

摘 要

每年国家自然科学基金委员会要受理大量的申请项目,公平合理的评价这些申请项目对我国的科技发展有巨大的影响。这不仅涉及到资金的合理分配和优化利用等问题,也是科研工作者十分关心的问题。本文以国家自然科学基金为主要研究对象,讨论了优化理论、复合序列比较方法、归一化方法在基金评审中的应用。并在此基础上首次提出了关于评审专家反评估的模型及算法。

本文主要工作如下:

首先应用复合序列比较方法,建立关于上述方法的自然科学基金项目评审模型,讨论了有关性质,并最终给出基金项目评审的复合序列比较方法及应用结果。

根据国家自然科学基金经常出现难于处理的可上可下的划分问题。提出了非共识度、最大非共识度的概念,论述了它们的数学性质,计算方法及在某学科面上项目基金中的应用。并在此基础上,提出了消除项目之间非共识度差异的归一化方法。

根据以往国家自然科学基金选择评审专家的境况,创新性地提出了专家反评估的概念和方法,以及命中率、成功率等概念。

最后,将关于项目评审的归一化方法及专家反评估方法,结合定性和定量的方法,开发国家自然科学基金项目评审软件,并在国家自然科学基金委员会的某学科项目评审中得到实际应用。

关键字: 复合序列比较; 归一化分析; 专家反评估; 基金评审

Abstract

Every year, the National Natural Science Foundation of China (NSFC) accepts many application projects. Fair and reasonable evaluation to these projects had an important impact on the development of our country's technology and science which is not only involved in the problem of the reasonable distribution of the funds and optimal exploitability, but also one that the researchers care much. In this paper, we discuss the procedure of evaluating the projects with the application of optimization theory, multiple sequence algorithm and normalization algorithm. On the base of these, we present the mathematical models and algorithms for the anti-evaluation analysis of the specialists in NSFC.

The main work is as follows:

First, set up the evaluation model of NSFC with application of multiple sequence algorithm, discuss the mathematical properties related, and present the final application result as well.

Then present the concepts of non-commonity and maximum non-commonity according to the no-consensus problem during the project evaluation of NSFC, discuss some properties and methods of them, find the application in the project evaluation of a certain field which overcome the adverse effect of different non-commonity.

Furthermore, present the concept and algorithm of the specialist anti-evaluation innovatively, which refer to the status of specialists in NSFC now, also present the concepts of hit ratio and success ratio etc.

At last, associated with the qualitative and quantitative method, develop the software of the project evaluation and the specialist anti-evaluation by the above algorithms, then get the perfect result after being applied actually in some field project evaluation in NSFC.

Key words: foundation evaluation; multiple sequence alignment; normalization analysis; specialist anti-evaluation

1 综述

1.1 总论

系统评价是利用现代科学的方法,对系统的价值作出综合评价,是系统工程研究的重要领域。系统评价的主要对象是社会经济和技术系统,系统科学和系统分析方法分别是系统评价的理论基础和工具,系统评价的目的是为科学决策提出充分可靠的依据。在系统评价众多准则中,量化分析准则是很重要的一个研究分支。尤其是有关项目评审的内容,涉及到现实生活中各种评比、考核等方法的研究。

本文就是在系统科学理论的基础上讨论并改进原有的系统分析方法,使之易于量化评价系统运行机制,明确评价目标含义,能充分运用评价信息。故本文研究的问题具有充实的理论基础和重要的实践意义。

1.2 评审方法综述

1.2.1 群体决策与偏好分析

目前,应用于项目评审的方法如应用群体决策理论,引进决策个体和决策群体关于两两供选方案的偏比度概念及利用偏比度进行群体偏爱排序的方法,及基金项目评审的方法如层次分析法、属性分析法或区间分析法等。

关于群体决策的研究,早期的文献可以追溯到法国数学家 J.C.Borda 于 1781 年发表的关于选举制度的探索,1785 年 M.de Condorcet 对投票选举的研究和 1882 年 E.J.Nanson 关于投票悖论的讨论。1938 年, A.Bergson 引进社会福利函数和 1947 年 P.A.Samuelson 对社会福利函数的研究,使得群体决策的研究更具系统化和理论化。尤其是 1951 年, K.J.Arrow 关于偏爱公理和不可能性定理的群体偏爱理论的发表使群体决策真正成为一门科学。从这以后, A.Sen 等许多学者对 K.J.Arrow 的不可能性定理进行了探讨,并提出各种变化形式。同时,一些数学工作者也开始从事研究经济学家社会学家们所关心的社会选择理论,使这方面的研究进入一个更深的层次。

随着研究的深入进行, R.E.Wendell 和 H.W.brock 将群体决策和多目标决策问题相结合,形成了群体多目标决策这一研究方向,使得群体决策的研究内容不再局限于投票或选举问题,而且备选方案集也不再是局限于有限集。

近年来,一些学者已意识到决策个体和决策群体的偏爱结构也可能是模糊的,即

对于供选方案, 决策个体和决策群体并不能清楚表达对其中两个方案的偏爱强度, 从而发展了模糊群体决策理论的研究, 成为目前群体决策中极有前途的方向。1998年, 胡毓达引进了决策个体和决策群体关于两两供选方案间的偏差度和群体偏差映射等概念, 构建了群体决策的偏差度公理体系, 建立了群体决策偏差度分析的基本理论。他在《群体决策的偏差分析》^[58]中给出群体决策的偏差度和偏差泛函, 并利用偏差度对序列进行排列。此后, 又引进了群体协调性指标, 提出了求群体决策问题的新方法, 通过求评价矩阵得到最终得群体间的排序关系, 及最终的评审结果。这些结果主要应用于项目排序或偏好分析方面。程建刚等将胡毓达的理论基础具体应用在基金评审上, 发表了《基金评审中的异同性分析、专家反评估及软件研制》^[51], 《划分临界项目的一种多目标决策方法, 决策科学理论、方法与应用》^[52]等。

实践证明, 简单多数规则和 Borda 数规则是两种最常用、最基本的规则。国内学者将这两种方法分别推广, 并且在 α -较多序和 k -较多序的基础上提出了群体决策的广义较多规则。

1.2.2 其他方法

人们在管理决策过程中常常要面临一个制约因素众多的复杂情况, 需要人们对定性描述的问题进行定量处理、分析。层次分析法首先应用于解决这类无结构决策问题。层次分析法 (Analytic Hierarchy Process), 简称AHP法, 1977年由T.L.Saaty在测量模型基础上提出, 解决问题的关键是如何通过元素的两两比较确定元素的排序。1982年由H.Gholam nezhaed在北京的一次国际学术会议上介绍到我国, 立即在许多部门得到广泛应用, 从理论到应用不断完善^[55]。在这方面已经有王汝华1987的《关于群体AHP方法探讨》^[37], 王莲芬、许树柏1990年的《层次分析法引论》^[38]等著作。1997年北京大学数学学院程乾生教授在属性测度基础上首次提出了一种新的无结构决策方法——属性层次模型 (Attribute Hierarchical Model, 简称AHM)。AHM法在管理决策上中有着非常广阔的应用前景。国家自然科学基金委员会的项目资助方案的制定是一个典型的管理决策过程, 在项目评审中应用AHM法可以有效地帮助资助方案的合理确定。在这方面, 王其冬、程建刚、冯恩民2000年发表的《划分临界项目的一种多目标决策方法》^[52]中, 将层次分析法应用于自然基金的评审, 主要思想是将基金评审过程分为三大层次, 将评审的指标和语义比较程度量化, 逐层比较。在王其冬、武佩珍、程建

刚、冯恩民在2001年发表的《层次分析法在国家自然科学基金项目评审中的应用》^[52]等文章中，对层次分析法在基金评审中的应用进行了更深入的探讨。之后，又提出了将区间分析法、属性分析法在基金评审中的应用算法模型，这方面有王其冬、冯恩民、李文娟2002年发表的《项目评价双层规划、性质及应用》^[54]和《评分区间方法、性质及应用》^[53]。而程建刚在2001年的《基金评审中的异同性分析、专家反评估及软件研制》^[51]中，应用群体决策理论，提出了非共识度的概念，引进决策个体和决策群体关于两两供选方案的偏比度概念及利用偏比度进行群体偏爱排序的方法，得出了有关基金评审的群体决策方法。

此外，近些年来，中科院的方为武教授在序列比对方面也作了一些工作，提出了有关完全信息集的概念，定义FDOD函数，讨论其性质并将其应用于蛋白质序列的比较方面^{[14]-[17]}。但由于FDOD函数对于越长序列的比对效果越好，故在项目评审这类的相对序列元素个数较少的比较应用中需要对函数进行修改，具体的内容将在下一章提到。

虽然项目评审有很多方法，但在实际操作中，用于评审异同性的软件系统目前还不够完善。

1.2.3 基金评审

科学基金项目的资助方案是建立在项目评审结果的基础上。中国的科学基金制主要借鉴美国国家科学基金会的经验。但从一开始，中国就采取完全有别于美国的两级同行评议决策的机制。先由项目主任选择不少于3名（一般为5名）与申请者的研究方向相同、或最为接近专家以通讯方式对申请项目进行评议，习惯称为“小同行”。在此基础上，由所属学科的资深专家（习惯称为大同行）组成的评审组（一般由7-15人组成）开会审议申请项目及其同行通讯评议的意见，经过讨论，采取投票的方式决定资助的项目。

项目主任由于属于基金会的固定编制（2001年开始实行少量的流动制度），虽然到基金会之前都有一定的研究员，但其身份是管理人员。在整个的评审过程中，他可以在同行评议之后发表意见，但不属于同行意见，一般是表述为对同行评议的分析，或依据学科发展政策向专家评审组提出建议。由于申请项目数量大，会期有限，所以，项目主任只能拿出一部分项目提交专家评审组评审。按照规定和管理，提交评审组审

议的项目在拟资助项目的 130%到 200%之间。由此可见，小同行、项目主任和评审组专家在评审系统中扮演着不同的角色。项目主任的意见虽然不属于同行意见，但因为在选择同行评议专家和选择提交审议的项目的两个关键环节具体操作，所以隐性作用大。

采取这样三方两级的评审机制的具体过程不得而知，但是可以推定的是：中国特有的文化背景和科技体制决定现在中国式的科学基金制。两级同行评议制表现出中国等级森严、遵奉权威的儒家传统和科学界希望吸纳西方民主制度心理；而对项目主任职责的设定反映科学共同体在科学基金制中希望抑制管理人员的权力，防止官员滥用职权用。

同行评议可以有两种方式：一是文字描述，二是量化打分（定级）。对评议专家评价意见的综合方法也分为文字综述、量化平均和综合排序三种。由于语言的模糊性，对文字评价的综合评论，难以确定项目的优先顺序，资助方案也就难以落实，在专家意见有较大分歧时更为困难。量化平均虽有操作简便，计算方法模型成熟等优点，但因评议专家对评价标准掌握、理解上不同，打出的分数之间差别很大，平均后的量化分数有表达抽象、简单和处理结果掩盖了评议的实质内容的缺点。综合以排序表示的专家意见的方法更加不一，一般要与打分的量化方法结合起来使用。为互补不足，经常是将三种方法结合起来，相互参照综合使用，但这样又受到“多重标准”、“随意性大”的批评。

管理专家也会遇到这样的问题。他们不参与对具体项目的评议，通过制定评议标准和调整标准之间的权重，对资助方案进行宏观调控。在斟酌不同方面的权重的过程中，虽然与专业评议内容不同，但性质是一样的。

1.3 国内外基金评审综述

1.3.1 科学基金制

科学基金制是科学、经济和社会发展到一定阶段的产物。19 世纪中期以后，这种制度首先在西方国家发展起来。1937 年，英国科学协会和科学工作者协会联合委员会向英国政府建议设立“国家科学研究基金”以“保持连贯和适当扩大科学研究”，他们认为科学研究与实际应用要有一段距离，科学研究完全直接依附于经济对科学的发展会损害科学的发展。二十世纪中叶，由国家财政拨款的科学基金得到普遍的发展。

科学基金被看作是对未来的投资。二次世界大战以后，以美国、德国、英国为代表的发达国家经济得以高速发展，以科学基金资助的科学研究、特别是基础研究为技术发展提供源源不断动力是一个关键而重要的因素。

建立科学基金，因为没有期望在近期得到回报，所以一般来说，政府行政干预、社会和经济等因素的影响会相对来说减为最低。决策机构由科学家组成，科学界可以根据自身对科学发展的理解做出资助的战略决策，制定资金使用规则，决定资助什么样的研究和什么样的人。毫无疑问，资金应该给想出最好研究课题的优秀科学家。因此资金分配最公平的方式就是通过竞争争取，需要研究经费的科学家提出自己的研究思路，经过比较选出最好的研究课题给予资助。

由国家财政大量拨款支持用于基础研究的科学基金的出现，则是在 20 世纪以后。西方国家在分配这种科学资源时，多采用同行评议方法。由于每个国家历史、文化、经济和社会发展的差异，不同国家的科学基金制各具特色，从而使各个国家的科学基金制既有共同点，又有不同处。美国是科学基金制开展较早的国家，借鉴它的成功经验，有助于完善我国的科学基金制，改进同行评议方法。

1.3.2 同行评议

那么怎样选择最好的？依靠什么来评价？科学研究是人类的一种全然的创新活动，科学研究活动中的一个发现、或一个发明都必须是人类知识领域所没有涵盖的才有意义，最有价值的研究应该是对人类认识的突破。这给选择和评价带来了难题，因为选择和评价没有现成的依据和标准，只能根据已有的知识和想象力，进行判断和预测。那么显然，最可靠的判断来自从事同样工作的人员，也就是同行的判断。所以同行评议（Peer Review）是取得科学界公认的判断和预测依据，它也是科学界遴选有价值的研究思路的唯一评价方法。

同行评议方法最早的使用是在十五世纪的专利申请，至今为止已使用了 500 年了。300 年前，英国皇家学会的会员以同行评议作为一种参照系统评审可以公开发表的科学论文。1937 年，美国癌症咨询委员会采用同行评议方法评审申请研究经费的科研项目。在美国海军总署采用同行评议方法资助大学的研究取得比较成熟的模式之后，50 年代成立的美国国家科学基金会正式将同行评议作为分派科学研究经费资源的依据，并逐步发展了一套比较完善成熟的方法。虽然在如何使用同行评议的意见方

面，各个国家的科学基金组织的方式不尽相同，但是同行评议的重要地位是基本一致的，制定的一切方针和政策均已此为核心。

对于同行评议的定义有很多种。美国国会技术评估办公室高级分析专家库宾在其专著《无同行的科学：同行评议和美国的科学政策》这样定义同行评议：“同行评议是用于评价科学工作的一种组织方法，这种方法常常被科学界用来判断工作程序的正确性、确认结果的可靠性以及对有限资源的分配”。吴述尧先生在总结美英科技政策专家有关研究论述以及美国基金会和英国研究理事会关于同行评议的提法，提出针对科研项目评审的同行评议定义：“由科学系统中的同行专家组成的群体根据一定的标准，对提交的申请项目进行评议，这样的评议得出的结果可以作为有关部门确定资助项目（科学资源的分配）的重要依据……”。美国国家科学基金会在其许多文件中对同行评议有不同的描述和解释，其实“同行评议”一词就明确了其含义。对于科学基金来说的进一步解释是“同行”系指申请资助的科学家的同行，“评议”即对申请项目所提供的评价。针对同行评议的争论是在决定资助项目的过程中如何使用这种方法。

我国基金会同样将同行评议作为资助项目决定过程的重要环节。2002年发布的《国家自然科学基金项目管理规定》第十四条明确规定：“基金项目的评审一般按照初审、同行专家评议、专家评审组或专业委员会评审的程序进行”。在第十六条对同行专家评议作了进一步的解释：“同行评议是指同行评议专家对申请项目的创新性、研究价值、研究目标、研究方案等做出独立的判断和评价，一般采用通讯评议方式。”。

其实，同行评议本身是很易于理解的方法，就是从事相同或相近的领域的科学家根据自己掌握的知识和对研究发展的判断去评价其他科学家的研究。由于判断的主体是人，所以评议人的学识、见解和判断能力，甚至个人的性格、品性都会对评价有一定的影响。因此它是一把“双刃剑”，关键是在资助项目决定过程中的如何使用。科学基金制实行同行评议虽然得到科学界的普遍赞同，但对其指责也是一直不断的，指责的焦点并不在同行评议的本身，而在其如何使用，本文也就此方向进行了一些讨论。

对于同行评议的研究，吴述尧先生在其专著《同行评议方法论》^[33]中有全面而细致的研究。他认为：同行评议的实质就是决策在科学共同体中的民主化。所以能得到科学界的广泛认同。如果不涉及经费资源，仅在涉及科学界内部的荣誉方面，如论文发表、会员选举和名称授予等，同行评议可以看成仅是科学界内部的一种“游戏规则”。然而，研究经费是一种资源，并且科学基金的经费来自国家财政税收，所以在

科学基金中采取同行评议就必须考虑其社会功能和公共政策方面的意义。1978 年美国纽约州立大学石溪分校社会学教授史蒂芬·科尔、全国科学院顾问利昂纳德·鲁宾和哥伦比亚大学社会学教授乔纳森·科尔针对美国基金会的同行评议专门为全国科学院科学与公共政策委员会撰写调研报告《美国国家科学基金会的同行评议》，按照美国全国科学院院长菲利普·汉德勒在其序言中的话，其目的是“探讨了一个科学家和广大公众均集中关注的问题——如何判断哪些基础研究项目和研究人员应得到国会为此所拨资金？”。吴述尧先生对于同行评议的社会功能有三点概括：一、同行评议有利于科学资源的合理分配；二，同行评议可以保证科学荣誉的正确授予；三、同行评议可对科学共同体进行社会控制。同行评议有一些难以解决的问题，但至今仍是科学界公认的唯一方法，因为没有其他方法可以替代它。

