

云制造背景下的服务匹配算法

张金广,李锋刚,张磊

(合肥工业大学管理学院,安徽合肥230009;
教育部过程优化与智能决策重点实验室,安徽合肥230009)

摘要:云制造背景下的服务匹配具有复杂性、海量性,因而需要高效的匹配算法来提高响应度。以往的基于本体的语义检索算法,往往只重视结果的查全率和查准率,不能有效满足云制造背景下的海量数据处理的效率要求。文中将基于语义距离的语义相似度算法应用到云制造背景下的服务匹配中,在满足查准率和查全率的同时,通过设置相应的过滤器,极大减少了不必要的匹配过程。实验结果表明本算法很好地满足了云制造背景下服务搜索的查全率和效率要求。

关键词:本体;云制造;服务匹配

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2013)03-0041-04

doi:10.3969/j.issn.1673-629X.2013.03.011

Services Matching Algorithm under Cloud Manufacturing Background

ZHANG Jin-guang, LI Feng-gang, ZHANG Lei

(School of Management, Hefei University of Technology, Hefei 230009, China;
Key Laboratory of Process Optimization and Intelligent Decision-making of Ministry of Education,
Hefei 230009, China)

Abstract: With the complexity and huge amount of service mapping based on cloud manufacturing background, an effective service mapping algorithm is needed to improve the response. Most service searching methods just focus on accuracy and recall, which can't meet the need of high efficiency based on cloud manufacturing background. A semantic similarity algorithm based on semantic distance is applied to service matching under cloud manufacturing background. The algorithm can reduce unnecessary mappings by setting a filter, which satisfies the need of service searching based on cloud manufacturing background. Experiment demonstrates the algorithm can meet the requirements of accuracy and recall.

Key words: ontology; cloud manufacturing; services matching

0 引言

近年来,随着云计算、物联网等技术的发展和日趋成熟,一种面向服务的网络化制造新模式——云制造应运而生,它是一种利用网络和云制造服务平台,按用户需求组织网上制造资源(制造云),为用户提供各类按需制造服务的一种网络化制造新模式^[1]。云制造环境下蕴含大量形式多样、复杂程度不同的制造服务信息,设计一种语义化、高效能服务搜索与动态匹配技术就成为实现云制造系统中的一个关键技术。目

前,实现语义搜索的最好的方式之一是通过领域本体来对概念进行扩展,从而提高服务的查全率。

目前,对于概念之间语义相似度的计算主要从以下3个不同的角度进行了研究^[2]:一是基于距离的语义相似度计算;二是基于信息内容的语义相似度计算;三是基于属性的语义相似度计算模型。文中通过基于语义距离的算法来确定服务中各元素的相似度,然后通过加权平均来最终计算出服务间的相似度。

为了改善匹配的效率,减少不必要的匹配,文中首先将信息检索领域的查询扩展技术应用到服务匹配过程中,通过基于本体的关系数据库关键词查询的语义查询扩展方法^[3],把用户提交的服务的功能、数据、服务质量和执行进行语义查询扩展,将其扩展为基于本体的语义关键词,并根据扩展后的关键词来确定一个初始结果集,所有的匹配都在该结果集内进行,通过减少大量的无效匹配来极大地提高匹配效率。

收稿日期:2012-06-25;**修回日期:**2012-09-27

基金项目:国家“863”云制造主题项目(2011AA040501);国家自然科学基金资助项目(70871033);安徽省教育自然科学基金重点项目(KJ2011A006)

作者简介:张金广(1984-),男,硕士生,研究方向为人工智能、机器学习。

1 相关概念

1.1 云制造

李伯虎^[4]等在“制造即服务”理念以及云计算和物联网等新兴技术发展的基础上提出云制造的概念。云制造,是先进的信息技术、制造技术以及新兴物联网技术等交叉融合的产品。采取包括云计算在内的当代信息技术前沿理念,支持制造业在广泛的网络资源环境下,为产品提供高附加值、低成本和全球化制造的服务,是一种面向服务的、高效低耗和基于知识的网络化敏捷制造新模式。

