

一种基于多态关联挖掘的位置服务优化查询方法

蔡朝晖^{1,2} 张健沛² 杨静²

(哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001)¹

(大庆师范学院计算机科学与信息技术学院 大庆 163712)²

摘要 高效可靠的位置服务查询方法是位置服务广泛应用的关键。传统的位置服务查询方法采用近邻查询的方法,在对近邻的分割时往往面临稳定性差的问题,无法实现高效查询。提出一种基于多态关联挖掘的位置服务优化查询方法,该方法根据查询标准,将数据分为多个状态,然后对多个状态进行关联挖掘,提取数据的深层次特征,根据深层次特征,对位置服务的查询方法进行优化查询来提高系统查询的效率和可靠性。采用查全率和查准率作为衡量标准,对200组数据进行了实际分析,结果显示,基于多态关联挖掘的位置服务查询方法使查全率和查准率分别提高9.5%和8.8%,研究成果在位置服务和信息挖掘中具有很好的应用价值。

关键词 多态性,关联挖掘,位置服务,优化查询

中图分类号 TP392 **文献标识码** A

Optimized Query Method of Location-based Service with Polymorphic Association Mining

CAI Zhao-hui^{1,2} ZHANG Jian-pei² YANG Jing²

(College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)¹

(College of Computer Science and Information Technology, Daqing Normal University, Daqing 163712, China)²

Abstract High efficient and reliable location-based service (LBS) is key for the wide use of LBS. In traditional method, the neighbor query of location-based services query is used, and the neighborhood stability is often faced with the problem of poor ability, so it is unable to achieve high efficient query. An optimized query method of location-based service with polymorphic association mining was proposed. According to the search criteria, the data is divided into more than one state, then the multiple state association mining is carried out, and the depth features of data were extracted, and according to the deep-seated characteristics of the location services available method, the efficiency and reliability of the system queries are improved. An actual analysis was carried out using a team of 200 groups of data, and the result shows that with polymorphic association mining method, the recall and precision rate are increased by 9.5% and 8.8%, so it has a good application value in LBS and information mining.

Keywords Polymorphism, Association mining, Location-based service, Optimized query

1 引言

近年来,随着先进计算机技术、现代化网络技术、无线通信技术和智能传感器技术的迅速发展,采用无线通信技术和空间定位技术实现位置信息服务(Location Based Services, LBS)逐渐发展起来。在位置信息服务过程中,用户通过向服务的提供方提供自身当前的位置信息,就可以获取相应的服务。在位置服务的过程中,一个很重要的问题是如何提高位置服务信息的高效查询。高效的位置服务信息查询可以给系统带来许多益处:一是高效的优化查询方法可以及时地将用户需要查询的信息反馈给用户,使得用户看起来是专享服务的,这对于用户的信息实时交互非常有好处;二是优化查询方

法可以提高用户对系统信息查询的查全率和查准率,作为衡量系统查询结果的两个最重要的指标,其对于用户数据的响应评判是非常重要的。

近年来,对位置信息服务和位置隐私保护等方面的研究逐渐兴起。胡磊、王佳俊等人提出了一种基于坐标和的保护位置隐私近邻查询方法,在查询时,客户端向LBS服务提供方发送其当前位置的二维坐标之和,通过加密实现对查询发起用户位置信息的保护^[1]。同样,在服务器处理端,LBS服务器通过设置基于坐标和的查询解密方案,将所有可能包含用户请求查询结果的候选信息反馈给客户,在客户端,通过进一步对候选解的有效剪枝,降低系统查询的请求和信息传输过程复杂度,方法简单易用,且具有一定的保密性质,但是在查

到稿日期:2013-10-30 返修日期:2013-10-30 本文受国家自然科学基金(61073041,61073043,61202274,61370083),高等学校博士学科点基金(20112304110011,20122304110012)资助。

蔡朝晖(1968—),女,博士生,副教授,主要研究领域为数据挖掘和隐私保护,E-mail:caozhaohuiheb@163.com;张健沛(1956—),男,博士,教授,博士生导师,主要研究领域为数据挖掘、机器学习和隐私保护;杨静(1962—),女,博士,教授,博士生导师,主要研究领域为社会计算、数据挖掘和隐私保护。