因此，下节是对美国和我国的资助机构的同行评议进行比较研究。

1.3.3 美国国家基金会的项目评审研究

1945 年 7 月，美国科学研究发展局（OSRD）局长 Vannevar. Bush 向总统提交《科学无止境的领域》的报告，倡议政府增加资助基础研究的力度。他认为“基础研究将导致新的知识，提供科学上的资本”。因此他建议成立国家级的科学基金会，资助以高等院校为主的的学术研究机构的基础研究。但因为当时美国科学界的“研究自由”意愿与联邦政府的计划性研究相矛盾，所以没有采纳他的建议。经过 5 年的的争论，1950 年杜鲁门总统签署了 S. 247 号提案，成为第 81 届国会的 507 号公法。根据这一法令正式成立国家科学基金会（NSF）。

美国国家基金会（简称 NSF），它本身不开展研究工作。其任务是：通过对基础研究项目的资助，促进科学教育，发展科技情报和增进国际科学合作等方式，来促进美国科学的发展。此外，国际科学基金会还参与国家科学技术政策的制定。

国家科学理事会（简称 NSB）是 NSF 的最高决策和领导机构。理事会的主要任务是确定国家科学基金会的政策，协助政府制定全国科学技术的主要政策，并负责对全国的科学研究状况进行研究。除基金会主任作为该理事会的当然理事外，其他所有理事会成员均需总统任命，并得到参议院的批准。

NSF 设有七个局（包括五个科学局和两个业务局），以及在秘书长领导下的七个室，负责例行行政业务和各科学的项目管理工作。除了上述组织机构外，NSF 还设立

了政策咨询委员会和管理咨询委员会，前者主要是针对 NSF 的发展和资助政策、计划和关键问题进行讨论，提出建议；后者主要是针对 NSF 管理业务中遇到的问题以及提出改进的方案，解决基金的内部科学管理问题。

NSF 主要支持高等院校中一般性的基础研究，这一点同美国能源部、国防部、宇航局等部门的机构有所区别，因为部门支持的科学研究并不只是为了科学本身，而是为了部门自身的任务需要。NSF 在项目管理方面的指导思想是尊重学术自由，鼓励优秀科学家的创新精神和求知愿望。除一部分大型研究计划有基金会组织协调并进行管理外，多数研究项目采取被动的方式，即依靠科学家和研究单位的自由申请。

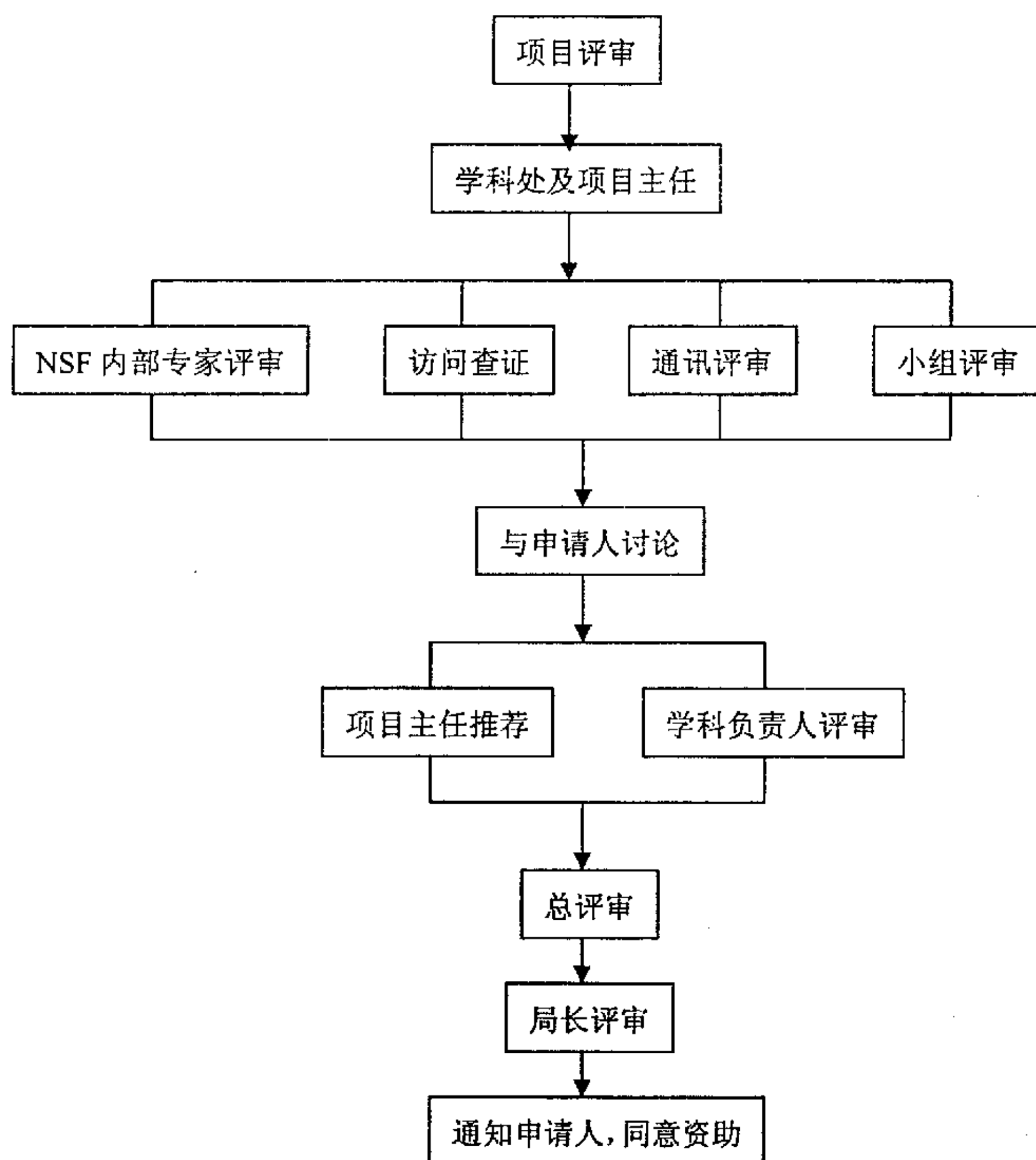


图 1.1 NSF 的项目评审和选择过程

Fig 1.1 NSF Project Evaluation and Selection Procedure

在美国国家科学基金会，由于相当一部分的项目主任是流动的，本身就是资深的科学家同行，所以在评审过程中起关键作用，来自外面的同行评议是参考性意见，最终决定权在项目主任。对于“高风险”的项目，项目主任甚至可以根据自己的判断决定是否资助，而不必经过同行评议。所以对其同行评议制度的指责意见是：容易以项目主任为核心在同行里形成一个不道德的任人唯亲的“圈子”，常会扼杀来自这个“圈子”之外的新思想。

美国科学基金会规定，除少数特殊情况外，申请项目一定要通过评审，评审前所有的申请必须经研究人员所在单位同意。评审一般采用通讯评审和小组评审两种方式进行。一些规模比较大的申请项目还需要组织基金会人员和一些科学家到研究单位进行考察。

与我国科学基金会不同的是，在 NSF 项目评审过程中，项目主任起着很大的作用，用什么方式进行评审，选择哪些专家参加同行评议，以及资助方案的决定，得到资助后资助金额的多少，都由项目主任负责。申请项目首先必须得到项目主任的首肯，在特别的情况下，如有充分的理由，项目主任还可以不采纳同行评议的意见。当然，项目主任做出决定后，还需要通过好几个审查步骤才能做出最后资助决定。如果项目规模不是很大，只须通过学科组即可。如果项目规模很大，还需要提交到助理主任一级来决定。上页图 1.1 是 NSF 的申请项目过程。

1.3.4 我国国家基金会的项目评审研究

1981 年，中国科学院 89 位学部委员（即现称为中国科学院院士）联名建议设立自然科学基金。1982 年，中央批准从 1982 年起设立中国科学院科学基金，面向全国科研人员。此后，地震、教育、卫生、邮电等部门成立了相应的学科或行业科技基金，有些省市地方成立地方科学基金。1985 年 3 月，中共中央关于科学技术体制改革的决定出台，明确指出：对基础研究和部分应用研究逐步试行科学基金制，主要从事基础研究和应用基础研究的机构，应力争几年之后做到科研经费主要靠申请基金。以往研究经费依靠行政拨款的方式改为通过竞争并依据同行评议分派的方式。1985 年 7 月针对建议成立独立的国家自然科学基金委员会的建议，邓小平作出重要指示：“成立国家自然科学基金委员会，大家都会赞成，不会反对。应该这样办。这是个新事物，办起来再说，取得经验。”1986 年 2 月，根据国务院发出的通知，国家自然科

学基金委员会正式成立。

科学家参与并且起着主导作用是我国科学基金管理体系的重要特点。咨询系统由国家自然科学基金全体委员会（以下简称全委会）、七个科学部专家咨询委员会（以下简称咨询委员会）和若干专家评审组组成，主要针对科学基金的发展战略提出建议，对资助项目的评审和立项给出评价意见。决策系统主要由基金委的正、副委主任和秘书长等组成，一般以委务会议的形式通过各种资助政策和决议。执行系统以基金委的常设部门机关为主，包括六个职能局室和七个科学部，其中，委办公室、计划局、政策局、财务局、人事局和国际合作局等六个职能局室按照所确定的职责负责各项决策的落实和贯彻，数学物理、化学、生命、地球、材料与工程、信息和管理科学部的主要职责是联络相应领域的科学家，组织项目的申请、评审和管理。科学基金项目依托单位的科研管理部门是执行系统的外延。监督系统由国家自然科学基金监督委员会负责，在基金委设立一个办公室（现称为纪检、监督、审计联合办公室）作为其事务机构。

在这个管理体系中，合理的资助分布格局是关键，公正的同行专家评议是核心，严谨的项目管理程序是保障。整个体系是否运作的良好、有效取决于这三个重要工作环节。

伴随着中国 20 年来的改革开放，中国的科学基金制从无到有，从试验走向成熟，逐步壮大，为推动中国的基础科学研究的发展发挥了巨大的作用。基金会成立当年的资助经费仅为 8000 万元，到 2003 年已达到 22 亿元。资助项目起到了对“未来投资”的种子作用，取得的成果不仅提升了中国科学研究方面的国际地位，而且大力推动了中国的经济和社会的发展，同时培养了一批高素质的科学研究人才和队伍。基金会以同行评议为核心方法，在实践中逐步地建立起一套比较完善的科学基金管理政策和办法体系，科学基金制从试验的阶段达到全面发展阶段。目前科学基金法的提案已到酝酿成熟阶段，近期准备提交全国人大。

国家自然科学基金会根据我国基础性研究的现状，从国情出发，把科学基金资助项目分成面上项目、重点项目、重大项目三个层次，分别给予不同程度的资助。其中面上项目是基于基础性研究、探索性强、不确定性大的特点而设立的，其目的在于为广大科技人员创造充分自由探索的条件，是科学基金项目中最主要的、基本的类型。面上项目规模较小，资助面广，几乎覆盖自然科学领域所有分支学科。

重点项目主要是针对我国科学发展和布局中的关键科学问题和学科发展的需要，

选择某些重要的学科前沿或领先的领域进行深入研究。重点项目学科性强，对各个分支学科深入的发展有重要作用。重点项目和面上项目联系密切，很多重点项目就是建筑在面上项目多年探索的基础上。同样，重点项目的深化和扩展，又可为重大项目提供很好的来源。

重大项目是以服务于我国社会主义建设总目标，针对经济、科技、社会发展中的重大学科问题，组织多学科、跨部门、跨单位的联合研究。重大项目体现了基础性研究对组织规模、集约形式的需要，以及自然科学高度综合而带来的多学科联合研究的需要，对一些待取得重大突破、投资多、周期长、任务性强、需要综合多个学科的研究，组织重大项目是一种行之有效的形式。它对边缘学科、综合性学科的发展也有较大的推动作用。

国家自然科学基金委员会三类资助项目的评审方法虽然具体做法不同，但根本上都是采用同行评议的方法，面上项目约占各类资助项目经费的 70%。因此本文主要研究面上项目的同行评议。

国家自然科学基金委员会对申请科学基金项目评审的原则是：依靠专家、发扬民主、择优支持、公平合理。

国家自然科学基金资助自然科学中的基础研究和应用基础研究。自组的学科范围覆盖了数学、物理学、化学、生命科学、地球科学、材料与工程科学、信息科学和管理科学等自然科学与工程技术科学的所有领域。因此，国家自然科学基金委员会只受理符合上述资助范围的申请项目。国家自然科学基金委员对面上项目评审的基本程序如下图 1.2。

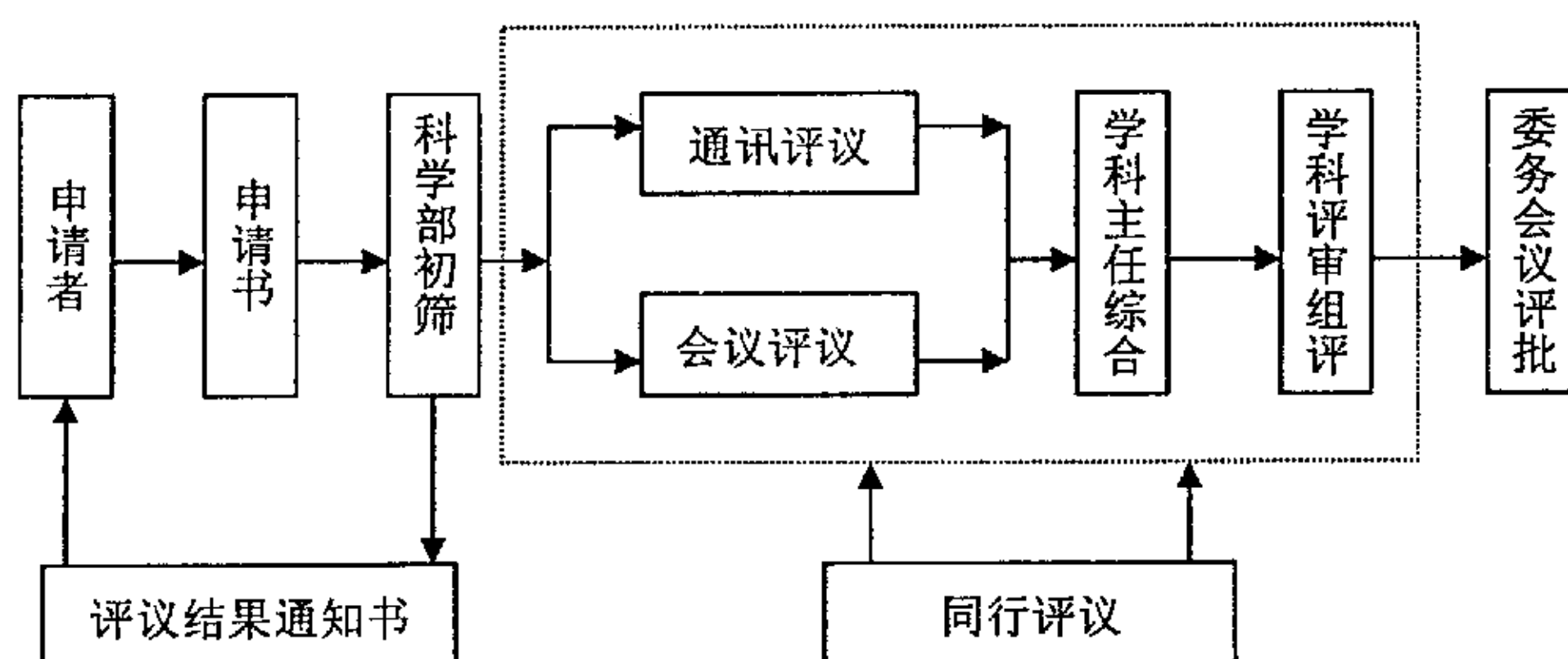


图 1.2 我国面上项目评审的基本程序

Fig 1.2 The Fundamental Procedure of Project Evaluation in China

1.4 面临的“非共识”问题

虽然我国的基金制一直得到科学界的普遍赞誉,被科技界公认为值得资助的研究项目和研究人员获得资助无疑比通过行政指令的方式更能得到科技界的欢迎。然而另一方面的问题逐渐显现出来,通过这个系统能够获得资助的项目一般须取得评审各方的同意,如果任何一方存有异议,就有被否定的可能。况且,两级同行评议的意见取决于若干个人,其中任何一人如果挑出项目存在的问题,那么不被通过的危险就很大,所以获准通过的极有可能是设计方案无错可挑、但缺乏创新思想的平庸保守项目,具有创新性的项目,往往因为其中的瑕疵而被淘汰掉。其关键的原因就在于对评审中非共识项目的处理。

在评审系统中取得共识赞同的项目获得资助是无疑的,其中大部分项目也是值得资助的,虽然有些项目的创新性不足,但对于促进科学研究的发展和通过研究项目培养训练年轻的科学家也是有价值的。如果对申请项目的意见一致是否定的,那么获得资助的可能性为零。即便可能是有创新、有价值的项目也无法得知。科学发展史表明,完全不被当时科学共同体承认的创新思想是超越时代的个例,同时代的社会无法负责。这里主要考虑的是,在科学界内引起争议的思想是否具有推动科学飞跃性发展价值?这样具有创新思想的研究如何通过一个良好的机制获得鼓励性支持。这就是我们研究科学基金评审系统中“非共识”现象和并对之采取科学态度的意义。

“小同行”的同行评议是项目评审的基础,在其中引起的“非共识”直接引起项目评审的后期结果。所以,我们直接分析其“非共识”现象。其中对项目的评议有定性和定量的两个方面,理论上说,定量和定性的评议应当相称,当然也有很多个例表明,某些评议人的定性和定量相互矛盾,这一问题一直困扰着基金委的项目主任。

