



(10) 授权公告号 CN 114676754 B

(45) 授权公告日 2024. 11. 29

(21) 申请号 202210205311.6

G06F 18/25 (2023.01)

(22) 申请日 2022.03.02

G06F 18/10 (2023.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 114676754 A

(56) 对比文件

AU 2020103923 A4, 2021.02.11

CN 113704853 A, 2021.11.26

(43) 申请公布日 2022.06.28

审查员 张裕

(73) 专利权人 香港理工大学深圳研究院

地址 518057 广东省深圳市南山区粤海街  
道高新技术产业园南区粤兴一道18号  
香港理工大学产学研大楼205室

(72) 发明人 史文中 王茂鲟 张安舒

(74) 专利代理机构 深圳市君胜知识产权代理事

务所(普通合伙) 44268

专利代理师 李可 王永文

(51) Int. Cl.

G06F 18/22 (2023.01)

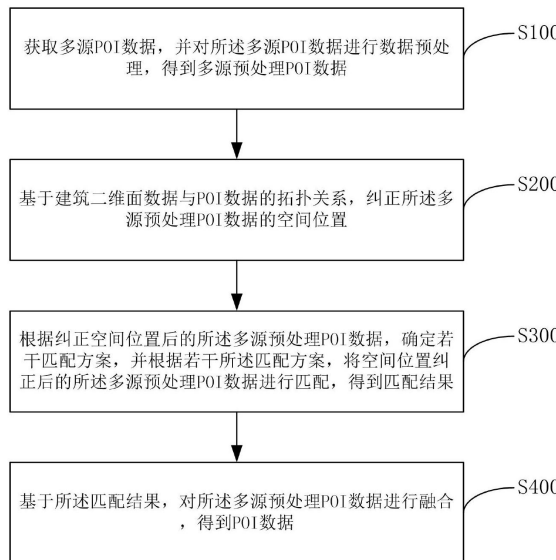
权利要求书2页 说明书11页 附图4页

(54) 发明名称

一种多源POI数据匹配和融合方法

(57) 摘要

本发明公开了一种多源POI数据匹配和融合方法,方法包括:获取多源POI数据,并对多源POI数据进行数据预处理,得到多源预处理POI数据;基于建筑二维面数据与POI数据的拓扑关系,纠正多源预处理POI数据的空间位置;根据纠正空间位置后的多源预处理POI数据,确定若干匹配方案,并根据若干匹配方案,将空间位置纠正后的多源预处理POI数据进行匹配,得到匹配结果;基于匹配结果,对多源预处理POI数据进行融合,得到POI融合数据。本发明通过补全不完整的POI地址,然后纠正POI的空间位置,接着根据若干匹配方案进行匹配,使得匹配全面且匹配计算量小,并在匹配之后进行融合,使得融合后的POI数据的数量更加完整。



1. 一种多源POI数据匹配和融合方法,其特征在于,所述方法包括:  
获取多源POI数据,并对所述多源POI数据进行数据预处理,得到多源预处理POI数据;  
基于建筑二维面数据与POI数据的拓扑关系,纠正所述多源预处理POI数据的空间位置;  
根据纠正空间位置后的所述多源预处理POI数据,确定若干匹配方案,并根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行匹配,得到匹配结果;  
基于所述匹配结果,对所述多源预处理POI数据进行融合,得到POI融合数据;  
所述基于建筑二维面数据与POI数据的拓扑关系,纠正所述多源预处理POI数据的空间位置包括:  
基于空间几何算法,提取所有建筑二维面数据的质心的空间位置;  
将所述建筑二维面数据的质心的空间位置作为属性值添加到所述多源预处理POI数据中;  
基于建筑二维面数据与POI数据的拓扑关系,将添加属性值的所述多源预处理POI数据输入预设的空间分析工具,得到定位错误的POI数据集;  
对所述定位错误的POI数据集中的质心进行重定位,以实现纠正所述多源预处理POI数据的空间位置;  
纠正空间位置后的所述多源预处理POI数据包括待匹配POI和若干异源POI;  
所述根据纠正空间位置后的所述多源预处理POI数据,确定若干匹配方案,并根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行匹配,得到匹配结果包括:  
解析纠正空间位置后的所述多源预处理POI数据,得到所述多源预处理POI数据对应的属性特征;  
根据所述属性特征,确定若干匹配方案;  
根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行匹配,得到匹配结果。  
2. 根据权利要求1所述的多源POI数据匹配和融合方法,其特征在于,所述对所述多源POI数据进行数据预处理,得到多源预处理POI数据包括:  
基于预设的标准,统一所述多源POI数据中的POI名称;  
解析所述多源POI数据的地址,得到若干地址段,并将若干所述地址段作为新的属性值添加到所述多源POI数据;  
对若干所述地址段数据集进行数据挖掘,得到挖掘结果,并基于预设的关联规则和所述挖掘结果,补充所述多源POI数据的地址。  
3. 根据权利要求1所述的多源POI数据匹配和融合方法,其特征在于,所述对所述定位错误的POI数据集中的质心进行重定位,以实现纠正所述多源预处理POI数据的空间位置之后,还包括:  
当重定位后的POI数据集有定位错误时,再一次对所述定位错误的POI数据集中的质心进行重定位。  
4. 根据权利要求1所述的多源POI数据匹配和融合方法,其特征在于,所述属性特征包括第一属性特征、第二属性特征或第三属性特征;所述根据所述属性特征,确定若干匹配方

案包括：

当所述属性特征为第一属性特征时，匹配方案为将待匹配POI与若干异源POI按照第一地址细粒度进行组合和匹配；

当所述属性特征为第二属性特征时，匹配方案为将待匹配POI与异源POI按照第二地址细粒度进行组合和匹配；

当所述属性特征为第三属性特征或者待匹配POI与异源POI未匹配时，匹配方案为将待匹配POI与异源POI以预设的距离值进行组合和匹配。

5. 根据权利要求4所述的多源POI数据匹配和融合方法，其特征在于，所述根据若干所述匹配方案，将空间位置纠正后的所述多源预处理POI数据进行匹配，得到匹配结果包括：

根据若干所述匹配方案，将空间位置纠正后的所述多源预处理POI数据进行两两组合；

计算组合后的点对名称的相似度；其中，所述点对名称为所述多源预处理POI数据进行两两组合后形成的点对的名称；

当点对名称的相似度大于预设的相似度阈值时，匹配结果为匹配成功，并将匹配成功的待匹配POI和异源POI存储到目标数据集。

6. 根据权利要求5所述的多源POI数据匹配和融合方法，其特征在于，所述基于所述匹配结果，对所述多源预处理POI数据进行融合，得到POI融合数据包括：

当目标数据集中的待匹配POI和异源POI中存在缺失的属性值时，将所述属性值进行补充；

当目标数据集中的待匹配POI和异源POI的名称不一致时，不执行任何操作；

当目标数据集中的待匹配POI和异源POI中同类属性值有若干时，将预设的第一属性值代替待匹配POI和异源POI中原有的属性值；

当目标数据集中的待匹配POI和异源POI中同类属性值具有唯一值时，将预设的第二属性值代替待匹配POI和异源POI中原有的属性值；

将经过属性值补充、属性值代替后的多源预处理POI数据作为POI融合数据。

7. 一种智能终端，其特征在于，包括有存储器，以及一个或者一个以上的程序，其中一个或者一个以上程序存储于存储器中，且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于执行如权利要求1-6中任意一项所述的方法。

