

基于贪心搜索的分组项目评审专家遴选方法^①



曹滔宇¹, 熊永平¹, 史梦洁², 徐会芳², 谷纪亭³

¹(北京邮电大学 网络与交换技术国家重点实验室, 北京 100876)

²(中国电力科学研究院有限公司 人工智能应用研究所, 北京 102209)

³(国网浙江省电力有限公司 经济技术研究院, 杭州 310001)

通讯作者: 熊永平, E-mail: ypxiong@bupt.edu.cn

摘要: 在科技项目评审环节中, 往往每组同时有多个项目和多位评审专家, 其中每个项目都有其覆盖的专业领域, 而每位专家又有其所研究的专业领域, 如何科学且自动地根据待评审科技项目所涵盖的专业领域, 从候选专家库中找出合适的评审专家组合团体具有很实际的研究意义. 对此本文提出了一种基于贪心算法的科技项目评审专家多重匹配模型, 该模型应用于已经建立起关联的“项目-领域”与“专家-领域”两个相关性矩阵上, 通过分别计算科技项目及评审专家团体所对应于专业领域上的离散分布, 并利用合适的评价函数综合衡量待审科技项目与评审专家组合之间的匹配度, 最终求出最优的评审专家组合来作为该期科技项目的最终评审专家团体. 本文通过使用电力行业数据集进行多次实验, 结果表明该模型能有效地进行科技项目评审专家的匹配, 并具有较高的合理性与准确性, 在解放人力成本提高评审工作效率的同时, 也杜绝了传统人工遴选专家过程中所出现纰漏以及评审不公现象的发生.

关键词: 专家遴选; 多重匹配模型; 贪心算法; 离散分布

引用格式: 曹滔宇, 熊永平, 史梦洁, 徐会芳, 谷纪亭. 基于贪心搜索的分组项目评审专家遴选方法. 计算机系统应用, 2021, 30(4): 168-174. <http://www.c-s-a.org.cn/1003-3254/7856.html>

Evaluation Experts Selecting Strategy Based on Greedy Algorithm

CAO Tao-Yu¹, XIONG Yong-Ping¹, SHI Meng-Jie², XU Hui-Fang², GU Ji-Ting³

¹(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

²(Artificial Intelligence Applications Institute, China Electric Power Research Institute Co. Ltd., Beijing 102209, China)

³(State Grid Zhejiang Economic Research Institute Co. Ltd., Hangzhou 310001, China)

Abstract: Multiple technology projects in a group are evaluated by several experts. Each project covers certain fields and each expert has his/her own advantageous fields. Thus, it is a significant challenge to scientifically and automatically group experts in suitable fields of relevant projects from a large number of candidates. This study proposes a multi-match model GIS of evaluation experts for technology projects based on the greedy algorithm. This model is applied to the two associated “project-field” and “expert-field” correlation matrices. Specifically, it separately calculates the discrete distribution of projects and experts in each field. Then it uses a proper evaluation function to measure the project-expert match and finally obtains the optimal team. The experiments based on the data sets in the power industry show that this model can match the experts with the technology projects and thus has high rationality and accuracy. It avoids the careless mistakes and unfairness in traditional expert selection while reducing human costs.

Key words: expert selection; multi-matching model; greedy algorithm; discrete distribution

① 基金项目: 国家电网公司总部科技项目 (SGTYHT/18-JS-206)

Foundation item: Science and Technology Project of State Grid Headquarters (SGTYHT/18-JS-206)

收稿时间: 2020-08-02; 修改时间: 2020-08-28; 采用时间: 2020-09-08; csa 在线出版时间: 2021-03-30

近些年来,从国家到地方,各领域各企事业单位都投入大量人力、物力开展科研和产业发展,在项目招标、申报、实施、验收等阶段,都离不开遴选出相关专家进行评审.随着国家经济发展和科技管理水平的提高,各级单位每年立项的科技项目数量快速增加,以自然科学基金委的科研项目为例,仅在2019年国家自然科学基金项目申请集中接收期间,国家自然科学基金委员会共计接收项目申请达到了240711项,而国家电网总公司每年立项的科技项目也有300余项,可见在短时间内对科技项目展开评审工作的任务量是极为繁重的.

随着项目管理流程的日益规范化,如今在项目的立项论证、中期检查、成果验收、成果评价等多个环节都需要组织相关专家进行会审.为了提高项目评审的效率,目前各公司及单位普遍采用分组评审的策略,但由于技术领域的日益细分和跨学科技术的广泛应用,导致从业务和管理维度进行分组评审往往会使每个组的项目跨越较多的技术领域,因此必须遴选出一组契合这些技术领域的专家才能实现对该组项目的有效评审.