所以非共识的量化问题是如今基金评审过程中亟待解决的关键问题,也是本文主要讨论的方向之一。

1.5 本文主要工作

每年国家自然科学基金委员会要受理大量的申请项目,公平合理的评价这些申请项目对我国的科技发展有巨大的影响。这涉及到资金的合理分配和优化利用的问题,也是科研工作者十分关心的问题。本文以国家自然科学基金为主要研究对象,应用优化理论、归一化方法、引入完全信息集理论,讨论了上述方法在基金评审中的应用。

并首次提出了关于评审专家的反评估模型及方法。主要工作如下：

本文在上述方法的基础上，引入完全信息集的概念，应用复合序列比较的方法，提出关于基金评审的复合序列比较方法；

对于由于项目非共识度分布不一致，从而影响评审尺度的问题，应用优化理论、归一化方法等，建立确定平均非共识度和最大非共识度的优化模型与算法，克服了因最大非共识度的不同影响项目评审公正性的弊端；

最后，提出专家评审的平均差异和绝对差异、命中率、成功率等概念，得到专家反评估的量化指标。并应用优化理论建立了专家反评估的数学模型和计算方法。

归一化方法和专家反评估在基金委的实际项目评审中得到采用，实际应用效果较为理想。

2 完全信息集方法

2.1 完全信息集方法

2.1.1 背景

复合序列的比对是数学及其它学科研究的基础和热点问题,它的应用背景也是非常广泛的。随着数学研究与其他学科相结合程度的加深,尤其是生物数学和金融数学等学科的发展,复合数列比较的得到了新的应用,也要求其在现有的理论结果上有更深入的研究。例如现在生物学比较热门的基因研究中,破译 DNA 片断是一个非常重要的科研方向。一般来说,比较的 DNA 链长度至少是 6000~8000 或更多,而且 DNA 链的比较不是单纯的单个基因的比较,而是某个基因片段的比较,这里要涉及到组合学和优化等多方面的知识,如果没有好的理论前提做依据,没有一个好的比较模型和算法,不仅计算机实现起来困难,更影响最终的研究结果。随着科学技术的不断进步,研究者们纷纷致力于改进现有的优化算法,或是另辟蹊径创立全新的理论分支,但其目的都是在实际中得到更好的应用。同样地,复合序列的比较在国家自然科学基金的项目评审中的应用更为重要。每年国家自然科学基金委员会要受理大量的申请项目,公平合理地评价这些项目对我国的科技发展有巨大的影响。项目评审涉及到资金的合理分配和优化利用的问题,同时也是科技工作者十分关心的问题。当然,复合序列的比对问题在医学、计算机等领域也有很广泛的应用。

目前国内外有关比对问题的比较成熟的理论有:

一、 比对方法 (Alignment approaches)。这种方法提出的较早,1972 年就有了较成熟的方法,但此种方法有相当大的不适宜性,如今已应用的很少;

二、 遍历与统计法 (Heuristic approaches and statistical methods)。此法被广泛应用到 DNA 测序中,是应用最多的方法之一。但遍历与统计法包含了某些主观因素,如选择最优准则,如何打分等,而“测量序列中的匹配、不匹配”这种方法往往不可能用于比对方法,而且常常要求比较的序列是固定不变且可遍历的,但在实际中这往往是不能实现的。

三、 信息理论法 (Information theory approaches)。是 Gatlin 于 1978 年提出的,稍晚于比对方法。虽然此法依靠序列的原始信息而不包含主观因素,但许多信息理论在基因测序研究中都不成立或应用很少,某些研究结果甚至否定了原有的信息理论结

果。

以下有几个基本信息学理论着重应用于测序：

a. 条件熵方法 (Conditional entropy approaches)；

在本算法中，将比对信息分为两类：潜在信息 (potential information) 和存储信息 (stored information)。定义整体的偏差 D_m ，使

$$D_m = H_{\max} - H_m$$

这里 $H_m = -N \sum_i \sum_j \sum_k \cdots \sum_m p_i p_{ij} p_{ijk} \cdots p_{ijk \cdots m} \log p_{ijk \cdots m}$ ， H_{\max} 表示不确定性的极大值， $p_{ij \cdots}$ 为条件概率。

b. 熵估计法 (The entropy estimation approaches)；

本文主要介绍 Shannon 熵和 Kullback-Leiber 熵。其中，Shannon 熵的定义如下：令

$$\Gamma_t := \left\{ (p_1, p_2, \dots, p_t)^T \mid \sum_{i=1}^t p_i = 1 \text{ 且 } p_i \geq 0 \right\}, (t = 2, 3, \dots), P := (p_1, p_2, \dots, p_t)$$

Γ_t 为一分布，则 Shannon 熵定义为

$$H(P) = -\sum_{i=1}^t p_i \log p_i, P \in \Gamma_t$$

但 Shannon 熵具有不确定性，只能用于一个分布的比对。而 Kullback-Leiber 熵可以用于两个分布的比对。Kullback-Leiber 熵定义为

$$H(P, Q) = \sum p_i \log \frac{q_i}{p_i}, P, Q \in \Gamma_t$$

其中 $P := (p_1, p_2, \dots, p_t)$ ， $Q := (q_1, q_2, \dots, q_t)$ 分别为两个分布。

对于比较一组序列的偏差来说，上述两种方法都不能很好的实现。而且对于例如基因序列的比较来说，基因可看成一个随机序列，故很难用一个简单的熵估计法把他们区分开。因此，有了以下的方法。

c. 完全信息集法 (Complete information set method)；

这种方法有许多以前的熵方法没有的好性质，在基因测序上的应用效果很好，但目前在其他领域没有效果较好的应用。

下面是完全信息集函数即 FDOD 函数与 Shannon 熵和 Kullback-Leiber 熵一些基本

性质的比较（见下表 2.1）

基本性质	Shannon 熵	Kullback-Leiber 熵	FDOD 函数
概念基本含义	不确定性, 无序性, ... 用于一个分布的偏差	用于两个分布的偏差比较	用于一组分布的偏差比较
数据量	1	2	$s \geq 2$
非负性	yes	yes	yes
一致性		yes	Yes
对称性	yes	no	Yes
有界性	yes	no	yes
一致连续	yes	no	yes
最大值	$\ln t$	∞	$s \ln s, (s \ln t)$
凸性或凹性	凹	凸	凸
单调性		no	yes
平均值恒定性		no	yes
线性奇次性		yes	yes
单调分布的有效性		no	yes

图 2.1 Shannon 熵、K-L measures 和 FDOD 函数的某些性质比较

Fig 2.1 A simple comparison with Shannon and K-L measures

下个章节将引入完全信息集，为了建立比较序列异同性的合理方法，需要引入完全信息集的概念和比较准则。并将完全信息集方法改进后应用基金项目评审中。

2.1.2 一个序列的完全信息集

设 $\Sigma = \{ a_1, a_2, \dots, a_m \}$ 为由 m 个不同符号构成的集合，令 $S = \{ S_1, S_2, \dots, S_s \}$ 为由符号集 Σ 中的元素组成的序列的集合。我们用 Θ^l 表示由集合 Σ 中的元素组成的长度为 l 的不同序列的集合，则 Θ^l 中所有不同序列的个数 $m(l) = m^l$ 。对任意 $S_k \in S$ ， L_k 表示 S_k 的长度，而 n_{ik}^l 表示与 Θ^l 中第 i 个序列相匹配的 S_k 中的相邻子列的个数， $l \leq L_k$ 。则显然当 $l \leq L_k$ 时，对所有 k 有，

$$\sum_{i=1}^{m(l)} n_{ik}^l = L_k - l + 1$$

令 $p_{ik}^l = n_{ik}^l / (L_k - l + 1)$ ，这样，对任意 k 和 $l \leq L_k$ ，就可以得到一个分布，

$$U_k^l := (p_{1k}^l, p_{2k}^l, \dots, p_{m(l)k}^l)^T, \quad \text{这里 } \sum_{i=1}^{m(l)} p_{ik}^l = 1.$$

令 $\Gamma^l := \{ (p_{1k}^l, p_{2k}^l, \dots, p_{m(l)k}^l)^T \mid \sum_{i=1}^{m(l)} p_{ik}^l = 1, p_{ik}^l \geq 0 \}, (l = 2, 3, \dots)$.

这样，对每个序列 S_k ，表示它分布的集合是唯一的，即

$$U_k = (U_k^1, U_k^2, \dots, U_k^{L_k}), \quad \text{其中 } U_k^1 \in \Gamma^1, U_k^2 \in \Gamma^2, \dots, U_k^{L_k} \in \Gamma^{L_k}.$$

上面这个集合包含了一个序列的所有结构信息，特别地，由 $U_k^{L_k}$ 能够唯一地确定这个序列，故我们称这个集合 U_k 为序列 S_k 的完全信息集。序列不相同，它们的完全信息集也不同。完全信息集间的偏差比较也意味着这些序列的所有结构信息的偏差比较。至此，我们可构造一种用复合序列的完全信息来度量复合序列本身偏差的方法。

为了更好地应用完全信息集方法，下一节将给出测量方法应满足的条件和衡量方法有效性的基本准则。

2.1.3 基于复合序列比较的准则

由 2.1.1 节可知，完全信息集包含了一个序列的所有结构信息，所以完全信息集的分布的比较可以用于序列间的偏差比较。这样，序列完全信息集的分布之间的偏差的一种度量，可看成不同序列之间的偏差的一种度量，而序列 S_k 中一个分布 U_k^l 是通过 S_k 中的长为 l 的子列与 Θ^l 中序列相比较得到的。所以，多个序列之间偏差的度量也是用这种方法进行比较，即以同样维数 $m(l)$ 的分布的比较来定义。我们主要研究 s 个序列的分布 $U_1^l, U_2^l, \dots, U_s^l$ 的偏差。

令 $D(U_1^l, U_2^l, \dots, U_s^l)$ 表示关于长为 l 的定义在分布集合 $\{(U_1^l, U_2^l, \dots, U_s^l) \mid U_k^l \in \Gamma^l, k = 1, 2, \dots, s\}$ 上的一种偏差测度。以下将列出偏差度量的一些条件和准则。

条件 A 对所有 $U_k^l \in \Gamma^l$ ， $D(U_1^l, U_2^l, \dots, U_s^l)$ 为实值、连续有界，且是非负的。

条件 B $D(U_1^l, U_2^l, \dots, U_s^l)$ 关于 K 或 L 的任意置换都是对称的，这里 $K = \{1, 2, \dots, s\}$ ， $L = \{1, 2, \dots, m(l)\}$ 。 K 的置换对应于 S 中的一个重新排序， L 的置换对应于 Θ^l 中所有序列的一个重新排序。

条件 C $D(U_1^l, U_2^l, \dots, U_s^l) = 0$ 当且仅当所有的 $U_1^l, U_2^l, \dots, U_s^l$ 彼此恒等。

在以上的序列比较的基本结果基础上, 我们还有完全信息集的关于序列比较的性质。

条件 D (增长性) 对任一组分布 $U_1^l, U_2^l, \dots, U_s^l$, $D(U_1^l, U_2^l, \dots, U_s^l)$ 是非减的, 即若 $l \leq n \leq \min_{k \in K} L_k$, $D(U_1^l, U_2^l, \dots, U_s^l) \leq D(U_1^n, U_2^n, \dots, U_s^n)$ 。

我们知道, 如果在—组序列中不存在任一个长为 l 的相同子列, 则这组序列中不存在任一个长度为 $n > l$ 的相同子列。这种情况下, 对长为 n 和长为 l 的偏差测量是相同的, 即下面提到的遗传性:

条件 E (遗传性) 若 s 个序列中不存在任一长度为 l 的相同子列, 则对 $n \geq l$, 有 $D(U_1^{n+1}, U_2^{n+1}, \dots, U_s^{n+1}) = D(U_1^n, U_2^n, \dots, U_s^n)$

由此可知在条件 E 的情况下比较序列时, 没必要比较完全信息集中的所有元素, 这就为计算机实现比较过程提供了理论前提。根据 $D(U_1^l, U_2^l, \dots, U_s^l)$ 的增长性和有界性, 我们定义偏离率:

$$R(U_1^l, U_2^l, \dots, U_s^l) := D(U_1^l, U_2^l, \dots, U_s^l) / \text{Max}D$$

这里, $\text{Max}D$ 是 s 个序列的最大偏差, 即

$$\text{Max}D := \text{Max} \left\{ \text{Max}D^l \mid l \leq \min_{k \in K} L_k \right\},$$

$$\text{Max}D^l := \text{Max} \left\{ D(U_1^l, U_2^l, \dots, U_s^l) \mid U_k^l \in \Gamma^l, k = 1, 2, \dots, s \right\}$$

显然 $R \leq 1$, 自然地, 若序列关于长为 l 的子列不相同得越多, $R(U_1^l, U_2^l, \dots, U_s^l)$ 的值就越快地等于或趋于 1。

第三个完全信息集的特性称为“不同随机序列的指数发散量特性”。因为一个随机序列含有长为 l 的指定子列的概率应小于 $(L_k - l + 1) / m^l$ (当 $L_k - l + 1 \leq m^l$ 时), 这说明不同序列含有共同子列的个数随着子列长度 l 的增大而按指数减小。在实际应用中这个性质可以作为衡量算法有效性的标准。则我们有下面的有效性准则。

有效性准则 对于不同的随机序列, $R(U_1^l, U_2^l, \dots, U_s^l)$ 的值随着 l 的增大而很快地趋于 1。

2.1.4 基于信息偏差的一个新方法

正如我们在 2.1.1 节提到的, 完全信息集包含一个序列的所有结构信息, 所以满足 2.1.2 节所给性质的偏差分析方法可以用于复合数列之间的偏差比较。尤其如本文

所述，应用到国家自然科学基金的项目评审上。

下面，我们将介绍一个比较复合序列偏差的函数，这个函数具有许多好的性质，如非负性、唯一性、对称性、有界性、一致连续性、单调性、齐线性、存在最值、凸性、均值不变性、递归性等等。而且，这个函数满足 2.1.2 节所给出的所有条件。则函数 B 如下定义。

假设给定 $G(t, s, r)$ 是由 s 个 $t \times r$ 阶矩阵组成的， u_{ik} 是第 k 个矩阵的 i 行 j 列元素且 $\sum_{i=1}^t \sum_{j=1}^r u_{ij} = 1, k = 1, 2, \dots, s$ 。进而， $G(t, s, r)$ 称为有对称结构的矩阵群当且仅当

$s \equiv 0 \pmod{t}$ 或 $n \neq 0$ 但 $n \cdot r \equiv 0 \pmod{t}$ 或 $n \cdot r \equiv 0 \pmod{s}$ 。当 $r = 1$ ，由 $G(t, s, r)$ 得到一组分布，即 $G(t, s, r) := (U_1, U_2, \dots, U_s)$ ，这里 $U_k := (u_{1k}, u_{2k}, \dots, u_{tk})^T$ 且 $u_{1k} + u_{2k} + \dots + u_{tk} = 1$ 。则函数 $B(t, s, r)$ 有如下定义：

$$B(U_1, U_2, \dots, U_s) = \sum_{k=1}^s \sum_{i=1}^t u_{ik} \log \frac{u_{ik}}{\sum_{k=1}^s u_{ik} / s}$$

这里，还要定义 $0 \cdot \log \frac{0}{0} = 0$ 。则有下列引理。

引理 1 对于给定的 t, s, r 及 $G(t, s, r)$ ，则 $B(U_1, U_2, \dots, U_s)$ 的上界为 $s \cdot \log t$ 。

引理 2 对任意的 t, s, r 及 $G(t, s, r)$ ， $B(t, s, r)$ 的最大值为：

- 1) $s \cdot \log s$ ，当 $s \leq t$ ；
- 2) $s \cdot \log t$ ，当 $s > t$ 且 $G(t, s, r)$ 是具有对称结构的矩阵群。

如果， $U_1^l, U_2^l, \dots, U_s^l$ 的完全信息集的偏差定义为：

$$B(U_1^l, U_2^l, \dots, U_s^l) = \sum_{k=1}^s \sum_{i=1}^{m(l)} p_{ik}^l \log \frac{p_{ik}^l}{\sum_{k=1}^s p_{ik}^l / s}, \text{ 这里 } U_k^l := (p_{1k}^l, p_{2k}^l, \dots, p_{m(l)k}^l) \in \Gamma^l,$$

则有下面的定理成立。

定理 1 当 $s \leq m(l)$ 时， $B(U_1^l, U_2^l, \dots, U_s^l)$ 的最大值为 $s \cdot \log s$ ；而当 $s > m(l)$ 时，其上界为 $s \cdot \log m(l)$ 。

通过引理 1 和引理 2 可以直接得到定理的结论。

定理 2 $B(U_1^l, U_2^l, \dots, U_s^l)$ 满足条件 A, B, C, D 和 E。

对于这个衡量序列偏差的新方法， s 个序列的最大偏离度 $MaxB$ 为：

$$MaxB := \text{Max} \left\{ \text{Max} B^l \mid l \leq \min_{k \in K} L_k \right\}, \text{ 其中}$$

$$Max B^l := \text{Max} \left\{ B(U_1^l, \dots, U_s^l) \mid U_k^l \in \Gamma^l, k = 1, 2, \dots, s \right\}$$

从定理 1 易知, 对于每个 $s \leq m(l)$ 及 $s > m(l)$, $MaxB \leq s \cdot \log s$, 故有以下推论:

推论 1 对于 s 个序列组成的群 s , 它的最大偏差 $MaxB$ 小于或等于 $s \cdot \log s$ 。

所以, 由推论 1 和第二节的内容, 我们有

$$\begin{aligned} R(U'_1, \dots, U'_s) &= \frac{B(U'_1, \dots, U'_s)}{s \cdot \log s} \\ &= \frac{\sum_{k=1}^s \sum_{i=1}^{m(l)} p'_{ik} \log(p'_{ik} / \sum_{k=1}^s p'_{ik} / s)}{s \cdot \log s} \leq 1 \end{aligned}$$

