·复杂性科学·

# 基于相对熵的多属性作者学术影响力排名研究

胡小军[1]，郭 强[1]，杨 凯[1]，王江盼[1]，刘建国[2]

(1. 上海理工大学复杂系统科学研究中心 上海 杨浦区 200093；2. 上海财经大学金融科技研究院 上海 杨浦区 200433)

**【摘要】**基于欧氏距离的多属性排序方法(TOPSIS-ED)可以综合考虑科研人员的不同属性并对其影响力进行评价，然而该方法无法对其中垂线上的点进行排序。考虑作者的发表文章数、总引用量、平均被引用量、I10指数、H指数等5种指标，该文提出了一种基于相对熵的多属性排序方法(TOPSIS-RE)。该方法通过计算作者的上述5种指标值与正理想解和负理想解的相对熵，根据其接近正理想解和远离负理想解的程度对作者进行排名。该文以美国物理学会APS数据集作为训练集，将获得诺贝尔奖的文章的作者作为测试数据集，用AUC值说明算法的准确性。实验结果表明，TOPSIS-RE方法算得的AUC值为0.932 1，比总引用量指标提高了2.047%，并且比TOPSIS-ED方法提高了0.833%。该文的工作为从多属性角度刻画科学家影响力提供了借鉴。

**关 键 词** H指数； I10指数； 影响力排名； 相对熵

**中图分类号** N949 **文献标志码** A doi:10.3969/j.issn.1001-0548.2018.02.019

# Multi-Attribute Researcher Academic Influence Ranking Based on Relative Entropy

HU Xiao-jun[1], GUO Qiang[1], YANG Kai[1], WANG Jiang-pan[1], and LIU Jian-guo[2]

(1. Complex Systems Science Research Center, University of Shanghai for Science and Technology Yangpu Shanghai 200093;
2. Fintech Institute, Shanghai University of Finance and Economics Yangpu Shanghai 200433)

**Abstract** The multi-attribute ranking method based on Euclidean distance (TOPSIS-ED) can evaluate researcher academic influence through taking into account the different attributes of researchers. However, the method also has a problem that the points on the vertical line cannot be sorted. In this paper, we propose a multi-attribute ranking method based on relative entropy (TOPSIS-RE) by considering five indicators including the total number of papers, the total number of citations, citations per papers, I10-index and H-index. By calculating the relative entropy of the five indicators to the positive-ideal solution and the negative-ideal solution, this method ranks the authors according to the measurement that the results are close to the positive-ideal solution and far away from the negative-ideal solution. We select the American Physical Society data set as the training set and the authors who have won the Nobel Prize in the American Physical Society data set as the testing set. The area under curve (AUC) value is used to illustrate the accuracy of the algorithm. The results show that the AUC value calculated by TOPSIS-RE is 0.9321, and increases by 2.047% and 0.833% respectively compared with the total number of citations and TOPSIS-ED. Our work may shed some lights for quantifying the influence of scientists from the multi-attribute perspective.

**Key words** H-index; I10-index; influence ranking; relative entropy

定量评价科研人员的学术影响力，对引进人才、晋升、科研成果报奖、科研项目申请等具有重要的指导意义[1-4]。科研人员发表的文章是学术产出的主要形式[5-6]，体现了科研人员的最新研究进展和成果，研究者公开发表的文章被学术界或同行重视、认可和引用的情况，一定程度上反映了学者的学术水平和影响力[7-9]。

目前，评价研究者的学术影响力有两种比较重要的方法，同行评议法[10]和文献计量法[11]。同行评议法只有少数专家参加且对参评专家的知识和经验要求较高，缺乏一个统一和公认的标准，容易受主观因素的影响，一定程度上影响着评价效果。发表文章数[12]、总引用量[13]、影响因子[14]等基本的文献计量指标在学术影响力评价中逐渐被认可和使用。

收稿日期：2017－05－09；修回日期：2017－11－02
基金项目：国家自然科学基金(61773248, 71771152)；上海市东方学者特聘教授和上海市“曙光学者”项目(14SG42)
作者简介：胡小军(1992－)，男，主要从事社会网络分析、网络科学方面的研究.

2005年，文献[5]在基本的文献计量指标的基础上提出了一种评价学术成就的指标H指数。一个人的H指数越高，则表明其论文影响力越大。但是该指标也存在一定的局限性，H指数越大，越难增长，对于发表文章数较少而总引用量较高的学者的评价缺乏科学性[15]。文献[16]于2006年提出了G指数，用于改进H指数的不足，G指数是基于研究者累积贡献的评价指标，对于发表文章数较少而总引用量较高的学者的评价结果更加公平。此外，一些用于优化H指数的指标也相继被提出，用于弥补或完善上述文献计量指标的不足。2011年，Google公司为了评价一个学者的学术影响力提出了I10指数[17]，I10指数是指作者发表的文章被引用10次以上的数目。在依据某一文献计量学指标对研究者进行学术影响力排名时，不但会由于研究者的指标值相同而无法排名的问题，而且会导致最后的评价结果具有一定的片面性。近年来，社会网络分析方法因其能够定量地反映出节点在网络中位置的重要性[18-22]，进而可以与引文分析方法相结合用于评价学者在网络中的重要性，处于引文网络中重要位置的作者，具有较高的学术影响力[23]。但社会网络分析法因其动态性较弱，并不能展现作者学术影响力的动态变化过程[24]。