云制造服务语义分为功能语义、数据语义、服务质量语义和执行语义等四种类型。其中:功能语义,指服务的实际功能描述,它是通过功能本体来表达且标注在 Web 服务描述语言文档中的操作上;数据语义,是指服务的输入/输出参数集合的语义表达,标注在 WSDL 文档中的操作上;服务质量语义,是指服务的众多质量性能的语义表达,在 WSDL 文档中可以进行单独的服务质量语义标注;执行语义,是指与服务执行相关的消息序列、任务流程、服务调用的前提条件/结果等语义信息,在 WSDL 文档中可以进行单独的服务过程执行语义标注。

1.2 Web 服务本体语言 OWL-S

因为制造服务也是 Web 服务的一种,因此,可以将 Web 服务匹配的算法及过程应用到云制造环境下的制造服务匹配。目前,Web 服务语义描述规范中最具有代表性的研究成果是 OWL-S^[5] (Web Ontology Language for Services)。OWL-S 是一种基于本体论思想构建的 Web 服务上层本体,如图 1 所示。

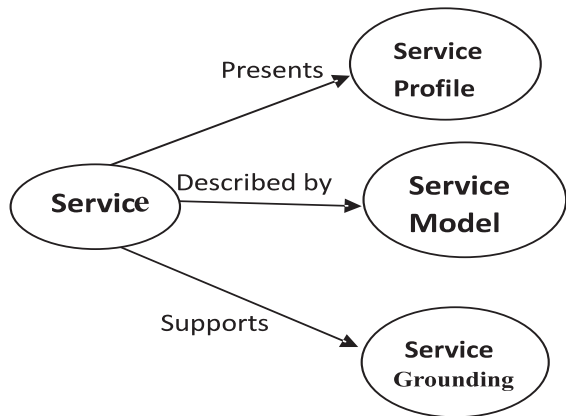


图 1 OWL-S 上层本体

OWL-S 包括三个组件:ServiceProfile,描述服务的功能,服务搜寻代理通过 ServiceProfile 实现服务匹配,寻找到满足服务请求者需求的 Web Service; Service-Model,服务的具体实现细节; ServiceGrounding,描述如何访问服务。文中主要针对 ServiceProfile,实现服务匹配过程。

2 云背景下的服务匹配算法

2.1 相关服务匹配算法分析

Paolucci 等人提出的 OWL-S/UDDI 匹配算法^[6]是一个经典算法,它将本体概念之间的匹配程度分为 4 个不同的等级:精确匹配、包含匹配、嵌入匹配以及失败匹配。下面以发布服务与请求服务的输出为例:

设请求服务的输出概念为 outR,发布服务的输出概念为 outA。

若 outR = outA 或者 outR 是 outA 的直接子类,则属于精确匹配;

若 outR 包含 outA,则属于包含匹配;

若 outA 包含 outR,则属于嵌入匹配;

若上述条件都不成立,则属于失败匹配。

这一算法具有局限性:它只是 4 个离散的匹配程度,不能给出服务之间的精确匹配,且不是所有没有包含关系的服务都为失败匹配。文献[7]通过定义一个分段函数计算本体中的两个概念相似度,并与 OWL-S/UDDI 算法的匹配结果进行比较,但概念的相似度匹配程度仍然不高。文献[8]考虑了概念之间继承关系、二元关系以及概念深度对相似度的影响,但其算法只能应用在由单个本体概念构成的服务接口。杨佳^[9]等人在计算语义相似度时考虑了概念间路径长度、概念深度和概念密度因素,并给出了相应改进算法,但计算量较大,不适合云制造特点。

2.2 改进的服务匹配算法

Lin 等人提出了基于语义距离的语义相似度计算公式^[10],如下:

$$\text{sim}(c_1, c_2) = e^{\alpha L} \times \frac{e^{\beta H} - e^{-\beta H}}{e^{\beta H} + e^{-\beta H}} \quad (1)$$

其中: c_1 和 c_2 表示两个概念, H 为 c_1 和 c_2 间的相距边数和 L 与最近公共节点概念在本体层次结构中所处的深度, α 和 β 用来调整 L 和 H 对概念相似度的影响因子;相似度 sim 关于 L 单调递减,关于 H 单调递增; $\alpha = 0.2$, $\beta = 0.6$ 时为最佳度量效果。文中采用公式(1)来计算云环境下的服务相似度。