8. 一种非临时性计算机可读存储介质，其特征在于，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行如权利要求1-6中任意一项所述的方法。

## 一种多源POI数据匹配和融合方法

### 技术领域

[0001] 本发明涉及空间大数据技术领域,尤其涉及的是一种多源POI数据匹配和融合方法。

### 背景技术

[0002] POI匹配和融合是通过多源数据丰富数字地图内容、提高地图质量的重要手段,已广泛渗透到国民生活的各个领域。但是现有的POI匹配方法会存在漏匹配,且现有的POI融合方法在融合之后出现数据不完整的问题。

[0003] 因此,现有技术还有待改进和发展。

### 发明内容

[0004] 本发明要解决的技术问题在于,针对现有技术的上述缺陷,提供一种多源POI数据匹配和融合方法,旨在解决现有技术中POI数据漏配和数据数量不完整的问题。

[0005] 本发明解决问题所采用的技术方案如下:

[0006] 第一方面,本发明实施例提供一种多源POI数据匹配和融合方法,其中,所述方法包括:

[0007] 获取多源POI数据,并对所述多源POI数据进行数据预处理,得到多源预处理POI数据;

[0008] 基于建筑二维面数据与POI数据的拓扑关系,纠正所述多源预处理POI 数据的空间位置;

[0009] 根据纠正空间位置后的所述多源预处理POI数据,确定若干匹配方案,并根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行匹配,得到匹配结果;

[0010] 基于所述匹配结果,对所述多源预处理POI数据进行融合,得到POI融合数据。

[0011] 在一种实现方式中,所述对所述多源POI数据进行数据预处理,得到多源预处理POI数据包括:

[0012] 基于预设的标准,统一所述多源POI数据中的POI名称;

[0013] 解析所述多源POI数据的地址,得到若干地址段,并将若干所述地址段作为新的属性值添加到所述多源POI数据;

[0014] 对若干所述地址段数据集进行数据挖掘,得到挖掘结果,并基于预设的关联规则和所述挖掘结果,补充所述多源POI数据的地址。

[0015] 在一种实现方式中,所述基于建筑二维面数据与POI数据的拓扑关系,纠正所述多源预处理POI数据的空间位置包括:

[0016] 基于空间几何算法,提取所有建筑二维面数据的质心的空间位置;

[0017] 将所述建筑二维面数据的质心的空间位置作为属性值添加到所述多源预处理POI数据中;

[0018] 基于建筑二维面数据与POI数据的拓扑关系,将添加属性值的所述多源预处理POI

数据输入预设的空间分析工具,得到定位错误的POI数据集;

[0019] 对所述定位错误的POI数据集中的质心进行重定位,以实现纠正所述多源预处理POI数据的空间位置。

[0020] 在一种实现方式中,所述对所述定位错误的POI数据集中的质心进行重定位,以实现纠正所述多源预处理POI数据的空间位置之后,还包括:

[0021] 当重定位后的POI数据集有定位错误时,再一次对所述定位错误的POI 数据集中的质心进行重定位。

[0022] 在一种实现方式中,纠正空间位置后的所述多源预处理POI数据包括待匹配POI和若干异源POI;所述根据纠正空间位置后的所述多源预处理POI数据,确定若干匹配方案,并根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行匹配,得到匹配结果包括:

[0023] 解析纠正空间位置后的所述多源预处理POI数据,得到所述多源预处理 POI数据对应的属性特征;

[0024] 根据所述属性特征,确定若干匹配方案;

[0025] 根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行匹配,得到匹配结果。

[0026] 在一种实现方式中,所述属性特征包括第一属性特征、第二属性特征或第三属性特征;所述根据所述属性特征,确定若干匹配方案包括:

[0027] 当所述属性特征为第一属性特征时,匹配方案为将待匹配POI与若干异源POI按照第一地址细粒度进行组合和匹配;

[0028] 当所述属性特征为第二属性特征时,匹配方案为将待匹配POI与异源 POI按照第二地址细粒度进行组合和匹配;

[0029] 当所述属性特征为第三属性特征或者待匹配POI与异源POI未匹配时,匹配方案为将待匹配POI与异源POI以预设的距离值进行组合和匹配。

[0030] 在一种实现方式中,所述根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行匹配,得到匹配结果包括:

[0031] 根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行两两组合;

[0032] 计算组合后的点对名称的相似度;其中,所述点对名称为所述多源预处理POI数据进行两两组合后形成的点对的名称;

[0033] 当点对名称的相似度大于预设的相似度阈值时,匹配结果为匹配成功,并将匹配成功的待匹配POI和异源POI存储到目标数据集。

[0034] 在一种实现方式中,所述基于所述匹配结果,对所述多源预处理POI数据进行融合,得到POI融合数据包括:

[0035] 当目标数据集中的待匹配POI和异源POI中存在缺失的属性值时,将所述属性值进行补充;

[0036] 当目标数据集中的待匹配POI和异源POI的名称不一致时,不执行任何操作;

[0037] 当目标数据集中的待匹配POI和异源POI中同类属性值有若干时,将预设的第一属性值代替待匹配POI和异源POI中原有的属性值;

[0038] 当目标数据集中的待匹配POI和异源POI中同类属性值具有唯一值时,将预设的第二属性值代替待匹配POI和异源POI中原有的属性值;

[0039] 将经过属性值补充、属性值代替后的多源预处理POI数据作为POI融合数据。

[0040] 第二方面,本发明实施例还提供一种多源POI数据匹配和融合装置,其中,所述装置包括:

[0041] 多源预处理POI数据获取模块,获取多源POI数据,并对所述多源POI数据进行数据预处理,得到多源预处理POI数据;

[0042] 空间位置纠正模块,用于基于建筑二维面数据与POI数据的拓扑关系,纠正所述多源预处理POI数据的空间位置;

[0043] 匹配结果确定模块,用于根据纠正空间位置后的所述多源预处理POI数据,确定若干匹配方案,并根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行匹配,得到匹配结果;

[0044] POI融合数据确定模块,用于基于所述匹配结果,对所述多源预处理POI 数据进行融合,得到POI融合数据。

[0045] 第三方面,本发明实施例还提供一种智能终端,包括有存储器,以及一个或者一个以上的程序,其中一个或者一个以上程序存储于存储器中,且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于执行如上述任意一项所述的多源POI数据匹配和融合方法。

