

## 基于多准则决策和相似度评价的 Web 服务推荐 SOA 系统

周 雄, 王莉莉

(重庆工商大学 融智学院, 重庆 401320)

**摘 要:**为解决 Web 服务决策过程中新增信息多且信息不精确的问题,提出一种基于 Web 服务推荐的面向服务体系结构(SOA)系统。针对多准则决策分析过程特点,构建基于层次 SOA 结构的服务评估系统。给出权重时变多准则相似评估算法,根据相似评估提高 Web 服务推荐精度,并结合权重时变过程深入考虑各时段准则权重,增加 Web 服务的决策科学性。通过 Yahoo 域名 Web 服务推荐数据集上的实验,验证所提算法具有较高的 Web 服务推荐精度和较低的排序损耗。

**关键词:**多准则决策;相似度评价;Web 服务;面向服务体系结构系统;域名

**中文引用格式:**周 雄,王莉莉. 基于多准则决策和相似度评价的 Web 服务推荐 SOA 系统[J]. 计算机工程,2017,43(3):187-192,199.

**英文引用格式:**Zhou Xiong, Wang Lili. SOA System for Web Service Recommendation Based on Multi-criteria Decision Making and Similarity Evaluation[J]. Computer Engineering, 2017, 43(3):187-192, 199.

### SOA System for Web Service Recommendation Based on Multi-criteria Decision Making and Similarity Evaluation

ZHOU Xiong, WANG Lili

(Rongzhi College, Chongqing Technology and Business University, Chongqing 401320, China)

**[Abstract]** In the Web service decision making process, there are problems of new and inaccurate information, so this paper proposes a Service-oriented Architecture (SOA) system for Web service recommendation. Based on the characteristics of the multi-criteria decision analysis, it constructs the hierarchical SOA structure based service evaluation system. It also proposes multi-criteria similarity evaluation algorithm of weight time variation to improve Web service recommendation accuracy. Combined with the weight-time-varying process, the criteria weights for each period are further considered, which can improve the scientific decision making of Web services. Experiments on the Web service recommendation data set of Yahoo domain name verifies that the proposed algorithm has a higher Web service recommendation accuracy and lower sorting loss.

**[Key words]** Multi-criteria Decision Making (MCDM); similarity evaluation; Web service; Service-oriented Architecture (SOA) system; domain name

**DOI:**10.3969/j.issn.1000-3428.2017.03.032

## 0 概述

常用的 Web 服务决策方法是层次分析法 (Analytic Hierarchy Process, AHP), 如: 文献[1]提出模糊优先顺序的方法, 文献[2]提出基于层次结构的模糊层次分析法改进项目风险评估, 文献[3]提出基于模糊层次分析法在虚拟维修训练评估中应用基于 AHP 分析法的指标评价体系等。但 AHP 算法存在缺点<sup>[4-5]</sup>: 1) 在方案拓展性上效果不佳; 2) 用于定量数据较少, 可靠性不佳; 3) 评价指标过多时, 数据量

大, 权重不好确定; 4) 特征精确表示较为复杂等。特别是 AHP 算法用在 Web 服务推荐中时, 存在计算复杂度过高的问题。

本文针对 Web 服务推荐算法中信息不精确及新增信息较多, 传统精确算法效果不理想且计算复杂度过高的问题, 提出一种基于多准则决策和相似度评价的 Web 服务推荐的面向服务的体系结构 (Service-oriented Architecture, SOA)。基于多准则决策分析过程特点, 构建基于层次 SOA 结构的服务评估系统, 解决 AHP 方法存在的缺陷。利用权重时

**作者简介:**周 雄(1980—), 男, 副教授, 主研方向为信息系统、数据库技术; 王莉莉, 讲师、硕士。

**收稿日期:**2016-04-28 **修回日期:**2016-06-07 **E-mail:**wanglil2006@qq.com

变多准则相似评估算法,并结合权重时变过程深入考虑各时段准则权重,增加 Web 服务的决策科学性。

## 1 相关技术

近年来,SOA<sup>[6]</sup>是集成业务应用程序的体系结构样式,由不同服务提供商提供的服务在 SOA 中扮演关键角色。服务可以是原子或综合服务。在复合服务中,单个服务的功能取决于另一个原子服务。服务组合已成为选择服务的决策问题,是一种新的服务,称为高可靠性的复合服务<sup>[7-8]</sup>。例如,文献[9]确定 SOA 结构影响可靠性的因素,通过使用行业审查,提出了影响系统整体可靠性的 3 个主要因素。文献[10]提出 SOA 结构的 4 个因素,即可重用性、用于业务文件配置、组件依赖性和应用程序的复杂性分析。文献[11]基于 SOA 框架,提出交流和服务监测方法,用于在故障服务的情况下,有效地执行命令的过程等。但是上述文献在应用 SOA 结构时,未考虑到服务间的相似度,并且采用的实验对象均为精确数据,不适于定性评价应用。而 Web 服务特点是存在大量定性评价元素,直接应用 SOA 框架对 Web 服务进行推荐的效果并不理想。同时,上述文献在进行 Web 服务决策时,并未考虑多准则问题<sup>[3]</sup>,这与真实情况不符。虽然,当前大多数 Web 服务采用了多准则决策:如文献[12]设计基于 SMICloud 框架的 Web 服务推荐过程,首先基于 SMI 系统实现 Web 服务的属性质量进行评估,然后采用 AHP 方法对 Web 服务评定等级。文献[13]设计 Web 服务识别程序,获得 Web 服务最佳匹配系统。文献[14]设计了完整 AHP 框架下的 Web 服务评估体系,实现 Web 服务有效评价等。但是这些文献在考虑多准则决策时,并未针对 SOA 系统进行设计。

