

一个基于 P2P 和本体分类的 Web 服务发现模型^{*}

赵泽宇 薛向阳 张成洪

(复旦大学计算机科学与工程系 智能信息处理开放实验室 上海200433)

摘要 服务发现是 Web 服务应用中极其重要的环节。随着 Web 服务数目的增长和对实时性需求的增加,集中式的服务发现机制难以满足实际需求,分布式服务发现机制成为 Web 服务应用的一个新研究方向。本文基于 P2P 的底部架构,实现了一个根据本体对服务注册节点进行语义分类的 Web 服务发现模型,并对其发现效率和节点的负载进行了评估。

关键词 Web 服务, Web 服务发现, P2P, 本体

An Ontology-Based Classification and P2P-Based Web Services Discovery Model

ZHAO Ze-Yu XUE Xiang-Yang ZHANG Cheng-Hong

(Dept. of Computer Science and Engineering, Laboratory of Intelligent Information Processing, Fudan University Shanghai200433)

Abstract Service discovery plays a very important role in Web service applications. Current Web service discovery mechanism requires replication of all Web service publications in all registries. Large growth in number of Web services as well as the growth in the number of registries makes this replication impractical. Distributed Web services discovery mechanism is put forward. We come out with a service discovery model which is based on semantic classification of registries and P2P infrastructure. The design and an experimental evaluation of the model are presented.

Keywords Web service, Web service discovery, Peer-to-Peer, Ontology

1 引言

随着分布式计算技术成为网络应用的主流,各种采用分布式计算技术的信息服务正越来越普及。然而,传统的分布式计算技术在网络应用中不断反映出来的种种缺陷促使研究人员不断探索新的分布式技术。Web 服务就是在这种环境下诞生的一种集成网上资源和实现业务重组的理想手段。

典型的 Web 服务采用了面向服务的体系结构,这种体系结构包含了三个重要角色:服务请求者(Requestor)、服务提供者(Provider)和注册中心(Registry)。服务提供者创建服务描述,并将其发布到一个或多个注册中心,服务请求者向注册中心检索各种服务描述,并根据服务描述获取要绑定的服务。

传统的 Web 服务发现机制使用集中的服务注册中心来存储和发现服务描述,这些注册中心的不同数据库周期性同步以保持数据的有效性(例如 UDDI^[1])。然而,随着 Web 服务数目的增长和对实时性需求的增加,集中式的服务发现机制难以满足实际需求,于是分布式的服务发现结构应运而生^[3,8]。

本文基于 P2P 的底部架构,实现了一个服务注册节点分类的 Web 服务发现模型。在模型中,每个注册节点根据需求和兴趣提供某个方面的服务描述信息服务,节点使用改进的移位弦环结构来组织,并按服务领域进行分类。每一个服务注册节点隶属于某个服务领域,并提供该领域的服务信息的存储和检索服务,所有注册节点共同提供注册中心的功能。

本文第2节介绍目前本领域内的相关研究工作;第3节介

绍模型实现的基本结构;第4节对模型的性能进行实验评估;最后进行小结。

2 相关工作

当前对 P2P 底层结构的研究基本分为两大类。第一类以 FreeNet、Gnutella^[10]为代表采用非结构化连接的底层结构。事实证明这类结构对可扩展性缺乏有效的支持,当节点的数目增大时,系统的性能急剧下降;第二类以 Chord^[5]、Pastry^[6]、Tapstry^[7]和 CAN^[11]为代表,采用结构化底层结构互连。这类结构对可扩展性和稳定性提供了可靠的保证。然而,这些资源发现模型没有对资源进行有效的分类,使得每个查询都会被分派到大量无关的节点中,降低了效率。

最近出现的一些 Web 服务的分布式发现方法不直接采用基本的 P2P 结构,而是引入服务节点分类的思想,将提供单一类型的信息服务的注册节点分类,以提高服务发现效率。METEOR^[4]和 EDUTELLA^[12,2]提出了在 P2P 的服务注册节点层上引入本体进行分类的方法,但是不得不引入“超节点”(分别是 GWP 和 Super-Peer)来支持对分类节点的操作和管理;文[1]提出了一个按服务描述特征空间对节点分类的方法,在节点的分类层上进行 P2P 的连接,同样未能摆脱对“超节点”的依赖。

