引文格式：张文宇,彭舒凡,叶乃夫. 综合多尺度的重点人多维空间数据关联分析[J]. 地理空间信息, 2023, 21(12): 14-20. doi:10. 3969/j.issn. 1672-4623.2023.12.003

综合多尺度的重点人多维空间数据关联分析

张文宇[1]，彭舒凡[1]，叶乃夫[1]
（1. 中国人民公安大学 信息网络安全学院，北京 100038）

摘　要：在公安数据库中，重点人的空间数据呈现出多维特性，并由空间文本与空间实体构成。在传统的空间关联计算方法中，通常将单一的空间文本关联度或空间实体关联度作为最终结果，缺乏对二者的综合分析，并较少考虑到空间尺度对空间实体关联度的影响。因此，结合空间文本与空间实体2个方面，提出重点人空间数据关联分析方法。其中，对于空间实体，结合空间尺度，提出综合多尺度的空间实体关联度的计算方法，并通过空间实体的空间拓扑关系与空间度量关系量化了每个尺度下的实体关联度。最后，通过实验证明，对于重点人多维空间数据的关联分析，该方法优于传统方法，更能满足公安业务的需求。

关键词：多维空间数据；重点人；空间拓扑关系；空间度量关系；关联分析；空间尺度
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2023）12-0014-07

Integrated Multi-scale Association Analysis of Multi-dimensional Spatial Data for Object Persons

ZHANG Wenyu[1], PENG Shufan[1], YE Naifu[1]
(1.School of Information Technology and Cyber Security, People's Public Security University of China, Beijing 100038, China)

Abstract: The spatial data of object persons in public security database exhibit multi-dimensional characteristics composed of spatial text and entities. Traditional spatial association calculation methods typically adopt either spatial text correlation or spatial entity correlation alone as the final result, while comprehensive analysis of both remains to be improved. Less consideration has been given to the influence of spatial scale on spatial entity relevance. Therefore, we proposed a focus on the human spatial data association analysis method by combining spatial text and spatial entities. We put forward a comprehensive multi-scale calculation of spatial entity relatedness that quantifies entity relatedness under each scale through the spatial topological relationship and spatial metric relationship of spatial entities. Experiments demonstrate that the method is superior to traditional methods for the association analysis of multi-dimensional spatial data of object persons and better meets the needs of public security applications.

Key words: multi-dimensional spatial data, object person, spatial topological relationship, spatial metric relationship, association analysis,spatial scale

犯罪活动在地理空间中展开[1]，因此，犯罪团伙成员之间具有大量的空间交集。公安侦查工作中，分析重点人空间数据之间的关联，是时空关联分析的重要步骤。空间数据关联分析的实质为计算空间数据关联度，根据空间数据的构成，分为2个方面，即空间文本关联度计算与空间实体关联度计算。目前，在空间文本关联度计算方面，主要包括：①计算文本类别关联度[2-7]；②计算文本专名关联度[8-10]；③综合计算类别与专名的关联度，得到空间文本的关联度[11]。公安数据库中，空间文本具有的文本类别及其专名通常超过一种，如“某省某市某区某街道”，然而已有研究认为空间文本只具有一种文本类别及其专名，不符合公安业务需求。

空间尺度包含空间范围与空间粒度[12]。计算空间实体关联度时所选的空间尺度，是影响结果的重要因素[13-14]，同一地理实体，在不同的尺度下得出的结论可能相差甚远[15]。在空间实体关联度计算方面，传统方法[16-20]主要基于空间拓扑关系计算。其中，赵红伟[16]在此基础上，使用空间度量关系辅助量化空间实体关联度，提高了计算的准确性。以上方法都较少考虑到空间尺度对计算结果的影响。陈祖刚[21]虽然考虑到空间尺度的影响，但只是分别计算了2个尺度下的关联度，并没有综合考虑各个尺度的影响。

综上，本文将重点人空间数据的关联分析分为2个方面：计算空间文本关联度与空间实体关联度。针对现有研究的不足，对于重点人的空间文本关联

收稿日期：2023-08-28。
第一作者简介：张文宇（1997—），硕士研究生，研究方向为警务时空数据分析，E-mail：1187907839@qq.com。

度，结合文本专名与文本类别2个多维特征，综合计算文本关联度。对于重点人的空间实体关联度，考虑到各个尺度对实体关联度的影响，通过空间实体之间的拓扑关系及度量关系对各尺度下的实体关联度进行量化，挖掘重点人与犯罪人在空间上的关联信息，提出了综合多尺度空间实体关联度的计算方法。

