

基于刻面与本体标识的语义 Web 服务发现方法

杜胜浩, 钱晓捷

(郑州大学 信息工程学院, 郑州 450001)

摘 要: 为快速且高效地从服务注册库中发现所需的语义 Web 服务, 提出一种语义 Web 服务发现方法。使用本体概念对语义 Web 服务进行刻面描述, 并将其转化为刻面树。在进行服务匹配之前对语义相似的服务进行聚类处理, 并在服务注册库中形成服务簇, 从而减少服务比较次数, 使服务匹配更高效。实验结果表明, 该方法的查准率和查全率分别为 87.6% 和 90.6%, 与基于聚类和二分图的方法相比, 服务响应时间减少 50%, 提高了服务发现的效率。

关键词: 语义 Web 服务; 刻面; 本体; 概念相似度; 聚类; 服务发现

中文引用格式: 杜胜浩, 钱晓捷. 基于刻面与本体标识的语义 Web 服务发现方法[J]. 计算机工程, 2018, 44(8): 224-229, 236.

英文引用格式: DU Shenghao, QIAN Xiaojie. Semantic Web service discovery method based on facet and ontology marker[J]. Computer Engineering, 2018, 44(8): 224-229, 236.

Semantic Web Service Discovery Method Based on Facet and Ontology Marker

DU Shenghao, QIAN Xiaojie

(School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China)

[Abstract] For quickly and efficiently find the required semantic Web service from the service register, a semantic Web service discovery method is proposed. The ontology concept is used to describe the semantic Web service and it is transformed into faceted tree. The semantic similarity service is clustered before the service matching, and the service cluster is formed in the service register. Then the number of service comparisons is reduced and the service matching is more efficient. Experimental results show that the precision and recall of the method is up to 87.6% and 90.6%, compared with clustering and two partite graph methods, the service response time is reduced by half, and the efficiency of service discovery is improved.

[Key words] semantic Web service; facet; ontology; concept similarity; clustering; service discovery

DOI: 10.19678/j.issn.1000-3428.0047537

0 概述

随着网络技术的不断发展, 人们提出 Web 服务的概念, 通过定义和建立一个 Web 服务界面就可以访问和共享计算资源或信息资源。Web 服务可以使用标准协议来组合、发布和定位, 以构建在一系列平台上运行的应用程序^[1]。语义 Web 服务依据本体概念对 Web 服务进行语义标注, 使服务信息能够被计算机理解并进行交互, 从而实现服务的自动发现和发现。随着语义 Web 服务的增加, 如何从规模不断扩大的服务注册库中发现所需服务是一个亟需解决的问题。文献[1]使用 UML 建模和术语扩展检索 Web 服务。文献[2]通过制定一个新的 Web 服务清晰度量表增强服务发现过程。文献[3]提出一

种上下文服务发现方法, 根据用户端绑定上下文来找出所需的服务。文献[4]基于本体的概念组之间的相关性度量模型分别评价服务请求与服务描述的语义匹配度, 从而进行语义 Web 服务的发现。文献[5]对发布的空间信息服务进行聚类, 计算服务请求与各服务簇中心的相似度, 确定最优匹配簇为服务发现结果。文献[6]采用空间向量模型表示语义 Web 服务, 对服务进行聚类, 并基于带权二分图思想对服务进行匹配。从本质上讲, Web 服务是一种轻量级的、松耦合的、与平台和语言无关的构件^[7]。因此, 对于服务注册库的管理与维护可以利用构件库技术来实现。大部分构件检索方法都使用刻面分类方法对构件进行描述和分类, 如文献[8]采用刻面分类法对 Web 应用构件进行描述和检索。一方面, 现

基金项目: 国家社会科学基金(14BYY096)。

作者简介: 杜胜浩(1990—), 男, 硕士研究生, 主研方向为信息检索、软件体系结构; 钱晓捷, 副教授。

收稿日期: 2017-06-09 **修回日期:** 2017-08-15 **E-mail:** du_shenghao@126.com

在的 Web 服务发现方法大都通过服务质量 (Quality of Service, QoS) 和功能来选择服务。即使具有相同输入和输出参数的 Web 服务的内部执行过程也会有很大不同,因此导致 Web 服务发现的准确率较低。另一方面,现有服务发现方法未将本体语言描述和服务聚类思想同时引入其设计思想中。本体语言使语义 Web 服务的描述更清晰,而且降低了计算复杂性、技术复杂性和概念复杂性。通过聚类形成服务簇,可以减少服务比较次数,优化服务发现过程。服务注册库中的很多服务能够完成相似的功能,具有比较近似的概念和语义信息,因此,可以对服务进行聚类。