本模型采用 P2P 的底层结构连接服务注册节点,引入服务领域本体对节点进行分类,并针对此结构实现了两个消息转发的算法。本模型弥补了上述两种方法的缺点:节点分类使得服务查询相对集中在特定节点中,提高发现的效率,同时分

^{*} 本文工作受到国家自然科学基金(60003017, 60373030, 69935010)、八六三计划(2001AA114120, 2002AA103065)以及上海市科技发展基金(03DZ15019, 03DZ14015)共同资助。赵泽宇 硕士研究生,主要研究方向为 Web 服务和语义 Web;薛向阳 教授,博士生导师,主要研究方向为多媒体信息处理和检索技术;张成洪 教授,博士生导师,主要研究方向为语义 Web 和 P2P 技术。

类中的每个注册节点之间是对等、自适应、自管理的,没有严格的互相依赖关系。实验结果表明,本模型具有良好的可扩展性。

3 模型的设计

模型的设计思路如下:模型中维护着大量的动态注册节点,每个注册节点隶属于一个服务领域,提供某一类服务描述的发布服务。每个服务领域由若干注册节点组成,这些节点共同提供该服务领域的 Web 服务信息的发布服务。节点的连接

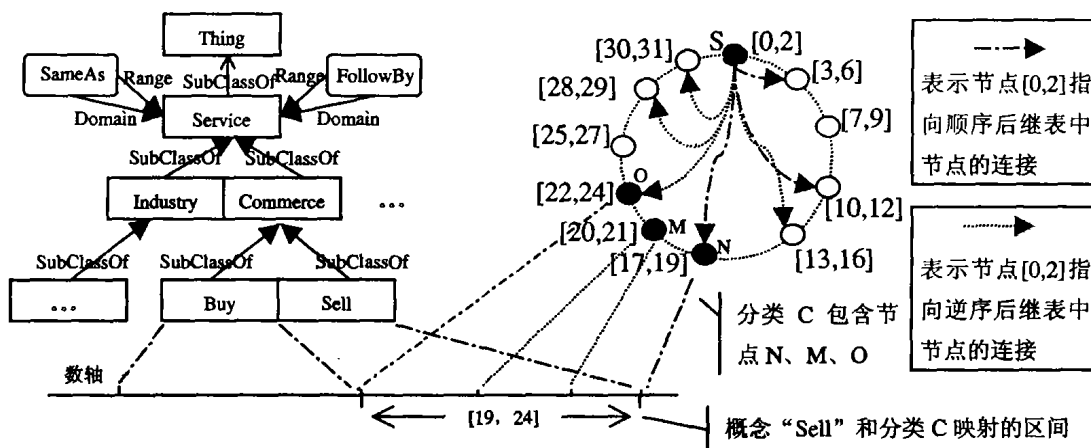


图1a 服务本体

图1b $GNS = [0, 2^5 - 1]$ 的移位弦环结构和节点 S 的后继节点示意

所有节点共同提供服务信息的存储和检索服务,而每个 Web 服务发现的请求都可以限制在相应的服务领域中,有效地提高了查询效率。在模型中,每个查询从代理节点传递到目标注册节点包括:a)从代理节点转发到目标服务领域的中介节点;b)从中介节点向所有目标服务领域内的节点转发。我们改进了移位弦环结构的两个算法来完成这两个步骤。在本节的后续部分,我们详细介绍模型的设计。

3.1 本体划分和服务领域描述

我们引入服务领域本体来为模型提供统一的语义规范。本体^[12]是用于描述或表达某一领域知识的一组概念或术语,它给出了领域实体概念及相互关系、领域活动以及该领域所具有的特性和规律的一种形式化描述。如图 1a 所示,服务本体的每个概念分别表示一类实例的集合,每个子类分别表示某个具体的服务概念,是其父类所涵盖实例集合的子集。每个服务概念被映射到一个节点集合,子类概念映射到其父类概念的节点集合的一个子集。这样一个节点集合称为一个服务领域。进行 Web 服务发现时可以把查询定位在相应的服务领域中,以提高查询的效率。服务领域本体还定义了概念之间的各种语义关系,例如,SameAs 关系表示两个服务概念相同,它们表示同一个服务领域的不同称谓;FollowBy 关系则表示服务概念在工作流表示中的顺序关系。

