

●王 筠

情报检索系统用户相关性判断的灰色聚类决策^{*}

摘 要 所谓灰色聚类决策,就是按多个不同的决策指标对决策对象进行综合评价,以确定决策对象是否满足给定的取舍原则。该方法可以用于对情报检索系统的检索结果进行相关性判断。参考文献4。

关键词 情报检索系统 检索系统评价 用户相关性 灰色聚类决策

分类号 G354.43

ABSTRACT Grey clustering decision is to comprehensively evaluate decision objects according to various decision indicators to determine whether the decision objects meet given criteria. It can be used for the relevance judgment of retrieval results of information retrieval system. 4 refs.

KEY WORDS Information retrieval system. Evaluation of retrieval system. End-user relevance. Grey clustering decision.

CLASS NUMBER G354.43

情报检索的主要任务是确定与用户情报需求相关的文献,然而相关文档本身是一个模糊概念,不能简单地用相关或不相关来判断。另外用户的相关性受多种因素的影响,其中有些属于确定的因素,有些属于非确定因素。本文尝试利用灰色聚类方法判定检索结果的相关性。它既可以把多种因素综合起来考虑,又能将检索结果按要求分为几个区域。比如:相关、部分相关、部分不相关、不相关等。在每个区域的检索结果中,还能给出其相关程度,根据这种程度,用户可选用其中的某些文献,同时也可对检索系统进行评价。

1 灰色聚类决策简介

所谓灰色聚类决策就是按多个不同的决策指标对决策对象进行综合评价,以确定决策对象是否满足给定的取舍准则。设有 n 个决策对象, m 个决策指标, s 个不同的灰类(即要把决策对象分为 s 类), x_{ij} ($i=1,2,\dots,n; j=1,2,\dots,m$) 为决策对象 i 关于决策指标 j 的量化评价, $f_j^k(x)$ 为 j 指标 k 子类白化权函数(用来描述一个灰数对其取值范围内不同数值的“偏爱”程度的函数。此函数的确定是研究者根据已知信息设计的,具体设计方法参见相关文献^[1]。这里的 $j=1,2,\dots,m; k=1,2,\dots,s$), η_j ($j=1,2,\dots,m$) 为决策指标 j 的综合决策

权,且

$$\sum_{j=1}^m \eta_j = 1,$$

则称 $\sigma_i^k = \sum_{j=1}^m f_j^k(x_{ij}) \cdot \eta_j$

为决策对象 i 属于 k 灰类的决策系数。

我们还称 $\sigma_i = (\sigma_i^1, \sigma_i^2, \dots, \sigma_i^s)$;

$i=1,2,\dots,n$ 为决策对象的决策系数向量。

若 $\max_{1 \leq k \leq s} \{\sigma_i^k\} = \sigma_i^{k^*}$,

则称决策 i 属于灰类 k^* 。

在实际问题中,常常会遇到多个决策对象同属于一个决策灰类,而该灰类所能容纳的对象个数又有一定限额的情况,此时可根据决策系数的大小决定取舍。

2 情报检索系统信息资源评价指标

鉴于目前图书情报检索主要使用了网络信息查询手段,因而情报检索系统评价指标也应采取网络资源评价指标。网络资源评价指标基本可沿用传统文献的评价指标,同时也应考虑网上资源的特点增设一些特定的标准。根据有关文献^[2],我们选取以下指标。

(1)内容:准确性、权威性、客观性、可靠性、独特性、新颖性、针对性、范围面和写作水平等。

^{*} 本文为辽宁省教育厅科研基金资助项目(20041009)研究成果。

(2)设计:用户界面的友好性、浏览和检索的难易、信息组织的科学性、页面设计的艺术性和适用性。

(3)运营:信息提供的保障性、可存取性、链接的可达性、设备使用的兼容性、费用的高低等。

需要说明的是决策指标可根据要求确定。比如可把上面3项大的指标分列成多项具体的指标。由于指标项数的多少不影响对决策方法的介绍,为了方便,我们不妨把决策指标确定为以上3大项。