1　多维空间数据的表达

重点人空间数据多维特征的表达，是空间数据关联分析的基础。重点人的多维空间数据，具有文本专名、文本类别、实体几何类型3个多维特征，使用三元组表达为：

$$s=[P,C,G] \tag{1}$$

1）P 为空间数据的文本专名，其值为空间文本的具体描述。

2）C 为空间数据的文本类别。若空间文本由单一的文本类别及其专名构成，则 C 的取值为该类；若空间文本由多种文本类别及其专名构成，则 C 的取值为最后一个文本类别。

空间数据的文本类别体现了空间实体的类型、隶属关系。本文根据《地名分类与类别代码编制规则（GB/T 18521—2001）》[22]，并基于上下位关系（IS-A），建立了轻量级的上层空间文本类别本体，涵盖了重点人空间数据的全部文本类别，如图1所示。

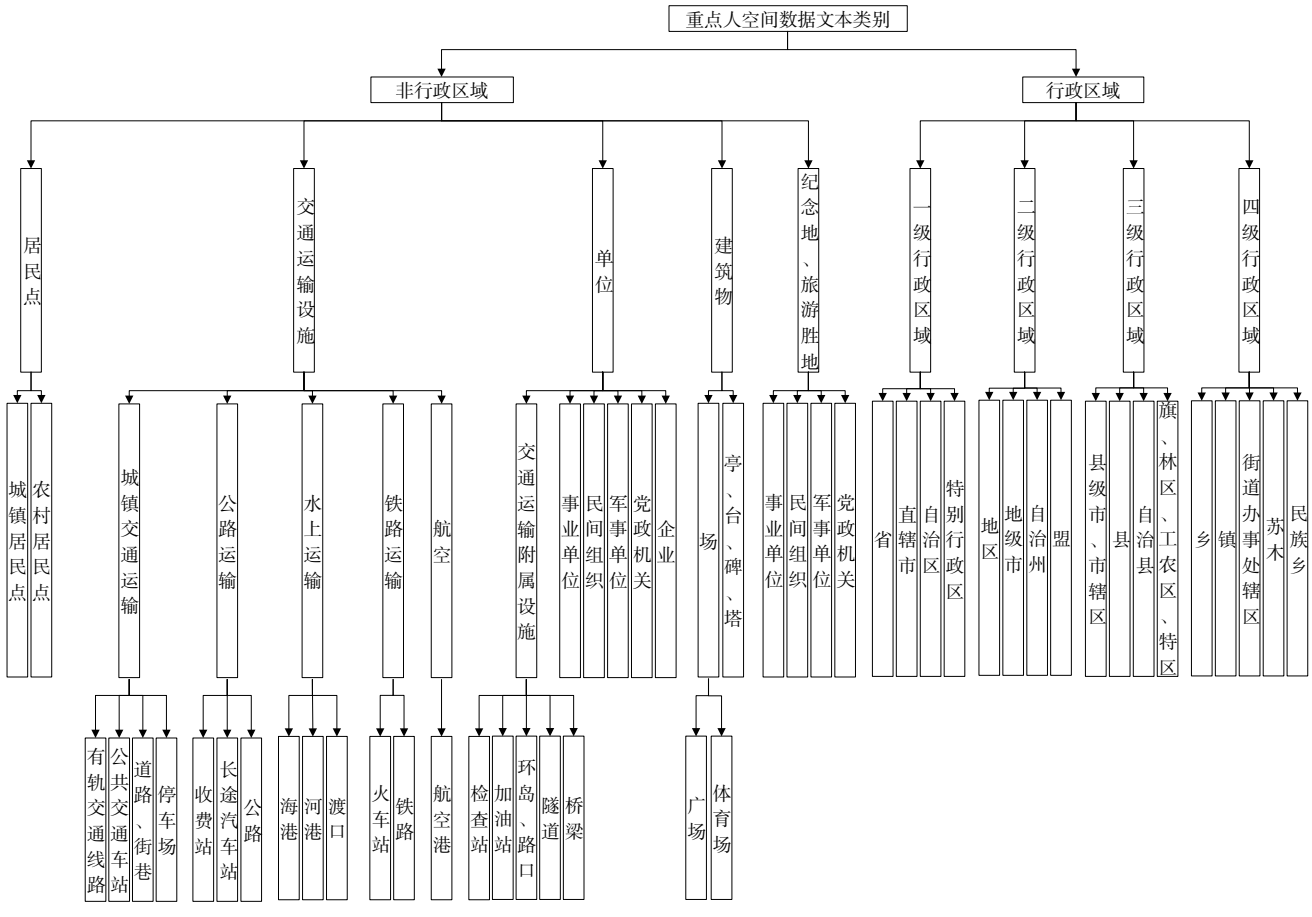


图1　轻量级的上层重点人空间数据的文本类别本体

3）G 为空间数据的实体几何类型，其取值存在3种情况：点（point）；线（line）；面（polygon）。各个类别具有固定的实体几何类型。

2　多维空间数据的关联分析

空间数据关联分析，就是计算出空间数据之间的关联度。由于重点人的多维空间数据由空间文本与空间实体构成，因此，将空间文本关联度与空间实体关联度结合，得到空间关联度：

$$C_{\text{Geo}}=W_{\text{Text}}C_{\text{Text}}+W_{\text{Entity}}C_{\text{Entity}} \tag{2}$$

式中，C_{Geo} 为空间关联度；C_{Text} 为空间文本关联度；C_{Entity} 为空间实体关联度；W_{Text} 与 W_{Entity} 分别为空间文本与空间实体关联度的权重值，满足 $W_{\text{Text}}+W_{\text{Entity}}=1$。关联分析的流程如图2所示。先结合专名关联度与类别关联度计算空间文本关联度。同时，通过空间拓扑关系与空间度量关系量化每个尺度