## 2 准则模糊化

模糊数可以表示不确定形式的数量,同时可表示为一个被称为隶属度的函数。这个隶属函数表示为 0 和 1 之间。三角模糊数可定义为三重态 $(a_1, a_2, a_3)$ ,如图 1 所示,其所对应的隶属函数为<sup>[15]</sup>:

$$\mu_A(x) = \begin{cases} \frac{(x-a_1)}{(a_2-a_1)}, & a_1 \leq x \leq a_2 \\ 0, & \text{otherwise} \\ \frac{(a_3-x)}{(a_3-a_2)}, & a_2 \leq x \leq a_3 \end{cases} \quad (1)$$

其中, $a_1, a_2, a_3$  被称为下限、最可能的值和上限,可表示为 $(l, m, u)$ 。

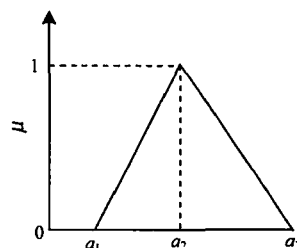


图 1 三角隶属函数

令  $A = (a_1, a_2, a_3)$  和  $B = (b_1, b_2, b_3)$  为 2 个三角模糊数,则其基本的算术运算过程为:

求逆操作:

$$A^{-1} = \left( \frac{1}{a_3}, \frac{1}{a_2}, \frac{1}{a_1} \right) \quad (2)$$

增补操作:

$$A + B = (a_1 + b_1, a_2 + b_2, a_3 + b_3) \quad (3)$$

减法操作:

$$A - B = (a_1 - b_3, a_2 - b_2, a_3 - b_1) \quad (4)$$

标量乘法:

$$\begin{cases} \forall k > 0, k \in R, kA = (ka_1, ka_2, ka_3) \\ \forall k < 0, k \in R, kA = (ka_3, ka_2, ka_1) \end{cases} \quad (5)$$

乘法:

$$AB = (a_1 b_1, a_2 b_2, a_3 b_3) \quad (6)$$

除法:

$$\frac{A}{B} = \left( \frac{a_1}{b_3}, \frac{a_2}{b_2}, \frac{a_3}{b_1} \right) \quad (7)$$

为进行定量评价的成对比较,可对三角模糊数 $\bar{x}$ 进行定义,如表 1 所示,其中, $1 \leq x \leq 9$ 。

表 1 模糊数定性评估

模糊数	隶属函数
$\bar{1}$	$(1, 1, 3)$
$\bar{x}$	$(x-2, x, x+2), x=3, 5, 7$
$\bar{9}$	$(7, 9, 11)$

决策问题一般包含多个备选方案,可表示为  $A_i$  ( $i=1, 2, \dots, n$ ),准则集为  $C_j$  ( $j=1, 2, \dots, m$ ),性能等级  $x_{ij}$ ,  $i=1, 2, \dots, n, j=1, 2, \dots, m$  表示在考虑准则  $C_j$  决策矩阵情况下,每个选项的性能表现。决策权重可表示为  $W = (w_1, w_2, \dots, w_m)$ 。

## 3 Web 服务推荐过程

如前所述,Web 服务推荐过程中,因为存在服务器工作负载、服务冗余计算、可重用性以及故障等评判准则,单纯采用准则叠加方式,容易造成准则特征的同化,不利于 Web 服务的合理组合推荐。为此,采用基于相似度的多准则评价方式,目的是兼顾各个准则对于 Web 服务推荐的影响,使得 Web 服务的推荐过程更为合理。

### 3.1 相似度评价步骤

相似度评价过程的具体步骤为:

**步骤 1** 对表 1 所描述模糊判断矩阵或是乘以权重矩阵( $W$ )或是乘以准则选项( $C_j$ ):

$$C_j \text{ or } W = \begin{bmatrix} \bar{a}_{11} & \bar{a}_{12} & \cdots & \bar{a}_{1k} \\ \bar{a}_{21} & \bar{a}_{22} & \cdots & \bar{a}_{2k} \\ \vdots & \vdots & & \vdots \\ \bar{a}_{k1} & \bar{a}_{k2} & \cdots & \bar{a}_{kk} \end{bmatrix} \quad (8)$$

