

顾及排序差异的对象群邻近关系相似性计算

孟妮娜¹ 艾廷华² 周校东³

(1 长安大学地质工程与测绘学院,西安市雁塔路 126 号,710054)

(2 武汉大学资源与环境科学学院,武汉市珞喻路 129 号,430079)

(3 西安测绘研究所,西安市雁塔路 1 号,710054)

摘要:针对空间邻近关系相似性的定量化计算问题,在确定邻近特征量的基础上,将相似性度量的要素模型和对准模型结合起来,建立了顾及排序差异的对象群邻近关系相似性计算方法,并验证了该方法的可用性。

关键词:邻近关系;相似性;排序差异;数据匹配;Delaunay 三角网

中图分类号:P208

存在于相离目标之间的几何邻近关系是空间关系分析和计算的重要内容。Delaunay 三角网及其对偶 Voronoi 图能够直接描述空间目标之间的相对位置关系,并形成对空间的一种剖分,是空间邻近问题的重要分析工具^[1]。众多学者定义了直接邻近、最邻近、侧向邻近^[2]、 k 阶 Voronoi 邻近^[3-4]、Delaunay 邻近^[5]等概念,在很多涉及空间邻近问题的分析和研究中都运用了 Delaunay 三角网和 Voronoi 图^[1,3,6]。

空间关系相似性的判断是空间认知的基本内容^[7]。已有空间关系相似性度量多采用定性的“概念邻域图”^[8-12]。从相似性的计算方法来看,Schwering 将对象及其特征层次上的相似性度量模型归纳为 5 种:几何模型、要素模型、网络模型、对准模型、转换模型^[13-14],而对对象群空间关系的相似性是对象与对象之间结构上的、比对象本身更高层次的比较^[11],这些通用模型的适用性还需要进一步研究。

本文以 Delaunay 三角网为工具,以对象和特征层次上的要素模型为基础,结合对准模型的思想,建立了顾及排序差异的对象群邻近关系相似性的计算方法。

1 基于相同特征量的邻近关系相似性计算

相似性度量中有两个主要概念:一是两个对象或概念之间的共性和差异,另一个是对象或概念之间的距离^[13]、空间关系的“概念邻域图”实质上就是运用两个空间关系之间的距离来度量相似程度。相似性度量的要素模型就是基于这种思想,最常用的是 Tversky 等^[15]提出的要素匹配模型。如图 1 所示, A 、 B 分别表示概念 a 、 b 的集

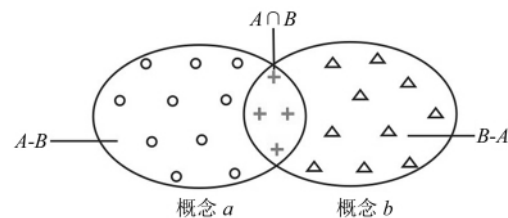


图 1 通过集合操作度量概念 a 和 b 的相似性
Fig. 1 Measures the Similarity of Concept a and b by Applying Set-theoretic Operations

合要素,则 a 和 b 的相似性可用的集合操作有 $A - B$ 、 $A \cap B$ 、 $B - A$,计算公式为:

$$s(a, b) =$$

收稿日期:2013-01-28。

项目来源:国家自然科学基金资助项目(41001229);中央高校基本科研业务费专项资金资助项目(CHD2012JC053);数字制图与国土应用工程国家测绘地理信息局重点实验室开放研究基金资助项目(GCWD201109)。

$$\frac{f(A \cap B)}{\theta * f(A \cap B) + \alpha * f(A - B) + \beta * f(B - A)} \quad (1)$$

其中,函数 f 是对集合取基数或者是集合中各种要素显著的、突出功能的反映。可见,两事物相同特征量的多少决定了它们的相似程度。

假设空间分布不同的两个对象群具有相同的对象类型和对象数目,以某对象的 k 阶邻近关系

$$S_k = \frac{f(k_A \cap k_B)}{\theta * f(k_A \cap k_B) + \alpha * f(k_A - k_B) + \beta * f(k_A - k_B)} \quad (2)$$

θ, α, β 分别为集合 $f(k_A \cap k_B), f(k_A - k_B), f(k_A - k_B)$ 的权重。

对象 P 的总体空间邻近关系相似度 Sim_P 由各阶邻近关系相似度共同决定。实际中,可根据具体情况给各阶邻近关系相似度赋以不同的权:

$$\text{Sim}_P = \frac{\sum_{k=1}^n W_k \times S_k}{k} \quad (3)$$

实际中 k 过大没有意义,并且对象的一阶邻近关系是最重要的。要素-匹配模型建立在对象具有相同邻近关系特征量的基础上,能够从宏观上粗粒度地比较空间邻近关系的变化。但当对象邻近关系特征量相同时,其排序关系等更细微的差异不能区分。