[0046] 第四方面,本发明实施例还提供一种非临时性计算机可读存储介质,当所述存储介质中的指令由电子设备的处理器执行时,使得电子设备能够执行如上述中任意一项所述的多源POI数据匹配和融合方法。

[0047] 本发明的有益效果:本发明实施例首先获取多源POI数据,并对所述多源POI数据进行数据预处理,得到多源预处理POI数据;然后基于建筑二维面数据与POI数据的拓扑关系,纠正所述多源预处理POI数据的空间位置;接着根据纠正空间位置后的所述多源预处理POI数据,确定若干匹配方案,并根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行匹配,得到匹配结果,最后基于所述匹配结果,对所述多源预处理POI数据进行融合,得到POI融合数据;可见,本发明实施例中充分挖掘POI已有信息,对不完整的属性进行了补全、对有冲突的属性进行了纠正;通过名称相似度就能高效匹配点对的方法,解决了由空间位置不准确导致的POI 漏配问题,对高密度城市POI数据尤其有效,可以解决由POI点过于密集导致的过大匹配量问题;纠正了POI错误的空间位置,POI融合结果更优,数据更完整。

## 附图说明

[0048] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明中记载的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0049] 图1为本发明实施例提供的多源POI数据匹配和融合方法流程示意图。

[0050] 图2为本发明实施例提供的一种实施例的建筑地址数据集关联规则的示意图。

- [0051] 图3为本发明实施例提供的POI匹配方案流程图。
- [0052] 图4为本发明实施例提供的一种实施例的POI空间位置纠正前后的空间分布示意图。
- [0053] 图5为本发明实施例提供的POI融合结果示意图。
- [0054] 图6为本发明实施例提供的多源POI数据匹配和融合装置的原理框图。
- [0055] 图7为本发明实施例提供的智能终端的内部结构原理框图。

## 具体实施方式

[0056] 本发明公开了一种多源POI数据匹配和融合方法,为使本发明的目的、技术方案及效果更加清楚、明确,以下参照附图并举实施例对本发明进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。

[0057] 本技术领域技术人员可以理解,除非特意声明,这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是,本发明的说明书中使用的措辞“包括”是指存在所述特征、整数、步骤、操作、元件和/或组件,但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、组件和/或它们的组。应该理解,当我们称元件被“连接”或“耦接”到另一元件时,它可以直接连接或耦接到其他元件,或者也可以存在中间元件。此外,这里使用的“连接”或“耦接”可以包括无线连接或无线耦接。这里使用的措辞“和/或”包括一个或更多个相关联的列出项的全部或任一单元和全部组合。

[0058] 本技术领域技术人员可以理解,除非另外定义,这里使用的所有术语(包括技术术语和科学术语),具有与本发明所属领域中的普通技术人员的一般理解相同的意义。还应该理解的是,诸如通用字典中定义的那些术语,应该被理解为具有与现有技术的上下文中的意义一致的意义,并且除非像这里一样被特定定义,否则不会用理想化或过于正式的含义来解释。

[0059] 由于现有技术中,最常用的POI匹配方法是基于POI的空间位置的POI 相似度计算办法,该方法首先计算不同来源POI的相似度,筛选出相似度最高的POI点对作为POI匹配点对,然后通过数据融合规则纠正和补全POI 属性值。

[0060] 目前的POI融合方法多将重心放于使用机器学习判断POI类别,再结合了POI的空间位置聚类和非空间属性相似度从不同来源的POI数据集中找出正确的融合对象。该方法用来判别POI类别时效果明显,但是由于数据预处理不充分,出现了以下的不足:1)未对POI的地址算法进行系统研究,缺少挖掘、补全地址信息和标准化地址格式的有效方案;2)对于引入了机器学习或人工智能模型的算法,鲁棒性较低,不能很好地适用于各个城市,真正地有效提高POI的融合精度。3)由POI的空间位置计算出的点对距离受POI空间位置信息质量影响较大,设置距离太小容易出现漏配问题,设置空间范围太大容易导致过大的匹配量。4)由于数据挖掘不够完全,导致很多POI漏配。

[0061] 为了解决现有技术的问题,本实施例提供了一种多源POI数据匹配和融合方法,通过上述方法充分挖掘POI已有信息,对不完整的属性进行了补全、对有冲突的属性进行了纠正;通过名称相似度就能高效匹配点对的方法,解决了由空间位置不准确导致的POI漏配问题,对高密度城市POI数据尤其有效,可以解决由POI点过于密集导致的过大匹配量问题;纠正了POI错误的空间位置,POI融合结果更优,数据更完整。

[0062] 示例性方法

[0063] 本实施例提供一种多源POI数据匹配和融合方法,该方法可以应用于空间大数据的智能终端。具体如图1所示,所述方法包括:

[0064] 步骤S100、获取多源POI数据,并对所述多源POI数据进行数据预处理,得到多源预处理POI数据;

[0065] 具体地,POI(Point of Interest),兴趣点,在地理信息系统中,一个 POI可以是一栋房子、一个商铺、一个邮筒、一个公交站等。多源POI数据可以在网站上获取,也可以在数据库中获取,多源POI数据可能存在格式不统一,或者信息错误的情况,故要对多源POI数据进行数据预处理,预处理的方式可以是将格式统一,也可以是将错误信息纠正,也可以是将不合规的数据进行剔除,具体不做限制。

[0066] 步骤S100包括如下步骤:

[0067] S101、基于预设的标准,统一所述多源POI数据中的POI名称;

[0068] S102、解析所述多源POI数据的地址,得到若干地址段,并将若干所述地址段作为新的属性值添加到所述多源POI数据;

