合考虑。具体来说，即到学校承办单位对应的官网进行信息捕捉与获取。其中学院一类的承办单位信息来源：[学院部门](#)，实验室一类承办单位信息来源：[科研平台](#)。

经过对各承办单位专业、研究方向、学科授位点的信息汇总，得到 49 个承办单位特色学科如下（以其中六个单位为例进行展示）：

机械与电气工程学院	光电科学与工程学院	自动化工程学院	资源与环境学院	计算机科学与工程学院（网络空间安全学院）	航空航天学院
机械设计制造及其自动化	光电信息	无线电测量	地球科学	计算机科学与技术	控制科学与工程
电气工程及其自动化	光电工程	自动控制	电子信息	网络空间安全	航空宇航科学技术
智能电网信息工程	电子器件	仪器科学	空间信息 数字技术	计算机软件与理论	探测制导与控制技术
工业工程	光电子技术	控制科学	地球信息科学	信息安全	航空航天工程
机器人工程	电真空器件	测控技术仪器	环境科学		无人驾驶航空器系统工程
		遥控遥测解算装置			飞行器控制与信息工程

5.2 项目最终学科的建模与实验

5.2.1 准备工作

(1) 项目名称预处理：

项目名称预处理即对项目文件中项目的完整名称进行关键词提取的操作，主要为解决网络爬虫过程中项目全名称输入返回为空的问题。

项目名称的预处理主要由以下三个步骤组成：

1. 进行停用词去除。去除与主题关联性不高的停用词，如“方法”“应用”等等

2. jieba 关键词的提取。利用 jieba 中文分析进行语句分词、关键词提取操作
3. 专用词汇添加。添加分词过程中湮没的专业词汇，但此专业词汇对学科信息的检索非常关键，如“大数据”“区块链”等

(2) 知网学科匹配及爬虫

知网学科的匹配与爬虫目的在于获得项目的所属学科列表（知网提供关键词文献分析分功能），处理好的项目名称用于知网爬虫匹配，以获取与该项目相关程度最高的排名前十的学科。

知网学科匹配与爬虫任务分四步进行：

1. 发送请求。以项目名称关键词为请求参数，向‘知网全部检索结果分析’目标站发送请求。
2. 获取检索结果，获取各学科的分布情况。
3. 解析响应内容。解析排名前十的学科，并解析各学科对应的占比及文献数量。
4. 保存解析数据结果。

(3) 承办单位二次匹配

承办单位二次匹配主要为解决部分项目名称由于某种原因无法反应项目基本情况的问题（如 XXX 合同），此时，一般情况下，项目所在的承办单位作为关键信息，也在一定程度上蕴含着项目的学科信息。故此时借助于承办单位二次匹配获得更加精准的学科归属划分。

承办单位二次匹配主要由三个步骤完成。

1. 获取每个项目的承办单位。
2. 根据承办单位进行相应的学科专业获取。在 5.1 所获取的‘承办单位特色学科’基础上去除“学院”、“与”等干扰词汇。
3. 进行知网学科和承办单位学科的相似文本匹配。运用 `fuzz.token.sort(s1,s2)` 方法对知网的每一学科均进行承担单位对应专业的列表匹配，选取最高匹配率作为匹配结果。

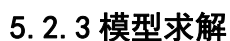
5.2.2 模型求解及检验

项目目前已知结果有：知网匹配率前十的学科、前十学科所对应比率、承办单位对知网每一个学科的匹配率。在此基础上求解项目的最终学科，运用的是“类决策树”的方法：即根据每个项目的学科实际情况分类分析得出最终结果。