2 邻近关系排序差异的计算

将要素的排列特征用于相似性度量产生于 Gentner^[16] 对排列结构和类比映射的研究。对准模型度量相似性的基本思想是计算要素所有可能的排列,并反复修正不一致的排列直到排列一致、相似性确定为止,修正次数或修正量决定了相似性的大小。

邻近关系排序差异的度量以对象在不同分布环境下所具有的共同邻近关系特征量为计算对象。因此,对不同对象群中 P 的 k 阶邻近关系集合取交集,令 $k_A \cap k_B = Q$,对集合 Q 中的元素进行排序。以对象 P 为极点,水平正轴方向为极轴建立极坐标系,计算交集中所有元素的极角,以极角的非降顺序排列 P 的 k 阶邻近对象。如果有两个以上的对象极角大小相同,则以最接近极点 P 的对象优先。由此可分别得到比较排序 $A_compare$ 和被比较排序 $A_compareby$ 。

排序差异可以定义为比较排序与被比较排序之间的差值。设在不断地调整比较排序与被比较排序不一致的排列次序时,如果规定每次只能交

为特征量,用不同对象群中该对象相同的 k 阶邻近关系数目来表达和计算该对象的空间邻近关系相似度。设被比较对象群中对象 P 的 k 阶邻近关系集合为 k_A ,比较对象群中的集合为 k_B ,两个对象群中对象 P 相同的 k 阶邻近关系集合为 $k_A \cap k_B$,其中 k 为所讨论的空间邻近关系的阶数。设 $f(X)$ 是对任意集合 X 求集合元素个数的取值函数,则 P 的 k 阶邻近关系相似度 S_k 为:

换一个位置,直到与被比较排序一致时,共经过了 t 次调整。根据顺序关系的基本原理,在所有不一致的排序中,与被比较排序完全成倒序排列时需要的调整次数最大。设 n 为邻近关系交集的元素个数,则最大调整次数 $\max = n(n-1)/2$,其中 $n = f(Q)$ 。排序差异度由实际调整次数与最大调整次数的比值确定,即 $2t/n(n-1)$ 。设 S_A 为邻近关系排序的相似度,比较排序修正成被比较排序的调整次数 t 由函数 $g(A_compareby, A_compare)$ 完成,则 S_A 的计算方法如下:1) 当 $k_A \cap k_B = \emptyset$ 时, $S_A = 0$;2) 当 $k_A \cap k_B \neq \emptyset, f(Q) \geq 2$ 时,则:

$$S_A = 1 - \frac{2t}{n(n-1)} = 1 - \frac{2g(A_compareby, A_compare)}{f(Q)(f(Q)-1)} \quad (4)$$

当 $f(Q) < 2$ 时, $S_A = 1$ 。

3 顾及排序差异的相似性计算实例

在基于相同邻近特征量的要素-匹配模型基础上,用邻近关系的排序特征来细化邻近关系的相似性,可以进一步精化对象 P_i 的第 k 阶邻近关系相似度 Sim_i :

$$\text{Sim}_i = \delta * S_k + \lambda * S_A \quad (5)$$

其中, δ, λ 分别为邻近关系特征量相似度 S_k 和邻近关系排序相似度 S_A 的权值, $\delta + \lambda = 1$ 。

对象群 P_n 的总体邻近关系相似度 Sim_{total} 可将对象群中每个对象邻近关系的相似度 Sim_i 计算出来,然后累加求和,再除以对象群中的对象数目 n ,用每个对象邻近关系的平均相似度来表示:

$$\text{Sim}_{total} = \sum_{i=1}^n \text{Sim}_i / n \quad (6)$$

以对象类型和数目相同而对象位置不同的 3 个空间场景为实验对象,如图 2 所示, A 为被比较场景, B 和 C 为比较场景。对象群中共有 12 个对象,用整数 1~12 为每个对象赋以不同的 ID。为 3 个对象群建立 Delaunay 三角网,如图 3 所示。

