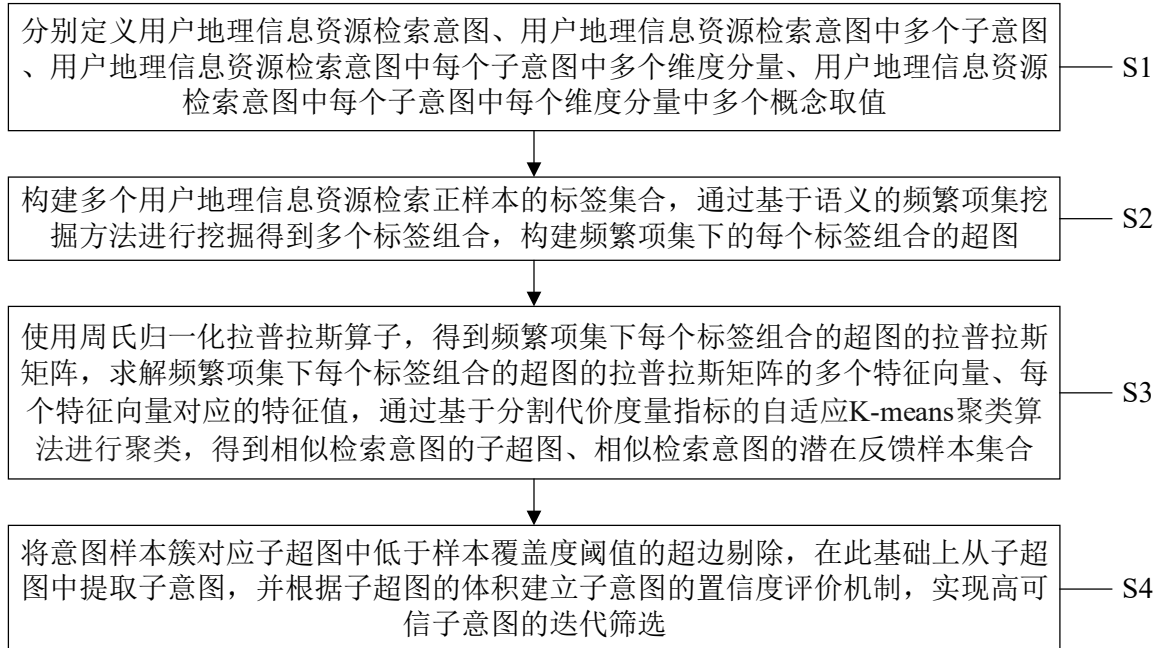


说 明 书 摘 要

本发明提出了一种地理信息资源检索意图检测方法及计算机可读介质。本发明定义用户地理信息资源检索意图的形式化表达机制；通过基于语义的频繁项集挖掘方法构建超图；基于分割代价度量指标实现自适应 K-means 聚类算法，得到相似检索意图的子超图、相似检索意图的潜在反馈样本集合；从相似检索意图的子超图中提取用户地理信息资源检索意图中的子意图，并建立用户地理信息资源检索意图中子意图的置信度评价机制。本发明利用超图建模反馈样本间的属性同义与上下位关系，显式描述子意图维度分量内的复合语义，并挖掘维度间的潜在关联，有助于发现隐藏意图并应对部分反馈样本元数据缺失导致意图无法完整表达的问题，支持各类地理信息资源的精准与鲁棒检索。

摘 要 附 图



权 利 要 求 书

1. 一种地理信息资源检索意图检测方法，其特征在于，包括以下步骤：

通过基于语义的频繁项集挖掘方法进行挖掘得到多个标签组合，构建频繁项集下的每个标签组合的超图；

5 结合周氏归一化拉普拉斯算子进行计算求解，得到频繁项集下的每个标签组合的超图的拉普拉斯矩阵的多个特征向量及对应的特征值，通过基于分割代价度量指标的自适应 K-means 聚类算法进行聚类，得到相似检索意图的子超图、潜在反馈样本集合；

构建用户地理信息资源检索意图中子意图中维度分量，计算用户地理信息资源检索意图中子意图的置信度。

10 2. 根据权利要求 1 所述的地理信息资源检索意图检测方法，其特征在于，包括以下步骤：

步骤 1：分别定义用户地理信息资源检索意图、用户地理信息资源检索意图中多个子意图、用户地理信息资源检索意图中每个子意图中多个维度分量、用户地理信息资源检索意图中每个子意图中每个维度分量中多个概念取值；

15 步骤 2：构建多个用户地理信息资源检索正样本的标签集合，通过基于语义的频繁项集挖掘方法进行挖掘得到多个标签组合，构建频繁项集下的每个标签组合的超图；

步骤 3：结合周氏归一化拉普拉斯算子进行计算，得到频繁项集下的每个标签组合的超图的拉普拉斯矩阵，求解频繁项集下的每个标签组合的超图的拉普拉斯矩阵的多个特征向量、每个特征向量对应的特征值，通过基于分割代价度量指标的自适应 K-means 聚类算法进行聚
20 类，得到相似检索意图的子超图、相似检索意图的潜在反馈样本集合；

步骤 4：计算相似检索意图的子超图中每条超边的覆盖度，在相似检索意图的子超图的多条超边中筛选出高于样本覆盖度阈值的多条超边作子超图的多条超边，构建用户地理信息资源检索意图中子意图中维度分量，计算用户地理信息资源检索意图中子意图的置信度。

3. 根据权利要求 2 所述的地理信息资源检索意图检测方法，其特征在于：

25 步骤 1 所述用户地理信息资源检索意图，定义为 I ；

步骤 1 所述用户地理信息资源检索意图中多个子意图，具体定如下：

$$I = \bigvee_{k=1}^M I^k$$

其中， M 为用户地理信息资源检索意图中子意图的数量， I^k 为用户地理信息资源检索意图中第 k 个子意图， \vee 表示逻辑或关系；

步骤 1 所述用户地理信息资源检索意图每个子意图中多个维度分量，具体定如下：

$$I^k = \bigwedge_{j=1}^N D_j^k$$

其中， N 为用户地理信息资源检索意图中每个子意图的维度分量的数量， D_j^k 表示用户地理信息资源检索意图中每个子意图第 k 个子意图中第 j 个维度分量， \wedge 表示逻辑与关系；

5 步骤 1 所述用户地理信息资源检索意图中每个子意图中每个维度分量中多个概念取值，具体定义如下：

$$D_j^k = \{c_{j,1}^k, c_{j,2}^k, \dots, c_{j,H}^k\}$$

其中， $c_{j,h}^k$ 为用户地理信息资源检索意图中第 k 个子意图中第 j 个维度分量中第 h 个概念取值， H 为用户地理信息资源检索意图中第 k 个子意图中第 j 个维度分量中概念取值的数量，

10 $c_{j,x}^k$ 、 $c_{j,y}^k$ 无同义与上下位关系， $h \in [1, H]$ ， $x \in [1, H]$ ， $y \in [1, H]$ 。

4. 根据权利要求 3 所述的地理信息资源检索意图检测方法，其特征在于：

步骤 2 所述构建多个用户地理信息资源检索正样本的标签集合，具体如下：

将每个用户地理信息资源检索正样本作为超图中每个顶点；

15 获取多个用户地理信息资源检索正样本以构建用户地理信息资源检索反馈样本集合，获取每个用户地理信息资源检索正样本的多个标签，所述的标签取通过外部知识库中的概念得到，与用户地理信息资源检索意图中每个子意图中每个维度分量中每个概念取值来源相同；

基于外部知识库中概念的同义和上下位关系，获取多个用户地理信息资源检索正样本的多个标签的多个最低公共祖先标签；

20 为每个用户地理信息资源检索正样本生成由多个标签和多个最低公共祖先标签构成的标签集合；

将多个用户地理信息资源检索正样本的标签集合通过基于语义的频繁项集挖掘方法进行挖掘得到频繁出现在多个用户地理信息资源检索正样本中的多个标签组合。

5. 根据权利要求 4 所述的地理信息资源检索意图检测方法，其特征在于：

步骤 2 所述通过基于语义的频繁项集挖掘方法进行挖掘得到多个标签组合，具体如下：

25 步骤 2.1：采用 Apriori 算法，将每个标签集合中的单个标签作为频繁项集挖掘中的单个项，计算所有单个项的支持度；

所述每个项的支持度，定义为：统计每个项在所有用户地理信息资源检索正样本的多个标签中出现的次数占用户地理信息资源检索正样本的数量的比例；

步骤 2.2: 判断所有单个项的支持度是否大于最小支持度阈值, 若是则保留, 若不是则剔除, 得到频繁 1 项集;

步骤 2.3: 以频繁 $n-1$ 项集为基础, $n>1$, 连接生成候选 n 项集, 每个项集是由 n 个标签组成的集合;

5 步骤 2.4: 基于外部知识库中概念的同义和上下位关系, 剔除候选 n 项集中不满足单个项取值无同义或上下位概念约束的项集;

步骤 2.5: 计算候选 n 项集中每个项的支持度;

判断每个项的支持度是否大于最小支持度阈值, 若是则保留, 若不是则剔除, 得到频繁 n 项集;

10 步骤 2.6: 若频繁 n 项集为空或只有一项, 输出所有频繁项集, 迭代结束;

步骤 2.7: 否则, 令 $n=n+1$, 转入步骤 2.3;

