

基于匹配度的网格服务组合及优化选择方法

李琳琳¹ 孙继银¹ 刘光斌² 梅生¹ 张艳红²

(第二炮兵工程学院指挥自动化系 西安 710025)¹ (第二炮兵工程学院自动控制系 西安 710025)²

摘要 为提高网格服务动态组合过程中服务间的匹配概率,提出了基于语义推理、相似度计算和图搜索的弱关联程度下的服务组合方法(Low Match Degree Service Composition Method, LMDSCM)。LMDSCM通过定义基于领域本体的概念集合匹配度作为服务匹配度计算的依据,来完成服务组合图的构建,并将服务组合的优化选择问题转换为最短路径选择问题。通过实验将LMDSCM与同类方法在服务组合满意度和算法效率上进行了对比,结果证明其在弱关联程度条件下的服务组合满意度高于同类算法,易于生成最优服务组合方案。

关键词 网格服务,服务发现,服务组合,领域本体,语义匹配

中图法分类号 TP311 **文献标识码** A

Match Degree-based Method for Grid Service Composition and Optimize Choose

LI Lin-lin¹ SUN Ji-yin¹ LIU Guang-bin² MEI Sheng¹ ZHANG Yan-hong²

(Command Automation Department, The Second Artillery Engineer College, Xi'an 710025, China)¹

(Automation Control Department, The Second Artillery Engineering College, Xi'an 710025, China)²

Abstract The match probability of automatic grid service composition is improved by Low Match Degree Service Composition Method(LMDSCM) based on semantic illation, similitude degree and graph search. With the concept aggregate degree based on domain ontology defined, it is used to the calculating of service composition and building of service composition graph. The optimize of service composition is transformed to shortest path problem. Finally, comparisons of efficiency and satisfaction shows that LMDSCM has higher satisfaction degree than other method and is easy to produce best service composition project.

Keywords Grid service, Service detect, Service composition, Domain ontology, Semantic match

Web服务组合作为Web服务的一项重要增值功能,为服务重用和自动化集成提供了应用的基础。它是根据用户需求,将多个小粒度的网格服务,按照它们间的语义关联关系组合成一个大粒度网格服务,达到服务增值的目的。

根据Web服务组合的动态性程度,分为静态Web服务组合和动态Web服务组合。静态组合主要基于 workflow思想,如eFlow和CAFISE,该类方法易于实现,但适应性较差。动态组合是在系统运行过程中选择和调用所需的服务组件,并将之组合成一个复合服务的过程。动态组合方法包括以语义为核心和基于图搜索的自动组合方法。文献[1]采用基于概念相似度和图搜索的方法,该方法通过引入服务聚类而具有较高的算法效率,但由于其对服务间关联程度定义过于简单,因此匹配精度较差;文献[2]采用基于语义推理和相似度的自动组合方法,该方法通过对本体库的预处理提高了匹配效率,但由于其未考虑概念集合匹配度,因此其效率较低;文献[3]采用基于图的语义服务组合法,该方法易于确定服务间的关联关系,但由于其未考虑服务匹配度问题,因此不具备服务组合的优化选择能力。

本文将基于语义推理、语义相似度计算和基于图搜索等服务组合技术结合起来,充分考虑概念集合同间的多种关联关

系,提出了“弱关联程度下的服务组合方法”(Low Match Degree Service Composition Method, LMDSCM)。

1 弱关联程度下的服务组合法

LMDSCM的核心思想是首先运用语义推理方法将服务间的关联程度划分为不同等级,其次运用语义相似度计算方法对服务匹配度进行量化处理,将其作为服务组合图及邻接矩阵生成的主要依据,最后运用图搜索方法对服务组合图进行遍历,获得最优服务组合方案。

LMDSCM依据网格服务概念集合同间的不同逻辑关系,将概念集合同间的匹配度划分为5个匹配等级,根据匹配等级和语义相似度计算结果对服务匹配度赋值,使其具有匹配概率高,易于生成服务组合图的特点。

2 基本定义

2.1 服务及服务请求

定义1(网格服务, Grid Service, GS) GS是基于网格体系架构,分布式的模块化构件,它执行特定的任务,遵守具体的技术规范,通过统一的技术规范达到多个兼容构件间的互操作能力。网格服务用如下二元组形式进行描述:GS(I_s, Q_s),其中

本文受国家863计划“十一五”重点课题(2010AA7010614)资助。

李琳琳(1974-),女,博士生,副教授,主要研究方向为指挥自动化系统工程与计算机应用, E-mail: Lilinlin74@yeah.net.