为了建立一个综合的评价体系去度量作者的学术影响力，文献[25]提出了综合考虑目标多属性的综合决策方法(又称TOPSIS方法)，对作者学术影响力的评估问题进行了细致研究。然而，TOPSIS方法对于正理想解与负理想解中垂线上的点无法进行排序。本文将作者的发表文章数、总引用量、平均被引用量、I10指数、H指数等5种评价指标作为TOPSIS的输入属性，根据每项指标对作者学术影响力评价的准确性(AUC值)进行加权来计算其综合评价值，对作者进行排序。由于相对熵(relative entropy)[26]并不对称也不满足三角不等式，因此可以用于两个概率分布差别的非对称性度量，从而解决正理想解和负理想解的中垂线上的点无法排序的问题。基于此思想，本文提出了一个基于相对熵的作者影响力排序方法(TOPSIS-RE)，用于评价研究者的学术影响力并对其做出排名。本文采用美国物理学会(American physical society, APS)的数据，将获得诺贝尔奖的文章的作者作为测试数据集，用AUC值说明算法的准确性。实验结果表明，基于相对熵的多属性排序方法(TOPSIS-RE)算得的AUC值为0.932 1，比总引用量指标提高了2.047%，并且比基于欧式距离的多属性排序方法(TOPSIS-ED)提高了0.833%。本文算法不仅解决了单个指标影响力值相同导致无法进行排序的问题，并且解决了基于欧氏距离的多属性排序方法(TOPSIS-ED)无法对中垂线上的点进行排序的问题，较全面准确地给出了作者学术影响力排名。

# 1 作者影响力评价指标

## 1.1 发表文章数

研究者 $v_i$ 的发表文章数目 $N_i$ 用于量化其学术影响力。

## 1.2 总引用量

作者 $v_i$ 一共发表了 $N_i$ 篇文章，每篇文章的被引用次数记为 $c_{ij}(j=1,2,\cdots,N_i)$，该作者 $v_i$ 的总引用量记为 $C_i$，即：

$$C_i=\sum_{j=1}^{N_i}c_{ij} \tag{1}$$

总引用量 $C_i$ 常被用于度量 $v_i$ 的学术影响力。

## 1.3 平均被引用量

作者 $v_i$ 一共发表了 $N_i$ 篇文章，总引用量为 $C_i$，则该作者的平均被引用量记为 $M_i$，即：

$$M_i=\frac{\sum_{j=1}^{N_i}c_{ij}}{N_i} \tag{2}$$

平均被引用量 $M_i$ 亦被用于度量 $v_i$ 的学术影响力。

## 1.4 I10指数

I10指数(I10-index)是由Google公司提出，并在Google学术网站上用以评价研究者学术影响力的指标。该指标是指作者 $v_i$ 已发表的文章中，被引用次数 $c_{ij}(j=1,2,\cdots,N_i)$ 大于10次的文章个数 $m$，记为 $I_i=m$。

## 1.5 H指数

H指数的计算基于研究者 $v_i$ 的论文数量 $N_i$ 及其论文被引用的次数 $c_{ij}(j=1,2,\cdots,N_i)$。一名科研人员的H指数是指其至多有 $h$ 篇论文分别被引用了至少 $h$ 次，则研究者 $v_i$ 的H指数为 $H_i=h$。

# 2 TOPSIS多属性排序方法

本文运用上述5种指标对美国物理学会APS数据集中所有作者的学术影响力进行量化，这5种指标具有一定的代表性，反映了科研文章的数量、质量以及领域研究情况等方面的特性。作者可以表示为 $V=\{v_1,v_2,\cdots,v_n\}$ 点的集合，发表文章数 $N$、总引用量 $C$、平均被引用量 $M$、I10指数 $I$、H指数 $H$ 作为作者的影响力属性，可以表示为 $F=\{f_1,f_2,$

$f_3,f_4,f_5\}=\{N,C,M,I,H\}$。$v_i(f_j)(i=1,2,\cdots,n;\ j=1,2,3,4,5)$表示研究者$v_i$的第$j$个评价指标值，由于各种指标的量纲不同，需要标准化各项指标的值：

$$t_{ij}=\frac{v_i(f_j)}{\sqrt{\sum_{i=1}^{n}v_i(f_j)^2}} \tag{3}$$

式中，$t_{ij}(i=1,2,\cdots,n;\ j=1,2,3,4,5)$表示作者$v_i$的第$j$个评价指标值标准化后的学术影响力值。

图1随机选取了15位作者并比较其各项指标的值，可以看出，不同作者的部分指标尤其是I10指数和H指数值相近或者相同(如11、12、13、14号作者的I10指标值相同；3、7、12号作者的平均被引指标值相同等)，这样就不能精确地区分不同作者的学术影响力。另外，由于每种指标的重要性程度不同，指标的选取对作者最终的排名有很大的影响。因此，本文提出多属性决策TOPSIS法，综合考虑多项指标来评价研究者的学术影响力，对作者进行排名。

