

文章编号: 1000-6788(2000) 10-0089-04

## 确定事例特征权值的方法研究

吴良刚<sup>1</sup>, 高 阳<sup>1</sup>, 张金隆<sup>2</sup>

(1. 中南工业大学工商管理学院, 湖南 长沙 410083; 2. 华中理工大学科研处, 湖北 武汉 430074)

**摘要:** 针对事例特征权值的确定问题, 提出了 Delphi 和 AHP 相结合的权值确定方法, 以及二次规划模型权值确定法和信用统计权值确定法。这些方法在事例特征权值的确定中, 有助于减少主观性, 增加客观性。

**关键词:** 事例推理; 事例特征权值; 层次分析法; 二次规划

**中图分类号:** TP18

## Research for Methods Detemining the Weights of Case Characteristics

WU L iang-gang<sup>1</sup>, GAO Yang<sup>1</sup>, ZHANG Jin-long<sup>2</sup>

(1. The Institute of Business and Adm inistration, CSU T, Changsha 410083; 2. Scientific Research Division, HU ST, Wuhan 430074)

**Abstract** In order to detem ine the weights of case characteristics , the paper presents the methods of combining Delphi with AHP, quadratic programming and credit statistics The methods contribute to decrease subjective and increase objective in detemining the weights of case characteristics

**Keywords:** case-based reasoning; weights of case characteristic; analytical hierarchy process; quadratic programming

### 1 引言

事例推理 (Case-Based Reasoning, CBR) 技术是 A I 中一种新兴的推理方法。CBR 基于人的认知过程, 即人类求解某一新问题时, 往往把以前使用过的与该问题类似的事例联系起来, 运用过去解决该事例的经验和方法, 来解决当前问题。CBR 就象人们找医生看病那样, 总是喜欢找老医生, 特别是老中医, 其原因就是老医生看过的病例多, 经验丰富。基于事例的方法为传统的知识处理系统问题的解决提供了一条有效的途径。

事例是由组成事例的特征结构来表达的。事例中各个特征的重要性是不同的。一般是用事例特征权值的大小来描述事例特征的重要程度。事例特征权值在 CBR 方法中有其广泛的应用, 例如事例的相似度计算、事例的索引等方面。在很多 CBR 应用系统中, 事例特征权值的确定都是由专家主观确定的, 缺乏一定的客观性。本文针对这一问题提出了事例特征权值确定的几种方法。这些方法是针对多个专家及大量的事例来确定事例特征权值。

### 2 Delphi法和AHP法相结合确定权值

由一个专家确定各特征的权值带有很大的主观性, 为了减少这种主观性, 我们采用 Delphi 法由多个专

收稿日期: 1999-03-17

资助项目: 国家自然科学基金 (79770036)

家确定各特征两两之间的相对重要程度,再由 AHP (Analytical Hierarchy Process) 法确定各特征的权值,其具体步骤如下:

设每个事例包含有  $n$  个特征,有  $p$  个专家.

1) 采用 Delphi 问卷调查方式向  $p$  个专家问卷调查,其问卷形式设计如下:

$$\begin{array}{c} \text{特征} \\ \begin{matrix} 1 & 2 & \dots & n \\ 1 & \left( \begin{array}{cccc} b_{11} & b_{12} & \dots & b_{1n} \\ & b_{22} & \dots & b_{2n} \\ & & \ddots & \\ & & & b_{nn} \end{array} \right) \\ 2 \\ \vdots \\ n \end{matrix} \end{array}$$

这里  $b_{ij}$  的取值范围 Saaty<sup>[1]</sup> 建议为 1~9, 且  $b_{ji} = 1/b_{ij}$ . 例如,  $b_{ij} = 5$  表示第  $i$  特征相对于第  $j$  特征的重要程度值为 5, 说明第  $j$  特征比第  $i$  特征重要;  $b_{ij} = 1/5$  表示第  $i$  特征相对于第  $j$  特征的重要程度值为 5, 说明第  $i$  特征比第  $j$  特征重要.

2) 综合  $p$  个专家的意见或观点. 我们用几何平均值法来综合专家的意见或观点. 由第 (1) 步, 可得  $p$  个专家对各事例特征两两之间相对重要程度的打分表, 设  $a_{ij}^r$  表示第  $r$  个专家确定的第  $i$  事例特征相对于第  $j$  事例特征的重要程度值, 则  $a_{ij}^1, a_{ij}^2, \dots, a_{ij}^p$  分别表示  $p$  个专家确定的第  $i$  事例特征相对于第  $j$  特征的重要程度值, 令

$$a_{ij} = \left( \prod_{r=1}^p a_{ij}^r \right)^{1/p}$$

$a_{ij}$  反映了  $p$  个专家确定第  $i$  事例特征相对于第  $j$  事例特征重要程度的平均值.