$I_s = \{I_{s1}, I_{s2}, \dots, I_{sm}\}$, $O_s = \{O_{s1}, O_{s2}, \dots, O_{sm}\}$ 分别为 GS 的输入和输出集合, 对应领域本体中的概念集合。

定义 2(用户请求, User Request, UR) 用户发出的某一使用网格服务的请求, 用如下二元组进行描述: $UR(I_r, O_r)$, 其中 $I_r = \{I_{r1}, I_{r2}, \dots, I_{rm}\}$, $O_r = \{O_{r1}, O_{r2}, \dots, O_{rm}\}$ 。

2.2 概念匹配及概念集合匹配度

服务参数 I_s 和 O_s 对应着本体域中的概念集合, 服务匹配的计算过程实际上是两个服务输入、输出概念集合间的匹配过程, 涉及的概念如下。

定义 3(基于语义推理的概念匹配度^[2]) 基于语义推理的概念匹配是建立在领域本体概念间的层次关系之上的, 匹配结果采用离散值来度量, 分为 *exact*、*plugin*、*subsumes* 和 *fail* 4 个级别。对领域本体中的两个概念 c_i 和 c_j 的概念匹配结果遵循如下规则:

$$CIDegree(c_i, c_j) = \begin{cases} exact, & c_i = c_j \text{ 或 } c_i \text{ 是 } c_j \text{ 的直接子类} \\ plugin, & c_i \subset c_j \text{ 即 } c_i \text{ 是 } c_j \text{ 的间接子类} \\ subsumes, & c_i \supset c_j \text{ 即 } c_j \text{ 是 } c_i \text{ 的直接或间接子类} \\ fail, & c_i, c_j \text{ 间没有关系} \end{cases}$$

定义 4(概念语义相似度^[2]) 作为基于语义推理匹配方法的改进, 基于语义相似度的匹配通过计算两个本体概念之间的相似度来得到更精确的匹配结果。本文采用文献[2]提供的概念语义相似度算法, 计算公式如下:

$$Sim(c_i, c_j) = \frac{2 \times depthof(parent(c_i, c_j))}{depthof(c_i) + depthof(c_j)} \quad (1)$$

式中, $parent(c_i, c_j)$ 表示概念 c_i, c_j 在概念树上的最低共同祖先, $depthof()$ 函数返回一个概念在概念树上的深度。

定义 5(概念匹配度) 概念匹配度是建立在语义推理概念匹配分级度量基础上, 运用相似度计算方法对概念匹配度进行精确计算而得到的, 计算公式如下:

$$CMDegree(c_i, c_j) = \begin{cases} 1, & CIDegree(c_i, c_j) = exact \\ Sim(c_i, c_j), & CIDegree(c_i, c_j) = plugin \\ k \times Sim(c_i, c_j), & CIDegree(c_i, c_j) = subsumes \\ 0, & CIDegree(c_i, c_j) = fail \end{cases}$$

式中, k 为调节系数, $0 < k < 1$, 表示 *subsumes* 的概念匹配度低于 *plugin*。

定义 6(概念集合匹配^[2]) 网格服务的输入、输出可能是一个概念集合而不是由单个概念构成的, 因此网格服务的匹配应考虑概念集合的匹配。对于概念集合 $C_i = \{c_i | c_i \text{ 为概念, 且 } 1 \leq i \leq m\}$ 和概念集合 $C_j = \{c_j | c_j \text{ 为概念, 且 } 1 \leq j \leq n\}$, 如果 c_i 和 c_j 满足如下关系: 概念 c_j 是 C_j 中与 C_i 中概念 c_i 匹配度最大的一个概念, 且 c_i 与 c_j 的匹配度大于一个预设的阈值, 则称 C_i 中的概念 c_i 与 C_j 中的概念 c_j 相匹配。