目前传统专家遴选的方式普遍由人工完成,在成千上万的专家库中检索出合适的专家组合十分具有挑战性.由于候选专家排列组合的方案数巨大,因此在有限时间内找出合适的专家团体变得十分困难,亟需找到一种能合理为科技项目评审工作匹配出评审专家组合的解决方案,以克服人工遴选专家方式所带来的种种弊端.

考虑到在分组项目评审过程中往往同时有多个项目和多位专家,应为科技项目选出有限数量的评审专家,使得这些专家组合成的评审团体可以较好地契合项目所涉及的各个相关领域.本文首先将该问题建模为一个典型的组合优化问题,通过将项目和专家映射到技术领域建立起科技项目和评审专家所对应于专业领域上的离散分布,进而基于余弦相似度函数来量化评价该组科技项目和评审专家组之间的匹配度.鉴于该类组合优化问题往往在多项式级时间复杂度上无法有效求解,因此本文提出了基于贪心迭代搜索的GIS算法,该算法主要采用了多轮迭代搜索最优部分解来组合形成全局最优解的策略以实现找出最优专家组合的目的.本文最终将GIS算法分别在国家电网专家库及其历史立项科技项目真实数据集上进行了实验验证,并研究了专家库大小、领域数量、项目数量等不同因素对算法的影响,结果表明本文提出的GIS算法能在较短的时间内找到较优的评审专家组合方案.

1 相关工作

近些年来,已有不少相关学者对专家推荐问题进行过深入的研究,其中也不乏一些十分具有代表性的专家推荐方法,目前的专家推荐方法大体可分为两类:第一类是针对专家独立推荐的研究,其主要思想是在推荐一位或多位专家时,主要考虑将专家以独立个体的方式推荐得出,即每位专家都与当前待评审项目存在一定程度上的强相关关系,但不考虑所推荐专家形成组合后的整体情况;第二类则是针对专家组合推荐的研究,一般是在限定若干约束条件的情况下推荐出由多位专家组成的专家团体,该专家团体实现了对于当前待评审项目能达到组合最优的评审效果,但其中每位专家则不必精通每个待评审项目,下面简要介绍在基于上述两种研究思路下当前已有的相关工作.

基于专家独立推荐的研究:文献[1-5]均通过提出一种科技项目评审专家推荐系统模型,该模型在挖掘文本信息的基础上运用关键词提取、特征权重计算等相关算法,得到科技项目的多维度特征信息,然后通过计算其与专家在词条上的相似度,并综合专家参评项目经验及专家业务能力等因素,最终使用基于内容推荐、协同过滤推荐以及专家评分加权因子相融合的混合推荐模型,计算出每位专家的综合评分,再根据设定的阈值以及推荐指数从高到低产生推荐专家名单实现了对科技专家的高效遴选;文献[6-11]均通过提出了一种基于文本分类模型的方式来实现专家自动推荐的效果,主要借助有监督或无监督的方式建立起专家知识模型来判断出评审专家的主要研究领域和评审项目的专业领域,再将评审项目的专业领域与评审专家的研究领域按相似性自动匹配,最终达到对评审专家精准推荐的目的;文献[12-14]主要提出了一种基于主题模型的评审专家协同推荐方法,即借助隐含狄利克雷分布模型构建主题特征空间,并利用特征提取算法分别获得项目文档与专家文档的主题特征向量,计算项目与专家主题特征向量的相关度并取项目相关度较高的专家作为推荐结果.

基于专家组合推荐的研究:文献[15-18]主要通过将项目与专家抽象为二分图网络模型,由网络节点的关联性出发,提出了一种基于相似度传播的复杂网络节点匹配方法,通过借助图论中的KM算法、最大流匹配算法等方式实现节点间分组匹配的目的,最终设计出项目与专家的多重匹配算法;文献[19]提出一种

基于语义挖掘的科技项目评审专家智能推荐方法,为一个或多个项目自动推荐生成候选专家列表;文献[20]提出一种适用于求解通用最大权完美匹配的智能优化方法,该方法能自适应地从改进的离散粒子群策略以及模拟退火策略中选择适用于当前演化过程的有效策略,并在保持种群稳定进化的同时促使种群快速收敛。

