

Research on Geospatial Service Discovery Based on GRUM and Clustering Relationship

Wu Du¹, Hao Tian², Kun Liu²

¹State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan Hubei

²School of Information Engineering, Hubei University of Economics, Wuhan Hubei
Email: haotian@whu.edu.cn

Received: Sep. 25th, 2017; accepted: Oct. 9th, 2017; published: Oct. 16th, 2017

Abstract

To solve the problem of the lack of semantic information processing and the low efficiency, a new method of geospatial service discovery is presented in this paper. Firstly, the lightweight geographic resources unified model GRUM is adopted to standardize service data. Then, the clustering relations of geospatial services are established based on ontology technology. Finally, the corresponding matching strategy and algorithm are designed. Experimental results show that the proposed approach can effectively improve the performance of the geospatial service discovery and perform the advantages of high speed and precision compared with other methods. It is proved to be an effective solution for geospatial service discovery.

Keywords

GRUM, Clustering Relationship, Geospatial Service, Discovery

基于GRUM与聚类的地理空间服务发现方法研究

杜武¹, 田浩², 刘坤²

¹武汉大学测绘遥感信息工程国家重点实验室, 湖北 武汉

²湖北经济学院信息工程学院, 湖北 武汉

Email: haotian@whu.edu.cn

收稿日期: 2017年9月25日; 录用日期: 2017年10月9日; 发布日期: 2017年10月16日

摘要

针对已有地理空间服务发现方法中语义信息处理手段不够有效, 算法效率低等问题, 提出了一种新的地理空间服务发现方法。首先采用了轻量级的地理资源统一模型GRUM来规范化服务数据, 其次基于本体技术建立了服务的聚类关系, 最后设计了相应的匹配策略和算法。实验结果表明, 本文提出的方法能有效提高地理空间服务发现的性能, 且较其它方法具有执行速度快和精度高的优点, 是一种有效的地理空间服务发现解决方案。

关键词

GRUM, 聚类关系, 地理空间服务, 发现

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

目前 Web 服务的空间数据服务规范主要有 Web 地图服务(Web Map Service, WMS)、Web 要素服务(Web Feature Service, WFS)、Web 覆盖服务(Web Coverage Service, WCS)等。虽然它们在服务的基本信息和操作接口信息等参数描述结构上有相同之处, 但由于使不同的元数据来描述具体的服务资源, 使得空间信息目录服务中的空间信息服务资源检索和共享变得困难。作为解决办法, 网络目录服务(Catalog Service for the Web, CSW)的元数据信息模型和地理标记语言(Geography Markup Language, GML)得到了广泛的应用, 但它们只是在语法和结构层面上实现了这些规范间的互操作, 并没有解决语义异构等深层次的问题。

为实现空间信息服务资源的有效检索和充分利用, 许多研究开展了不同的尝试。为缓解服务发现受自然语言模糊性的限制, 文献[1]提出了一种新的空间信息服务发现方法, 它使用描述地理空间操作的本体来创建对需求和服务功能的描述。这些描述之间的匹配是基于函数子类型的, 文中还研究了如何将该方法集成到现有的空间数据基础结构体系结构中, 并给出了一个典型的实现。文献[2]针对 WFS 资源的检索提出了一种新的方法。该方法将发现请求和 Web 处理服务都注释为逻辑编程语言中的连接查询, 而发现过程则是基于这些描述之间的逻辑编程查询控制。该方法将输入和输出之间的关系形式化, 使之能够更精确地捕获 WFS 的功能并能使用逻辑编程查询在发现过程中进行有效的推理。文献[3]在北极问题研究中利用基于知识的方法和空间 Web 门户技术, 提出了一个基于分布式 Web 目录和动态 Internet 的高效服务发现混合方法, 建立了一个面向科学数据与服务之间潜在语义关系建模的领域知识库, 设计了一个智能逻辑推理机制来实现服务的自动选择与链接。文献[4]构建一个面向服务的体系结构, 实现分布式地理空间数据的广泛共享和无缝集成, 能将城市经济建模和有效的空间规划等复杂的问题分解为原子空间分析任务, 生成一个 Web 服务链来解决这些复杂问题, 可捕获和表示地理空间数据的来源, 以追踪其在建模任务中的流程。文献[5]提出了一个地理空间知识基础设施的信息模型, 在该模型中使用一系列不同类型的本体来表示域的知识概念, 资源要素和 Web 服务等对象的语义细节。通过增强语义信息来进行地理空间资源的智能搜索, 并支持知识推理。