由定理 2, 我们知道如果复合序列的 $B(U'_1, U'_2, \dots, U'_s)$ 或 $R(U'_1, U'_2, \dots, U'_s)$ 的值对于某个 l 不为 0, 则这些复合序列就不相同。随 l 的增加, 只需重复比较 $B(U'_1, U'_2, \dots, U'_s)$ 或 $R(U'_1, U'_2, \dots, U'_s)$ 的值。并且从定理 2 我们还可以知道, 一个随机序列 $R(U'_1, U'_2, \dots, U'_s)$ 的值会随 l 的增大很快地趋于 1。这个性质在计算机实现过程中是十分有效的。

2.2 评审模型及算法

设某学科的面上基金申请项目数为 t , 每个项目请 n 位专家评审, 有 m 个定量评价指标, 专家们根据每个评价指标逐项给出评分, 则对第 k 个项目可构成 $m \times n$ 阶指标评分矩阵:

$$X^k = (x_{ij}^k)_{m \times n} \in R_+^{m \times n}$$

其中 x_{ij}^k 表示评审专家 j 依评价指标 i 对第 k 个项目的评分。为了应用完全信息集, 且考虑到项目间的评分具有可比性, 设评分可分为 A, B, C, D , 为便于各项目之间评分序列比较分值大小, 再增加字母 E , 则字母集为 $\Sigma_1 = \{A, B, C, D\}$ 及 $\Sigma = \Sigma_1 \cup \{E\} = \{A, B, C, D, E\}$ 。这样 $x_{ij}^k \in \Sigma_1$, 另外, 令元素 A, B, C, D, E 对应的评分为 $5, 4, 3, 2, 1$, 即 $x_{ij}^k \in I_5 = \{1, 2, 3, 4, 5\}$ 。我们将每个专家的评分重新排列如下: 设第 j

位专家对项目 k 的总评分为 $x_j^k = \sum_{i=1}^m x_{ij}^k, j = 1, 2, \dots, n$, 对第 k 个项目, 各专家的总分

为 $x_1^k, x_2^k, \dots, x_n^k$, 从大到小排序。将排序结果记为 $y_j^k, j = 1, 2, \dots, n$, 即 $y_1^k \geq y_2^k \geq y_3^k \geq \dots \geq y_n^k$ 。而 y_j^k 表示第 j 位专家依评价指标 i 对项目 k 的评分。这样可得由 Σ_1 中元素形成的项目 k 的信息序列为:

$$y^k = \{y_{11}^k, y_{21}^k, \dots, y_{m1}^k, y_{21}^k, y_{22}^k, \dots, y_{m2}^k, \dots, y_{1n}^k, y_{2n}^k, \dots, y_{mn}^k\}, k = 1, 2, \dots, t$$

为能使序列 y^k 中的元素(字母)体现对应的评分值,作如下变化,变换 $f: \Sigma_1 \mapsto \Sigma$ 为:

$$f(A) = AAAA, f(B) = BBBE, f(C) = CCEE, f(D) = DEEE,$$

并把经上述变换后的序列记为 z^k , 即 $f(y^k) = z^k, k \in I_t = \{1, 2, \dots, t\}$, 且 $|y^k| = mn, |z^k| = 4mn$ 。令 $L = 4mn, I_L = \{1, 2, \dots, L\}, \forall l \in I_L$, 定义

$$\Theta^l := \{s^i \mid s^i \text{ 为 } \Sigma \text{ 中形成的长为 } l \text{ 的序列, 即 } |s^i| = l\}$$

把 Θ^l 中不同序列的个数记为 $m(l)$, 则 $m(l) = 5^l, \Theta^l = \{s_1^l, s_2^l, \dots, s_{m(l)}^l\}$ 。令 $I_{m(l)} = \{1, 2, \dots, m(l)\}, \forall l \in I_L, \forall k \in I_t, \forall i \in I_{m(l)}$, 定义

$$z_{i,j}^k = \{z_{i,j}^k \mid z_{i,j}^k \text{ 为 } z^k \text{ 中与 } \Theta^l \text{ 中的第 } i \text{ 个序列 } s_i^l \text{ 一致的子列}\}$$

把 $z_{i,j}^k$ 中不同子列的个数 $|z_{i,j}^k|$ 记为 $n_{i,j}^k$, 即 $z_{i,j}^k = \{z_{i,j,1}^k, z_{i,j,2}^k, \dots, z_{i,j,q}^k\}$, 其中 $q = n_{i,j}^k$ 。从

上述定义可得: $\forall l \in I_L, \forall k \in I_t$, 有 $\sum_{i=1}^{m(l)} n_{i,j}^k = L - l + 1$ 。令 $p_{i,j}^k = n_{i,j}^k / (L - l + 1), \forall i \in I_{m(l)}$,

得到项目 k 对应的序列 z^k 与 Θ^l 中序列一致的子列的分布为:

$$z_l^k := \{p_{i,j}^k, i \in I_{m(l)}\}, \text{ 其中 } \sum_{i=1}^{m(l)} p_{i,j}^k = 1, l \in I_L, \Gamma^l := \left\{ z_l^k \mid \sum_{i=1}^{m(l)} p_{i,j}^k = 1, p_{i,j}^k \geq 0 \right\},$$

项目 k 对应序列 z^k 的分布集为:

$$P^k = \{z_1^k, z_2^k, \dots, z_L^k\}, k \in I_t, \text{ 其中, } z_l^k \in \Gamma^l, \forall l \in I_L.$$

这样可以计算 t 个项目同样维数的偏差:

$$D(z_1^1, z_1^2, \dots, z_1^t) = \sum_{k=1}^t \sum_{i=1}^{m(l)} p_{i,j}^k \log(p_{i,j}^k / (\sum_{k=1}^t p_{i,j}^k / t)), \text{ 其中, 对 } l \in I_L, 0 \cdot \log \frac{0}{0} \text{ 定义为 } 0.$$

为了对 t 个项目进行剖分, 确定评审专家非共识性较大的项目及资助项目数 N 。定义各项目的均值, 即对 $x_{ij}^k \in I_5, y_{ij}^k \in I_5$, 计算项目 k 的均值为:

$$\bar{x}^k = \frac{\sum_{i=1}^m \sum_{j=1}^n x_{ij}^k}{(m \times n)} = \frac{\sum_{i=1}^m \sum_{j=1}^n y_{ij}^k}{(m \times n)}$$

依据 $\bar{x}^k, k \in I_t$ 及 $D(z_1^1, z_1^2, \dots, z_1^t), l \in I_L$, 选择资助项目数 N 和剖分 t 各项目的均值分界值 $T_1, T_2, T_1 < T_2$, 使当项目的平均值 $\bar{x}_k < T_1$ 时, 不资助该项目; 当项目的平均值 $\bar{x}_k > T_2$ 时, 资助该项目。设肯定不资助的项目的集合为 S_1 , 肯定资助项目的集合为 S_2 , 即

$$S_1 = \{k \in I_t \mid \bar{x}_k < T_1\}, \quad S_2 = \{k \in I_t \mid \bar{x}_k > T_2\} \quad (2.2-1)$$

设 $S_0 = I \setminus (S_1 \cup S_2)$, 为可上可下项目的集合。

$$\text{令 } t_1 = |S_1|, t_2 = |S_2|, t_0 = |S_0|, n_0 = N - t_2 \quad (2.2-2)$$

现在的问题是: 根据基金评价指标, 如何从可上可下项目集合 S_0 中选出要资助的 n_0 项。这是本文主要解决的问题。

对可上可下项目集 S_0 中的项目, 必须计算每位专家对同一项目评价的偏差。为此, $\forall k \in S_0$, 分别考虑每位专家 $j \in I_n$, 对项目 k 的评价序列: $y_j^k = \{y_{1,j}^k, y_{2,j}^k, \dots, y_{m,j}^k\}$, 令 $L_1 = 4m, I_{L_1} = \{1, 2, \dots, L_1\}$, 经变换 $f: \Sigma_1 \longrightarrow \Sigma$ 后, 有

$$f(y_j^k) = z_j^k = \{z_{1,j}^k, z_{2,j}^k, \dots, z_{L_1,j}^k\}, j \in I_n, k \in S_0$$

令 $S_{i,j}^{k,l} = \{z_{i,j}^{k,l} | z_{i,j}^{k,l} \text{ 为 } z_j^k \text{ 中与 } \Theta^l \text{ 中第 } i \text{ 个序列 } s_i^l \text{ 一致的子列}\}$, 把 $S_{i,j}^{k,l}$ 中不同子列个数 $|S_{i,j}^{k,l}|$ 记为 $m_{i,j}^{k,l}$, 令 $q_1 = m_{i,j}^{k,l}$, $S_{i,j}^{k,l} = \{z_{i,j,1}^{k,l}, z_{i,j,2}^{k,l}, \dots, z_{i,j,q_1}^{k,l}\}$,

从上述可得, $\forall l \in I_{L_1}, \forall k \in S_0, \forall j \in I_n$ 有, $\sum_{i=1}^{m(l)} m_{i,j}^{k,l} = L_1 - l + 1$,

$$\text{令 } q_{i,j}^{k,l} = m_{i,j}^{k,l} / (L_1 - l + 1), \forall i \in I_{m(l)}, j \in I_n,$$

于是第 j 位专家对项目 k 的序列 z_j^k 与 Θ^l 中序列一致的子列的分布为:

$$z_j^{k,l} = \{q_{i,j}^{k,l}, i \in I_{m(l)}\}, \sum_{i=1}^{m(l)} q_{i,j}^{k,l} = 1, \forall l \in I_{L_1}$$

令 $\Gamma^l = \left\{ z_j^{k,l} \mid \sum_{i=1}^{m(l)} q_{i,j}^{k,l} = 1, q_{i,j}^{k,l} \geq 0 \right\}$, 则第 j 位专家对项目 k 的评价序列 z_j^k 的分布集为:

$$Q_j^k = \{z_j^{k,1}, z_j^{k,2}, \dots, z_j^{k,L_1}\}, \text{ 其中, } z_j^{k,l} \in \Gamma^l, \forall l \in I_{L_1}$$

对 S_0 中的项目, 有同样维数的, n 位专家的评价序列偏差可有如下定义:

$$f_{k,l} = D(z_1^{k,l}, z_2^{k,l}, \dots, z_n^{k,l}) = \sum_{j=1}^n \sum_{i=1}^{m(l)} q_{i,j}^{k,l} \cdot \log(q_{i,j}^{k,l} / \sum_{j=1}^n q_{i,j}^{k,l}), \forall l \in I_{L_1},$$

同样定义 $0 \cdot \log \frac{0}{0}$ 为 0。

选择适当的 $l \in I_{L_1}$, 根据 \bar{x}^k 与 $f_{k,l}$ 的值 ($k \in S_0$), 即同时考虑专家评分均值和非共识性的多目标评价: $f_k = \alpha_1 \bar{x}^k + \alpha_2 f_{k,l}$, 其中, α_1, α_2 为关于项目均值与非共识度的权重, 且 $\alpha_1, \alpha_2 \in (0, 1), \alpha_1 + \alpha_2 = 1$ 。依上式分别对 S_0 中的项目从大到小排序, 在从中选出 n_0 项予以资助。

具体算法如下:

1. 读入已知数据: $m, n, t, N \in Z^+$, (已知 $N < t$), $x^k \in R^{m \times n}$, $a_i, b_i \in R^+, i \in I_m$, 计算 \bar{x}^k ;
2. 选择 $T_1, T_2, T_1 < T_2$;
3. 根据 (2.2-1) 式及可上可下项目的定义计算 S_1, S_2 及 S_0 , 根据 (2.2-2) 计算 t_1, t_2, t_0 和 n_0 ;
4. 分别计算每个专家对项目 k 的总分, 从大到小重新排序后得 y^k , 经变换 $f: \Sigma_1 \longrightarrow \Sigma$, 得 z^k . 计算 $f_{k,l}$;
5. 若 $n_0 < 0$, 则给 T_2 适当增量, 转 3; 若 $n_0 = 0$, 则算法结束, 否则选权重 α_1, α_2 ;
6. 分别计算 S_0 中项目的 $\alpha_1 \bar{x}^k + \alpha_2 f_{k,l}$ 的值, 从大到小排序, 分别记为 $f_{1,j_1}, f_{1,j_2}, f_{1,j_3}, \dots, f_{1,j_{n_0}}$, 其中, $j_1, j_2, \dots, j_{n_0} \in S_0, k \in S_0$;
7. 若 $f_{1,j_{n_0}} > f_{1,j_{n_0+1}}$, 则取 $j_1, j_2, \dots, j_{n_0} \in S_0$ 为资助项目, 运算结束; 否则, 若 $f_{1,j_{n_0}} = f_{1,j_{n_0+1}}$, 则令 $S_{n_1} = \{f_{1,j_h} | f_{1,j_h} = f_{1,j_{n_0}}\}, h_0 = |S_{n_1}|$, 修改权系数 α_1, α_2 (前者减小, 后者增大), 并令 $S_0 = S_{n_1}, t_0 = h_0$, 转 5.

可以证明, 算法 CM 或有限步结束, 即找出 n_0 个资助项目, 或无穷多步下, 使 $n_0 > 0$, 这时可人工处理 S_0 中的项目, 因为此时 $|S_0| = n_0$ 很小。

2.3 应用实例

以某年国家自然科学基金的项目评审为例。国家自然科学基金申请项目同行评议卡中的定量评审分为四部分共 14 项指标, 评审专家对每个申请项目按各指标选择 A, B, C, D 。根据 A, B, C, D 得对应的分数。下表为某学科 7 个项目同行评议定量评议的得分(见下页表 2.2)。其中, 科学部编号为项目的编号, 1-4 部分的得分是项目四部分部份指标的各自得分。

根据算法, 将数据读入后, 首先将专家评委对项目 k 的评审总分从大到小排列, 再根据变换 $f: \Sigma_1 \longrightarrow \Sigma$ 将其变化成 $5 \times (14 \times 4) = 5 \times 56$ 阶的评分矩阵。我们研究 $l \leq 2$ 的情况。当 $l = 1$ 时, $\Theta^1 = \{A, B, C, D, E\}$, 计算每个序列的偏差(即异同性) $f_{k,l}$,

科学部编号	专家对项目的14个指标的 打分	总分	1部分的得 分	2部分的得 分	3部分的得 分	4部分的得 分	专家编号
39660357	BCCCCDCDDDCDD	406	310	81	15	0	11458
39660357	CBBABCBBBCBCBC	686	256	235	95	4	11457
39660357	CCCCCDDCBCDCD	289	150	69	70	0	11459
39660357	CCCCDCCDDDDDD	254	150	89	15	0	11018
39660357	BCBBCCBCCCCD	520	300	175	45	0	11019
39670598	BBBBBCCAABCBC	685	400	145	140	5	11458
39670598	BCBBBBBBBBBBBD	700	300	280	120	0	11457
39670598	AABBAAAAABBBAB	944	464	350	130	9	11459
39670598	BBAABBBAAABBBAB	858	436	292	130	7.5	11018
39670598	CCCCCCCCCCCCC	300	150	105	45	0	11766
39670599	BBCBDDCCBCBCD	167	350	47	40	0	11766
39670599	ABBABBBBBAABAA	860	440	280	140	9.9	11458
39670599	AAABBAAAAAAAAAA	968	484	334	150	9.9	11459
39670599	BBBAABBABABBAA	854	416	308	130	7.5	11018
39670599	BBAAAABAAABBAB	902	436	326	140	9.9	11457
39660357	ABBBBCBCAABBBC	769	424	205	140	0	11457
39660357	DCCCDCCDCCCD	250	126	89	35	0	11018
39660357	AABABABBBAABAA	918	480	298	140	10	11019
39660357	BBBCCCBCBCCCC	605	400	135	70	5	11458
39660357	ABBBDDCDBBCCD	612	424	103	85	0	11766
39674059	BBBBBCBBCBBBBC	730	400	235	95	8	11766
39674059	AAAAAAAAAAAAAAAA	1000	500	350	150	15	11459
39674059	CBBBBBBABBBCBD	75562	340	292	120	0	11457
39674059	BAAAAAABBABAA	968	476	350	130	10	11018
39674059	BBBABBBSBBBBSBC	816	416	280	120	10	11019
39674065	BBBBBCCABACCC	685	400	145	140	8	11766
39674065	BABAABBBBBAABB	882	456	296	130	10	11458
39674065	AAABAAAAAAAAAAB	984	484	350	150	15	11459
39674065	CCBCCCCADDCCD	365	200	105	60	0	11018
39674065	AABBABABABAAAA	924	464	320	140	14.5	11019
39675607	BCBBBBBBBBBBCBD	700	300	280	120	0	11019
39675607	BCBBCCBCCDCCD	510	300	175	35	0	11018
39675607	BBABABBAABBAAB	858	420	308	130	9	11459
39675607	ABAAABBAABAA	918	460	308	150	9	11458
39675607	CCBBDDCCBBBCCD	487	240	127	120	0	11766

表 2.2 参评项目评分表

Fig 2.2 The Scores of Evaluated Projects

当 $l = 2$ 时, $\Theta^2 = \begin{bmatrix} AA & AB & AC & AD & AE \\ BA & BB & BC & BD & BE \\ CA & CB & CC & CD & CE \\ DA & DB & DC & DD & DE \\ EA & EB & EC & ED & EE \end{bmatrix}$ 。但由于变换 $f: \Sigma_1 \longrightarrow \Sigma$ 的特殊性,