其中:

$$\bar{a}_{ls} = \begin{cases} \bar{1}, \bar{3}, \bar{5}, \bar{9}, l < s \\ 1, l = s, l, s = 1, 2, \cdots, k \\ 1/\bar{a}_{sl}, l > s \end{cases} \quad (9)$$

**步骤 2** 决策矩阵( $X$ )和权重向量( $W$ )可计算为:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad (10)$$

$$W = (w_1, w_2, \cdots, w_m) \quad (11)$$

其中:

$$x_{ij} \text{ or } w_j = \frac{\sum_{l=1}^k \bar{a}_{ls}}{\sum_{l=1}^k \sum_{s=1}^k \bar{a}_{ls}} \quad (12)$$

其中,  $i = 1, 2, \cdots, n, j = 1, 2, \cdots, m, k = m \text{ or } n, x_{ij}$  和  $w_j$  分别为考虑准则  $C_j$  的模糊性能,以及考虑总体目标问题的准则  $C_j$  的模糊权重。

**步骤 3** 模糊评价矩阵  $Z$  为考虑每个准则的,所有选项的整体性能表现,可通过乘以权重向量获得决策矩阵:

$$Z = \begin{bmatrix} w_1 x_{11} & w_2 x_{12} & \cdots & w_m x_{1m} \\ w_1 x_{21} & w_2 x_{22} & \cdots & w_m x_{2m} \\ \vdots & \vdots & & \vdots \\ w_1 x_{n1} & w_2 x_{n2} & \cdots & w_m x_{nm} \end{bmatrix} \quad (13)$$

**步骤 4** 利用性能矩阵的  $\alpha$  割方法,查找区间性能矩阵,其中,  $0 \leq \alpha \leq 1$ :

$$Z_\alpha = \begin{bmatrix} [z_{11l}^\alpha, z_{11r}^\alpha] & [z_{12l}^\alpha, z_{12r}^\alpha] & \cdots & [z_{1ml}^\alpha, z_{1mr}^\alpha] \\ [z_{21l}^\alpha, z_{21r}^\alpha] & [z_{22l}^\alpha, z_{22r}^\alpha] & \cdots & [z_{2ml}^\alpha, z_{2mr}^\alpha] \\ \vdots & \vdots & & \vdots \\ [z_{n1l}^\alpha, z_{n1r}^\alpha] & [z_{n2l}^\alpha, z_{n2r}^\alpha] & \cdots & [z_{nml}^\alpha, z_{nmr}^\alpha] \end{bmatrix} \quad (14)$$

其中:

$$\begin{cases} \alpha_l = [\alpha \cdot (m - l)] + l \\ \alpha_r = u - [\alpha \cdot (u - m)] \end{cases} \quad (15)$$

**步骤 5** 使用乐观指数  $\lambda$  进行脆性矩阵计算:

$$Z_\alpha^{\lambda'} = \begin{bmatrix} z_{11\alpha}^{\lambda'} & z_{12\alpha}^{\lambda'} & \cdots & z_{1m\alpha}^{\lambda'} \\ z_{21\alpha}^{\lambda'} & z_{22\alpha}^{\lambda'} & \cdots & z_{2m\alpha}^{\lambda'} \\ \vdots & \vdots & & \vdots \\ z_{n1\alpha}^{\lambda'} & z_{n2\alpha}^{\lambda'} & \cdots & z_{nm\alpha}^{\lambda'} \end{bmatrix} \quad (16)$$

其中,  $0 \leq \lambda \leq 1, z_{ija}^{\lambda'} = \lambda z_{ija}^\alpha + (1 - \lambda) z_{ija}^\beta$ 。因为  $\lambda$  为固定值,并可表示为  $\lambda = 1, \lambda = 0.5$  和  $\lambda = 0$ ,所有这些值可被决策者用来表示乐观、温和或悲观的看法。在下面的例子中,使用  $\lambda = 0.5$  来计算脆矩阵。

**步骤 6** 将标准化方法应用于脆性矩阵中,可得归一化性能矩阵:

$$Z_\alpha^\lambda = \begin{bmatrix} z_{11\alpha}^\lambda & z_{12\alpha}^\lambda & \cdots & z_{1m\alpha}^\lambda \\ z_{21\alpha}^\lambda & z_{22\alpha}^\lambda & \cdots & z_{2m\alpha}^\lambda \\ \vdots & \vdots & & \vdots \\ z_{n1\alpha}^\lambda & z_{n2\alpha}^\lambda & \cdots & z_{nm\alpha}^\lambda \end{bmatrix} \quad (17)$$

其中:

$$z_{ija}^\lambda = \frac{z_{ija}^{\lambda'}}{\sqrt{\sum_{i=1}^n (z_{ija}^{\lambda'})^2}} \quad (18)$$