步骤 2.8: 在所有频繁项集中, 最大程度合并任意单个项取值无同义与上下位概念关系的项, 生成频繁项集下的多个标签组合。

6. 根据权利要求 5 所述的地理信息资源检索意图检测方法, 其特征在于:

15 步骤 2 所述构建频繁项集下的每个标签组合的超图, 具体如下:

将每个标签组合中的每个标签作为频繁项集下的每个标签组合的超图中每条超边的取值;

将频繁项集下的每个标签组合的超图中每条超边的取值依次与每个用户地理信息资源检索正样本的多个标签进行语义判断得到频繁项集下的每个标签组合的超图中每条超边包含的多个顶点、频繁项集下的每个标签组合的超图中每个顶点关联的多条边, 具体如下:

20 使用以频繁项集下的每个标签组合的超图中每个顶点为行元素、频繁项集下的每个标签组合的超图中每条超边为列元素的点边关联矩阵描述频繁项集下的每个标签组合的超图中顶点和超边的关系, 满足如下条件:

若频繁项集下的每个标签组合的超图中超边的取值为用户地理信息资源检索正样本的多个标签中任意一个标签的同义或上位概念, 则用户地理信息资源检索正样本对应的顶点属于频繁项集下的每个标签组合的超图中对应的超边包含的顶点、频繁项集下的每个标签组合的超图中对应的超边属于用户地理信息资源检索正样本对应的顶点关联的超边, 在点边关联矩阵中顶点和超边对应的行列取值为 1, 否则为 0;

25 多个标签中任意一个标签的同义或上位概念, 则用户地理信息资源检索正样本对应的顶点属于频繁项集下的每个标签组合的超图中对应的超边包含的顶点、频繁项集下的每个标签组合的超图中对应的超边属于用户地理信息资源检索正样本对应的顶点关联的超边, 在点边关联矩阵中顶点和超边对应的行列取值为 1, 否则为 0;

30 将频繁项集下的每个标签组合的超图中每条超边包含的顶点的数量定义为频繁项集下的每个标签组合的超图中每条超边的度;

将频繁项集下的每个标签组合的超图中每条超边的取值通过概念信息量度量方法计算

得到频繁项集下的每个标签组合的超图中每条超边包含的每个顶点到每条超边的距离，将每条超边包含的每个顶点到每条超边的距离通过归一化方法计算得到每条超边包含的每个顶点到每条超边的归一化后距离，将每条超边包含的多个顶点到每条超边的归一化后距离进行均值计算，得到频繁项集下的每个标签组合的超图中每条超边的权重，具体计算过程如下：

$$w_{t,q} = \frac{1}{\delta(e_{t,q})} \sum_{v_{t,q}=1}^{L_{t,q}} \exp\left(-\frac{Dis(v_{t,q}, e_{t,q})^2}{\mu}\right)$$

其中， $w_{t,q}$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边的权重， $e_{t,q}$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边， $\delta(e_{t,q})$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边的度， $v_{t,q}$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边包含的第 v_q 个顶点， $Dis(v_{t,q}, e_{t,q})$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边包含的第 v_q 个顶点到第 q 条超边的距离， μ 为权重的归一化系数， $q \in [1, Q]$ ， $t \in [1, K]$ ， Q 为频繁项集下的每个标签组合的超图中超边的数量， $L_{t,q}$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边包含的顶点的数量；

将频繁项集下的每个标签组合的超图中每个顶点关联的超边的权重之和定义为频繁项集下的每个标签组合的超图中每个顶点的度；

构建频繁项集下的每个标签组合的超图中超边的权重矩阵，所述频繁项集下的每个标签组合的超图的权重矩阵定义如下：

$$W_t \in \mathbb{R}^{|E_t| \times |E_t|}$$

其中， $|E_t|$ 表示频繁项集下的第 t 个标签组合的超图中超边集合的模长，即频繁项集下的第 t 个标签组合的超图中超边的数量， W_t 表示频繁项集下的第 t 个标签组合的超图的权重矩阵，为对角阵， W_t 上的每个元素对应每条超边的权重， \mathbb{R} 表示实数集；

构建频繁项集下的每个标签组合的超图中超边的度矩阵，所述频繁项集下的每个标签组合的超图中超边的度矩阵定义如下：

$$D_{t,e} \in \mathbb{R}^{|E_t| \times |E_t|}$$

其中， $|E_t|$ 表示频繁项集下的第 t 个标签组合的超图中超边集合的模长，即频繁项集下的第 t 个标签组合的超图中超边的数量， $D_{t,e}$ 表示频繁项集下的第 t 个标签组合的超图中超边的度矩阵，为对角阵， $D_{t,e}$ 上的每个元素对应每条超边的度；

构建频繁项集下的每个标签组合的超图中顶点的度矩阵，所述频繁项集下的每个标签组合的超图中顶点的度矩阵定义如下：

$$D_{t,v} \in \mathbb{R}^{|V_t| \times |V_t|}$$

其中, $|V_t|$ 表示频繁项集下的第 t 个标签组合的超图中顶点集合的模长, 即频繁项集下的第 t 个标签组合的超图中顶点的数量, $D_{t,v}$ 表示频繁项集下的第 t 个标签组合的超图中顶点的度矩阵, 为对角阵, $D_{t,v}$ 上的每个元素对应每个顶点的度。

7. 根据权利要求 6 所述的地理信息资源检索意图检测方法, 其特征在于:

5 步骤 3 所述结合周氏归一化拉普拉斯算子进行计算, 得到频繁项集下的每个标签组合的超图的拉普拉斯矩阵, 具体如下:

将频繁项集下的每个标签组合的超图的点边关联矩阵、频繁项集下的每个标签组合的超图中超边的权重矩阵、频繁项集下的每个标签组合的超图中超边的度矩阵、频繁项集下的每个标签组合的超图中顶点的度矩阵结合周氏归一化拉普拉斯算子进行计算, 得到频繁项集下的
10 的每个标签组合的超图的拉普拉斯矩阵;

所述频繁项集下的每个标签组合的超图的拉普拉斯矩阵的行、列的数量均为频繁项集下的每个标签组合的超图中顶点的数量;

步骤 3 所述频繁项集下的每个标签组合的超图的拉普拉斯矩阵的每个特征向量的长度为顶点的数量;

15 步骤 3 所述频繁项集下的每个标签组合的超图的拉普拉斯矩阵的每个特征向量的每个元素对应每个顶点的特征。

8. 根据权利要求 7 所述的地理信息资源检索意图检测方法, 其特征在于:

步骤 3 所述通过基于分割代价度量指标的自适应 K-means 聚类算法进行聚类, 具体过程如下:

20 初始化 $K=2$;

步骤 3.1: 频繁项集下的每个标签组合的超图的拉普拉斯矩阵的多个特征向量的特征值由小到大排序得到排序后多个特征值, 在排序后多个特征值中选择前 K 个排序后特征值对应的特征向量, 通过 K 个排序后特征向量中每个排序后特征向量中每个元素即 K 个, 构建对应频繁项集下的每个标签组合的超图的每个顶点的特征序列;

25 步骤 3.2: 将频繁项集下的每个标签组合的超图的多个顶点的特征序列通过 K-means 聚类方法进行聚类, 得到频繁项集下的每个标签组合的超图的顶点特征序列子集合;

所述频繁项集下的每个标签组合的超图的顶点特征序列集合为:

$$data_t = (T_{t,1}, T_{t,2}, \dots, T_{t,K})$$

其中, $data_t$ 为频繁项集下的第 t 个标签组合的超图的顶点特征序列集合, $T_{t,z}$ 为频繁项集下的第 t 个标签组合的超图的顶点特征序列集合中第 z 个顶点的特征序列子集合, K 为聚类的数量, $z \in [1, K]$;

步骤 3.3: 计算频繁项集下的每个标签组合的超图的分割代价度量指标;

$$5 \quad CC_t = \begin{cases} \left(\sum_{z=1}^K \frac{vol\partial(T_{t,z})}{vol(T_{t,z})} \right) / (K - 1), & K > 1 \\ 0, & K = 1 \end{cases}$$

其中, CC_t 为频繁项集下的第 t 个标签组合的超图的分割代价度量指标, $T_{t,z}$ 为频繁项集下的第 t 个标签组合的超图的顶点特征序列集合中第 z 个顶点的特征序列子集合, vol 表示使用标准化超图剪切方法计算超图体积, $vol\partial$ 表示使用标准化超图剪切方法计算超图边界体积, K 为聚类的数量, $z \in [1, K]$, 更新 $K=K+1$;

10 重复执行步骤 3.1-3.3, 直至 $CC_t \geq \gamma$ 或者 $K \geq num$ 时, γ 表示迭代阈值, num 表示最大迭代次数;

在频繁项集下的多个标签组合的多个超图和超图分割方案中, 选择频繁项集下的标签组合的超图的分割代价度量指标最小对应的频繁项集下的标签组合的超图被分割后的子超图作为相似检索意图的子超图;

15 将对应的分割代价度量指标最小对应的频繁项集下的标签组合的超图的顶点子集合作为相似检索意图的潜在反馈样本集合。

9. 根据权利要求 8 所述的地理信息资源检索意图检测方法, 其特征在于:

步骤 4 所述计算相似检索意图的子超图中每条超边的覆盖度, 具体如下:

每条超边在子超图中包含的顶点个数占子超图顶点总数的比例;

20 步骤 4 所述构建用户地理信息资源检索意图中子意图中维度分量, 具体如下:

获取多条相似检索意图的子超图中超边的值, 将每条相似检索意图的子超图中超边的值作为用户地理信息资源检索意图中子意图的维度分量中的每个概念取值, 以构建用户地理信息资源检索意图中子意图中维度分量;

步骤 4 所述计算用户地理信息资源检索意图中子意图的置信度, 具体如下:

$$25 \quad Conf(I^a) = \frac{vol(g_a)}{\max(vol(g_a))}$$

其中, $Conf(I^a)$ 表示第 a 个子超图对应的第 a 个子意图的置信度, vol 表示使用标准化超图剪切方法计算超图体积, g_a 表示频繁项集下的标签组合的超图的分割代价度量指标最小对应的频繁项集下的标签组合的超图被分割后的第 a 个子超图作为相似检索意图的第 a 个子超图, a

$\in [1, K]$; $\max(\text{vol}(g_a))$ 为子超图 g_a 的顶点个数乘以超边条数, K 为聚类的数量。

10. 一种计算机可读介质, 其特征在于, 其存储电子设备执行的计算机程序, 当所述计算机程序在电子设备上运行时, 使得所述电子设备执行如权利要求 1-9 任一项所述方法的步骤。

说明书

一种地理信息资源检索意图检测方法及其计算机可读介质

技术领域

本发明属于地理信息检索技术领域，尤其涉及一种地理信息资源检索意图检测方法及其计算机可读介质。

背景技术

随着地理与地球科学数据采集与共享技术日益完善，互联网上各式地理信息资源大量涌现，为地学研究与应用提供丰富数据资源的同时，也给准确的资源检索与发现带来巨大挑战。目前，地理信息资源的检索方法大致分为基于语义的检索、基于内容的检索和基于用户相关反馈的检索。基于语义的检索通过匹配用户输入的检索词与资源元数据中的标题、描述等字段实现检索，这类方法可解释性强但过度依赖用户对文本化检索条件的精准表述，亦难以应对数据资源元数据字段缺失导致的匹配失效问题。基于内容的检索利用图像的多模态特征信息进行相似度匹配完成检索，能够应对元数据缺失或错误的问题，但容易导致视觉特征的“过拟合”（例如将线状河流识别为等高线）与“欠匹配”（例如将闭合的等高线识别为面状区域），难以跨越低层图像特征与高层语义之间的“语义鸿沟”。基于用户相关反馈的检索，利用用户对当前检索结果的相关反馈数据训练分类器或构建典型样本进行重检索，实现检索结果的精化，该方法降低了用户检索需求表达的成本，但由于缺少对用户检索需求的显式建模，导致检索需求理解不准确且检索结果可解释性差。

基于意图的检索通过显式建模与识别用户需求提高检索精度，可为地理信息资源检索提供新的技术途径。基于预定义类别的意图识别方法在建立意图分类体系的基础上，利用决策树、支持向量机或深度学习模型识别潜在意图，但这种方法仅能表达检索意图的大致范围，无法准确刻画用户需求的具体内容；基于词项向量的方法，先从用户相关反馈数据中提取关键词集合作为意图模型，然后将关键词集合作为查询条件重检索，但该方法未考虑词项的语义及词项间的逻辑关系，难以全面准确定位用户兴趣资源。此外，上述意图识别方法，难以应对一次检索任务中存在多个检索需求的复杂地理信息资源检索场景，容易导致意图识别不全或失败。

发明内容

针对现有技术的不足，本发明提出了一种地理信息资源检索意图检测方法及其计算机可读介质。

本发明使用超图建模反馈样本间的属性同义与上下位关系，显式描述子意图维度分量内

的复合语义，并挖掘维度间的潜在关联，有助于发现隐藏意图并应对部分反馈样本元数据缺失导致意图无法完整表达的问题，以支持各类地理信息资源的精准与鲁棒检索。

本发明方法的技术方案为一种地理信息资源检索意图检测方法，包括以下步骤：

步骤 1：分别定义用户地理信息资源检索意图、用户地理信息资源检索意图中多个子意图、用户地理信息资源检索意图中每个子意图中多个维度分量、用户地理信息资源检索意图中每个子意图中每个维度分量中多个概念取值；

步骤 2：构建多个用户地理信息资源检索正样本的标签集合，通过基于语义的频繁项集挖掘方法进行挖掘得到多个标签组合，构建频繁项集下的每个标签组合的超图；

步骤 3：结合周氏归一化拉普拉斯算子进行计算，得到频繁项集下的每个标签组合的超图的拉普拉斯矩阵，求解频繁项集下的每个标签组合的超图的拉普拉斯矩阵的多个特征向量、每个特征向量对应的特征值，通过基于分割代价度量指标的自适应 K-means 聚类算法进行聚类，得到相似检索意图的子超图、相似检索意图的潜在反馈样本集合；

步骤 4：计算相似检索意图的子超图中每条超边的覆盖度，在相似检索意图的子超图的多条超边中筛选出高于样本覆盖度阈值的多条超边作子超图的多条超边，构建用户地理信息资源检索意图中子意图中维度分量，计算用户地理信息资源检索意图中子意图的置信度；

作为优选，步骤 1 所述用户地理信息资源检索意图，定义为 I ；

步骤 1 所述用户地理信息资源检索意图中多个子意图，具体定如下：

$$I = \bigvee_{k=1}^M I^k$$

其中， M 为用户地理信息资源检索意图中子意图的数量， I^k 为用户地理信息资源检索意图中第 k 个子意图， \vee 表示逻辑或关系；

步骤 1 所述用户地理信息资源检索意图每个子意图中多个维度分量，具体定如下：

$$I^k = \bigwedge_{j=1}^N D_j^k$$

其中， N 为用户地理信息资源检索意图中每个子意图的维度分量的数量， D_j^k 表示用户地理信息资源检索意图中每个子意图第 k 个子意图中第 j 个维度分量， \wedge 表示逻辑与关系；

步骤 1 所述用户地理信息资源检索意图中每个子意图中每个维度分量中多个概念取值，具体定义如下：

$$D_j^k = \{c_{j,1}^k, c_{j,2}^k, \dots, c_{j,H}^k\}$$

其中， $c_{j,h}^k$ 为用户地理信息资源检索意图中第 k 个子意图中第 j 个维度分量中第 h 个概念

取值, H 为用户地理信息资源检索意图中第 k 个子意图中第 j 个维度分量中概念取值的数量, $c_{j,x}^k$ 、 $c_{j,y}^k$ 无同义与上下位关系, $h \in [1, H]$, $x \in [1, H]$, $y \in [1, H]$;

作为优选, 步骤 2 所述构建多个用户地理信息资源检索正样本的标签集合, 具体如下:
将每个用户地理信息资源检索正样本作为超图中每个顶点;

- 5 获取多个用户地理信息资源检索正样本以构建用户地理信息资源检索反馈样本集合, 获取每个用户地理信息资源检索正样本的多个标签, 所述的标签取通过外部知识库中的概念得到, 与用户地理信息资源检索意图中每个子意图中每个维度分量中每个概念取值来源相同;

基于外部知识库中概念的同义和上下位关系, 获取多个用户地理信息资源检索正样本的多个标签的多个最低公共祖先标签;

- 10 为每个用户地理信息资源检索正样本生成由多个标签和多个最低公共祖先标签构成的标签集合;

将多个用户地理信息资源检索正样本的标签集合通过基于语义的频繁项集挖掘方法进行挖掘得到频繁出现在多个用户地理信息资源检索正样本中的多个标签组合;

步骤 2 所述通过基于语义的频繁项集挖掘方法进行挖掘得到多个标签组合, 具体如下:

- 15 步骤 2.1: 采用 Apriori 算法, 将每个标签集合中的单个标签作为频繁项集挖掘中的单个项, 计算所有单个项的支持度;

所述每个项的支持度, 定义为: 统计每个项在所有用户地理信息资源检索正样本的多个标签中出现的次数占用户地理信息资源检索正样本的数量的比例;

- 20 步骤 2.2: 判断所有单个项的支持度是否大于最小支持度阈值, 若是则保留, 若不是则剔除, 得到频繁 1 项集;

步骤 2.3: 以频繁 $n-1$ 项集为基础, $n > 1$, 连接生成候选 n 项集, 每个项集是由 n 个标签组成的集合;

步骤 2.4: 基于外部知识库中概念的同义和上下位关系, 剔除候选 n 项集中不满足单个项取值无同义或上下位概念约束的项集;

- 25 步骤 2.5: 计算候选 n 项集中每个项的支持度;

判断每个项的支持度是否大于最小支持度阈值, 若是则保留, 若不是则剔除, 得到频繁 n 项集;

步骤 2.6: 若频繁 n 项集为空或只有一项, 输出所有频繁项集, 迭代结束;

步骤 2.7: 否则, 令 $n=n+1$, 转入步骤 2.3;

- 30 步骤 2.8: 在所有频繁项集中, 最大程度合并任意单个项取值无同义与上下位概念关系的项, 生成频繁项集下的多个标签组合;