下的实体关联度，再将各尺度下的空间实体关联度综合，求得空间实体关联度。最后，通过上式将二者加权求和，得到重点人空间数据的关联度。

在公安领域的应用中，若重点人与犯罪人对应的空间实体不关联，即二者不存在于同一空间范围内，说明二者共同实施犯罪的可能性较低，不为犯罪团伙。此时无论文本关联度 C_{Text} 是否为0，均认为2个重点人在空间上不关联，即空间关联度 C_{Geo} 为0。

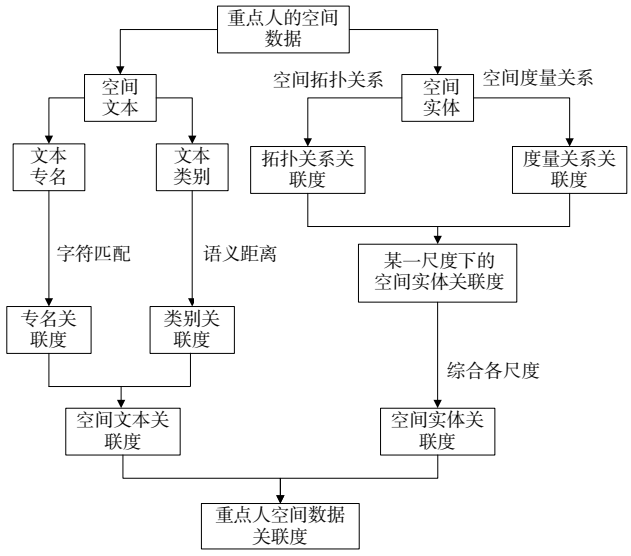


图2 重点人多维空间数据关联分析流程

2.1 空间文本关联度计算

空间文本具有文本专名与文本类别2个多维特征，将其多维特征的关联度综合计算，得到空间文本关联度：

$$C_{\text{Text}} = W_{\text{P}} C_{\text{P}} + W_{\text{C}} C_{\text{C}} \tag{3}$$

式中，C_{P} 与 C_{C} 分别为专名关联度与类别关联度；W_{P} 与 W_{C} 分为为空间专名关联度与空间类别关联度的权重，满足 $W_{\text{P}} + W_{\text{C}} = 1$。

2.1.1 文本专名的关联度计算

本文将文本专名看作字符串，使用叶鹏[8]提出的方法计算文本专名关联度。设2个空间数据的文本专名分别为 p_1 与 p_2，字符总数分别为 m、n，其关联度计算如下：

$$C_P(p_1,p_2) = \alpha \times \frac{1}{2}\left(\frac{c}{m}+\frac{c}{n}\right) + \beta \times \min\left(\frac{m}{n},\frac{n}{m}\right) \times \frac{1}{2}\left(\frac{\sum_{i=1}^{c} L_1(i)}{\sum_{t=1}^{m} t} + \frac{\sum_{i=1}^{c} L_2(i)}{\sum_{k=1}^{n} k}\right) \tag{4}$$