**步骤 7** 考虑所有准则选项的正理想解  $A_\alpha^{+\lambda}$  和负理想解  $A_\alpha^{-\lambda}$ :

$$\begin{cases} A_\alpha^{+\lambda} = (z_{1\alpha}^{+\lambda}, z_{2\alpha}^{+\lambda}, \cdots, z_{m\alpha}^{+\lambda}) \\ A_\alpha^{-\lambda} = (z_{1\alpha}^{-\lambda}, z_{2\alpha}^{-\lambda}, \cdots, z_{m\alpha}^{-\lambda}) \end{cases} \quad (19)$$

其中:

$$\begin{cases} z_{ja}^{+\lambda} = \max(z_{1ja}^\lambda, z_{2ja}^\lambda, \cdots, z_{mja}^\lambda) \\ z_{ja}^{-\lambda} = \min(z_{1ja}^\lambda, z_{2ja}^\lambda, \cdots, z_{mja}^\lambda) \end{cases} \quad (20)$$

**步骤 8** 求取正、负理想解的距离。根据文献[9]提出的 2 个三角模糊数  $A_1 = (a_1, b_1, c_1)$  和  $A_2 = (a_2, b_2, c_2)$  的距离公式进行计算:

$$\begin{aligned} d(A_1, A_2) &= \sqrt{\frac{1}{3}[(a_1 - a_2)^2 + (b_1 - b_2)^2 + (c_1 - c_2)^2]} \\ &\times \begin{cases} d_i^+ = \sum_{j=1}^k d(\tilde{v}_{ij}, \tilde{v}_j^+), i = 1, 2, \cdots, m \\ d_i^- = \sum_{j=1}^k d(\tilde{v}_{ij}, \tilde{v}_j^-), i = 1, 2, \cdots, m \end{cases} \end{aligned} \quad (21)$$

**步骤 9** 求取每个备选 Web 服务的接近系数(CC)并对其进行排序,选取具有最高接近系数的备选 Web 服务:

$$CC_i = \frac{d_i^-}{d_i^+ + d_i^-} \quad (22)$$

上述基于相似度的组合推荐方式,优点是无需设置 Web 推荐过程的目标,降低人为参与导致推荐过程缺乏客观公正,可实现全自动的组合推荐。

### 3.2 准则权重时变

在进行 Web 服务推荐过程中,服务的时间属性对

于其推荐评价具有直接影响。在指导网络中,早期的 Web 服务或者版本较低,或者已被淘汰,在进行 Web 服务推荐过程中,应该尽量选取较近时间的 Web 服务。为实现上述目的,采用时变准则权重,目的是区分各时段准则的不同影响。若  $n$  个权重时段分别为  $t_1, t_2, \dots, t_n$ , 则  $t_i$  对应权重时变逻辑衰减过程为:

$$\omega_i = A + \frac{K - A}{(1 + e^{-B(\Delta t_i - M)})^{0.5}} \quad (23)$$

其中,  $\Delta t_i$  是准则时段及对比时段之间的距离;  $A, K, B, M$  为影响系数,  $A$  为衰减过程的下包络线;  $K$  为衰减过程的上包络线;  $B$  是增长率影响系数;  $M$  为最大间隔。权重选取时设置时段最近权重为 1, 时段最远权重为 0.4。

为令所有时段 Web 服务决策结果可用,需进行布尔矩阵构建:

$$U = \begin{matrix} & t_1 & t_2 & \cdots & t_n \\ \begin{matrix} S_1 \\ S_2 \\ \vdots \\ S_m \end{matrix} & \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mn} \end{bmatrix} \end{matrix} \quad (24)$$

式(24)所示矩阵元素  $u_{ij}$  分别为 Web 服务  $S_i$  及其对应时段  $t_j$ , 如果 Web 服务  $S_i$  对于时段  $t_j$  为对应时段 Web 服务等级最高, 那么  $u_{ij} = 1$ 。则式(24)对于矩阵每列对应该时段 Web 可用服务决策输出, 矩阵行表示所有时段 Web 服务决策输出。基于上述矩阵可得 Web 服务  $S_i$  对应融合等级输出结果  $R_i$ :

$$R_i = \sum_{j=1}^n \omega_j u_{ij} \quad (25)$$

其中,  $\omega_j$  是准则影响权重时变值。对所有的 Web 服务循环执行上述过程, 可得所有 Web 服务对于整个时段的等级排名。上述计算过程可基于式(24)矩阵, 对权重列向量执行乘积操作:

$$\begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_m \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mn} \end{bmatrix} \cdot \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{bmatrix} \quad (26)$$

根据式(26)所得融合等级  $R$  结果, 选择等级最大输出  $R_k$  的 Web 服务  $S_k$  当作决策最终输出。

### 3.3 Web 服务评估

进行 Web 服务排序时, 融合过程选用持续时间内准则进行判定, 并考虑时变权重进行分析, 而非文献[10]所提权重准则均值方式, 可实现 Web 服务排序过程局部极值问题, 以及信息时变的性能丢失问题。所设计的 Web 服务排序过程为:

#### 步骤 1 时段划分

利用 SOA 系统构建目标可靠性估计, 将持续 Web 服务进行时段划分, 使用者利用 Web 服务准则  $C_j (j=1, 2, \dots, n)$  进行判定, 而每时段 Web 服务可利用决策模块从准则库内进行信息提取, 如图 1 所示。