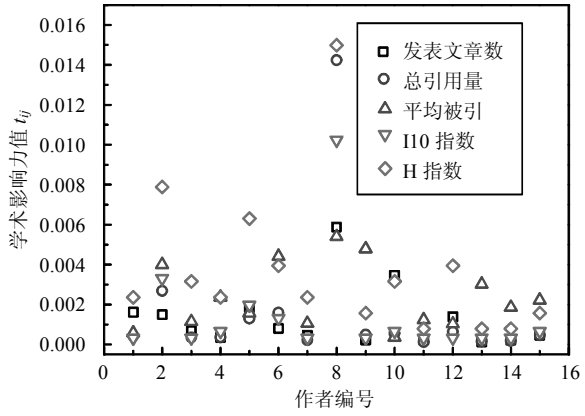


图1　5种指标的作者影响力值

TOPSIS[27]通过构造多属性问题的理想解，并以接近正理想解和远离负理想解这两个基准作为评价各可行方案的依据。正理想解是设想各指标属性都达到最满意的解，负理想解就是设想各指标属性都达到最不满意的解。本文首先计算作者的学术影响力值与正理想解和负理想解的距离，再根据算得的距离计算该作者影响力值贴近正理想解的程度，对作者进行排名具体步骤如下。

### 2.1　构造属性矩阵

属性矩阵可以表示为$\boldsymbol{P}$：

$$\boldsymbol{P}=\begin{pmatrix} v_1(f_1) & v_1(f_2) & v_1(f_3) & v_1(f_4) & v_1(f_5) \\ v_2(f_1) & v_2(f_2) & v_2(f_3) & v_2(f_4) & v_2(f_5) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ v_n(f_1) & v_n(f_2) & v_n(f_3) & v_n(f_4) & v_n(f_5) \end{pmatrix} \tag{4}$$

### 2.2　标准化矩阵

因为每种评价指标的量纲不同，作者的属性矩阵$\boldsymbol{P}$应该转换成标准化矩阵$\boldsymbol{T}$：

$$\boldsymbol{T}=\begin{pmatrix} t_{11} & t_{12} & t_{13} & t_{14} & t_{15} \\ t_{21} & t_{22} & t_{23} & t_{24} & t_{25} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{n1} & t_{n2} & t_{n3} & t_{n4} & t_{n5} \end{pmatrix} \tag{5}$$

式中，每个元素$t_{ij}(i=1,2,\cdots,n;\ j=1,2,3,\ 4,5)$表示作者$v_i$的第$j$个评价指标的学术影响力值，计算公式如式(3)所示。

### 2.3　加权标准化矩阵

第$j$个评价指标的权重系数为$w_j(j=1,2,3,4,5;\ \sum w_j=1)$，因此，加权标准化矩阵$\boldsymbol{R}$可以通过权重向量$\boldsymbol{W}^{\mathrm{T}}$和标准化矩阵$\boldsymbol{T}=(t_{ij})_{n\times 5}$重新构建：

$$\boldsymbol{R}=(r_{ij})=(w_jt_{ij})=\begin{pmatrix} w_1t_{11} & w_2t_{12} & \cdots & w_5t_{15} \\ w_1t_{21} & w_2t_{22} & \cdots & w_5t_{25} \\ \vdots & \vdots & \ddots & \vdots \\ w_1t_{n1} & w_2t_{n2} & \cdots & w_5t_{n5} \end{pmatrix} \tag{6}$$

式中，$\boldsymbol{W}^{\mathrm{T}}=[w_N\ w_C\ w_M\ w_I\ w_H]$根据层次分析法[26,28]计算得到，具体过程如下:

1) 评价指标的比较矩阵

根据5种单指标对研究者学术影响力评价的准确性，可以区分各项指标的重要程度。发表文章数$N$只是对作者工作量简单的量化，并不能体现所发表文章的质量，因此这一指标最不重要。文章的总引用量$C$和平均被引用量$M$表现为作者发表的文章被别人继续研究的情况，从一定程度上说明了作者的学术影响力，所以总引用量$C$和平均被引用量$M$是非常重要的两个指标，并且重要程度相当。而I10指数，只是部分统计了作者已发表的文章中被引用量大于10的文章数目，不能完全量化作者的全局影响力，H指数的评价性质与I10指数类似，因此I10指数和H指数的重要性略逊于总引用量$C$和平均被引用量$M$。

表1列出了按照式(7)三标度值方法构建的比较矩阵**CV**中的值。

**表1　指标重要性比较结果**

| **CV** | $N$ | $C$ | $M$ | $I$ | $H$ | $b_i$ |
|---|---|---|---|---|---|---|
| $N$ | 1 | 0 | 0 | 0 | 0 | 1 |
| $C$ | 2 | 1 | 1 | 2 | 2 | 8 |
| $M$ | 2 | 1 | 1 | 2 | 2 | 8 |
| $I$ | 2 | 0 | 0 | 1 | 1 | 4 |
| $H$ | 2 | 0 | 0 | 1 | 1 | 4 |

