

特征加权集对分析方法

朱红宁 张 斌

(东北大学信息科学与工程学院 沈阳 110004)

摘 要 提出对集对分析理论的一种改进方法——特征加权集对分析。对于 Web 服务的 QoS 评价中常见的两类问题,描述了常权与变权的处理方法。通过应用比较说明了特征加权集对分析方法比联系数加权集对分析方法更适用于基于 QoS 评价的 Web 服务选取。

关键词 集对分析,特征加权,QoS 评价,服务选取

中图法分类号 TP393 **文献标识码** A

Method of Character Weighted Set Pair Analysis

ZHU Hong-ning ZHANG Bin

(College of Information Science and Engineering, Northeastern University, Shenyang 110004, China)

Abstract This paper proposed an improved method of SPA theory named character-weighted set pair analysis, described constant and variable weighted processing methods to the two kinds of familiar problems in QoS evaluation of Web services. It indicated that the method of character-weighted SPA was fitter than the method of connection number weighted SPA in the aspect of Web service selection based on QoS evaluation.

Key words SPA, Character weighted, QoS evaluation, Service selection

由于自然界中普遍存在不确定性,人类对自然的认识一直受不确定性的困扰,科学家们不断推出新的理论和方法来研究不确定性。大样本、多数据的不确定性问题,即随机不确定性问题,可以用概率论和数理统计解决;认识不确定性问题,即模糊不确定性问题,可以用模糊集^[1]和粗糙集^[2]解决;少数据、小样本、信息不完全和经验缺乏的不确定性问题,即少数据不确定性(灰性)问题,可以用灰色系统理论^[3]解决;另外,集对分析理论从两个集合同(同一)、异(差异不确定)、反(对立)的角度来研究不确定性,采取了对不确定性加以客观承认、系统刻画、具体分析的态度,从而使研究结果更加贴近实际^[4]。

本文在分析近年来有关集对分析理论的典型研究的基础上,提出对集对分析理论的一种改进方法——特征加权集对分析方法(Character Weighted SPA, W-SPA),并通过 Web 服务的 QoS 评价表述了方法的具体步骤。

1 SPA 的基础概念

集对分析^[5]是中国学者赵克勤于 20 世纪 80 年代末提出的,它提出从同、异、反 3 方面表征两集合(集对) $H(A, B)$ 的关系。它主要通过同、异、反联系数来描述。

(1) 联系度

$$\mu(w) = S/N + F/N i + P/N j \quad (1)$$

其中, N 指集对的特性总数, S 指共同具备的特性总数, P 指具有相对立的特性总数, $F = N - S - P$ 指既不相对立,又不

共同具备的特性总数。

S/N 为两集合在问题 w 下的同一度;

F/N 为两集合在问题 w 下的差异度;

P/N 为两集合在问题 w 下的对立度。

其中, $i \in [-1, 1], j$ 一般情况下取 -1 。可简记为 $\mu = a + bi + cj$; a, b, c 满足归一化条件 $a + b + c = 1$ 。

(2) 集对势

当联系度 $\mu = a + bi + cj$ 中 $c = 0$ 时,同一度 a 与对立度 c 的比值 a/c 为所论集对在指定问题背景下的联系势或集对势,用“ shi ”表示:

$$shi(H) = a/c \quad (2)$$

$a/c > 1$, 则称 $shi(H)$ 为集对同势;

$a/c = 1$, 则称 $shi(H)$ 为集对均势;

$a/c < 1$, 则称 $shi(H)$ 为集对反势。

(3) 四元联系数

四元联系数^[6]是一种扩展的联系数,表达式为

$$\mu = a + bi + cj + dk \quad (3)$$

其中, $a, b, c, d \in [0, 1], i \in [0, 1], j \in [-1, 0], k = -1$, a 表示同一度, b 表示同异度, c 表示反异度, d 表示相反度。

此时的集对势为

$$shi(H) = \frac{a+bi}{cj+dk} \quad (4)$$

2 相关工作

赵克勤的研究是原创的、基础的,他自 1989 年包头市召

到稿日期:2008-10-15 返修日期:2009-02-25 本文受国家自然科学基金(60773218)资助。

朱红宁(1968-),男,博士研究生,主要研究方向为 Web 服务 QoS, E-mail: honeyzhu@126.com; 张 斌(1964-),男,博士生导师,主要研究方向为服务行为、演进。