3 建立用户相关性判断的灰色聚类模型

设某检索系统就用户的一项检索提问共输出6个检索结果。用户相关性判断的决策指标为:内容、设计、运营。各个指标的决策权分别为0.708, 0.138, 0.154^[3],试按相关、部分相关、部分不相关和不相关作聚类决策。根据用户的需要若只需选用其中的两个检索结果,试确定选出的结果。

采用专家鉴定的方法,得到3个决策指标的量化评价价值,

$$x_{ij} (i = 1, 2, 3, 4, 5, 6; j = 1, 2, 3)$$

$$C = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \\ x_{51} & x_{52} & x_{53} \\ x_{61} & x_{62} & x_{63} \end{bmatrix} = \begin{bmatrix} 86 & 96 & 70 \\ 85 & 95 & 20 \\ 63 & 77 & 40 \\ 84 & 86 & 66 \\ 35 & 78 & 80 \\ 30 & 74 & 86 \end{bmatrix}$$

设3个决策指标关于4个灰类的白化权函数分别为

$$f_j^1(x) = \begin{cases} 0 & x < 65 \\ \frac{x-65}{20} & 65 \leq x \leq 85 \\ 1 & x > 85 \end{cases}$$

$$f_j^2(x) = \begin{cases} 0 & x < 50 \\ \frac{x-50}{15} & 50 \leq x \leq 65 \\ \frac{80-x}{15} & 65 < x \leq 80 \\ 0 & x > 80 \end{cases}$$

$$f_j^3(x) = \begin{cases} 1 & 0 \leq x \leq 35 \\ \frac{60-x}{25} & 35 < x \leq 60 \\ 0 & x > 60 \end{cases}$$

$$f_j^4(x) = \begin{cases} 0 & x < 70 \\ \frac{x-70}{20} & 70 \leq x \leq 90 \\ 1 & x > 90 \end{cases}$$

这里的 $j = 1, 2, 3$, 即不同指标的白化权函数完全相同。

下面求各检索结果对各灰类的决策系数。

$$\begin{aligned} \sigma_1^1 &= \sum_{j=1}^3 f_j^1(x_{1j}) \eta_j \\ &= f_1^1(x_{11}) \eta_1 + f_2^1(x_{12}) \eta_2 + f_3^1(x_{13}) \eta_3 \\ &= f_1^1(86) \times 0.708 + f_2^1(96) \times 0.138 + f_3^1(70) \times 0.154 \\ &= 1 \times 0.708 + 1 \times 0.138 + 0.25 \times 0.154 \\ &\approx 0.885 \end{aligned}$$

$$\begin{aligned} \sigma_1^2 &= \sum_{j=1}^3 f_j^2(x_{1j}) \eta_j \\ &= 0 \times 0.708 + 0 \times 0.138 + 0.667 \times 0.154 \\ &\approx 0.103 \end{aligned}$$

$$\sigma_1^3 = 0$$

$$\begin{aligned} \sigma_1^4 &= 0.8 \times 0.708 + 1 \times 0.138 + 0 \times 0.154 \\ &\approx 0.704 \end{aligned}$$

所以

$$\begin{aligned} \sigma_1 &= (\sigma_1^1, \sigma_1^2, \sigma_1^3, \sigma_1^4) \\ &= (0.885, 0.103, 0, 0.704)。 \end{aligned}$$

类似地可以算出

$$\begin{aligned} \sigma_2 &= (\sigma_2^1, \sigma_2^2, \sigma_2^3, \sigma_2^4) \\ &= (0.846, 0, 0.154, 0.669), \\ \sigma_3 &= (\sigma_3^1, \sigma_3^2, \sigma_3^3, \sigma_3^4) \\ &= (0.083, 0.642, 0.123, 0.048), \\ \sigma_4 &= (\sigma_4^1, \sigma_4^2, \sigma_4^3, \sigma_4^4) \\ &= (0.819, 0.144, 0, 0.606), \\ \sigma_5 &= (\sigma_5^1, \sigma_5^2, \sigma_5^3, \sigma_5^4) \\ &= (0.206, 0.018, 0.708, 0.132), \\ \sigma_6 &= (\sigma_6^1, \sigma_6^2, \sigma_6^3, \sigma_6^4) \\ &= (0.216, 0.055, 0.708, 0.151)。 \end{aligned}$$