上述研究在不同程度上解决了空间信息服务资源的检索问题，但由于语义异构的存在，仍存在着一些不足，如增强地理资源层面的语义信息的手段还不够有效，算法效率低等。因此，本文在前期研究的基础上，基于本体技术，采用轻量级的地理资源统一模型，提出一种新的地理空间服务发现方法并通过实验验证了其性能。

2. 背景分析

2.1. GRUM

GRUM (Geographic Resource Uniform Model)是文献[6]中提出的一种面向空间信息服务检索的地理资源统一模型。它是一套关于空间信息服务元数据的概念模型，遵循简单化、轻量级的原则，能够为空间信息服务的描述提供基础的元模型和统一表示。GRUM 中顶层概念是抽象地理资源(Geo Resource)，它派生出空间信息服务(Geo Service)、服务绑定(Service Binding)和服务参数(Service Parameter)三个子概念。除此之外，GRUM 还设计了数据类型 Data Type 和服务类别 Service Taxonomy 两个待扩展的部分。这样的话，GRUM 模型的基础部分主要用来对空间信息服务原有元数据进行移植和映射，扩展部分则用来增强数据类型和服务功能的语义信息。其本体模型如图 1 所示。

WFS、WCS 及 WPS 等不同类型的空间信息服务核心元数据描述都可以映射到 GRUM 模型上，因此，本文选择 GRUM 来规范化不同类型的空间信息服务。

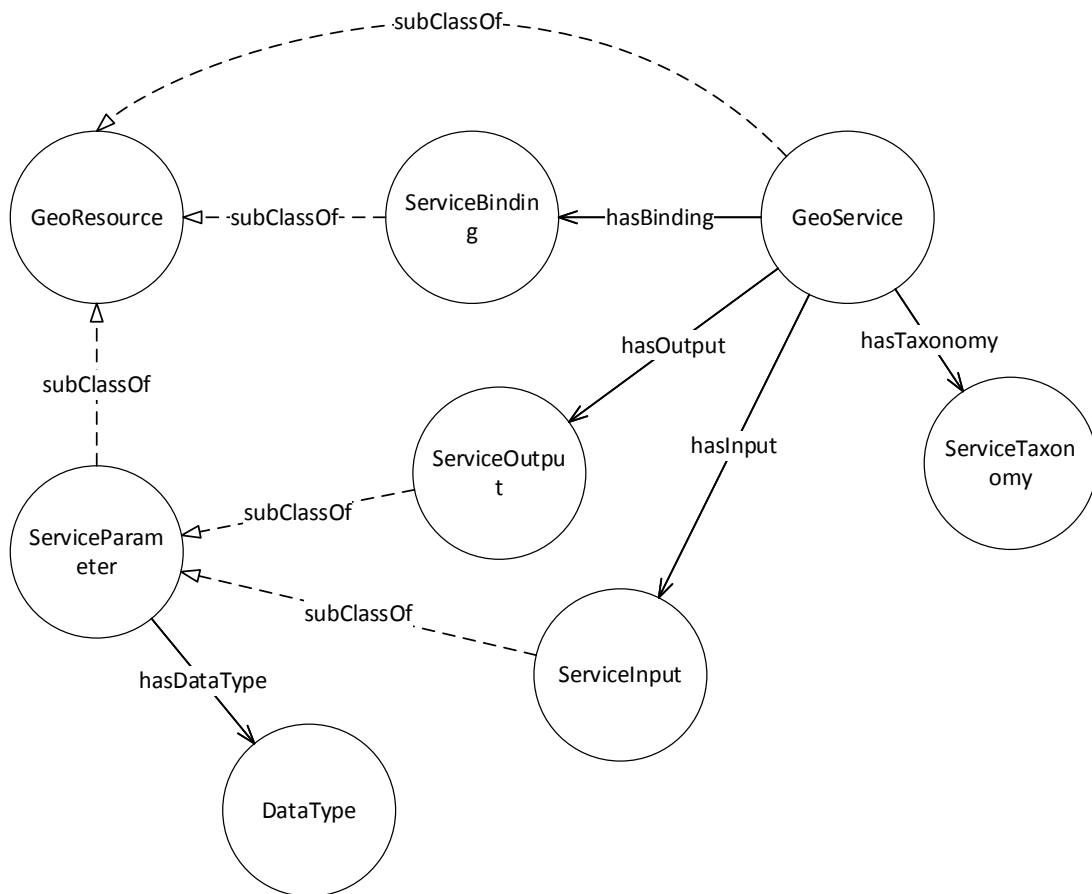


Figure 1. Core classes and properties of the GRUM ontology

图 1. GRUM 本体的核心类和属性

2.2. 聚类关系

文献[7]提出,在众多可用的 Web 服务中,会存在着许多相似的服务,这些服务根据某些参数指标可以形成一个聚类关系。该研究基于本体技术将服务之间的功能属性和服务名的综合相似度作为服务聚类的依据。

首先,一个服务 s 可以表述为一个二元组:

$$s = (sn, O) \quad (1)$$

$$O = (IN, OUT) \quad (2)$$

sn 是服务 s 的名称; O 是 s 的操作集合,用 $O.IN$ 、 $O.OUT$ 分别表示该操作的输入集合以及操作的输出集合。

那么对于服务 s_i 与 s_j 而言,它们之间的综合相似度计算公式可以表示为:

$$SSim_{ij} = \omega_1 \times NSim(S_i.sn, S_j.sn) + \omega_2 \times FSim(S_i.O, S_j.O) \quad (3)$$

$$0 < \omega_1, \omega_2 < 1, \omega_1 + \omega_2 = 1$$

函数 $NSim(\cdot)$ 用来计算服务的名称相似度, $FSim(\cdot)$ 用来计算服务的操作相似度。两者均采用文献[7]中提出的定义和计算方法。