[0069] S103、对若干所述地址段数据集进行数据挖掘,得到挖掘结果,并基于预设的关联规则和所述挖掘结果,补充所述多源POI数据的地址。

[0070] 具体地,预设的标准为POI数据主要语言,如中文,也即基于中文统一所述多源POI数据的POI名称。在本实施例中,本发明使用网页爬虫技术在 Google map中自动批量搜索POI的英文名称,再使用正则表达式识别并存储其对应的中文名称,将各来源POI数据集中的英文名称统一为简体中文。此外,通过对比为名称标准化前后的数据集,评价POI名称的准确性,不准确的判断标准:在Google map上搜索不到的POI和相比Google map上的中文名称有错别字的名称为不准确的POI名称。统一所述多源POI数据中的POI名称后,解析所述多源POI数据的地址,得到若干地址段,并将若干所述地址段作为新的属性值添加到所述多源POI数据。在本实施例中,利用正则表达式将每个POI地址解析成6个地址段并作为新的属性值添加到各来源POI数据集:街道,街道号码,大楼名称,大楼座号,楼层,房间号;得到地址完全拆分的多个数据集。此外,一个POI地址被解析后的6个部分缺少任意一部分都被认为是地址不完整,通过统计POI数据中的不完整地址所占的比例,评价各来源POI地址的完整性,一个POI地址已经被解析成了6个部分,缺一个部分都被认为地址不完整。对若干所述地址段数据集进行数据挖掘,得到挖掘结果为从地址解析后的数据集中筛选出所有大楼名称不为空的地址,通过重操作建立建筑地址的数据集,其中包括街道、街道号、大楼名称(建筑名称)。目的是为了补全POI地址中的缺失部分、纠正错误的地址、在匹配过程中用地址属性更快地搜索小范围内的POI点对。最后基于预设的关联规则和所述挖掘结果,补充所述多源POI数据的地址。在本实施例中,预设的关联规则为街道、街道号码与建筑的关联规则,具体如图2所示。对比不同来源关联规则的数量,评价不同来源的关联规则的相对完整性。同时,为了纠正POI街道名称的书写错误,从地址解析后的数据集中筛选出所有街道级别的地址信息,建立街道名称数据集。使用建筑地址数据集,对于本来位于建筑内且地址中缺少建筑名称的POI,通过POI的街道和街道号,补全POI缺失的建筑信息;对于地址中只有房间号却缺少建筑座号、楼层号的 POI,可根据地址解析后数据集中的房间号信息,挖掘得到POI所在的建筑的座号、楼层、房号;使用街道地址数据集,通过模糊匹配算法,也即用 python语

言编写算法在CSV中的数据集中操作,去标准化各来源POI数据集的街道名称。这样得到匹配前名称和地址数据质量都被提高的待匹配POI 数据集A,和其他多源POI数据集 $S_i$  ( $i \geq 1$ )。在一种实现方式中,通过数据更新时间判断不同来源POI数据的时效性。

[0071] 得到多源预处理POI数据后,就可以执行如图1所示的如下步骤:S200、基于建筑二维面数据与POI数据的拓扑关系,纠正所述多源预处理POI数据的空间位置;相应的,步骤S200包括如下步骤:

[0072] S201、基于空间几何算法,提取所有建筑二维面数据的质心的空间位置;

[0073] S202、将所述建筑二维面数据的质心的空间位置作为属性值添加到所述多源预处理POI数据中;

[0074] S203、基于建筑二维面数据与POI数据的拓扑关系,将添加属性值的所述多源预处理POI数据输入预设的空间分析工具,得到定位错误的POI数据集;

[0075] S204、对所述定位错误的POI数据集中的质心进行重定位,以实现纠正所述多源预处理POI数据的空间位置。

[0076] 具体地,基于空间几何算法,提取所有建筑二维面数据的质心的空间位置为通过几何计算出所有建筑面数据质心的空间位置(centro\_lon, centro\_lat),然后将所述建筑二维面数据的质心的空间位置作为属性值添加到所述多源预处理POI数据中,如同时将其保留为新的属性用于导航。接着,基于建筑二维面数据与POI数据的拓扑关系,将添加属性值的所述多源预处理POI数据输入预设的空间分析工具,得到定位错误的POI数据集,如使用 ArcGIS的空间分析工具,得到被匹配源数据集A中POI点数据和Building二维面数据的拓扑关系,输出定位绝对错误的POI数据集A\_wrong\_loc,如位于建筑外的餐厅和酒店或位于建筑内的公交车站;最后对所述定位错误的 POI数据集中的质心进行重定位,以实现纠正所述多源预处理POI数据的空间位置,例如,根据A\_wrong\_loc中POI的“建筑名称”对应的建筑在2D数据集中的质心,通过以下公式对POI进行重定位:

[0077]  $[lon1, lat1] = [centro\_lon + \text{random}(-0.000001, 0.000001), centro\_lat + \text{random}(-0.000001, 0.000001)]$ 。

[0078] 其中,  $[centro\_lon, centro\_lat]$  表示二维平面上的建筑面数据质心的空间位置,经纬度0.000001在现实生活中的距离为10米,将POI重定位于建筑质心的10m范围内是为了保证纠正后的POI在电子地图缩放时即位于建筑内,又不会重叠。

[0079] 在一种实现方式中,统计不同来源空间位置的准确性(如位置绝对错误的POI所占的比例),得出不同来源POI空间位置的可信度。

[0080] 在一种实现方式中,将纠正后的空间位置 $[lon1, lat1]$ 替换掉A和 $S_i$ 中的空间位置 $[lon, lat]$ 。

[0081] 在一种实现方式中,对所述定位错误的POI数据集中的质心进行重定位,以实现纠正所述多源预处理POI数据的空间位置之后,还包括如下步骤:当重定位后的POI数据集有定位错误时,再一次对所述定位错误的POI数据集中的质心进行重定位。

[0082] 具体地,当重定位后的POI数据集仍有错误时,需要进一步处理。在本实施例中,判断空间纠正后的POI是否位于2D建筑面内,若在建筑面外,则使用更小的空间范围使其落到到建筑面内。

[0083] 纠正所述多源预处理POI数据的空间位置后,就可以执行如图1所示的如下步骤:

S300、根据纠正空间位置后的所述多源预处理POI数据,确定若干匹配方案,并根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行匹配,得到匹配结果;相应的,步骤S300包括如下步骤:

[0084] S301、解析纠正空间位置后的所述多源预处理POI数据,得到所述多源预处理POI数据对应的属性特征;

[0085] S302、根据所述属性特征,确定若干匹配方案;

[0086] S303、根据若干所述匹配方案,将空间位置纠正后的所述多源预处理 POI数据进行匹配,得到匹配结果。

[0087] 具体地,如图3所示,解析纠正空间位置后的所述多源预处理POI数据,然后识别预处理中经过解析的POI地址,基于parallel的思想将POI分为三个类数据集:a类有“建筑”信息;b类“建筑”属性为空值,而有“街道”和“街道号”信息;c类“街道”,“街道号”和“建筑”信息均为空,无法通过地址信息识别其所在的小尺度空间范围。其中地址细粒度: $a>b>c$ 。将有“建筑”信息作为第一属性特征,将“建筑”属性为空值,而有“街道”和“街道号”信息作为第二属性特征;将“街道”,“街道号”和“建筑”信息均为空作为第三属性特征。

[0088] 所述属性特征包括第一属性特征、第二属性特征或第三属性特征;步骤S302包括如下步骤:当所述属性特征为第一属性特征时,匹配方案为将待匹配POI与若干异源POI按照第一地址细粒度进行组合和匹配;当所述属性特征为第二属性特征时,匹配方案为将待匹配POI与异源POI按照第二地址细粒度进行组合和匹配;当所述属性特征为第三属性特征或者待匹配POI 与异源POI未匹配时,匹配方案为将待匹配POI与异源POI以预设的距离值进行组合和匹配。