图 2 为服务匹配模型。

一般情况下,语义 Web 服务匹配是通过遍历的方式,逐个将发布的 Web 服务与用户请求的 Web 服务的文本描述、输入、输出、前提条件、效果和服务质量匹配,计算它们的综合匹配度。最后依据综合匹配度对服务进行筛选和排序并返回给用户。但在云环境下,由于服务数量较大,运用比较复杂的算法会导致响应速度极低。因此,必须通过设置过滤器来缩小匹配范围,从而减少响应时间。文中通过功能过滤器^[11],根据用户输入的四服务语义的重要性(高、中、低)等级,选择等级为高的语义类型来设置过滤器,比如:用

户将功能和服务质量的等级设为高,将过滤器种类设置为功能和服务质量,过程如下:

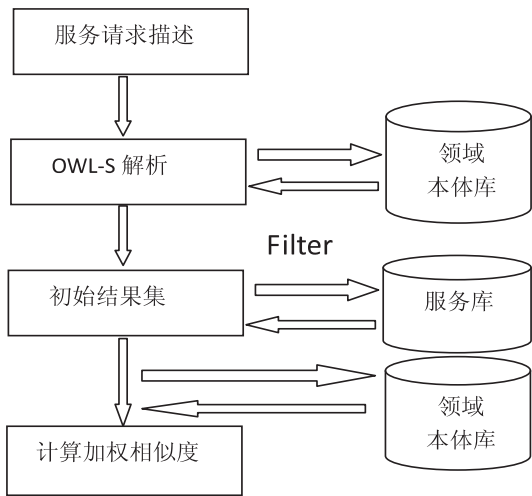


图2 服务匹配模型

- 1) 功能过滤器。通过设定一个最低相似度阈值,当服务的功能名称与用户请求的功能名称的相似度高于该值时,将该服务放入初始结果集,否则舍弃该服务。
- 2) 服务质量过滤器。通过功能过滤器后,再通过质量过滤器,把低于设置阈值的服务从初始结果集中去除。
- 最后得到的初始结果集就会小很多,然后再将初始结果集中的服务通过公式(2)、(3)、(4)来计算最终的匹配度。

$$n_h w_{h0} x + n_m w_{m0} x + n_l w_{l0} x = 1 \Rightarrow x = ? \tag{2}$$

$$w_h = w_{h0} \bullet x, w_m = w_{m0} \bullet x, w_l = w_{l0} \bullet x \tag{3}$$

$$\text{sim}(s_1, s_2) = w_f \times s_f + w_d \times s_d + w_q \times s_q + w_p \times s_p \tag{4}$$

因为用户检索时可能出现多个等级为高(或中、低)的语义项,因此需要通过式(2)和(3)来进行归一化处理。其中公式(2)中: n_h 、 n_m 、 n_l 分别为用户请求中服务语义项等级为高、中、低的数量; w_{h0} 、 w_{m0} 、 w_{l0} 分别为等级高、中、低的初始权重,文中使用0.5,0.3,0.2(可根据需要调整), x 为单位权重。公式(3)中: w_h 、 w_m 、 w_l 分别代表本次用户检索时,权重为高、中、低的服务语义项经过归一化处理的具体权重。公式(4)中: s_1 为用户请求的服务, s_2 为初始结果集中

的某具体服务, w_f 、 w_d 、 w_q 、 w_p 分别为用户输入的功能、数据、服务质量、执行的权重,比如功能和质量的等级为高,那么 w_f 、 w_q 即为 w_h ; s_f 、 s_d 、 s_q 、 s_p 分别为用户检索模型(功能、数据、服务质量、执行)与初始结果集中的某具体服务(功能、数据、服务质量、执行)的语义相似度,为公式(1)计算所得。