因此,服务的聚类关系可以定义为所有关于服务及它们之间综合相似度的集合:

$$CR_i = \left\{ (s_i, s_j, SSim_{ij}) \mid s_i \in S, s_j \in S, SSim_{ij} > SSim_0 \right\} \quad (4)$$

3. 方法设计

3.1. 服务匹配策略

基于文献[7]中的方法,将服务请求 R 也表示为一个二元组,如下所示:

$$R = (C, O) \quad (5)$$

C 是一个由本体概念组成的概念集,用来归纳和表达用户对目标服务的功能需求;而 O 的形式与之前定义的服务 s 的操作集合相似,用来在语义层面上描述用户对目标服务输入和输出的要求。

服务 s 与服务请求 R 之间的匹配相似度可计算为:

$$Sim(S, R) = \alpha \times FSim(S.O, R.O) + \beta \times NSim(S.sn, R.C) \quad (6)$$

$$0 < \alpha, \beta < 1, \alpha + \beta = 1$$

3.2. 算法设计

在以上基础上,本文设计一个基于 GRUM 与聚类的地理空间资源发现方法(简称 GDGCR),其算法内容可以描述为:

算法: GDGCR

初始化: 基于 GRUM 的空间信息服务核心元数据映射,设定阈值 Sim_0 、 $SSim_0$

输入: 服务请求 R

输出: 候选服务列表 L_o

- 1) initiate CR_i
- 2) set $L_n = \phi$ and $L_o = \phi$

- 3) **for** each candidate s_i **do**
- 4) **if** $s_i \notin L_n \wedge s_i \notin L_o$ **do**
- 5) calculate $Sim(s_i, R)$; //进行服务匹配
- 6) **if** $Sim(s_i, R) \geq Sim_0$ **do** //若相似度达到要求
- 7) add s_i to L_o
- 8) **else**
- 9) add s_i to L_n
- 10) **for** each $s_j \in CR_i \wedge s_j \notin L_n$ **do**//聚类关系中的其他服务
- 11) add s_j to L_n
- 12) **end for**
- 13) **end if**
- 14) **end if**
- 15) **end for**
- 16) remain top k services in L_o based on $Sim(s_i, R)$ //基于相似度升序排序并保留前 k 个
- 17) **return** L_o