步骤 2 所述构建频繁项集下的每个标签组合的超图，具体如下：

将每个标签组合中的每个标签作为频繁项集下的每个标签组合的超图中每条超边的取值；

5 将频繁项集下的每个标签组合的超图中每条超边的取值依次与每个用户地理信息资源检索正样本的多个标签进行语义判断得到频繁项集下的每个标签组合的超图中每条超边包含的多个顶点、频繁项集下的每个标签组合的超图中每个顶点关联的多条边，具体如下：

使用以频繁项集下的每个标签组合的超图中每个顶点为行元素、频繁项集下的每个标签组合的超图中每条超边为列元素的点边关联矩阵描述频繁项集下的每个标签组合的超图中顶点和超边的关系，满足如下条件：

10 若频繁项集下的每个标签组合的超图中超边的取值为用户地理信息资源检索正样本的多个标签中任意一个标签的同义或上位概念，则用户地理信息资源检索正样本对应的顶点属于频繁项集下的每个标签组合的超图中对应的超边包含的顶点、频繁项集下的每个标签组合的超图中对应的超边属于用户地理信息资源检索正样本对应的顶点关联的超边，在点边关联矩阵中顶点和超边对应的行列取值为 1，否则为 0；

15 将频繁项集下的每个标签组合的超图中每条超边包含的顶点的数量定义为频繁项集下的每个标签组合的超图中每条超边的度；

将频繁项集下的每个标签组合的超图中每条超边的取值通过概念信息量度量方法计算得到频繁项集下的每个标签组合的超图中每条超边包含的每个顶点到每条超边的距离，将每条超边包含的每个顶点到每条超边的距离通过归一化方法计算得到每条超边包含的每个顶点到每条超边的归一化后距离，将每条超边包含的多个顶点到每条超边的归一化后距离进行均值计算，得到频繁项集下的每个标签组合的超图中每条超边的权重，具体计算过程如下：

$$w_{t,q} = \frac{1}{\delta(e_{t,q})} \sum_{v_{t,q}=1}^{L_{t,q}} \exp\left(-\frac{Dis(v_{t,q}, e_{t,q})^2}{\mu}\right)$$

其中， $w_{t,q}$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边的权重， $e_{t,q}$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边， $\delta(e_{t,q})$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边的度， $v_{t,q}$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边包含的第 v_q 个顶点， $Dis(v_{t,q}, e_{t,q})$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边包含的第 v_q 个顶点到第 q 条超边的距离， μ 为权重的归一化系数， $q \in [1, Q]$ ， $t \in [1, K]$ ， Q 为频繁项集下的每个标签组合的超图中超边的数量， $L_{t,q}$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边包含的顶点的数量；

将频繁项集下的每个标签组合的超图中每个顶点关联的超边的权重之和定义为频繁项集下的每个标签组合的超图中每个顶点的度；

构建频繁项集下的每个标签组合的超图中超边的权重矩阵，所述频繁项集下的每个标签组合的超图的权重矩阵定义如下：

$$5 \quad W_t \in \mathbb{R}^{|E_t| \times |E_t|}$$

其中， $|E_t|$ 表示频繁项集下的第 t 个标签组合的超图中超边集合的模长，即频繁项集下的第 t 个标签组合的超图中超边的数量， W_t 表示频繁项集下的第 t 个标签组合的超图的权重矩阵，为对角阵， W_t 上的每个元素对应每条超边的权重， \mathbb{R} 表示实数集；

构建频繁项集下的每个标签组合的超图中超边的度矩阵，所述频繁项集下的每个标签组合的超图中超边的度矩阵定义如下：

$$D_{t,e} \in \mathbb{R}^{|E_t| \times |E_t|}$$

其中， $|E_t|$ 表示频繁项集下的第 t 个标签组合的超图中超边集合的模长，即频繁项集下的第 t 个标签组合的超图中超边的数量， $D_{t,e}$ 表示频繁项集下的第 t 个标签组合的超图中超边的度矩阵，为对角阵， $D_{t,e}$ 上的每个元素对应每条超边的度；

构建频繁项集下的每个标签组合的超图中顶点的度矩阵，所述频繁项集下的每个标签组合的超图中顶点的度矩阵定义如下：

$$D_{t,v} \in \mathbb{R}^{|V_t| \times |V_t|}$$

其中， $|V_t|$ 表示频繁项集下的第 t 个标签组合的超图中顶点集合的模长，即频繁项集下的第 t 个标签组合的超图中顶点的数量， $D_{t,v}$ 表示频繁项集下的第 t 个标签组合的超图中顶点的度矩阵，为对角阵， $D_{t,v}$ 上的每个元素对应每个顶点的度；

作为优选，步骤3所述结合周氏归一化拉普拉斯算子进行计算，得到频繁项集下的每个标签组合的超图的拉普拉斯矩阵，具体如下：

将频繁项集下的每个标签组合的超图的点边关联矩阵、频繁项集下的每个标签组合的超图中超边的权重矩阵、频繁项集下的每个标签组合的超图中超边的度矩阵、频繁项集下的每个标签组合的超图中顶点的度矩阵结合周氏归一化拉普拉斯算子进行计算，得到频繁项集下的每个标签组合的超图的拉普拉斯矩阵；

所述频繁项集下的每个标签组合的超图的拉普拉斯矩阵的行、列的数量均为频繁项集下的每个标签组合的超图中顶点的数量；

步骤3所述频繁项集下的每个标签组合的超图的拉普拉斯矩阵的每个特征向量的长度为顶点的数量；

步骤3所述频繁项集下的每个标签组合的超图的拉普拉斯矩阵的每个特征向量的每个元素对应每个顶点的特征；

步骤3所述通过基于分割代价度量指标的自适应 K-means 聚类算法进行聚类，具体过程如下：

5 初始化 $K=2$ ；

步骤 3.1：频繁项集下的每个标签组合的超图的拉普拉斯矩阵的多个特征向量的特征值由小到大排序得到排序后多个特征值，在排序后多个特征值中选择前 K 个排序后特征值对应的特征向量，通过 K 个排序后特征向量中每个排序后特征向量中每个元素即 K 个，构建对应频繁项集下的每个标签组合的超图的每个顶点的特征序列；

10 步骤 3.2：将频繁项集下的每个标签组合的超图的多个顶点的特征序列通过 K-means 聚类方法进行聚类，得到频繁项集下的每个标签组合的超图的顶点特征序列子集合；

所述频繁项集下的每个标签组合的超图的顶点特征序列集合为：

$$data_t = (T_{t,1}, T_{t,2}, \dots, T_{t,K})$$

15 其中， $data_t$ 为频繁项集下的第 t 个标签组合的超图的顶点特征序列集合， $T_{t,z}$ 为频繁项集下的第 t 个标签组合的超图的顶点特征序列集合中第 z 个顶点的特征序列子集合， K 为聚类的数量， $z \in [1, K]$ ；

步骤 3.3：计算频繁项集下的每个标签组合的超图的分割代价度量指标；

$$CC_t = \begin{cases} \left(\sum_{z=1}^K \frac{vol\partial(T_{t,z})}{vol(T_{t,z})} \right) / (K - 1), & K > 1 \\ 0, & K = 1 \end{cases}$$

20 其中， CC_t 为频繁项集下的第 t 个标签组合的超图的分割代价度量指标， $T_{t,z}$ 为频繁项集下的第 t 个标签组合的超图的顶点特征序列集合中第 z 个顶点的特征序列子集合， vol 表示使用标准化超图剪切方法计算超图体积， $vol\partial$ 表示使用标准化超图剪切方法计算超图边界体积， K 为聚类的数量， $z \in [1, K]$ ，更新 $K=K+1$ ；

重复执行步骤 3.1-3.3，直至 $CC_t \geq \gamma$ 或者 $K \geq num$ 时， γ 表示迭代阈值， num 表示最大迭代次数；

25 在频繁项集下的多个标签组合的多个超图和超图分割方案中，选择频繁项集下的标签组合的超图的分割代价度量指标最小对应的频繁项集下的标签组合的超图被分割后的子超图作为相似检索意图的子超图；

将对应的分割代价度量指标最小对应的频繁项集下的标签组合的超图的顶点子集合作为相似检索意图的潜在反馈样本集合；

作为优选，步骤4所述计算相似检索意图的子超图中每条超边的覆盖度，具体如下：

每条超边在子超图中包含的顶点个数占子超图顶点总数的比例；

步骤4所述构建用户地理信息资源检索意图中子意图中维度分量，具体如下：

获取多条相似检索意图的子超图中超边的值，将每条相似检索意图的子超图中超边的值

5 作为用户地理信息资源检索意图中子意图的维度分量中的每个概念取值，以构建用户地理信息资源检索意图中子意图中维度分量；

步骤4所述计算用户地理信息资源检索意图中子意图的置信度，具体如下：

$$Conf(I^a) = \frac{vol(g_a)}{\max(vol(g_a))}$$

其中， $Conf(I^a)$ 表示第 a 个子超图对应的第 a 个子意图的置信度， vol 表示使用标准化超图

10 剪切方法计算超图体积， g_a 表示频繁项集下的标签组合的超图的分割代价度量指标最小对应的频繁项集下的标签组合的超图被分割后的第 a 个子超图作为相似检索意图的第 a 个子超图， $a \in [1, K]$ ； $\max(vol(g_a))$ 为子超图 g_a 的顶点个数乘以超边条数， K 为聚类的数量；

