

# 基于数据分析模型的项目负责人评估机制研究

贾异<sup>1,2</sup>, 童杨<sup>1</sup>, 刘栋<sup>1</sup>, 宋阳<sup>1</sup>, 卞曙光<sup>1</sup>

(1.科学技术部 高技术研究发展中心,北京 100044;2.天津大学 管理与经济学部,天津 300072)

**摘要:**国家科技计划项目的实施是国家推动科技发展,突破关键技术,实现科技强国的重要途径之一,而项目负责人的综合能力直接影响项目的实施及综合绩效评价,项目负责人评估机制的完善优化有助于促进国家科技计划项目的管理成效。基于此,提出了一种利用数据分析建模预测项目实施效果的方法,通过对项目负责人综合能力的分析,实现了对项目实施效果的预测,借助预测结果对项目负责人综合能力进行评估,利用此方法对项目负责人的评估机制进行优化,提高项目管理效率,优化资源配置。通过国家科技计划项目实例对模型进行了验证,证明了其准确性及有效性。

**关键词:**国家科技计划;数据分析;项目负责人;综合绩效评价;评估机制

**中图分类号:**G311;D922.17

**文献标志码:**A

科学技术是推动世界发展的关键因素。由于世界经济的迅速全球化,科学技术从根本上决定了国家的核心竞争力。国家科技计划项目的实施是国家提升相应领域科学技术的重要手段,而项目负责人对科技计划项目的执行完成情况有着举足轻重的影响,这就决定了项目负责人在科技计划项目中的重要地位及项目管理部门对项目负责人评估决策的重要性。近些年,大数据的发展已渗透到各行各业<sup>[1]</sup>,国家科技项目管理中也产生了大量的数据,在科技项目管理中融入大数据理念<sup>[2]</sup>,有利于发挥大数据的作用,提高项目管理效率,优化资源配置,提升科技项目执行完成水平,最终达到提高国家核心竞争力的目标。

当前,国家科技计划项目评审工作是一个收集项目评审信息,并对评审信息进行分析 and 归纳的系统过程。例如国家重点研发计划项目评审在着重评价的研发团队及工作基础中对项目负责人有明确的评估要求,而评审工作主要采取同行评议的方式进行,如预评审及视频答辩评审均采用同行评议方式,而同行评议专家对于评估指标的理解程度及各自的评估意见存在差异性,从而导致评议结果具有较大模糊性和不确定性。

利用大数据分析辅助科技计划项目管理正受到越来越多的关注,很多学者从不同的角度给出了研究方法和理论体系。一些研究人员分析了科技项目管理数据的基本特征,并深入研究了这些数据的应用模式。耿国庆<sup>[3]</sup>分析了科技计划项目大数据的新特征,并研究了大数据在科技项目管理中的应用及作用,进而提出了科技计划项目大数据背景下项目管理的策略。刘蔚等<sup>[4]</sup>等通过构建科技计划项目大数据分析模型,形成多种属性特定服务场景的分析产品和决策支撑工具,为利用科技管理大数据支撑科技计划管理和决策提供了参考。

科技计划项目评审必须结合基于事实数据的定量分析和专家智慧的定性分析<sup>[5]</sup>。本文利用大数据分析处理方法,建立科技计划项目负责人与项目效果预测关联模型,探讨基于大数据分析建模的方法来辅助评审专家对项目负责人进行评估的方法。通过对科技计划项目资料的整理和收集,以及对项目负责人及其衍生数据的收集分析,建立模型并根据项目负责人数据预测项目执行完成效果,借助现代科学分析方法,为科技计划