本文使用 Web 服务本体语言 (Ontology Web Language for Service, OWL-S) 描述的语义 Web 服务信息,基于本体库将语义 Web 服务转化为刻面描述树,用于实现服务聚类的预处理。

1 语义 Web 服务的描述和发现框架

1.1 语义 Web 服务描述

OWL-S 是构建于 Web 本体语言 (Web Ontology Language, OWL) 之上用于描述 Web 服务的标记语言。OWL-S 的 Web 服务描述模型可以描述语义 Web 服务,它对服务的基本描述包括服务的基本信息、功能信息和属性等。Web 服务的管理与维护可以使用构件库技术实现,而对构件的描述和分类通常由刻面来完成。每个刻面都由一组基本的术语构成,这组术语称为术语空间^[9]。因此,语义 Web 服务中也可由一个刻面集合对其进行描述。

定义 1 用一个三元组 $WS = \{ WS_Name, WS_Description, WS_Facets \}$ 表示语义 Web 服务,其中, WS_Name 是语义 Web 服务的标识名,即服务的名称, $WS_Description$ 是语义 Web 服务使用文本描述的基本信息, WS_Facets 表示语义 Web 服务的刻面集。

定义 2 语义 Web 服务的刻面集由一组刻面术语的集合组成,每个刻面术语都用于进行本体标注。该服务集合中定义了具体的语义 Web 服务刻面,包括 QoS 刻面、功能刻面、参数刻面、应用刻面等。

使用本体标注的各个语义 Web 服务刻面包含不同的属性信息:

1) QoS 刻面,即服务质量刻面,是用户在使用语义 Web 服务时涉及的体验信息,即用户对语义 Web 服务的可靠性、安全性、可用性等非功能性属性的期望信息。自 QoS 被引入到 Web 服务发现领域后,为 Web 服务发现研究人员提供了新的研究方向^[10-11]。

2) 功能刻面,包括语义 Web 服务的接口信息和调用方式等。

3) 参数刻面,包括语义 Web 服务的输入参数集和输出参数集,分别由一组参数组成,每个参数都用于进行本体标注。

4) 应用刻面,是在应用语义 Web 服务时,服务运行前置条件描述的环境状况和服务运行对相关资源对象产生的效果等。

提取语义 Web 服务的语义信息并进行本体标注,形成的语义 Web 服务刻面集包括 4 个主刻面和若干子刻面,可用图 1 所示的刻面树表示,根节点用一个虚拟节点表示。

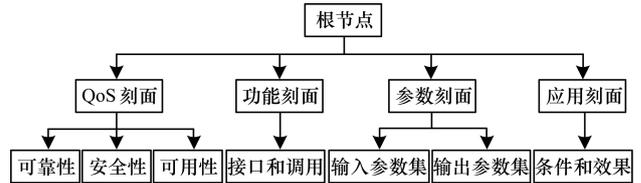


图 1 语义 Web 服务刻面集形成的刻面树

传统描述服务的方式通过一个四元组 $\{ Name, Description, Input, Output \}$ 来表示,其中, $Name$ 是服务的名称, $Description$ 表示 Web 服务的基本描述, $Input$ 和 $Output$ 分别表示服务的输入参数和输出参数的集合。但是,这种表示方式对于具有相同输入和输出的服务不能很好地进行区分。本文采用多刻面和本体对 Web 服务进行语义描述,增加了 QoS 刻面和应用刻面的描述,加强了 Web 服务的语义约束,从而提高了 Web 服务发现的准确率。

1.2 语义 Web 服务发现框架

本文提出的语义 Web 服务发现框架如图 2 所示,其中,本体库为共享本体库,本文中所有语义 Web 服务刻面描述都是基于此本体库。在服务提供者提供的语义 Web 服务经过服务分析器处理并注册到服务注册库后,通过服务注册库中的服务聚类器对语义 Web 服务进行聚类,能够得到多个服务聚类簇,减少服务匹配器在进行服务匹配时的比较次数,从而提高服务发现的效率。

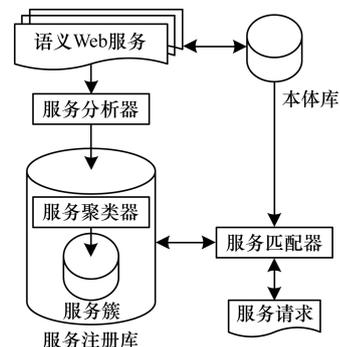


图 2 语义 Web 服务发现框架

该语义 Web 服务发现框架主要包含 3 个组件:

1) 服务分析器。根据服务提供者提供的语义 Web 服务的描述,提取语义信息并进行本体标注,进而将语义 Web 服务刻面集用一棵刻面树表示,为语义 Web 服务在服务注册库中的聚类提供数据。