3.2 基本拓扑结构

模型底部采用基于改进的移位弦环空间分割的 P2P 路由结构,它可以保证节点的加入、离开、节点间消息传递的数量都很少。这个结构在 Chord^[5]等一些分布式散列表系统(DHT)中得到应用。在 Chord 中,每个节点使用一个键值区间进行标识,同时维护一个顺时针方向的后继表(Finger),并提供了给定一个键值,从任意节点路由到键值所映射节点的方法。在此基础上,我们将环路中的相邻节点根据其与服务概念的映射关系分类,使每个节点在原后继表的基础上增加一个逆时针方向的后继表,实现了从任意节点将消息转发到目

采用基于 Chord^[5]改进的移位弦环结构实现,隶属于相同服务领域的结点邻接相连。我们采用服务领域本体来描述服务领域,其中每个概念描述一个服务领域,并映射到一个由连续整数组成的区间。服务节点分类采用其对应的服务领域概念映射的整数区间进行标识,其中的每一节点分别用区间的不相交子集标识。一个节点分类中所有节点的标识区间的并集等于该分类的标识区间。如图 1a、b 所示,概念“Sell”映射到整数区间^[19,24],分类 C 由节点 N、M 和 O 组成,节点的标识区间分别是分类区间的不相交子集。

标节点分类的区间定位算法和在一个分类中进行消息广播的多点传送算法。

在一个以 2^m 为模,由 $n(n \leq 2^m)$ 个节点依次连接而成的移位弦环结构中,节点的全局命名空间 GNS(Global Name Space)是整数区间 $[0, 2^m - 1]$;每一个子节点都将 GNS 的子空间作为自己的私有命名空间 PNS(Private Name Space),表示为区间 $[DB, UB]$,区间中的所有整数被称为该节点的 IDs。该环路的子节点满足下面的关系:A) $\bigcup_{0 \leq i < n} PNS_i = GNS$ (n 是环路中节点总数);B) $PNS_i \cap PNS_j = \emptyset$, 其中 $i \neq j$ 且 $1 \leq i, j \leq n$;C) $(UB_i + 1) \% 2^m = DB_j$, 其中 $(i + 1) \% 2^m = j$ 且 $1 \leq i, j \leq n$ 。图 1b 示例了一种节点的 PNS 的分布情况。在本文的后续章节中,我们用函数 $DB(N)$ 和 $UB(N)$ 分别表示节点 N 的上确界和下确界,用 $PNS(N)$ 表示节点 N 的私有空间。

每个节点 L 维护其 PNS 和后继节点集之间连接关系的两个后继表,沿顺指针方向的称为顺序表,沿逆时针方向的为逆序表。表项包含 {后继节点 IP 地址,后继节点 PNS}。若在顺序表中存在指向后继节点 N 的表项,则意味着满足 $(UB(N) - UB(L)) \% 2^m = 2^i$, 其中 $0 \leq i \leq m - 1, GNS = 2^m$ 。类似的,逆序表中指向后继节点 N 的表项意味着 $(DB(L) - DB(N)) \% 2^m = 2^i$ 。为了节约存储空间和减轻节点的负担,采用长度动态变化的后继表,每个不同的后继节点仅在后继表中存储一次。图 1b 给出了节点 S 连接顺序表和逆序表中的后继节点的示意图。

Chord 的键值定位算法^[5]从任意节点将键值定位到目标节点。将目标节点扩大成目标分类,结合逆序表,得到从任意节点 L 提交一个请求 $Q([A, B])$,将该请求快速转发到对应于 $Target = [A, B]$ 的分类中的节点的方法,即区间定位算法:定义 L 和 $Target$ 的顺时针距离 $Dist$ 为 L 的上确界和 $Target$ 的下确界的距离,即 $Dist = (A + GNS - UB(L)) \% GNS$;逆时针距离为 L 的下确界和 $Target$ 的上确界的距离即 $iDist = (DB(L) + GNS - B) \% GNS$;定义任意节点 L 的转发操作