综上所述,对于独立推荐专家的方式来说,其主要关注点为挖掘专家与项目的知识信息,并使用混合加权的方式来计算每个专家的综合得分,最终基于该分数实现推荐。但这种方式难以保证由多个高分专家组合而成的团体也能契合实际项目需求,具体来说,当这些高分专家均仅擅长于特定领域且彼此相似时,那么此时的专家组合虽能满足每位专家最优,但却无法保证该组合整体足够适合于当前项目的评审需求。而基于专家组合推荐的方式则避免了这种情况的发生,其实现方式主要是将项目和专家的关联关系抽象成二部图网络模型,进而考虑使用如完美匹配、最大流匹配等图论算法实现专家遴选,但这类算法常常由于具有较高的时间复杂度而难以将其应用到大规模数据集上。为此基于上述考虑,本文设计了一种抛弃传统二部图网络结构的专家组合推荐策略,并最终将其运用到较大的数据集上实现了科技项目与评审专家的多重匹配。

2 问题定义

本文在考虑实际分组项目评审的情况下将该匹配问题进一步抽象描述如下:

当前共有 n 个待审批科技项目,用集合 $P = \{P_1, P_2, \dots, P_i, \dots, P_n\}$ 来表示;专家库中共有 m 位专家,用集合 $E = \{E_1, E_2, \dots, E_i, \dots, E_m\}$ 来表示;项目集合 P 与专家集合 E 共涉及 l 个专业领域,用集合 $F = \{F_1, F_2, \dots, F_i, \dots, F_l\}$ 来表示。

由于任意一个科技项目 P_i 都与若干专业领域有一定的相关性,这里记矩阵 W^{PF} 来表示项目集合 P 与领域集合 F 的相关性矩阵,其中 W^{PF} 中的第 i 行 j 列的元素 w_{ij} 表示科技项目 P_i 与专业领域 F_j 之间的相关度,特别地,当 w_{ij} 的值为零时表示科技项目 P_i 与专业领域 F_j 之间没有关联关系。

$$W^{PF} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1l} \\ w_{21} & w_{22} & \dots & w_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nl} \end{bmatrix} \quad (1)$$

同理,因为任意一名候选专家 E_i 都有其擅长的研究领域,所以仍可以得到矩阵 W^{EF} 来表示专家集合 E 与领域集合 F 的相关性,矩阵 W^{EF} 表示如下:

$$W^{EF} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1l} \\ w_{21} & w_{22} & \dots & w_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ w_{m1} & w_{m2} & \dots & w_{ml} \end{bmatrix} \quad (2)$$

假设每个科技项目 P_i 都有其关联的专业领域 $F^{P_i} = \{F_{x_1}, F_{x_2}, \dots\}$, 对应于 W^{PF} 中第 i 行数据 $W_i^{PF} = (w_{i1}, w_{i2}, \dots, w_{il})$, 那么对于当前待评审的项目集合 $P = \{P_1, P_2, \dots, P_i, \dots, P_n\}$ 来说,该组项目所关联的专业领域 F^P 可表示为 $F^P = \bigcup_{i=1}^n F^{P_i}$, 对应到矩阵 W^{PF} 即可得到能反映出项目集合所关联专业领域的离散分布,记该离散分布为 $D(P)$, 则其计算方式可定义为:

$$D(P) = \sum_{P_i \in P} W_i^{PF} = \left(\sum_{i=1}^n w_{i1}, \sum_{i=1}^n w_{i2}, \dots, \sum_{i=1}^n w_{il} \right) \quad (3)$$

同理,若对已选出的 k 位专家所组成的集合 $E^{(k)} = \{E_{x_1}, E_{x_2}, \dots, E_{x_k}\}$ 进行考虑,其中每位专家 E_i 所擅长的专业领域 $F^{E_i} = \{F_{x_1}, F_{x_2}, \dots\}$, 那么同样可以找到能反映该专家团体 $E^{(k)}$ 主要研究方向的离散分布 $D(E^{(k)})$, 其计算方法如下所示:

$$D(E^{(k)}) = \sum_{E_x \in E^{(k)}} W_x^{EF} = \left(\sum_{i=x_1}^{x_k} w_{i1}, \sum_{i=x_1}^{x_k} w_{i2}, \dots, \sum_{i=x_1}^{x_k} w_{il} \right) \quad (4)$$

为了定义所遴选出的专家与待评审项目的匹配程度,本文提出了一种评价函数 $S(P, E^{(k)})$ 来衡量当前选出的专家子集 $E^{(k)}$ 对项目集合 P 的匹配度。该评价函数能够满足:当项目集合 P 所涉及专业领域与专家子集 $E^{(k)}$ 研究方向足够契合时, $S(P, E^{(k)})$ 始终能给出较高的评价,反之则会给出较低的评价,这样即可认为选定专家子集 $E^{(k)}$ 来评审该组科技项目是比较合适的。