式中，c 为 p_1 与 p_2 的匹配字符数；$L_1(i)$ 与 $L_2(i)$ 分别为匹配字符 i 在 p_1 与 p_2 中的匹配序；α 与 β 分别为匹配度与匹配序的权重，二者常按黄金分割率，分别取值为0.6与0.4[8–9]。匹配序从左至右，根据字符依次递增。例如，p_1 =“北京市西城区”，p_2 =“西城区”，p_1 与 p_2 的匹配字符为“西”、“城”、“区”，在 p_1 与 p_2 中的匹配序分别为4（西）、5（城）、6（区）和1（西）、2（城）、3（区）。p_1 与 p_2 的关联度定义为：

$$C_P(p_1,p_2) = 0.6 \times \frac{1}{2}\left(\frac{3}{6}+\frac{3}{3}\right) + 0.4 \times \min\left(\frac{6}{3},\frac{3}{6}\right) \times \frac{1}{2}\left(\frac{4+5+6}{1+2+3+4+5+6} + \frac{1+2+3}{1+2+3}\right) = 0.62 \tag{5}$$

2.1.2 文本类别的关联度计算

当前研究主要通过建立本体，计算空间文本类别的关联度。根据方法细分为：基于信息论、基于语义距离、基于本体属性。基于信息论的方法，以概念出现的频率量化关联度；基于本体属性的方法需对文本类别进行严格属性定义，这两类方法均不适用于空间数据的文本类别的关联度计算，因此，本文基于第1节建立的轻量级上层本体，并采用基于语义距离的方法[23]进行计算。其中，影响关联度的主要因素分别为：语义距离、概念深度、概念密度。设2个空间数据的文本类别分别为 c_1 与 c_2。

在本体中，连接 c_1 与 c_2 的最短路径的长度为二者之间的语义距离，记为 $\text{len}(c_1, c_2)$。语义距离越大，c_1 与 c_2 的关联度越低；语义距离为0时，c_1 与 c_2 是同一类别；语义距离为无穷大时，c_1 与 c_2 的关联度为0，表示 c_1 与 c_2 没有关联。语义距离对类别关联度的影响因子记为 $\text{Dis}(c_1, c_2)$，公式如下：

$$\text{Dis}(c_1, c_2) = \frac{1}{1+\text{len}(c_1, c_2)} \tag{6}$$

文本类别 c_1 的概念深度是指 c_1 与根节点的最短路径中所包含的边数，记为 $\text{Dep}(c_1)$，同理，文本类别 c_2 的概念深度记为 $\text{Dep}(c_2)$。在本体中，下层概念都是对上层概念的细化，概念的含义随着深度的增加而具体。语义距离相等时，c_1 与 c_2 的深度越大，其关联度越高；反之越低。概念深度对类别关联度的影响因子记为 $\text{Dep}(c_1, c_2)$，公式如下：

$$\text{Dep}(c_1, c_2) = \frac{\text{Dep}(c_1)+\text{Dep}(c_2)}{\left|\text{Dep}(c_1)-\text{Dep}(c_2)\right|+1} \tag{7}$$

c_1 与 c_2 的概念密度，为 c_1 与 c_2 最近共同祖先的直接子节点的数量，记为 $\text{degree}(c_1, c_2)$。在本体中，某节点的直接子节点数量越多，对应概念的分类越详尽，其子节点间的关联度越高，反之越低。概念密度对类别关联度的影响因子记为 $\text{Den}(c_1, c_2)$，公式如下：

$$\mathrm{Den}(c_1, c_2)=\frac{\mathrm{degree}(c_1, c_2)}{\mathrm{degree}(O)} \tag{8}$$

式中，degree(O) 为本体树 O 中各节点的直接子节点数量的最大值。

因此，c_1 与 c_2 的类别关联度计算如下：

$$C_c(c_1, c_2)=(\mathrm{Dis}(c_1, c_2))^{\delta}\times(\mathrm{Dep}(c_1, c_2))^{\varepsilon}\times(\mathrm{Den}(c_1, c_2))^{\varphi} \tag{9}$$

式中，δ、ε、φ 分别为3个因子对文本类别关联度影响的权重值，且 $\delta+\varepsilon+\varphi=1$。计算类别关联度时，语义距离在3个因素起主要作用，因此 δ 相对较大，ε 与 φ 相对较小，通常取为0.9、0.05、0.05[23-24]。

2.2　空间实体关联度计算

考虑到空间尺度对空间实体关联度的影响，本文将空间尺度分为5个，从大到小依次为："一级行政区域尺度"、"二级行政区域尺度"、"三级行政区域尺度"、"四级行政区域尺度"、"其他区域尺度"，与空间文本类别中的"一级行政区域"、"二级行政区域"、"三级行政区域"、"四级行政区域"、"非行政区域"及其子类别相对应。具有多种文本类别的重点人的空间数据，在不同尺度下，映射为各自对应的空间实体。如"贵州省安顺市西秀区"，在"一级行政区域尺度"时，映射为"贵州省"的空间实体；"二级行政区域尺度"时，映射为"安顺市"的空间实体；"三级行政区域尺度"时，映射为"西秀区"的空间实体；"四级行政区域尺度"、"其他区域尺度"时，无可映射的空间实体。