开的全国系统理论上提出集对分析以来,就不断地从哲学以及科学技术各个角度论述集对分析的思想、方法。其中文献[4]重点介绍了集对分析对不确定性的描述和处理方法,构成了一种新的不同于传统概率论和模糊集理论但又可能在一定条件下包含这两者的不确定性理论;文献[7]重点研究了集对分析中关于对立的分类与度量,指出了“倒数对立”的基础性、重要性;文献[8]指出,集对分析的核心思想是把确定性与不确定性恒作为一个确定不确定系统来处理,在这个确定不确定系统中,确定性与不确定性互相联系、互相影响、互相制约,并在一定条件下互相转化;文献[9]提出用集对分析研究诸如“弃权”、“和局”这类中介不确定的方法。

另外,国内的一些学者结合其它理论对集对分析进行了大量的研究与实践。其中代表性的有:文献[10]将模糊数与集对分析中的联系系数结合起来,提出模糊联系系数,并对其在工程建筑中的应用做了初步探讨;文献[11]利用模糊数学、物元分析、集对分析对传统预测方法加以改造,创立了集对索引法、集对类比法,为预测提供了更加精确科学的工具;文献[12]提出了基于集对分析的模式识别方法,给出了集对模式识别的3个判定原则,对SPA的发展及其在模式识别中的应用提供了理论支持;文献[13]提出了基于集对分析的集对决策支持系统(SP-DSS),并阐述了SP-DSS的决策原理和模型以及在病体系统诊断中的应用。

本文针对的是集对分析中的加权问题。下面重点探讨在集对分析中加权的典型研究。

文献[13]对联系系数 μ 中 a, b, c 加权,并归一化得联系值:

$$v = a + bi + cj \quad (5)$$

同理,对于四元联系系数有

$$v = a + bi + cj + dk \quad (6)$$

这种加权形式将权重加到了同一度、差异度和对立度上,试图增强对集合之间关系描述的准确度,提高决策的质量。我们称其为联系系数加权集对分析方法。

文献[14]在集对分析中引入了变权理论,在具体问题背景下,通过动态调整权向量,不但考虑各基本因素的相对重要次序,也考虑集对状态均衡程度,从而在一定程度上解决了“状态失衡”问题。具体地,构造均衡函数,通过求偏导得到变权向量 $(W(v))$,再将变权向量作用于由影响因素与目标之间同一度、差异度和对立度构成的矩阵 $(M(A, B, C))$,从而得到集对联系度 (μ) ,即

$$\mu = W(v) \times M(A, B, C) \times (1, i, j)^T = a + bi + cj \quad (7)$$

这种变权形式在具体问题的分析上能起到一定的积极作用,其关键在于均衡函数的构造。

以上研究的实质,不论常权还是变权都作用于同一度、差异度和对立度上,是对集对联系系数的一种扩展。尽管如此,在某些情况下,这种加权的联系度还是不能准确刻画集合之间的不确定性关系。

考虑下列基于QoS评价的Web服务选取的两个例子。

例1 相同QoS评价者(评价标准)

假设, QoS评价的考察因素为响应时间(Duration)、安全性(Security)、可靠性(Reliability)和费用(Cost),即评价指标集为 $\{D, S, R, C\}$;对于功能属性相同的备选Web服务 WS_1 和 WS_2 的评价指标向量分别为 $(5s, 4, 50\%, 1\$)$ 和

$(10s, 5, 30\%, 2\$)$;约束为响应时间10s以内,费用1\$以下,不考虑安全性和可靠性。则在备选Web服务与约束构成的集对 $H(ws, c)$ 中,联系度与集对势为

$$\mu_1 = \mu_2 = \frac{1}{4} + \frac{1}{2}i + \frac{1}{4}j, shi_1(H) = shi_2(H) = 1 \quad (8)$$

问题:无法分辨 WS_1 和 WS_2 ,其实 WS_1 优。

例2 不同QoS评价者(评价标准)