因此可以借助前文所定义的离散分布 $D(E^{(k)})$ 来表示专家子集 $E^{(k)}$ 的专业能力分布, $D(P)$ 用来表示项目集合 P 所涉及到的研究领域分布,这样通过将两者信息映射到共同的专业领域维度上之后,便可以进一步分析 $D(E^{(k)})$ 与 $D(P)$ 两个离散分布间的匹配度。显然当两个离散分布越“相似”时匹配度应当越高,但考虑到用于衡量 $D(E^{(k)})$ 和 $D(P)$ 两个离散分布相似性的方式有很多,如基于欧氏距离、交叉熵、余弦相似度等函数,然而对于描述了专家子集 $E^{(k)}$ 专业能力和项目集合 P 研究

领域的两个离散分布来说, $D(E^{(k)})$ 和 $D(P)$ 之间的差异不应受到其具体数值大小的影响, 而应该侧重关注于两分布间整体趋势及结构上的相似性, 那么选用余弦相似度来定义此需求下两种离散分布的相似性是较为合适的. 因为根据余弦相似度函数的特性可知, 当把两个离散分布映射成高维空间上的向量后, 此时这两个向量的相似性将不再受到自身模值的影响, 而仅仅取决于其夹角的大小. 反映到离散分布上而言, 只有当两个分布的整体趋势及结构足够相似时, 即使两个分布之间具体数值可能相差若干倍, 但在余弦相似度函数的度量下, 仍会认为这两个离散分布是相似的, 这样也就限制了评价函数将侧重关注专家子集 $E^{(k)}$ 的所包含的主要研究领域与项目集合 P 所涉及的研究方向的契合性, 以便能保证选定该专家团体来评审当前科技项目是完全合适的, 而若采用如欧氏距离、交叉熵等作为评价函数时则无法满足此项特性. 故综合上述考虑, 本文最终定义用于衡量当前选出的专家子集 $E^{(k)}$ 与项目集合 P 之间匹配度的评价函数 $S(P, E^{(k)})$ 为:

$$S(P, E^{(k)}) = \cos[D(P), D(E^{(k)})] = \frac{D(P) \cdot D(E^{(k)})}{|D(P)| |D(E^{(k)})|} \quad (5)$$

离散分布间结构相似性的度量方式如图1.

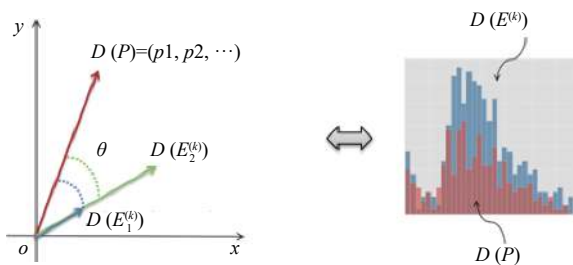


图1 离散分布间结构相似性的度量方式

3 贪心迭代搜索算法

通过上一节的定义, 显然能计算出任意一组科技项目与评审专家集之间的匹配度大小, 那么 $E^{(k)}$ 便可以通过枚举 E 的所有 k 元素子集并代入评价函数 $S(P, E^{(k)})$ 中以找到最优的匹配方案. 但这样做其实在实际应用中是无法实现的, 因为通过穷举集集合 E 的所有 k 元素子集 $E^{(k)}$ 其解的数量便高达 C_m^k 种, 而现实评审状况则往往是专家库内候选评审专家数目 m 是比较大的, 同时也需要选出一定数量的专家构成最终的评审专家团体, 那么上述方案将无法在可接受的时间范围内求解, 对

此本节将介绍一种贪心迭代搜索算法 (Greedy Iterative Search, GIS) 以实现最优专家组合的高效遴选.

假设本组待审的科技项目集合记为 P , 候选专家库内所有评审专家集合记为 E , 最终需要在 E 中匹配到一个包含 k 名评审专家的组 $E^{(k)}$ 来完成本期科技项目的评审工作. GIS算法的主要思想则是找出某个专家团体 $E^{(k)}$ 使得 $S(P, E^{(k)})$ 最大, 在保证当前所选出的评审专家团体能达到较高匹配度的前提下, 算法每轮都会从未选择的专家库中挑选出若干名专家加入到当前的评审专家团体中, 并从中删去评价较低的专家组合方案, 下一轮将继续在本轮更新后的解集中继续加入更多的专家实现评审团体的扩充, 以此类推不断迭代直至产生出若干组评价较高且人数符合预期的评审专家组合, 最终GIS算法将在该集合中选出最优的专家团体 $E^{(k)}$ 来作为其所找出的评审专家团体.

由此定义GIS算法的具体实现步骤如算法1.

