基于本体带 QoS 约束的数据挖掘服务选择*

熊安萍，黄美珂，蒋 溢
（重庆邮电大学 计算机科学与技术学院，重庆 400065）

摘 要：为提高数据挖掘服务选择的匹配度，提出了一种基于数据挖掘本体的带 QoS 约束的数据挖掘服务选择方法。方法引入了数据挖掘服务本体，给出了数据挖掘服务描述模型，定义了一种基于数据挖掘本体和 QoS 约束的数据挖掘服务匹配层次分类方法和服务匹配度计算方法，有效解决了数据挖掘服务选择中服务匹配的问题。实验表明，该方法可行且具有较高的查全率和查准率。

关键词：数据挖掘服务；服务选择；服务质量；数据挖掘本体；服务匹配度
中图分类号：TP391 **文献标志码**：A **文章编号**：1001-3695(2012)04-1395-04
doi:10.3969/j.issn.1001-3695.2012.04.054

Data mining service selection based on ontology and QoS constraint

XIONG An-ping, HUANG Mei-ke, JIANG Yi
(*College of Computer Science & Technology, Chongqing University of Posts & Telecommunications, Chongqing* 400065, *China*)

Abstract: To improve the matching degree of data mining service selection, this paper proposed an ontology-based data mining service selection approach with QoS constraint. Firstly, the paper introduced the ontology of data mining service and data mining service description model. Secondly, the paper defined a classification of data mining service matching degree based on data mining ontology with QoS constraint and the arithmetic of service matching degree, which solved the problem of service matching in data mining service selection. Experimental results show that the approach is feasible and obtains high recall and precision rates.

Key words: data mining service; service selection; quality of service (QoS); data mining ontology; service matching degree

0 引言

面向服务的体系结构(service oriented architecture, SOA)是一种分布式异构环境下构建数据挖掘系统的有效解决方案，它具有灵活、可扩展等特点。数据挖掘服务(data mining service, DMS)作为一种构建面向服务数据挖掘系统的关键技术，它与 Web 服务[1~3](Web service)在服务描述模型、服务选择方法和服务评价因子等方面存在较大的差异。而数据挖掘服务选择作为实现数据挖掘服务共享、复用的前提，是 SOA 架构下数据挖掘系统的一个重要组成部分。因此，研究数据挖掘服务选择方法对构建面向服务的数据挖掘系统具有重要的意义。

文献[4]提出了一套数据挖掘服务质量评价本体，给出了基于数据挖掘服务质量评价的动态数据挖掘服务选择方法，但该方法未考虑语义匹配的问题。文献[5]引入领域知识定义了数据挖掘服务质量本体和方法本体，并提出了根据领域知识和用户需求进行数据挖掘服务发现的算法，但该算法未考虑 QoS 对服务选择的影响。文献[6]通过构造数据挖掘服务 QoS 用户约束元，提出了一种可用于评估相似服务的方法，但该方法主要存在两方面的问题：a)只能根据 QoS 相似度比较待选服务在某项 QoS 指标上的性能优劣，无法对待选服务的 QoS 相似度进行整体评价；b)由于传统语义知识词典缺少数据挖掘领域词汇，影响了待选数据挖掘的服务功能性语义相似度，从而导致该方法的查全率不高。

针对以上不足并结合现有研究成果，本文首先根据数据挖掘领域基本概念构建了数据挖掘本体，并建立了数据挖掘服务描述模型；其次，根据数据挖掘本体概念关系和 QoS 约束将匹配结果分为等价匹配、泛化匹配和匹配失败三种，研究了基于数据挖掘服务本体概念语义相关度和 QoS 感知的数据挖掘服务匹配度的计算方法，以提高匹配结果的查准率和查全率；最后，构造对比实验验证了方法的有效性。