假设, 同例1;评价者甲,强调响应时间10s以内,不考虑其它指标;评价者乙,强调安全等级在4以上,不考虑其它指标。则在备选Web服务与约束构成的集对 $H(ws, c)$ 中,对于甲(用(1)表示),联系度为

$$\mu_1(1) = \mu_2(1) = \frac{1}{4} + \frac{3}{4}i \quad (9)$$

对于乙(用(2)表示),联系度为

$$\mu_1(2) = \mu_2(2) = \frac{1}{4} + \frac{3}{4}i \quad (10)$$

由式(9)、式(10)可得

$$\mu_1(1) = \mu_1(2), \mu_2(1) = \mu_2(2) \quad (11)$$

问题:无法分辨 WS_1 和 WS_2 ,其实对于甲, WS_1 优;对于乙, WS_2 优。

例1中的问题可以用“变权”解决,例2中的问题可以用“常权”解决,即它们都可以通过加权来解决,关键是将权加在哪里。

笔者认为除了在同一度、差异度和对立度上加权^[13,14]外,还可以在特性上加权,而且这种加权更基础、更说明问题。

3 特征加权集对分析

对于相关集合 A 与 B 构成的集对 $H(A, B)$,假设集对总共有 n 个特性,分别是 p_1, p_2, \dots, p_n ,对应的特征值为 v_1, v_2, \dots, v_n ,特性权重为 w_1, w_2, \dots, w_n 。在 n 个特性中,共同具备的特性总数为sameness,具有相对立的特性总数为opposition,既不相对立又不共同具备的特性(差异属性)总数为deference,即有 $s + d + o = n$ 。不妨设 p_1, p_2, \dots, p_s 为共同特性, $p_{s+1}, p_{s+2}, \dots, p_{s+d}$ 为差异属性; $p_{s+d+1}, p_{s+d+2}, \dots, p_n$ 为对立特性,于是有以下概念。

3.1 基本定义

(1)特征加权联系度

$\mu_w = a + bi + cj$, where

$$a = \frac{S(s)}{S(n)}, b = \frac{S(d)}{S(n)}, c = \frac{S(o)}{S(n)}$$

$$S(n) = \sum_{k=1}^n w_k v_k, S(s) = \sum_{k=1}^s w_k v_k, \quad (12)$$

$$S(d) = \sum_{k=s+1}^{s+d} w_k v_k, S(o) = \sum_{k=s+d+1}^n w_k v_k$$

and, $a + b + c = 1$

注意:一般地,特征加权求和之前对特征值要进行不确定性、无量纲、等数量级和一致性极性等一系列处理,这些不是本文的研究重点,恕不详陈。

在式(12)中, μ_w 为特征加权联系度, $S(n)$ 为总特征值加权和, $S(s)$ 为同一特征值加权和, $S(d)$ 为差异特征值加权和, $S(o)$ 为对立特征值加权和; a 为特征加权同一度, b 为特征加权差异度, c 为特征加权对立度。 $i \in [-1, 1], j$ 一般情况下取-1。

同理,可得到加权四元联系系数及集对势。

如果式(12)中的权是等权,通过计算可知式(12)与式(1)等价。

(2) 集对势

当式(12)中的对立特征值加权和 $S(o) = 0$ 时,特征加权同一度 a 与特征加权对立度 c 的比值 a/c 为所论集对在指定问题背景下的特征加权联系势或特征加权集对势,用“ shi_w ”表示:

$$shi_w(H) = S(s)/S(o) \quad (13)$$

$S(s)/S(o) > 1$, 则称 $shi_w(H)$ 为特征加权集对同势; $S(s)/S(o) = 1$, 则称 $shi_w(H)$ 为特征加权集对均势; $S(s)/S(o) < 1$, 则称 $shi_w(H)$ 为特征加权集对反势。

3.2 比较说明

(1) 与经典集对分析的比较

我们称赵克勤先生提出的集对分析理论为经典集对分析。

笔者认为,经典集对分析是从哲学观点出发,对事物普遍联系的宏观分析,是奠基性的理论。而由于人类认知的局限,对现实生活中的相关事物之间的联系的刻画往往受客观和主观因素的影响。因此,笔者在对集对分析理论研究的基础上,为了更准确地描述事物之间的联系,认为主观因素是不可或缺的。