询时,系统对于近邻的衡量和评判需要一个较严格的尺度标准,这恰恰是系统的瓶颈所在。文献[2]中,张德干、王冬提出了面向 LBS 的服务匹配方法,采用 LBS 匹配的方法,在位置服务中,首先分析了主流的发现协议,并基于分析比较,提出了一种新的服务匹配方法,新的服务匹配方法中,通过分层匹配和用户喜好程度两个匹配要素来提高服务匹配成功率,从而最终达到提高位置服务质量的目的,算法将用户的喜好程度加入评判系统,是一种创新应用,但考虑衡量系统的评判要素较少,不具备普遍的代表性。文献[3]提出一种基于动态交通信息的移动位置服务系统,针对当前位置服务系统主要是基于静态数据的服务,并且服务缺乏实时性的缺点,力图采用新的方法将动态位置信息应用到交通系统中,系统属于在新的应用平台上的实现,稳定性有待进一步的考证。薛姣、刘向宇等人研究了一种面向公路网络的位置隐私保护方法,提出采用隐匿环和隐匿树进行公路网络位置信息模糊查询,并有效地保护位置隐私,为移动用户对目前位置的近邻查询提供了一种合理思路,隐匿环和隐匿树计算复杂,单行线公路网络位置查询的可靠性无法保证^[4]。另外,毛典辉、蔡强等人^[5]也提出了一种基于自适应情景的位置隐私保护方法,对当前 LBS 定位服务系统中出现的隐私保护度与通信开销不平衡的问题进行探讨和改进,但其方法是通过上下文来感知位置情景,在位置情景的动态变化下多状态位置信息挖掘值得进一步研究,且对路网密度与终端性能的自适应平衡度有待提高。同时,关于位置隐私保护协议的研究和位置服务优化挖掘和保护的组网构建的研究也在广泛开展和进行^[6-10]。

针对传统研究成果的不足和存在的问题,为有效可靠地进行位置服务查询和动态挖掘,本文提出了一种基于多态关联挖掘的位置服务优化查询方法,采用深度数据挖掘的方法提取数据之间多个状态的关联特征,然后根据关联特征优化系统位置服务的查询方法,大大提高系统的查询可靠性和准确率。

2 多态关联挖掘方法

多态关联挖掘是一种数据处理的新挖掘方法,采用数据多态性深度挖掘的方法,在对数据的多态关联度进行衡量考虑的基础上,对数据特征进行挖掘^[11]。

多态关联挖掘时,首先假设数据待挖掘区域为 W ,该区域为需要进行数据挖掘的区域,采用挖掘处理的方法对该区域进行挖掘分析,把它视为一个分析区域,该分析区域为 $p \times q$ 分析元素。在待挖掘区域 W 中,假设需要设置 Q 同等性能指标的数据,各个数据在区域 W 中的坐标参数假设为 (x_i, y_i) ,各个数据由于性能参数相同,故其挖掘半径是相同的,这里假设为 r ,因此,最终能有效实现全区域挖掘的半径为 $R = 2r$,相应的每个数据坐标 (x_i, y_i) 的圆心可表示为 $W_i = \{x_i, y_i, r\}$,其中 W_i 的半径表示为 r ,为挖掘半径。区域内所需的目标分析元素点假设为 (x, y) ,根据欧式距离的定义,得到该目标点与源节点之间的距离为:

$$h(w_i, k) = \sqrt{(x_i - x)^2 + (y_i - y)^2} \quad (1)$$

假设 $K\{r_i\}$ 为区域中的分析元素点 (x, y) 能被节点 W_i 挖掘的改善因子。该改善因子具有一定的规律性,可用如下公式表示:

$$K_{upg}(x, y, w_i) = \begin{cases} 1, & d(w_i, k) \leq r - r_u \\ u(\frac{-\alpha_1 \theta_1}{\beta_2 \beta_2} + \alpha_2), & r - r_u < d(w_i, k) < r + r_u \\ 0, & \text{else} \end{cases} \quad (2)$$