1 数据挖掘本体及数据挖掘服务描述模型

1.1 数据挖掘本体

数据挖掘本体是指数据挖掘领域词汇的基本术语和关系，以及结合这些术语和关系来定义词汇表外延的规则。在数据挖掘服务选择过程中，客观存在认知异构问题，即不同的数据挖掘服务描述表达同一数据挖掘服务。构建数据挖掘本体以建立数据挖掘概念之间的映射关系，可有效解决认知异构问题。本文将数据挖掘本体分为顶层本体[7]和应用本体两种。顶层本体定义了数据挖掘领域基本概念及其横向逻辑和拓扑关系；应用本体是顶层本体抽象概念的实例化，是数据挖掘领域内不同基本术语和关系的纵向层次划分。

收稿日期：2011-09-19；**修回日期**：2011-10-29 **基金项目**：国家自然科学基金资助项目(61003247)

作者简介：熊安萍(1970-)，女，四川泸县人，副教授，硕士，主要研究方向为面向服务计算、分布式计算与并行处理(xiongap@cqupt.edu.cn)；黄美珂(1986-)，女，硕士研究生，主要研究方向为数据挖掘；蒋溢(1969-)，男，副教授，硕士，主要研究方向为计算机软件体系结构.

数据挖掘主要有六种基本概念:数据挖掘主体、数据挖掘对象、任务相关数据集、数据挖掘预处理过程、数据挖掘算法和数据挖掘结果。这六种基本概念互相作用构成了整个数据挖掘过程的行为集合,六者之间的作用关系如图 1 所示。这六种基本概念关系框架构成了数据挖掘领域的顶层本体。

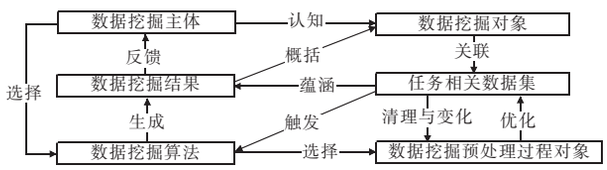

图 1 数据挖掘服务顶层本体

根据顶层本体的六种基本概念的行为特征,可以将数据挖掘领域六种基本概念划分为三类基本成分:数据挖掘实体、数据挖掘实体执行的活动和数据挖掘活动作用对象。数据挖掘实体是指数据挖掘活动参与的主体以及主体需要获取的信息,如数据挖掘结果等;数据挖掘实体执行的活动是指实体为获得其所需信息而采取的相关操作,如预处理过程等;数据挖掘作用对象的本质即数据挖掘任务相关数据集,它是数据挖掘实体获得信息的来源,是数据挖掘实体执行的活动的作用域。以此构建数据挖掘系统应用本体,以树的形式表示数据挖掘系统本体概念类的抽象层次结构,如图 2 所示。

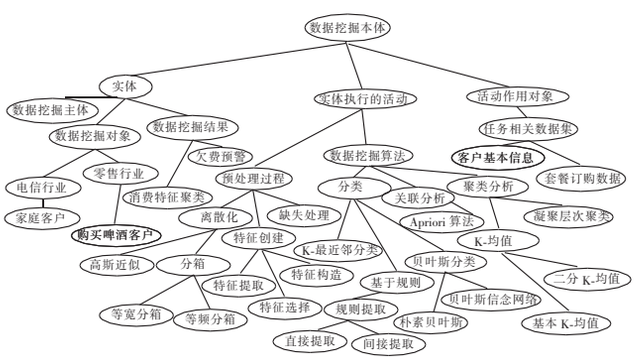

图 2 数据挖掘系统本体概念类抽象层次结构(部分)

1.2 数据挖掘服务描述模型

在面向服务数据挖掘系统中,数据挖掘服务请求应充分包含数据挖掘服务请求者对服务的需求信息,以便数据挖掘服务注册中心提供满意的服务。为此,本文根据数据挖掘服务的特点,定义了数据挖掘服务描述模型。

定义 1 数据挖掘服务描述模型

$$\mathrm{DMS} = \{B, F, S, D, \mathrm{QoS}\}$$

其中:

B 表示服务的基本属性,包括服务 ID、服务提供商 ID、服务描述信息、服务注释等;