只具有部分尺度所对应文本类别的空间文本，可能存在同名不同地的情况；具有全部尺度所对应的文本类别的空间文本，其映射的空间实体唯一。故空间实体的关联度，应由5个尺度下的空间实体关联度加权计算得到：

$$C_{\mathrm{Entity}}=W_{\mathrm{First}}C_{\mathrm{First}}+W_{\mathrm{Second}}C_{\mathrm{Second}}+W_{\mathrm{Third}}C_{\mathrm{Third}}+W_{\mathrm{Fourth}}C_{\mathrm{Fourth}}+W_{\mathrm{Other}}C_{\mathrm{Other}} \tag{9}$$

式中，C_{First}、C_{Second}、C_{Third}、C_{Fourth}、C_{Other} 分别为5种尺度下空间数据所对应空间实体关联度，通过量化各尺度下的空间拓扑关系与空间度量关系得到；W_{First}、W_{Second}、W_{Third}、W_{Fourth}、W_{Other} 分别为5种尺度下的实体关联度对应的权重值，满足 $W_{\mathrm{First}}+W_{\mathrm{Second}}+W_{\mathrm{Third}}+W_{\mathrm{Fourth}}+W_{\mathrm{Other}}=1$。

2个空间数据在计算关联度时存在以下情况：①2个空间数据均具有某一尺度所对应的文本类别，则直接计算两者在该尺度下的实体关联度；②2个空间数据均无某一尺度所对应的文本类别，则在该尺度下的实体关联度为0；③2个空间数据中只有一个具有某一尺度所对应的类别，则该尺度下的实体关联度需要跨尺度计算。如"贵州省"与"贵州省安顺市"，只有后者具有"二级行政区域尺度"所对应的文本类别"地级市"，因此，该组数据在"二级行政区域尺度"下的实体关联度为："贵州省"与"安顺市"对应实体的关联度。

某一尺度下的空间实体关联度由该尺度下的空间拓扑关系关联度与空间度量关系关联度加权得到：

$$C_{\mathrm{S}}=W_{\mathrm{T}}C_{\mathrm{T}}+W_{\mathrm{M}}C_{\mathrm{M}} \tag{10}$$

式中，C_{S} 为某一尺度下的空间实体关联度，可为 C_{First}、C_{Second}、C_{Third}、C_{Fourth}、C_{Other}；C_{T}、C_{M} 分别为对应尺度下的空间拓扑关系关联度与空间度量关系关联度；W_{T}、W_{M} 分别为拓扑关系关联度与度量关系关联度对应的权重值，且满足 $W_{\mathrm{T}}+W_{\mathrm{M}}=1$。

2.2.1　空间拓扑关系的关联度计算

空间实体具有点、线、面3种实体几何类型，空间拓扑关系用于描述3种实体几何类型之间的关系。目前，普遍采用4I模型[25]与9I模型[26]进行描述。由于4I模型对于线线之间、线面之间的关系描述不具备唯一性[27]，因此本文采用9I模型对实体之间拓扑关系进行表达，任意2种实体几何类型的空间实体之间的拓扑关系如表1所示。

表1　空间实体拓扑关系

拓扑关系	点点	点线	点面	线线	线面	面面
相离（Disjoints）	• •	•	•			
相接（Touches）	***					
相交（Crosses）	***	***	***			***
重叠（Overlaps）	***	***	***			
相等（Equals）	•	***	***		***	
包含/被包含（Contains/Within）	***					

注：***表示该类型的空间拓扑关系不存在。

若重点人与犯罪人的空间数据，在某一尺度上对应空间实体的拓扑关系为相离，即在该尺度下二者不存在于同一空间范围内，所对应的空间实体不关联，此时 C_S 为0；反之，若拓扑关系不为相离，即在该尺度下二者存在于同一空间范围内，所对应的空间实体存在关联，此时 C_T 为1，C_M 通过计算得到；若2个重点人空间数据，在各尺度下的拓扑关系均为相离，则空间实体关联度 C_{Entity} 为0，故空间关联度 C_{Geo} 也为0。

2.2.2 空间度量关系的关联度计算

空间度量关系用于辅助空间拓扑关系，更详细地描述空间实体之间的关系[28]。对空间度量关系关联度的量化分为2个方面，一是空间实体的重叠，二是空间实体的距离。相关定义如下：