式中, r_u ($0 < r_u < r$) 为挖掘能够实现通信传输和监测的有效性能参数; $\alpha_1, \alpha_2, \beta_1, \beta_2$ 对应于数据检测系数,该系数与其特征相关; θ_i 为输入参数变量,表示式为:

$$\theta_i = r_u - r + d(w_i, k) \quad (3)$$

在整个挖掘中,通过多个数据节点,实现对区域的多态关联挖掘,非此即彼,协同进行,从而提高了对区域的挖掘效果,多态关联挖掘的挖掘改善因子可表示为^[6]:

$$K_{upg}(W_{pg}) = 1 - \prod_{w_i \in W_{pg}} (1 - K_{upg}(x, y, w_i)) \quad (4)$$

在挖掘系统中,设数据中待挖掘区域为 W ,同样为 $p \times q$ 分析元素。如上所述,各个节点集合的多态关联挖掘的改善因子为 $K_{upg}(W_{pg})$,通过该改善因子可以得到区域内的各个分析元素是否为关联挖掘信息,由此得到信息待挖掘区域为 W 的节点集合的关联挖掘率为:

$$Q_w(\omega) = \frac{\sum K_{upg}(W_{pg})}{p \times q} \quad (5)$$

在利用上述方法重构的数据集合中,数据 x_j 到除 x_i 本身外数据 x_i 的距离小于 r 的 x_j 的数据类似度定义为:

$$Q = \sum_{j \neq i} H(r - \|x_i - x_j\|) \quad (6)$$

此处设定关联函数概念,定义为把所有有可能在距离上比给定的距离 r 小的点相对于占有总的点总数的比例称为关联函数^[12,13],表示为:

$$C_N(r) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N H(r - \|x_i - x_j\|) \quad (7)$$

式中分子为 2 是为了排除重复计数,用范数表示两数据之间的类似度。可得到两数据类似度为两个矢量的最大差分量:

$$\|x_i - x_j\| = \max_{1 \leq k \leq m} |x_{i-(k-1)r} - x_{j-(k-1)r}| \quad (8)$$

对于数据类似不大于 r 的挖掘,可称为具有关联性的矢量。在此,假设一维数据的数据序列为 n ,则多态关联挖掘中的挖掘点个数为 $N = n - (m-1)r$ 。计算挖掘点中有关联的挖掘点对数,占一切可能的 $N(N-1)/2$ 种配对的比例称为关联积分,表示为:

$$C_m(r) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N H(r - \|x_i - x_j\|) \quad (9)$$

通过上面的多态关联挖掘方法,可以对待挖掘区域的数据进行深度分析,提取有效信息。

3 基于多态关联挖掘的位置服务优化查询

位置服务查询系统性能的好坏直接取决于系统查询匹配算法的优劣。为了适应各种情形下的系统查询,构建系统的匹配算法标准为弹性匹配。弹性匹配算法主要是考虑了基于功能性的位置服务匹配查询方法,考虑系统查询服务的输入、输出参数的匹配程度,给出匹配结果。

匹配查询算法共定义了 5 种不同的匹配等级,分别是:完全匹配(Exact)、插拔匹配(Plugin)、包含匹配(Subsume)、相干匹配(Intersection)和匹配失败(Fail)。5 种匹配的关系图如图 1(a)~(e)所示。

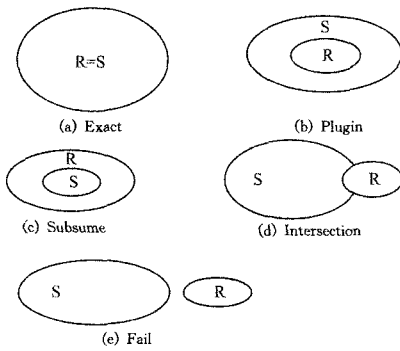


图1 5种不同的匹配等级

实际匹配查询时,给定两个位置服务优化查询的时间序列分别为 Q 与 P (设 Q 数据序列长为 m , P 数据序列长为 n), 采用归一化后的距离来评价位置服务优化查询系统中两个位置服务优化查询序列间的相似性, 定义为:

$$s = \exp[-\gamma(m, n)/G\sigma] \quad (10)$$

式中, $\gamma(m, n)$ 为位置服务优化查询两个序列 Q 与 S 的最小累加距离; G 为位置服务优化查询最佳学习路径长度, 其取值范围为:

$$\max(m, n) \leq G \leq m+n-1$$

其中, σ 为常数因子。

在位置服务优化查询系统中, 定义系统中多个查询样本之间的方差贡献率 μ 为:

$$\mu = \frac{\sum_{i=1}^L \lambda_i}{\sum_{j=1}^H \lambda_j} \quad (L \leq H) \quad (11)$$