**收稿日期:**2021-10-17;**修回日期:**2021-11-16.

**基金项目:**博士后科学基金特别资助(站中)(2021T140287)

**作者简介:**贾异(1991—),男,河南南阳人,科学技术部高技术研究发展中心助理研究员,研究方向为公共管理,E-mail:jiayi@htrdc.com.

**通信作者:**卞曙光(1964—),男,内蒙古呼和浩特人,科学技术部高技术研究发展中心高级工程师,研究方向为项目管理研究,E-mail:jeanlosg@htrdc.com.

项目管理决策提供参考,以提高科技计划项目管理效率,优化资源配置,落实科技管理“放管服”精神,促进科技计划项目高质量完成,提升科研绩效。

## 1 数据获取

为了促进科技计划项目高质量地完成,建立起项目负责人各因素与项目执行完成效果的关联模型,将数据分为输入数据与输出数据两大类,如表 1 所示,此两大类信息皆为已知数据,其中输入数据为可以反映项目负责人对整个项目影响较大的信息数据,包括:项目负责人个人情况、项目负责人的科研能力、项目负责人的组织能力及边缘效应。输出数据为项目结题情况信息,包括:项目完成情况、项目产出情况、项目创造利润、项目产出投入效益比及项目单位时间产出效益。项目结题情况信息显示的是项目最终完成效果,可以反映出项目负责人的综合能力在项目执行期间对整个项目的影响,因此项目负责人信息与项目的结题情况存在着明显的因果关系。利用大数据分析的方法将两者关联起来,分析两者的关联性,从而达到利用已知项目负责人信息预测项目执行完成情况,并依据预测结果评估项目负责人个人综合能力,为项目负责人评估机制优化提供理论数据依据。

表 1 模型输入输出数据表  
Tab. 1 Input and output data sheet of the model

| 类型   | 一级指标       | 备注   |
|------|------------|--|
| 输入数据 | 项目负责人个人情况  | 包括负责人性别、年龄、学位、技术职称、从事专业、所属单位性质及科研职务等   |
|      | 项目负责人科研能力  | 包括负责人 H 因子*、G 指数*、获得专利数量、承担/参与的项目数量、获得的科研奖励及关注的学术领域等                             |
|      | 项目负责人组织能力  | 包括负责人作为项目负责人的次数、承担项目的质量、承担项目参与单位数量、承担项目参与人员数量、承担项目经费额度及担任的领导职务等                  |
|      | 边缘效应       | 包括负责人与同机构的合作人数、与其他机构的合作人数、合作最紧密的 6 人的学术能力、负责人的学术影响力(参评过国家科技计划项目的数目以及个人获得学术/荣誉头衔) |
| 输出数据 | 项目完成情况     | 专家评议打分   |
|      | 项目产出情况     | 包括项目产出的论文、专利、标准草案、专著、培养的硕博,博士后及工程师等数量、项目执行期间获得的奖励或荣誉等                            |
|      | 项目创造利润     | 项目执行完成期间创造的利润  |
|      | 项目产出投入效益比  | 总产出量化分数与总投入(包括立项经费等投入)量化分数之比   |
|      | 项目单位时间产出效益 | 总产出量化分数与项目执行时间年限之比   |

\* H 因子:又称为 h 指数或 h 因子(h-factor),是一种评价学术成就的新方法,h 代表“高引用次数”(high citations),一名科研人员的 h 指数是指他至多有 h 篇论文分别被引用了至少 h 次;\* G 指数:论文按被引次数排序后相对前排的累积被引至少  $G^2$  次的最大论文序次 G,亦即第  $(G+1)$  序次论文对应的累积引文数将小于  $(G+1)^2$ 。

如表 1 所示输入数据信息包括 4 类,其中负责人个人基本信息在项目申报书中提取,信息包括项目负责人年龄、性别、学位、技术职称、所属单位性质及科研职务等。

而负责人的科研能力不仅代表着其完成整个项目目标的潜力,也代表着其在学术领域的带头作用大小。对负责人科研能力的考量主要从其 H 因子、G 指数、授权专利、承担/参与项目数量、科研奖励情况及关注领域数等方面。H 因子及 G 指数均以百度学术数据库为准;授权专利主要以 soopat 专利数据库为准;参与或承担过的项目数据以中国知网数据库为准;科研奖励数据参考科技专家库服务管理系统;关注领域数据以中国知网数据库为准。