表1中：

$$\mathbf{CV}=(\mathrm{cv}_{ij})=\begin{cases}2 & \text{指标}i\text{ 比指标}j\text{ 更重要}\\ 1 & \text{指标}i\text{ 和指标}j\text{ 同等重要}\\ 0 & \text{指标}i\text{ 没有指标}j\text{ 重要}\end{cases}\quad(7)$$

2) 判断矩阵

用极差法构造判断矩阵，因为 $f(b_i,b_j)=d_{ij}=d_b^{(b_i-b_j)/B}$，所得的矩阵 $\boldsymbol{D}=(d_{ij})_{n\times n}$ 为一致性判断矩阵，其中 $d_b$ 为一常量，是按某种标准预先给定的极差元素对的相对重要程度(一般在实践应用中常取 $d_b$=9)；$b_i=\sum_{j=1}^{5}\mathrm{cv}_{ij}$ (其中，$\mathrm{cv}_{ij}$ 为表1中比较矩阵 **CV** 中的元素)；$B=b_{\max}-b_{\min}$，称为极差，其中 $b_{\max}=\max(b_1,b_2,\cdots,b_5)$，$b_{\min}=\min(b_1,b_2,\cdots,b_5)$。一致性判断矩阵 $\boldsymbol{D}$ 为：

$$\boldsymbol{D}=(d_{ij})=\begin{pmatrix} d & N & C & M & I & H\\ N & 1 & 9^{-1} & 9^{-1} & 9^{-3/7} & 9^{-3/7}\\ C & 9 & 1 & 1 & 9^{4/7} & 9^{4/7}\\ M & 9 & 1 & 1 & 9^{4/7} & 9^{4/7}\\ I & 9^{3/7} & 9^{-4/7} & 9^{-4/7} & 1 & 1\\ H & 9^{3/7} & 9^{-4/7} & 9^{-4/7} & 1 & 1\end{pmatrix}\quad(8)$$

权重系数 $\overline{W}$ 确定如下：

$W'$=[$9^{-20/7}$, $9^{15/7}$, $9^{15/7}$, $9^{-5/7}$, $9^{-5/7}$]

$W$=[$9^{-4/7}$, $9^{3/7}$, $9^{3/7}$, $9^{-1/7}$, $9^{-1/7}$]

$\overline{W}$=[0.041 4, 0.373 0, 0.373 0, 0.106 3, 0.106 3]

式中：$W_i'=\prod_{j=1}^{5}d_{ij}$；$W_i=\sqrt[5]{W_i'}$；$\overline{W}_i=W_i\Big/\sum_{i=1}^{5}W_i$，$(i,j=1,2,3,4,5)$。

3) 一致性检验

一致性检验是为了检验各指标重要程度之间的协调性，避免出现前后矛盾的情况。按一致性检验指标 $P_{\mathrm{C.I.}}=(\lambda_{\max}-n)/(n-1)\leqslant\varepsilon$ ($\varepsilon$ 为满足一致性要求所允许的最大值，一般根据具体情况来确定)，进行一致性检验。其中 $\lambda_{\max}$ 为一致性判断矩阵 $\boldsymbol{E}=(e_i)_{5\times1}=\boldsymbol{D}\times\overline{W}$ 的最大特征向量值，$n$ 为评价指标个数。$\lambda_{\max}$ 为：

$$\lambda_{\max}=\sum_{i=1}^{5}\frac{e_i}{5\overline{W}_i}\quad(9)$$

通过计算得到 $P_{\mathrm{C.I.}}=\dfrac{\lambda_{\max}-5}{5-1}=0.000\,15$，因为 $\varepsilon$=0.000 15<0.1(当 $\varepsilon\leqslant0.1$ 时，判断矩阵符合一致性要求[29])，所以满足一致性检验。由以上计算可知，各评价指标相应的权重系数为 $\boldsymbol{W}^{\mathrm{T}}$=[0.041 4 0.373 0 0.373 0 0.106 3 0.106 3]。

### 2.4 确定正理想解和负理想解

本文将每种评价指标计算得到的作者学术影响力的最大值和最小值分别作为该项指标的正理想解 $r_j^+$ 和负理想解 $r_j^-$ ($j=1,2,\cdots,5$)。根据式(6)加权标准化矩阵 $\boldsymbol{R}$，可求得正理想解 $A^+$ 和负理想解 $A^-$ 分别为：

$$A^+=\left\{\max_{i\in\{1,2,\cdots,n\}}(r_{i1}),\max_{i\in\{1,2,\cdots,n\}}(r_{i2}),\cdots,\max_{i\in\{1,2,\cdots,n\}}(r_{i5})\right\}=\left\{r_1^+,r_2^+,\cdots,r_5^+\right\}\quad(10a)$$

$$A^-=\left\{\min_{i\in\{1,2,\cdots,n\}}(r_{i1}),\min_{i\in\{1,2,\cdots,n\}}(r_{i2}),\cdots,\min_{i\in\{1,2,\cdots,n\}}(r_{i5})\right\}=\left\{r_1^-,r_2^-,\cdots,r_5^-\right\}\quad(10b)$$