通过上面的参数可以反映出位置服务优化查询系统各个学习元素之间相似度的整体情况, 取值越小, 表明样本之间的相似度越高, 反之, 表明样本之间的相似度越小。

位置服务优化查询系统中, 高斯混合函数反映的是系统的分布情况, 则位置服务优化查询系统的高斯混合函数定义为:

$$p(y|\alpha, \theta) = \sum_{k=1}^K \alpha_k p_k(y|\mu_k, \Sigma_k) \quad (12)$$

式中, y 为位置服务优化查询系统各分量相互独立的 D 维随机变量; $p_k(y|\mu_k, \Sigma_k)$ 为位置服务优化查询系统第 k 个成分的高斯密度函数; μ_k 为位置服务优化查询序列的均值; Σ_k 为位置服务优化查询系统的协方差; θ 为由位置服务优化查询系统成分密度函数的参数所构成的向量, 包括各个高斯密度函数的均值 μ 、协方差 Σ_k ; $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_K)$ ($\sum_{k=1}^K \alpha_k = 1$) 为位置服务优化查询系统的混合系数向量; α_k 为位置服务优化查询系统的第 k 个成分被选择的概率。

位置服务优化查询系统中各隐藏状态之间的转移概率定义为:

$$a_{ij} = (1 - a_{ii}) / (N - 1) \quad (i \neq j) \quad (13)$$

式中, a_{ii} 为位置服务优化查询系统第 i 个隐藏学习节点的自转移概率。

设位置服务优化查询系统中, 学习系统当前学习序列为 y , 第 i 个位置服务优化查询的参数集合定义为:

$$\lambda_i \quad (i=1, 2, 3, \dots, C)$$

式中, λ_i 为包括第 i 个位置服务优化查询的初始状态的分布向量。

位置服务优化查询的学习状态转移矩阵和学习输出分布矩阵是反映位置服务优化查询系统中转移概率特性的描述

量, 位置服务优化查询系统的阈值模型参数为 λ_T , 如果位置服务优化查询的 y 属于第 i 类, 则定义:

$$i = \max_j \{P(Y|\lambda_j) / P(Y|\lambda_T) > P(Y|\lambda_T)\} \quad (j=1, 2, 3, \dots, C) \quad (14)$$

式中, $P(Y|\lambda_j)$ 为第 j 个位置服务优化查询产生观察序列 y 的概率; $P(Y|\lambda_T)$ 为位置服务优化查询阈值模型产生学习观察序列 y 的概率。

在位置服务优化查询数据处理时, 根据最大后验原理, 我们选择使位置服务优化查询的后验概率密度产生最大参数时的值作为当前位置服务优化查询参数估计的一个原则, 即:

$$\theta_{MAP} = \arg \max_{\theta} p(\theta|X) = \arg \max_{\theta} p(X|\theta)P(\theta) \quad (15)$$

式中, $p(X|\theta)$ 为位置服务优化查询中相对于样本集 $X = \{x_1, x_2, \dots, x_N\}$ 的最大似然函数; $P(\theta)$ 为位置服务优化查询的元素先验分布; θ 为位置服务优化查询的似然函数参数。

通过上面的基于高斯分布的系统位置服务优化查询方法, 可以从概率统计的角度, 充分查询系统的需求, 实现位置服务的优化查询, 通过优化查询的方法, 可以大大提高整个系统的查询效率和查询准确率, 即查全率和查准率。