具体的决策方法为：

1. 若知网匹配仅包含一个学科，该学科即为项目对应学科；
2. 若知网匹配包含两个学科，则对学科一和学科二分别进行“ $\text{知网比率} \times 0.4 + \text{承担单位匹配率} \times 0.6$ ”的计算得出 `result1`、`result2`，若 `result1` 与 `result2` 相差小于 0.1，则说明两个学科势均力敌同时选取，否则选取 `result` 大所对应学科；
3. 若知网匹配出三个及以上数目的学科，那首先按照知网匹配率和承担单位匹配率由高到低进行排序得到 `sort1`、`sort2`。接下来取 `sort1` 和 `sort2` 的交集 `m`，若交集 `m` 个数大于等于 2，则选取交集部分学科作为项目学科。否则进行进一步判定：在 `sort1` 和 `sort2` 中选取前三学科，对于 `sort1` 中前三学科选取其在 `sort2` 中最靠前的学科 `x`，

“类决策树”方法执行流程图:



```
for i in range(project_num):
    #获取学科数目
    subject_num=10
    while subject_num>0:
        if str(data["学科"+str(subject_num)])[3*i]!='nan':
            break
        else:
            subject_num=subject_num-1
    print(subject_num)
#根据学科数目开始进行处理
```



```

if subject_num<=0:
    print("进入 0 学科判断")
    continue
elif subject_num <=1:
    print("进入 1 学科判断")
    final_subject=data["学科 1"][3*i]
    data["学科 1"][3*i+2]=set([final_subject])
elif subject_num <=2:
    print("进入 2 学科判断")
    result_1=0.4*float(data["学科 1 比率"][3*i])+0.6*float(data["学科 1 比率"][3*i+1])
    result_2=0.4*float(data["学科 2 比率"][3*i])+0.6*float(data["学科 2 比率"][3*i+1])
    if abs(result_1-result_2) <=0.1:
        final_subject=[data["学科 1"][3*i],data["学科 2"][3*i]]
    else:
        final_subject=[data["学科 1"][3*i] if (result_1 >result_2) else data["学科 2"][3*i]]
    data["学科 1"][3*i+2]=set(final_subject)
else:
    print("进入 3 学科判断")
    cnki_subject_order=[" " for x in range(subject_num)]
    cnki_subject_ratio=[0 for x in range(subject_num)]
    third_for_cnki=[" " for x in range(3)] #知网前三学科
    third_ratio_cnki=[0 for x in range(3)] #用以存储知网前三学科的比率
    organizer_subject_order=[" " for x in range(subject_num)]
    organizer_ratio=[0 for x in range(subject_num)]#原始承担单位比率
    organizer_subject_ratio=[0 for x in range(subject_num)]#排序后单位比率
    third_for_organizer=[" " for x in range(3)] #承担单位前三学科
    third_ratio_organizer=[0 for x in range(3)]
    #对知网和承担单位进行匹配度排序
    for j in range(subject_num):
        cnki_subject_order[j]=data["学科"+str(j+1)][3*i]
        cnki_subject_ratio[j]=data["学科"+str(j+1)+"比率"][3*i]
        organizer_ratio[j]=data["学科"+str(j+1)+"比率"][3*i+1]
    #判断是否所有学科匹配均为 0
    if (np.array(organizer_ratio) ==0).all():
        final_subject=data["学科 1"][3*i]
        data["学科 1"][3*i+2]=set([final_subject])
        continue
    cnki_ratio_dict=dict(zip(cnki_subject_order,cnki_subject_ratio))
    organizer_ratio_dict=dict(zip(cnki_subject_order, organizer_ratio))
    dic1SortList = sorted(organizer_ratio_dict.items(),key = lambda x:x[1],reverse = True)
    #排序后的承担单位学科和匹配率列表
    for j in range(subject_num):