3) 构造事例特征相互重要程度的判断矩阵. 由第 (2) 步的方法确定每个  $a_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, n$ , 则就构成了事例特征相互重要程度的判断矩阵.

$$A = (a_{ij})_{n \times n}$$

其中  $a_{ij} = 1/a_{ji}, a_{ii} = 1$

4) 用 AHP 法确定各特征的权值. AHP 法确定各特征权值用如下公式计算:

$$W_i = \frac{1}{n} \sum_{j=1}^n \left( \frac{a_{ji}}{\sum_{k=1}^n a_{kj}} \right) \quad i = 1, 2, \dots, n \quad (1)$$

由 (1) 式可得各特征权值的向量  $W = (W_1, W_2, \dots, W_n)^T$ ,  $W$  具有如下性质:

性质 1  $AW = \lambda_{\max} W$ , 这里  $\lambda_{\max}$  是矩阵  $A$  的最大特征值.

证明见参考文献 [1].

性质 2  $\sum_{i=1}^n W_i = 1$

证明

$$\begin{aligned} \sum_{i=1}^n W_i &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left( \frac{a_{ji}}{\sum_{k=1}^n a_{kj}} \right) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left( \frac{a_{ji}}{\sum_{k=1}^n a_{kj}} \right) \\ &= \frac{1}{n} \sum_{j=1}^n \frac{\sum_{i=1}^n a_{ji}}{\sum_{k=1}^n a_{kj}} = \frac{1}{n} \sum_{j=1}^n 1 = 1 \end{aligned}$$

上述步骤的计算过程见表 1, 表 2. 由表 2 可知五个特征的各数值分别为: 0.182, 0.155, 0.151, 0.273, 0.239

表 1

$a_{ij}$ 特征 特征	1	2	3	4	5
1	1.000	3.000	3.000	0.333	0.500
2	0.333	1.000	3.000	1.000	0.250
3	0.333	0.333	1.000	3.000	0.333
4	3.000	1.000	0.333	1.000	3.000
5	2.000	4.000	3.000	0.333	1.000
$\sum K$	6.666	9.333	10.333	5.666	5.083

表 2

$a_{ik}/\sum K$	特征 1	特征 2	特征 3	特征 4	特征 5	$\sum_j$	$W_i = \frac{1}{5} \sum_j$
特征 1	0.150	0.321	0.291	0.059	0.098	0.911	0.182
特征 2	0.050	0.107	0.291	0.177	0.049	0.647	0.155
特征 3	0.050	0.036	0.097	0.528	0.045	0.756	0.151
特征 4	0.450	0.107	0.030	0.177	0.591	1.365	0.273
特征 5	0.300	0.429	0.291	0.059	0.197	1.294	0.239

3 二次规划模型权值确定法

上述用 Delphi 法和 AHP 法进行事例特征权值的确定主要是针对某一个事例的事例特征权值的确定。对于同类型的问题,在事例库中可能存在很多事例,每个事例均有它们各自的事例特征权值。为了更有效地确定该类型问题的事例特征权值,需要综合该类问题各事例特征权值。为此,我们设该类型问题有  $m$  个事例,每个事例有  $n$  个事例特征。 $W_i = (W_{i1}, W_{i2}, \dots, W_{ij}, \dots, W_{in}) \quad i = 1, 2, \dots, m$ , 其中  $W_{ij}$  表示第  $i$  事例第  $j$  事例特征的权值,  $W_i$  表示第  $i$  事例的事例特征权值的分布,且  $\sum_{j=1}^n W_{ij} = 1$ 。此外,为了说明第  $i$  事例的事例特征权值分布  $W_i$  的好坏,我们引入一个评价好坏的评估值(或信用值)  $r_i$ ,  $r_i$  值越大,就说明  $W_i$  的数值分布就越正确,反之亦然。这样就可得如下一组数据,用  $W$  矩阵表示为:

$$W = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_m \end{pmatrix} = \begin{pmatrix} W_{11} & W_{12} & \dots & W_{1n} & r_1 \\ W_{21} & W_{22} & \dots & W_{2n} & r_2 \\ & & \ddots & & \\ W_{m1} & W_{m2} & \dots & W_{mn} & r_m \end{pmatrix} \tag{2}$$