[0089] 具体地,第一地址细粒度为a类,第二地址细粒度为b类,当所述属性特征为第一属性特征时,匹配方案为将待匹配POI与若干异源POI按照第一地址细粒度进行组合和匹配为将处于同一空间尺度的异源POI两两组合和匹配;待匹配源A中的a类POI与异源 $S_i$ 中a类中处于同一“建筑”内的POI两两组合和匹配,作为被匹配和融合的来源A,A中一个POI会和 $S_i$ 中多个POI两两组合和匹配。当所述属性特征为第二属性特征时,匹配方案为将待匹配POI与若干异源POI按照第二地址细粒度进行组合和匹配为将处于同一空间尺度的异源POI两两组合和匹配;A中待匹配源的b 类POI与 $S_i$ 中b类中处于同一“街道”和“街道号”的POI两两组合和匹配,作为被匹配和融合的来源A,A中一个POI会和 $S_i$ 中多个POI两两组合和匹配。当所述属性特征为第三属性特征或者待匹配POI与异源POI未匹配时,匹配方案为将待匹配POI与异源POI以预设的距离值进行组合和匹配,如,在A和 $S_i$ 数据集中删除Match1中被匹配的POI,剩余的POI 和c类放到同一个数据集d类,再以A源POI为单位与其600m范围内的  $S_i$ 源点对两两组合和匹配。

[0090] 步骤S303包括如下步骤:根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行两两组合;计算组合后的点对名称的相似度;其中,所述点对名称为所述多源预处理POI数据进行两两组合后形成的点对的名称;当点对名称的相似度大于预设的相似度阈值时,匹配结果为匹配成功,并将匹配成功的待匹配POI和异源POI存储到目标数据集。

[0091] 具体地,当匹配方案为将待匹配POI与若干异源POI按照第一地址细粒度进行组合和匹配时,处于同一地址细粒度且同一空间尺度的异源POI两两组合生成POI预匹配点对

集。待匹配源A中的a类POI与异源 $S_i$ 中a类中处于同一“建筑”内的POI两两组合,然后计算组合后的点对名称的相似度;其中,所述点对名称为所述多源预处理POI数据进行两两组合后形成的点对的名称;当点对名称的相似度大于预设的相似度阈值时,匹配结果为匹配成功,并将匹配成功的待匹配POI和异源POI存储到目标数据集。计算点对的名称相似度,使用的算法是基于DTW动态时间序列理论的非语义有序字符串的相似度算法。以A源POI为单位,筛选出名称相似度最大的POI点对。对上一步筛选出的点对的名称相似度进行排序,发现名称相似度大于0.7时,POI 点对被认为是被成功匹配的同一POI。将这些点对的所有属性被放入匹配结果数据集Match1。当匹配方案为将待匹配POI与异源POI按照第二地址细粒度进行组合和匹配时,将空间位置纠正后的所述多源预处理POI数据进行两两组合,如:A中待匹配源的b类POI与 $S_i$ 中b类中处于同一“街道”和“街道号”的POI两两组合。作为被匹配和融合的来源A,A中一个POI会和 $S_i$ 中多个POI匹配。然后计算组合后的点对名称的相似度;其中,所述点对名称为所述多源预处理POI数据进行两两组合后形成的点对的名称;当点对名称的相似度大于预设的相似度阈值时,匹配结果为匹配成功,并将匹配成功的待匹配POI和异源POI存储到目标数据集,例如:计算点对的名称相似度,使用的算法是基于DTW动态时间序列理论的非语义有序字符串的相似度算法。以A源POI为单位,筛选出名称相似度最大的POI点对。对上一步筛选出的点对的名称相似度进行排序,发现名称相似度大于0.7时,POI点对被认为是被成功匹配的同一POI。将这些点对的所有属性被放入匹配结果数据集 Match1。当匹配方案为将待匹配POI与异源POI以预设的距离值进行组合和匹配时,将空间位置纠正后的所述多源预处理POI数据进行两两组合,如在 A和 $S_i$ 数据集中删除Match1中被匹配的POI,剩余的POI和c类放到同一个数据集d类,再以A源POI为单位与其600m范围内的 $S_i$ 源点对两两组合生成待匹配点对数据集。再计算组合后的点对名称的相似度;其中,所述点对名称为所述多源预处理POI数据进行两两组合后形成的点对的名称;当点对名称的相似度大于预设的相似度阈值时,匹配结果为匹配成功,并将匹配成功的待匹配POI和异源POI存储到目标数据集,如:通过计算POI点对的名称相似度进行匹配。再次筛选名称相似度大于0.7的点对被认为是通过一个POI,它们被存入匹配结果Match1。名称相似度最大值小于0.5的点被认为没有与其能够匹配的点被存入结果No\_Match。

[0092] 得到匹配结果后,就可以执行如图1所示的如下步骤:S400、基于所述匹配结果,对所述多源预处理POI数据进行融合,得到POI融合数据。相应的,步骤S400包括如下步骤:

[0093] S401、当目标数据集中的待匹配POI和异源POI中存在缺失的属性值时,将所述属性值进行补充;

[0094] S402、当目标数据集中的待匹配POI和异源POI的名称不一致时,不执行任何操作;

[0095] S403、当目标数据集中的待匹配POI和异源POI中同类属性值有若干时,将预设的第一属性值代替待匹配POI和异源POI中原有的属性值;

[0096] S404、当目标数据集中的待匹配POI和异源POI中同类属性值具有唯一值时,将预设的第二属性值代替待匹配POI和异源POI中原有的属性值;

[0097] S405、将经过属性值补充、属性值代替后的多源预处理POI数据作为POI 融合数据。

[0098] 具体地,首先判断被匹配POI是否成功匹配上另一个来源的POI,将成功匹配上的POI点对放入数据集Result1,没有匹配到的POI放入数据集 Result2,被匹配源中没有的

POI,但是其他来源中的POI,即缺失的POI 放入数据集Result3,缺失的POI为其它来源中多次都没被匹配到的POI。

[0099] 当目标数据集中的待匹配POI和异源POI中存在缺失的属性值时,将所述属性值进行补充具体为:对于Result1中被匹配POI中缺失的属性,使用属性缺失信息补全方案,也即用其他来源的信息来补充这个空值,直接用匹配点的属性去补充被融合数据集。

[0100] 当目标数据集中的待匹配POI和异源POI的名称不一致时,不执行任何操作为:对于Result1中POI匹配点对的名称冲突,由于在预处理过程中名称已经获得了纠正,故采取冲突忽略策略(即不进行任何操作),将涉及冲突的属性值一并进行记录显示。例如,同一POI在A源和B源的名称为中文名name1和英文名name2,将name1与name2均作为名称属性值进行记录存储。

[0101] 当目标数据集中的待匹配POI和异源POI中同类属性值有若干时,将预设的第一属性值代替待匹配POI和异源POI中原有的属性值具体为:针对POI可以具有多个值的属性,如本发明实施中酒店住宿类POI的商家联系方式可以有多个号码,可以采取冲突消解政策,使用预设的第一属性值(更权威的数据源)去丰富POI属性。