定义1，空间实体的重叠：2个空间实体重叠部分的长度/面积。

定义2，空间实体的距离：2个空间实体之间的距离。

实体几何类型为点–点、点–面、面–面的2个空间实体，其距离指实体几何中心之间的欧氏距离；实体几何类型为点–线、线–面的2个空间实体，其距离指点和面的几何中心到线的最短距离；实体几何类型为线–线的2个空间实体，其距离指实体之间的最短距离。

定义3，某一尺度下实体关联度的基本权重：在各尺度下，2个空间实体为某种拓扑关系时，该拓扑关系能决定的关联度的最大值，即 W_T。

定义4，某一尺度下实体关联度的控制权重：在各尺度下，2个空间实体为某种拓扑关系时，空间度量关系能决定的关联度变化的最大值，即 W_M。

将空间实体在某一尺度下的度量关系关联度记为 C_M；重叠记为 O；距离记为 D。C_M 与 O、D 有关，重叠 O 的占比越高，C_M 越大；根据地理学第一定理，距离 D 越小则关联度越高，同时距离对度量关系关联度的影响与尺度相关。空间度量关系在某一尺度下的关联度 C_M 计算公式如下：

$$C_M = W_O\left(W_1\frac{O}{E_1} + W_2\frac{O}{E_2}\right) + W_D\left(1 - \frac{D}{D+S}\right) \quad (11)$$

式中，E_1、E_2 为2个空间实体在该尺度下的周长/面积；S 为该尺度的尺度因素，各尺度下的 S 取值不一；W_1、W_2 为重叠部分分别在2个实体中的占比所对应的权重值，满足 $W_1 + W_2 = 1$；W_O、W_D 为空间实体的重叠与距离的权重值，满足 $W_O + W_D = 1$。

空间实体的几何类型与空间实体之间的拓扑关系，决定了2个空间实体的重叠，进而影响度量关系关联度的计算。

1）2个空间实体的重叠为点时，有3种可能：①2个空间实体中存在几何类型为点的实体，并且2个空间实体具有交集；②2个几何类型为线的空间实体，拓扑关系为Touches或Crosses；③2个几何类型为线–面的空间实体，拓扑关系为Touches或Crosses或Overlaps。此时，重叠O不具有实际意义，则度量关系的关联度由距离 D 控制。

2）2个空间实体的重叠部分为线时，有2种可能：①2个几何类型为线的空间实体，拓扑关系为Overlaps或Equals或Contains/Within；②2个几何类型为面的空间实体，拓扑关系为Touches。此时，度量关系关联度由重叠O与距离 D 共同控制，O取重叠的长度，E_1、E_2 分别取2个实体的周长。

3）2个空间实体的重叠为面时，2个几何类型为面的空间实体，拓扑关系为Overlaps或Equals或Contains/Within。此时，度量关系关联度由重叠O与距离 D 共同控制，O取重叠的面积，E_1、E_2 分别取2个实体的面积。

3 实验与分析

2021年某日，某市公安局接到报警：在该市某小区，某住户的财物遭到盗窃。经过犯罪现场勘验，发现是规模为两人的团伙作案，并已锁定犯罪嫌疑人。该犯罪嫌疑人经常于“GZ省GY市NM区HGY街道”一带活动。本文从公安重点人数据库中，抽取出100名重点人经常到访的空间数据。以这100名重点人为例，本文将其空间数据分别与“GZ省GY市NM区HGY街道”进行空间关联分析，得到与犯罪人的空间关联度，为侦查提供思路。

本文邀请地理科学、犯罪学领域专家对多维空间数据关联分析方法中的各权重进行打分，平均结果如表2所示。

表2 本文权重取值表

权重	值	权重	值
W_{Text}	0.2	W_{Other}	0.27
W_{Entity}	0.8	W_T	0.7
W_P	0.6	W_M	0.3
W_C	0.4	W_O	0.6
W_{First}	0.12	W_D	0.4
W_{Second}	0.16	W_1	0.5
W_{Third}	0.2	W_2	0.5
W_{Fourth}	0.25	W_{Other}	0.27

由于篇幅限制，将计算结果中关联度大于等于0.2的重点人及其空间数据排序，如表3所示，其中，将空间文本关联度与根据文献[16]方法计算得到的空间实体关联度结合，结果作为传统方法的空间关联度，与本文方法进行对比分析。