L. Location(Q[A,B]),其转发规则如下:(1)节点 L 接收到查询 Q([A,B]),若 PNS(L)与[A,B]相交,则 L 自身作为查找结果返回;(2)若 PNS(L)与[A,B]不相交,则查找后继节点,若存在后继节点 N 使得 PNS(N)与[A,B]相交,则将 N 作为结果返回;(3)若 L 无与[A,B]相交的后继节点,则选择最靠近[A,B]的后继节点转发请求:比较 Dist 和 iDist,若 $Dist \leq iDist$,则在顺序表中选择从顺时针方向最接近[A,B]的节点 N,调用 N.Location(Q[A,B])方法;若 $Dist > iDist$,则在逆序表中选择从逆时针方向最接近[A,B]的节点 N',调用 N'.Location(Q[A,B])方法。

从上述转发规则知查询请求每经一次转发后都会更接近目标区间,所以每个查询请求从源节点转发到目标区间必然沿恒定方向,即顺时针或逆时针方向。现证明在一个均匀分布的环路中,区间定位算法将请求转发到目标区间的转发次数不超过 $O(\log H/2)$,H 为节点分类的数目。这里,均匀分布是指环路的长度为 2^m 的 GNS 近似平均分割为 H 个类的区间,每个区间的长度为 $\lfloor 2^m/H \rfloor$ 或 $\lfloor 2^m/H \rfloor + 1$ 。

我们仅证明消息沿顺时针方向到达目标区间的情况,逆时针方向情况的证明类同。假设消息 Q 从源节点 S 沿顺时针方向转发到目标区间 T,定义 T 和 S 的距离 Distance(T,S) 为两个区间的下确界的差值。若 PNS 与 T 不相交,S 的下一个转发节点是 P_1 ,则 $Distance(S,P_1) > Distance(P_1,T)$,否则 S 可以在 P_1 和 T 之间找到另外一个更接近 T 的后继节点。假设 P_1 的 PNS 的长度为 L_{P_1} ,则 P_1 到 T 的距离 $Distance(P_1,T) < ((1/2) * Distance(T,S))$ 。若 P_1 不能直接到达 T,则 P_1 和 T 之间必然存在 P_1 的下一个转发节点 P_2 ,如此下去,在 k 次转发后到达区间 T 中的节点,此时有 $(1/2)^k * Distance(S,T) < Length(T)$,其中 Length(T) 表示区间 T 的长度。假设环路中有 H 个分类,每个分类有 N 个节点,则共有节点 $H \times N$ 个。取 $GNS = [0, 2^m - 1]$ 环路的最长转发距离为最远的两个节点的距离, $Distance(S,T) = 2^m * (\frac{HN/2-1}{HN})$, $Length(T) = 2^m * 1/H$,则有 $(\frac{1}{2})^k * 2^m * (\frac{HN/2-1}{HN}) < 2^m * \frac{1}{H}$,结果 $k > \log(H/2-1/N)$,转发次数 k 近似 $O(\log H/2)$ 。

移位弦环结构的广播算法^[14]沿环路单向向所有节点广播消息,其缺陷在于无法从任意节点向其所属分类的其他节点转发消息。我们基于该广播算法实现了从分类中任意节点沿环路双向向分类中的节点转发消息的方法,即多点传送算法:设一个相邻节点集合 T 的区间为 $[a,b]$,从其中任意节点 P $[a_1,b_1]$ (P 为 T 中一个节点)向 T 中所有节点转发消息的规则如下:

(a1) 设 $\{N_1, N_2, \dots, N_k\}$ 是 P 的顺序表中所有后继节点的集合,若 N_i 的 PNS 与 T 相交,则 P 向 N_i 转发查询 Q $[(DB(N_i), \max(DB(N_{i+1})-1, b))]$,其中 $i \leq k-1$,若 $DB(N_k) \leq b$,则 P 向 N_k 转发查询 Q $[(DB(N_k), \max(UB(N_k), b))]$;

(a2) 类似地,设 $\{M_1, M_2, \dots, M_q\}$ 是 P 的逆序表中所有后继节点的集合,若 M_i 的 PNS 与 T 相交,则 S 向 M_i 转发查询 Q $[\min(UB(M_{i+1})+1, a), UB(M_i)]$,其中 $i \leq q-1$;若 $UB(M_q) \geq a$,则 P 向 M_q 转发查询 Q $[\min(DB(M_q), a), UB(M_q)]$;