F 表示服务的功能信息,包括数据挖掘的目标,数据挖掘服务应使用的数据挖掘算法等;

S 表示数据挖掘专有属性[8],如关联分析需要的支持度和置信度,K-均值聚类算法中的初始质心参数 K 等;

D 表示数据挖掘服务的数据集约束信息,包括需要待挖掘数据的范围等,如用户请求挖掘“年龄在 40 岁以上手机用户消费特征”,D 表示 age > 40;

QoS 表示服务的 QoS 约束条件。

文献[4]提出了一些数据挖掘 QoS 评价指标,但是考虑到不同数据挖掘服务请求包含 QoS 指标的多样性,服务注册中心无法预见请求中包含的所有 QoS,故服务注册中心将 QoS 分为基本 QoS 和附属 QoS,规定所有的数据挖掘服务请求 QoS 和待选数据挖掘服务注册信息必须包含基本 QoS 集中所有的元素。若数据挖掘服务请求的 QoS 基准值为 $Q_r = \{q_{r1}, q_{r2}, \cdots, q_{rn}\}$,数据挖掘服务注册中心规定的服务基本 QoS 集为 $Q_b = \{q_{b1}, q_{b2}, \cdots, q_{bm}\}$,$Q_b$ 的元素代表了数据挖掘服务注册中心对基本 QoS 集中每项 QoS 指标的默认值(即最差值)。在每个 Q_r 提交服务注册中心后都需与 Q_b 进行并集运算以规范请求,规范后的数据挖掘服务请求基本 QoS 为:$Q_{sb} = Q_r \cup Q_b = \{q_{si} \mid q_{si} \in Q_r \vee q_{si} \in Q_b\}$。若 $q_{ri} \in Q_b$ 且 $q_{ri} \in Q_r$,即请求 QoS 集中包含该项基本 QoS,则 $q_{si} = q_{ri}$,否则,添加基本 QoS 至规范后的 QoS 集中,即 $q_{si} = q_{bi}$。

2 基于本体带 QoS 约束的数据挖掘服务选择

2.1 数据挖掘服务匹配层次

由于数据挖掘服务提供者与数据挖掘服务请求者之间的认知异构和 QoS 差异,数据挖掘服务注册中心通常无法精确选择出恰当的服务提供给服务请求者。为了选择恰当的数据挖掘服务,不仅仅需要在语义上理解请求服务与待选服务,还需要考虑服务 QoS 约束允许的状态,从而指导数据挖掘服务的动态选择。本文根据数据挖掘本体概念包含关系的推理及 QoS 约束,按照匹配精确程度差异性,可定义如下匹配层次。

定义 2 服务等价匹配。对于请求的服务概念 Q 和发布的服务概念 A,若在数据挖掘本体中,Q 与 A 是同一概念或 Q 与 A 是等价类(即语义等价),且发布的服务 QoS 均满足请求的服务 QoS 约束,则称请求服务与发布服务等价。

定义 3 服务泛化匹配。对于请求的服务概念 Q 和发布的服务概念 A,若在数据挖掘本体中,Q 与 A 为等价类或存在父子关系(即语义相关),且发布的服务至少满足请求对基本 QoS 属性的基准值,则称提供的服务与请求服务匹配泛化。

定义 4 服务匹配失败。对于请求的服务概念 Q 和发布的服务概念 A,若在数据挖掘本体中,Q 与 A 之间不存在直接的关联关系(即语义无关),或发布的服务不满足请求对基本 QoS 属性的基准值,则称提供的服务与请求服务匹配失败。

以上三组定义定性了数据挖掘服务请求与待选数据挖掘服务之间的匹配层次。但是当同一匹配层次有多项待选服务满足数据挖掘服务请求时,需要综合量化待选服务的本体概念语义相关度和 QoS,以便选出效果最好的服务。因此,本文提出数据挖掘服务匹配度(matching of service, MoS)的概念,用于定量描述数据挖掘服务请求与待选数据挖掘服务间的匹配程度。