[0102] 当目标数据集中的待匹配POI和异源POI中同类属性值具有唯一值时,将预设的第二属性值代替待匹配POI和异源POI中原有的属性值具体为:针对POI具有唯一值的属性采取冲突避免策略,在本发明实施中的包括POI 空间位置,使用预设的第二属性值(纠正后的空间位置或者更权威数据源如政府数据中的空间位置)作为融合后的POI位置。

[0103] 在一种实现方式中,对于被融合POI中没有的属性,采用属性缺失信息补全方案。如本发明实施中酒店住宿服务类POI缺少的牌照信息和房间数量,直接将来自于政府匹配点的相关信息添加到被融合数据集中。又比如被融合来源的POI的属性中没有商家的联系方式,丰富后被融合来源的 POI数据的属性中就多了商家联系方式的属性信息。

[0104] 在另外一种实现方式中,对比融合前后数据集,评价融合前单源POI 的数量完整性,和融合后数据集POI的数据质量。通过对比融合前被融合源的POI数量和融合后的POI数量,得到融合前的POI数量完整性。然后,评价数据融合后的POI数据质量,包括名称准确性,地址完整性,属性。

[0105] 为了说明本发明实施例提供的POI匹配和融合方法的效果,以将高德地图酒店住宿服务的POI为例,使用源于香港地区的POI数据去匹配和融合高德地图的POI,融合后POI数量上有明显增加,属性值丰富效果显著。

[0106] 如图4给出了POI空间位置纠正前后点的位置,可以看出那些多数位于建筑外的酒店住宿服务POI都被纠正到了建筑内部。

[0107] 表1进一步给出了本发明实施例预处理后的数据质量评估结果。

[0108]

指标	精度
名称准确度	91.4%
地址完整性	42.7%
关联规则完整性	72.6%
空间位置准确性	83%

[0109] 表1

[0110] 表2进一步给出了本发明实施例匹配结果的定量分析。

[0111]

类别	占比
已匹配POI	49%
缺失POI	26%
确定未匹配POI	32.5%

[0112] 表2

[0113] 表3进一步给出了本发明实施例融合后的数据质量评估结果。

[0114]

指标	精度
POI数量完整性	64.4%
名称准确度	80%
地址完整性	42.7%
数据源的时效性	13.7%
空间位置准确性	86.7%

[0115] 表3

[0116] 如图5为POI最终的融合结果,可以发现本发明有效纠正和融合了POI 属性值。

[0117] 示例性设备

[0118] 如图6中所示,本发明实施例提供一种多源POI数据匹配和融合装置,该装置包括多源预处理POI数据获取模块501、空间位置纠正模块502、匹配结果确定模块503和POI融合数据确定模块504,其中:

[0119] 多源预处理POI数据获取模块501,获取多源POI数据,并对所述多源POI 数据进行数据预处理,得到多源预处理POI数据;

[0120] 空间位置纠正模块502,用于基于建筑二维面数据与POI数据的拓扑关系,纠正所述多源预处理POI数据的空间位置;

[0121] 匹配结果确定模块503,用于根据纠正空间位置后的所述多源预处理 POI数据,确定若干匹配方案,并根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行匹配,得到匹配结果;

[0122] POI融合数据确定模块504,用于基于所述匹配结果,对所述多源预处理POI数据进行融合,得到POI融合数据。

[0123] 基于上述实施例,本发明还提供了一种智能终端,其原理框图可以如图7所示。该智能终端包括通过系统总线连接的处理器、存储器、网络接口、显示屏、温度传感器。其中,该智能终端的处理器用于提供计算和控制能力。该智能终端的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该智能终端的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种多源POI数据匹配和融合方法。该智能终端的显示屏可以是液晶显示屏或者电子墨水显示屏,该智能终端的温度传感器是预先在智能终端内部设置,用于检测内部设备的运行温度。

[0124] 本领域技术人员可以理解,图7中的原理图,仅仅是与本发明方案相关的部分结构的框图,并不构成对本发明方案所应用于其上的智能终端的限定,具体的智能终端可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0125] 在一个实施例中,提供了一种智能终端,包括有存储器,以及一个或者一个以上的

程序,其中一个或者一个以上程序存储于存储器中,且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行以下操作的指令:

[0126] 获取多源POI数据,并对所述多源POI数据进行数据预处理,得到多源预处理POI数据;

[0127] 基于建筑二维面数据与POI数据的拓扑关系,纠正所述多源预处理POI 数据的空间位置;

[0128] 根据纠正空间位置后的所述多源预处理POI数据,确定若干匹配方案,并根据若干所述匹配方案,将空间位置纠正后的所述多源预处理POI数据进行匹配,得到匹配结果;

[0129] 基于所述匹配结果,对所述多源预处理POI数据进行融合,得到POI融合数据。

[0130] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本发明所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器 (ROM)、可编程ROM (PROM)、电可编程ROM (EPROM)、电可擦除可编程ROM (EEPROM) 或闪存。易失性存储器可包括随机存取存储器 (RAM) 或者外部高速缓冲存储器。作为说明而非局限,RAM以多种形式可得,诸如静态RAM (SRAM)、动态RAM (DRAM)、同步DRAM (SDRAM)、双数据率SDRAM (DDRSDRAM)、增强型SDRAM (ESDRAM)、同步链路 (Synchlink) DRAM (SLDRAM)、存储器总线 (Rambus) 直接RAM (RDRAM)、直接存储器总线动态RAM (DRDRAM)、以及存储器总线动态RAM (RDRAM) 等。

[0131] 综上所述,本发明公开了多源POI数据匹配和融合方法,所述方法包括:获取多源POI数据,并对多源POI数据进行数据预处理,得到多源预处理POI 数据;基于建筑二维面数据与POI数据的拓扑关系,纠正多源预处理POI数据的空间位置;根据纠正空间位置后的多源预处理POI数据,确定若干匹配方案,并根据若干匹配方案,将空间位置纠正后的多源预处理POI数据进行匹配,得到匹配结果;基于匹配结果,对多源预处理POI数据进行融合,得到POI融合数据;本发明实施例中充分挖掘POI已有信息,对不完整的属性进行了补全、对有冲突的属性进行了纠正;通过名称相似度就能高效匹配点对的方法,解决了由空间位置不准确导致的POI漏配问题,对高密度城市 POI数据尤其有效,可以解决由POI点过于密集导致的过大匹配量问题;纠正了POI错误的空间位置,POI融合结果更优,数据更完整。

[0132] 基于上述实施例,本发明公开了一种多源POI数据匹配和融合方法,应当理解的是,本发明的应用不限于上述的举例,对本领域普通技术人员来说,可以根据上述说明加以改进或变换,所有这些改进和变换都应属于本发明所附权利要求的保护范围。