(a3) 若节点 $R[r_1, r_2]$ 接收到查询 Q $[q_1, q_2]$ 满足 $r_1 \leq q_1 \leq q_2 \leq r_2$,则不再继续往后继节点转发查询。如图 2 所示,节点 S 将接收到的请求向顺序表中的后继节点 R1、R3 和向逆序表

中的 L1、L3 转发,整个消息在分类节点中的传播沿树状结构。

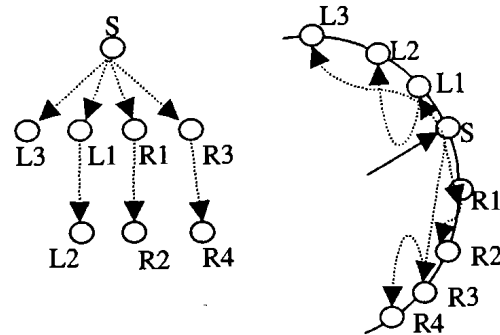


图 2 多点转发示意图

目标区间中每个节点都能收到消息,且仅有一次。证明如下:

(b1) 源节点 S 仅接收到一次消息;

(b2) 假设某个节点非源节点 G 接收到消息 Q $[G1, G2]$ 至少两次,根据(a1)、(a2),S 至少存在两条不完全相同的转发路径 $Path1 = \{S, P11, P12, \dots, G\}$ 和 $Path2 = \{S, P21, P22, \dots, G\}$ 分别可到达 G。用 $Path1[i]$ 和 $Path2[j]$ 分别表示 Path1 和 Path2 中第 i 和第 j 个节点。假设 $Path1[i] = Path2[i]$ ($0 \leq i$), $Path1[i+1] \neq Path2[i+1]$,则 $Path1[i+1]$ 和 $Path2[i+1]$ 都是 $Path[i]$ 的后继节点,且 $Path1[i+1]$ 与 $Path2[i+1]$ 的区间不相交。由(a1)、(a2)和 Path1 可知,G 的区间必然包含于 $Path1[i+1]$,同理 G 的区间包含于 $Path2[i+1]$,上述结论产生矛盾。所以,每个非源节点至多接收到消息一次;

(b3) 同(b2)的假设,若某个节点 G 不能接收到消息 Q $[G1, G2]$,则不存在一条转发路径 $Path = \{S, P1, P2, \dots, G\}$ 可从 S 到达 G。然而,由环路的构造规则知,若节点 $N1 \rightarrow N2 \rightarrow N3$ 相邻,则 $N3$ 必然是 $N2$ 的顺序表中的节点, $N1$ 必然是 $N2$ 逆序表中的节点。所以,至少存在一条转发路径 $Path1 = \{S, N1, N2, \dots, G\}$ (路径中的相邻节点为环路中的相邻节点) 可从 S 到达 G,所以每个非源节点至少接收到消息一次。

因此,每个节点都将收到消息,且仅有一次。在一个节点数为 n 的区间中,该算法消息转发次数为 $O(n)$;消息转发的树状结构最多不超过 $O(\log(n))$ 层,所以消息转发时间不超过 $O(\log(n))$ ^[14]。

3.3 节点的加入和离开

我们提供了一个通过均匀分割区间方法来标记节点的 PNS,使得 PNS 在环上的节点中分布均匀。

节点的加入过程 新节点 N 首先向系统中任意节点 P 发出加入请求 Q_{add}, Q_{add} ,中包含类型属性 Type,其值为查询所属服务领域;代理节点 P 将 Q_{add} 映射到服务本体的相应概念,获得目标区间 $[A_0, A_0]$,而后通过区间定位算法将 Q_{add} 转发到目标服务领域的某个节点 P' , P' 通过多点传送算法将 Q_{add} 向该服务领域中所有节点传送。每个节点接收到请求后,根据多点传送算法向其后继节点转发,并接收返回的结果,无法向后继节点转发消息的节点直接返回本节点的标识区间的长度 d 和 IP;每个接收到返回结果的节点比较结果的区间长度 d 和自身的长度 c,若 $c \geq d$,则返回本节点的标识区间长度和 IP,否则将后继节点的返回结果向上一级节点返回。新节点 N 最终接收到一个节点 N_p 的 IP,设 N_p 的标识区间为 $[a, b]$,则分割为等长的两段: $[a, \lfloor a + (b-a)/2 \rfloor]$ 和 $[\lfloor a + (b-a)/2 \rfloor + 1, b]$, N_p 保留后者作为自身的标识区间,N 取前者