2) 服务聚类器。语义 Web 服务经过服务分析器处理后,基于刻面对语义 Web 服务进行本体语义相似度的聚类,属于同一类别的语义 Web 服务聚集到一个服务簇中,在每一个服务簇中都选取一个语义 Web 服务作为服务簇中心,这些服务簇存放在服务注册库中。

3) 服务匹配器。对服务请求者的服务请求信息进行分析并依据本体库进行本体标注,依次计算请求服务与服务注册库中每个服务簇中心的相似度,取相似度最大的服务簇作为服务发现集合,然后计算请求服务与集合中的语义 Web 服务的相似度,将满足阈值要求的服务集合输出给服务请求者。

1.3 刻面的动态更新

随着 Web 服务的更新和扩展,语义 Web 服务的功能性和非功能性属性可能发生改变,对应的刻面值也要动态地更新。在服务提供者提供 Web 服务时,如果发现 Web 服务刻画集中没有对某一刻面的刻画,则服务提供者可以在已有 Web 服务刻画集的基础上定义新的刻面和对应的属性值,并添加到语义 Web 服务刻画集中。如果语义 Web 服务刻画集中某一刻面的描述错误或冗余,服务提供者可以将其修改或删除。以上语义 Web 服务刻画更新必须满足刻面的描述规约,否则更新无效。

以 QoS 主刻面子刻面的更新为例,在刻画描述中已有服务的可靠性、安全性和可用性,服务提供者发现没有对服务响应时间的刻画,于是服务提供者可根据刻画描述规约添加服务响应时间刻面到 QoS 主刻面子刻面下,扩展语义 Web 服务刻画集,同时更新刻面子刻面,实现刻面的动态更新。

2 语义 Web 服务聚类

2.1 本体概念语义相似度

本体作为 Web 核心支撑技术之一^[12],它是一个概念框架,用来对领域知识概念进行抽象和描述,形成领域术语,然后对概念进行标识,通过定义概念以及概念与概念之间的关系来描述语义信息^[13-14]。本体可表示成一棵有向树,例如 Food 本体有向树的部分信息如图 3 所示,树中每个节点表示一个概念。

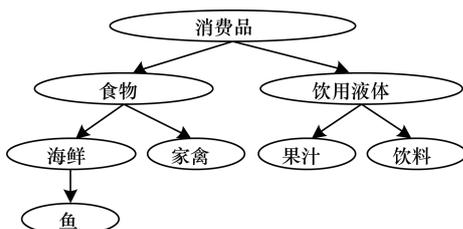


图 3 Food 本体有向树部分信息示例

本文使用文献[15]中的定义来计算本体距离,设有 2 个概念 C_1 和 C_2 ,则这 2 个概念的本体距离 $Dis(C_1, C_2)$ 为本体有向树中连接它们的最短路径的边数。例如在图 3 中,概念海鲜 (Seafood) 和概念果汁 (Juice) 之间的本体距离 $Dis(\text{Seafood}, \text{Juice}) = 4$ 。于是,2 个本体概念的语义相似度可基于这种本体距离来计算,计算公式为:

$$Sim(C_1, C_2) = \frac{Dis_{\max} - Dis(C_1, C_2)}{Dis_{\max} - Dis_{\min}} \quad (1)$$

其中, $Sim(C_1, C_2)$ 表示概念 C_1 和概念 C_2 的语义相似度, Dis_{\max} 表示本体有向树中所有概念的本体距离最大值, Dis_{\min} 表示本体有向树中所有概念的本体距离最小值, $Dis(C_1, C_2)$ 表示概念 C_1 和概念 C_2 的本体距离。

2.2 服务聚类

从使用 OWL-S 描述的语义 Web 服务中提取语义信息并进行本体标注,将其表示为一棵语义 Web 服务刻画树,完成语义 Web 服务聚类的预处理。将语义 Web 服务按照子刻面的描述聚集到所属刻面分类下,形成子刻面服务集合,然后聚类得到服务簇,同一个子刻面下的语义 Web 服务的相似度较大,不同子刻面下的语义 Web 服务相似度较小。因此,当服务请求者在查找所需服务时,在某种程度上缩小了服务请求者的查找范围,提高了服务发现的效率。本文采用类似于 K 近邻算法的聚类方法,使用本体概念语义相似度作为聚类标准,依据式(2)计算 2 个语义 Web 服务的语义相似度,相似度较大的服务聚成一类。式(2)的形式如下:

$$Sim(w_1^{WS}, w_2^{WS}) = \frac{1}{n} \sum_{i=1}^n Sim(w_{1i}^{WS}, w_{2i}^{WS}) \quad (2)$$