2.2 数据挖掘服务匹配度

根据定义 2 ~ 4,待选数据挖掘服务与服务请求间的匹配层次主要取决于本体概念语义的相关性和 QoS 的满意度。本体概念语义的相关性可通过数据挖掘系统本体概念类抽象层次结构树中概念节点间的距离量化。QoS 的满意度可用各个 QoS 指标值线性加权后的数值度量。通过语义概念间的距离和 QoS 满意度即可综合量化数据挖掘服务匹配度。

2.2.1　QoS 约束的综合效果评价方法

假定待选数据挖掘服务集中有 m 项待选服务,记为 $S=\{s_1,s_2,\cdots,s_m\}$,$m\in N$,N 为自然数,服务 $s_i(i=1,2,\cdots,n)$ 对应的 QoS 向量为 $\boldsymbol{Q}_i$,$s_i\in S$,$Q_i=\{q_{i1},q_{i2},\cdots,q_{in}\}$,$n\in N$, QoS 向量 $\boldsymbol{Q}_i$ 的每个元素对应数据挖掘服务的一项 QoS,向量 $\boldsymbol{\eta}_j(j=1,2,\cdots,n)$ 对应一项 QoS,其每个元素表示了该项 QoS 下该元素对应服务的量值,由此可建立服务集 S 的 QoS 评价矩阵:

$$\boldsymbol{E}=(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n)=(\boldsymbol{Q}_1,\boldsymbol{Q}_2,\cdots,\boldsymbol{Q}_m)^{\mathrm{T}}=\begin{pmatrix}q_{11} & \cdots & q_{1n}\\ \vdots & & \vdots\\ q_{m1} & \cdots & q_{mn}\end{pmatrix}$$

QoS 指标评价矩阵记录了所有待选数据挖掘服务的 QoS 信息,存储于数据挖掘服务注册中心,其每行表示了一项待选服务的 QoS 指标值,其每列表示了一项 QoS 对应每项数据挖掘服务的值,其值由服务注册中心动态收集或服务提供者注册。为评价每项服务 QoS 的满意度,定义如下:

定义 5　若 QoS 评价矩阵中列向量 $\boldsymbol{\eta}_i$ 对应的 QoS 对数据挖掘服务 QoS 综合评价值的权重为 w_i,p_{ij} 为该项 QoS 下服务 s_i 对应 QoS 值的比重,$\sum_{j=1}^{n}p_{ij}w_j$ 称为服务 s_i 的 QoS 综合效果评价值,w_i 为效果强度因子。

根据定义 5,待选数据挖掘服务的 QoS 综合效果评价值主要取决于待选服务的各项 QoS 值的比重和效果强度因子。下文将分别给出它们的计算方法。

1)计算待选服务的各项 QoS 值的比重

考虑不同的 QoS 在纲量上存在较大的差异,如响应时间的单位可能是秒,噪声比可能是一个无度量单位百分比。为了确定 p_{ij} 的值,本文对 QoS 指标值进行了归一化处理,以消除纲量对 QoS 综合效果评价值的影响。

根据不同 QoS 的效果,将所有 QoS 属性分为两类:a)积极 QoS,即属性值越大越好,如超空间中数据密度等;b)消极 QoS,即属性值越小越好,如服务的成本等。本文根据最大化—最小化原理[9] 及文献[10]的归一化公式,针对积极 QoS 和消极 QoS 分别得到如下变换规则:

$$\begin{cases}\dfrac{q_{ij}-q_j^{(s)}}{q_j^{(h)}-q_j^{(s)}}\times\alpha+(1-\alpha) & q_j \text{ 是积极 QoS 且 } q_j^{(h)}\neq q_j^{(s)}\\ 1 & q_j^{(h)}=q_j^{(s)}\\ \dfrac{q_j^{(h)}-q_{ij}}{q_j^{(h)}-q_j^{(s)}}\times\alpha+(1-\alpha) & q_j \text{ 是消极 QoS 且 } q_j^{(h)}\neq q_j^{(s)}\end{cases}\tag{1}$$