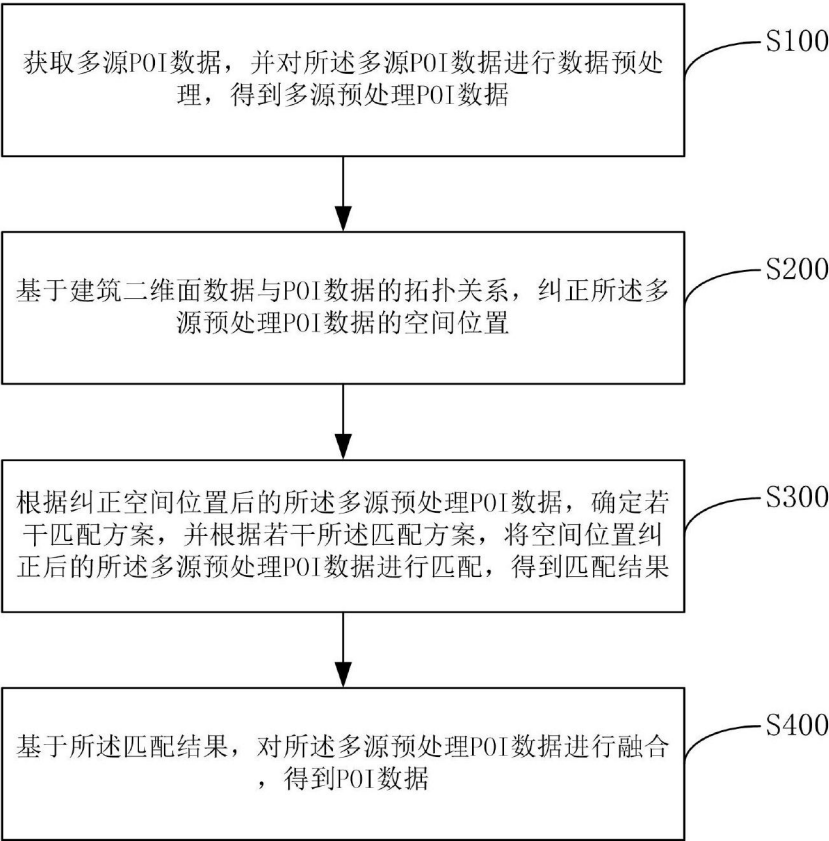


图1

街道名称	街道号	关联规则	建筑
Name1	A-B	若 A,B,C,D,E 均为奇数或者偶数; 当 $C \in [A,B]$ 或 $C \in [D,E]$ 或 $[D,E] \in [A,B]$ 或 $[A,B] \in [D,E]$	同一个建筑
	C		
	D-E		