(2) 与可加权和评价的比较

加权求和是求解多属性评价、多目标决策的可加性系统问题最常用、最主要的方法。

笔者认为通过属性取值、权重分析和目标确定的加权求和方法在处理评价、决策问题时,缺乏对同一、差异和对立的衡量。尽管通过对权重的调整可以增加对主观因素的考虑,但它是模糊的、不确定的。因此,笔者认为在处理可加权求和评价的问题时,应充分分析同一、差异和对立因素,以期提高评价的准确性。

3.3 权重分析

权重的分析应考虑主观和客观两方面因素,主观因素一般指使用者偏好,可以是偏好排序的梯度向量;客观因素一般指事物的内在联系,可以是数据相关性的分析。权分为常权和变权。

(1) 常权

权值为常数的权为常权。常权方法在一定程度上反映了影响事物各因素的相对重要次序,由于其简单和一定的合理性而被广泛应用。然而,无论各因素的组态(各因素的取值状况)如何,权向量都保持固定不变,这往往会造成实际问题中出现不合理的综合结果,即出现“状态失衡”问题。基于常权的特征加权集对分析方法也存在这样的问题。

(2) 变权

权值是变数的权为变权,即在分析时权重可以被动态调整。当特征值与目标值变化率(一阶导数)存在函数关系时,一般使用变权来做评价和决策的分析。

4 应用比较

4.1 特征加权集对分析在 Web 服务 QoS 评价中的应用

(1) 特征变权集对分析方法

考虑第 2 节例 1 中提到的问题,响应时间、约束与用户认可程度之间不是简单的二值关系。

如图 1 所示,当 Web 服务响应时间超过约束值($t_c = 10s$)时,随着时间增长,用户认可程度迅速下降,最后趋近于阈值($a_1 < 0$),代表放弃;相反地,当 Web 服务响应时间没有超过约束值时,随着时间减缩,用户认可程度迅速上升。这里有两个特殊部分:当 Web 服务响应时间刚好满足约束值时,用户认可程度直接跃进到某一阈值($0 < a_0 < 1$);当 Web 服务响应时间缩短到某一阈值($t_0 < t_c$)时,用户的认可程度并不上升,而是趋于一致的 100% 认可,表现为一条水平线。

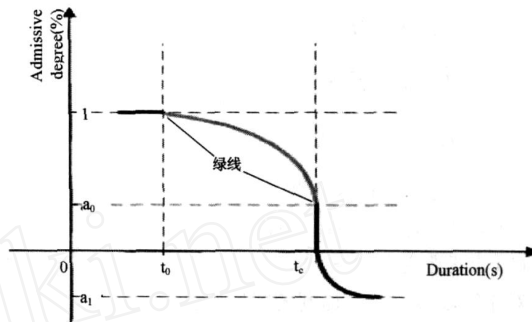


图 1 Web 服务响应时间、约束与用户认可程度关系

例 1 中两个 Web 服务的响应时间分别为 5s 和 10s, 虽然都没有超出约束值($t_c = 10s$), 但用户对其的认可程度是不一样的。根据 3.3 小节的讨论, 此处应该使用特征变权集对分析方法。

具体地, 首先构造均衡函数解决状态失衡。为不失一般性, 对图 1 绿线部分做如下描述。

用户认可程度(A)关于响应时间(D)的函数为 $A = f(t)$, 其中 $A \in [a_0, 1]$, $t \in [t_0, t_c]$; $t = t_0$ 时 $A = 1$, $t = t_c$ 时 $A = a_0$; $f'(t) < 0$ 。

由于用户认可程度的变化率与响应时间成反比, 则有

$$f'(t) = \frac{k}{t} \quad (14)$$

由微分方程求解, 得

$$f(t) = k \ln t + C \quad (15)$$

代入边界值, 得

$$k = \frac{1 - a_0}{\ln(\frac{t_0}{t_c})}, C = 1 - \frac{1 - a_0}{\ln(\frac{t_0}{t_c})} \ln t_0 \quad (16)$$