#### 步骤 2 准则选取

查询者对 Web 服务准则  $C_j (j=1, 2, \dots, n)$  进行重要性排序, 并根据特定偏好进行 Web 服务决策。这里提出准则权重自适应时变选取方式, 可使得准则  $C_j (j=1, 2, \dots, n)$  选取过程简化。

#### 步骤 3 准则排序

对 Web 服务所有时段准则进行决策矩阵性能构建, 基于权重准则进行 Web 服务决策。因为时段具有不交叉特征, 上述排序过程可在各时段循环执行。

#### 步骤 4 权重时变决策

根据时段不同对时变权重进行对应赋值, 并按照时段远近进行赋值, 权重降低方式为  $1 \rightarrow 0.4$  逐渐降低, 代表近时段段的准则要比远段准则更为重要。基于该权重时变结果赋值可对步骤 3 所示各时段最佳 Web 服务执行决策融合。

在上述研究基础上, 给出基于 SOA 的服务排序总体框架, 该框架基于模糊成对比较和相似度计算方法, 并结合层次化结果进行设计。所有这些工作都是在一个层次结构中, 如图 2 所示。

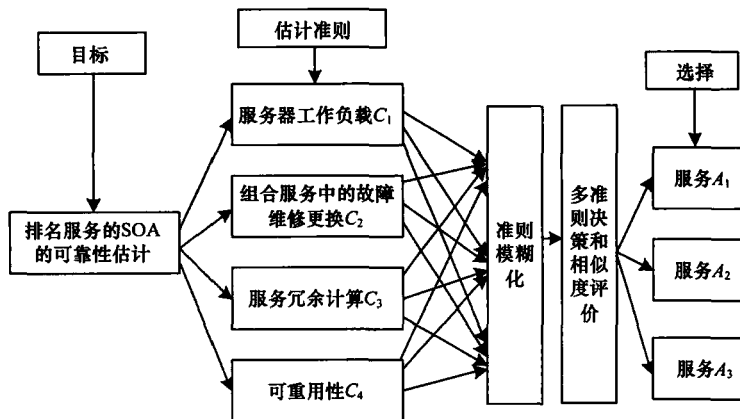


图 2 Web 服务评估算法框架

4 实验分析

实验对象选取 Yahoo 网站的 11 组 Web 服务排序数据集<sup>[13]</sup>,并对算法排序推荐进行实验验证。对比排序方法选择概率联合 Web 服务排序方法(JPMLC)<sup>[8]</sup>、logistic 回归细粒度 Web 服务排序方法(IBLR-ML)<sup>[14]</sup>、RBF 网络最速 Web 服务排序优化(SDRBF)<sup>[15]</sup>。硬件设置:处理器 i7-6300HQ 3.5 GHz,内存 16 GB ddr4,2 400 GHz,仿真平台 Matlab2013a。

对选取的 Web 服务排序数据集执行特征提取,并对 Web 服务排序数据集进行降维,仅约有 2% 数量及出现频次最高的文本进行排序服务推荐验证,其余数据则进行删除处理。将单独的文本形成矢量形式,每个矢量维度代表该字在文中出现的频次。每组实验数据集约含 2 500 组样本用于模型训练,有 3 500 组数据用于测试,类别数均值设为 30。所涉及其余参数如表 1 所示。

表 1 数据信息						
序号	数据集	$T$	$C$	$DC/\%$	$MNC$	$RC/\%$
1	Arts	452	26	44.48	11	19.23
2	Business	443	30	42.19	10	50.00
3	Computers	683	33	29.58	17	39.39
4	Education	553	33	33.47	7	57.58
5	Entertainment	639	21	29.29	9	28.57
6	Health	613	32	48.07	7	53.13
7	Recreation	611	22	30.18	13	18.18
8	Reference	796	33	13.76	5	51.52
9	Science	753	40	34.75	7	35.00
10	Social	1 017	39	20.95	9	56.38
11	Society	646	27	41.87	13	25.93

在表 1 所示数据信息中, $T$  是条目数量, $C$  是类别的数量, $DC$  是样本属多类别占比, $MNC$  是个体样本分配的类别最大数量, $RC$  是类别稀有占比。实验数据可分为 1 500 组数据训练集,数据集 中的 1 000 组数据对分类器进行构建,500 组进行数据测试。调整参数主要有: $\varphi_{ini},\tau_1,\tau$ 。其余参数有: $n_{max}\rightarrow\infty,\rho=100,\alpha=0.2,\eta=0.8,\tau_2=1/300$ ,上述参数借鉴相关标签推荐的参数设置,在 Web 服务排序实际应用中,这些参数对于算法的性能具有较小的影响度。评估指标选取如下:

指标 1 海明损耗( $hl$ ),表征实例  $d_j$  分类错误数量:

$$hl=\frac{1}{p}\sum_{j=1}^p\frac{1}{|C|}|P_j\Delta C_j|$$

(27)