本发明还提供了一种计算机可读介质，所述计算机可读介质存储电子设备执行的计算机程序，当所述计算机程序在电子设备上运行时，执行所述地理信息资源检索意图检测方法的
15 步骤。

与现有技术相比，本发明具有如下优点和有益效果：

本发明提出一种基于超图分割的地理信息资源检索意图识别方法。将“意图-子意图-维度分量”检索意图表达模型中维度分量上的单标签约束改为多标签，以便表达意图中的复合概念，提升意图识别的准确性；融合语义的频繁项集挖掘，能够更好地捕捉反馈样本各维度
20 分量内及维度间的语义关联，而利用超图建模反馈样本间的属性同义与上下位关系，有助于挖掘隐藏意图，克服部分反馈样本元数据缺失导致意图无法完整表达的问题；基于超图分割的意图识别方法，能够通过建立意图样本簇、子超图、子意图之间的映射关系，应对一次检索任务中涉及多个需求的情况；建立子意图的置信度评价机制，可为迭代式相关反馈中高可信意图的筛选提供依据。本发明方法可推广应用于各类地理信息门户，实现地理信息资源精
25 准、智能的检索，提升地理信息共享与发现的服务品质。

附图说明

图1：本发明实施例的方法流程图。

图2：本发明实施例的检索意图形式化表达模型图。

图3：本发明实施例的基于超图建模反馈样本间属性同义与上下位关系的示意图。

30 图4：本发明实施例的基于频繁项集挖掘的超图构建示意图。

图 5: 本发明实施例的意图提取及置信度评价机制示意图。

图 6: 本发明实施例的四种典型检索场景下意图识别效果图。

具体实施方式

下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

具体实施时，本发明技术方案提出的方法可由本领域技术人员采用计算机软件技术实现自动运行流程，实现方法的系统装置例如存储本发明技术方案相应计算机程序的计算机可读存储介质以及包括运行相应计算机程序的计算机设备，也应当在本发明的保护范围内。

下面结合图 1-6 介绍本发明实施例的技术方案为一种地理信息资源检索意图检测方法，如图 1 所示为本发明的方法流程图，包括以下步骤：

步骤 1：如图 2 所示，分别定义用户地理信息资源检索意图、用户地理信息资源检索意图中多个子意图、用户地理信息资源检索意图中每个子意图中多个维度分量、用户地理信息资源检索意图中每个子意图中每个维度分量中多个概念取值；

步骤 1 所述用户地理信息资源检索意图，定义为 I ；以下步骤具体介绍如何表达“使用分级统计图法制作且主题为地质关于土壤中矿物质含量的地图，或使用分级统计图法制作且主题为农业与生物多样性的渔业地图”的地图检索意图；

步骤 1 所述用户地理信息资源检索意图中多个子意图，具体定如下：

$$I = \bigvee_{k=1}^M I^k$$

其中， M 为用户地理信息资源检索意图中子意图的数量， I^k 为用户地理信息资源检索意图中第 k 个子意图， \bigvee 表示逻辑或关系；可根据此意图模型将其形式化表达为“ $I = [\{\text{地图内容：土壤矿物质，制图方法：分级统计图法，地理主题：地质}\}, \{\text{地图内容：渔业，制图方法：分级统计图法，地理主题：农业、生物多样性}\}]$ ”，其中“ $I^1 = \{\text{地图内容：土壤、矿物质，制图方法：分级统计图法，地理主题：地质}\}$ ”和“ $I^2 = \{\text{地图内容：渔业，制图方法：分级统计图法，地理主题：农业、生物多样性}\}$ ”分别为两个子意图，子意图与子意图之间是“或”关系；

步骤 1 所述用户地理信息资源检索意图每个子意图中多个维度分量，具体定如下：

$$I^k = \bigwedge_{j=1}^N D_j^k$$

其中, $N=3$ 为用户地理信息资源检索意图中每个子意图的维度分量的数量, D_j^k 表示用户地理信息资源检索意图中每个子意图第 k 个子意图中第 j 个维度分量, \wedge 表示逻辑与关系; 实施例选取“地图内容 (Content, C)”、“制图方法 (Style, S)”与“地理主题 (Topic, T)” 3 个维度分量, 在后文中简称为 C、S、T;

步骤 1 所述用户地理信息资源检索意图中每个子意图中每个维度分量中多个概念取值, 具体定义如下:

$$D_j^k = \{c_{j,1}^k, c_{j,2}^k, \dots, c_{j,H}^k\}$$

其中, $c_{j,h}^k$ 为用户地理信息资源检索意图中第 k 个子意图中第 j 个维度分量中第 h 个概念取值, H 为用户地理信息资源检索意图中第 k 个子意图中第 j 个维度分量中概念取值的数量, $c_{j,x}^k$ 、 $c_{j,y}^k$ 无同义与上下位关系, $h \in [1, H]$, $x \in [1, H]$, $y \in [1, H]$; 以“地图内容: 土壤、矿物质”为例, 表示第 1 个子意图 I' 第 1 个维度分量中的第 1 个概念取值为“土壤”, 第 1 个子意图 I' 第 1 个维度分量中的第 2 个概念取值“矿物质”, $H=2$;

步骤 2: 构建多个用户地理信息资源检索正样本的标签集合, 通过基于语义的频繁项集挖掘方法进行挖掘得到多个标签组合, 构建频繁项集下的每个标签组合的超图;

步骤 2 所述构建多个用户地理信息资源检索正样本的标签集合, 具体如下:

以图 3a 所示的 SWEET 本体片段作为本例的外部知识库;

将每个用户地理信息资源检索正样本作为超图中每个顶点, 以图 3b 中的 5 个样本 s_1, s_2, s_3, s_4, s_5 为例, 作为超图中的 5 个顶点, 即 v_1 (淡水湖, 置底法, 水资源), v_2 (湖泊, 置底法, 水资源), v_3 (湾, 置底法, 水资源), v_4 (淡水湖, 河流, 线状制图, 地质), v_5 (河流, 线状制图, 地质);

获取多个用户地理信息资源检索正样本以构建用户地理信息资源检索反馈样本集合, 获取每个用户地理信息资源检索正样本的多个标签, 所述的标签取通过外部知识库中的概念得到, 与用户地理信息资源检索意图中每个子意图中每个维度分量中每个概念取值来源相同;

基于外部知识库中概念的同义和上下位关系, 获取多个用户地理信息资源检索正样本的多个标签的多个最低公共祖先标签, 如图 3a 的 SWEET 本体片段所示, “河流”和“淡水湖”的最低公共祖先概念为“水体”;

为每个用户地理信息资源检索正样本生成由多个标签和多个最低公共祖先标签构成的标签集合如图 3b;

将多个用户地理信息资源检索正样本的标签集合通过基于语义的频繁项集挖掘方法进行挖掘得到频繁出现在多个用户地理信息资源检索正样本中的多个标签组合；

步骤 2 所述通过基于语义的频繁项集挖掘方法进行挖掘得到多个标签组合，具体如下：

5 步骤 2.1：采用 Apriori 算法，将每个标签集合中的单个标签作为频繁项集挖掘中的单个项，计算所有单个项的支持度；如图 3b 中所示，样本 s_1 的标签集合为 s_1 {淡水湖,湖泊,水体,置底法,水资源}，将该集合中的每个标签作为频繁项集挖掘中的单个项，即图 3c 中候选 1 项集 C1；

10 所述每个项的支持度，定义为：统计每个项在所有用户地理信息资源检索正样本的多个标签中出现的次数占用户地理信息资源检索正样本的数量的比例；如图 3c 中候选 1 项集 C1 所示，{河流} 在 s_4, s_5 中均出现，此处共 5 个样本，{河流}的支持度为40%；

步骤 2.2：判断所有单个项的支持度是否大于最小支持度阈值，若是则保留，若不是则剔除，得到频繁 1 项集；为了举例及说明方便，此处取最小支持度阈值 $\lambda = 30\%$ ，由于{湾}的支持度小于 λ ，剔除该项集，剩余项集为频繁项集，如图 3c 中频繁 1 项集 L1 所示；

15 步骤 2.3：以频繁 $n-1$ 项集为基础, $n>1$ ，连接生成候选 n 项集，每个项集是由 n 个标签组成的集合；

步骤 2.4：基于外部知识库中概念的同义和上下位关系，剔除候选 n 项集中不满足单个项取值无同义或上下位概念约束的项集；参考图 3a 的 SWEET 本体关系可知“水体”是“河流”、“湖泊”的上位概念，因此项集{河流, 水体}和{水体, 湖泊}不满足约束条件，剔除此类项集；

20 步骤 2.5：计算候选 n 项集中每个项的支持度；

判断每个项的支持度是否大于最小支持度阈值，若是则保留，若不是则剔除，得到频繁 n 项集；以项集{河流, 湖泊}为例，该二项集共同出现在样本 s_4 中，支持度为 20%，小于支持度阈值 λ ，剔除该项集；最终输出如图 3d 中的频繁 2 项集 L2；

步骤 2.6：若频繁 n 项集为空或只有一项，输出所有频繁项集，迭代结束；

25 步骤 2.7：否则，令 $n=n+1$ ，转入步骤 2.3；

步骤 2.8：在所有频繁项集中，最大程度合并任意单个项取值无同义与上下位概念关系的项，生成频繁项集下的多个标签组合；如图 3e 中，共合并出两个标签组合，分别为 $\{R_1(C: 水体), (S: 置底法, 线状制图), (T: 水资源, 地质)\}$ 和 $R_2 \{(C: 河流, 湖泊), (S: 置底法, 线状制图), (T: 水资源, 地质)\}$ ；