其中, w_1^{WS} 和 w_2^{WS} 为 2 个使用了 n 个刻面(本体概念)描述的语义 Web 服务,即 $w_1^{WS} = \{w_{11}^{WS}, w_{12}^{WS}, \dots, w_{1n}^{WS}\}$, $w_2^{WS} = \{w_{21}^{WS}, w_{22}^{WS}, \dots, w_{2n}^{WS}\}$, $Sim(w_1^{WS}, w_2^{WS})$ 表示 2 个服务的相似度, n 为刻面的个数, $Sim(w_{1i}^{WS}, w_{2i}^{WS})$ 表示 2 个刻面的本体概念的语义相似度,使用式(1)计算得到。

子刻面具体聚类步骤如下:

步骤 1 初始化聚类中心。对服务注册库中已表示成服务刻画树的语义 Web 服务按照刻面描述进行分类,从所要聚类的子刻面下的语义 Web 服务中选取 k 个类作为初始聚类中心,在选择聚类中心时尽量使得 k 个作为中心的语义 Web 服务的相似度很小,得到初始聚类中心 $c = \{c_1, c_2, \dots, c_k\}$ 。

步骤 2 根据式(1)和式(2),分别计算当前子刻面下其他语义 Web 服务与这 k 个作为类中心的语义 Web 服务的相似度,并归类于与其相似度最大的那个类。

步骤3 分别对得到的 k 个类取新的聚类中心。根据中位数的思想更新聚类中心,依次对 k 个类进行如下操作:按照相似度大小对语义 Web 服务进行排序,假如当前类有 n 个 Web 服务,取第 $\lfloor \frac{n}{2} \rfloor$ 个构件作为新的聚类中心,于是得到新的聚类中心 $c' = \{c_1', c_2', \dots, c_k'\}$ 。并且初始化一个累加变量 $count$, 赋初值为 0。

步骤4 判断新得到的聚类中心 c' 与原聚类中心 c 是否完全相同,若完全相同则结束聚类,跳到步骤 7; 否则判断 $count$ 的值,若 $count$ 的值大于 1 000, 则跳到步骤 7, 否则继续向下执行。

步骤5 将所有语义 Web 服务与其所在类的聚类中心的相似度累加到一起,记为 $SumSim$ 。多次实验结果表明,相似度总和 $SumSim$ 越大,对语义 Web 服务的聚类效果越好。

步骤6 根据新得到的聚类中心按照步骤 2 中的聚类方式进行聚类,得到 k 个新的类。按照步骤 5 的方式计算相似度总和,得到 $sumSim'$ 。若 $sumSim$ 小于 $sumSim'$, 则表示新得到的 k 个类的聚类效果优于之前的类,将聚类中心 c' 赋予 c , 让 c 保存拥有最好聚类效果的聚类中心,然后跳转到步骤 3; 否则,直接按照步骤 3 的方式得到新的聚类中心 $c' = \{c_1', c_2', \dots, c_k'\}$, 并令 $count = count + 1$, 跳转到步骤 4。

步骤7 聚类完成,结束聚类算法。

在服务注册库中对每个子刻面下的语义 Web 服务都按照上述聚类算法进行聚类,从而将相似度较高的语义 Web 服务聚集到一个类中,形成相应的服务簇。最终聚类的个数,即服务簇的个数取决于子刻面的个数和聚类时选取的 k 值的大小。本文的服务发现方法在对语义 Web 服务进行聚类时,需要提前给出聚类的个数(即 k 的值),而且聚类的个数直接影响聚类结果,因此,要经过多次不同聚类个数的实验才能确定较好的聚类个数。

3 语义 Web 服务发现过程

基于领域本体知识对语义 Web 服务进行多刻面表示,多角度描述语义 Web 服务。在服务注册库中对语义 Web 服务进行基于语义相似度的聚类,使得服务发现方法具备一定的模糊匹配功能。另外,服务请求者对要请求的语义 Web 服务的描述信息可能不完整,没有包含所有语义 Web 服务刻面集中的刻面描述信息(即本体概念),服务匹配器为服务请求者没有输入的信息设置默认值,保证服务发现方法具备一定的张弛能力,使 Web 服务发现结果具有良好的查准率与查全率。

服务注册库中的语义 Web 服务通过聚类处理

并形成服务簇后,由服务匹配器对服务请求者提供的服务请求进行处理,之后进行本体标注和刻面描述,用一组刻面描述表示服务请求者的服务请求,即 $d_{QS} = \{d_1^{QS}, d_2^{QS}, \dots, d_n^{QS}\}$, d_i^{QS} 为服务请求的一个刻面描述。于是,服务请求和语义 Web 服务就可以进行语义相似度的计算,公式如式(3)所示。

$$\text{sim}(w_{ws}, d_{QS}) = \frac{1}{n} \sum_{i=1}^n \text{sim}(w_i^{ws}, d_i^{QS}) \quad (3)$$