作为自身的标识区间。节点加入方法基本保持同一分类中节点的 PNS 的均匀分布,一个节点加入的时间量级为 $O(2m)$ 。 N_i 加入后, N_i 需要向所有前继节点发送新节点的 PNS 和 IP, 前继节点重新分配后继节点。

节点的离开过程 一个节点 N_i 退出环路时, 通过多点传送算法向所有节点发送消息(取消息的目标区间为整个 GNS), 每个接收到消息的节点修改自身的后继表, 并向其后继节点转发消息。最后 N_i 的 PNS 并入其后继节点或前任节点。

3.4 服务查询

服务查询的过程如下: 当一个查询 Q 提交到任意一个注册节点 A 时, A 负责将 Q 映射到服务领域本体的概念, 取得相应的区间范围 T, 并通过区间定位算法将查询转发到 T 中某个服务注册节点 R; R 通过多点传送算法将查询转发到目标区间中每个注册节点 S; 每个接受到查询的节点 S 将本地与后继节点返回的查询结果进行综合筛选, 选取最为合适的结果向来源节点返回; 最终查询结果沿查询转发过的节点原路返回到代理节点。

4 性能评估

基于对等节点分类的服务发现模型的主要目标是高效地将服务查询分发到目标服务节点, 并使得所有节点的负载相对较为均衡。我们开发了一个原型, 对模型从服务发现效率和负载情况两方面性能进行评估。

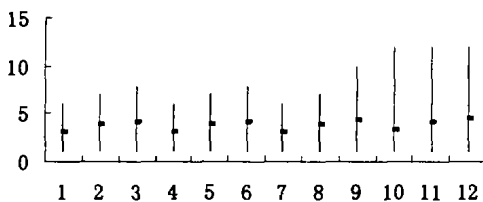


图3 分别在2100/4200/8400/16800个节点, 36/96/156个聚类情况下的平均区间定位次数

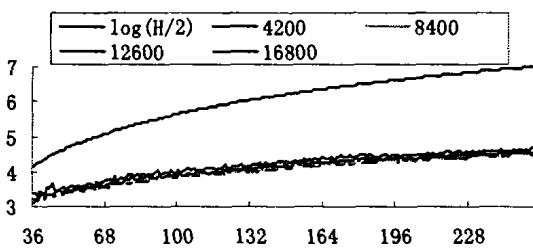


图4 从源节点到目标节点的平均定位次数示意

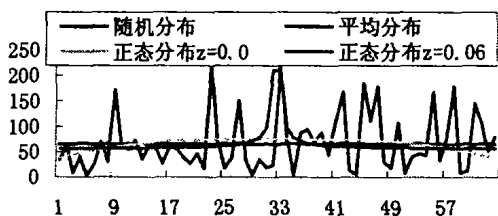


图5 节点分布

4.1 区间定位算法评估

实验1 评估模型中将每个查询转发到目标区间的效率。每个查询从源节点转发到目标节点包括两个步骤: 首先从代理节点转发到目标分类中某个节点, 其次再由该节点将查询

转发到分类中所有其他节点。从本文第三部分的论述已知, 在目标分类内的查询转发的次数与区间的节点数相同, 为向分类中所有节点广播的最小转发次数。所以, 查询从代理节点到目标分类的转发次数成为衡量模型的服务效率的一个重要方面。分别统计在不同的节点数目和分类数目下区间定位算法的最大次数和平均次数。如图3所示, 系列集 {1, 2, 3}, {4, 5, 6}, {7, 8, 9}, {10, 11, 12} 分别表示节点数为2100/4200/8400/16800的情况下, 当分类数分别为36/96/156时每60000次随机查询的区间定位次数的最大值和平均值。从图3中可看出, 不同的序列集随着节点数从2100增长到16800的情况下, 平均区间定位次数并无明显变化; 而在同一个序列集中, 相同节点总数的情况下, 不同的分类数目导致平均定位次数不同。图4描绘了节点数分别为4200、8400、12600、16800, 分类总数从36上升到256时, 每60000个随机查询的平均定位次数曲线和曲线 $y = \log(H/2)$ 的关系图。可以看出, 当节点数变化时, 各条曲线基本重合, 平均转发次数只有理论最大值的一半左右, 节点数的变化对平均转发次数影响并不大; 在相同节点数目下, 平均定位次数随着分类数目的增长, 这种增长的速度远远低于服务领域总数的增长速度, 并随着服务领域总数的增加而放缓。理论证明和实验结果表明影响模型的消息转发次数的主要因素是服务本体中概念划分粒度而非节点总数。