其中:$q_j^{(s)}$ 表示 q_{ij} 所属 QoS 指标的最小值,$q_j^{(h)}$ 表示 q_{ij} 所属 QoS 指标的最大值;α 为可调节因子,它可以避免变化后的数据出现零值,取值范围为(0,1)。

对评价矩阵 $\boldsymbol{E}$ 的每个元素按式(1)进行变换,可得变换后的评价矩阵 $\boldsymbol{E}'=(q'_{ij})_{m\times n}$。QoS 评价矩阵第 j 列对应 QoS 下服务 s_i 的 QoS 值的比重为

$$p_{ij}=\frac{q'_{ij}}{\sum_{i=1}^{m}q'_{ij}}$$

2)计算效果强度因子

为保证在数据挖掘服务选择时,服务注册中心能够自动化进行多目标决策,本文借鉴熵值法[10] 的思想确定效果强度因子 w_j。熵值法是一种以指标的信息熵评价决策指标权重的经典方法。对于任意的可选服务,考察某项 QoS 属性值 $\boldsymbol{\eta}_j=(p_{1j},p_{2j},\cdots,p_{mj})^{\mathrm{T}}$,若所有的服务该属性值差异性越大,则在满足基本 QoS 约束条件下,该属性包含的信息量越大,对综合评价值的影响越大;反之,若属性的差异性越小,则该属性包含的信息量越少,对综合效果评价值的影响越小。

QoS 评价矩阵第 j 列对应 QoS 的熵值为

$$e_j=-\sum_{k=1}^{m}p_{kj}\ln p_{kj}\tag{2}$$

且

$$\sum_{k=1}^{m}p_{kj}=1\tag{3}$$

其中:p_{kj} 表示 QoS 评价矩阵第 j 列对应 QoS 下服务 s_k 的 QoS 值的比重。

为保证后续运算的合理性,e_j 应满足 $e_j\in[0,1]$,所以须求得 e_j 的最大值,以便对 e_j 进行归一化处理。由式(2)(3)构造如下拉格朗日函数:

$$F_j=-\sum_{i=1}^{m}p_{ij}\log p_{ij}+\lambda\left(\sum_{i=1}^{m}p_{ij}-1\right)\tag{4}$$

根据式(4)可由如下公式:

$$\begin{cases}\dfrac{\partial F_j}{\partial p_{ij}}=0,i=1,2,\cdots,m\\ \dfrac{\partial F_j}{\partial\lambda}=0\end{cases}\tag{5}$$

求得,当 $p_{ij}=\dfrac{1}{m}$ 时,其中 $i=1,2,\cdots,m$,e_j 取得极大值为 $(e_j)_{\max}=\ln m$。

当熵值越大时,指标的差异性越小,对决策的价值也就越小,故差异性系数定义为

$$g_j=1-\frac{e_j}{(e_j)_{\max}}=1+\frac{\sum_{i=1}^{n}p_{kj}\ln p_{kj}}{\ln m}$$

则 QoS 评价矩阵第 j 列所对应 QoS 的效果强度因子为

$$w_j=\frac{g_j}{\sum_{j=1}^{n}g_j}$$

2.2.2　数据挖掘服务匹配度计算方法

数据挖掘系统本体概念类抽象层次结构树中每个节点表示了数据挖掘领域的一个概念。树的任意两个节点之间的距离越小,节点所表示的概念之间的语义越相似。为了量化概念间语义的相关度,定义如下:

定义 6　对于数据挖掘系统本体概念类抽象层次结构树中任意两个概念 C_1 和 C_2 的语义相关度 relevance(C_1,C_2),定义如下:a)若 C_1 和 C_2 为树中相同节点,即概念 C_1 和 C_2 语义等价,则 relevance(C_1,C_2)=0;b)若概念 C_1 和 C_2 语义相关,则 relevance(C_1,C_2)=|level(C_1)−level(C_2)|,其中 level(C_i)表示概念节点 C_i 的层次;c)若概念 C_1 和 C_2 语义无关,则 relevance(C_1,C_2)=∞。