2.3 服务匹配过程

- 文中使用的服务匹配过程如下:
- 步骤1 过滤器设置。首先用 SPARQL^[12] 语句对等级为高的语义项进行语义扩展,这时得到的结果集 tempSet 与检索请求概念之间的边距为 1,通过递归调用分别对结果集 A 中的概念进行扩展,此时得到的结果集 B 与检索请求概念之间的边距就为 2,将边距带入公式 1,即可计算出扩展结果集与检索请求之间的相似度。通过设定一个阈值,当扩展结果集与检索请求之间的相似度低于此阈值时,不再继续扩展,此时得到结果集 funSet。
- 步骤2 通过 sql 语句检索所有服务功能在结果集 funSet 中的服务,得到结果集 Set。
- 步骤3 根据公式(1)依次计算输入/输出、执行等的相似度,并通过与步骤 1 中计算得到的相似度通过公式(4)计算最终的相似度。
- 步骤4 按照最终的相似度大小对 Set 中的服务进行排序,展现给用户。

3 实验与分析

目前在云制造领域还没有与之相关的标准平台和标准测试数据集,文中通过 protégé 4.1 建立了关于导弹的本体(如图3)以及100个与导弹本体相关的云制造服务集。为了证明文中算法的优越性,实验同时采用了基于关键词服务匹配方案以及 OWL-S/UDDI 匹配算法的结果进行了对比并对每个用户请求用三种实验方法来检索,并分别从查准率和查全率以及检索所用时间进行了对比。

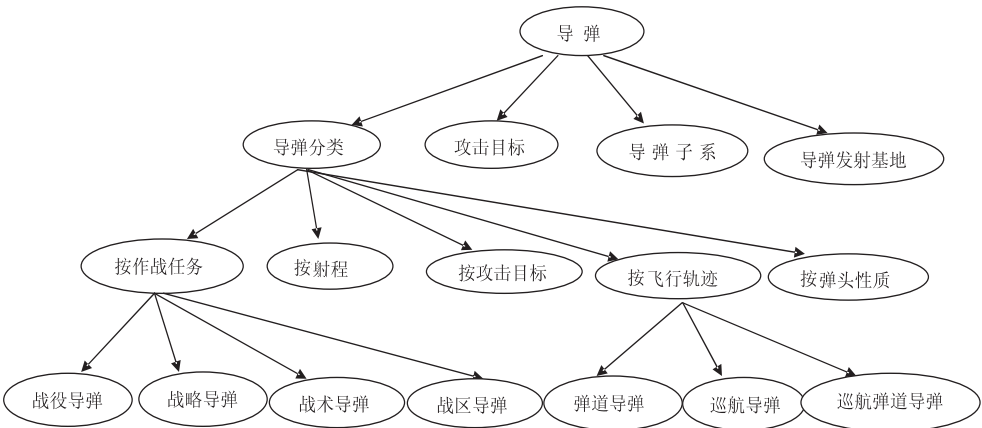


图3 部分导弹领域本体

文中试验在 Windows XP 环境下开发,硬件配置为 Inter(R) core(TM) 2 duo CPU E7500 @ 2.93GHz, 1.96GB 内存,采用 myeclipse805 作为开发工具。试验结果如图 4。

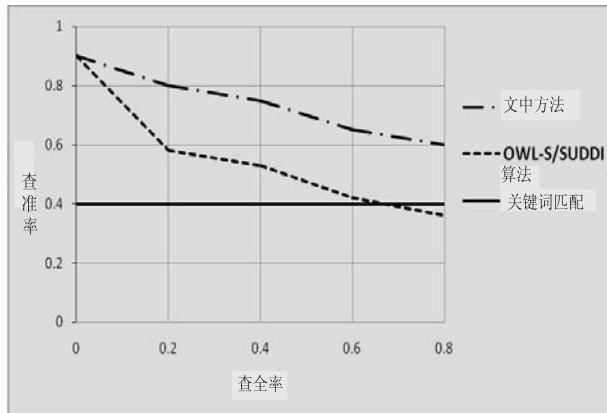


图 4 查全率与查准率对比

通过分析图 4 可以发现,与基于关键词匹配方法相比较,由于引入了领域本体,使得可以从语义上搜索相关的服务而不仅仅是关键词的匹配,因此提高了服务的查全率和查准率;与 OWL-S/UDDI 匹配算法相比,文中算法随着查全率的提高,查准率始终高于 OWL-S/UDDI 匹配算法,这时因为 OWL-S/UDDI 匹配算法只是 4 个离散的匹配程度,不能给出服务之间的精确匹配且没有包含关系的服务都为失败匹配,而文中考虑了两个概念间的边数及与最近公共节点概念在本体层次结构中所处的深度,从而实现查全率与查准率的提高。