其中, w_{ws} 和 d_{QS} 分别是使用了 n 个刻面描述的语义 Web 服务和请求, $\text{sim}(w_{ws}, d_{QS})$ 表示语义 Web 服务和请求的相似度, n 为刻面的个数, $\text{sim}(w_i^{ws}, d_i^{QS})$ 表示通过式(1)计算得到的 2 个刻面的本体概念语义相似度。

从服务注册库中的多个服务簇的簇中心中选取与请求服务相似度最大的簇中心,记录该服务簇中的所有语义 Web 服务。根据刻面划分原理,一个主刻面下不同子刻面包含的语义 Web 服务不同,与服务请求拥有较高相似度的簇中心所在的服务簇中的所有语义 Web 服务都可作为候选服务发现结果,于是对同一主刻面不同子刻面下被记录的语义 Web 服务集合求并集。由于一个语义 Web 服务含有多个刻面描述,同一个语义 Web 服务会出现在不同主刻面下,因此再对不同主刻面下满足条件的所有语义 Web 服务求交集,使得到的服务发现结果中的语义 Web 服务不重复。计算所得语义 Web 服务集合中的每一个服务与服务请求的相似度,满足要求的语义 Web 服务与服务请求的相似度记为 S_1 ($0 \leq S_1 \leq 1$)。上述语义 Web 服务集合中的服务在不同主刻面下都有其与所属服务簇的簇中心的相似度。为了体现聚类和簇中心在计算服务请求与服务发现结果中的语义 Web 服务的最终相似度所起到的作用,将语义 Web 服务在不同主刻面下的相似度求和取平均数,记为 S_2 ($0 \leq S_2 \leq 1$),于是可以得到服务请求与上述语义 Web 服务集中服务的最终相似度,记为 S ($0 \leq S \leq 1$),令 $S = (S_1 + S_2)/2$ 。最后,对相似度大于阈值的语义 Web 服务按照相似度进行排序,得到最终服务发现结果。

服务发现算法流程描述如下:

输入 服务请求描述 d_{QS} , 服务簇 s_{sc} , 服务本体库 l_{sol} , 服务刻面树 T

输出 相似度从大到小的 N 个构件

1. 对 d_{QS} 进行基于 l_{sol} 的刻面描述
2. for s_{sc} 中的所有簇中心
3. $Sim_QW = \text{sim}(d_{QS}, w_{ws})$;
4. end for;
5. for T 的子刻面层的所有节点
6. $maxsim = \max(Sim_QW, k)$;
7. 相似度最大的服务簇中的所有语义 Web 服务保存到

```

maxSim_WS 中;
8. end for;
9. for T 的主刻面下的所有节点
10. unionQW = Union( maxSim_WS, n );
11. unionQW 中的语义 Web 服务与服务簇中心的相似度的值保存到 unionQW_Sim;
12. end for;
13. array_mixQW = mix( unionQW, n );
14. array_mixQW_Sim 记录 array_mixQW 中的语义 Web 服务与簇中心的相似度;
15. for array_mixQW 中的所有语义 Web 服务
16.  $S_1 = \text{sim}(w_{ws}, d_{qs})$ ;
17.  $S_2 = n$  个主刻面下的相似度的平均值;
18. if(  $(S_1 + S_2) / 2 \geq \text{阈值}$  )
19. result = ( array_mixQWi,  $(S_1 + S_2) / 2$  );
20. end for
21. sort( result );

```

在服务发现算法中,步骤1是将语义Web服务请求进行基于本体库的刻面表示;步骤2~步骤4用式(3)计算服务请求与服务簇的簇中心的相似度;步骤5~步骤8是选取与服务请求相似度最大的簇中心,步骤6中的 k 为聚类个数;步骤9~步骤12是同一主刻面下不同子刻面中选出的服务簇求并集;步骤13和步骤14是对不同主刻面下得到的服务簇求交集,同时记录语义Web服务与所属服务簇中心的相似度;步骤15~步骤20是计算服务请求与符合要求语义Web服务的总相似度,大于阈值条件的加入服务发现结果集中;步骤21是对得到的服务发现结果集中的服务按照相似度排序,并将结果呈现给用户。该算法中刻面的个数相比于语义Web服务的个数可以忽略不计,故算法的时间复杂度为 $O(n)$ 。