4. 实验及分析

4.1. 实验准备

为验证本文提出的方法，设计以下实验来检验其性能。实验环境的搭建参照文献[6]，采用 GI-cat 作为空间信息目录服务，空间数据服务 400 个，主要来源于 52 north [8]中的 WPS 服务并采用 GeoServer 来管理。将这些数据服务随机划分为两部分，训练集占比 70%，测试集占比 30%；同时生成服务请求 15 个。参与比较的对象方法选择了基于 CSW 的原始查询方法(Original Query Method, OQM)以及 Ranking SVM [6]和 LanbdaMART [6]。比较指标 3 个：平均完成时间(Average Completion Time, ACT)、平均精度均值(Mean Average Precision, MAP)和平均折扣增益值(Normalized Discounted Cumulative Gain, NDCG)，后两者采用文献[6]中的定义，具体如下：

$$MAP = AP(\pi, l) = \frac{\sum_{k=1}^m I_{\{l_{\pi^{-1}(k)}=1\}}}{m_1} \quad (7)$$

其中 m 是本次查询产生时所有可用服务的数量， m_1 是指与查询相关的服务数量。

$$NDCG = \frac{1}{Z_k} \sum_{j=1}^k G\left(l_{\pi^{-1}(j)}\right) \eta(j) \quad (8)$$

$$G(\zeta) = 2^\zeta - 1$$

$$\eta(j) = 1/\log_2(j+1)$$

4.2. 结果分析

图 2 显示了这四种方法的平均精度均值指标对比情况。从图可以看出，GDGCR、Ranking SVM 和 LanbdaMART 的 MAP 值都明显高于 OQM，说明这三种方法检索的服务都与服务请求有较高的相关度，在这当中，LanbdaMART 的 MAP 为 0.486，Ranking SVM 的 MAP 为 0.459，本文提出的 GDGCR 方法 MAP 值为 0.479，位居第二。

NDCG 指标可以衡量这几种算法优先推荐和排序相关程度高的空间服务的能力。从图 3 可以看出, OQM 的 NDCG 指标始终是这四种方法中最弱的; Ranking SVM 表现好于 OQM 但不及另外两者; GDGCR 在排序 1 的 NDCG 指标值高于 LambdaMART, 在其他排序位置略低于 LambdaMART, 两者整体表现相当。

平均完成时间 ACT 用来比较算法的执行效率。图 4 的结果表明, OQM 的执行效率是这四者中最高的, 这是因为传统基于 CSW 查询的方法没有额外计算代价, 仅基于服务请求做直接简单的检索。为取得高质量的检索结果, GDGCR、Ranking SVM 和 LambdaMART 的执行时间都大幅增加。从图中可以看出, GDGCR 的执行效率是这三者中最高的, Ranking SVM 次之, LambdaMART 的效率最低。