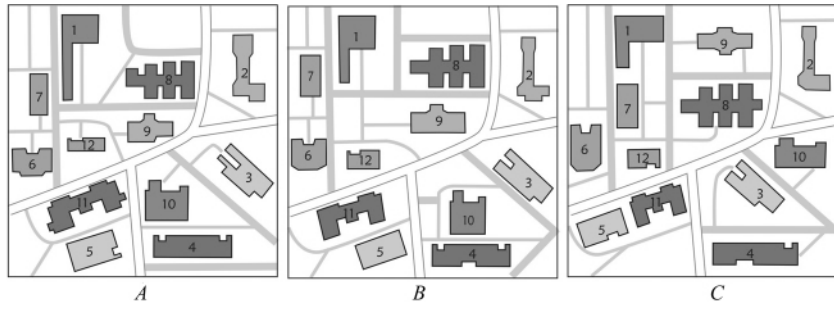


图 2 不同分布的空间场景 A、B、C

Fig. 2 Spatial Scene A, B and C with Different Distribution

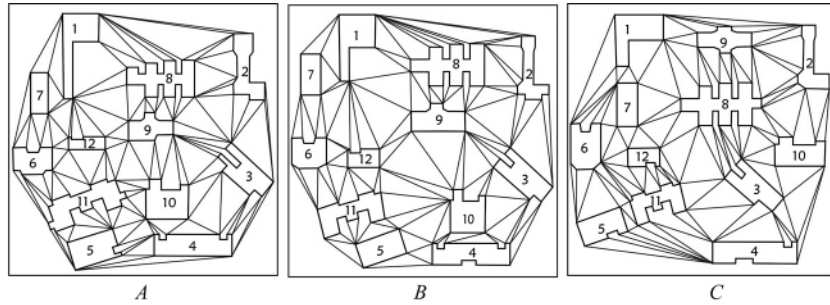


图 3 空间场景 A、B、C 中对象群的 Delaunay 三角网

Fig. 3 Delaunay Triangulation of Group Objects in Spatial Scene A, B and C

根据 Delaunay 三角网,搜索各个对象的一阶邻近关系,运用式(2)分别计算比较场景 B、C 与被比较场景 A 各个对象的一阶邻近关系相似度 S_k , 邻近特征量集合 k_A 和 k_B 相同元素的数目与不同元素的数目具有相同的重要性,因此各分量权值均等于 1,计算过程及结果分别如表 1 和表 2 所示。

运用式(4)分别计算比较场景 B、C 与被比较场景 A 各个对象的一阶邻近关系排序相似度 S_A , 计算过程和结果如表 3 所示。

运用式(5)计算每个对象结合了排序特征的一阶邻近关系相似度 Sim_i 。由于邻近特征量是邻近关系差异的本质,在对象群具有相同邻近特

征量的基础上讨论排序差异才有意义,因此 S_k 的权值应大于 S_A 的权值。这里以 $\delta=0.6, \lambda=0.4$ 为例计算 Sim_i ,在此基础上运用式(6)计算对象群的总体平均邻近关系相似度 Sim_{total} ,见表 4。

从计算结果可知,空间场景 B 中对象群与场景 A 中对象群邻近关系相似度大,总体邻近关系相似度为 0.81;而场景 C 与场景 A 的对象群邻近关系相似度小,为 0.58。将上述 3 个空间场景进行了空间关系差异的认知实验,在被调查的 20 个人中,有 85% 的人认为场景 B 中对象的位置分布比场景 C 的分布更接近于场景 A。可见,计算结果与大部分人的认知情况相一致。

表 1 空间场景 B 与 A 各个对象的一阶邻近关系相似度

Tab. 1 1-order Adjacency Relation Similarity of Every Object Between Spatial Scene B and A

i	k_A	k_B	$k_A \cap k_B$	$f(k_A \cap k_B)$	$f(k_A - k_B)$	$f(k_B - k_A)$	S_k
1	7 12 9 8 2	7 6 12 9 8 2	7 12 9 8 2	5	0	1	0.83
2	1 8 3	18 9 3	18 3	3	0	1	0.75
3	2 8 9 10 4	2 9 10 4	2 9 10 4	4	1	0	0.8
4	3 10 5	3 10 5	3 10 5	3	0	0	1
5	4 10 11 6	10 11 4	4 10 11	3	1	0	0.75
6	12 7 5 11	1 7 11 12	12 7 11	3	1	1	0.6
7	1 6 12	1 6	1 6	2	1	0	0.67
8	2 1 9 3	2 1 9	2 1 9	3	1	0	0.75
9	8 1 12 11 10 3	2 8 1 12 10 3	8 1 12 10 3	5	1	1	0.72
10	3 9 11 5 4	3 9 12 11 5 4	3 9 11 5 4	5	0	1	0.83
11	10 9 12 6 5	12 6 5 10	12 6 5 10	4	1	0	0.8
12	9 1 7 6 11	9 1 6 11 10	9 1 6 11	4	1	1	0.67