4 实验结果及分析

为验证本文所提出的语义Web服务发现方法

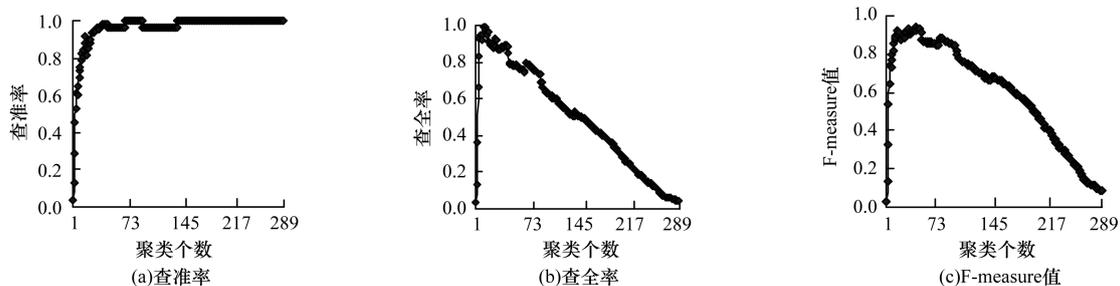


图4 不同聚类个数查准率、查全率和F-measure值趋势

根据图4中聚类个数对查准率、查全率和F-measure值影响的分析确定合适的聚类个数,然后在不同领域下进行实验,采用OWLS-TC4测试集中提供的请求,同时与文献[6]所提出的基于聚类和二分图匹配的Web服务发现方法作对比,最后对每个领域的多次服务请求所得的查准率和查全率分别计算平均

在查准率、查全率以及执行效率方面的性能,使用OWLS-TC4测试集进行实验。OWLS-TC4测试集提供了来自9个不同领域的总共1083个语义Web服务,包括教育、医疗保健、食品、旅游、通信、经济、武器、地理和仿真。本文选取服务个数和请求个数较多的4个领域的语义Web服务进行实验,实验数据如表1所示。

表1 语义Web服务的实验数据

序号	领域	服务个数	请求个数
1	地理	60	10
2	旅游	197	6
3	教育	286	6
4	经济	395	12

利用Protégé工具处理OWLS-TC4测试集中服务的本体概念,借助MySQL数据库,使用Java语言在Eclipse中对算法进行编码和实现。实验中服务发现的效果通过评测查准率、查全率和执行时间3个指标来体现,查准率和查全率的计算公式分别为:查准率 $P = n_r / N_q$,查全率 $R = n_r / N_r$ 。其中, n_r 是检索结果中满足服务请求的语义Web服务个数, N_q 是检索得到的服务总个数, N_r 是服务注册库中满足服务请求的服务个数。另外,通过在不同相似度阈值下进行的实验测试,发现阈值取为0.7,即服务相似度 $S \geq 0.7$ 时,检索结果的查准率和查全率较好,故在本文中服务相似度阈值设为0.7。

考虑到聚类个数对聚类结果存在直接影响,进而影响实验结果的查准率和查全率,于是使用本文提出的服务发现方法进行多次不同聚类个数的实验,选取教育领域的286个语义Web服务,不同聚类个数的查准率、查全率和F-measure值如图4所示。

值。不同领域下2种服务发现方法的查准率和查全率分别如图5和图6所示。通过2种服务发现方法的查准率和查全率的对比可以发现,相较于基于聚类与二分图匹配服务的发现方法,本文方法对服务的查准率和查全率较高。实验结果表明,本文方法的平均查准率在85%以上,查全率是90%左右。

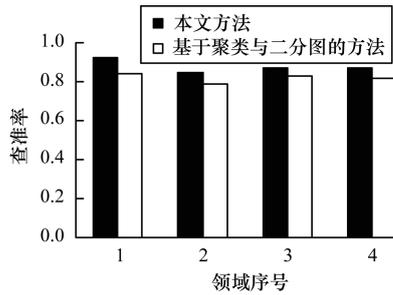


图5 不同领域查准率对比

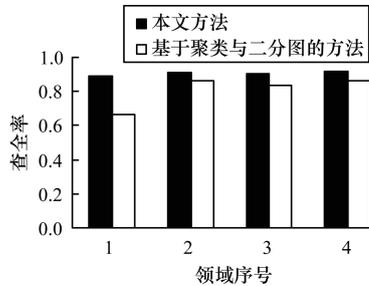


图6 不同领域查全率对比

实验中算法的执行效率(即服务请求的响应时间)除了和服务发现方法有关,还与进行实验时使用的仿真平台及硬件配置有关,本文进行实验时使用的电脑处理器是 AMD Athlon X4。对不同领域中的服务请求分别进行响应时间的测试,选取 OWLS-TC4 测试集中经济领域的 12 个服务请求,2 种服务发现方法的服务请求响应时间对比如图 7 所示。