显然变换后的序列中, 下列组合是不存在的: $\{AE, BA, BC, BD, CA, CB, CD, DA, DB, DC, DD\}$ 。故在实际计算中可以不必考虑。所以, $\Theta^2 = \{AA, AB, AC, AD, BB, BE, CC, CE, DE, EA, EB, EC, ED, EE\}$, 同样计算得到 $f_{k,2}$; 根据算法取 $\alpha_1 = \alpha_2 = 0.5$, 则由 $f_k = \alpha_1 \bar{x}_k + \alpha_2 (f_{k,1} + f_{k,2})$, $k \in S_0$, 得到结果(见表 2.3)。

排序名次	科学部编号	申请者	项目名称	准则 f_k
7	39660357	XXX	气候变化对亚热带岩溶区作物生产的影响及对策研究	22.17 (G)
5	39670598	XXX	作物植被冠层 CO2 源流变化与其调控机理的研究	74.70 (D)
3	39670599	XXX	作物光能利用的太阳分光辐射的研究	77.56 (B)
6	39671919	XXX	气候变化对中国作物基因资源及育种的影响	68.04 (F)
1	39674059	XXX	华北平原粮食生产气候风险评估和决策方法研究	107.44 (A)
2	39674065	XXX	大气涡能场的地球物理学基础及其在减灾中的应用研究	89.97 (C)
4	39675607	XXX	水稻洪涝灾害的危害机制及其抗灾对策	76.91 (E)

表 2.3 某学科申请项目同行评议意见量化排序表

Fig 2.3 The final result of the project evaluation in a certain field

注: 括号中字母顺序、为原综合评审分数排名

3 关于非共识度的归一化方法

3.1 应用背景

随着基金项目评审过程越来越透明，评审方法的客观性要求越来越高，势必使方方面面面对评审方法的分歧越来越小，所以评审方法在实际运用中应不断改进。在实际评审中，就是因为对影响评审的各种因素考虑得越来越细，我们注意到了专家打分时存在的分歧问题，这种分歧不仅不能被上述方法在项目最终得分中体现出来，而且还影响项目评审。我们发现上文介绍的评审方法的一般评审准则总是将多目标转化为单目标排序，很少考虑各目标之间及目标本身的关系，解决此类问题的工作量很大且不直观。如属性分析法或区间分析法、层次分析法等，都是在一维空间中给出项目的排序，没有注意到项目自身的空间分布，从而无法考虑项目的非共识度问题，忽略了非共识度尺度不一致而影响项目评审的公平与合理性的问题。此外，以上述方法为理论基础的评审软件系统还不够完善。

就基金评审中存在的上述问题，本章提出平均非共识度和最大非共识度的概念，针对项目评分的最大非共识度分布不一致的问题，应用优化理论、归一化方法等，建立了确定平均非共识度和最大非共识度的优化模型与算法，即平均分不同的项目对应的非共识度二次规划和最大非共识度不可微规划。依据归一化后的项目平均分、非共识度和最大非共识度的加权评分大小得出量化排序。这一结果可以科学、合理地评价申请项目，克服了因最大非共识度的不同影响，为评审委员会清楚、直观、简易地选择优秀项目提供了可靠的定量依据。项目评审的归一化方法已在国家自然科学基金委的评审工作中得到应用，收到很好效果。

评议的量化分为两种，一是评议者直接给出整体的综合评价指数，二是评议者针对不同的指标给出评价指标，然后，通过对指标给与不同的权重，得出综合评议指标。目前已有的方法是：

单指标： y

多指标(n 个)： $x_i, i=1,2,\dots,n$

综合指标：

一般为加权线性相加， $y = \sum_{i=1}^n \alpha_i \times x_i, i=1,2,\dots,n$

特例韦加权相乘, $y = \prod_{i=1}^n \alpha_i \times x_i, i = 1, 2, \dots, n$

群体 (m 人) 评价的平均:

加权线性相加, $Y = \sum_{j=1}^m \beta_j \times Y_j, j = 1, 2, \dots, m$

关于多指标的问题, 专家认为对于科学基金主要和标志性资助项目——面上项目并不合适, 这是因为科学基金资助的项目是个人独立的、小规模、单独问题的研究活动, 评价不宜细化, 多指标评议适合于大项目的评审。美国国家科学基金会针对评议指标问题开展过长期的、大规模的研究, 但一直坚持给出一个综合定量指标。所以, 下面的讨论主要针对项目的综合指标。

3.2 归一化分析

3.2.1 基金项目评审的非共识度概念和项目评审模型

本节, 对同行评议的非共识量化描述是这样的:

设某学科的面上基金申请的项目集合记为 S , 项目数为 n , 其指标集为 I_n , 显然 $|I_n| = n$ 。每个项目请 m 位专家评审, 专家们依每个评价指标逐项给出评分。由此可计算出专家 j 对项目 k 的评议综合分数为 y_{kj} , 评审项目 k 的专家指标集记为 O_k , 显然 $|O_k| = m$ 。定义项目 k 的平均分数为:

$$\bar{y}_k = \sum_{j \in O_k} y_{kj} / m。$$

首先要对同行评议的非共识量进行量化描述, 由此引入下面几个相关定义:

定义 1 令

$$\Delta y_k = \sum_{j \in O_k} |y_{kj} - \bar{y}_k| / m$$

则称 Δy_k 为 m 个专家对于项目 k 的平均非共识度或平均分歧度。

定义 2 令

$$\Gamma y_k = \max_{j \in O_k} \{y_{kj}\} - \min_{j \in O_k} \{y_{kj}\}$$

则称 Γy_k 为最大非共识度或最大分歧度。

对于任意项目 k , $\bar{y}_k, \Delta y_k, \Gamma y_k$ 可以构成三维空间, 在这个三维空间里能比较全面地描述项目的评议分数特征。其中, $\Delta y_k, \Gamma y_k$ 描述的是同行评议中对项目 k 的非共识

情况。根据多年的基金评审实践经验，在项目平均分数 \bar{y}_k 相近的情况下，往往专家分歧越大的项目，其创新性就越强。因为在科学的探索中，人的思维有两种主要的方式：趋同和异化。当一个新的学科或分支出现后，会有越来越多的人参加研究，称之为趋同。趋同思维方式使得该领域的研究愈发深入，使其优点、缺点、适用性很快地为人所知。而也有一些人不断地探索新的研究领域，提出新的研究途径，解决新的问题，称之为异化。异化思维使科学不断地创新。扩展探究领域，认识更多的未知领域。这两种不同的思维方式的交互作用推动了科学越来越快的发展。当然，在社会生活中，人们的思维方式要复杂得多，但是趋同和异化依然是其中的主要方式。基金项目评审中，通过专家打分得到的项目平均得分就是趋同思维的体现，相对的，专家间存在评分差异就是异化思维的作用。项目评审方法要体现出这两种思维的综合作用才能更公平公正。因此，要找到一种方法，包含项目平均分和非共识度的综合度量。

在评选项目的过程中，当项目得分之间相差很小时，要优先考虑选择分歧度较大的项目。所以，对于所有参评项目，可以根据其平均分和分歧度分为三个集合：

首先，设 P 为综合分数域，是有界闭集， $\forall a, b \in P$ ， $S_1(a) = \{k | \bar{y}_k \geq a, k \in I_n\}$ ， $S_2(b) = \{k | \bar{y}_k < b, k \in I_n\}$ 。

选择 α_1 (α_1 为入选比例)，求 a (a 为入选分数线)，使 $|S_1(a)| = \alpha_1 \cdot |I_n|$ ，并把 $S_1(a)$ 称为入选项目集合，这类项目将予以资助；

选择 α_2 (α_2 为落选比例)，求 b (b 为落选分数线)，使 $|S_2(b)| = \alpha_2 \cdot |I_n|$ ，并把 $S_2(b)$ 称为落选项目集合，这类项目将不予考虑；

称 $S_3(a, b) = I_n \setminus (S_1(a) \cup S_2(b))$ 为可上可下项目集合。此类项目分数处在中游，不易明确是入选资助还是落选。

3.2.2 归一化分析与算法

尽量消除专家打分的差异，使项目评审尽可能具有创新性、公平性和公正性，是本节重点解决的问题。项目的平均分数不同，对应的平均非共识度最大值就不同，造成评价尺度不同，从而使相同的平均非共识度反映的非共识性是不同的。为此我们对项目平均评分、平均非共识度、最大非共识度应用归一化方法，消除分歧，使评价尺度一致。实践中我们发现 S_3 中的项目往往是平均评分差距不大，得分在中游，如果存在个别专家打分偏高或偏低的现象，对整个评比结果的影响很大。所以应重点对这

类项目进行归一化分析。具体地，通过平均分、平均非共识度、最大非共识度与各自最大值相比的方法消除刻划尺度的不同，达到归一化的目的。

首先，对平均分进行归一化。已知平均评分的最大值，例如某学科的满分 1000 分。所以用项目的实际得分与满分 1000 分相比，就能实现平均分的归一化。

其次，我们考虑平均非共识度和最大非共识度的归一化问题。归一化问题中，平均非共识度和最大非共识度的最大值可以通过分别求解二次规划和不可微规划得到。对于平均非共识度而言，先对项目进行归一化分析，对任意给定的项目平均分 $\bar{y} \in [0,1000]$ ，求使项目平均非共识度最大的二次规划问题（记为 P_1 ），表示为：

$$P_1: \quad \max J_1(\bar{y}) = \left(\sum_{j=1}^m (y - \bar{y}) \right)^2 / m$$

$$\text{s.t.} \quad \sum y = m\bar{y}$$

$$y \in [0,1000]$$

这个规划含 m 个优化变量， \bar{y} 为已知正数， y 为优化变量。目标为二次函数，且具有线性约束，所以问题 P_1 为二次规划。显然对于任意给定的 $\bar{y} \in [0,1000]$ ，问题 P_1 的最优解存在但不唯一，最优值 $J_1(\bar{y}) \in R_+$ 存在且唯一，即 $P_1: \bar{y} \in [0,1000] \rightarrow J_1(\bar{y}) \in R_+$ 为一一映射。所以，对平均非共识度的归一化结果是存在且唯一的，用平均非共识度与对应问题 P_1 的最优解相比得到。

同理，求最大非共识度问题：对任意给定的 $\bar{y} \in [0,1000]$ ，求使项目最大非共识度最大的问题（记为 P_2 ），可表示为：

$$P_2: \quad \max J_2(\bar{y}) = \{\max\{y\} - \min\{y\}\}$$

$$\text{s.t.} \quad \sum y = m\bar{y}$$

$$y \in [0,1000]$$

其中， \bar{y} 为已知正数， y 为优化变量，它含 m 个优化变量。显然，目标函数不可微，但约束是线性的，所以 P_2 为不可微规划。且对于任意给定的 $\bar{y} \in [0,1000]$ ，问题 P_2 的

最优解存在但不唯一，而最大值存在且唯一，所以对最大非共识度的归一化结果是存在且唯一的，用最大非共识与对应问题 P_2 的最优值相比得到。

最后，选取关于项目平均分、平均非共识度、最大非共识度的权系数 $\delta_1, \delta_2, 1 - \delta_1 - \delta_2$ ，使 $\delta_1, \delta_2, (1 - \delta_1 - \delta_2) \in (0, 1)$ ，令项目的加权评分 y' 为：

$$y' = \delta_1 \frac{\bar{y}}{1000} + \delta_2 \frac{\Delta y}{J_1(\bar{y})} + (1 - \delta_1 - \delta_2) \frac{\Gamma y}{J_2(\bar{y})}$$

根据 y' 的大小，分别对 S_3 中项目从大到小排序，再从中选取项目予以资助。

具体的归一化算法的主要计算步骤为：

- 1). 读入项目数据，并根据实际要求整理出有效数据；
- 2). 输入入选比例 α_1 和落选比例 α_2 ；
- 3). 根据项目平均分的定义及定义 1, 2 计算项目的平均分 \bar{y} ，平均非共识度 Δy 和最大非共识度 Γy ，并按平均分的大小对项目排序；
- 4). 由落选比例和入选比例得到项目入选集合 $S_1(a)$ 、落选集合 $S_2(b)$ 和可上可下集合 $S_3(a, b)$ ；
- 5). 对任意的 $\bar{y}_n = 0 + 10n \in [0, 1000], n = 0, 1, \dots, 100$ ，以 0 为初始点，10 为步长，解规划 P_1, P_2 ，得到对应的最优解点列 $J_1(\bar{y}_n), J_2(\bar{y}_n), n = 0, 1, \dots, 100$ ；根据点列拟合出对于任意 $\bar{y} \in [0, 1000]$ ，问题 P_1, P_2 的最优解及图像；
- 6). 对于 $\bar{y} \in S_3$ ，通过项目平均评分 \bar{y} 与评分最大理论值的比，即 $\bar{y}/1000 \in [0, 1]$ ；项目平均非共识度 Δy 与对应 P_1 问题的最优值 $J_1(\bar{y})$ 的比，即 $\Delta y/J_1(\bar{y}) \in [0, 1]$ ；项目最大非共识度 Γy 与 P_2 问题的最优值 $J_2(\bar{y})$ 的比，即 $\Gamma y/J_2(\bar{y}) \in [0, 1]$ ，得到对项目平均分、平均非共识度、最大非共识度归一化结果；
- 7). 取权系数为 $\delta_1, \delta_2, 1 - \delta_1 - \delta_2, \delta_1, \delta_2, (1 - \delta_1 - \delta_2) \in (0, 1)$ ，对于 $\bar{y} \in S_3$ ，根据

$$y' = \delta_1 \frac{\bar{y}}{1000} + \delta_2 \frac{\Delta y}{J_1(\bar{y})} + (1 - \delta_1 - \delta_2) \frac{\Gamma y}{J_2(\bar{y})}$$

求每一项的归一化加权评分。并按加权得分的大小排序。

- 8). 根据实际要求，按加权评分大小确定 S_3 中有哪些项目入选；

由上述算法可知，任一项目 k 的平均评分 \bar{y}_k 与所对应的非共识度最优值 $J_1(\bar{y}_k)$ 所形成的区域、平均评分 \bar{y}_k 与对应的最大非共识度的最优值 $J_2(\bar{y}_k)$ 所形成的区域分别反映了两种不同二维空间的分布状态。这样使决策者更直观，形象地分析这些项目。

这一结果可以科学、合理地评价申请项目，为评审委员会清楚、直观、简易地选择优秀项目提供了可靠的依据。

3.3 应用实例

项目归一化方法在国家自然科学基金委员会的部分学科项目评审中已得到应用。以某学科 95、96、98、99 年的项目评审为例。图 3.1 显示了这些项目的得分情况。

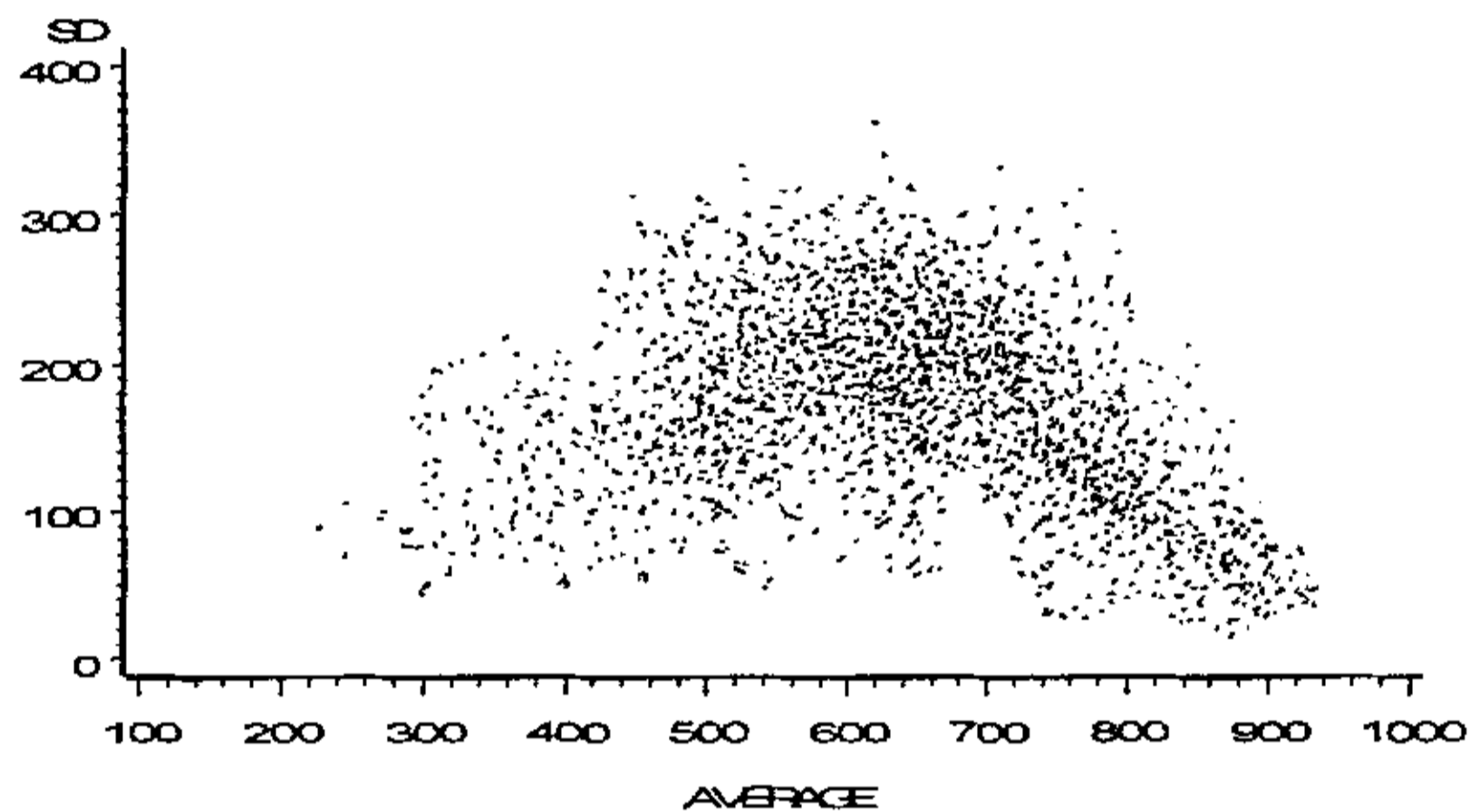


图 3.1 参评项目得分点图

Fig 3.1 The Scores of the Evaluated Projects

图中坐标横轴表示项目的平均分 \bar{y}_k ，坐标竖轴表示项目的平均非共识度 Δy_k ， \bar{y}_k 和 Δy_k 构成的二维空间可以显示出平均分数和平均非共识度的关系，可以看出申请项目的同行评议非共识的情况。