4 系统实验与结果分析

4.1 实验环境描述

为了测试基于多态关联挖掘的位置服务优化查询方法相对于传统方法的查询性能改进, 构建一组 200 组数据的查询测试系统, 采用数据查全率与数据查准率为衡量系统的标准, 查全率为实际查询的数据与请求查询的数据之比, 查准率为实际查询的准确数据与查询到的数据之比。通过这两个衡量标准, 可以很好地对系统的查询结果进行性能优劣的判定。

4.2 结果分析

查询时查询响应信号随响应时间的变化特性如图 2 所示。

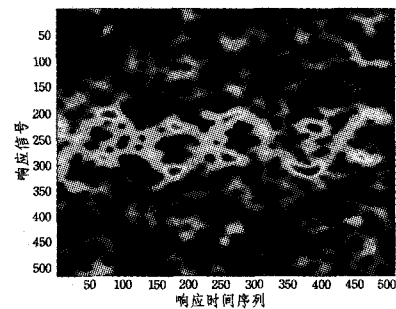


图2 查询信号的响应特性

从图 2 可以看出, 在位置优化服务数据查询的响应时间内, 查询信号与响应信号在时间轴平面上前后均匀地分布开来, 先产生系统的查询访问信号, 然后响应的服务器就会产生用户需要的响应信号。由于系统访问和请求的随机性, 随意系统的响应特性在平面上也是散开分布的。

查全率表征了系统响应用户请求时能够完全响应的能力, 传统方法与多态关联挖掘的查全率比较如图 3 所示。

从图 3 可以看出, 与传统的查询方法相比, 基于多态关联挖掘的优化查询方法具有更高的查全率, 平均高于传统方法 9.5%。

查准率表征了系统对于用户请求的正确响应能力, 传统方法与多态关联挖掘的查准率比较如图 4 所示。

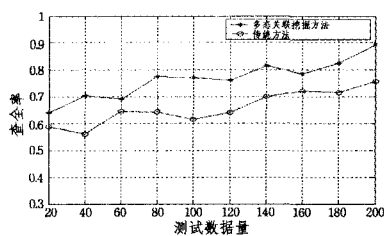


图3 查全率比较

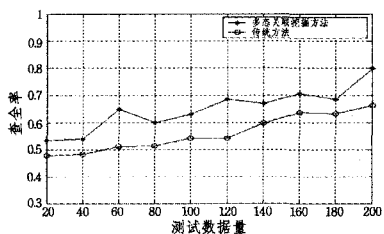


图4 查准率比较

从图4可以看出,与传统的查询方法相比,基于多态关联挖掘的优化查询方法具有更高的查准率,平均高于传统方法8.8%。

结束语 随着现代网络技术和定位技术的高速发展,基于位置服务的定位技术的广泛应用成为了可能,而高效可靠的位置服务查询方法是位置服务广泛应用的关键。传统的位置服务查询方法采用近邻查询的方法,在小范围近邻中应用可靠,但在对近邻的分割时往往面临稳定性差的问题,无法实现高效查询。本文针对传统方法的问题,提出了一种基于多态关联挖掘的位置服务优化查询方法,在系统优化查询时,该方法首先根据查询标准,在数据多态分割的基础上实现深度的数据挖掘,提取数据的深层次特征,根据深层次特征,对位置服务的查询方法进行优化查询,以提高系统查询的效率和可靠性。最后采用200组数据进行实际的分析,将查全率和查准率作为衡量标准,实验结果显示,基于多态关联挖掘的位