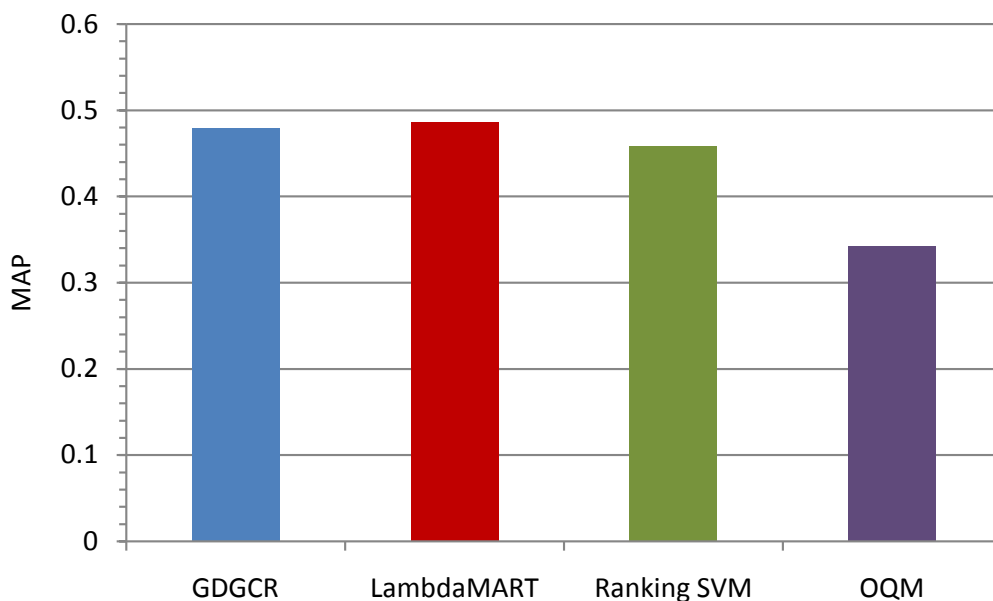


Figure 2. Comparison of MAP

图 2. MAP 对比结果图

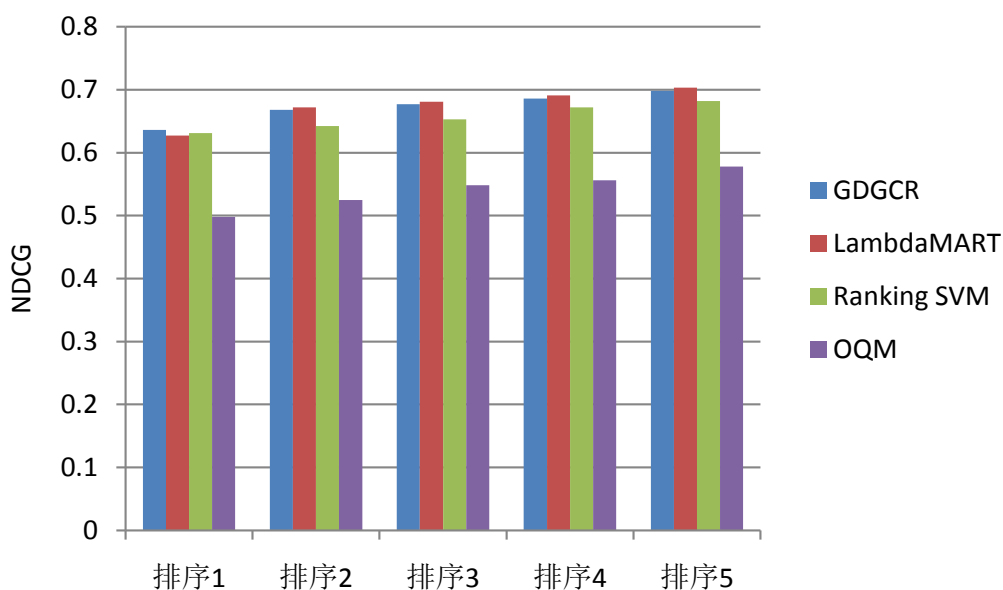


Figure 3. Comparison of NDCG

图 3. NDCG 对比结果图

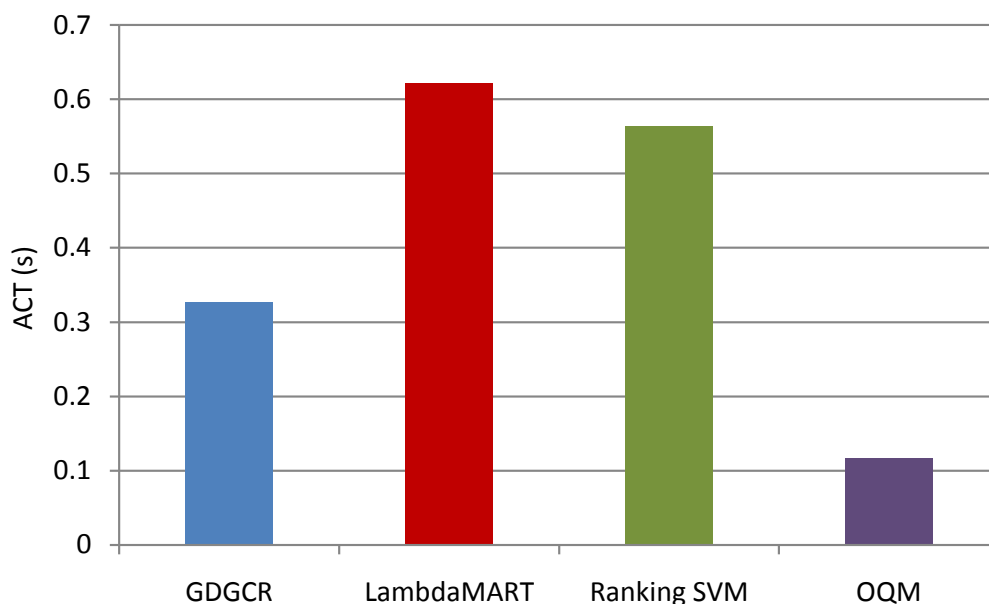


Figure 4. Comparison of ACT
图 4. ACT 对比结果图

综合以上三个比较结果可以看出，本文提出的 GDGCR 方法能够实现满意的平均精度均值和平均折扣增益值，并有着较高的执行效率，体现出了良好的算法性能，能有效解决存在的问题，达到了预期的设计目标。

5. 总结

为增强地理资源层面的语义信息，改善地理空间服务发现的结果，提升发现的效率，本文提出了一种新的地理空间服务发现方法。首先采用了轻量级的地理资源统一模型 GRUM 来规范化服务数据，其次基于本体技术建立服务的聚类关系，最后设计了相应的匹配策略和算法。实验结果表明，本文提出的方法能有效提高地理空间服务发现的性能，且较其它方法具有执行速度快和精度高的优点，表现出了良好的优越性。

致 谢

本文受到湖北省自然科学基金项目(2016CKB714)资助。

参考文献 (References)

- [1] Lutz, M. (2005) Ontology-Based Service Discovery in Spatial Data Infrastructures. *Proceedings of the 2005 Workshop on Geographic Information Retrieval*, Bremen, 4 November 2005, 45-54. <https://doi.org/10.1145/1096985.1096997>
- [2] Fitzner, D., Hoffmann, J. and Klien, E. (2011) Functional Description of Geoprocessing Services as Conjunctive Data-log Queries. *GeoInformatica*, **15**, 191-221. <https://doi.org/10.1007/s10707-009-0093-4>