定义 7(基于语义推理的网格服务概念集合匹配度) 结合网格服务的输入输出概念集合间的匹配要求, 设 $UR(I_r, O_r)$ 为用户请求, $GS(I_s, O_s)$ 为网格服务, 采用 *exactM*、*supperM*、*subM*、*partM*、*failM* 5 个匹配级别^[2] 来确定 UR 与 GS 间的匹配度, UR 和 GS 间的匹配规则如下:

规则 1 对于用户请求的输入概念集合 I_r 和服务输入概念集合 I_s :

$$CAIDegree(I_r, I_s) = \begin{cases} exactM & I_r \equiv I_s \\ supperM, & I_r \supset I_s \\ failM & I_r \not\supseteq I_s \end{cases}$$

规则 2 对于用户请求的输出概念集合 O_r 和服务输出概念集合 O_s :

$$CAODegree(O_r, O_s) = \begin{cases} exactM, & O_r \equiv O_s \\ supperM, & O_r \subset O_s \\ subM, & O_r \supset O_s \\ partM, & O_r \otimes O_s \\ failM, & O_r \cap O_s = \phi \end{cases}$$

规则 3 对于中间服务的匹配, PO_i 作为前驱服务的输出集合, SI_i 作为后序服务的输入集合:

$$CAOIDegree(PO_i, SI_i) = \begin{cases} exactM, & PO_i \equiv SI_i \\ supperM, & PO_i \subset SI_i \\ subM, & PO_i \supset SI_i \\ partM, & PO_i \otimes SI_i \\ failM, & PO_i \cap SI_i = \phi \end{cases}$$

定义 8(基于语义相似度的概念集合匹配度) 根据概念集合匹配定义, 采用语义相似度计算方法, 对基于语义推理后确定的输入、输出概念集合的匹配度进行量化计算。假设 C_i, C_j 为网格服务或用户请求对应的输入和输出概念集合, $CADegree$ 为经过规则 1、2、3 的语义推理后获得的匹配等级, 则 C_i, C_j 概念集合的语义相似度计算公式如下:

$$CAMDegree(C_i, C_j, CADegree) = \begin{cases} 1, & CADegree = exactM \\ a, & CADegree = supperM \\ \frac{1}{m} \sum_{c_j \in C_j, c_i \in C_i} \max(CMDegree(c_i, c_j)), & CADegree = subM \\ b \times \frac{1}{m} \sum_{c_j \in C_j, c_i \in C_i} \max(CMDegree(c_i, c_j)), & CADegree = partM \\ 0, & CADegree = failM \end{cases}$$

式中, m 为概念集合 C_j 的概念总数。 a, b 为概念集合匹配度的调节系数, $0 < a < 1, 0 < b < 1$, 调节后 *supperM* 的匹配度介于 *exactM* 和 *subM* 之间, *partM* 的匹配度低于 *subM*。

2.3 服务组合及服务组合图

定义 9(服务组合^[3]) 根据用户请求, 将多个小粒度的网格服务, 按照它们间的语义关联关系组合成一个大粒度网格服务, 以满足用户请求目标, 达到服务增值的目的。

定义 10(服务之间的关系及关联程度^[1]) 对于网格服务集合中的 2 个服务 $GS_i(I_{s_i}, O_{s_i})$ 和 $GS_j(I_{s_j}, O_{s_j})$, 如果 $CAOIDegree(O_{s_i}, I_{s_j}) \neq failM$, 那么称它们之间存在偏序关系, 称 GS_i 为 GS_j 的前序服务, GS_j 为 GS_i 的后序服务, 记为 $GS_i > GS_j$, GS_i 与 GS_j 间的关联程度用 $CAMDegree(GS_i, GS_j, CADegree)$ 来表示。

定义 11(用户请求与服务之间的关系及关联程度) 对于某一用户服务请求 $UR(I_r, O_r)$ 和一个网格服务 $GS_i(I_{s_i}, O_{s_i})$, 如果有 $I_r \supseteq I_{s_i}$, 那么称 GS_i 为 UR 的后序服务, 记为 $GS_i > UR$, UR 与 GS_i 间的关联程度用 $CAMDegree(UR, GS_i, CADegree)$ 来表示。

定义 12(服务组合图) 以用户请求提出的 I_r 和 O_r 概念集合作为起点和终点, 通过发现服务集合中服务参数间的语义关联构造服务组合图, 服务组合图定义为一个二元组