令 $t_0 = 2s$, $t_c = 10s$, $a_0 = 0.5$, 则: $k = -0.31$, $C = 1.215$ 。

用户认可程度与响应时间的关系函数为

$$A = f(t) = -0.31 \ln t + 1.215 \quad (17)$$

表 1 中, t 是响应时间 t 经过无量纲、一致性极性处理后的特征值, C-W Value 是变权 A 与 t 的乘积。

表 1 Web 服务响应时间、约束与用户认可程度关系

t(s)	A (%)	t	C-W Value
2	1.000124374	0.500	1.442874475
3	0.874430191	0.333	0.795940660
4	0.785248748	0.250	0.566437237
5	0.716074247	0.200	0.444921945
6	0.659554565	0.167	0.368104411
7	0.611767854	0.143	0.314386486
8	0.570373122	0.125	0.274291492
9	0.533860381	0.111	0.242970330
10	0.501198621	0.100	0.217667796

例 1 中, WS_1 和 WS_2 的费用 C 分别为 1 \$ 和 2 \$, 经过无量纲、一致性极性处理后, C 分别为 1 和 0.5。在不考虑安全

性和可靠性的情况下,简便计算如下:

不妨设 WS_1 和 WS_2 的费用权重为 0.7 和 0.5,这与表 1 中的变权满足归一化条件。

将以上特征值和权重代入式(12)可得

$$\begin{aligned}\mu_5 &= \frac{0.44}{0.44+0.28 \times 1} + \frac{0.28 \times 1}{0.44+0.28 \times 1^j} \\ &= 0.61+0.29j \\ \mu_{10} &= \frac{0.22}{0.22+0.5 \times 0.5} + \frac{0.5 \times 0.5}{0.22+0.5 \times 0.5^j} \\ &= 0.47+0.53j \\ shi_w^{(5)}(H) &= \frac{0.61}{0.29} = 2.1 \\ shi_w^{(10)}(H) &= \frac{0.47}{0.53} = 0.9\end{aligned}\quad (18)$$

可见,无论从特征加权联系度,还是从特征加权集对势来评价,其结果都是 WS_1 优于 WS_2 。

(2) 特征常权集对分析方法

对于第 2 节例 2 中提到的问题,可以简单地使用特征常权集对分析方法处理。

评价者甲重视响应时间,他对于 Web 服务特征的偏好排序为响应时间 费用 可靠性 安全性,则不妨设相应的权重为 $w_D^{(1)}=0.4, w_C^{(1)}=0.3, w_R^{(1)}=0.2, w_S^{(1)}=0.1$ 。

评价者乙重视安全性,他对于 Web 服务特征的偏好排序为安全 可靠性 费用 响应时间,则不妨设相应的权重为 $w_S^{(2)}=0.4, w_R^{(2)}=0.3, w_C^{(2)}=0.2, w_D^{(2)}=0.1$ 。

经过无量纲、一致性极性、等数量级处理, WS_1 的特征值为 $v_D^{(1)}=1, v_S^{(1)}=0.8, v_R^{(1)}=1, v_C^{(1)}=1$; WS_2 的特征值为 $v_D^{(2)}=0.5, v_S^{(2)}=1, v_R^{(2)}=0.6, v_C^{(2)}=0.5$ 。

由式(12)得

甲对 WS_1 和 WS_2 评价(特征加权联系度)为

$$\mu_1(1) = \frac{2}{49} + \frac{47}{49}i, \mu_2(1) = \frac{2}{57} + \frac{55}{57}i$$

乙对 WS_1 和 WS_2 评价(特征加权联系度)为

$$\mu_1(2) = \frac{8}{23} + \frac{15}{23}i, \mu_2(2) = \frac{40}{73} + \frac{33}{73}i$$

可见,对于甲 WS_1 较优,对于乙 WS_2 较优。这个结论符合实际。

4.2 与联系系数加权的比较

由第 2 节例 1 和例 2 的结论,经典集对分析无法分清 WS_1 和 WS_2 的优劣。

文献[13]提到的联系系数加权方法,联系度由式(5)确定,其中 μ_1, μ_2 由文献[15]确定,即需要专家对联系系数进行半定

性、半定量的分析、评价,用区间评价值叠加成区间权重,再采取平均准则处理区间权重,最后归一化得到联系度权重。这种方法受专家影响太大,不适合 Web 服务的 QoS 评价。