其中, $|C|$  是所属类别量; $\Delta$  为  $d_j$  类别预测  $P_j$  与类别合理度  $C_j$  集合对称差异,类别预测排序级别高于设

定的阈值  $\tau$ 。  
指标 2 错误率( $E_{error}$ ),主要评估实例  $d_j$  类别排序最高是否属于类别合理集  $C_j$ :

$$E_{error}=\frac{1}{p}\sum_{j=1}^pE_{error}^j$$

(28)

$$E_{error}^j=\begin{cases}0, & \text{if}[\arg\max_{c\in C}f(d_j,c)\in C_j] \\ 1, & \text{otherwise}\end{cases}$$

(29)

其中, $[\arg\max_{c\in C}f(d_j,c)\in C_j]$  输出实例  $d_j$  类别最高排序。

指标 3 覆盖率( $C_{cover}$ )主要评估类别等级所需降低阈值,保证测试实例  $d_j$  属于所有类别:

$$C_{cover}=\frac{1}{p}\sum_{j=1}^p(\max_{c\in C_j}r(d_j,c)-1)$$

(30)

其中, $\max_{c\in C_j}r(d_j,c)$  为测试实例  $d_j$  的最大等级类别集。

指标 4 目标排序损耗( $C_{rloss}$ ),主要评估类别对  $\langle c_k,c_l\rangle$  实例  $d_j$  的排序分数:

$$C_{rloss}=\frac{1}{p}\sum_{j=1}^p\frac{|\{(c_k,c_l)|f(d_j,c_k)\leq f(d_j,c_l)\}|}{|C_j|\cdot|\bar{C}_j|}$$

(31)

其中, $(c_k,c_l)\in C_j\times\bar{C}_j;\bar{C}_j$  为  $C_j$  的类别补集。

指标 5 排序精度( $C_{avep}$ ),主要评估实例  $d_j$  排序精度:

$$C_{avep}=\frac{1}{p}\sum_{j=1}^p\frac{1}{|C_j|}\sum_{k=1}^{|C_j|}N_{precis}^j(R_{jk})$$

(32)

其中, $R_{jk}$  是位置  $k$  离等级最高的目标距离,对于实例  $d_j,c_i\in C_j,N_{precis}^j(R_{jk})$  为  $R_{jk}$  类别相关数。

在以上所述指标中,除排序精度  $C_{avep}$  外,越小代表 Web 服务的排序效果越好。排序最佳为: $hl=E_{error}=C_{cover}=C_{rloss}=0$  且  $C_{avep}=1$ 。

图 3 所示是选取的验证算法的海明指标数据对比情况,此指标数值越小代表排序过程的结果越佳。

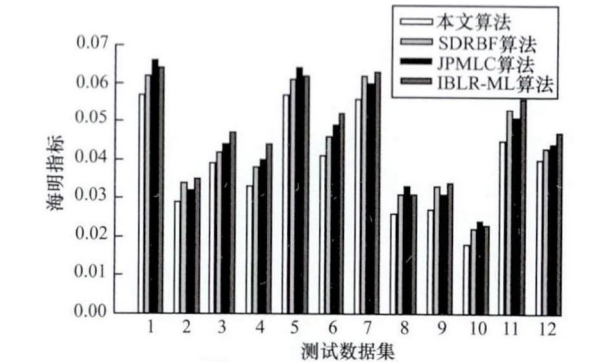


图 3 海明指标比较

图 3 所示横坐标数字分别对应上述表 1 所示实验数据集的编号,数字 12 表示算法的指标均值。可知在海明 Web 服务排序指标上,本文算法相对于 JPMLC,IBLR-ML 和 SDRBF 这 3 种对比算法性能更佳,SDRBF 算法的海明 Web 服务排序指标相对另

2 种算法性能更好,排名第 2,其他 2 种算法性能较为接近,互有优势。

图 4 所示为 JPMLC,IBLR-ML,SDRBF 和本文算法的错误率指标对比情况。从对比数据可知,本文算法的错误率指标要优于 JPMLC,IBLR-ML,SDRBF 3 种对比算法。3 种对比算法的错误率指标非常接近,互有优势。

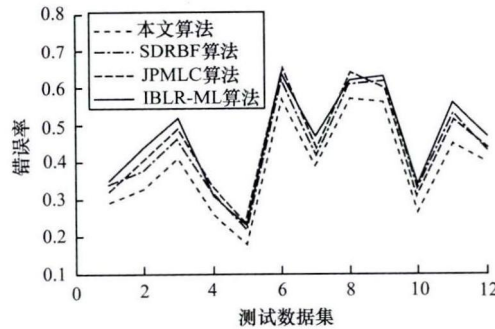


图 4 错误率比较

图 5 所示为 JPMLC,IBLR-ML,SDRBF 和本文算法的覆盖率指标对比情况。该指标可表征排序阈值需降低值,因该值越小表征算法的排序性能越佳,从图 5 所示曲线可知,本文算法的覆盖效果要优于 JPMLC,IBLR-ML,SDRBF 3 种对比算法。