置服务查询方法使查全率和查准率分别提高9.5%和8.8%,具有很好的应用价值。

参考文献

- [1] 胡磊,王佳俊,倪巍伟.一种基于坐标和保护位置隐私近邻查询方法[J]. 计算机科学,2012,39(8):173-177
- [2] 张德干,王冬.面向LBS的服务匹配方法研究[J]. 计算机科学,2012,39(3):17-21
- [3] 何丽娜,刘岳峰,晏磊,等.基于动态交通信息的移动位置服务系统[J]. 计算机工程与应用,2007,43(7):108-111
- [4] 薛姣,刘向宇,杨晓春,等.一种面向公路网络的位置隐私保护方法[J]. 计算机学报,2011,34(5):865-878
- [5] 毛典辉,蔡强,李海生,等.一种自适应情景的位置隐私保护方法[J]. 中南大学学报:自然科学版,2013,44(7):279-283
- [6] 王华东,李巍.混沌粒子群算法在WSN覆盖优化中的应用[J]. 科技通报,2012,28(8):114-116
- [7] 吕庆聪,周集良,杨帆,等.普适计算服务匹配技术研究[J]. 计算机科学,2009,36(11):182-185
- [8] 黄毅,霍峥,孟小峰. CoPrivacy:一种用户协作无匿名区域的位置隐私保护方法[J]. 计算机学报,2011,34(10):1976-1985
- [9] 陈娟,方滨兴,殷丽华.传感器网络中基于源节点有限洪泛的源位置隐私保护协议[J]. 计算机学报,2010,33(9):1736-1746
- [10] 张建明,赵玉娟,江浩斌,等.车辆自组网的位置隐私保护技术研究[J]. 通信学报,2012,33(8):180-189
- [11] 刘燕.基于云计算信息处理系统体系结构设计[J]. 科技通报,2012,28(18):100-102
- [12] 侯丽敏,张瑞坤.基于Agent的QoS组播路由算法及仿真[J]. 计算机仿真,2011,28(1):140-144
- [13] 林祝亮,冯远.基于粒子群算法的无线传感网络覆盖优化策略[J]. 计算机仿真,2009,26(4):190-193
- [14] 钱建平,张志涌,阮秀凯.基于复值多态连续Hopfield网络的MPSK信号盲检测[J]. 重庆邮电大学学报:自然科学版,2012,24(3):292-296

(上接第285页)

给出并证明了两个转化定理,然后给出了外P-概念格的一种构建算法。下一步将研究利用外P-概念格进行概念聚类、规则抽取。文献[14]提出的P-集合理论已经在诸多领域得到了深入的应用,相信P-形式概念分析理论将会具有重要的研究价值。

参考文献

- [1] Wille R. Restructuring Lattice theory: An approach based on hierarchies of concepts[C]// Rival I, ed. Ordered Sets. Dordrecht: Reidel, 1985:445-470
- [2] 陈湘,吴跃.基于概念格挖掘GIS中的关联规则[J]. 计算机工程,2011,31(3):686-689
- [3] 胡明涵,张俐,任飞亮.模糊形式概念分析与模糊概念格[J]. 东北大学学报:自然科学版,2007,28(9):1274-1277
- [4] 谢志鹏,刘宗田.概念格的快速渐进式构造算法[J]. 计算机学报,2002,25(5):490-496
- [5] 蒋义勇,张继福,张素兰.基于链式结构的概念格渐进式构造[J]. 计算机工程与应用,2007,43(11):178-180
- [6] 刘利峰,吴孟达,王丹.基于属性约简的概念格构造[J]. 计算机

工程与科学,2007,29(6):140-142

- [7] Medina J, Ojeda-Aciego M, Ruiz-Calvino J. Relating generalized concept lattices and concept lattices for non-commutative conjunctors[J]. Applied Mathematics Letters, 2008, 21:1296-1300
- [8] Ghosh P, Kundu L, Sarker D. Fuzzy graph representation of a fuzzy concept lattice[J]. Fuzzy Sets and Systems, 2010, 161(12):1669-1675
- [9] 张继福,张素兰.加权概念格及其渐进式构造[J]. 模式识别与人工智能,2005,18(2):171-176
- [10] 张继福,张素兰,胡立华.约束概念格及其构造算法[J]. 智能系统学报,2006,1(2):31-38
- [11] 姜峰,范玉顺.基于扩展概念格的Web关系挖掘[J]. 软件学报,2010,21(10):2432-2444
- [12] 刘保相,张春英.一种新的概念格结构——区间概念格[J]. 计算机科学,2012,39(8):273-277
- [13] 杨海峰,张继福.粗糙概念格及构造算法[J]. 计算机工程与应用,2007,43(24):172-175
- [14] 史开泉.P-集合与它的应用特征[J]. 计算机科学,2010,37(8):1-8
- [15] 张丽,崔玉泉,史开泉.外P-集合与数据内-恢复[J]. 系统工程与电子技术,2010,32(6):1233-1238