算法1. 贪心迭代搜索算法

- 1) 定义搜索参数 $topK$, 初始化当前解集集合 $G^0 = \{E_1^{(0)}, E_2^{(0)}, \dots, E_{topK}^{(0)}\}$;
- 2) 遍历解集集合 G , 对每个专家团体 $E_i^{(t)}$ 尝试加入专家 $e_i (e_i \notin E_i^{(t)})$, 使得发生 $E_i^{(t)} \rightarrow E_i^{(t+1)}$ 的转变, 并对当前得到的所有新专家团体 $E_i^{(t+1)}$ 计算 $S(P, E_i^{(t+1)})$, 并取其中评价最高的 $topK$ 组加入到集合 G^{t+1} 中;
- 3) 重复步骤2) 使得所有专家团体 $E_i^{(t)}$ 都求出相应的 $E_i^{(t+1)}$ 并将其全部加入集合 G^{t+1} 中, 最终将得到 $G^{t+1} = \{E_{11}^{(t+1)}, E_{12}^{(t+1)}, \dots, E_{topK}^{(t+1)}, \dots, E_{topK1}^{(t+1)}, E_{topK2}^{(t+1)}, \dots, E_{topKtopK}^{(t+1)}\}$;
- 4) 根据 $S(P, E^{(t+1)})$ 的评价进一步削减集合 G^{t+1} 的大小, 使该解集集合所包含可能解的数量仍为最优的 $topK$ 组;
- 5) 至此由上述步骤已完成了一轮 $G^t \rightarrow G^{t+1}$ 的转变, 算法将继续迭代直到产生 $G^0 \rightarrow G^1 \rightarrow \dots \rightarrow G^k$;
- 6) 在集合 G^k 中找出能使 $S(P, E^{(k)})$ 评价最高的专家团体 $E^{(k)} = \text{Max}(G^k)$, 此专家团体 $E^{(k)}$ 即为GIS算法的最终输出.

分析GIS算法的执行流程可知, 该算法的主要运算成本集中在步骤2)~5)上, 其中步骤2)将会迭代 $topK$ 次, 步骤3)和步骤4)迭代 m 次, 步骤5)迭代 k 次, 故GIS算法整体的平均复杂度为 $O(topK \times m \times k)$, 该复杂度远小于枚举法的时间复杂度 (约为 $O(m^k)$). 考虑到在实际的分组项目评审需求背景下, 一般 m 的范围是 10^4 量级, k 的实际取值最大不会超过50, $topK$ 的可选区间亦一般不超过 10^2 量级, 故本文提出的GIS算法在极端条件下的运算成本约为 $10^7 \sim 10^8$ 次, 这在目前的计算设备下普遍可以在分钟内完成, 已经具有一定的实际可行性.

4 实验分析

为了验证GIS算法的有效性, 本文使用了真实的

电力行业科技项目评审数据集进行实验。图 2 是 GIS 算法的搜索过程。该数据集共包含有 8364 名电力行业资深专家, 涉及 127 个电力行业相关研究领域, 项目数据包含过去 3 年内国家电网总公司立项的科技项目共 912 项, 其中科技项目共分为 90 组, 每组包含的项目从 5 个到 20 个不等, 每个项目所涉及的专业领域也均在上述 127 个电力行业研究领域之内。本节将通过使用该数据集来测试 GIS 算法对于解决专家遴选问题的表现, 并结合蛮力搜索算法、RandomSelect 算法和 GradualSubtract 算法作为 baseline 构成对比实验以综合分析出 GIS 算法的有效性。其中 RandomSelect 算法的基本思想是每次从专家库中随机挑选一名专家, 使得该专家加入当前专家团体后, 专家团体的评价分数能得到提升, 依此原则不断地挑选出指定数目的专家即可; GradualSubtract 算法的主要思想则是每次都会从当前未被选择的专家中找出一名研究领域最多覆盖于当前项目的专家, 并将其加入到评审专家团体中, 然后从项目中删除这些技术领域, 接下来再重复地寻找下一位专家, 直至构建出最终的评审专家团体。

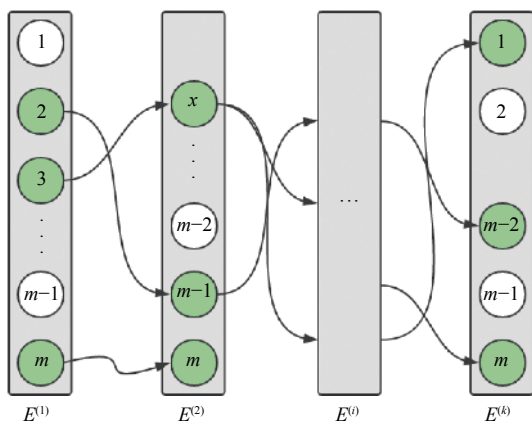


图 2 GIS 算法的搜索过程

4.1 蛮力搜索算法的可行性

本节通过使用蛮力搜索算法 (Brute Force Search, BFS) 构成对比实验, 分析其与本文提出的 GIS 算法应用在真实场景下的可行性。鉴于 BFS 算法在专家总量稍大时便会带来巨大的耗时, 所以本次实验仅选取了 150 位专家为全部候选专家, 测试其在面对包含 1 至 10 个科技项目的评审工作时, BFS 算法和 GIS 算法所能达到的匹配度及耗时情况。特别地, 由于分组评审工作中所需的评审专家数目与其所包含的科技项目数目一般为 1:1 配置, 所以本章所进行的实验也将默认采取