表2 空间场景C与A各个对象的一阶邻近关系相似度

Tab. 2 1-order Adjacency Relation Similarity of Every Object Between Spatial Scene C and A

i	k_A	k_B	$k_A \cap k_B$	$f(k_A \cap k_B)$	$f(k_A - k_B)$	$f(k_B - k_A)$	S_k
1	7 12 9 8 2	6 7 8 9	7 8 9	3	2	1	0.5
2	1 8 3	9 8 10	8	1	2	2	0.2
3	2 8 9 10 4	10 8 11 4	8 10 4	3	2	1	0.5
4	3 10 5	10 3 11 5	3 10 5	3	0	1	0.75
5	4 10 11 6	11 6 4	11 6 4	3	1	0	0.75
6	12 7 5 11	7 1 5 11 12	7 5 11 12	4	0	1	0.8
7	1 6 12	8 1 6 12	1 6 12	3	0	1	0.75
8	2 1 9 3	2 9 1 7 12 11 3 10	2 1 9 3	4	0	4	0.5
9	8 1 12 11 10 3	1 8 2	1 8	2	4	1	0.29
10	3 9 11 5 4	2 8 3 4	3 4	2	3	2	0.29
11	10 9 12 6 5	3 8 12 6 5 4	12 6 5	3	2	3	0.38
12	9 1 7 6 11	8 7 6 11	7 6 11	3	2	1	0.5

表3 一阶邻近关系排序相似度

Tab. 3 Sequence Similarity of 1-order Adjacency Relations

i	Scene B-Scene A					Scene C-Scene A				
	A_compareby	A_compare	t	max	S_A	A_compareby	A_compare	t	max	S_A
1	7 12 9 8 2	7 12 9 8 2	0	10	1	8 9 7	7 8 9	2	3	0.3
2	1 8 3	1 8 3	0	3	1	8	8	0	0	1
3	2 9 10 4	2 9 10 4	0	6	1	8 10 4	10 8 4	1	3	0.67
4	3 10 5	3 10 5	0	3	1	3 10 5	10 3 5	1	3	0.67
5	4 10 11	10 11 4	2	3	0.33	4 11 6	11 6 4	2	3	0.33
6	12 7 11	7 11 12	2	3	0.33	12 7 5 11	7 5 11 12	3	6	0.5
7	1 6	1 6	0	1	1	1 6 12	1 6 12	0	3	1
8	2 1 9	2 1 9	0	3	1	2 1 9 3	2 9 1 3	1	6	0.83
9	8 1 12 10 3	8 1 12 10 3	0	10	1	8 1	1 8	1	1	0
10	3 9 11 5 4	3 9 11 5 4	0	10	1	3 4	3 4	0	1	1
11	10 12 6 5	12 6 5 10	3	6	0.5	12 6 5	12 6 5	0	3	1
12	9 1 6 11	9 1 6 11	0	6	1	7 6 11	7 6 11	0	3	1

表4 总体邻近关系相似度

Tab. 4 Total Similarity of Adjacency Relations

Scene	Sim ₁	Sim ₂	Sim ₃	Sim ₄	Sim ₅	Sim ₆	Sim ₇	Sim ₈	Sim ₉	Sim ₁₀	Sim ₁₁	Sim ₁₂	Sim _{total}
B-A	0.9	0.9	0.9	1	0.6	0.5	0.8	0.9	0.8	0.9	0.7	0.8	0.81
C-A	0.4	0.5	0.6	0.7	0.6	0.7	0.8	0.6	0.2	0.6	0.6	0.7	0.58

4 结 语

针对空间关系的相似性,本文以通用的要素模型为基础,用不同场景中对象相同的邻近特征量来计算邻近关系的本质差异,再用排序差异细化度量邻近关系的相似性,建立了顾及排序差异的对象群邻近关系相似性计算方法,并用实例证明了该方法在邻近关系相似性计算中的有效性,同时也验证了对象和特征层次上的要素模型在对象与对象之间高层次的空间关系相似性计算中的可用性。