由表3可知，本文方法综合了空间数据在各尺度下的关联度，只有在5个尺度下均一致时，空间关联度为1。同时，在各尺度下实体关联度的计算中，本文更加注重拓扑关系的影响，而度量关系起辅助作用。在具有空间关联时，与空间实体具有映射关系的空间文本描述越细致，即空间文本具有的文本类别越多，其关联度可能越高；反之，越低，与事实相符。如，65、5、92号重点人的空间文本均对应同一空间实体，但描述的细致程度不同，其关联度分别为：0.651、0.626、0.598。在相同拓扑关系下，重叠的占比越大、距离越近，关联度越高。

文献[16]的传统方法，没有考虑到空间尺度对关联度的影响，只计算2个空间数据在其最小尺度下的关联度，并更加关注实体之间距离对关联度的影响。因此，该方法面对即便是相离的2个空间数据，在最小尺度下其关联度也可能较高。如，48号重点人的空间数据与目标空间数据在“四级行政区域尺度”上相离，50号重点人的空间数据与目标空间数据在“四级行政区域尺度”上包含，因此50号重点人的关联度理应大于48号重点人的关联度，但传统方法却得到了大相径庭的结论。同时，传统方法在

表3　与空间数据“GZ省GY市NM区HGY街道”关联度≥0.2的重点人空间数据排序

重点人编号	空间数据	关联度（本文方法）	关联度（传统方法）
76	[GZ省GY市NM区HGY街道SZT,塔,polygon]	0.875	0.699
23	[GZ省GY市NM区HGY街道,街道办事处辖区,polygon]	0.784	1.000
44	[HGYSD公园,公园,polygon]	0.737	0.736
50	[HGYL1区,城镇居民点, polygon]	0.735	0.707
19	[HGY街道,街道办事处辖区, polygon]	0.685	0.942
65	[GZ省GY市NM区,市辖区,polygon]	0.651	0.784
5	[GY市NM区,市辖区,polygon]	0.626	0.768
92	[NM区,市辖区,polygon]	0.598	0.753
22	[GY市,地级市,polygon]	0.582	0.757
48	[GZ省GY市NM区XY街道,街道办事处辖区, polygon]	0.555	0.708
97	[GZ省GY市NM区XHL街道,街道办事处辖区, polygon]	0.554	0.546
26	[GZ省GY市YY区WQ街道,街道办事处辖区, polygon]	0.509	0.595
86	[GY市WD区GXL街道,街道办事处辖区, polygon]	0.478	0.272
79	[GZ省GY市WD区,市辖区,polygon]	0.433	0.636
12	[GZ省GY市HX区,市辖区,polygon]	0.433	0.622
40	[GZ省GY市YY区QNS公园,公园,polygon]	0.420	0.420
57	[GZ省GY市DCSD广场,广场, polygon]]	0.415	0.235
61	[GY市HB公园,公园,polygon]	0.407	0.513
53	[GY市HX区,市辖区,polygon]	0.406	0.604
72	[GY市YY区HX路,道路,line]	0.393	0.424
39	[GY火车站,火车站, polygon]	0.380	0.499
70	[GZ省JXSWL,企业, point]	0.379	0.223
35	[YY区SB路,道路,line]	0.358	0.082
29	[HX区QH路,道路,line]	0.350	0.181
83	[GY市MW街道,街道办事处辖区, polygon]	0.347	0.599
8	[JXL,风景名胜区, point]	0.327	0.338
55	[GZ省GY市BY区KC路,道路,line]	0.290	–0.427
17	[GZ省GY市GSH区YCD路,道路,line]	0.287	0.366
74	[GSH公园,公园,polygon]	0.232	0.297
37	[DY路,道路,line]	0.216	0.269

量化空间实体的度量关系关联度时，只要2个空间数据对应的空间实体距离较远，就会存在关联度可能为负的不足。如55号重点人使用传统方法计算得到的关联度为–0.427。

4 结 论

本文将重点人的空间数据定义为空间文本与空间实体，并在此基础上定义了空间文本的文本专名、文本类别、空间实体的实体几何类型等3个多维特征及其表达。其次，提出了多维空间数据关联分析的方法，通过计算空间文本关联度和空间实体关联度，再将其结合得到空间数据的关联度。采用字符串相似度的方法计算文本专名的关联度，并建立了重点人空间数据文本类别的轻量级上层本体，以计算文本类别关联度，由此得到空间文本关联度。本文考虑了空间尺度对实体关联度的影响，认为空间实体关联度由2个实体在各个尺度下的实体关联度加权得到。并且空间实体的拓扑关系和度量关系决定了实体关联度的大小，因此通过拓扑关系和度量关系量化了每个尺度下的实体关联度。