由

$$\begin{aligned} \max_{1 \leq k \leq 4} \{ \sigma_1^k \} &= 0.885 = \sigma_1^1 \\ \max_{1 \leq k \leq 4} \{ \sigma_2^k \} &= 0.846 = \sigma_2^1 \\ \max_{1 \leq k \leq 4} \{ \sigma_3^k \} &= 0.642 = \sigma_3^2 \\ \max_{1 \leq k \leq 4} \{ \sigma_4^k \} &= 0.819 = \sigma_4^1 \\ \max_{1 \leq k \leq 4} \{ \sigma_5^k \} &= 0.708 = \sigma_5^3 \\ \max_{1 \leq k \leq 4} \{ \sigma_6^k \} &= 0.708 = \sigma_6^3 \end{aligned}$$

知,编号为1、2、4的检索结果属于相关灰类;编号为3的检索结果属于部分相关灰类;编号为5、6的检索

●王云娣

关于我国大学图书馆的电子资源利用教育*

摘 要 对我国130所大学图书馆的网络调研结果表明,我国大学图书馆电子资源利用教育目前存在的主要问题有:电子资源利用教育不平衡现象明显,缺乏统一管理,教学内容的针对性比较模糊,远距离用户的教育亟待解决。要加强对电子资源利用教育的统一管理,拟定完善的教育计划,提高师资队伍素质,加强合作,共享电子资源利用教育的资源与经验。参考文献5。

关键词 大学图书馆 电子资源 用户教育 信息检索

分类号 G252

ABSTRACT According a survey in 130 university libraries in China, the author finds some problems in the education for the use of electronic resources. Then, she proposes her own points of view to solve these problems, such as good educational plans, high teachers' quality and resources sharing. 5 refs.

KEY WORDS University library. Electronic resource. User education. Information retrieval.

CLASS NUMBER G252

虽然网络环境改变了用户到图书馆查阅文献、依靠咨询馆员检索数据库的局面,但电子资源的良莠不齐、庞杂无序,给用户的取舍和利用带来了极大不便。尤其是电子信息资源在形式、资源组织方式、资源储存方式、资源检索方式以及资源使用的时空

等方面与传统资源存在较大的差异,使得读者的信息寻求行为发生变化。大学图书馆必须了解电子资源对读者的影响,利用切实可行的手段和工具开展电子资源利用的教育,使之在教学科研中发挥更大作用。

结果属于部分不相关灰类;没有检索结果属于不相关灰类。又由于 $\sigma_1^2 = 0.885 > \sigma_2^2 = 0.846 > \sigma_4^2 = 0.819$, 所以若选择两个相关结果,应首先考虑编号为1、2的检索结果,4号可作为候补。

已有学者提出如何用灰色聚类决策的方法对检索系统进行评价^[4]。这里只不过将其用户对检索结果评价的定性方法,改为现在的定量方法,避免了评价时的主观性、片面性以及用户面对检索结果难做决断的现象。

灰色聚类决策法按多个不同的决策指标对决策对象进行综合评价,将可供选择的检索结果按要求分为几个灰类,便于从整体上掌握检索结果的相关程度和决定取舍的范围。对于情报检索系统检索结果的相关性判断是一种定量的科学方法。而且决策

指标和所分灰类均可根据要求设定,使所得结果更符合用户的特点,并便于对检索系统进行评价。

参考文献

- 1 刘思峰等.灰色系统理论及应用(第二版).北京:科学出版社,1999
- 2 栗慧.网络资源评价:评价标准及元数据和 CORC 系统的应用.情报学报,2002(3)
- 3 陆宝益.网络信息资源的评价.情报学报,2002(1)
- 4 杨立英等.评价情报检索系统性能的新方法——加权中值法.情报学报,2003(1)

王 筠 辽宁师范大学数学系资料室副研究馆员。通讯地址:大连。邮编 116029。

(来稿时间:2003-05-26)

* 本文系浙江省教育厅科研课题“网络环境下的高校信息素质教育模式研究”成果之一。