首先对原始数据进行整理，共有 1587 位专家对 3345 个项目进行评审，每个项目由 5 位专家打分，每个项目得到 5 项综合评分，即 $n = 3345, |O_k| = m = 5$ ，第 k 项的评分集合 $y_k = \{y_{kj}, j \in O_k\}$ ，则该项平均分为：

$$\bar{y}_k = \sum_{j \in O_k} y_{kj} / 5。$$

按平均分的从大到小给项目排序。入选项目的比例为 $\alpha_1 = 20\%$ ，落选比例为 $\alpha_2 = 30\%$ ，根据项目的平均分高低排序确定：入选分数线为 748.08，落选分数线为 575，则参评项目可分为以下集合：

$$S_1(a) = \{k | \bar{y}_k \geq 748.08, k \in I_n\},$$

$$S_2(b) = \{k | \bar{y}_k < 575, k \in I_n\},$$

$$S_3(a, b) = I_n \setminus (S_1(a) \cup S_2(b)),$$

其中 $|S_1(a)| = \alpha_1 \cdot |I_n| = 3345 \times 20\% \approx 670$,

$$|S_2(b)| = \alpha_2 \cdot |I_n| = 3345 \times 30\% \approx 1004。$$

$$|S_3(a, b)| = n - |S_1(a)| - |S_2(b)| = 1671。$$

计算 $S_3(a, b)$ 中各项目的平均非共识和最大分共识度，求关于这些项目的优化问题 P_1 和 P_2 ，得到下列两个最优值图像（见图 3.2、图 3.3）。

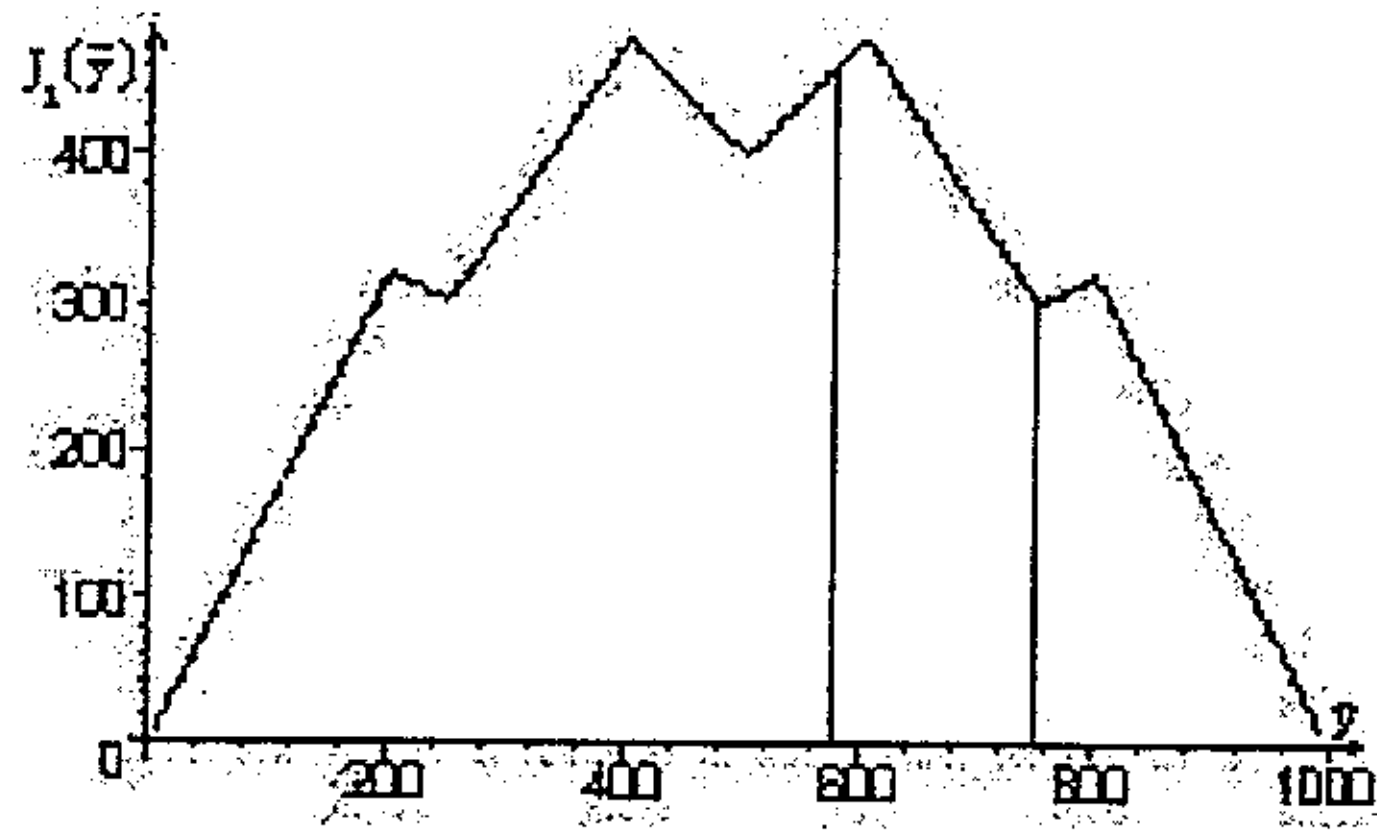


图 3.2 问题 P_1 的最优解的图像
Fig 3.2 The Optimization of P_1

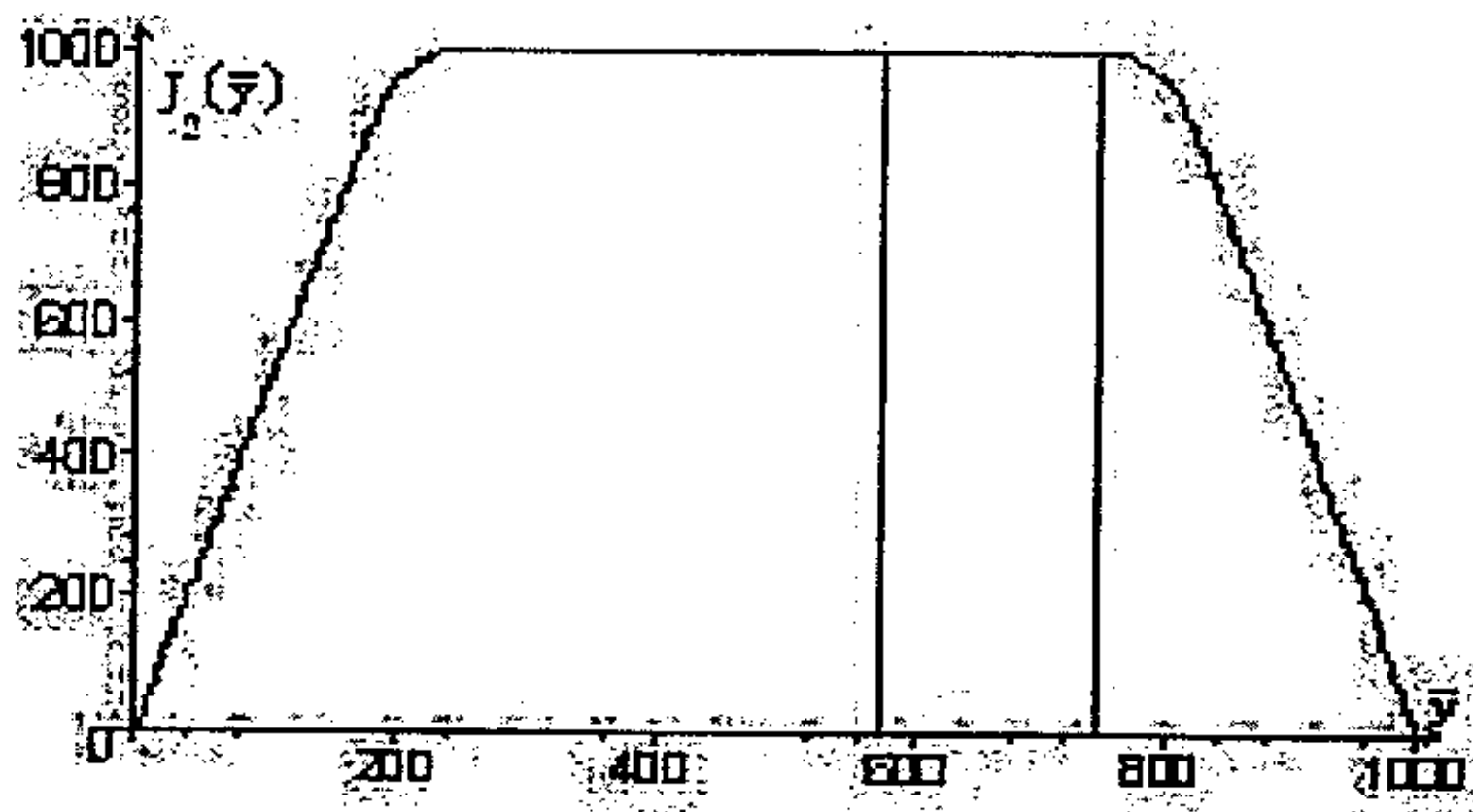


图 3.3 问题 P_2 的最优解的图像
Fig 3.3 The Optimization of P_2

图 3.2 为问题 P_1 在区间 $[0,1000]$ 上得到的最优值的图像。其中, x 轴表示项目的平均得分 \bar{y} , y 轴表示对应 x 轴项目平均得分的非共识度最优值 $J_1(\bar{y})$ 。两条与 y 轴平行的线段表示入选和落选分数线, 线段间的部分是 $S_3(a,b)$ 中个项目的最优解。

图 3.3 为问题 P_2 在区间 $[0,1000]$ 上得到的最优值的图像。其中, x 轴表示项目的平均得分 \bar{y} , y 轴表示对应 x 轴项目平均得分的最大非共识度最优值 $J_2(\bar{y})$ 。两条与 y 轴平行的线段依然表示入选和落选分数线, 两者之间的图像即 S_3 中个项目的最优解。

根据评审情况, 项目平均分、平均非共识度、最大非共识度对应的权系数分别是 0.7, 0.2, 0.1。则项目的归一化加权结果是:

$$y' = 0.7 \frac{\bar{y}}{1000} + 0.2 \frac{\Delta y}{J_1(\bar{y})} + 0.1 \frac{y}{J_2(\bar{y})}$$

根据上述结果大小将 S_3 内的项目排序, 批准资助的项目评议情况见图 3.4。其中曲线部分是 \bar{y}_k 和 Δy_k 构成的两维空间的值域界限, 即优化问题 P_1 的解。点图部分是入选项目的平均得分与非共识度的空间情况。

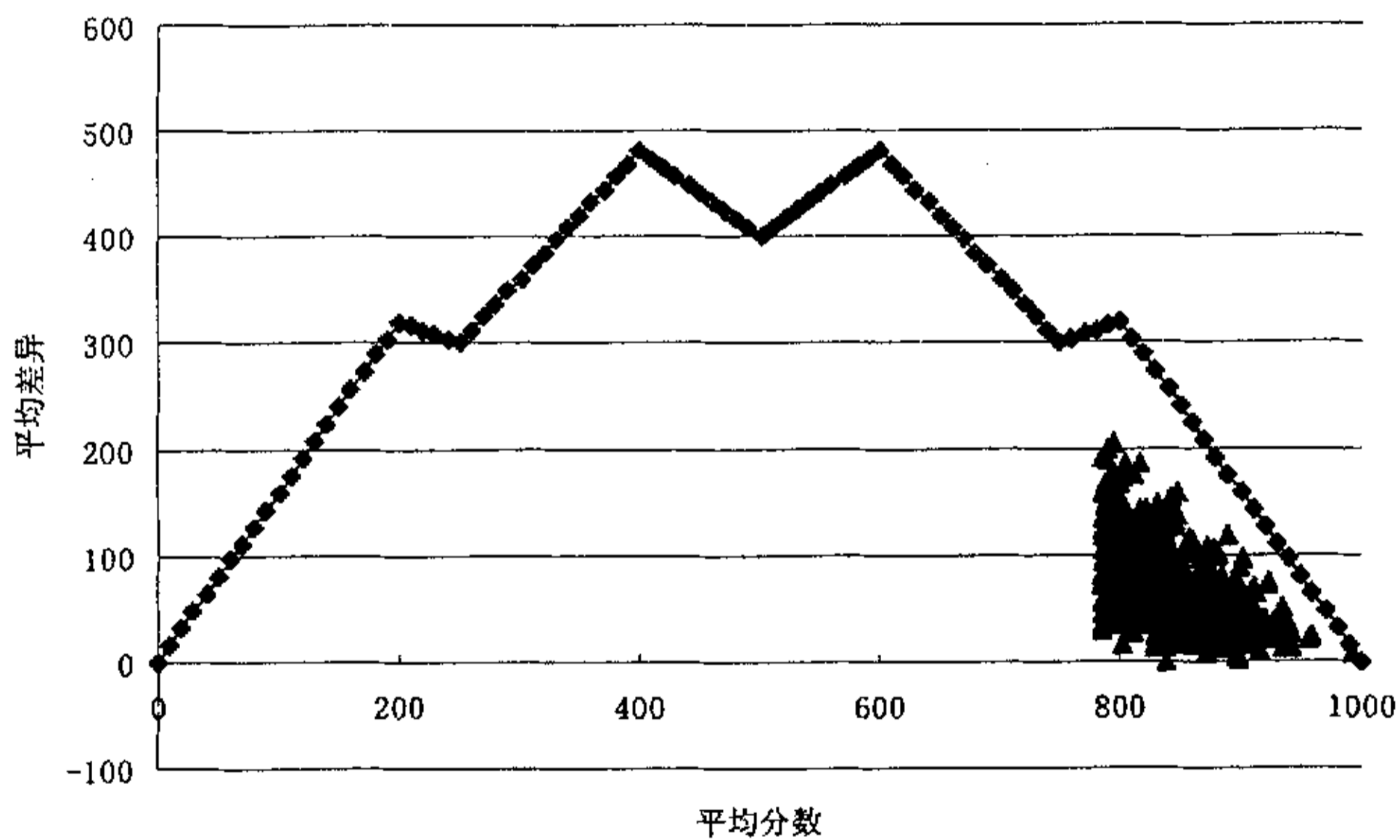


图3.4 平均差-非共识的值域

Fig 3.4 The Value of Average & Non-commonity

3.4 结论

本章分析了项目评审中的非共识度问题，从模型入手，提供一套方法解决非共识的问题。实践证明，发挥作用的评审系统中的人，单纯使用数学方法可能会导致走向错误的方向。因为引起评审中的非共识，创新不是唯一的因素，需要在认真分析定性的评价，而这样的分析和判断必须由人作出。

但是对非共识的量化分析研究证明了评审系统的确存在的问题。归一化方法在实际工作中提出处理非共识比较妥善的原则和方法，打破了评审系统的“共识”性的绝对平衡，给具有争议的创新性项目更大的资助机会。

非共识评审系统已在基金评审委员会得到应用，收到很好效果，并正式被纳入评审政策。基于归一化方法的应用软件也已完成，并在与基金委的合作中得到不断完善和维护。

4 专家反评估

4.1 引言

选择专家是国家自然科学基金项目评审的重要构成因素，也是国家基金充分发挥作用的根本保证。当前，国内外公众对国家基金评审制度的批评，实质上很多是针对同行评议专家的。因此，如何建立科学与可行的评议专家体系，以便选好、用好评议专家，是当前亟待解决的问题。

评审专家以往对基金项目评议的情况，在一定程度上反映了评审专家的水平 and 态度。将评议专家的工作业绩用有关指标进行分析，列入评议专家选择的指标体系，这对于挑选、调整评议专家是十分必要的，而且可使同行评议专家处于动态优化过程中。但是，以往关于评审专家的评审质量问题，即反评估问题仅限于定性描述，到目前，还没有见到对评审专家评审质量进行定量描述的研究。用“非共识”的方法可以分析评议专家群体的评议特征，可以发现专家中的另类，他们的评议与群体有较大的差异，是群体中的另类。具有这样特征的专家，要么是很傲气，不肯说太多的好话，要么是“好好”先生，习惯于说好话。掌握专家的评议特征，对于项目主任分析其评价意见有很多帮助，虽然一些资深的项目主任对评议专家的特点很熟悉，但是量化分析更能使他们“心中有数”，掌控评审质量。

本章对评议专家的选择问题提出平均差异、绝对差异、成功率、命中率等的概念用以量化描述专家的评分质量。在此基础上建立了关于评审专家评审质量的数学模型与优化算法。研究的软件已应用于对基金评审专家的选择。