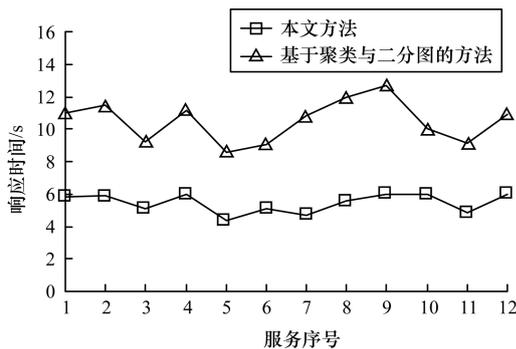


图7 服务请求响应时间对比

实验结果表明,本文方法的服务请求响应时间比基于聚类和二分图匹配语义的 Web 服务方法要少很多,服务发现效率明显提高。综合服务的查准率、查全率和服务请求响应时间,本文提出的服务发现方法是有效的,而且拥有较好的服务发现效果。

5 结束语

在软件开发过程中对现有软件进行复用,可以在很大程度上节约开发成本和开发时间。随着人们对 Web 服务的研究与开发,从服务注册库中快速且

有效地找到用户所需要的服务对软件检索技术提出了更高的要求。本文提出基于刻面和本体标识的语义 Web 服务发现方法,通过服务分析器将语义 Web 服务转化成基于本体概念的刻面树,再使用服务聚类器对服务注册库中的语义 Web 服务按照刻面进行聚类,得到多个服务簇。实验结果证明本文的服务发现方法具有较好的查准率、查全率和执行效率。

语义 Web 服务的研究依赖于语义分析技术,对语义 Web 服务进行刻面描述时,刻面的选取以及相关的本体概念都对本文的服务发现方法有影响。另外,本文都是在同一个领域中进行服务发现实验,而且数据也不够多。因此,后续工作将对语义分析技术、语义 Web 服务的刻面分类描述以及聚类方法进行研究,完善服务发现方法、增加实验数据,进一步提高检索效果。

参考文献

- [1] LEE W L, MA S, TSAI Y. Retrieval of Web service components using UML modeling and term expansion[J]. *Journal of Information Science and Engineering*, 2017, 33(1):17-36.
- [2] MAABED U M, EL-FATATRY A, EL-ZOGHABI A. Enhancing interface understandability as a means for better discovery of Web services [J]. *International Journal of Information Technology and Web Engineering*, 2016, 11(4):1-23.
- [3] MA Shangpin, LAN Ciwei, LI Chia-Hsueh. Contextual service discovery using term expansion and binding coverage analysis [J]. *Journal of Future Generation Computer Systems*, 2015, 48(3):73-81.
- [4] 夏昊翔,王玉影,宣照国,等. 基于概念相关性度量的语义 Web 服务匹配:模型及匹配器原型[J]. *情报学报*, 2008, 27(3):375-385.
- [5] 陈科,成毅,谢明霞,等. 基于服务簇的空间信息服务自动发现[J]. *计算机工程*, 2012, 38(24):182-187.
- [6] 刘一松,朱丹. 基于聚类与二分图匹配的语义 Web 服务发现[J]. *计算机工程*, 2016, 42(2):157-163.
- [7] TSALGATIDOU A, PILIOURA T. An overview of standards and related technology in Web services[J]. *Distributed and Parallel Databases*, 2002, 12(2/3):135-136.
- [8] 彭博. 一种基于刻面分类法的 Web 应用构件库的研究与实现[D]. 北京:北京工业大学, 2009.
- [9] 王映辉. 构件式软件技术[M]. 北京:机械工业出版社, 2012.
- [10] RAN Shuping. A model for Web services discovery with QoS[J]. *ACM SIGecom Exchanges*, 2003, 4(1):1-10.
- [11] 杨胜文,史美林. 一种支持 QoS 约束的 Web 服务发现模型[J]. *计算机学报*, 2005, 28(4):589-594.

以上5组实验结果表明,本文建立的 BiLSTM-CNN-CRF 模型通过使用字符向量、词性向量和词向量组合的混合向量,在维吾尔文命名实体识别任务中达到了最好的性能。

4 结束语

针对维吾尔文命名实体识别任务,本文以传统的 CRF 统计模型作为基准进行实验,总结维吾尔文命名实体识别中出现的问题,进而构建基于 BiLSTM-CNN-CRF 框架的神经网络模型。该模型在 CNN 层捕获字符级特征向量,在 BiLSTM 层获取当前词语的过去和将来的上下文信息,在 CRF 层对 BiLSTM 层的输出进行解码,最终输出最优的标记序列。基于 UNERDATA 语料的实验结果进一步验证了 BiLSTM-CNN-CRF 框架对维吾尔文命名实体识别的有效性。