这种设定。

图 3 和图 4 展示了在逐渐改变实验中每组评审工作中所包含的科技项目数量的情况下, BFS 算法和 GIS 算法各自的耗时及匹配度表现。不难发现当采用 BFS 算法后, 虽候选专家的总量仅有 150 个, 但面对包含 4 个科技项目的评审需求时, 便难以在可接受的时间尺度上找出一种合适的评审专家组合方案; 相比之下, 本文提出的 GIS 算法则能保证在尽量少的耗时下达到与 BFS 算法一致的匹配度表现, 可见该算法有处理较大数据集的潜力, 能应用于现实情景下的实际需求。

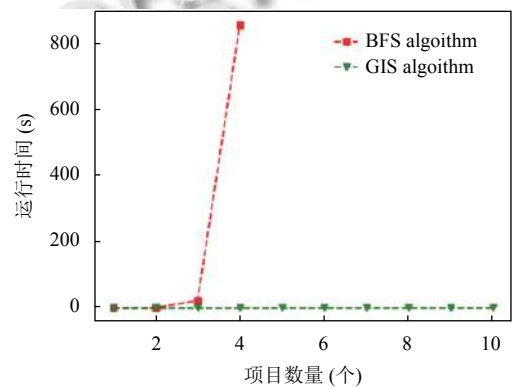


图 3 BFS 算法和 GIS 算法的耗时对比

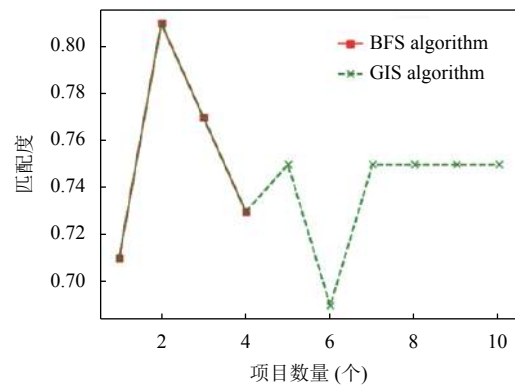


图 4 BFS 算法和 GIS 算法的匹配度对比

4.2 专家库大小的影响

本节将测试在逐渐改变专家库大小的情况下, 分析 GIS 算法对于包含 20 个科技项目的评审工作需求时的表现情况, 图 5 和图 6 记录了实验过程中 RandomSelect 算法、GradualSubtract 算法和 GIS 算法的匹配度及耗时表现。通过对比不难分析出当专家总量逐渐增大时, GIS 算法所找出的专家组合方案的匹配度会逐渐升高, 直至专家总量达到 2000 左右时趋于稳定; 相比之下,

RandomSelect 算法与 GradualSubtract 算法的匹配度表现则较差,其中 RandomSelect 算法的匹配度一直处于 0.5 到 0.6 的范围内上下波动,而 GradualSubtract 算法的匹配度虽在不断升高,但其提升速度及稳定上限相比于 GIS 算法仍有一定差距。

考虑算法的运行时间而言,当 K 值为 30 时 GIS 算法的耗时最多,而将 K 值下调至 5 后 GIS 算法的整体耗时便能显著下降。值得注意的是,虽然此时 GIS 算法的整体耗时仍略多于 RandomSelect 算法和 GradualSubtract 算法,但不难发现其匹配度表现已经有了较大幅度的提升。