```

```

organizer_subject_order[j]=dic1SortList[j][0]
organizer_subject_ratio[j]=dic1SortList[j][1]
for j in range(3):
    third_for_cnki[j]=cnki_subject_order[j]
    third_for_organizer[j]=organizer_subject_order[j]
set_common=set(third_for_cnki).intersection(set(third_for_organizer))
if len(set_common)>=2:
    final_subject=set_common
    data["学科 1"][3*i+2]=final_subject
else:
    #分别得到知网前三对应的承担单位匹配率和承担单位前三对应的知网匹配率
    for j in range(3):
        third_ratio_cnki[j]=organizer_ratio_dict[third_for_cnki[j]]
        third_ratio_organizer[j]=cnki_ratio_dict[third_for_organizer[j]]
    index_1=organizer_subject_ratio.index(max(third_ratio_cnki))
    index_2=cnki_subject_ratio.index(max(third_ratio_organizer))
    set_1=[organizer_subject_order[index_1]]
    set_2=[cnki_subject_order[index_2]]
    final_subject=set(set_1).union(set(set_2))
    final_subject=set(final_subject).union(set_common)
    data["学科 1"][3*i+2]=final_subject

```

5.3 方法评价与改进

5.3.1 模型优点

- (1) 项目名称和承担单位双重出发，更巧妙地避开了单一元素决策的偶然性。
- (2) 结合每一项目的具体情况做判断，减小普适性方法可能存在的误差。
- (3) 学科归属各过程均为自动化程序处理，便捷且执行效率高。

5.3.2 模型缺点

- (1) 承担单位二次匹配基于已有的知网匹配学科，耦合性强，而不能单纯以承办单位自主推举出认定学科。
- (2) 知网学科匹配比率与该学科在知网中的文献数量存在不可分离的联系，文献数量不均衡可能会对知网学科匹配的结果产生部分影响。
- (3) 知网匹配与项目名称处理、承担单位二次匹配与知网处理结果均存在前因后果关系，各处理结果之间独立性弱、联系紧密度高，其中一环的失误会造成连带影响。

5.3.3 模型改进方向

- (1) 项目名称处理部分再加强可能会使知网学科匹配结果更加准确。
- (2) 可综合考虑多数据库，例如增加“万方数据知识服务平台”的学科匹配结果，多中数据库减少单一可能带来的误差，增加结果的可参考价值。

(3) 学科归属划分的“类决策树”可进行再细化使结果更具有针对性。

6. “评价指标及模型建立”方向 TOPSIS 解法的建模与实验

6.1 学校全部科研团队实力量化评价的建模与实验

6.1.1 准备工作——建立评价指标

(1) 评价指标的确定

通过第一阶段项目学科划分模型的建立与实验，能够初步准确得到各项目所属学科，以此为基础，建立学校所有科研团队的评价模型。利用已有数据条件，选择适当的评价指标，能够合理反映各科研团队的综合实力水平。根据已知各科研项目信息及其所属学科，本模型采用各科研团队的“年均经费”、“项目数量”、“学科交融性”为三大评价指标，构建 TOPSIS 模型。

(2) 评价指标的建立

本模型使用“年均经费”、“项目数量”、“学科交融性”三大指标评价科研团队的综合实力，为合理构建并量化模型指标，使其能够更好地用于模型中，本模型采用“项目总经费/项目进行时间”获得单个项目的年均经费；统计外协、横向、纵向项目的各团队数量，获得各团队项目总数量；以项目为单位提取每个项目对应的学科，计算同一团队所作项目中包含的学科总数，作为对科研团队学科交融性的量化指标。

6.1.2 模型建立——建立 TOPSIS 模型

(1) 数据预处理

本模型所采用指标均为效益型指标，采用向量规范化的方法能够有效地将效益值属性达到规范化，归一化。无论成本型属性还是效益型属性，设原始指标数据决策矩阵为 $A = (a_{ij})_{m \times n}$ ，变换后的矩阵为 $B = (b_{ij})_{m \times n}$ 向量规范化均用下式进行变换：

$$b_{ij} = a_{ij} / \sqrt{\sum_{i=1}^m a_{ij}^2}, i = 1, \dots, m, j = 1, \dots, n。$$