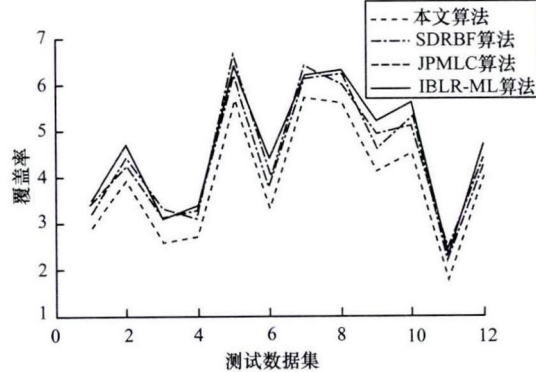


图 5 覆盖率比较

图 6 所示为 JPMLC,IBLR-ML,SDRBF 和本文算法的排序损耗对比情况。该指标数值越小代表算法的性能越佳,从图 6 所示对比数据可知,本文算法在性能上要优于选取的对比算法。

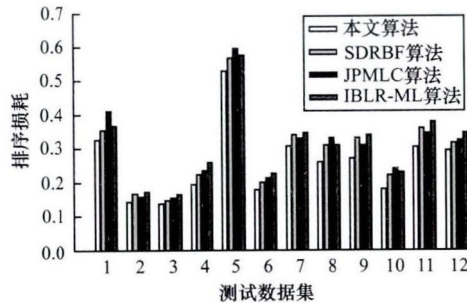


图 6 排序损耗比较

图 7 所示为 JPMLC,IBLR-ML,SDRBF 和本文算法的排序精度对比情况。从图 7 所示对比数据可知,本文算法在预测精度指标上要优于选取的对比算法。

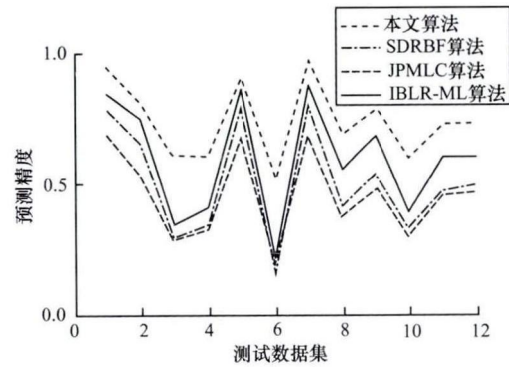


图 7 预测精度比较

从表 2 所示对比运行时间数据可知,本文算法在运行时间指标上要优于 JPMLC,IBLR-ML,SDRBF 3 种对比算法,这表明所提算法具有更加优异的执行效率。

表 2 运行时间 s

算法	计算时间
SDRBF 算法	42.32
IBLR-ML 算法	37.54
JPMLC 算法	12.39
本文算法	13.47

5 结束语

本文提出一种基于多准则决策和相似度评价的 Web 服务推荐的 SOA 系统,解决 Web 服务决策过程效果不理想的问题。基于多准则决策分析过程特点,构建基于层次 SOA 结构的服务评估系统,提出时变权重多准则相似度评价改进算法,通过 Yahoo 域名的 Web 服务推荐数据集上的实验,验证了所提算法在推荐精度和质量上的优势。

参考文献

[1] Chen Yen-cheng, Yu Tung-han, Tsui Pei-ling, et al. Erratum to: A Fuzzy AHP Approach to Construct International Hotel SPA Atmosphere Evaluation Model[J]. Quality & Quantity, 2014, 48 (4): 2371-2378.

[2] 翁迟迟,齐法制,陈刚. 基于层次分析法与云模型的主机安全风险评估[J]. 计算机工程,2016,42(2): 1-6.

[3] 矫永康,李小民,毛琼. 改进模糊层次分析法在虚拟维修训练评估中的应用[J]. 计算机工程,2014, 40(10):314-317.

(下转第 199 页)



KNN 算法有较大提高。然而,由于每个短信文本的长度非常有限、垃圾短信特殊语法和句法格式的干扰,在特征提取算法上如何进行更加有效的特征词提取和短文本特征扩展,以及提取出特殊语法句法格式包含的真正含义,将是下一步研究的方向。