### 2.5 计算距离

记 $S_i^+$ 和 $S_i^-$ 分别为作者 $v_i$ 的学术影响力值与正理想解 $A^+$ 和负理想解 $A^-$ 的欧氏距离(Euclidean distance)，称为TOPSIS-ED，计算公式如下：

$$S_i^+=\sqrt{\sum_{j=1}^{5}(r_{ij}-r_j^+)^2}\quad(11a)$$

$$S_i^-=\sqrt{\sum_{j=1}^{5}(r_{ij}-r_j^-)^2}\quad(11b)$$

式中，$r_{ij}$ 表示作者 $v_i$ 的第 $j$ 个指标对其学术影响力的量化值；$r_j^+$ 和 $r_j^-$ 分别表示第 $j$ 个指标对所有作者学术影响力评估值中的最大值和最小值。

当作者 $v_i$ 的学术影响力值处于正理想解 $A^+$ 和负理想解 $A^-$ 的中垂线上时，上面介绍的欧氏距离的计算方法无法对作者进行排序。由于相对熵并不对称也不满足三角不等式，可以用于两个概率分布差别的非对称性度量。考虑作者学术影响力值与正理想解和负理想解相对熵，可以解决欧氏距离中垂线上的点无法排序的问题。作者 $v_i$ 的学术影响力值与正理想解 $A^+$ 和负理想解 $A^-$ 的相对熵(relative entropy)[26,30]，称为TOPSIS-RE，计算公式如下：

$$S_i^+=\left\{\sum_{j=1}^{5}\left[r_j^+\log\frac{r_j^+}{r_{ij}}+(1-r_j^+)\log\frac{1-r_j^+}{1-r_{ij}}\right]\right\}^{1/2}\quad(12a)$$

$$S_i^-=\left\{\sum_{j=1}^{5}\left[r_j^-\log\frac{r_j^-}{r_{ij}}+(1-r_j^-)\log\frac{1-r_j^-}{1-r_{ij}}\right]\right\}^{1/2}\quad(12b)$$

### 2.6 计算接近度

根据 $S_i^+$ 和 $S_i^-$ 可以计算作者 $v_i$ 的学术影响力与理想方案的相对接近程度，记为 $A_i$：

$$A_i=\frac{S_i^-}{S_i^+ + S_i^-} \qquad i=1,2,\cdots,n \tag{13}$$

$A_i$ 值越大，表示作者 $v_i$ 的学术影响力越大。如果 $A_i$=0，表示作者 $v_i$ 最不重要；相反，如果 $A_i$=1，则表示作者 $v_i$ 最重要。根据 $A_i$ 的值降序排序，可以综合评估作者 $v_i$ 的学术影响力排名。

## 3　数值结果

### 3.1　数据描述

本文采用美国物理学会(American Physical Society, APS)的数据，包括从1893年～2009年，超过46万篇已发表的文章。每篇文章包含唯一的文章编号、文章名、发表时间(年-月-日)、作者名字、以及每位作者的所属机构。另一个数据集用文章编号，提供了超过470万条引用关系。为了研究作者的学术影响力，本文最终处理完的数据包含10万多位作者，包括他们的发表文章情况以及被引用情况。

### 3.2　实验结果

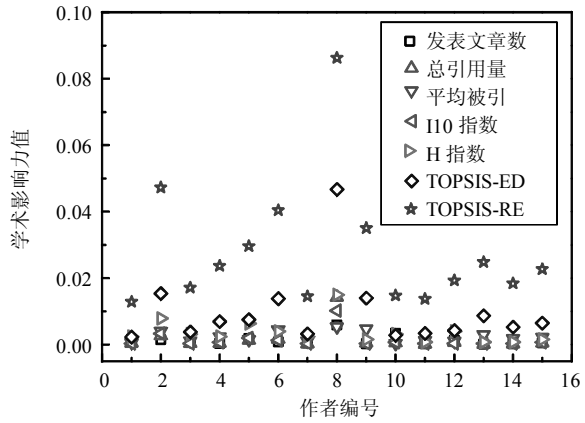


图2　TOPSIS-RE和其他指标比较

在APS数据集中，本文提出TOPSIS-RE方法，综合考虑5种指标对作者的影响力进行计算，并与单个指标及TOPSIS-ED方法作了对比。图2随机选取了15位作者并比较其各项指标的值，纵坐标表示作者学术影响力值(5种单指标的数值为式(3)中的 $t_{ij}$ 值，TOPSIS-ED和TOPSIS-RE的数值为式(13)中的 $A_i$ 值)。依据指标值的大小，本文可以对作者进行排名。本文提出的TOPSIS-RE方法对不同作者的影响力值有相对明显的区分，解决了单个评价指标值相同导致无法进行排名的问题(如11、13、14号作者的H指数值相同)；并且解决了TOPSIS-ED方法中垂线上的点无法排名的问题(如6、9号作者的TOPSIS-ED值相同)。再者，本文对不同指标进行加权，分析了各项指标在评价作者影响力时的重要程度，提高了依靠作者指标值进行排名的准确性。