30 步骤 2 所述构建频繁项集下的每个标签组合的超图，具体如下：

将每个标签组合中的每个标签作为频繁项集下的每个标签组合的超图中每条超边的取值；根据步骤 2.8 示例中的两个标签组合，生成如图 4a 及图 4b 所示的超图，其中图 4a 所示超图中的 5 条超边为 $e_1(C, \text{水体}, w_1)$, $e_2(S, \text{置底法}, w_2)$, $e_3(T, \text{水资源}, w_3)$, $e_4(S, \text{线状制图}, w_4)$, $e_5(T, \text{地质}, w_5)$ ，图 4b 所示超图中的 6 条超边为 $e_1(C, \text{湖泊}, w_1)$, $e_2(S, \text{置底法}, w_2)$, $e_3(T, \text{水资源}, w_3)$, $e_4(S, \text{线状制图}, w_4)$, $e_5(T, \text{地质}, w_5)$, $e_6(C, \text{河流}, w_6)$ ；

将频繁项集下的每个标签组合的超图中每条超边的取值依次与每个用户地理信息资源检索正样本的多个标签进行语义判断得到频繁项集下的每个标签组合的超图中每条超边包含的多个顶点、频繁项集下的每个标签组合的超图中每个顶点关联的多条边，具体如下：

使用以频繁项集下的每个标签组合的超图中每个顶点为行元素、频繁项集下的每个标签组合的超图中每条超边为列元素的点边关联矩阵描述频繁项集下的每个标签组合的超图中顶点和超边的关系，满足如下条件：

若频繁项集下的每个标签组合的超图中超边的取值为用户地理信息资源检索正样本的多个标签中任意一个标签的同义或上位概念，则用户地理信息资源检索正样本对应的顶点属于频繁项集下的每个标签组合的超图中对应的超边包含的顶点、频繁项集下的每个标签组合的超图中对应的超边属于用户地理信息资源检索正样本对应的顶点关联的超边，在点边关联矩阵中顶点和超边对应的行列取值为 1，否则为 0；图 4a 所示的超图可由图 4c 所示的点边关联矩阵刻画，图 4b 所示的超图可由图 4d 所示的点边关联矩阵刻画；

将频繁项集下的每个标签组合的超图中每条超边包含的顶点的数量定义为频繁项集下的每个标签组合的超图中每条超边的度；

将频繁项集下的每个标签组合的超图中每条超边的取值通过概念信息量度量方法计算得到频繁项集下的每个标签组合的超图中每条超边包含的每个顶点到每条超边的距离，将每条超边包含的每个顶点到每条超边的距离通过归一化方法计算得到每条超边包含的每个顶点到每条超边的归一化后距离，将每条超边包含的多个顶点到每条超边的归一化后距离进行均值计算，得到频繁项集下的每个标签组合的超图中每条超边的权重，具体计算过程如下：

$$w_{t,q} = \frac{1}{\delta(e_{t,q})} \sum_{v_{t,q}=1}^{L_{t,q}} \exp\left(-\frac{Dis(v_{t,q}, e_{t,q})^2}{\mu}\right)$$

其中， $w_{t,q}$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边的权重， $e_{t,q}$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边， $\delta(e_{t,q})$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边的度， $v_{t,q}$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边包含的第 v_q 个顶点， $Dis(v_{t,q}, e_{t,q})$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边包含的第 v_q 个顶点到第 q

条超边的距离， μ 为权重的归一化系数， $q \in [1, Q]$ ， $t \in [1, K]$ ， Q 为频繁项集下的每个标签组合的超图中超边的数量， $L_{t,q}$ 为频繁项集下的第 t 个标签组合的超图中第 q 条超边包含的顶点的数量；

5 将频繁项集下的每个标签组合的超图中每个顶点关联的超边的权重之和定义为频繁项集下的每个标签组合的超图中每个顶点的度；

构建频繁项集下的每个标签组合的超图中超边的权重矩阵，所述频繁项集下的每个标签组合的超图的权重矩阵定义如下：

$$W_t \in \mathbb{R}^{|E_t| \times |E_t|}$$

10 其中， $|E_t|$ 表示频繁项集下的第 t 个标签组合的超图中超边集合的模长，即频繁项集下的第 t 个标签组合的超图中超边的数量， W_t 表示频繁项集下的第 t 个标签组合的超图的权重矩阵，为对角阵， W_t 上的每个元素对应每条超边的权重， \mathbb{R} 表示实数集；

构建频繁项集下的每个标签组合的超图中超边的度矩阵，所述频繁项集下的每个标签组合的超图中超边的度矩阵定义如下：

$$D_{t,e} \in \mathbb{R}^{|E_t| \times |E_t|}$$

15 其中， $|E_t|$ 表示频繁项集下的第 t 个标签组合的超图中超边集合的模长，即频繁项集下的第 t 个标签组合的超图中超边的数量， $D_{t,e}$ 表示频繁项集下的第 t 个标签组合的超图中超边的度矩阵，为对角阵， $D_{t,e}$ 上的每个元素对应每条超边的度；

构建频繁项集下的每个标签组合的超图中顶点的度矩阵，所述频繁项集下的每个标签组合的超图中顶点的度矩阵定义如下：

20
$$D_{t,v} \in \mathbb{R}^{|V_t| \times |V_t|}$$

其中， $|V_t|$ 表示频繁项集下的第 t 个标签组合的超图中顶点集合的模长，即频繁项集下的第 t 个标签组合的超图中顶点的数量， $D_{t,v}$ 表示频繁项集下的第 t 个标签组合的超图中顶点的度矩阵，为对角阵， $D_{t,v}$ 上的每个元素对应每个顶点的度；

25 步骤 3：结合周氏归一化拉普拉斯算子进行计算，得到频繁项集下的每个标签组合的超图的拉普拉斯矩阵，求解频繁项集下的每个标签组合的超图的拉普拉斯矩阵的多个特征向量、每个特征向量对应的特征值，通过基于分割代价度量指标的自适应 K-means 聚类算法进行聚类，得到相似检索意图的子超图、相似检索意图的潜在反馈样本集合；

步骤 3 所述结合周氏归一化拉普拉斯算子进行计算，得到频繁项集下的每个标签组合的超图的拉普拉斯矩阵，具体如下：

30 将频繁项集下的每个标签组合的超图的点边关联矩阵、频繁项集下的每个标签组合的超

图中超边的权重矩阵、频繁项集下的每个标签组合的超图中超边的度矩阵、频繁项集下的每个标签组合的超图中顶点的度矩阵结合周氏归一化拉普拉斯算子进行计算，得到频繁项集下的每个标签组合的超图的拉普拉斯矩阵；

5 所述频繁项集下的每个标签组合的超图的拉普拉斯矩阵的行、列的数量均为频繁项集下的每个标签组合的超图中顶点的数量；

步骤3所述频繁项集下的每个标签组合的超图的拉普拉斯矩阵的每个特征向量的长度为顶点的数量；

步骤3所述频繁项集下的每个标签组合的超图的拉普拉斯矩阵的每个特征向量的每个元素对应每个顶点的特征；

10 步骤3所述通过基于分割代价度量指标的自适应 K-means 聚类算法进行聚类，具体过程如下：

初始化 $K=2$ ；

步骤 3.1：频繁项集下的每个标签组合的超图的拉普拉斯矩阵的多个特征向量的特征值由小到大排序得到排序后多个特征值，在排序后多个特征值中选择前 K 个排序后特征值对应的特征向量，通过 K 个排序后特征向量中每个排序后特征向量中每个元素即 K 个，构建对应频繁项集下的每个标签组合的超图的每个顶点的特征序列；

步骤 3.2：将频繁项集下的每个标签组合的超图的多个顶点的特征序列通过 K-means 聚类方法进行聚类，得到频繁项集下的每个标签组合的超图的顶点特征序列子集合；

所述频繁项集下的每个标签组合的超图的顶点特征序列集合为：

$$20 \quad data_t = (T_{t,1}, T_{t,2}, \dots, T_{t,K})$$

其中， $data_t$ 为频繁项集下的第 t 个标签组合的超图的顶点特征序列集合， $T_{t,z}$ 为频繁项集下的第 t 个标签组合的超图的顶点特征序列集合中第 z 个顶点的特征序列子集合， K 为聚类的数量， $z \in [1, K]$ ；

步骤 3.3：计算频繁项集下的每个标签组合的超图的分割代价度量指标；

$$25 \quad CC_t = \begin{cases} \left(\sum_{z=1}^K \frac{vol\partial(T_{t,z})}{vol(T_{t,z})} \right) / (K-1), & K > 1 \\ 0, & K = 1 \end{cases}$$

其中， CC_t 为频繁项集下的第 t 个标签组合的超图的分割代价度量指标， $T_{t,z}$ 为频繁项集下的第 t 个标签组合的超图的顶点特征序列集合中第 z 个顶点的特征序列子集合， vol 表示使用标准化超图剪切方法计算超图体积， $vol\partial$ 表示使用标准化超图剪切方法计算超图边界体积， K 为聚类的数量， $z \in [1, K]$ ，更新 $K=K+1$ ；