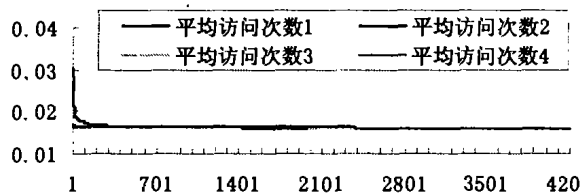


图6 节点平均访问次数统计

4.2 节点负载评估

实验2 评估在不同服务领域的不同节点数分布下, 模型中每个节点的负载情况。我们考虑作为消息区间定位通路中的转发节点所转发的查询数和作为目标区间中的查询处理节点所处理过的节点数的总和。模型中包含4200个节点和64个分类时, 每产生600000次查询, 统计每个节点的负载的平均值。图5为节点数目在64个分类中的分布曲线, 包括随机分布、平均分布和正态分布的情况, 图6显示了4200个节点中每个节点的平均负载。可以看出, 在随机产生查询的情况下, 节点的平均访问次数基本相同, 访问量曲线趋近水平直线; 只有在随机分布中, 有少部分节点的平均访问量高于其他节点一倍左右。实验结果表明在随机产生查询的情况下, 无论各个分类的节点数目均匀分布或随机分布, 各个节点的负载都趋向于均衡。

从测试和分析的结果来看, 消息定位所需的转发次数主要由概念本体映射到数值空间的粒度决定。当服务领域数目增加时, 消息定位的平均转发次数同时增长, 且这种增长的幅度相当缓慢。而且, 所有节点的负载在各种节点分布情况下都趋近均衡。所以, 本结构适用于大量注册节点的情况, 并且在规模增大时仍能保持相对较好的性能。

小结 本文实现了一个基于移位弦环结构改进的 P2P 的 Web 服务发现结构。本结构设计的主要特点是将服务节点基于服务概念本体分类, 每个查询请求可以集中在一些注册节点的分类中被处理, 避免使每个查询请求都会被分派到大量无关的节点中的情况减轻了整个网络的负担, 提高了服务

(下转第201页)

$$P_B(X_1, X_2, \dots, X_N) = \prod_{i=1}^N P(x_i | \text{parents}(X_i))$$

用这个联合概率值对所有候选路径进行排序, 概率值最大的为最优路径。

4 分析和展望

本文综述了粘连字符的切分算法以及切分路径评价方法。

在现有的切分算法中, 基于细化的算法具有较好的适应性, 而且可以处理复杂的粘连情况, 如接触面倾斜、字符间相互重叠等, 对于复杂的汉字粘连, 也能进行有效的处理^[10]。

在现有的切分评价方法中, 基于识别的方法依赖于具体的识别器, 识别器的错误率决定了评价的错误率。基于特征的方法是相对独立的一类方法, 能获得较高的正确率和效率。对复杂字符(如汉字)的切分, 如能结合评价、识别和上下文关系, 将进一步减少切分错误。

参考文献

- 1 Strathy N W, Suen C Y, Kryzak A. Segmentation of handwritten digits using contour features. In: Second Int. Conf. Document Anal. Recognition, Oct. 1993. 577~580
- 2 Dimauro G, Impedovo S, Pirlo G, Salzo A. Automatic Bankcheck processing: A New Engineered System. International Journal of Pattern Recognition and Artificial Intelligence, 1997, 11(4): 467~504
- 3 Congedo G, Dimauro G, Impedovo S, Pirlo G. Segmentation of Numeric strings. In: Proc. of Third Int. Conf. on Document Analysis and Recognition, Montreal, Aug. 1995. 1038~1041

(上接第195页)