采用向量规范化的最大特点是，规范化后，各方案的同一属性值的平方和为 1，因此常用于计算各方案与某种虚拟方案（如理想点或负理想点）的欧几里得距离的场合。

(2) 加权向量规范化

由于模型所选取的指标对团队综合实力评价的影响大小各不相同，故构建权重向量，给予属性不同的权重，能够更为合理准确地达到评价团队综合实力的目的。

构造加权矩阵 $C = (c_{ij})_{m \times n}$ ，由决策人设定权重向量 $w = [w_1, w_2, \dots, w_n]^T$ ，则有：

$$c_{ij} = w_j \cdot b_{ij}, i = 1, 2, \dots, m; j = 1, 2, \dots, n。$$

根据对科研团队实力影响因素考量及相关资料，决定分别给予“年均经费”、“项目数量”、“学科交融性”，0.5、0.3、0.2的权重。即权重向量为 $w = [0.5, 0.3, 0.2]^T$ 。

(3) 确定正理想解与负理想解

正理想解即为各科研团队经过向量规范和权重规范后属性值中的最大值，而负理想解即为各科研团队经过向量规范和权重规范后属性值中的最小值。

确定正理想解 C^* 和负理想解 C^0 。设正理想解 C^* 的第 j 个属性值为 c_j^* ，负理想解 C^0 第 j 个属性值为 c_j^0 ，则

$$\text{正理想解} : c_j^* = \max_i c_{ij}, j = 1, 2, \dots, n \text{ 其中 } j \text{ 为效益性属性。}$$

$$\text{负理想解} : c_j^0 = \min_i c_{ij}, j = 1, 2, \dots, n \text{ 其中 } j \text{ 为效益性属性。}$$

(4) 计算各方案到达各理想解的距离

科研团队 d_i 到正理想解的距离为

$$s_i^* = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^*)^2}, i = 1, 2, \dots, m;$$

科研团队 d_i 到负理想解的距离为

$$s_i^0 = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^0)^2}, i = 1, 2, \dots, m。$$

(5) 计算排列指标值

排列指标值，即为对各科研团队的综合评价指数，所得综合评价排序，则以此为依据进行倒序排序。具体的综合评价指数如下所示：

$$f_i^* = s_i^0 / (s_i^0 + s_i^*), i = 1, 2, \dots, m。$$

(6) 按 f_i^* 由大到小排列方案的优劣次序。

6.1.3. 模型求解

根据模型原理，即对相关算法实现程度考量，使用 Python 语言对模型过程进行了算法实现，具体程序内容见附页。模型及其算法实现过程共分为三部分：

1. 数据的读取与写入
2. 数据的规范化与预处理
3. 项目评价指标的求解

6.2 学校各学科综合实力量化评价的建模与实验

6.2.1 准备工作——建立评价指标

(1) 评价指标的确定

本模型在第一阶段项目学科划分模型建立与实验的基础上，将各项目划归到其所属学科下，统计各学科下所有项目的相关信息，建立学校各学科综合实力量化的评价模型。利用已有数据条件，选择适当的评价指标，能够合理反应各学科的综合实力水平。根据已知各科研项目信息及其所属学科，本模型采用各学科下所有所属项目的“项目总经费”、“团队总数量”、“各科研团队平均综合评价指数”为三大评价指标，构建 TOPSIS 模型。

(2) 评价指标的建立

本模型使用各学科下所有所属项目的“项目总经费”、“团队总数量”、“各科研团队平均综合评价指数”三大指标作为评价学校各学科的综合实力，为合理构建并量化模型指标，使其能够更好地用于模型中，本模型采用“各学科下的所有项目的经费之和”作为学校项目总经费指标；统计学科下所属团队总数量，作为另一评价指标；以“学校所有科研项目团队评价模型”中获得的综合评价指数，统计学科下所有科研团队的平均综合评价指数，作为综合指标。