图2

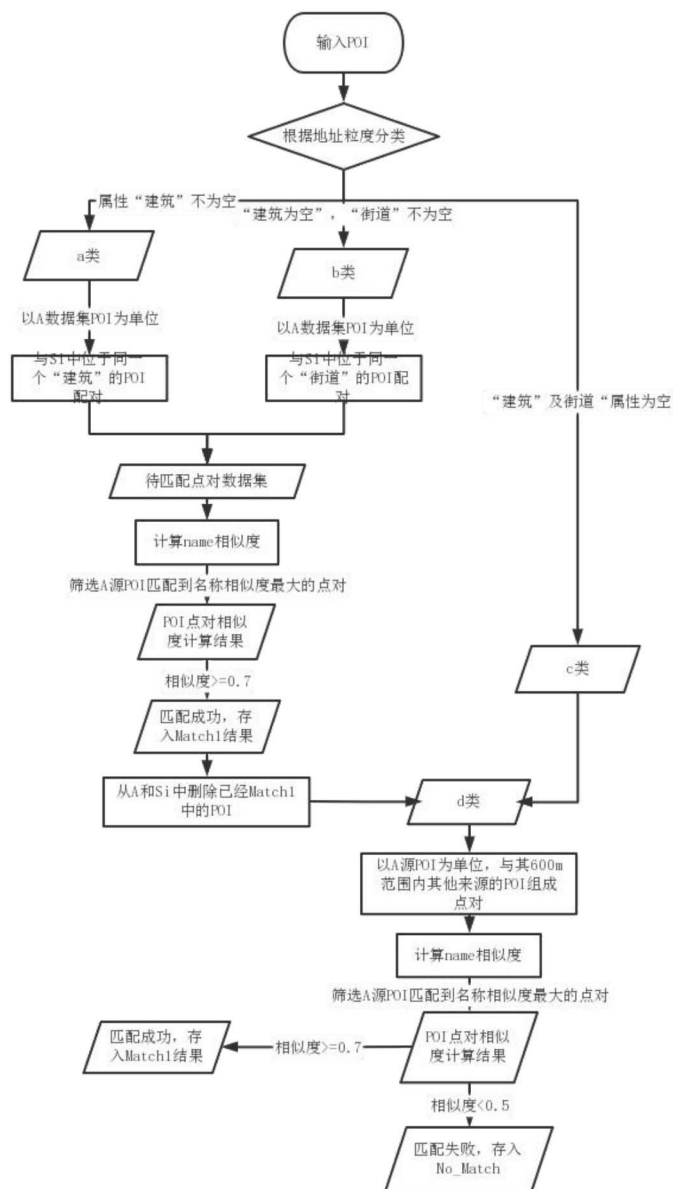


图3



图4

ANAP_ID	name	牌照号	房间数	牌照数	地址	住宿服务类型	地区	lon	lat	联系电话	传真号码
B0FFG0GVYC	遨帆酒店 (THE OLYMPIAN HONG KONG)	H/5813	32	1	大角咀海辉路18号一号银海地下(部分)及地下高层(部分)	宾馆酒店	油尖旺区	114.1559388	22.31775518	3199 8888	3199 8899
B073C04284	朗廷酒店	H/0022	600	1	尖沙咀北京道8号	五星级宾馆	油尖旺区	114.1696155	22.29666087	23751133	23756611
B073C03VRZ	海旁嘉福酒店	H/1003	572	1	尖沙咀东么地道70号	五星级宾馆	油尖旺区	114.1793199	22.29906315	27215161	27322233
B073C03RPF	马哥孖罗港威酒店 (the marco polo gateway)	H/0206	471	1	尖沙咀广东道13号海港城	四星級宾馆	油尖旺区	114.1685285	22.29777158	21130888	21130022
B0FFFA8D11	帝金酒店	H/6106	69	1	尖沙咀赫德道4号-1157号,天台及天台高层	宾馆酒店	油尖旺区	114.1748423	22.29785239	2391 0000	2391 0001
B0FFFS8BTY	大都酒店	H/2793	41	1	尖沙咀金巴利道16号香港大厦B座1字楼A-10室	旅馆招待所	油尖旺区	114.1728275	22.30028196	27398233	23997039
B0FFK7SDQ1	page148精品酒店	H/6457	197	1	尖沙咀柯士甸道148号地下-3字楼,5-12字楼,15-23字楼,25字楼及天台-天台顶层	宾馆酒店	油尖旺区	114.1748266	22.30321695	經營人士未有經營人士未有提供	
B0FF72H453	悠悦宾馆	H/4640	8	1	尖沙咀乐道6号23字楼	旅馆招待所	油尖旺区	114.1714884	22.29687738	23691151	23926999
B0FF1883C2	尚酒店	H/5845	6	1	尖沙咀弥敦道101号海防大厦2楼207室,15字楼1501室	旅馆招待所	油尖旺区	114.1717698	22.29802457	66931388	經營人士未有提供
B073C080FX	美丽华酒店 (THE MIRA HONG KONG)	H/0049	470	1	弥敦道118-130号	五星级宾馆	油尖旺区	114.1724296	22.30027322	23681111	23691788
B073C040EZ	伊丰酒店 (PREBUTAL HOTEL)	H/0153	432	1	尖沙咀弥敦道222号	四星級宾馆	油尖旺区	114.1719637	22.30430086	23118222	23114760
B0FFHLXWV2	hotelvenus	H/5668	7	1	尖沙咀弥敦道40号重庆大厦C座12楼C6室, D座11字楼D7室B座11字楼1室	旅馆招待所	油尖旺区	114.1730896	22.29653456	23667468, 236	2723 7960
B073C0Q0W	金域假日酒店	H/0034	621	1	尖沙咀弥敦道50号	四星級宾馆	油尖旺区	114.1724297	22.29679643	23693111	27221931
B0FF748523	福建酒店	H/4485	4	1	九龙尖沙咀弥敦道68号美丽都大厦A座15字楼A1室	旅馆招待所	油尖旺区	114.1725795	22.29735194	23684525, 2368	23684526, 2778 1776
B0FFFRNB9K	森之旅馆	H/5506	6	1	尖沙咀弥敦道54-56号美丽都大厦8字楼A3室	旅馆招待所	油尖旺区	114.172585	22.29753694	經營人士未有經營人士未有提供	
B0FFGV1U9B	迪高酒店	H/5169, H/5504	12	2	尖沙咀弥敦道80号金堂大厦11字楼B5室, C座10字楼C4室,	旅馆招待所	油尖旺区	114.1723812	22.29857546	6741 5201, 636	23675756
B073C0XWTS	珊瑚酒店	H/5265	6	1	尖沙咀弥敦道87号华源大厦12字楼B4室, 10字楼A室	旅馆招待所	油尖旺区	114.1717698	22.29776268	5993 2220	34936934
B0FFFA6XDA	酒店易	H/4374	9	1	尖沙咀弥敦道87号华源大厦2字楼A1-A2室, C座8字楼C1室, C座9字楼C3室	旅馆招待所	油尖旺区	114.1717698	22.29776268	2368 3886	3691 9020
B0FF1838BE	百佳酒店	H/0424, H/1129, H/2601, H/2780, H/3217	142	5	尖沙咀金巴利道27-33号永利大厦6字楼D室, 湾仔轩尼诗道275-285号立兴大厦地下C楼A1-3字楼	旅馆招待所	湾仔区	114.1734583	22.3008512	2770 1028, 236	2381 0870, 2587 9922
B073C031LY	名都酒店	H/6082, H/4903, H/2697	30	3	旺角弥敦道607号新兴大厦26字楼2605室, 新界荃湾沙咀道242-244号1字楼, 九龙油麻地弥敦道578-580号恒隆大厦2字楼, 26字楼2601室, 1字楼	旅馆招待所	荃湾区	114.1155396	22.3701761	6693 3221, 918	經營人士未有提供
B0FFHSZ29B	君立酒店	none	none	none	九龙九龍彌敦道15号地庫-地下(部份), 1-6字楼, 天台及天台高层(部份)	宾馆酒店	观塘区	114.2057408	22.32367802	25933282	26793222
B0FF1I23A2M	尖沙咀阳光酒店	none	none	none	尖沙咀弥敦道101号海防大厦13字楼1302室, 2字楼A室, 1502室	旅馆招待所	油尖旺区	4N/A	4N/A	23878718	經營人士未有提供
B073C044VL	海都宾馆	none	none	none	尖沙咀弥敦道36-44号重庆大厦A座15字楼A4室, B座15字楼B8室, 3字楼B1室	旅馆招待所	油尖旺区	4N/A	4N/A	27226708	經營人士未有提供
B0FFHLFCHU	香港嘉里酒店	H/6225	546	1	九龙红磡红磡道38号地庫(部份), 地下(部份), 1字楼(部份), 2字楼(部份), 3字楼(部份), 4字楼(部份), 5-6字楼, 7字楼(部份), 8-15字楼和16字楼(部份)	宾馆酒店	九龙城区	114.1885879	22.30117591	25993000	經營人士未有提供

图5

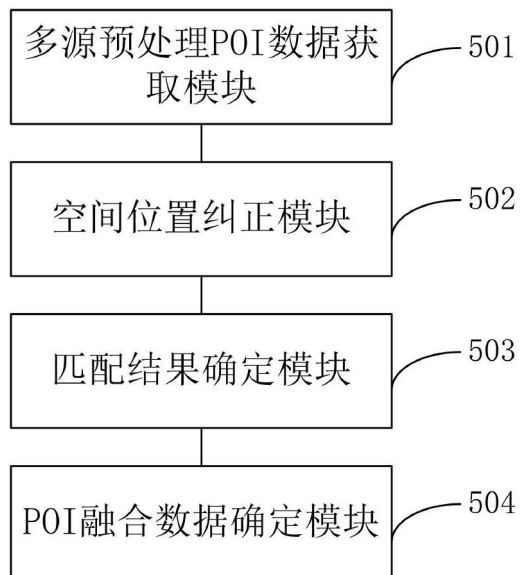


图6

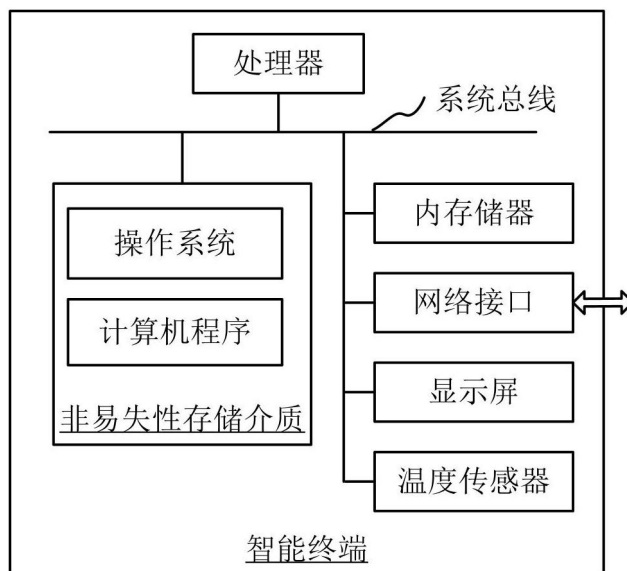


图7