

面向个性化推荐的空间信息服务三层模型研究*

李方林, 邬群勇

(福州大学 空间数据挖掘与信息共享教育部重点实验室, 福建 福州 350002)

摘要: 分析了当前空间信息服务在个性化推荐方面所存在的问题, 充分考虑用户行为对空间信息服务个性化推荐的约束, 匹配过程中的“需求”与空间信息服务“操作”的映射关系, 提出针对不同用户需求的三层模型。该模型将用户的历史操作抽象为二元组, 以此二元组作为约束条件, 通过语义的相似度匹配, 有效且准确地进行空间信息服务的个性化推荐。

关键词: 空间信息服务; 三层模型; 个性化推荐

中图分类号: P208

文献标识码: A

文章编号: 1674-7720(2013)12-0077-04

Research on three layers model oriented to the spatial information webservices personalized recommendation

Li Fanglin, Wu Qunyong

(Key Lab of Spatial Data Mining & Information Sharing, Ministry of Education, Fuzhou University, Fuzhou 350002, China)

Abstract: This paper analyses the current space information service existing problems of personalized recommendation, gives full consideration to the constraint user's historic behaviors on to spatial information services personalized recommendation and mapping relationship between “need” and “operate” of spatial information service during the matching process, then puts forward a space information service three layers organization model conforms to user's different needs. The model abstracts the users' historical operation as two-tuples and can effectively and rapidly process spatial information personalized recommendation with two-tuples as constraint condition through the semantic similarity matching.

Key words: spatial information service; three layers organization model; personalized recommendation

人类社会进入信息时代, 服务提供商提供的规范化、标准化服务并不能满足用户的差异性需求。个性化服务通过收集和分析用户信息来学习用户的兴趣和行为, 从而实现主动推荐的目的^[1], 成为目前研究的热点。

国内外研究人员对空间信息服务匹配^[2-4]、空间信息服务组合^[5-7]、空间信息服务质量评价^[8]以及个性化系统中用户建模^[9-12]等方面进行了大量研究。这些研究主要关注于如何对空间信息服务进行语义表达、空间信息服务的有效匹配、服务组合的有序展开以及怎样通过各种质量属性参数来衡量服务的好坏等。然而它们却存在着一个共同的问题, 即都是面向服务本身, 忽略了用户需求, 分离用户与服务两者之间的映射关系, 使得每个服务的匹配没有结合其用户的需求而独立进行或是仅仅凭借用户所输入的几个服务参数来检索服务, 其结果

必然导致最后所得到的服务无法满足用户的复杂需求, 迫使用户做出二次或多次服务检索, 不利于个性化服务的有效开展。

本文提出一种面向用户复杂需求、针对单个空间信息服务的三层组织模型, 在空间信息服务个性化推荐过程中, 根据用户的历史行为, 通过源服务拆分以及用户建模等步骤, 准确地将具体的用户行为映射到抽象的空间信息服务上, 满足了动态变化的用户人群对空间信息服务的不同需求。

1 几个定义

为了建立起针对用户约束的个性化推荐三层模型, 首先给出以下几个概念:

定义 1: 空间信息源服务: 服务提供者提供的单个具体空间信息服务称为源服务。一个源服务必须包含以下变量, 即:

$$So = \text{Name} + \text{URL} + \text{Operation} + \text{QoS} + \text{DomainFeature}$$

* 基金项目: 福建省自然科学基金项目(2012J01168)

技术与方法 Technique and Method

其中, S_o 表示某一目标服务; $Name$ 为服务的名称; URL 表示服务的访问地址; $Operation$ 为该服务所支持的各种操作; QoS 为该服务的质量, 包含服务响应时间、价格、稳定度、可用性等参数; $DomainFeature$ 为空间信息服务的领域特性, 如空间数据服务应包含 $Data\ format$ 、 $Spatial\ resolution$ 、 $Spatial\ reference\ system$ 、 $Extent$, 描述内容参考了 ISO19115 元数据标准。

定义 2: 空间信息原子服务: 只封装了一个操作的空间服务为原子服务。原子服务使用一个三元组来表示, 即:

$$Sa = \langle f, Op, IO \rangle$$

其中, f 表示服务 Sa 能完成的功能, 它以语句形式描述该原子服务的功能信息; Op 表示服务 Sa 中封装的单一操作; IO 表示操作 Op 中输入输出接口参数, 其中从参数名 (param Value)、参数描述 (description)、参数类型 (type)、单位 (unit) 和比例系数 (scale) 5 个方面来描述。