本文构建的 BiLSTM-CNN-CRF 深度学习模型能够在维吾尔文命名实体识别语料库上得到较好的实验结果,并已应用于维吾尔文网络舆情分析系统,有效识别出了文本中的人名、地名和机构名,提高了舆情系统分析能力。后续将进一步完善语料库,加入二级词性标注特征信息,并在新语料库的基础上测试本文模型的性能。

参考文献

- [1] CHIU J P C, NICHOLS E. Named entity recognition with bidirectional LSTM-CNNs [EB/OL]. [2017-05-11]. <https://arxiv.org/pdf/1511.08308v1.pdf>.
- [2] 张海楠,伍大勇,刘悦,等.基于深度神经网络的中文命名实体识别[J].中文信息学报,2017,31(4):28-35.
- [3] 艾斯卡尔·肉孜,宗成庆,姑丽加玛丽·麦麦提艾力,等.基于条件随机场的维吾尔人名识别方法[J].清华大学学报(自然科学版),2013,53(6):873-877.
- [4] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [EB/OL]. [2017-05-11]. <https://arxiv.org/pdf/1310.4546.pdf>.
- [5] REI M, CRICHTON G K O, PYYSALO S. Attending to characters in neural sequence labeling models [EB/OL]. [2017-05-11]. <https://arxiv.org/pdf/1611.04361v1.pdf>.
- [6] DERNONCOURT F, LEE J Y, SZOLOVITS P. NeuroNER: an easy-to-use program for named-entity recognition based on neural networks [EB/OL]. [2017-05-11]. <https://arxiv.org/pdf/1705.05487.pdf>.
- [7] MA X Z, HOVY E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF [EB/OL]. [2017-05-11]. <https://arxiv.org/pdf/1603.01354.pdf>.
- [8] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition [EB/OL]. [2017-05-11]. <https://arxiv.org/pdf/1603.01360.pdf>.
- [9] LI L S, MAO T, HUANG D, et al. Hybrid models for Chinese named entity recognition [C] // Proceedings of the 5th SIGHAN Workshop on Chinese Language Processing. Beijing, China: [s. n.], 2006: 72-78.
- [10] COLLOBERT R, BOTTOU J W L, KARLEN M, et al. Natural language processing (almost) from scratch [J]. The Journal of Machine Learning Research, 2011, 12: 2493-2537.
- [11] HUANG Z H, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging [EB/OL]. [2017-05-11]. <https://arxiv.org/pdf/1508.01991v1.pdf>.
- [12] 陈斌,周勇,刘兵.基于卷积长短期记忆网络的事件触发词抽取方法 [J/OL]. 计算机工程: 1-7 [2018-02-17]. <http://kns.cnki.net/kcms/detail/31.1289.TP.20180206.1410.002.html>.
- [13] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [EB/OL]. [2017-05-11]. <https://arxiv.org/pdf/1310.4546.pdf>.
- [14] 喻靖民,向凌云,曾道建.一种基于 Word2vec 的自然语言隐写分析方法 [J/OL]. 计算机工程: 1-7 [2018-02-17]. <http://doi.org/10.19678/j.issn.1000-3428.0050407>.
- [15] 王洪亮.基于词向量聚类的中文微博产品命名实体识别 [J]. 兰州理工大学学报, 2017, 43 (1): 104-110.
- [16] YANG Z, SALAKHUTDINOV R, COHEN W. Multi-task cross-lingual sequence tagging from scratch [EB/OL]. [2017-05-11]. <https://arxiv.org/pdf/1603.06270v1.pdf>.
- [17] MIKOLOV T, KARAFIAT M, BURGET L, et al. Recurrent neural network based language model [C] // Proceedings of INTERSPEECH ' 10. Makuhari, Japan: [s. n.], 1045-1048.
- [18] HAMMERTON J. Named entity recognition with long short-term memory [C] // Proceedings of HLT-NAACL ' 03. [S. l.]: ACL, 2003: 172-175.

编辑 金胡考

(上接第 229 页)

- [12] 岳昆,王晓玲,周傲英,等. Web 服务核心支撑技术研究综述 [J]. 软件学报, 2004, 15 (3): 428-442.
- [13] GRUBERT R. A translation approach to portable ontology specifications [J]. Knowledge Acquisition, 1993, 5 (2): 199-220.
- [14] 吴健,吴朝晖,李莹,等.基于本体论和词汇语义相似度的 Web 服务发现 [J]. 计算机学报, 2005, 28 (4): 595-602.
- [15] 孙萍,蒋昌俊.利用服务聚类优化面向过程模型的语义 Web 服务发现 [J]. 计算机学报, 2008, 31 (8): 1340-1353.

编辑 顾逸斐