参考文献

[1] 史安,刘庆芳,宋金平. 北京市重大刑事犯罪时空分布特征与影响因素分析[J]. 地理空间信息,2023,21(7):23-26,64

[2] 李红梅,翟亮,朱煜. 基于本体的地理空间实体类型语义相似度计算模型的研究[J]. 测绘科学,2009,34(2):12-14,24

[3] 杨娜娜,张青年,牛继强. 基于本体结构的空间实体语义相似度计算模型[J]. 测绘科学,2015,40(3):107-111,84

[4] 宋佳,王卷乐,诸云强,等. 基于地理空间本体的语义检索相关度研究[J]. 计算机工程与应用,2011,47(5):114-117

[5] 索俊锋,郑海晨. 基于地名本体的加权语义相关度算法研究[J]. 西北民族大学学报(自然科学版),2019,40(3):1-8,25

[6] 赵云鹏,孙群,刘新贵,等. 面向地理实体的语义相似性度量方法及其在道路匹配中的应用[J]. 武汉大学学报(信息科学版),2020,45(5):728-735

[7] 陈永佩,杜震洪,刘仁义,等. 一种引入实体的地理语义相似度混合计算模型[J]. 浙江大学学报(理学版),2018,45(2):196-204

[8] 叶鹏,张雪英,杜咪. 顾及字符特征的中文地名词典查询方法[J]. 地球信息科学学报,2018,20(7):880-886

[9] 张雪英,闾国年. 基于字面相似度的地理信息分类体系自动转换方法[J]. 遥感学报,2008,12(3):433-441

[10] 廖一兰,王劲峰,马家奇,等. 基于BPM-BM算法的地名数据匹配[J]. 测绘通报,2008(6):22-25

[11] 程钢,卢小平. 顾及通名语义的汉语地名相似度匹配算法[J]. 测绘学报,2014,43(4):404-410,418

[12] João E. The Tmportance of Data and Scale Issues for Strategic Environmental Assessment (SEA)[J]. Environmental Impact Assessment Review, 2007,27(5):361-364

[13] 薛存金. 基于空间尺度的海洋信息提取[D]. 武汉:武汉大学,2005

[14] Robinson W S. Ecological Correlations and the Behavior of Individuals[J]. International journal of Epidemiology, 2009,38(2):337-341

[15] 魏高磊,吴太夏,王树东,等. 基于无人机数据的植被指数空间尺度效应研究[J]. 地理空间信息,2021,19(4):4-9,40

[16] 赵红伟,诸云强,杨宏伟,等. 地理空间数据本质特征语义相关度计算模型[J]. 地理研究,2016,35(1):58-70

[17] 吴烨,陈荦,熊伟,等. 面向高效检索的多源地理空间数据关联模型[J]. 计算机学报,2014,37(9):1 999-2 010

[18] 蓝秋萍. 地图数据多尺度级联更新方法研究[D]. 武汉:武汉大学,2010

[19] 张桥平,李德仁,龚健雅. 城市地图数据库面实体匹配技术[J]. 遥感学报,2004,8(2):107-112

[20] Li B, Fonseca F. TDD:A Comprehensive Model for Qualitative Spatial Similarity Assessment[J]. Spatial Cognition and Computation, 2006, 6(1):31-62

[21] 陈祖刚,杨雅萍. 耦合尺度的地理实体空间相关度算法的建立与应用[J]. 地球信息科学学报,2018,20(1):37-47

[22] 中华人民共和国民政部. 地名分类与类别代码编制规则:GB/T18521-2001[S].

[23] 刘景方,邹平,张朋柱,等. 一种改进的本体概念语义相似度算法研究[J]. 武汉理工大学学报,2010,32(20):112-117

[24] 陈沈焰,吴军华. 基于本体的概念语义相似度计算及其应用[J]. 微电子学与计算机,2008,25(12):96-99

[25] Egenhofer M J, Franzosa R D. Point-set Topological Spatial Relations[J]. International Journal of Geographical Information System,1991,5(2):161-174

[26] Egenhofer M J, Herring J. Categorizing Binary Topological Relations Between Regions Lines and Points in Geographic Databases the 9-intersection:Formalism and Its Use for Naturallanguage Spatial Predicates[J]. Santa Barbara CA National Center for Geographic Information and Analysis Technical Report, 1994,1(1):94-102

[27] 李成名,陈军. 空间关系描述的9-交模型[J]. 武汉测绘科技大学学报,1997,22(3):21-25

[28] 郭庆胜,杜晓初,刘浩. 空间拓扑关系定量描述与抽象方法研究[J]. 测绘学报,2005(2):123-128