重复执行步骤 3.1-3.3, 直至 $CC_t \geq \gamma$ 或者 $K \geq num$ 时, γ 表示迭代阈值, num 表示最大迭代次数;

在频繁项集下的多个标签组合的多个超图和超图分割方案中, 选择频繁项集下的标签组合的超图的分割代价度量指标最小对应的频繁项集下的标签组合的超图被分割后的子超图

5 作为相似检索意图的子超图;

将对应的分割代价度量指标最小对应的频繁项集下的标签组合的超图的顶点子集合作为相似检索意图的潜在反馈样本集合;

步骤 4: 计算相似检索意图的子超图中每条超边的覆盖度, 在相似检索意图的子超图的多条超边中筛选出高于样本覆盖度阈值的多条超边作子超图的多条超边, 构建用户地理信息资源检索意图中子意图中维度分量, 计算用户地理信息资源检索意图中子意图的置信度;

步骤 4 所述计算相似检索意图的子超图中每条超边的覆盖度, 具体如下:

每条超边在子超图中包含的顶点个数占子超图顶点总数的比例; 将图 4b 中的超图通过步骤 3 分割为图 5a 中所示, 子超图 g_1 和 g_2 的超边覆盖度均大于阈值, 全部保留

步骤 4 所述构建用户地理信息资源检索意图中子意图中维度分量, 具体如下:

15 获取多条相似检索意图的子超图中超边的值, 将每条相似检索意图的子超图中超边的值作为用户地理信息资源检索意图中子意图的维度分量中的每个概念取值, 以构建用户地理信息资源检索意图中子意图中维度分量; 如图 5b 中所示, 子超图 g_2 的 $e_1(C, \text{湖泊}, w_1)$ 和 $e_6(C, \text{河流}, w_6)$ 对应子意图 I^2 的第 1 个维度分量 D_1^2 (地图内容 C) 的两个取值 {湖泊, 河流};

步骤 4 所述计算用户地理信息资源检索意图中子意图的置信度, 具体如下:

$$20 \quad Conf(I^a) = \frac{vol(g_a)}{\max(vol(g_a))}$$

其中, $Conf(I^a)$ 表示第 a 个子超图对应的第 a 个子意图的置信度, vol 表示使用标准化超图剪切方法计算超图体积, g_a 表示频繁项集下的标签组合的超图的分割代价度量指标最小对应的频繁项集下的标签组合的超图被分割后的第 a 个子超图作为相似检索意图的第 a 个子超图, $a \in [1, K]$; $\max(vol(g_a))$ 为子超图 g_a 的顶点个数乘以超边条数, K 为聚类的数量; 为了说明方便, 本发明以权重 $w = 1$ 为例, 如图 5a 所示, 计算子意图 I^1 和 I^2 的置信度, 最后按步骤 S1 中扩展的检索意图模型输出如图 5b 所示的检索意图;

步骤 5: 从意图语义距离的角度采用意图查准率 (Precision) 和意图查全率 (Recall) 验证意图识别的效果; 从样本覆盖的角度采用杰卡德系数 (Jaccard coefficient) 验证意图识别的效果。

30 使用如下所示的公式计算用户地理信息资源检索意图中子意图的维度分量间的距离

$Dis(dim)$ ，继而计算用户地理信息资源检索意图中子意图间的距离 $Dis(intention)$ ，最后将用户地理信息资源检索意图中子意图间的距离归一化至用户地理信息资源检索意图中子意图间的相似度 $Sim(intention)$ ；定义如下：

$$Dis(dim) = \min(\sum dis(c_u, c_w))$$

$$Dis(intention) = \frac{1}{3} \sqrt{Dis(C)^2 + Dis(S)^2 + Dis(T)^2}$$

$$Sim(intention) = \exp(-\frac{Dis(intention)^2}{\mu})$$

用户地理信息资源检索意图中子意图的维度分量间的距离 $Dis(dim)$ 等于维度分量中概念取值间的最小距离和，其中 dim 指维度，使用匈牙利算法确定两个子意图的同一维度分量中概念取值与概念取值之间的映射关系， $dis(c_u, c_w)$ 是使用概念信息量度量方法计算的两个概念取值的距离；用户地理信息资源检索意图中子意图间的距离 $Dis(intention)$ 中 $Dis(\cdot)$ 指两个子意图同一维度分量间的距离； $Sim(intention)$ 中， μ 为归一化系数；

杰卡德系数 $Jaccard$ 是一种用于衡量两个集合相似度的指标，杰卡德系数的取值范围在 0 和 1 之间，其中 0 表示两个集合没有共同元素，1 表示两个集合完全相同，定义如下：

$$J(I_{true}, I_{recognize}) = \frac{|S_{I_{true}} \cap S_{I_{recognize}}|}{|S_{I_{true}} \cup S_{I_{recognize}}|}$$

$J(I_{true}, I_{recognize})$ 表示真实意图 I_{true} 与识别意图 $I_{recognize}$ 的杰卡德系数， $S_{I_{true}}$ 指真实意图覆盖的地理信息资源检索样本集合， $S_{I_{recognize}}$ 指识别意图覆盖的地理信息资源检索样本集合，分子为二者的交集，分母为二者的并集；

意图查准率 $Precision$ 指识别的意图中，正确意图的比例，定义如下：

$$Precision = \frac{TP}{TP + FP}$$

TP 表示识别的意图对应真实意图的最大语义相似度之和，使用匈牙利算法确定识别意图和真意图最大语义相似度之和的映射关系， $TP + FP$ 表示识别的意图数量；

意图查全率 $Recall$ 指正确识别的意图占真意图的比例，定义如下：

$$Recall = \frac{TP}{TP + FP'}$$

TP 表示识别的意图对应真实意图的最大语义相似度之和，使用匈牙利算法确定识别意图和真意图最大语义相似度之和的映射关系， $TP + FP'$ 表示真意图的数量；

现将正反馈标签定义为意图维度分量的概念标签及其上位概念、下位概念和其他同义概念标签，具有正反馈标签的样本为正样本，其它样本为负样本。考虑到用户相关反馈中会出

现语义泛化（即样本标签为用户检索意图维度分量取值的上位概念）、语义特化（即样本标签为用户检索意图维度分量取值的下位概念）等模糊匹配行为，以及多标签判断（即反馈样本的多个标签中有一定数量的标签满足用户检索意图维度分量取值即标记为正样本）和误选（即用户错误的将负样本标记为正样本）的情况，设置语义模糊度、取值满足率和噪声率模拟用户相关反馈行为，为每个检索意图匹配正负相关反馈样本。其中，语义模糊度为上位概念、下位概念和其他同义概念标签占正反馈标签的比例，用于刻画用户模糊匹配行为；取值满足率为正反馈标签占总标签的比例，用于度量多标签判断行为；噪声率为负样本占总样本的比例，用于度量误选行为。

如图 6 所示为“单意图单维度”、“单意图多维度”、“多意图单维度”和“多意图多维度”共 4 种典型检索场景下开展验证实验的意图识别效果图。横轴为取值满足率的范围，分为[0, 0.2]、[0.2, 0.4]、[0.4, 0.6]、[0.6, 0.8]和[0.8, 1]共 5 组，纵轴为杰卡德系数 Jaccard、意图查准率 Precision 和意图查全率 Recall 的值。整体而言，随着取值满足率的增大，各个检索场景中三个指标均以接近 1 的趋势增大，说明本发明方法具有正确识别意图的能力。另外，在多意图场景中三个指标最终接近 1，说明本发明方法可以应对一次检索任务中涉及多个需求的情况。

本发明的具体实施例还提供了一种计算机可读介质。

所述计算机可读介质为服务器工作站；

所述服务器工作站存储电子设备执行的计算机程序，当所述计算机程序在电子设备上运行时，使得所述电子设备执行本发明实施例的地理信息资源检索意图检测方法的步骤。

应当理解的是，本说明书未详细阐述的部分均属于现有技术。

应当理解的是，上述针对较佳实施例的描述较为详细，并不能因此而认为是对本发明专利保护范围的限制，本领域的普通技术人员在本发明的启示下，在不脱离本发明权利要求所保护的范围情况下，还可以做出替换或变形，均落入本发明的保护范围之内，本发明的请求保护范围应以所附权利要求为准。