- [3] Li, W., Yang, C., Nebert, D., Raskin, R., Houser, P. and Wu, H. (2011) Semantic-Based Web Service Discovery and Chaining for Building an Arctic Spatial Data Infrastructure. *Computers & Geosciences*, **37**, 1752-1762. <https://doi.org/10.1016/j.cageo.2011.06.024>
- [4] Li, W., Li, L., Goodchild, M.F. and Anselin, L. (2013) A Geospatial Cyberinfrastructure for Urban Economic Analysis and Spatial Decision-Making. *ISPRS International Journal of Geo-Information*, **2**, 413-431. <https://doi.org/10.3390/ijgi2020413>
- [5] Stock, K., Stojanovic, T., Reitsma, F., Ou, Y., Bishr, M., Ortmann, J. and Robertson, A. (2012) To Ontologies or Not to Ontologies: An Information Model for a Geospatial Knowledge Infrastructure. *Computers & Geosciences*, **45**,

98-108. <https://doi.org/10.1016/j.cageo.2011.10.021>

- [6] 杜武. 基于排序学习的空间信息服务检索关键技术研究[D]: [博士学位论文]. 武汉: 武汉大学, 2017.
- [7] 田浩. 以用户为中心的 Web 服务发现方法及其在金融服务中的应用研究[D]: [博士学位论文]. 武汉: 武汉大学, 2014.
- [8] 52 North (2017). <http://52north.org/resources>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2286, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: sea@hanspub.org