2 空间信息服务三层组织模型的构建

传统的三层服务模型^[5-6], 主要针对原子服务如何组合以及服务链的有序展开等问题, 其思想是将整条服务链拆分为多个独立的原子服务, 研究这些原子服务之间的接口关系, 便于前置服务输出接口所得到的参数与后置服务的输入接口所需要的参数相匹配, 从而使服务组合有效开展。而本文中的三层模型, 针对的是复杂多变的用户人群以及用户动态的需求, 从用户需求的角度来研究如何组织服务, 从而实现空间信息领域的个性化推荐。

用户对空间信息服务的需求往往不是单一的, 而是由多个单纯需求构成的复杂体, 反映在用户操作或行为上就是多个操作或行为的复杂集合体, 反映在服务上就是服务的各类参数和所包含的原子服务。空间信息服务的个性化推荐, 其实质是服务本身所含有的原子服务和用户模型耦合作用的结果。

根据以上分析, 从功能角度组织服务, 将源服务(对象服务)拆分为原子服务, 再将用户需求拆分为单个行为或操作, 通过原子服务与用户单个操作或行为的映射关系, 实现一个服务-用户的分级模型, 便于源服务与多用户的不同需求和单个用户的多个需求进行匹配; 实现针对用户历史行为或操作集合所做出的服务个性化推荐。具体模型如图 1 所示, 即源服务层-原子服务层-用户层。三层模型实现两个映射关系: 源服务与原子服务之间的映射, 原子服务与用户之间的映射。

2.1 源服务到原子服务的映射

源服务拆分为原子服务, 是要将源服务的功能进行分解, 而这些功能实质上是由原子服务来实现的, 例如在 ArcWeb 服务中位置查找 Web 服务(The Place Finder Sample Web Service), 客户端调用某一源服务中的 Find-Place 原子服务, 传递位置名称, 然后得到这个位置的确切定位。过程如下: FindPlace(placeName:string, placeFinder-Options:placeFinder Options):LocationInfo。反之, 如果一个

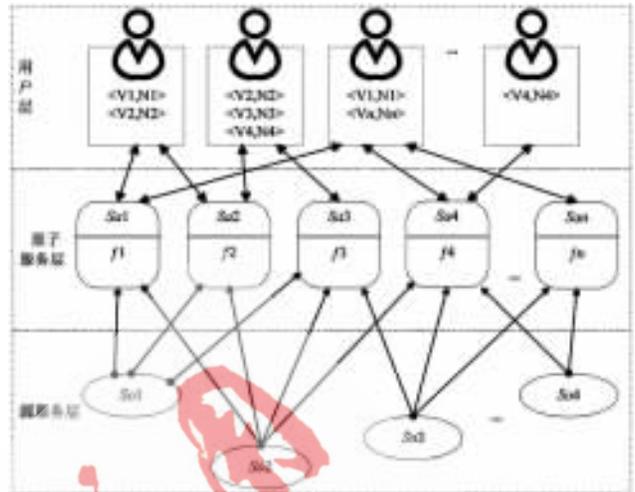


图 1 空间信息服务的个性化推荐三层模型

源服务中没有 FindPlace 原子服务, 则该服务就不能提供这一功能。

在源服务拆分为原子服务的过程中, 并不是真正地对某个具体的源服务进行拆分, 而是为了便于与用户的单个需求进行匹配所做的工作。服务提供者发布源服务, 通过对其 WSDL 进行解析, 建立起由源服务-原子服务的映射关系, 原子服务与源服务的关系可以表示为:

$$S_o = Sa_{a0} + Sa_{a1} + \dots + Sa_{a_n} \quad (1)$$

$$S_o = Sa_{a0} + Sa_{a1} + \dots + Sa_{a_m} \quad (2)$$

其中 S_o 为某一源服务, 可以表示为 Sa_{a0} 、 Sa_{a1} 、 \dots 、 Sa_{a_n} 等原子服务的总和; 同样源服务 S_o 也可以表示为 Sa_{a0} 、 Sa_{a1} 、 \dots 、 Sa_{a_m} 等原子服务的总和。式(1)、式(2)中的原子服务个数 m 与 n 可以相等或不等, 原子服务项可以相同或不同。