发现的效率。分类中的每个注册节点之间是对等、自适应、自管理的, 没有严格的互相依赖关系。实验结果表明, 本模型具有良好的可扩展性。

参考文献

- 1 Schmidt C, Parashar M. A Peer-to-Peer Approach to Web Service Discovery, submitted for publication, February 2003. <http://cite-seer.nj.nec.com/schmidt03peertopeer.html>
- 2 Schlosser M, Sintek M, Decker S, Nejd W. A Scalable and Ontology-Based P2P Infrastructure for Semantic Web Services. Second Intl. Conf. on Peer-to-Peer Computing (P2P'02), Linkoping, Sweden, Sept. 2002
- 3 Hoschek W. Peer-to-Peer Grid Databases for Web Service Discovery. Grid Computing: Making the Global Infrastructure a Reality. Nov. 2002
- 4 Verma K, et al. METEOR-S WSDI: A Scalable Infrastructure of Registries for Semantic Publication and Discovery of Web Services. <http://lstdis.cs.uga.edu/lib/download/MWSDI.pdf>, 2002
- 5 Stoica, Morris R, Karger D, Kaashoek M F, Balakrishnan H. Chord: A scalable peer-to-peer lookup service for internet applications. Submission to ACM SIGCOMM, 2001
- 6 Druschel P, A Rowstron A. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. Submission

- 4 Pal U, Belad A, Choisy C. Touching numeral segmentation using water reservoir concept. Pattern Recognition Letters, 2003, 24: 261~272
- 5 Lu Zhongkang, et al. Zheru Chi, Wan-Chi Siu, Pengfei Shi. A background-thinning-based approach for separating and recognizing connected handwritten digit strings. Pattern Recognition, 1999, 132: 921~933
- 6 Chen Y-K, Wang J-F. Segmentation of Single- or Multiple-Touching Handwritten Numeral String Using Background and Foreground Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11)
- 7 Tseng L Y, Chen R C. Segmenting handwritten Chinese characters based on heuristic merging of stroke bounding boxes and dynamic programming. Pattern Recognition Letters, 1998, 19: 963~973
- 8 Lu Y, Shridhar M. Character Segmentation in Handwritten Words-An Overview. Pattern Recognition, 1996, 29(1): 77~96
- 9 Casey R G, Lecolinet E. A Survey of Methods and Strategies in Character Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(7)
- 10 Maragoudakis M, Kavallieratou E, Fakotakis N, Kokkinakis G. How Conditional Independence Assumption Affects Handwritten Character Segmentation. ICDAR 2001
- 11 Lu Y, Shridhar M. Character Segmentation in Handwritten Words-An Overview. Pattern Recognition, 1996, 29(1): 77~96
- 12 Casey R G, Lecolinet E. A Survey of Methods and Strategies in Character Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(7)

to ACM SIGCOMM, 2001

- 7 Zhao B Y, Kubiawicz J, Joseph A D. Tapestry: An Infrastructure for Fault-tolerant Widearea Location and Routing. Technical Report No. UCB/CSD-01-1141, Computer Science Division, University of California at Berkeley, April 2001
- 8 Sivashanmugam K, Verma K, Mulye R, Zhong Z, Sheth A. Speed-R: Semantic P2P Environment for diverse Web Service Registries
- 9 Universal Description, Discovery and Integration: UDDI Technical White paper, September 2000
- 10 Gnutella website. www.gnutella.com
- 11 Ratnasamy S, Francis P, Handley M, Karp R, Shenker S. A Scalable Content-Addressable Network. In: Proc. of ACM SIGCOMM, 2001
- 12 Nejd W, et al. Super-peer-based routing and clustering strategies for RDF-based peer-to-peer networks. In: Proc. of the Intl. World Wide Web Conf. (Budapest, Hungary, May 2003)
- 13 Gandon F. Ontology Engineering: A Survey and a Return on Experience, Institut de Recherche en Informatique et Automatique (INRIA), March 2002, ISSN 0249-6399. <ftp://ftp-sop.inria.fr/acacia/fgandon/FabienGandon-RR4396.pdf>
- 14 El-Ansary S, Alima L O, Brand P, Haridi S. Efficient Broadcast in Structured P2P Networks. In: 2nd Intl. Workshop on Peer-to-Peer Systems (IPTPS '03)