项目负责人作为国家科技计划项目的负责人,其应具备相应的组织能力,对国家级项目进行统筹协调,规划并督促国家科技计划项目按时按标完成。项目负责人的组织能力从承担过国家科技计划项目负责人的次数、承担项目质量、承担项目参与单位数量、承担项目参与人员数量、承担的项目经费、担任的领导职务等方面数据体现。承担国家科技计划项目次数数据参考科技专家库服务管理系统;承担项目质量数据参照中国知网数据库;承担项目参与单位数量、参与人数及项目经费数均来自项目任务书数据;项目负责人在社会中担任领导职务数据从科技专家库服务管理系统中获取。

在信息数据激增的时代,除了项目负责人的主要科研背景对项目执行有一定影响外,负责人的衍生关联信息对项目负责人在项目执行中的影响也不可小视.在这里称之为边缘效应.包括项目负责人的活跃网络数据、学术影响力、社会属性以及项目负责人与项目承担单位间的关系.项目负责人的活跃网络中 6 人学术水平主要从合作人的发文量、下载量及承担/参与基金等方面进行考量.活跃网络数据来自中国知网数据库.学术影响力则包括负责人曾作为专家参评过国家科技计划项目的数目以及个人获得学术/荣誉头衔(中国科学院院士、中国工程院院士、863 首席专家、973 首席专家等),与项目负责人在社会学术组织中的任职情况数据均从科技专家库服务管理系统中获取.

项目负责人与项目承担单位间的关系对于项目实施进展同样有着不可忽视的影响,项目负责人所在单位与项目承担单位一致抑或不一致,都对项目的执行完成情况有正面或负面的影响.

输出数据主要为项目结题情况信息,这一数据主要对照项目立项时任务书规定的目标和任务,关键技术指标等是否达到.这一数据包括 5 类,其中项目完成情况作为固定数据可以定性比较项目完成优劣,同时作为项目负责人评估的重要参考依据.项目执行期间产出包括发表的论文数量,论文级别,申请和授权的专利情况,产生的标准与草案,出版的教材和学术专著,培养的博士后、博士、硕士及工程师等数量,项目团队通过项目实施获得的奖励或荣誉情况.

以上两类信息中,项目负责人信息属于输入数据,而项目结题情况是输出数据,这些数据都需要进行细致谨慎的量化与归一化,以便于后续模拟处理.这些数据中有的数据本身即是数字化的数值,量化比较容易,可以直接采用或者经过简单变换后采用.有的数据是非数值化数据,需要进行处理,可以按照简单直观的原则进行等级分制量化.

## 2 模型建立

本研究为建立一个全面的项目负责人评估模型,选择一个多输入和多输出变量的网络模型.项目负责人的个人情况、科研能力、组织能力及边缘效应 4 类输入数据可以反映出作为一个项目负责人在项目执行中应具备的个人基本素养及能力,其对项目的执行完成情况有着较直接的影响,而代表项目执行完成情况的项目产出、项目创造利润、项目产出投入效益比及项目单位时间产出效益 4 类输出数据可以反映出项目最终的成果呈现,项目完成情况作为项目最终成果的综合评定打分数据可以定性比较项目完成优劣,同时作为项目负责人评估的重要参考依据.具体考虑利用 python 将各输入变量与输出变量建立关联度模型,根据各个输入变量与输出变量的关联度情况,最终形成一个 4 输入、4 输出的拟合模型.主要大项数据包含的子项数据量化如表 2 所示.