2.2 用户分析与建模

空间信息 Web 服务是在 Web 服务技术和标准基础之上实现的地理空间信息网上在线服务。主要的应用客户包括发现客户、地图浏览客户、影像使用客户、增值客户和传感器客户等。这些人群大致可以分为 3 类: (1) 面向程序的开发人员: 指对服务进行增值服务开发的用户。这类用户更关注于服务质量(QoS); (2) 面向空间数据的用户: 指通过 GIS 相关的软件产品使用空间信息服务所提供数据的用户。这类用户关注于服务所含有的空间实体的属性(如位置信息、几何性质、拓扑关系等); (3) 空间信息服务的终极用户: 指只使用空间信息服务, 而不对服务进行二次处理的用户, 也即普通大众。这类用户可以通过查询空间信息服务解决方案得到最新的空间信息报告、地图和感兴趣的应用程序。

因为前两种用户和第三种用户对服务的关注程度存在递进关系, 即开发人员 > 数据用户 > 大众用户, 因此针对第三种用户进行分析与建模可对构建另外两种用户模型提供帮助。以下对第三类用户进行分析, 这类用户在使用空间信息服务时, 其兴趣将反应在其行为上,

技术与方法 Technique and Method

具体表现为在示例地图上(示例地图是能提供所有操作的服务,且该服务不参与服务推荐)的操作。对用户所进行的操作进行采集和预处理,根据用户记录拆分其操作集合为离散的操作,每一个操作采用一个二元组(动词,名词)来表示,其中操作的本身用动词表示,操作的对象用名词表示,用户二元组简称为 $U(V, N)$,如检索一个位置或查询某一地点在规定时间内可达性,分别表述为 (search, place)和 (compute, extent)。用户模型通常由多个二元组组成,表达了用户的复杂需求。基于以上分析,三类用户模型中包含的具体内容如表 1 所示。

表 1 用户分类和模型具体内容分类

用户分类	模型内容
大众用户	多个二元组 $\langle V, N \rangle$
数据用户	多个二元组 $\langle V, N \rangle$ + 服务的 DomainFeature
开发人员	多个二元组 $\langle V, N \rangle$ + 服务的 DomainFeature + QoS

2.3 原子服务与用户兴趣的双向映射

原子服务与用户操作的匹配,主要通过挑选出 f 中的动_名关键词,与用户二元组(动_名词)进行语义匹配。本文将动_名词匹配等级分为 exact、subsume、sibling 和 irrelevant,其中 exact 表示完全匹配关系,subsume 表示包含关系,sibling 表示类似关系,irrelevant 表示不相关关系,在匹配过程中,优先级为 exact > subsume > sibling > irrelevant。

为定量表示原子服务与用户二元组的相关关系,分别对动、名词的 4 个等级赋予权值 ω ,即 exact 级, $\omega=1$; subsume 级, $\omega=0.6$; sibling 级, $\omega=0.4$; irrelevant 级, $\omega=0$ 。其中定义 ω 的标准应符合:(1)人们对原子服务匹配成功的语义相似度要求,相似度越高,匹配越准确,一般要求相似度要大于 0.4;(2)便于区分 4 个等级的差异性。定义原子服务与用户二元组的相似度 Φ 及动名词在二元组中的权重 ω ,有如下关系:

$$\Phi = \omega_V \times \Phi_V + \omega_N \times \Phi_N \quad (3)$$

其中 Φ_V 表示原子服务与用户二元组的动词相似度, Φ_N 表示原子服务与用户二元组的名词相似度, ω_V, ω_N 分别表示用户二元组的动词与名词在二元组中的权值,有约束条件 $\omega_V + \omega_N = 1$ 。

2.4 个性化推荐过程

一般的个性化推荐是根据用户所发出的明确请求,向用户推荐其感兴趣的信息和商品的过程。而空间信息服务的个性化推荐是根据用户在示例地图中的历史操作或使用网络服务的历史情况,向用户推荐其感兴趣的网络服务过程。本文提出的模型在个性化推荐中的具体应用如图 2 所示。在有效时间内,通过对用户历史行为的记录展开筛选服务,进而对该服务进行推荐。