SCRG(V, E)。V 是图中顶点集合, 表示满足用户请求所需的一个网格服务; E 为图中边的集合, E 中的每一个元素 $e(e \in E)$ 都是一条有向边, $E = \{ \langle x, y \rangle | (S(x, y), S(x, y), Degree) \wedge (x, y \in V) \}$, $S(x, y)$ 表示服务 x 到服务 y 的一条有向边, $S(x, y)$. Degree 是 x 到 y 有向边的权值, 用 x, y 服务间的匹配度表示。具体计算时结合网格服务输入输出概念集合要求, 运用基于语义相似度的概念集合匹配度计算结果对其进行赋值。

定义 13(服务组图邻接矩阵) 网格服务集中所有服务构成的网格服务组图对应的 $n \times n$ 阶邻接矩阵 S, 可由如下计算公式确定:

$$s_{ij} = \begin{cases} 1, & i=j \\ CAMDegree(GS_i, GS_j, CADegree), & i \neq j \end{cases}$$

式中, $1 \leq i \leq n, 1 \leq j \leq n, n$ 为服务集中的服务个数。

如果把用户请求 UR 的 I_r 和 O_r 分别作为服务组合图的起点和终点, 作为邻接矩阵 S 的第 0 节点和第 $n+1$ 节点, 生成的 $(n+2) \times (n+2)$ 阶邻接矩阵 S' , 需要增加的元素可由如下计算公式确定^[1]:

$$\begin{cases} s'_{00} = 1 \\ s'_{0(n+1)} = 0 \\ s'_{0i} = CAMDegree(UR, GS_i, CADegree) \\ s'_{i0} = 0 \\ s'_{i(n+1)} = CAMDegree(UR, GS_i, CADegree) \\ s'_{(n+1)i} = 0 \\ s'_{(n+1)0} = 0 \\ s'_{(n+1)(n+1)} = 0 \end{cases}$$

定义 14(服务组合满意度) 文献[1]描述一个网格服务集合对一个用户服务请求的满足程度。对一个用户请求 UR 和一个网格服务集合 GSA(GS_1, GS_2, \dots, GS_n), 其满意度为:

$$SAT = CAMDegree(UR, GS_i, CADegree) \cdot \prod_{i=1}^n CAMDegree(GS_i, GS_{i+1}, CADegree) \cdot CAMDegree(UR, GS_n, CADegree)$$

3 服务组合图的生成及最优服务组合选择

3.1 服务组合图的生成算法

输入: 用户请求 $UR(I_r, O_r)$, 服务集合 $GS = \{GS_1, GS_2, \dots, GS_n\}$, GS_i 用二元组 $GS_i(I_{S_i}, O_{S_i})$ 描述。

输出: 以 I_r 为第 0 个元素, O_r 为第 $n+1$ 个元素的 $(n+2) \times (n+2)$ 阶服务组图邻接矩阵 $SCRGM[i, j]$ 。

具体算法如下:

```
SCRGMCreate(UR, GS)
{for(i = 0; i <= n+1; i++) /* 以  $I_r$  为起点,  $O_r$  为终点, 遍历服务集合 S 中的所有服务 */
{for(j = 0; j <= n+1; j++)
{if(i == j) /* 对角线元素均为 1, 表示服务于自身匹配度最高 */
{SCRGM[i, j] = 1; continue;}
else if((i == 0 && j == n+1) || (i != 0 && j == 0) || (i == n+1 && j != n+1))
{SCRGM[i, j] = 0; continue;} /*  $I_r$  与  $O_r$  间无任何关联 */
else if(i == 0 && j != n+1) /* 用户服务请求与服务集合输入概念集合间的匹配度 */
{CADegree = CAIDegree( $I_r, I_{S_j}$ );
```

```
SCRGM[i, j] = CAMDegree( $UR, GS_j, CADegree$ ); continue;}
else if(i != n+1 && j == n+1) /* 用户服务请求与服务集合输出概念集合间的匹配度 */
{CADegree = CAODegree( $O_r, O_{S_i}$ );
SCRGM[i, j] = CAMDegree( $UR, GS_i, CADegree$ ); continue;}
else /* 服务集合内任意两个中间服务间的匹配度 */
{CADegree = CAOIDegree( $O_{S_i}, I_{S_j}$ );
SCRGM[i, j] = CAMDegree( $GS_i, GS_j, CADegree$ ); continue;} } }
```