参 考 文 献

[1] 艾廷华,刘耀林. 保持空间分布特征的群点化简方法[J]. 测绘学报,2002,2:175-181

[2] 李成名,陈军,朱英浩. 基于 Voronoi 图的空间邻近定义与查询[J]. 武汉测绘科技大学学报,1998,23(2):128-131

[3] 陈军,赵仁亮,乔朝飞. 基于 Voronoi 图的 GIS 空间分析研究[J]. 武汉大学学报·信息科学版,2003,28(特刊):33-37

[4] 宋晓眉,程昌秀,周成虎,等. 利用 k 阶空间邻近图的空间层次聚类方法[J]. 武汉大学学报·信息科学版,2010,35(12):1 496-1 499

[5] 杜晓初,郭庆胜. 基于 Delaunay 三角网的空间邻近关系推理[J]. 测绘科学,2004,28(6):65-67

[6] Ai Tinghua, Zhang Xiang. The Aggregation of Urban Building Clusters Based on the Skeleton Partitioning of Gap Space[J]. Lecture Notes in Computer Science,2007(4):153-170

[7] Taylor E G, Hummel J E. Finding Similarity in a Model of Relational Reasoning[J]. Cognitive Systems Research, 2010, 11(2):229-230

[8] Egenhofer M J, Sharma J. Topological Consistency [C]. International Symposium on Spatial Data Handling, Charleston S C, 1992

[9] Randell, Cui D A, Cohn A G. Spatial Logic Based on Regions and Connection[C]. The 3rd International Conference on Knowledge Representation and Reasoning, San Mateo, 1992

[10] Bruns H T, Egenhofer M J. Similarity of Spatial Scenes[C]. The Seventh International Symposium on Spatial Data Handling, Delft, The Netherlands, 1996

- [11] Nedas K A, Egenhofer M J. Spatial-Scene Similarity Queries[J]. Transactions in GIS, 2008, 12(6): 661-681
- [12] 吕秀琴, 吴凡. 多尺度空间对象拓扑相似关系的表达与计算[J]. 测绘信息与工程, 2006, 31(2): 29-31
- [13] Schwering A. Approaches to Semantic Similarity Measurement for Geo-Spatial Data: a Survey[J]. Transactions in GIS, 2008, 12: 5-29
- [14] Schwering A, Raubal M. Spatial Relations for Semantic Similarity Measurement[J]. Lecture Notes in Computer Science, 2005, 3 770: 259-269
- [15] Tversky A. Features of Similarity[J]. Psychological Review, 1977, 84(4): 327-352
- [16] Gentner D, Markman A B. Structure Mapping in Analogy and Similarity[J]. American Psychologist, 1997, 52(1): 45-56

第一作者简介: 孟妮娜, 讲师, 博士, 主要研究方向: 地图综合、空间关系、空间相似性理论与方法。
E-mail: zhouxdm@163.com

Similarity Calculation of Adjacency Relations Attended to Sequence Difference Between Group Objects

MENG Nina¹ AI Tinghua² ZHOU Xiaodong³

(1 College of Geology Engineering and Geomatics, Chang'an University, 126 Yanta Road, Xi'an 710054, China)

(2 School of Resources and Environmental Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(3 Xi'an Institute of Surveying and Mapping, 1 Yanta Road, Xi'an 710054, China)

Abstract: Aiming at the similarity quantitative calculation of adjacency relations, we combine the feature model and the alignment model of similarity measures based on confirming characteristic quantity of adjacency relations, and build the similarity calculation method of adjacency relations between group objects attended to sequence difference. Finally, the method is verified by an example of usability.

Key words: adjacency relation; similarity; sequence difference; data matching; Delaunay triangulation

About the first author: MENG Nina, Ph. D., lecturer, majors in map generalization, spatial relation and theory of spatial similarity.
E-mail: zhouxdm@163.com

(上接第 736 页)

Dynamic Statistical Algorithm of Spatial Data Access Laws

PAN Shaoming¹ XU Zhengquan¹ LIU Xiaojun¹

(1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing,

Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: The tile access has dynamic features (server peer capability, storage device and hot tiles have dynamic features) and the Hotmap model based on the historical log information can not reflect the real system's current global information. Reproduction and distribution will produce a huge network flow rate in a large scale distributed nodes environment. A dynamic statistics algorithm for the distribution rule of the spatial data based on P2P is proposed to resolve above-mentioned problems. The service capabilities of the service nodes are calculated in this algorithm. The node agents with good service capabilities are chosen preferentially in the group to fuse dynamic statistical information. The experimental results show that the algorithm can meet the need of dynamic statistics in large scale distributed nodes environment with high efficiency.

Key words: spatial data; access law; dynamic statistics; global; P2P

About the first author: PAN Shaoming, associate professor, majors in multimedia communication and massive spatial information storage.
E-mail: pansm@lmars.whu.edu.cn