3 模拟应用实验

3.1 实验准备

设计源服务组织模型并下载 Esri 以及本地的服务



图 2 模型在个性化推荐中的应用

等 200 多个服务 WSDL,组成源服务数据;解析 WSDL 得到源服务所包含的操作组成原子服务,借鉴 ArcGIS Soap 标准对原子服务功能 f 进行描述。个性化推荐通过隐性方式收集用户在页面上的行为,抽象出用户二元组。

3.2 实验过程

本文以一个仿真实验为例进行分析。对用户的历史操作或行为进行采集,论证能否通过三层模型筛选出满足用户需求的源服务。

过程简述为:首先采集用户的历史行为,并对这些用户行为进行抽象,组成代表用户复杂兴趣的二元组集合,计算单个二元组与原子服务的相似度,取相似度最大的原子服务。因为用户操作不仅是一个,通常情况下是多个,以每个二元组所对应的原子服务为索引,组成索引集,检索出源服务并进行推荐。

将实验源服务拆分为原子服务,如表 2 所示。

表 2 源服务与原子服务关系表

编号	源服务	包含的原子服务
1	Esri_StreetMap_World_2D	Find、ToMapPoints、GetLayerTile、Identify
2	SanFrancisco	Find、Solve、ToMapPoints、GetLayerTile、Identify、GetNALayerNames
3	Esri_Imagery_Word_2D	Find、ToMapPoints、GetLayerTile、Identify
4	Asia_China	Find、ToMapPoints、GetLayerTile、Identify
5	Portland	Find、Solve、ToMapPoints、GetLayerTile、Identify、GetNALayerNames

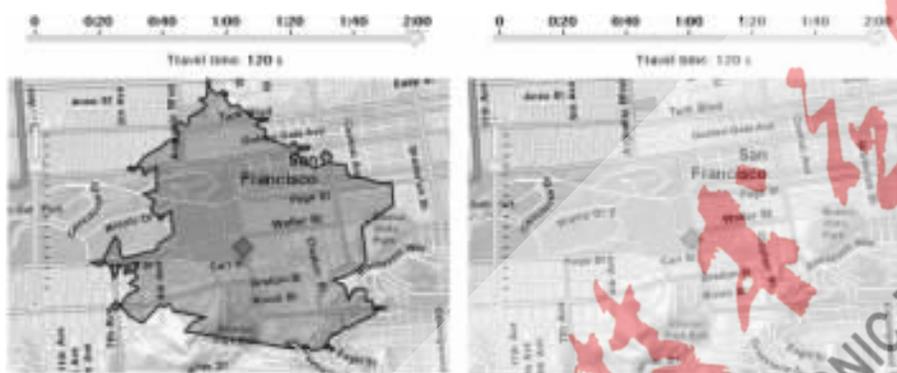
通过 javascript 脚本以及 session 的建立与销毁来记录用户的行为日志,发现在有效的时间段,用户在示例地图上进行了如下操作:(1)用鼠标在地图上进行定位;(2)分析该位置在 2 min 内所能到达的区域,构建其模型为:用户= $\langle \text{Locate}, \text{Point} \rangle + \langle \text{Analyse}, \text{Extent} \rangle$ 。其中,借助 wordnet 工具包将原子服务 f 语句中的动词和名词挑选出来作为关键词匹配计算,采用 Lin 算法计算原子服务与用户二元组的相似度 (Φ_V, Φ_N) ,设定动词与名词在二元组中的权重都为 0.5,即 $\omega_V = \omega_N = 0.5$ 。用户与原子服务的相似度计算结果如表 3 所示。

抽取出分别对应用户两个二元组的最大相似度原子服务(表 3 中原子服务 3 和 2),将其作为索引集,检索源服务,得到源服务结果(表 2 中服务 2 和 5,服务效果