如表 2 所示,项目负责人的个人情况中,技术职称如初级、副高级、高级分别取 1、2、3 分.学位按照学士、硕士、博士分别取 1、2、3 分.科研职务按照项目负责人所在单位中担任的职务高低,适当加分;项目负责人的学术能力,如 H 因子、G 指数、专利、承担/参与科研项目数得分则采用原数据暂不做处理,后期统一归一化;项目负责人的组织能力中,获得的科研奖励则按照国家级、省部级、地市级分别计分并累加.作为项目负责人承担过的项目数、参与单位数量、人数及项目经费数都采用原数据,数据处理时归一化;项目负责人的边缘效应,与负责人合作紧密的同机构及不同机构人员数及他们的学术水平(发文量、下载量、基金数)得分为原始数据累加,项目负责人参评过的国家级科技计划项目数量得分为累加数,而其获得的学术/荣誉头衔得分按照层次级别划分并分别赋值累加,最终做归一化处理;项目产出的综合情况,科研成果及奖励中项目产出的论文、专利、标准草案、专著等得分为累加得分.项目周期内获得的荣誉或奖励量化得分按照奖励的级别及数量进行数据累加.项目周期内培养的人才数量及创造的效益、利润等数据累加并最后归一化.

为了建立数据模型,对 70 个已结题国家高技术研究发展计划(863 计划)重大共性关键技术类课题进行回溯整理,将项目任务书及结题报告中对应的数据按照表 1 进行搜集整理,部分数据从网上数据库获取,量化并归一化.将数据库中 70% 作为建立模型的训练集,另外 30% 项目数据作为测试集,通过使用 python 的 pandas 模块,将训练集中项目数据抽象成数组,通过计算皮尔森相关性系数和线性回归算法,计算得到各输入变量与输出变量的关联度模型及定量关系,建立 4 个输入变量分别与 4 个输出变量间关联模型,并形成一

个输出变量对应最终项目完成情况关联模型.利用测试集数据对建立模型进行验证,并通过对结果数据的对比模型进行调整,最终得到模型结果.模型训练过程如图 1 所示.

表 2 详细子项数据表

Tab. 2 Detailed data sheet

| 一级指标   | 对应主要子项数据         | 备注                             |
|--------|------------------|--------------------------------|
| 项目负责人  | 学位、技术职称数据        | 学位及技术职称越高,则分数越高                |
| 个人情况   | 科研职务数据           | 担任科研职务的可适当加分                   |
| 项目负责人  | H 因子和 G 指数数据     | H 因子和 G 指数越高,则得分越高             |
| 科研能力   | 专利数据、承担/参与科研项目数据 | 专利数量及承担/参与科研项目数量越多则分数越高        |
|        | 科研奖励数据           | 根据获得的科研奖励级别及数量累计加分             |
| 项目负责人  | 承担项目负责人数据        | 承担项目负责人次数越多,分数越高               |
| 组织能力   |                  | 承担的项目根据国家级、省级、地市企业分级积分累加       |
|        |                  | 承担项目参与单位数量及人数累加;承担项目经费数        |
|        | 担任领导职务数据         | 在社会中担任领导职务适当加分                 |
| 边缘效应   | 负责人活跃网络数据        | 负责人同机构和不同机构合作人数越多分数越高          |
|        |                  | 负责人合作人学术水平累加计分                 |
|        | 负责人学术影响力数据       | 负责人参评的国家科技计划项目数累加计分            |
|        |                  | 负责人个人获得学术/荣誉头衔根据数量及级别计分        |
| 项目产出情况 | 科研成果、奖励数据        | 项目周期内产生的论文、专利、标准草案、专著等量化得分     |
|        |                  | 项目周期内获得的荣誉或奖励量化得分              |
|        | 培养人才数据           | 项目实施周期内培养的硕博、博士后及工程师等人才量化得分    |
|        | 社会效益             | 项目周期内获得的社会效益,包括创造的利润或成果转化等量化得分 |

2.1 皮尔森相关性系数和线性回归算法

2.1.1 皮尔森相关性系数

皮尔森相关系数<sup>[6]</sup>的定义为:  $\rho_{x,y} = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{\sqrt{N \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2} \sqrt{N \sum_{i=1}^N y_i^2 - (\sum_{i=1}^N y_i)^2}}$ ,由公式可