6.2.2 模型的建立与求解

根据对科研团队实力影响因素考量及相关资料，决定分别给予“项目总经费”、“团队总数量”、“各科研团队平均综合评价指数”，0.5、0.3、0.2 的权重。即权重向量为 $w = [0.5, 0.3, 0.2]^T$ 。

在各学科下分别构建 TOPSIS 模型，具体模型的建立与求解过程与 6.1.2，6.1.3 无异，本文对此不再赘述。

6.3 各学科下科研团队实力量化评价的建模与实验

6.3.1 准备工作——建立评价指标

(1) 评价指标的确定

本模型在第一阶段项目学科划分的模型建立与实验的基础上，将各项目划归到其所属学科下，在各学科下建立各学科下科研团队的评价模型。利用已有数据条件，选择适当的评价指标，能够合理反应各学科下科研团队的综合实力水平。根据已知各科研项目信息及其所属学科，本模型采用各学科下各科研团队的“项目数量”、“年均经费”、“综合评价指数”为三大评价指标，构建 TOPSIS 模型。

(2) 评价指标的建立

本模型使用“项目数量”、“年均经费”、“综合评价指数”三大指标作为评价同一学科下科研团队的综合实力，为合理构建并量化模型指标，使其能够更好地用于模型中，本模型采用“特定团队在特定学科下的所有项目的年均经费”获得单个项目的年均经费；统计特定团队在特定学科下的项目数量，获得各团队项目总数量；以“学

校所有科研项目团队评价模型”中获得的综合评价指数，作为对科研团队发展前景的量化指标。

6.3.2 模型的建立与求解

根据对科研团队实力影响因素考量及相关资料，决定分别给予“项目数量”、“年均经费”、“综合评价指数”，0.3、0.5、0.2的权重。即权重向量为 $w = [0.3, 0.5, 0.2]^T$ 。

在各学科下分别构建 TOPSIS 模型，具体模型的建立与求解过程与 6.1.2，6.1.3 无异，本文对此不再赘述。

6.4 方法评价与改进

6.4.1 模型优点

(1) 对数据的处理尽可能避免了主观性影响。模型采用了 TOPSIS 综合评价模型，避免了对数据指标的主观处理，在不需要设定目标函数，不需要通过检验的过程下，依然后能够很好的刻画多个影响指标的对相关综合实力的影响力度。

(2) 对数据分布及样本含量没有严格限制，数据计算简单易行。模型对于数据分布及样本量、指标多少无严格限制，既适于小样本资料，也适于评价多单元、多指标的大系统，较为灵活、方便。对数据分布及样本含量没有严格限制，数据计算简单易行。

(3) 模型能充分利用原始数据的信息，其结果能精确地反映各评价方案之间的差距，并能够给出具体的评价指数信息。

6.4.2 模型缺点

(1) 模型虽已尽可能避免主观性的影响，但在对应的量化指标选取与权重的赋予上依然存在一定的主观性。受获得项目信息影响，在量化指标的选取上具有一定的局限性。

(2) 模型建立在科研团队划分基础上，受科研项目所属学科影响较大，对科研项目所属学科划分的准确性要求较高。

6.4.3 模型改进方向

在获得更多项目信息的基础上，合理拓展所选评价指标，能够更进一步提高模型的合理性与准确性。

参考文献

[1] 司守奎，孙兆亮. 数学建模算法与引用[M]. 北京：国防工业出版社，2020：369-375.

[2] TOPSIS 法(优劣解距离法)介绍及 python3 实现

[EB/OL]. <https://zhuanlan.zhihu.com/p/37738503>, 2021-02-01.

[3] 评价模型-TOPSIS 法

[EB/OL]. https://blog.csdn.net/qq_39742734/article/details/82053127, 2018-08-25.

附件

附件清单

附件一：项目源代码

附件二：科研团队总排名

附件三：特定学科下科研团队排名

附件四：学校学科总体排名