设  $X_1, X_2, \dots, X_n$  分别表示所需求的事例特征权值,我们希望该权值的分布尽量接近评估值高的事例特征权值的分布。因此可构造如下模型:

$$\begin{aligned} \min & \sum_{i=1}^m r_i \sum_{j=1}^n (X_j - W_{ij})^2 \\ \text{s.t.} & \sum_{j=1}^n X_j = 1 \\ & X_j \geq 0 \end{aligned} \tag{3}$$

上述模型是一个二次规划模型

例 1  $W_1 = (0.3, 0.5, 0.2), W_2 = (0.3, 0.4, 0.3), W_3 = (0.2, 0.6, 0.2)$

$r_1 = 0.9 \quad r_2 = 0.5 \quad r_3 = 0.1$

代入上述模型,用二次规划模型求解方法,可得到  $X_1 = 0.28, X_2 = 0.48, X_3 = 0.24$

#### 4 信用统计法

针对(2)式的问题,用二次规划模型求解事例特征权值需要具备数学规划领域的知识,且求解过程比较繁琐.下面我们提出一种信用统计法来确定事例特征权值,该方法简单易懂,其公式构造如下:

$$W_j = \frac{1}{R} \sum_{i=1}^m r_i \cdot W_{ij} \quad j = 1, 2, \dots, n \quad (4)$$

$$R = \sum_{i=1}^m r_i$$

$W_j$  表示第  $j$  事例特征的权值,并且有  $\sum_{j=1}^n W_j = 1$ .

$$\begin{aligned} \sum_{j=1}^n W_j &= \frac{1}{R} \sum_{j=1}^n \sum_{i=1}^m r_i \cdot W_{ij} = \frac{1}{R} \sum_{i=1}^m \sum_{j=1}^n r_i \cdot W_{ij} = \frac{1}{R} \sum_{i=1}^m r_i \sum_{j=1}^n W_{ij} \\ &= \frac{1}{R} \sum_{i=1}^m r_i = 1 \end{aligned}$$

例 2 由例(1)中的数据,代入(4)式中,可得  $W_1 = 0.293, W_2 = 0.473, W_3 = 0.234$ .

#### 5 结束语

事例中各事例特征权值的正确确定,在 CBR 推理过程中起着重要的作用.例如在 CBR 中的相似度计算、事例的检索等方面都需要用到事例特征的权值.而传统的事例特征的权值确定带有很大的主观性.为减少主观性,增加客观性,针对某一单个事例,本文提出了 Delphi 和 AHP 相结合的权值确定法;而针对事例库中的大量事例,本文提出了二次规划模型权值确定法和信用统计权值确定法.这些方法提供了解决权值问题的一条途径.

#### 参考文献

- [1] Saaty T L. The Analytic Hierarchy Process: Planning, Priority Setting [M]. Resource Allocation McGraw-Hill, New York, 1980
- [2] Meade L M, Liles D H, Sarkis J. Justifying Strategic Alliances and Partnering: a Prerequisite for Virtual Enterprising[J]. Omega, Int J Mgmt Sci, 1997, 125(1): 29~ 42
- [3] Lawrence M Seifert, Joe Zhu. Identifying Excesses and Deficits in Chinese Industrial Productivity (1953- 1990): a Weighted Data Envelopment Analysis Approach [J]. Omega, Int J. Mgmt Sci, 1998, 26(2): 279~ 296.

(上接第 83 页)

- [5] Casdagli M. Nonlinear prediction of chaotic time series[J]. Physica D, 1989, 35: 335~ 356
- [6] Gencay R. Nonlinear prediction of noisy time series with feedforward networks[J]. Phys. Lett A, 1994, 187: 397~ 403
- [7] Smith L A. Identification and prediction of low dimensional dynamics[J]. Physica D, 1992, 58: 50~ 76
- [8] Albano A M, Passamante A, Hediger T, Farrell M E. Using neural nets to look for chaos[J]. Physica D, 1992, 58: 1~ 9
- [9] Farmer J D. Chaotic attractor of an infinite-dimension chaotic system [J]. Physica D, 1982, 4: 366~ 393
- [10] Sugihara G, May R M. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series[J]. Nature, 1990, 344: 734~ 741.
- [11] Barbanel H D I, Brown R, Sidorowich J J, L. Sh. Tsingring. The analysis of observed chaotic data in physical systems[J]. Rev Mod Phys, 1993, 65: 1331~ 1392