4 结束语

云制造语义匹配有其特殊之处,文中根据其语义类型,运用当前比较行之有效的相似度匹配算法,提出了能在云制造背景下适用的服务匹配算法,实验结果证实,该算法能基本满足云制造系统下的服务匹配需

求。

参考文献:

- [1] 李伯虎,张霖,王时龙,等. 云制造-面向服务的网络化制造新模式[J]. 计算机集成制造系统,2010,16(1):1-7.
- [2] 黄果,周竹荣,周亨. 基于领域本体的语义相似度计算研究[J]. 计算机工程与科学,2009(5):112-117.
- [3] 郝君甫,刘国华,唐军军,等. 基于本体的关系数据库关键词语义查询扩展方法[J]. 燕山大学学报,2010,34(3):231-235.
- [4] 李伯虎. 云制造-制造领域中的云计算[J]. 中国制造业信息化,2011(10):24-26.
- [5] OWL-S Coalition. OWL-S 1.0 Release [EB/OL]. 2004-08-17. <http://www.daml.org/services/owl-s/1.0>.
- [6] Paolucci M, Kawamura T, Payne T, et al. Semantic matching of web services capabilities[C]//Proceedings of the First International Semantic Web Conference. Sardinia, Italy: Springer-Verlag, 2002:333-347.
- [7] 彭晖,史忠植,邱莉榕,等. 基于本体概念相似度的语义 Web 服务匹配算法[J]. 计算机工程,2008,34(15):51-53.
- [8] 杨惠荣,刘珊珊,尹宝才,等. 基于语义距离的 Web 服务匹配算法[J]. 北京工业大学学报,2011,37(4):591-595.
- [9] 杨佳,张金广,杨龙,等. 基于本体概念集合的 Web 服务匹配[J]. 计算机技术与发展,2012,22(8):56-59.
- [10] Lin Y H, Bander Z A, Mclean D. An approach for measuring semantic similarity between words using multiple information sources[J]. IEEE Transactions on Knowledge and Data Engineering, 2003, 15(4):871-882.
- [11] 冯勇,方欣,徐红艳. 带有高效索引的语义 Web 服务 I/O 匹配优化方法[J]. 计算机应用,2011,31(3):677-679.
- [12] Sirin E, Parsia B. SPARQL-DL: SPARQL Query for OWL-DL [C]//Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions. Innsbruck, Austria; [s. n.], 2007.

(上接第 40 页)

- [5] 朱俊良,王茂芝,郭科. 基于形态预处理和标记提取的分水岭分割算法[J]. 信息技术,2010(9):17-20.
- [6] 卢中宁,强赞霞. 基于梯度修正和区域合并的分水岭分割算法[J]. 计算机工程与设计,2009,30(8):2075-2077.
- [7] 高丽,杨树元,夏杰,等. 基于标记的 Watershed 图像分割新算法[J]. 电子学报,2006,34(11):2018-2023.
- [8] 徐伟,王希常,郑志宽. 一种基于改进分水岭算法的图像分割方法[J]. 计算机技术与发展,2008,18(12):39-40.
- [9] 李然. 基于改进分水岭算法的图像分割[J]. 电脑知识与

技术,2011,7(16):3920-3922.

- [10] Solie P. Morphological Image Analysis Principles and Applications [M]. Berlin, Germany: Springer Verlag, 1999:123-140.
- [11] Solie P. Morphological image analysis applied to crop field mapping [J]. Image and Vision Computing, 2000, 18(13):1025-1032.
- [12] 余旺盛,侯志强,宋建军. 基于标记分水岭和区域合并的彩色图像分割[J]. 电子学报,2011,39(5):1007-1012.

云制造背景下的服务匹配算法

作者：[张金广](#)，[李锋刚](#)，[张磊](#)
作者单位：[合肥工业大学 管理学院, 安徽 合肥230009](#)
刊名：[计算机技术与发展](#)
英文刊名：[Computer Technology and Development](#)
年，卷(期)：2013(3)

本文链接：http://d.g.wanfangdata.com.cn/Periodical_wjtz201303013.aspx