特征加权集对分析中的联系度和集对势对于 Web 服务的 QoS 评价服务事物的客观规律,能更贴切地进行基于 QoS 评价的 Web 服务的比较、选取。

结束语 在对集对分析相关理论和可加权和评价研究的基础上,笔者提出了特征加权集对分析方法。它不同于目前的联系系数加权,是对特征的加权。经应用比较分析,特征加权集对分析方法比联系系数加权集对分析方法在描述相关事物之间的联系方面更准确、合理。

参考文献

- [1] Zadeh L. A. Fuzzy sets [J]. Information and Control, 1965, 8: 338-356
- [2] DENG Julong. Control problem of grey systems [J]. System & Control Letters, 1982(3)
- [3] Pawlak Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11: 341-356
- [4] 赵克勤. 集对分析对不确定性的描述和处理 [J]. 信息与控制, 1995, 24(3): 162-162
- [5] 赵克勤. 集对分析及其应用初探 [M]. 浙江: 浙江科技出版社, 2000
- [6] 赵克勤. 四元联系系数的事态排序及应用 [M]. 集对分析与界壳论的研究与应用. 北京: 北京气象出版社, 2002
- [7] 赵克勤. 基于集对分析的对立分类、度量及应用 [J]. 科学技术与辩证法, 1994, 11(2): 26-30
- [8] 赵克勤. 集对分析中的不确定性理论 [J]. 大自然探索, 1995, 14(4): 87-88
- [9] 赵克勤. 集对论——一种新的不确定性理论方法与应用 [J]. 系统工程, 1996, 14(1): 18-23
- [10] 黄树林, 张江, 李华, 等. 模糊化的联系系数及其应用初探 [M]. 模糊集理论与应用. 河北: 河北大学出版社, 1998
- [11] 朱军. 基于物元的集对分析预测 [C]. 第四届全国集对分析论文. 1998
- [12] 李光军, 贺仲雄. 基于 SPA 的模式识别 [J]. 通讯和计算机, 2007, 4(6)
- [13] 周泽南, 贺仲雄. 基于集对分析的决策支持系统及其应用 [J]. 通讯和计算机, 2006, 3(4)
- [14] 单鑫, 董文洪. 基于变权 SPA 的空中目标威胁综合评估方法 [J]. 现代防御技术, 2007, 35(5)
- [15] 闫皓, 贺仲雄. 权重分析系统 [J]. 系统工程与电子技术, 1992, 4: 41-45

(上接第 207 页)

- [3] Farrell J, Lausen H. Semantic annotations for WSDL and XML schema [EB/OL]. <http://www.w3.org/TR/sawSDL>, 2007
- [4] Akkiraju R, Sapkota B. Semantic annotations for WSDL and XML schema—Usage guide [EB/OL]. <http://www.w3.org/TR/sawSDL-guide/>, 2007
- [5] Cabreira F, Copeland G, Cox B. <http://www.ibm.com/developerworks/library/ws-transpec/>
- [6] Chen Shyi-Ming. Fuzzy backward reasoning using fuzzy Petri nets [J]. IEEE Transaction on System, Man, and Cybernetics-part B, 2000, 30(6): 846-856
- [7] 乐晓波, 葛敬军. 自动制造系统设计和仿真中的死锁控制 [J]. 计

算机工程, 2004, 31(4): 192-194

- [8] GAO Mei-mei, WU Zhi-ming. A fuzzy reasoning Petri net model and its reasoning algorithm [J]. Journal of Shanghai Jiao tong University, 1999, 4(2): 5-9
- [9] Men Peng, Duan Zhang-hua. Backward Reasoning Algorithm for Petri Nets Based on Algebra [J]. Journal of System Simulation, 2007, 19: 161-163
- [10] Borujij A, Boutayeb M, Cecchin T. A decentralized approach for computing invariants in large scale and interconnected Petri nets [C]. Proc. of the 1997 IEEE Int'l Conf. on System, Man and Cybernetics. Orlando: IEEE Computer Society, 1997: 1741-1746