### 3.3　评价方法

为了评价本文方法对作者学术影响力排名的准确性，本文选取物理领域获得诺贝尔奖的文章的作者(去重之后共142位)作为测试数据集，如表2所示，其他的作者作为非测试数据集。将实验求得的作者学术影响力排名和测试数据集作比对，计算其AUC的值[31-32]，当AUC=1时，表明测试数据集里作者的排名都高于非测试数据集里作者的排名；当AUC=0.5时，表明所有作者的排名是随机的。较大的AUC值代表了较好的实验准确性，其计算公式如下：

$$\text{AUC}=\frac{n_1+n_2\times 0.5}{n} \tag{14}$$

式中，$n$ 表示比较次数(取$10^5$)；$n_1$ 表示测试数据集里作者影响力值高于非测试数据集里作者影响力值的次数；$n_2$ 表示测试数据集里作者影响力值等于非测试数据集里作者影响力值的次数。

**表2　Physics领域获得诺贝尔奖的文章**

| 文章DOI | 作者人数 | 获奖年份 |
|---|---|---|
| PhysRevLett.35.1489(1975) | 36 | 1995 |
| PhysRevLett.55.48(1985) | 5 | 1997 |
| PhysRevLett.61.826(1988) | 5 | 1997 |
| PhysRevLett.61.169(1988) | 6 | 1997 |
| PhysRevLett.48.1559(1982) | 3 | 1998 |
| PhysRev.140.A1133(1965) | 2 | 1998 |
| PhysRevLett.75.3969(1995) | 7 | 2001 |
| PhysRevLett.20.1205(1968) | 3 | 2002 |
| PhysRevLett.58.1490(1987) | 23 | 2002 |
| PhysRevLett.9.439(1962) | 4 | 2002 |
| PhysRevLett.84.5102(2000) | 10 | 2005 |
| PhysRevLett.84.3232(2000) | 6 | 2005 |
| PhysRevLett.61.2472(1988) | 9 | 2007 |
| PhysRevLett.57.2442(1986) | 5 | 2007 |
| PhysRev.122.345(1961) | 2 | 2008 |
| PhysRevLett.53.1951(1984) | 4 | 2011 |
| PhysRevLett.77.4887(1996) | 8 | 2012 |
| PhysRevLett.76.1796(1996) | 5 | 2012 |
| PhysRevLett.13.321(1964) | 2 | 2013 |

AUC计算结果如表3所示，本文作者影响力评价指标中，总引用量指标相对于其他几个指标能较好的反应作者的学术影响力水平，其AUC值为0.913 4。基于相对熵的多属性排序方法(TOPSIS-RE)，比单个指标中最高的总引用量提高了2.047%，并且比经典的TOPSIS-ED方法提高了约0.833%，对作者影响力排名的评估更加准确。

**表3　各指标AUC值**

| | 发表文章数 | 总引用量 | 平均被引 | I10指数 | H指数 | TOPSIS-ED | TOPSIS-RE |
|---|---|---|---|---|---|---|---|
| AUC | 0.754 2 | 0.913 4 | 0.881 3 | 0.822 7 | 0.809 3 | 0.924 4 | **0.932 1** |

为了直观地看出本文提出的TOPSIS-RE方法能更准确地从120 000位作者中识别出获得诺贝尔奖文章的142位作者，本文将各种指标的排名结果进行了对比分析。如图3所示，横坐标表示排名列表中的前$k$个作者，纵坐标表示前$k$个作者中获得诺贝尔奖文章的作者数目。例如，按各指标值排名的前1 000位作者中，发表文章数指标、总引用量指标、平均被引指标、I10指数指标、H指数指标、TOPSIS-RE指标分别包含了8位、27位、25位、14位、32位、41位获诺贝尔奖文章的作者；前10 000位作者中，发表文章数指标、总引用量指标、平均被引指标、I10指数指标、H指数指标、TOPSIS-RE指标分别包含了59位、112位、86位、77位、83位、119位获诺贝尔奖文章的作者。图3可以看出本文提出的TOPSIS-RE方法能够使获得诺贝尔奖的文章的作者排名较其他指标更靠前，对作者排名的准确性比单个指标高，并且高于TOPSIS-ED方法。