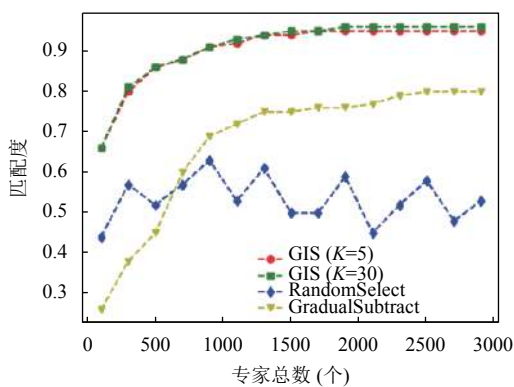


图5 改变专家库大小时 GIS 算法的匹配度表现

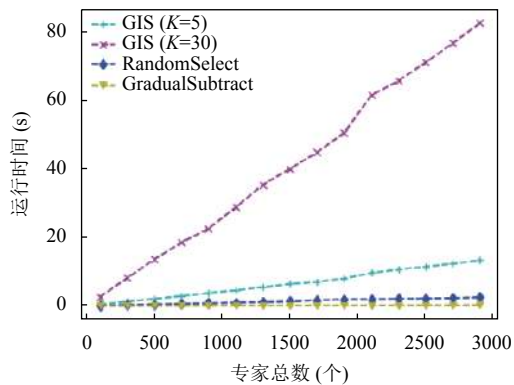


图6 改变专家库大小时 GIS 算法的耗时表现

4.3 领域数量的影响

本节将通过调整每组评审工作中所包含的 20 个科技项目所涉及专业领域的数目,探究领域数量对于本文所提出的 GIS 算法的影响,如图 7 和图 8。由实验中匹配度曲线和耗时曲线不难看出,当逐渐增大项目所涉及的领域数量后,RandomSelect 算法、GradualSubtract 算法和 GIS 算法的耗时曲线基本处于稳定状态,虽个别情况下有小范围的波动,但总体上这 3 种算法的耗时表现均不会受领域数量变化的影响。

同时根据匹配度曲线亦可以发现,RandomSelect 算法的匹配度表现会随着领域数量的增多而出现明显下降,GradualSubtract 算法的表现则相对稳定,其匹配度曲线在发展趋势上并未出现明显变化。相比之下,本文提出的 GIS 算法的表现最好,其匹配度曲线始终能保持在较高位且整个过程中十分稳定,所以本实验证实了通过改变领域数量不会对 GIS 算法的匹配度表现产生根本性影响。

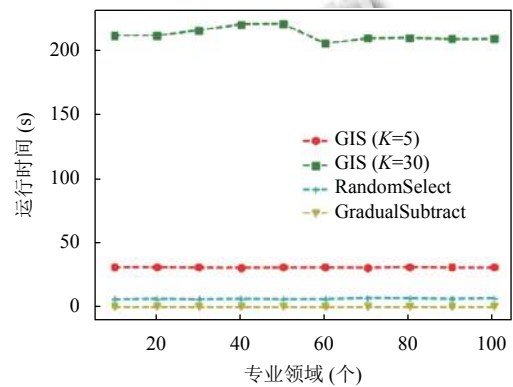


图7 改变领域数目时 GIS 算法的耗时表现

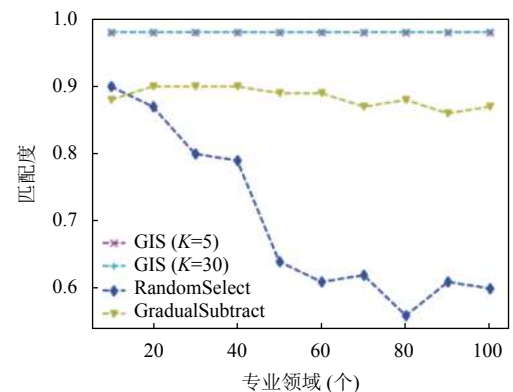


图8 改变领域数目时 GIS 算法的匹配度表现

4.4 项目数量的影响

本节最后将通过改变每期评审工作中所包含科技项目的数量测试其对 GIS 算法的影响,图 9 和图 10 记录了实验过程中在逐渐增大项目数量后 RandomSelect 算法、GradualSubtract 算法和 GIS 算法的表现。不难看出当项目数量逐渐增大时,RandomSelect 算法和 GradualSubtract 算法的耗时相对较少,而 K 值为 5 的 GIS 算法耗时则略多,且随着科技项目数量的增加, K 值设置的越大 GIS 算法的耗时增长越为迅速。

同样对比这 3 种算法的匹配度曲线可知, GIS 算法相比于 RandomSelect 算法和 GradualSubtract 算法

的表现更好,且增大 K 值后 GIS 算法的表现仍会有小幅提升.其中 GradualSubtract 算法的匹配度表现最高时可达 0.88 的匹配度,而 RandomSelect 算法最高时仅达到 0.66 的匹配度,且其整体表现较不稳定;相比之下 GIS 算法的匹配度表现则更加稳定且优异,其匹配度始终能保持在 0.95 左右.