以看出,  $-1 \leq \rho_{x,y} \leq 1$ ,当  $\rho_{x,y} = 0$  时,  $x$  和  $y$  不具有线性相关的关系;当  $\rho_{x,y} > 0$ ,  $x$  和  $y$  具有正相关的关系,值越大则两者相关性越高;当  $\rho_{x,y} < 0$ ,则  $x$  和  $y$  具有负相关的关系;当  $\rho_{x,y}$  接近  $\pm 1$  时,两者的相关性最高.

分别计算项目产出、产出效率、项目创造利润、总投入与产出效益比与项目负责人的个人情况、学术能力、组织能力和边缘效应的皮尔森相关性系数,研究项目负责人的各项客观因素与项目执行完成效果间的相关性,定性分析项目负责人的各项客观因素对项目完成结果影响大小.

2.1.2 线性回归算法

线性回归(linear regression)<sup>[7]</sup>是一种应用于统计学中的回归分析方法,它能够归纳出自变量和因变量之间相互依赖的定量关系,这种关系通过线性回归方程来表示,其中自变量可以是一个也可以是多个.

线性回归中通常习惯用  $\mathbf{X} = (x_1, x_2, \dots, x_n)^T \in \mathbf{R}^{n \times p}$  表示数据矩阵,其中  $x_i \in \mathbf{R}^p$  表示一个  $p$  维度长的数据样本; $\mathbf{Y} = (y_1, y_2, \dots, y_n)^T \in \mathbf{R}^n$  表示数据的预测值,这里只考虑每个样本为一类的情况,线性回归的

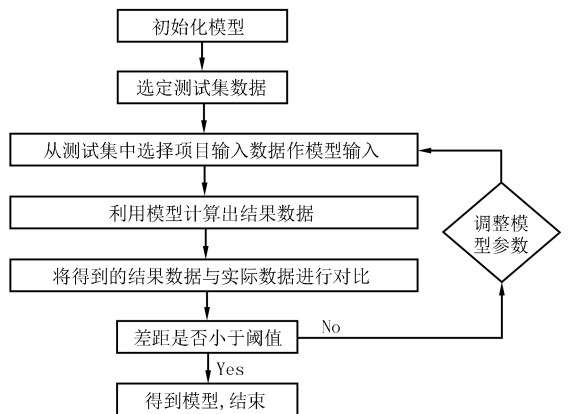


图1 模型训练流程示意图

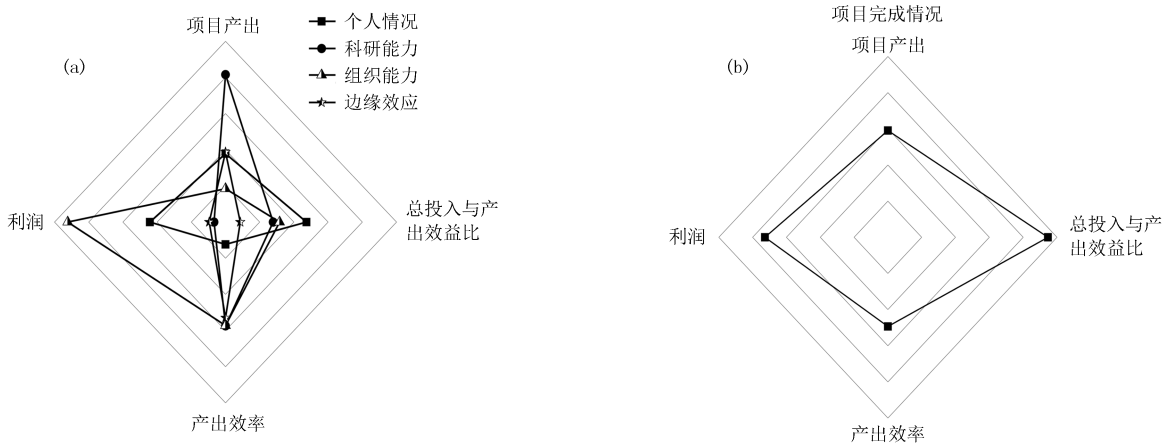
Fig.1 Schematic diagram of model training process

模型如下:  $f(x_i) = \sum_{m=1}^p \omega_m x_{im} + \omega_0 = \omega^T x_i$ , 其中,  $\omega_0$  为截距,  $x_i$  有  $p + 1$  维度.