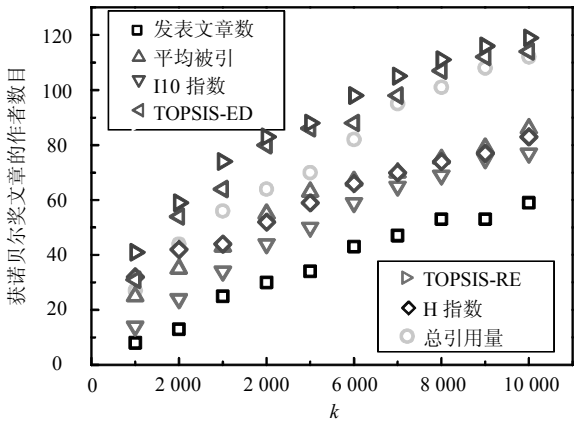


图3 TOPSIS-RE与其他指标的排名结果对比

## 4 结 束 语

本文综合考虑作者的发表文章数、总引用量、平均被引用量、I10指数、H指数等5种指标，通过计算作者的学术影响力值与正理想解和负理想解的相对熵，提出了一种基于相对熵的多属性排序方法。在美国物理学会APS数据集上的实验结果表明，TOPSIS-RE方法算得的AUC值为0.932 1，比总引用量指标提高了2.047%，并且比TOPSIS-ED方法提高了0.833%。从图3可以看出TOPSIS-RE方法比其他指标能更好地识别出获得诺贝尔奖文章的作者。本文算法不仅解决了单个评价指标值相同导致无法进行排名的问题，同时解决了TOPSIS-ED方法中垂线上的点无法排序的问题。运用相对熵的距离计算方法还会存在少部分作者的学术影响力相同的情况：由于数据集的限制，存在一部份作者，他们是某篇或者某几篇文章的合作关系，而数据集中没有他们发表的其他的文章信息，这会导致这部分作者的5种单指标的值完全一致，所以不管何种计算方法，都不能把他们区分开，需要更丰富的发表文章信息。

多属性排序方法的有效运用，取决于所选指标的优劣和赋予权重系数的合理性，因此可以进一步研究作者的学术影响力评价指标以及更优的赋权方法，使作者的排名更为准确。在研究作者学术影响力时，时间因素也会对作者的排名结果产生重大的影响，未来的工作中，将通过年份的划分更细化地研究作者学术影响力动态变化。另外，通过引文网络，研究网络的结构对理解作者的学术地位和合作模式具有重要意义。

## 参 考 文 献

[1] HICKS D, WOUTERS P, WALTMAN L, et al. The Leiden Manifesto for research metrics[J]. Nature, 2015, 520(7548): 429.

[2] SHEN H W, BARABÁSi A L. Collective credit allocation in science[J]. Proceedings of the National Academy of Sciences, 2014, 111(34): 12325-12330.

[3] FORTIN J M, CURRIE D J. Big science vs. little science: How scientific impact scales with funding[J]. PloS One, 2013, 8(6): e65263.

[4] GILES C L, COUNCILL I G. Who gets acknowledged: Measuring scientific contributions through automatic acknowledgment indexing[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(51): 17599-17604.

[5] HIRSCH J E. An index to quantify an individual's scientific research output[J]. Proceedings of the National academy of Sciences of the United States of America, 2005, 102(46): 16569-16572.

[6] 胡枫, 赵海兴, 何佳倍, 等. 基于超图结构的科研合作网络演化模型[J]. 物理学报, 2013, 62(19): 178901.
HU Feng, ZHAO Hai-xing, HE Jia-pei, et al. An evolving model for hypergraph-structure-based scientific collaboration networks[J]. Acta Phys Sin, 2013, 62(19): 178901.

[7] 高志, 张志强. 个人学术影响力定量评价方法研究综述[J]. 情报理论与实践, 2016, 39(1): 133-138.
GAO Zhi, ZHANG Zhi-qiang. A Summary of quantitative evaluation methods of personal academic influence[J]. Information Studies: Theory & Application, 2016, 39(1): 133-138.

[8] 宣照国, 苗静, 党延忠, 等. 科研领域关联网络的社团结构分析[J]. 上海理工大学学报, 2008, 30(3): 249-252.
XUAN Zhao-guo, MIAO Jing, DANG Yan-zhong, et al. Community structure of Chinese nature science basic research weighted networks[J]. Journal of University of Shanghai for Science and Technology, 2008, 30(3): 249-252.

[9] 郑佳之, 张杰. 一种个人学术影响力的评价方法[J]. 中国科技期刊研究, 2007, 18(6): 957-960.
ZHEN Jia-zhi, ZHANG Jie. Method to estimate academic impact of individuals[J]. Chinese Journal of Scientific and Technica, 2007, 18(6): 957-960.

[10] 龚旭. 美国国家科学基金会的同行评议制度及其启示[J]. 中国科学基金, 2005, 18(6): 373-376.
GONG Xu. Peer review system of the National Science Foundation and its implications[J]. Bulletin of National Natural Science Foundation of China, 2005, 18(6): 373-376.

[11] 崔宇红. 从文献计量学到 Altmetrics: 基于社会网络的学术影响力评价研究[J]. 情报理论与实践, 2013, 36(12): 17-20.
CUI Yu-hong. From bibliometrics to altmetrics: a study of academic impacts based on social networks[J]. Information Studies: Theory & Application, 2013, 36(12): 17-20.