3.2 基于最短路径的服务组合选择算法

运用 3.1 节生成的服务组图 $SCRGM(V, E)$, 按照最短路径法, 以 I_r 为起点, O_r 为终点对服务组图进行遍历, 得到节点间边权值乘积最大的一条连通路, 即为对用户请求服务组合满意度最高的服务组合路径, 作为满足该用户服务请求的最优服务组合方案。

4 实例分析

4.1 实例数据

本实例主要用于验证 3.1 节提出的服务组图生成算法的有效性, 同时通过实例数据证明本文提出的服务组图生成算法和最优服务组合选择方法较文献[1]提出的 TPM 算法具有更高的服务组合满意度和匹配概率。本实例选取如表 1 所列的 7 个服务实例。

表 1 服务实例数据

服务名	输入参数	输出参数
GS ₁	P ₁	P ₂ , P ₃ , P ₄
GS ₂	P ₂ , P ₄	P ₅ , P ₇
GS ₃	P ₂ , P ₃ , P ₄	P ₅
GS ₄	P ₆ , P ₇	P ₈ , P ₉
GS ₅	P ₅ , P ₇ , P ₁₀	P ₈ , P ₁₁
GS ₆	P ₈ , P ₁₁	P ₁₂ , P ₁₃
GS ₇	P ₈	P ₁₃

4.2 服务组图生成

以用户请求 $UR(I_r, O_r)$, $I_r = \{P_1, P_2\}$, $O_r = \{P_{13}\}$ 的输入、输出作为服务组合图的起点和终点, 按 3.1 节提供的服务组图算法生成服务组图, 如图 1 所示。按文献[1]提供的服务组图算法 TPM 生成的服务组图如图 2 所示。

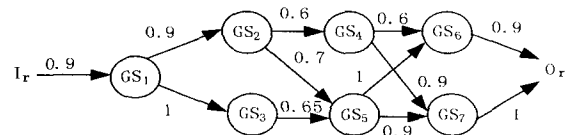


图 1 基于服务匹配度的服务组图

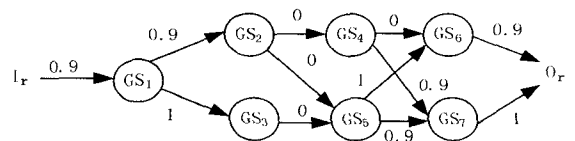


图 2 基于 TPM 的服务组图

4.3 最优服务组合的选择及服务组合满意度

依据图 1、图 2 生成的服务组图, 利用最短路径法生成最优服务组合路径和计算服务组合满意度, 图 1 生成的最优服务组合路径为 $I_r \rightarrow GS_1 \rightarrow GS_3 \rightarrow GS_5 \rightarrow GS_6 \rightarrow O_r$, 服务组合满意度为 0.5265。图 2 由于未考虑中间服务的输出参数与输入参数间的子匹配和部分匹配情况, 因此服务匹配过程到

(下转第 436 页)

- [11] 娄申,干晓蓉. 基于 BP 神经网络的水质评价[J]. 云南民族大学:自然科学版,2007,16(2):165-167
- [12] 黄国宏,熊志化,邵惠鹤. 一种新的基于构造型神经网络分类算法[J]. 计算机学报,2005,28(9):1519-1523
- [13] 邓万宇,郑庆华,陈琳,等. 神经网络极速学习方法研究[J]. 计算机学报,2010,33(2):279-287
- [14] 王建军,徐宗本. 多元多项式函数的三层前向神经网络逼近方法[J]. 计算机学报,2009,32(12):2482-2488
- [15] Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: Theory and applications [J]. Neurocomputing, 2006, 70 (1-3): 489-501
- [16] Huang G-B. Learning capability and storage capacity of two hidden-layer feedforward networks [J]. IEEE Transactions on Neural Networks, 2003, 14(2): 224-229
- [17] 张德丰. MATLAB 神经网络应用设计[M]. 北京:机械工业出版社,2009
- [18] Chen T, Chen H, Liu R-W. Approximation Capability in $C(\bar{R}^n)$ by Multilayer Feedforward Networks and Related Problems [J]. IEEE Transactions on Neural Networks, 1995, 6(1): 25-30
- [19] 褚辉,赖惠成. 一种改进的 BP 神经网络算法及应用[J]. 计算机仿真,2007,24(4):75-77