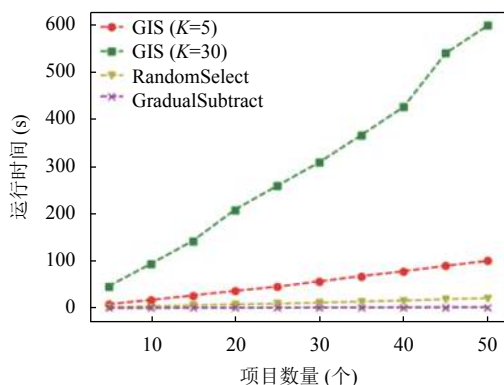


图9 改变项目数量时 GIS 算法的耗时表现

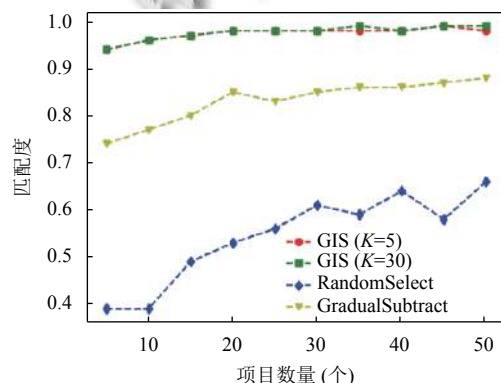


图10 改变项目数量时 GIS 算法的匹配度表现

5 结束语

本文提出了在分组项目评审的背景下求解最优评审专家组合的 GIS 算法,该算法主要通过约束贪心算法的搜索空间,多轮迭代后找出契合本期项目评审需求的专家团体.通过借助电力行业数据集对该分组项目评审专家遴选问题进行实验分析,结果表明本文提出的 GIS 算法在计算耗时和计算效果上均有较好的表现,可以将其应用到实际的科技项目评审工作中.

参考文献

1 胡斌.科技项目评审专家推荐系统的研究与实现 [硕士学位论文]

位论文].杭州:杭州电子科技大学,2011.

2 胡斌,徐小良.科技项目评审专家推荐系统模型.电子科技,2012,25(7):1-5.[doi:10.3969/j.issn.1007-7820.2012.07.001]

3 黄敏.科技项目评审专家推荐系统研究 [硕士学位论文].杭州:杭州电子科技大学,2013.

4 黎坚.科技专家遴选方法研究与实现 [硕士学位论文].广州:华南理工大学,2017.

5 余峰.项目评审专家推荐方法研究 [硕士学位论文].昆明:昆明理工大学,2013.

6 李明,刘鲁,王君,等.基于模糊文本分类的多知识领域专家推荐方法.北京航空航天大学学报,2009,35(10):1254-1257.

7 郑新宇,徐建良.基于文本相似度的评审专家推荐方法研究.科技资讯,2019,17(17):173-176.

8 魏捷.论文审稿专家推荐系统的设计与实现 [硕士学位论文].北京:北京邮电大学,2019.

9 刘一星.论文投稿系统评审专家自动推荐模型研究 [硕士学位论文].重庆:重庆大学,2009.

10 陶砾,刘恒初,杨朔,等.专家遴选系统设计与实现.计算机时代,2019,(7):36-39.

11 刘一星,梁山.基于改进 ATSV M 算法的评审专家自动推荐模型.重庆科技学院学报(自然科学版),2010,12(1):134-136.

12 石林宾.基于主题分析的评审专家推荐方法研究 [硕士学位论文].昆明:昆明理工大学,2014.

13 李春英,汤庸,陈国华,等.面向学术社区的专家推荐模型.智能系统学报,2012,7(4):365-369.[doi:10.3969/j.issn.1673-4785.201205041]

14 余峰,余正涛,杨剑锋,等.基于主题信息的项目评审专家推荐方法.计算机工程,2014,40(6):201-205.[doi:10.3969/j.issn.1000-3428.2014.06.043]

15 陈泽亚,王庆,郭静,等.基于二分图网络的项目与专家多重匹配策略.小型微型计算机系统,2016,37(3):545-550.[doi:10.3969/j.issn.1000-1220.2016.03.029]

16 杜方,宣琦,吴铁军.基于相似度传播的复杂网络间节点匹配算法.信息与控制,2011,40(3):331-337,342.

17 孙吴良.科技项目管理中专家与申请书分组匹配算法的研究 [硕士学位论文].北京:北京交通大学,2008.

18 毛晚堆,谷千军,褚蓓蓓,等.科技项目评审专家分组匹配算法.北京理工大学学报,2014,34(5):523-527.

19 吴仁克.科技项目评审专家智能检索与推荐系统的研究及实现 [硕士学位论文].杭州:杭州电子科技大学,2014.

20 印桂生,崔晓晖,董红斌,等.量子协同的二分图最大权完美匹配求解方法.计算机研究与发展,2014,51(11):2573-2584.[doi:10.7544/issn1000-1239.2014.20130687]