[12] NEWMAN M E J. Coauthorship networks and patterns of scientific collaboration[J]. Proceedings of the National Academy of Sciences, 2004, 101(suppl 1): 5200-5205.

[13] PETERSEN A M, WANG F, STANLEY H E. Methods for measuring the citations and productivity of scientists across time and discipline[J]. Physical Review E, 2010, 81(3): 036114.

[14] GARFIELD E. Citation analysis as a tool in journal evaluation[J]. Science, 1972, 178(4060): 471-479.

[15] BORNMANN L, DANIEL H D. Does the H-index for ranking of scientists really work?[J]. Scientometrics, 2005, 65(3): 391-392.

[16] EGGHE L. Theory and practise of the g-index[J]. Scientometrics, 2006, 69(1): 131-152.

[17] DELGADO L C E, ROBINSON G N, TORRES S D. The Google scholar experiment: How to index false papers and manipulate bibliometric indicators[J]. Journal of the Association for Information Science and Technology, 2014, 65(3): 446-454.

[18] 刘建国, 任卓明, 郭强, 等. 复杂网络中节点重要性排序的研究进展[J]. 物理学报, 2013, 62(17): 178901.
LIU Jian-guo, REN Zhuo-ming, GUO Qiang, et al. Node importance ranking of complex networks[J]. Acta Phys Sin, 2013, 62(17): 178901.

[19] 于会, 刘尊, 李勇军. 基于多属性决策的复杂网络节点重要性综合评价方法[J]. 物理学报, 2013, 62(2): 020204.
YU Hui, LIU Zun, LI Yong-jun. Key nodes in complex networks identified by multi-attribute decision-making method[J]. Acta Phys Sin, 2013, 62(2): 020204.

[20] 邵凤, 郭强, 曾诗奇, 等. 微博系统网络结构的研究进展[J]. 电子科技大学学报, 2014, 43(2): 174-183.
SHAO Feng, GUO Qiang, ZENG Shi-qi, et al. Research progress of the microblog system structures[J]. Journal of University of Electronic Science and Technology of China, 2014, 43(2): 174-183.

[21] LIU J G, LIN J H, GUO Q, et al. Locating influential nodes via dynamics-sensitive centrality[J]. Scientific Reports, 2016, 6: 21380.

[22] 狄增如. 系统科学视角下的复杂网络研究[J]. 上海理工大学学报, 2011, 33(2): 111-116.
DI Zeng-ru. Research of complex networks from the view point of systems science[J]. Journal of University of Shanghai for Science and Technology, 2011, 33(2): 111-116.

[23] 孟祥保, 钱鹏. 国际图书情报学研究群体结构——以核心作者互引分析为视角[J]. 情报科学, 2015, 33(5): 124-128.
MENG Xiang-bao, QIAN Peng. Research groups structure of international library and information science based on core author cross-citation analysis[J]. Information Science, 2015, 33(5): 124-128.

[24] 李旋, 郝继英. 学者的学术影响力评价方法[J]. 中华医学图书情报杂志, 2016(8): 48-52.
LI Xuan，HAO Ji-ying. Methods for evaluating the academic impact of scholars[J]. Chin J Med Libr Inf Sci, 2016(8): 48-52.

[25] 金晶, 何苗, 王孝宁, 等. 不同学科领域自然科学论文学术影响力评价与比较的可行性研究[J]. 科技管理研究, 2010, 30(14): 279-284.
JIN Jing，HE Miao，WANG Xiao-ning，et al. Feasibility research of evaluation and comparison of natural science papers in different fields[J]. Science and Technology Management Research, 2010, 30(14): 279-284.

[26] KULLBACK S, LEIBLER R A. On information and sufficiency[J]. The Annals of Mathematical Statistics, 1951, 22(1): 79-86.

[27] YOON K P, HWANG C L. Multiple attribute decision making: an introduction[M]. Sage Publications, 1995, 104: 38-44

[28] 朱茵, 孟志勇, 阚叔愚. 用层次分析法计算权重[J]. 北京交通大学学报, 1999, 23(5): 119-122.
ZHU Yin, MENG Zhi-yong, KAN Shu-yu. Determination of weight value by AHP[J]. Journal of Northern Jiaotong University, 1999, 23(5): 119-122.

[29] GOLDEN B L, WASIL E A, HARKER P T. The analytic hierarchy process: Applications and studies[J]. Computers & Operations Research, 1993, 20(5): 562-563.

[30] 赵萌, 邱菀华, 刘北上. 基于相对熵的多属性决策排序方法[J]. 控制与决策, 2010(7): 1098-1100.
ZHAO Meng, QIU Wan-hua, LIU Bei-shang. Relative entropy evaluation method for multiple attribute decision making[J]. Control and Decision, 2010(7): 1098-1100.

[31] HANLEY J A, MCNEIL B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve[J]. Radiology, 1982, 143(1): 29-36.

[32] LIU X L, GUO Q, HOU L, et al. Ranking online quality and reputation via the user activity[J]. Physica A: Statistical Mechanics and its Applications, 2015, 436: 629-636.

编 辑 蒋 晓