(上接第 400 页)

GS_2 、 GS_3 即停止,无法完成服务组合,服务组合满意度为 0,未满足用户需求。

由此可以看出,本文提出的基于语义推理和相似度相结合的服务组合算法,与文献[1]提出的 TPM 比较,在服务间语义关联度较弱的情况下,具有更强的服务匹配能力,提高了用户请求与服务间的输出参数以及中间服务间输出参数与输入参数的匹配概率。

4.4 算法效率实验

算法效率实验中,选取如表 1 所列的 7 个服务实例以及随机产生一定数量的干扰服务作为测试用例,当干扰服务样本数量分别为 100、400、600、800、1000、1200、1400、1600、1800、2000 时,与 TPM 算法的算法效率的仿真结果如图 3 所示。

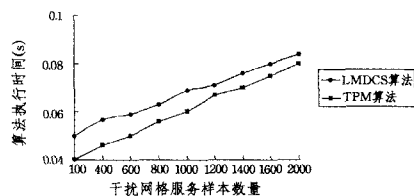


图 3 服务组合算法效率比较

由图 3 可以看出,由于算法 TPM 采用了服务聚类方法对网络服务组合进行了预处理,且在服务匹配过程中直接采用语义相似度的计算方法来获得服务间的关联程度,因此其在执行效率上优于本文算法。可以考虑在本文算法中加入文献[1]提出的服务聚类方法,以降低服务匹配次数从而提高算法效率。

实例分析表明:本文提出的 LMDSCM 算法,由于将用户请求与服务之间、服务与服务之间输入输出参数的匹配程度划分为: *exactM*、*superM*、*subM*、*partM*、*failM* 5 个匹配等级,并运用基于语义相似度的计算方法对匹配度进行量化计算,将其作为服务组合图构建时边的权值进行赋值,为利用最短路径法确定最优服务组合提供了计算依据。由于采用由粗粒度离散量匹配到细粒度服务匹配度量化计算的两阶段匹配算法,因此服务间的匹配概率明显优于仅考虑语义关联度计

算的 TPM 算法,具有较高的服务组合满意度。

结束语 本文基于语义推理和语义相似度计算方法,提出了弱关联程度下的服务匹配算法 LMDSCM,其以服务匹配度作为服务组合图构建的依据,采用最短路径法完成最优服务组合路径的选择和服务组合满意度的计算。LMDSCM 算法与基于 workflow 模型的服务组合方法相比,具有较高的自动化程度,与一般基于语义的服务组合算法相比具有更高的匹配概率,与文献[1]提出的 TPM 算法相比具有更高的服务组合满意度。但由于算法采取粗粒度和细粒度的分阶段计算方法,因此算法效率低于 TPM。可采用 TPM 提出的服务聚类方法对服务集合进行预处理,达到提高算法效率的目的。

参考文献

- [1] 倪晚成. 基于概念关联程度的网络服务组合方法[J]. 清华大学学报:自然科学版,2007,47(10):581-585
- [2] 李玉鑫. 基于本体的动态 Web 服务组合的研究与应用[D]. 大连:大连理工大学,2008:21-22
- [3] Pfeffer H, Linner D, Steglich S. Modeling and Controlling Dynamic Service Compositions[C]// Computing in the Global Information Technology, The Third International Multi-Conference on (ICCGI '08). 2008:210-216
- [4] Papazoglou M P, Traverso P, Dustdar S, et al. Service Oriented Computing: State of the Art and Research Challenges[J]. Computer, 2007, 40(11): 38-45
- [5] Narayanan S, McIlraith S A. Simulation, Verification and Automated Composition of Web Services[C]// Proceedings of the eleventh international conference on World Wide. 2006:77-88
- [6] McDermott D. Estimated-regression planning for interactions with services[C]// Proc. of the 6th International Conference on AI Planning and Scheduling. Toulouse France, AI Press, 2005
- [7] Shina D-H, Lee K-H, Sudab T. Automated generation of composite Web services based on functional semantics[J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2009(7): 332-343
- [8] 宋巍,马晓星,吕建. Web 服务组合动态演化的实例可迁移[J]. 计算机学报,2009,32(9):1816-1831