为了量化请求服务与待选服务之间的匹配程度,本文定义了一种基于本体的带 QoS 约束的,定义如下:

定义 7　待选数据挖掘服务 s_c 与请求服务 s_r 之间的数据挖掘服务匹配度为

$$\mathrm{MoS}(s_c,s_r)=\frac{\lambda}{1+\mathrm{relevance}(s_c,s_r)}+\frac{\lambda}{2}\sum_{i=1}^{n}w_ip_{ij}$$

其中:relevance(s_c,s_r)表示请求数据挖掘服务与待选服务的语义相关度,$\sum_{i=1}^{n}w_ip_{ij}$ 表示 QoS 综合效果评价值,λ 为调节因子。

若发布的服务不满足请求对基本 QoS 属性的基准值或请求的服务概念和发布的服务概念语义无关则 $\lambda = 0$;若发布的服务 QoS 不能完全满足请求的服务 QoS 约束,但能够满足服务基本 QoS 约束,则 $\lambda = 1/2$;否则 $\lambda = 1$。

由定义 5 可知,$\sum_{i=1}^{n} w_i p_{ij}$ 的范围为(0,1]。故请求数据挖掘服务与待选服务间的匹配程度是一个介于 0 ~ 1.5 的实数。该值越大,服务间的匹配度就越高,反之,服务间的匹配度就越低。数据挖掘服务匹配度与服务匹配层次之间的关系如表 1 所示。通过数据挖掘服务间的匹配度就可以定量描述不同待选数据挖掘服务对服务请求的满足程度。

表 1 服务匹配度与服务匹配层次关系列表

服务匹配层次	服务匹配度范围
等价匹配	(1,1.5]
泛化匹配	(0,1]
匹配失败	0

3 实验结果及分析

为了验证本文提出的数据挖掘服务匹配方法的可行性及其在实践应用中的有效性,本文分别验证了请求数据挖掘服务和待选服务间的服务匹配度和匹配结果的查全率、查准率。仿真实验环境为 Inter Pentium Dual 1.6 GHz CPU,2 GB RAM,Windows XP 和 JDK 1.6 环境,测试程序采用 Java 语言开发。

a)实验 1 验证待选服务匹配度

实验选取了某电信企业的六项服务作为待选服务,数据挖掘服务注册中心获得的待选服务信息如表 2 所示,数据挖掘服务注册中心规定的基本 QoS 集为(成本,响应时间,可视化,主观兴趣度,信誉度)。用户向数据挖掘服务注册中心提交数据挖掘服务请求——“我的 E 家客户消费分析”,请求 QoS 基准值为(1000,50,1,0.6,0.9,0.8)。经过数据服务注册中心计算,待选数据挖掘服务的服务匹配度如表 3 所示。实验结果表明,不同的待选数据挖掘服务的服务匹配度与实际情况一致,服务匹配度和服务匹配层次符合表 1 的关系。本文的数据挖掘服务选择方法具有可行性。

表 2 待选数据挖掘服务信息

待选服务	成本	响应时间	可视化	主观兴趣度	信誉度	数据完整性
A 市我的 E 家客户消费分析	360	33	1	0.98	0.99	0.99
B 省我的 E 家消费分析	429	19	1	0.75	1.0	0.87
C 市 E6 手机客户增值业务消费分析	984	49	1	0.96	0.95	0.94
D 区我的 E 家客户消费分析	565	10	1	0.71	0.98	未知
E 市我的 E 家客户流失分析	746	29	0	0.58	0.91	0.86
A 区 E6 宽带用户消费分析	940	76	0	0.98	0.81	0.95