4.2 基金项目评审的非共识度概念和项目评审模型

设某学科的面上基金申请的项目数为 n ，其指标集为 I_n ，显然 $|I_n| = n$ 。每个项目请 m 位专家评审，设 $U = \{e_1, e_2, \dots, e_l\}$ 为评审专家的集合， $I_l = \{1, 2, \dots, l\}$ ，称 U 为论域， I_l 为专家标号集。专家们依每个评价指标逐项给出评分。设专家 j 对项目 k 的评议综合分数为 $y_{kj} \in P \subset R$ ，其中 P 为综合分数域，是有界闭集。设评审项目 k 的专家指标集记为 O_k ，则 $|O_k| = m$ 。定义项目 k 的平均分数为：

$$\bar{y}_k = \sum_{j \in O_k} y_{kj} / m。$$

首先要对同行评议的非共识量进行量化描述，由此引入下面几个相关定义：

定义1 设

$$\Delta y_k = \sum_{j \in O_k} |y_{kj} - \bar{y}_k| / m$$

则称 Δy_k 为 m 个专家对于项目 k 的非共识度或分歧度。

定义2 设

$$\Gamma y_k = \max_{j \in O_k} \{y_{kj}\} - \min_{j \in O_k} \{y_{kj}\}$$

则称 Γy_k 为最大非共识度或最大分歧度。

对于任意项目 k , $\bar{y}_k, \Delta y_k, \Gamma y_k$ 可以构成三维空间, 在这个三维空间里能比较全面地描述项目的评议分数特征。其中, $\Delta y_k, \Gamma y_k$ 描述的是同行评议中对项目 k 的非共识情况。根据多年的基金评审实践经验, 在项目平均分数 \bar{y}_k 相近的情况下, 往往专家分歧越大的项目, 其创新性就越强。因此, 在评选项目的过程中, 当项目得分之间相差很小时, 要优先考虑选择分歧度较大的项目。对于所有参评项目, 可以根据其平均分和分歧度分为三个集合:

首先, $\forall a, b \in P$, 设 $S_1(a) = \{k | \bar{y}_k \geq a, k \in I_n\}$, $S_2(b) = \{k | \bar{y}_k < b, k \in I_n\}$ 。

选择 α_1 (α_1 为入选比例), 求 a (a 为入选分数线), 使 $|S_1(a)| = \alpha_1 \cdot |I_n|$, 并把 $S_1(a)$ 称为入选项目集合, 这类项目将予以资助;

选择 α_2 (α_2 为落选比例), 求 b (b 为落选分数线), 使 $|S_2(b)| = \alpha_2 \cdot |I_n|$, 并把 $S_2(b)$ 称为落选项目集合, 这类项目将不予资助;

称 $S_3(a, b) = I_n \setminus (S_1(a) \cup S_2(b))$ 为可上可下项目集合。此类项目分数处在中游, 不易明确是入选资助还是落选。

4.3 评审专家的反评估

为了考察专家在评审过程中的学术水平, 评比公正性等情况, 除了以上的量化描述外, 还应着重对评议者的“非共识”特征进行量化分析。所以我们有以下概念。

4.3.1 平均差异和绝对差异

设 j 专家评议了 n_j 个项目, 第 j 位专家评审的项目指标集为 S_j , 故 $|S_j| = n_j$ 。对于项目 k 的评分 y_{kj} 与同行评议专家的平均分数 \bar{y}_k 的差异可以用下面定义来描述:

定义3 设

$$\Delta y_j = \sum_{k \in S_j} (y_{kj} - \bar{y}_k) / n_j$$

$$\text{令 } y_{\max j} := \max_{k \in S_j} \{y_{kj}\}, y_{\min j} := \min_{k \in S_j} \{y_{kj}\}, \text{ 则 } \Delta y_j \in (-y_{\max j} + y_{\min j}, y_{\max j} - y_{\min j}).$$

则称 Δy_j 为第 j 位专家评审的平均差异即平均非共识度。

它不仅能够反映出某评议专家与其他专家对被评审项目在认识上的差异性,更重要的是它还能较好地反映出该专家在评审过程中表现出来的水平与公正性。当 j 专家的平均差异为正时,说明其总体上打分偏高;而当平均差异为负时,则说明其总体上打分偏低;平均差异的值接近 0 时,说明专家打分与项目平均分差别不大,打分正常。

定义4 设

$$\Gamma y_j := \sum_{k \in S_j} |y_{kj} - \bar{y}_k| / n_j,$$

则称 Γy_j 为第 j 位专家评审的绝对差异即绝对非共识度。这里 $\Gamma y_j \in (0, y_{\max j} - y_{\min j})$ 。

4.3.2 命中率

定义5 在同行评议中,评议专家 j 对申请项目的评议结果与最终评议结果的一致性的 大小称为评议专家 j 的评审命中率,记为 P_j ,即

$$P_j = (x_{0j} / n_j) \times 100\%,$$

其中 x_{0j} 表示专家 j 的评审结果与最终结果一致的项目数; n_j 表示专家 j 的评议项目数。

国家自然科学基金委员会成立以来资助了大量科研项目,从已完成的项目来看,资助是有效的。从这个角度看,同行评议专家的评议结果是否与基金委员会资助的政策相符,是对同行评议专家评议能力的一种检验。如果专家政策水平高,深刻了解基金资助宗旨;学术水平高,准确判断申请项目的价值;同时认真负责,公平合理,那么所评议项目的命中率一定高。命中率的高低反映了同行评议专家的能力和水平。

命中率的高低还与该专家一次评议项目的多少有关。通常,评议项目数多的专家的命中率要高于评议项目少的专家。一次评审项目数较多,可以使专家在较大范围内进行比较,从而找出其中的最优方案。但一次评审项目数过多也会出现相反情况。所以一般专家评审项目数在 10 项左右为宜。

4.3.3 成功率

在基金数额有限的情况下,我们希望获准资助的项目能获得成功,从而促进我国科技事业的发展,这也是检验基金工作的重要标准。但是,有些资助项目可能会由于种种原因最终没有获得成功,这也从一个侧面反映了同行评议结果的正确与否。因此,有必要提出成功率的概念,以作为评议专家工作业绩检查的重要标准。

定义 6 实践证明取得成功的资助项目与评审中专家 j 评出的资助项目的比率,称为成功率。记为 r_j , 即

$$r_j = (r_{0j}/r_{1j}) \times 100\%$$

其中 r_{0j} 表示评议专家 j 认可获得赞助,且经实践证明是成功的项目个数; r_{1j} 表示评议专家 j 认为应获得赞助的项目数。

成功率是从最终结果的角度对评议专家的水平进行总体反映的指标。成功率的高低从根本上反映了评议专家水平的高低,较高的成功率反映了同行评议系统的功能,使基金资助的宗旨得以实现。因此,它与其他指标相比具有更强的说服力,是所有评价指标体系中非常重要的一个指标。但是,作为基础研究和应用项目,其成功与否不是短时间能够反映的,这类项目研究的成功标准是一个涉及面极广,非常复杂的项目后期管理问题。我国对此问题的研究还不是特别深入。这为成功率的判断无形中带来较大困难,随着基础工作的完善,基金资助工作的发展,这一指标的量化描述将更为合理,贴近事实。

4.3.4 专家反评估系统

上几节讨论的是各位专家的一维信息。我们知道作为学科主任,每年要研究大量的申请项目和评审专家。为了更好的挑选评审专家,仅用命中率、成功率这几个数字是不足以说明问题的。随着计算机的普及以及办公自动化程度的提高,基金评审及专家选择越来越依赖于计算机。因此本节研究评审专家反评估系统及在计算机里的二维表示。以图文的形式使基金管理者更加清楚各位专家的评审情况。

对于每个专家 j , $(\Delta y_j, \Gamma y_j)$ 可以构成二维空间的向量,用来描述 j 专家的评议特征。由上面章节的定义 3 与定义 4 可以证明:

$$(\Delta y_j, \Gamma y_j) \in D := \{(\Delta y, \Gamma y) | \Gamma y \geq \Delta y, \text{当 } \Delta y \geq 0 \text{ 时}, \Gamma \geq -\Delta y, \text{当 } \Delta y \leq 0 \text{ 时}\},$$

即所有 $(\Delta y_j, \Gamma y_j)$ 都位于二维直角坐标系的第一二象限角平分线之间的区域内 (如图 4.1 所示)。所以, 若选择平均差异允许值 σ_1 和绝对差异允许值 σ_2 , 及 σ_3 , 使 $\sigma_1 \in (0, 2y_{\max} - 2y_{\min})$ 且 $\sigma_2 \in (0, y_{\max} - y_{\min})$, σ_3 表示某一给定的绝对差异值, 把上述区域分为几个部分:

$$D_1 = \left\{ (\Delta y_j, \Gamma y_j) \in D \mid \Delta y_j \in \left[-1000, -\frac{\sigma_1}{2} \right], \Gamma y_j \in [\sigma_2 + \sigma_3, 1000] \right\},$$

$$D_2 = \left\{ (\Delta y_j, \Gamma y_j) \in D \mid \Delta y_j \in \left[\frac{\sigma_1}{2}, 1000 \right], \Gamma y_j \in [\sigma_2 + \sigma_3, 1000] \right\},$$

$$D_3 = \left\{ (\Delta y_j, \Gamma y_j) \in D \mid \Delta y_j \in \left[-\frac{\sigma_1}{2}, \frac{\sigma_1}{2} \right], \Gamma y_j \in [\sigma_2 + \sigma_3, 1000] \right\},$$

$$D_4 = \left\{ (\Delta y_j, \Gamma y_j) \in D \mid \Delta y_j \in [-1000, 1000], \Gamma y_j \in [\sigma_3, \sigma_2 + \sigma_3] \right\},$$

$$D_5 = \left\{ (\Delta y_j, \Gamma y_j) \in D \mid \Delta y_j \in [-1000, 1000], \Gamma y_j \in [0, \sigma_3] \right\},$$

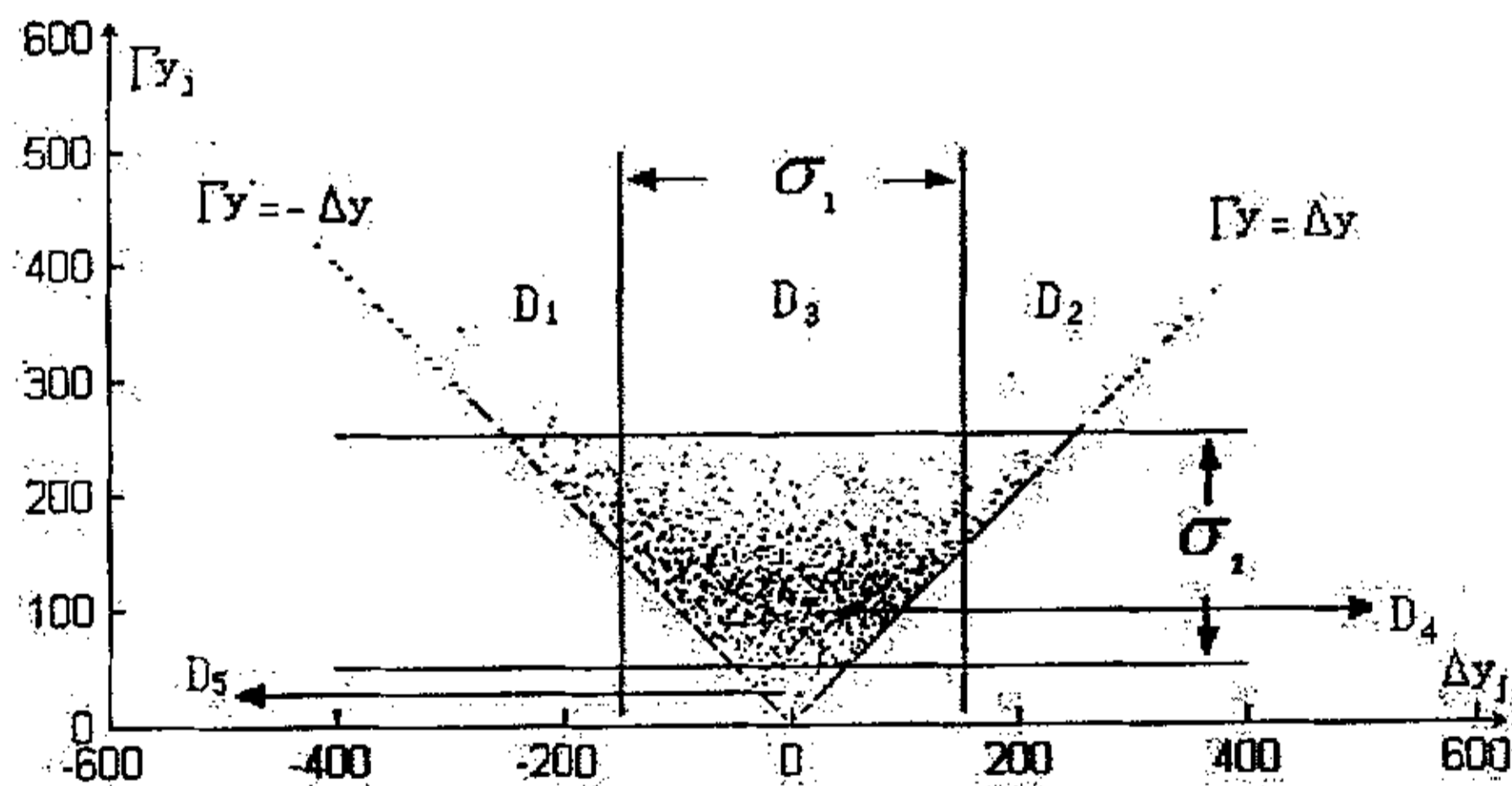


图 4.1 $(\Delta y_j, \Gamma y_j)$ 在直角坐标系中的分布情况

Fig 4.1 The Distribution of $(\Delta y_j, \Gamma y_j)$ in the Cartesian Coordinates

当 $(\Delta y_j, \Gamma y_j) \in D_1$ 部分内时, 说明该专家习惯给低分; 以此类推, $(\Delta y_j, \Gamma y_j) \in D_2$ 部分内时, 说明该专家习惯给高分; $(\Delta y_j, \Gamma y_j) \in D_3$ 部分内时, 为有时给高分, 有时给低分; $(\Delta y_j, \Gamma y_j) \in D_4$ 部分为给分正常; $(\Delta y_j, \Gamma y_j) \in D_5$ 部分为给分接近平均分。

在以上定义及评审项目分类的基础上, 设 q_j 为第 j 位专家所有评审项目编号的集合, h 表示给定的专家评审项目最小值, 即被考察专家评审项目数不得少于 h , 我们将专家分为以下集合进行研究:

$Q_1 = \{j \mid |q_j| \geq h, q_j \cap S_1(a) \neq \phi, q_j \cap S_2(b) \neq \phi, j \in I_1\}$, 表示所评审项目既有入选予以资助的, 也有落选被淘汰的专家的集合;

$Q_2 = \{j \mid |q_j| \geq h, q_j \cap S_1(a) = \phi, q_j \cap S_2(b) \neq \phi, j \in I_1\}$, 表示所评审项目中没有入选项目但有落选项目的专家的集合;

$Q_3 = \{j \mid |q_j| \geq h, q_j \cap S_1(a) \neq \phi, q_j \cap S_2(b) = \phi, j \in I_1\}$, 表示所评审项目中有入选项目但没有落选项目的专家的集合;

$Q_4 = \{j \mid |q_j| \geq h, q_j \cap S_1(a) = \phi, q_j \cap S_2(b) = \phi, j \in I_1\}$, 表示所评审项目既没有入选项目又没有落选项目的专家的集合。在这些集合中, 我们重点研究 Q_1 。

A 表示关于论域 U 上属性的非空集合, 即为衡量评审专家的标准的集合, 因此 $A = \{(P_j, r_j) \mid j \in I_1\}$, 即通过考察评审专家的命中率和成功率的大小及与给定的差异允许值考察评审专家评分是否公正与其学术水平高低。故称 $S = (U, A)$ 为评审专家信息系统。

评审专家反评估的主要计算步骤为:

- <1>. 读入数据, 并按要求整理出有效数据;
- <2>. 根据所给定义 1~6 计算各项指标, 依项目平均分大小进行排序;
- <3>. 根据实际情况给定入选和落选比例 α_1, α_2 , 确定入选和落选分数线 a, b , 得入选集合 $S_1(a)$ 、落选集合 $S_2(b)$ 及可上可下集合 $S_3(a, b)$;
- <4>. 统计每位评审专家评选的入选项目数、落选项目数和总评审项目数, 分别记为 x_{0j}, x'_{0j}, n_j ; 挑选评审项目数大于 h 项的专家为研究对象;
- <5>. 将符合条件的专家按其评分特点分为 Q_1, Q_2, Q_3, Q_4 集合;
- <6>. 对集合 Q_1 进行重点考察, 分别计算集合 Q_1 中专家的成功率和命中率;
- <7>. 给定 $\sigma_1, \sigma_2, P_j, r_j$, 考察评审专家的命中率和成功率的差异允许范围 σ_1, σ_2 之间的大小关系, 得出有关专家打分高低、是否客观的结论;

4.4 应用实例

以某学科 95、96、98、99 年的项目评审过程中的专家反评估为例。首先对原始

数据进行整理,共有 1587 位专家对 3345 个项目进行评审,每个项目由 5 位专家打分,即每个项目得到 5 项综合评分,即 $n = 3345, |O_k| = m = 5$, 第 k 个项目的评分集合为 $y_k = \{y_{kj}\}, j \in O_k$, 则这一项的平均分为 $\bar{y}_k = \sum_{j \in O_k} y_{kj} / 5$ 。按平均分的大小给项目排序。