利用线性回归算法可以定量计算出项目负责人的各个客观因素与项目完成成果间的关系,并最终计算得到项目完成情况的量化得分,通过分数对比判断项目负责人综合客观情况.

### 2.2 模型

分别计算项目负责人个人情况、科研能力、组织能力、边缘效应与项目产出、产出效率、项目创造利润、总投入与产出效益比的皮尔森相关性系数,可以得到每项输出变量与输入变量的相关性关联模型,可以定性分析输入变量对输出变量的影响力大小,同时,建立项目完成情况与各输出变量的相关性关联模型,定性比较各输出变量对项目完成情况的影响.结果如图 2 所示.



(a) 输入数据与输出数据的相关性雷达图; (b) 最终项目完成情况与输出数据的相关性雷达图.

图2 输入与输出数据相关性雷达图

Fig.2 Radar chart of correlation between input and output data

图 2 为建立的各输入数据与输出数据间的模型,由图 2 可以看出,项目产出与项目负责人的科研能力关联度较高,项目负责人科研能力越强,则项目整体的产出就越多;项目产出效率与项目负责人的科研能力、组织能力和边缘效应都有较大关联度.在一定时间内,项目的整体产出除了靠项目负责人的个人科研能力外,项目负责人对整个项目团队的组织越强,项目团队高效协同地完成项目目标成效越好,则产出效率也越高.项目负责人的边缘效应对产出效率影响也较大;项目创造利润与项目负责人的组织能力关联度较大;总投入与产出效益比与项目负责人的个人情况关联度最高,与负责人的科研能力和组织能力也有较强的关联度;最终的项目完成情况则与各个结果数据都有较强的关联度.

利用软件对数据进行分析,通过多元线性回归算法,可以得到各个自变量与因变量间的定量关系,同时还可以确定线性回归关系的显著性,计算过程中将置信度设置为 0.95.假设项目负责人个人情况为  $a$ ,科研能力为  $b$ ,组织能力和边缘效应为  $c$ ,则各个自变量与因变量间的定量关系为:

$$S_{\text{项目产出}} = 0.353 + 0.014a + 0.183b + 0.072c - 0.026d, \tag{1}$$

$$S_{\text{产出效率}} = 0.337 - 0.029a + 0.101b + 0.230c + 0.059d, \tag{2}$$

$$S_{\text{项目创造利润}} = 0.070 + 0.096a - 0.033b + 0.319c + 0.026d, \tag{3}$$

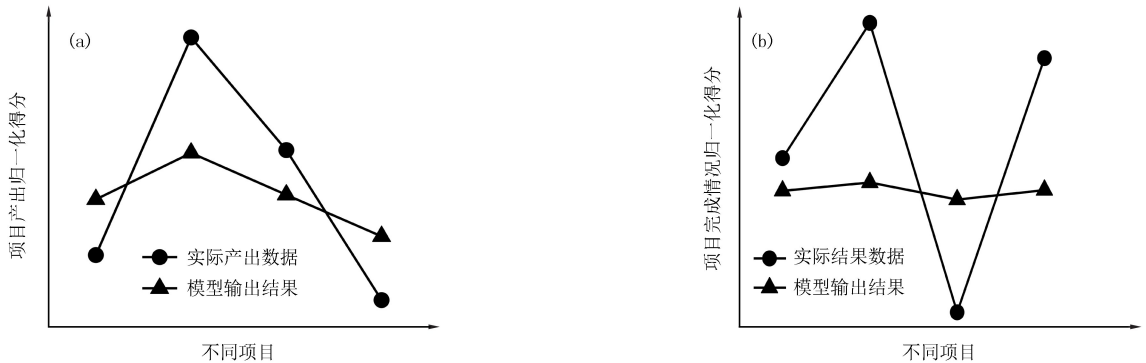
$$S_{\text{总投入与产出效益比}} = 0.081 + 0.074a + 0.030b + 0.089c - 0.012d, \tag{4}$$