表 3 用户与原子服务的相似度计算结果

编号	原子服务	原子服务描述	相似度(Φ)	
			<Locate, Point>	<Analyse, Region>
1	Find	在地图服务中以给定的查询字符串,按照数据格式,查询一个或多个图层的属性	0.320 65	0.268 65
2	Solve	基于输入的解析参数执行网络分析	0.161 1	0.550 5
3	ToMapPoints	将屏幕坐标转化为地图坐标	0.5	0.452 1
4	GetLayerTile	以特定的位置,从图层缓存中返还影像名称	0.430 8	0.166 7
5	Identify	基于空间索引返还地图特征以及属性值	0.499 35	0.057 35
6	GetNALayerNames	以特定分析类型的网络分析图层序列	0.453 9	0.176 8

如图 3(a)所示),推荐给用户;对于噪声源服务(表 2 中服务 3 等,服务效果如图 3(b)所示),拒绝推荐。



(a)满足用户需求的 SanFrancisco 服务 (b)不满足用户需求的 Esri_StreetMap_World_2D 服务

图 3 服务效果图

3.3 实验分析与结论

通过模拟实验,用户模型代表了用户的两方面需求:定位和分析通达性。根据表 3 所示用户与原子服务的相似度计算结果,用户的这两方面需求由原子服务 ToMapPoints 和 Solve(表 3 中原子服务 3 和 2)实现。根据源服务与原子服务的关系(表 2)可知,包含这两个原子服务的源服务(表 2 中源服务 2 和 5)能提供用户所要求的功能,符合推荐条件。

由以上分析得出以下结论:传统空间信息服务模型以服务本身为考虑对象,忽略了对用户信息的挖掘,检索出的服务其功能并不一定满足用户的需求,迫使用户做出二次甚至多次的检索,影响服务质量;本文模型有效地克服了这一缺点,在对用户历史信息进行挖掘后,经过原子服务这一桥梁,把用户需求与服务功能有机地结合在一起,所筛选出的源服务也必然具有与用户需求相一致的功能。

传统的空间信息服务组织模式已难以满足广大不同用户的需要,个性化在空间信息服务中的应用是当前研究的一大难点。本文尝试以用户为研究对象,提出面向用户需求的三层组织模型,该模型充分挖掘了用户信息,将用户需求与服务功能有机地联系起来,为实现空间信息服务领域的个性化推荐提供一种更为便利的组织形式,且实例证明该三层模型能有效且迅速地对用户行为作出响应,从众多的空间服务中筛选出最满足用户需要的服务,为未来空间信息服务领域的智能化和个性

化研究提供新的思路。

参考文献

- [1] 曾春,邢春晓,周立柱.个性化服务技术综述[J].软件学报,2002,13(10):1952-1961.
- [2] 郭群勇,郑孝苗,康凌骏.语义地理信息服务的三级匹配发现算法[J].厦门大学学报(自然科学版),2012,51(2):53-57.
- [3] 王强,王家耀,姜艳媛,等.基于 P2P 和本体的空间信息服务发布与发现[J].测绘科学技术学报,2009,26(6):68-71.
- [4] 王强,王家耀,姜艳媛.本体支持的智能化空间信息服务发现[J].信息工程大学学报,2010,11(2):46-50.
- [5] 高岩,那俊,张斌,等.支持动态服务组合的 Web 服务三层组织模型[J].小型微型计算机系统,2006,27(10):89-92.
- [6] 张永胜,徐丽丽,齐峰,等.基于三层组织模型的一种 Web 服务组合策略[J].计算机工程与应用,2008,44(22):89-91.
- [7] 罗安,王艳东,龚健雅.顾及上下文的语义匹配方法[J].武汉大学学报(信息科学版),2011,36(3):368-372.
- [8] 章汉武,龚俊,吴华意.地理空间信息服务质量评价的概念与方法[J].测绘科学,2012,37(1):161-164.
- [9] 裴仰军.个性化服务中用户兴趣模型的研究[D].重庆:重庆大学,2005.
- [10] 唐灿,朱征宇.基于模糊兴趣模型的个性化推荐算法[J].计算机工程与应用,2006,42(9):170-172.
- [11] 韩旭.个性化推荐系统用户兴趣建模方式的研究[J].数字技术与应用,2010(11):46,48.
- [12] 牛亚真,祝忠明.个性化服务中跨系统用户建模方法研究综述[J].现代图书情报技术,2012,28(5):5-10.

(收稿日期:2013-01-21)

作者简介:

李方林,男,1985 年生,硕士研究生,主要研究方向:地理信息,网络共享与服务技术,地理信息系统软件设计与开发。