### 参考文献

- [1] Zhang Liumei, Ma Jianfeng, Wang Yichuan. Content Based Spam Text Classification: An Empirical Comparison Between English and Chinese [C]//Proceedings of the 5th International Conference on Intelligent Networking and Collaborative Systems. Washington D. C., USA: IEEE Press, 2013: 69-76.
- [2] Sharma N, Verma A. Survey on Text Classification (Spam) Using Machine Learning [J]. International Journal of Computer Science & Information Technology, 2014, 5(4): 5098-5102.
- [3] Shahi T B, Yadav A. Mobile SMS Spam Filtering for Nepali Text Using Naïve Bayesian and Support Vector Machine [J]. International Journal of Intelligence Science, 2013, 4(1): 24-28.
- [4] Jiang Liangxiao, Li Chaoqun, Wang Shasha, et al. Deep Feature Weighting for Naive Bayes and Its Application to Text Classification [J]. Engineering Applications of Artificial Intelligence, 2016, 52(C): 26-39.
- [5] La Lei, Guo Qiao, Yang Dequan, et al. Multiclass Boosting with Adaptive Group-based kNN and Its Application in Text Categorization [J]. Mathematical Problems in Engineering, 2012, 16(3): 473-486.
- [6] Chang C C, Lin C J. LIBSVM: A Library for Support Vector Machines [J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 389-396.
- [7] Harrag F, Al-Qawasmah E, Hamdi-Cherif A. Performance of Decision Trees on Arabic Text Categorization [J]. Journal of Digital Information Management, 2009, 7(6): 377-382.
- [8] Ghiassi M, Olschmke M, Moon B, et al. Automated Text Classification Using a Dynamic Artificial Neural Network Model [J]. Expert Systems with Applications, 2012, 39(12): 10967-10976.
- [9] 张永军,刘金岭. 基于特征词的垃圾短信分类器模型 [J]. 计算机应用, 2013, 33(5): 1334-1337.
- [10] 蔡永泉,晋月培,葛安生,等. 基于关联分类的中文短信分类 [J]. 北京工业大学学报, 2015, 41(7): 1020-1027.
- [11] 杨柳,于 剑,景丽萍. 一种自适应的大间隔近邻分类算法 [J]. 计算机研究与发展, 2013, 50(11): 2269-2277.
- [12] Shi Kansheng, Li Lemin, Liu Haitao, et al. An Improved KNN Text Classification Algorithm Based on Density [C]//Proceedings of IEEE International Conference on Cloud Computing and Intelligence Systems. Washington D. C., USA: IEEE Press, 2011: 113-117.
- [13] Shang Changxing, Li Min, Feng Shengzhong, et al. Feature Selection via Maximizing Global Information Gain for Text Classification [J]. Knowledge-based Systems, 2013, 54(4): 298-309.
- [14] Santos I, Laorden C, Sanz B, et al. Enhanced Topic-based Vector Space Model for Semantics-aware Spam Filtering [J]. Expert Systems with Applications, 2012, 39(39): 437-444.
- [15] Hu Jianjun, Yao Yuxian. Research on the Application of an Improved TFIDF Algorithm in Text Classification [J]. Journal of Convergence Information Technology, 2013, 8(7): 639-646.

编辑 刘 冰

(上接第 192 页)

- [4] 刘延华,陈国龙,吴瑞芬. 基于云模型和 AHP 的网络信息系统可生存性评估 [J]. 通信学报, 2014, 35(8): 107-115.
- [5] Zhu Keyu, Yang Shanlin, Chen Wei. An Innovative Orders-of-Magnitude AHP Approach of Leading Industries Selection: Case Study of Jiangbei Industrial Clusters in China [J]. Journal of Systems Science and Systems Engineering, 2014, 23(1): 1-19.
- [6] 赵会群,孙 晶. 一种 SOA 软件系统可信性评价方法研究 [J]. 计算机学报, 2010, 33(11): 2202-2209.
- [7] 孙万驹,何安勇,朱海燕. SOA 技术在区域医疗机构信息共享中的应用研究 [J]. 计算机应用与软件, 2013, 30(7): 255-258.
- [8] 黄 强,王 薇,倪少权. 基于和的铁水联运信息平台构架设计 [J]. 计算机应用与软件, 2013, 30(6): 124-126.
- [9] Mazzarolo C, Martins V, Toffanello A, et al. A Method for SOA Maturity Assessment and Improvement [J]. IEEE Latin America Transactions, 2015, 13(1): 204-213.
- [10] Chen Wenyi, Ker Ming-dou. Characterization of SOA in Time Domain and the Improvement Techniques for Using in High-voltage Integrated Circuits [J]. IEEE Transactions on Device and Materials Reliability, 2012, 12(2): 382-390.
- [11] 龚艳冰. 方案偏好已知的三角模糊数型多属性决策方法 [J]. 控制与决策, 2012, 27(2): 281-284.
- [12] Liman N, Boutaba R. Assessing Software Service Quality and Trust Worthiness at Selection Time [J]. IEEE Transactions on Software Engineering, 2010, 36(4): 559-574.
- [13] 陈 军,卢涵宇,姚丹丹. 基于处理时间的网络地图 Web 服务调度算法 [J]. 计算机测量与控制, 2013, 21(7): 1987-1989.
- [14] Nie Guihua, She Qiping, Chen Donglin. Evaluation Index System of Cloud Service and the Purchase Decision Making Process Based on AHP [C]//Proceedings of International Conference on Informatics, Cybernetics, and Computer Engineering. Berlin, Germany: Springer, 2012: 345-352.
- [15] 王坚强,李康健. 基于直觉正态模糊集结算子的多准则决策方法 [J]. 系统工程理论与实践, 2013(6): 1501-1508.

编辑 顾逸斐