$$S_{\text{项目完成情况}} = 0.568 + 0.265S_{\text{项目产出}} - 0.044S_{\text{产出效率}} - 0.050S_{\text{项目创造利润}} + 0.377S_{\text{总投入与产出效益比}}, \tag{5}$$

(1)~(4)式为 4 个输入数据与输出数据间的定量关系公式,(5)式为最终结果项目完成情况与 4 个输出数据间的定量关系公式,各定量关系的相关性判定系数  $R^2$  分别为 0.825 85, 0.831 14, 0.873 89, 0.893 58 及 0.999 03,均接近于 1,方程的显著性检验  $F$  统计值分别为 77.063 87, 79.984 92, 136.453 44 及 16 797.938 75,数值均较大,说明各自变量与因变量间均存在显著的线性回归关系.依据结果来看,样本是符合线性关系的,故用线性模型来表示是合理的,至于其他模型是否也会符合样本数据,需要更进一步的探究.通过公式结果可以看出,项目负责人的科研能力对项目产出影响较大,项目负责人的组织能力对产出效率、项目创造利润

及总投入与产出效益比结果影响较大,这基本与各输出与输入间的相关性分析结果一致。

通过将输入数据输入模型,可以得到项目产出、产出效率、项目创造利润及项目投入与产出效益比情况,最终可以预测项目完成情况得分.利用此模型可以根据项目负责人的总体情况对项目最终实施效果及结果进行预测,以此判断项目申请团队中项目负责人对整个项目执行效果的影响,并利用此模型对项目负责人进行评估.选取项目实际数据对所建立的模型及数据进行验证.图3为4个项目的实际产出数据、实际结果数据与模型输出结果数据对比图.对于结果数据中项目产出的准确率较高,因进行模拟及计算的数据有限,所以结果有一定误差,但最终的结果数据即项目完成情况得分趋势准确,可依据最终预测结果对项目负责人综合情况进行判断,客观地对项目负责人评价评估,能有效支撑对科技计划项目负责人的评估。



(a) 4个项目实际产出数据与模型输出结果的对比图; (b) 4个项目实际结果数据与模型输出结果对比图.

图3 模型输出结果与实际结果对比图

Fig. 3 Comparison between modeling output results and actual results

### 3 评估机制

国家科技计划项目在评审立项时,依然采取申报方填写项目申报任务书,同行评议专家通过对任务书的审阅及个人理解对项目负责人进行评估,大数据分析模型为项目负责人的评估提供了一种新思路,事先通过提交项目负责人个人情况信息,利用大数据分析建模的方法,结合皮尔森相关性系数分析、线性回归分析等计算方法,综合模拟预测出项目完成情况,进而对项目负责人进行评估,评估结果可以提供给同行评议专家,辅助专家在评审立项时对项目负责人评估打分,减少项目评审立项时专家主观性、倾向性的影响,实现国家科技计划项目的智慧评审、精准评审,进而提升国家科技计划项目的科学化管理水平。

### 4 结论与展望

本文借助大数据分析处理方法,对国家科技计划项目负责人总体评价体系进行了建立大数据模型的初步尝试,经验证,模型在项目负责人评估中具有一定参考价值.在本文提出的方法框架下,通过对国家科技计划项目信息进行收集,并搜索补充相关数据,对项目负责人的总体情况数据进行整理分析,通过建立模型对项目实施效果预测,以此来评估项目负责人.该方法为国家科技计划项目评审中对项目负责人的评估提供了一种新的思路,可以利用此方法消除评估中专家主观判断存在的不全面、不客观、不公正的问题,辅助评审专家对项目负责人的评估,从而使项目评审过程更加合理,最终达到促进国家科技发展的目的。