表 3 待选数据挖掘服务匹配度和服务匹配层次

待选服务	服务匹配度	匹配层次
A 市我的 E 家客户消费分析	1.021 5	精确匹配
B 省我的 E 家消费分析	1.020 7	精确匹配
C 市 E6 手机客户增值业务消费分析	0.536 1	泛化匹配
D 区我的 E 家客户消费分析	0.289 3	泛化匹配
E 市政企客户消费分析	0	匹配失败
A 区 E6 宽带用户消费分析	0	匹配失败

b)实验 2 验证服务匹配结果的查全率和查准率

实验选取了某电信企业的 249 项数据挖掘服务作为实验样本,分别以文献[4 ~ 6]和本文提出的方法等四种方式构建了数据挖掘服务注册中心,数据挖掘服务请求者向数据挖掘服务注册中心发出 20 次服务请求,通过考察数据挖掘服务匹配结果的平均查全率和查准率衡量本文方法的效果。实验结果如表 4 所示。

表 4 不同方法的查全率和查准率

方法名称	查全率/%	查准率/%
基于质量的数据挖掘服务选择	30.41	96.17
基于领域本体的数据挖掘服务选择方法	97.56	52.35
基于物元和 QoS 约束的数据挖掘服务选择	68.48	90.87
本文方法	83.53	91.96

基于质量的数据挖掘服务选择方法的查全率不高,特别是在泛化匹配的层次上,该方法查全率为 0。基于领域本体的数据挖掘服务选择方法在查全率效果较好,但是由于其未考虑 QoS 的影响,匹配的结果会包含一些 QoS 不能满足请求的服务,影响了查准率。基于物元和 QoS 约束的数据挖掘服务选择比前两种方法,在查全率和查准率上虽有了较明显的提升,但查全率上依然不高;本文的数据挖掘服务选择方法在查全率和查准率上均取得了较好的效果。可见,本文提出的数据挖掘服务选择方法提高了服务选择的效果。

4 结束语

本文提出了一种基于本体概念关系、带 QoS 约束的数据挖掘服务选择方法。本文方法与前人方法相比具有如下优点:a)匹配结果的查全率和查准率同时取得了较好的效果;b)提出了基于语义概念相关度和 QoS 综合效果评价值的数据挖掘服务匹配度量化方法,适合计算机自动化处理。下一步工作重点将关注数据挖掘服务组合及数据挖掘服务过程建模的问题。

参考文献:

[1] 李金忠,夏洁武,唐卫东,等. 基于 QoS 的 Web 服务选择算法综述[J]. 计算机应用研究,2010,27(10):3622-3627.

[2] PAOLUCCI M, KAWAMURA T, PAYNE T R, *et al*. Semantic matching of Web services capabilities[C]//Proc of the 1st International Semantic Web Conference. London: Springer-Verlag, 2002: 333-347.

[3] 吴健,吴朝晖,李莹,等. 基于本体论和词汇相似度的 Web 服务发现[J]. 计算机学报,2005,28(4):595-601.

[4] 李玉华,陈云开,卢正鼎. 基于质量的数据挖掘服务选择[J]. 计算机科学,2007,34(8):159-164.

[5] 陈英,顾国昌. 基于领域本体的数据挖掘服务发现算法[J]. 计算机工程与应用,2008,44(18):150-152.

[6] 陈增科,肖基毅,陈灵娜,等. 基于物元和 QoS 约束的数据挖掘服务选择[J]. 计算机工程,2009,35(24):90-92.

[7] GANGEMI A, GUARINO N, MASOLO C. A understanding top-level onto-logical distinctions[C]//Proc of Workshop on Ontologies and Information Sharing. Palo Alto: AAAI Press, 2001: 26-33.

[8] SOMAN K P, DIWAKAR S, AJAY V. Insight into data mining theory and practice[M]. [S.l.]: Prentice-Hall of India Pvt. Ltd, 2006: 121-129.

[9] HAN Jia-wei, KAMBER M. Data mining concepts and techniques[M]. 2nd ed. San Francisco: Morgan Kaufmann, 2005: 46-47.

[10] 乔家君. 改进的熵值法在河南省可持续发展能力评估中的应用[J]. 资源科学,2004,26(1):113-119.