并且根据定义 1~6 计算有关项目和专家的指标。实际给定的入选项目比例为 $\alpha_1 = 20\%$, 落选比例为 $\alpha_2 = 30\%$, 根据项目的平均分高低排序得: 入选分数线为 $a = 748.08$, 落选分数线为 $b = 575$, 则参评项目分为以下集合:

$$S_1(a) = \{k | \bar{y}_k \geq 748.08, k \in I_n\},$$

$$S_2(b) = \{k | \bar{y}_k < 575, k \in I_n\},$$

$$S_3(a, b) = I_n \setminus (S_1(a) \cup S_2(b)),$$

$$\text{其中 } |S_1(a)| = \alpha_1 \cdot |I_n| = 3345 \times 20\% \approx 670,$$

$$|S_2(b)| = \alpha_2 \cdot |I_n| = 3345 \times 30\% \approx 1004。$$

为了尽量减少随机性,我们研究评审项目数不少于 10 项的专家,即评审项目数 $h \geq 10$, 得符合条件的专家 1451 位。分别求出这些专家的评审总项目数、评审入选项目数、落选项目数,计算各位专家的成功率和命中率。则可将上述专家分为以下集合:

$$Q_1 = \{j | |q_j| \geq 10, q_j \cap S_1(a) \neq \phi, q_j \cap S_2(b) \neq \phi, j \in I_1\}, \text{ 其中 } |Q_1| = 433;$$

$$Q_2 = \{j | |q_j| \geq 10, q_j \cap S_1(a) = \phi, q_j \cap S_2(b) \neq \phi, j \in I_1\}, \text{ 其中 } |Q_2| = 144;$$

$$Q_3 = \{j | |q_j| \geq 10, q_j \cap S_1(a) \neq \phi, q_j \cap S_2(b) = \phi, j \in I_1\}, \text{ 其中 } |Q_3| = 137;$$

$$Q_4 = \{j | |q_j| \geq 10, q_j \cap S_1(a) = \phi, q_j \cap S_2(b) = \phi, j \in I_1\}, \text{ 其中 } |Q_4| = 19;$$

其中,考察各集合的特征可知, Q_2 中的专家总体上给分偏低,好打低分; Q_3 的专家总体上给分偏高,好打高分;应继续考察上述两个集合中的专家的平均差异和最大差异,做出是否任用的结论; Q_4 中的专家命中率和成功率都为零,所以,除去主观偏见的因素外,说明本身的学术水平不高或对评审规则的理解不够透彻,不擅于发现有重大科学意义的项目,也不能淘汰科学价值不大的项目,所以评审水平较差,在以后的评审活动中应慎重使用;最后重点考察 Q_1 。

Q_1 中的专家既有评审的入选项又有落选项,总体上属于打分正常,应通过综合分析入选成功率和评审落选成功率得出专家反评估结论。具体的,给定评审入选成功率和落选的成功率都为 50%, Q_1 中的专家又可分为以下子集:

$$\begin{aligned}
 d_1 &= \{j | r_j \in (0.5, 1], r'_j \in (0.5, 1], j \in Q_1, j \in I_1\}; \\
 d_2 &= \{j | r_j \in (0, 0.5], r'_j \in (0.5, 1], j \in Q_1, j \in I_1\}; \\
 d_3 &= \{j | r_j \in (0.5, 1], r'_j \in (0, 0.5], j \in Q_1, j \in I_1\}; \\
 d_4 &= \{j | r_j \in (0, 0.5], r'_j \in (0, 0.5], j \in Q_1, j \in I_1\};
 \end{aligned}$$

其中 r_j 表示 j 专家的评审入选成功率, r'_j 表落选成功率, 则若专家 $j \in d_1$, 说明这些专家评审水平较高, 主观性较小, 能够在众多项目中分清优良, 评审结果与最终实际资助结果达到相当高得一致性; 若专家 $j \in d_4$, 他们能够发现项目的优劣, 但是成功率较低, 不排除主观偏见和客观水平不高的因素; 若专家 $j \in d_2$ 或 $j \in d_3$, 说明这些专家能够客观地评审项目, 可按成功率的大小选择留用。而若入选和落选成功率都在 0.5 附近波动, 说明这些专家在评审中能够客观地评价项目, 但打分有所保留, 无法在评分过程中准确地评选出那些具有重大革新意义的项目。

通过对比基金委的评审情况, 上述结果比较理想。

4.5 结论

本文提出了几个用于专家反评估的重要指标及应用评审模型。在实际工作中, 只要给定相应的入选比列, 成功率或留用专家人数等量, 就能得到项目评审及专家反评估结果。对于超出正常评分值的专家, 判断其是否是由于革命性创新而造成非共识问题, 可组织专家进一步评审; 在排除了这类问题以后, 则大致可分为主客观两类原因: 如果客观原因造成的, 即由于该专家本身水平有限, 不能正确认识申请项目本质, 从而不能对其进行正确评价, 反映的是该专家的水平, 应待其学术水平有所提高之后, 在考虑是否继续聘用。如果是主观原因造成的, 即该专家具有较高评审水平, 但由于主观上对申请项目有偏见, 而不能对其进行合理的评价, 反映的是该专家的公正性。对于打分偏颇。有失公允的专家在以后的评审过程中予以注意, 而那些打分公正, 能够正确评审具有重大革新项目的专家, 可以继续聘用参与项目评审。

此外, 有关专家反评估的评审软件已经完成, 并在国家基金委的实际评审中投入使用。使专家的反评估工作实现办公自动化, 易于今后评审工作和专家的聘用。

参考文献

- [1] Sen A., Interpersonal Aggregation and Partial Comparability, *Econometrica*, 1970, 38:393-409
- [2] Sen A., Social Theory, In: K.J. Arrow and M.D. Intriligator, *Handbook of Mathematical Economics* III, Amsterdam: North-Holland, 1986
- [3] Bacharach M., Group Decision in Face of Difference of Opinion, *Mgmt. Sci.*, 1975,22:182
- [4] Dyer J.S., and Surlin R.K., Group Preference Aggregation Rules Based on Strength of Preference, *Mgmt. Sci.*, 1979, 25(9):22-34
- [5] Wendell R.E., Multiple Objective Mathematical Programming with Respect to Multiple Decision Makers, *Ops. Res.*, 1980,28:1100
- [6] Greenberg J., Consistent Majority Rules over Compact Sets of Alternatives, *Econometrica*, 1979,47(3):627-636
- [7] Arbel A., and Vargas L.G., Preference Simulation and Preference Programming: Robust Issues in Priority Derivation, *European J. Ops. Res.*, 993,69(2):209-222
- [8] Cho S.K., and Kang T.G., Decision Maker's Personality Type and Risk Attitude, *J. Korea Ops. Res.*, 1996, 21(2): 1-16
- [9] Sen A., *Collective Choice and Social Choice*, Princeton Univ. Press, Princeton, NJ, 1973
- [10] Keeney R.L. and Kirkwood C.W., Group Decision Making Using Cardinal Social Welfare Functions, *Management Science*, 1975, 22(4): 430-437
- [11] S. Cole, et., Peer Review and the Support of Science, *Scientific American*, Oct. 1987, Vol 237, No. 4, p37
- [12] U.S. General Accounting Office, *University Funding Information-the Role of Peer Review at NSF and NIH*, PB67-180949, p12
- [13] T.L. Saaty, *The Analytic Hierarchy Process*. McGraw Hill, Inc., 1980
- [14] Weiwu Fang, Fred S. Roberts, Zhengrong Ma, A measure of discrepancy of multiple sequences, *Information Sciences*, 137(2001), 75-102
- [15] Weiwu Fang, On a Global Optimization Problem in the Study of Information Discrepancy, *Journal of Global Optimization* 11: 387-408, 1997
- [16] Weiwu Fang, The Characterization of a Measure of Information Discrepancy, *Information Sciences* 125(2000) 207-232
- [17] Weiwu Fang, Some Properties of FDOD Function and Kullback Measure, *CSIAM'96 VII:9*, 1996

- [18] R.Rockafellar, *Convex Analysis*, Princeton University Press, Princeton,1970
- [19] Jian Ma, Qi-Dong Wang, En-Min Feng, A Measure of Discrepancy of Multiple Sequences and Applications in Evaluation of NFC, *Advanced in Systems Science and Applications*, Vol.3, 1(2003), 45-51.
- [20] A.Ullah, Entropy, Divergence and Distance Measures with Econometric Applications, *Journal of statistical planning and inference* 49, 137-162, 1996
- [21] S.B. Needleman, C.D. Wunsch, A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins, *Journal of Molecular Evolution* 19(1972) 153-170
- [22] L.Gatlin, *Information Theory and the Living System*, Columbia University Press, New York, 1978
- [23] Wang L., Jiang T., On the Complexity of Multiple Sequence Alignment, *J. Comput. Biol.*, 1994;(1):337-348
- [24] Wang L., Gusfield D., Improved Approximation Algorithms for Tree Alignment, *J. of Algorithms.*, 1997; 25: 255-273
- [25] Lipman D.T., Altschul S. F, Kececiglu J.D., A Tool for Multiple Sequence Alignment *Proc, Nat. Acad Sci. USA*, 1989; 86:4412-4415
- [26] Jiang T., E.L.Lawler, and L.Wang, Aligning Sequences Via an Evolutionary Tree: Complexity and Approximation, *Proceeding of the 26th ACM Symposium on the Theory of Computing*, 760-769 1994, to appear in *Algorithmica*
- [27] Gusfield D., Efficient Methods for Multiple Sequence Alignment with Guaranteed Error Bounds, *Bulletin of Mathematical Biology* 2:3, 459-472
- [28] Kececiglu J., The Maximum Trace Problem in Multiple Sequence Alignment, *Proceeding of the 4th Symposium on Combinatorial Pattern Matching*, Springer Verlag Lecture Notes in Computer Science, 684, 106-119, 1993
- [29] Hein J., Unified Approach to Alignment and Phylogenies, *Method in Enzynology* 188, 626-645, 1990
- [30] Pe vzuer P.A., MultipleAlignment, Communication Cost, and Graph Matching *SIAM, Journal on Applied Mathematics* 52:6,1763-1779, 1992
- [31] Arrow K.J., *Social Choice and Individual Values*, New York: Wiley, 1951.2nd, ed., New York: Wiley, 1963
- [32] Dr Fiona Q. Wood. *The Peer Review Process/Commissioned Report No. 54.* Canberra : Australian Government Publishing Service, 1997:
- [33] 吴述尧著, 同行评议方法论, 科学出版社, 1996
- [34] A. 乔伊科夫等, 多目标决策分析及其在工程和经济中的应用, 航空工业出版社, 1987
- [35] 姜启源, 数学模型, 北京: 高等教育出版社, 1987

- [36] 梅家骝, 辜介田, 多目标规划强有效解的充分性条件, 高等学校计算数学学报, 1983, 5(4): 328-335
- [37] 王汝华, 关于群体 AHP 方法探讨, 全国第一届层次分析法学术会议论文, 1987
- [38] 王莲芬, 许树柏, 层次分析法因论, 中国人民大学出版社, 1990
- [39] 赵焕臣, 许树柏, 和金生, 层次分析法——一种简易的新决策方法, 科学出版社, 1986
- [40] 王雪华, 秦学志, 杨德礼, AHP 中判断矩阵一致性修正的模式识别法, 系统工程理论与实践, 1997, 7(1): 56-59
- [41] 于英川, 弱有效解的分层与评价序模型, 中国管理科学, 1994, 3
- [42] 商玉生, 十岁的思考, 中国言实出版社, 1997
- [43] 美国国家研究委员会、美国科学基金会著, 阮祖启等译, 科学质量的评估, 1994
- [44] 于维栋主编, 科学基金制——科学研究永葆活力的催化剂, 科学技术文献出版社, 1994
- [45] 吴信东编著, 专家系统技术, 电子工业出版社, 1988
- [46] 国家自然科学基金委员会, 国家自然科学基金资助要览(1986-1990), 北京大学出版社, 1992
- [47] 国家自然科学基金委员会政策局编译, 加拿大自然科学与工程研究理事会咨询委员会的调查报告, 1992
- [48] 德意志研究联合会著, 汪平忠等译, 德意志研究联合会科学研究资助任务和财政计划(第八卷), 科学出版社, 1994
- [49] 沈世镒, 多重序列比对的信息度量准则, 工程数学学报, Vol. 19, No.3, 1-9, 2002
- [50] 张文修、吴伟志、梁吉业、李德玉编著, 粗糙集理论与方法[M], 科学出版社, 2001.
- [51] 程建刚, 基金评审中的异同性分析、专家反评估及软件研制, 大连理工大学硕士学位论文, 2001
- [52] 王其冬、程建刚、冯恩民, 划分临界项目的一种多目标决策方法, 决策科学理论与应用, 全国决策科学与多目标决策研讨会论文集, 香港卓越出版社, 2000, 25-28.
- [53] 冯恩民、王其冬、李文娟, 评分区间方法、性质及应用, 温州大学学报, Vol.15, No.3, 2002, 1-3
- [54] 王其冬、冯恩民、李文娟, 项目评价双层规划、性质及应用, 运筹与管理, Vol.11, No.3, 2002, 44-48
- [55] 王其冬、武佩珍、程建刚、冯恩民, 层次分析法在国家自然科学基金项目评审中的应用, 系统工程理论与实践, Vol.21, No.2, 2001, 119-123
- [56] 冯英俊, 多目标优化问题的 Fuzzy 解, 科学通报, 1981, 26(17): 1028-1030
- [57] 罗中, 朱志光, 模糊系统评价及应用, 培训与研究——湖北教育学院学报(自然科学版), 1999, Vol. 2, 5-7
- [58] 胡毓达, 群体决策的偏差度分析, 运筹学学报, 1982, (2): 77-83

- [59] 胡毓达, 群体决策理论和方法, 上海交通大学讲义, 1988
- [60] 胡毓达, 向量空间的较多序类, 数学年刊, 1990, 11A(3): 319-378
- [61] 胡毓达, 实用多目标最优化, 上海: 上海科学技术出版社, 1990
- [62] 胡毓达, 多目标规划有效性理论, 上海: 上海科学技术出版社, 1994
- [63] 胡毓达, 多目标群体决策的优先数法, 贵州大学学报, 1994, 11(4): 193-198
- [64] 胡毓达, 群体多目标决策的联合有效解类及其最优性条件, 上海交通大学学报, 1999, 33(6): 1-5
- [65] 孙明玺. 预测与评价. 浙江教育出版社, 1986
- [66] 许国志, 顾基发, 车宏安. 系统科学. 上海: 上海科技教育出版社, 2000
- [67] 《中外专家论课题制》编委会. 中外专家论课题制. 北京: 中信出版社, 2003
- [68] 胡毓达, 丁鸿生, 群体决策的惩罚二次评分方法, 上海交通大学学报, 1999, 33(6): 6-9
- [69] 邱菀华等. 项目管理学. 北京: 科学出版社, 2001
- [70] 马剑, 王其冬, 冯恩民, 基金评审中的归一化分析方法及应用, 运筹与管理, 待发表

附录:

硕士研究生期间发表论文情况

1. Jian Ma, Qi-Dong Wang, En-Min Feng, A Measure of Discrepancy of Multiple Sequences and Applications in Evaluation of NFC, Advanced in Systems Science and Applications, Vol.3, 1(2003), 45-51.
2. 马剑, 王其冬, 冯恩民, 基金评审中的归一化分析方法与应用, 运筹与管理, 已接收, 待发表。
3. 马剑, 王其冬, 冯恩民, 基金评审中的专家反评估分析, 已投稿。

致谢

在近三年的硕士学习阶段中，我首先要感谢我的导师冯恩民教授。冯老师高尚的品质、渊博的知识、严谨的治学态度和不断进取的精神一直感染并激励着我，并将使我终生受益。我能够在研究生学习期间取得优异的成绩并顺利完成学业得益于冯老师对严格要求。

本文就是在导师冯恩民教授的精心指导下完成的。在导师的耐心教导下，不但使我学到许多的知识，更重要的是为将来的科研与学习工作打下了良好的基础。在此向导师致以崇高的敬意和衷心的感谢。祝他在工作中取得更好更新的成果。

衷心感谢夏尊铨教授，唐焕文教授在平日学习中的大力帮助和指导，并给本文提出了宝贵意见；感谢张立卫教授，张宏伟教授，蔡晶老师，在学习上给了我很大帮助。

感谢中科院的方为武教授，本文中的部分结果都是基于方教授的一些学术成果完成的；对国家自然科学基金委员会的王其冬处长表示谢意，王处长提供了大量的实际数据，在研究过程中凭借实际工作经验提出了许多宝贵的算法修改意见，使整个课题得以顺利进行，并促成了与基金委的合作；虽与基金委的同志们未曾谋面，但我们的合作是愉快的。

我特别要感谢我的父母对我无微不至的照顾和无私的奉献，家里是我永远坚强的后盾和温暖的港湾，学业的完成也蕴含着他们殷勤的期盼和无尽的牵挂，对此我除了在以后的工作学习中更加勤奋努力外无以回报；衷心感谢我的爱人朱春钢，在学习上对我的帮助和促进，在生活上对我的体贴和谦让，在精神思想上的鼓励和支持，我的成绩里有他的一半。

我还要感谢钱伟懿、罗诚新、张序、谭欣欣、王丽艳等师兄、师姐在平日里的指导帮助；丛龙飞、李文娟、韦艳华等已毕业的老同学，同级的王宗涛、宫昭华、徐贡贤，同寝的黄玉娟、邹美凤、赵晶等朋友在共同生活这三年里的关爱。

感谢我们班级，以及所有朋友们，有你们的关心，我才会有今天的成绩。

最后向以上的以及匆忙之间没有提到的所有人衷心表示感谢！

大连理工大学学位论文版权使用授权书

本学位论文作者及指导教师完全了解“大连理工大学硕士、博士学位论文版权使用规定”，同意大连理工大学保留并向国家有关部门或机构送交学位论文的复印件和电子版，允许论文被查阅和借阅。本人授权大连理工大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，也可采用影印、缩印或扫描等复制手段保存和汇编学位论文。

保密，在____年解密后适用本授权书。

本学位论文属于

不保密.

(请在以上方框内打“√”)

作者签名: 马刚

指导导师签名: 冯思民

2004年3月10日