### 参 考 文 献

- [1] 徐宗本,冯芷艳,郭迅华,等.大数据驱动的管理与决策前沿课题[J].管理世界,2014(11):158-163.  
XU Z B, FENG Z Y, GUO X H, et al. Big data driven management and decision-making frontier topics[J]. Management World, 2014(11): 158-163.
- [2] 黄冕,李林.大数据背景下项目管理理论与模式创新研究[J].湘潭大学学报(哲学社会科学版),2020,44(3):25-30.

- HUANG M, LI L. Research on project management theory and model under the background of big data[J]. Journal of Xiangtan University (Philosophy and Social Sciences), 2020, 44(3): 25-30.
- [3] 耿国庆. 大数据时代的科技项目管理[J]. 通讯世界, 2019, 26(11): 10-12.
- GENG G Q. Science and technology project management in big data era[J]. Telecom World, 2019, 26(11): 10-12.
- [4] 刘蔚, 屈宝强, 陈白雪, 等. 基于科技计划项目大数据的情报分析模型研究[J]. 情报理论与实践, 2020, 43(4): 89-93.
- LIU W, QU B Q, CHEN B X, et al. Research on Intelligence Analysis Model Based on Big Data of Science and Technology Projects[J]. Information Studies (Theory & Application), 2020, 43(4): 89-93.
- [5] 姚长青, 张均胜, 董诚, 等. 科技项目评审辅助系统框架与关键技术研究[J]. 情报学报, 2016, 35(9): 932-938.
- YAO C Q, ZHANG J S, DONG C, et al. Research on framework and key technology of science and technology project evaluation auxiliary system[J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(9): 932-938.
- [6] 盛骤, 谢式千, 潘承毅. 概率论与数理统计[M]. 4版. 北京: 高等教育出版社, 2008: 106-112.
- [7] 王惠文, 孟洁. 多元线性回归的预测建模方法[J]. 北京航空航天大学学报, 2007, 33(4): 500-504.
- WANG H W, MENG J. Predictive modeling on multivariate linear regression[J]. Journal of Beijing University of Aeronautics and Astronautics, 2007, 33(4): 500-504.

## Exploration on the evaluation mechanism of project leader assisted by data analysis model

Jia Yi<sup>1,2</sup>, Tong Yang<sup>1</sup>, Liu Dong<sup>1</sup>, Song Yang<sup>1</sup>, Bian Shuguang<sup>1</sup>

(1. High Technology Research and Development Center, Ministry of Science and Technology of the People's Republic of China, Beijing 100044, China; 2. College of Management and Economics, Tianjin University, Tianjin 300072, China)

**Abstract:** The implementation of the National Science and Technology Plan is one of the important ways for our country to promote the development of science and technology, break through the key technologies, and realize the powerful country of science and technology. The comprehensive ability of the project leader directly affects the implementation and integrated performance evaluation of the project. The improvement and optimization of the evaluation mechanism of the project leader will help to promote the management effect of the National Science and Technology Plan project. This paper proposed a method for predicting the implementation effect of the project by using data analysis and modeling. The project implementation effect could be predicted through the analysis of the comprehensive ability of the project leader. With the help of the prediction results, the comprehensive ability of the project leader is evaluated. This method is used to optimize the evaluation mechanism of the project leader and the allocation of resources, and improve the efficiency of project management. The model is verified by the example of national science and technology plan project, which proves its accuracy and effectiveness.

**Keywords:** the National Science and Technology Plan; data analysis; project leader; integrated performance evaluation; evaluation mechanism

[责任编辑 陈留院 赵晓华]