说明书附图

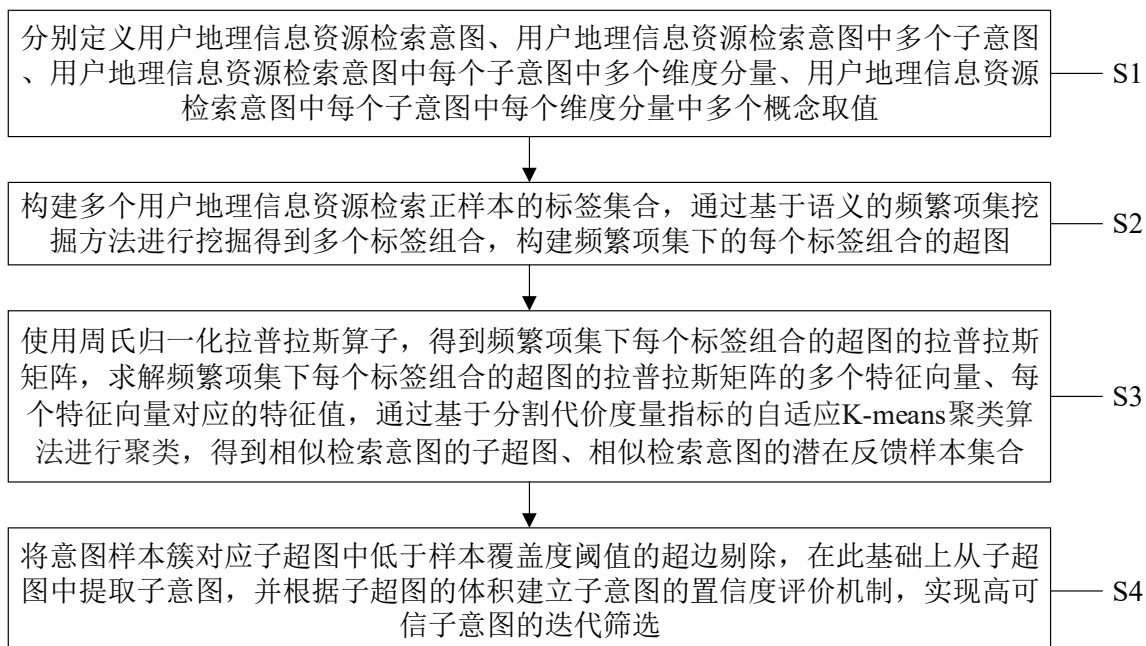


图 1

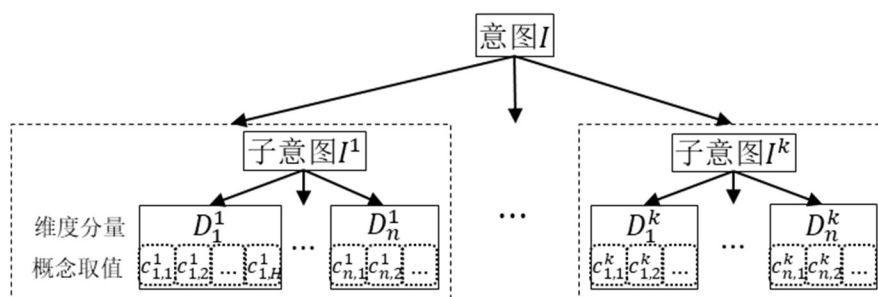
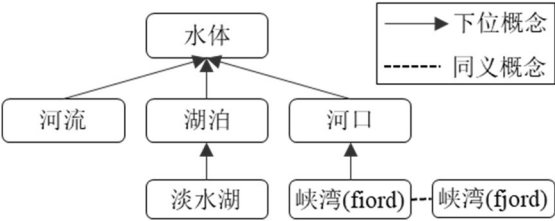


图 2



a

样本	标签集合			
	地图内容		制图方法	地理主题
	标签	LCA上位标签		
s_1	淡水湖	湖泊、水体	置底法	水资源
s_2	湖泊	水体	置底法	水资源
s_3	湾	水体	置底法	水资源
s_4	淡水湖、河流	湖泊、水体	线状制图	地质
s_5	河流	水体	线状制图	地质

b

候选1项集C1	支持度	频繁1项集L1	支持度
河流	2, 40%	河流	2, 40%
淡水湖	2, 40%	淡水湖	2, 40%
湾	1, 20%	水体	5, 100%
水体	5, 100%	湖泊	3, 60%
湖泊	3, 60%	线状制图	2, 50%
线状制图	2, 50%	置底法	3, 60%
置底法	3, 60%	水资源	3, 60%
水资源	3, 60%	地质	2, 40%
地质	2, 40%		

C	S	T
水体	置底法, 线状制图	水资源, 地质
湖泊, 河流	置底法, 线状制图	水资源, 地质
淡水湖, 河流	置底法, 线状制图	水资源, 地质

c

候选2项集C2	支持度	频繁2项集L2	支持度
河流, 水体		水体, 置底法	3, 60%
河流, 湖泊	1, 20%	水体, 水资源	3, 60%
河流, 线状制图	2, 40%	水体, 地质	2, 40%
河流, 地质	2, 40%	水体, 线状制图	2, 40%
湖泊, 地质	1, 20%	湖泊, 置底法	2, 40%
水体, 湖泊		湖泊, 水资源	2, 40%
水体, 置底法	3, 60%	河流, 地质	2, 40%
水体, 水资源	3, 60%	河流, 线状制图	2, 40%
...	...	置底法, 水资源	3, 60%
		线状制图, 地质	2, 40%

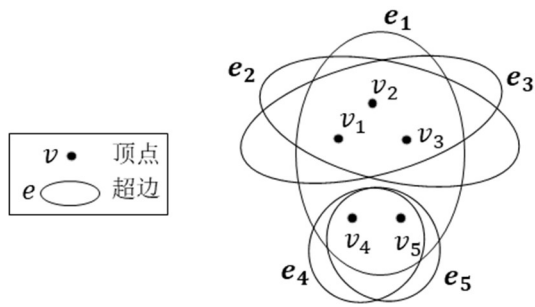
C	S	T
水体	置底法, 线状制图	水资源, 地质
湖泊, 河流	置底法, 线状制图	水资源, 地质

d

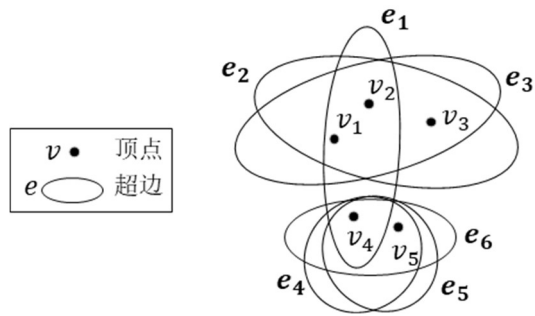
候选3项集C3	支持度	频繁3项集L3	支持度
水体, 置底法, 水资源	3, 60%	水体, 置底法, 水资源	3, 60%
水体, 置底法, 地质	0, 0%	水体, 线状制图, 地质	2, 40%
水体, 置底法, 线状制图	0, 0%	湖泊, 水资源, 置底法	2, 40%
水体, 线状制图, 地质	2, 40%	河流, 线状制图, 地质	2, 40%
水体, 水资源, 湖泊		C 水体	湖泊, 河流
水体, 线状制图, 河流		S 置底法, 线状制图	置底法, 线状制图
湖泊, 水资源, 置底法	2, 40%	T 水资源, 地质	水资源, 地质
河流, 线状制图, 地质	2, 40%		
...	...		

e

图 3



a



b

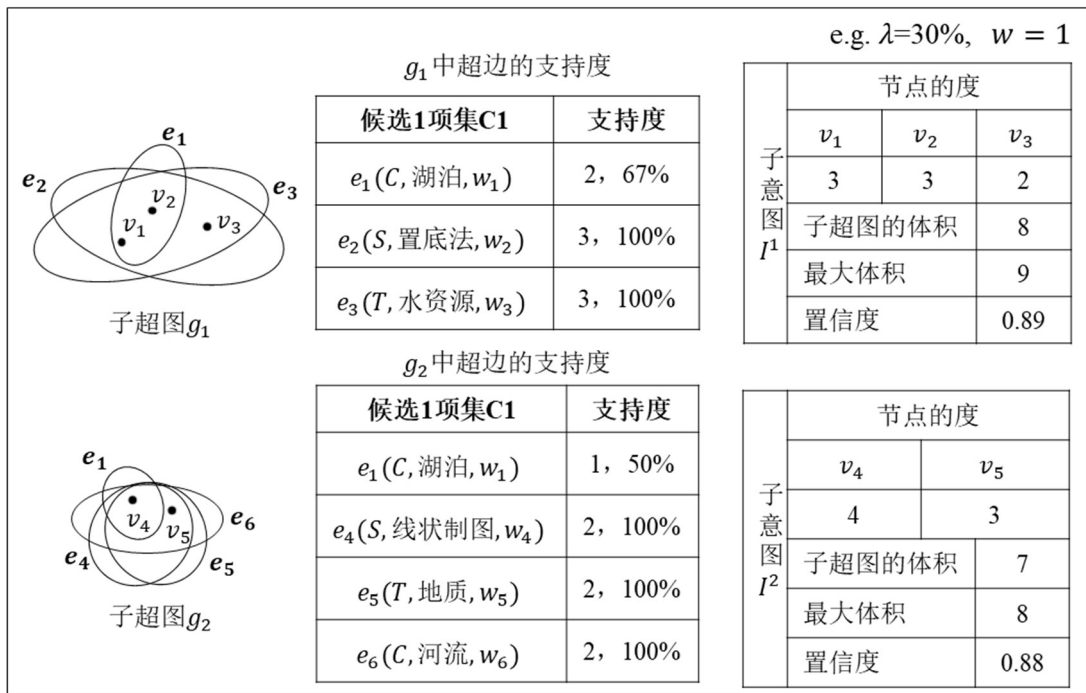
点边关联矩阵	e_1 (C, 水体, w_1)	e_2 (S, 置底法, w_2)	e_3 (T, 水资源, w_3)	e_4 (S, 线状制图, w_4)	e_5 (T, 地质, w_5)
v_1 (淡水湖, 置底法, 水资源)	1	1	1	0	0
v_2 (湖泊, 置底法, 水资源)	1	1	1	0	0
v_3 (湾, 置底法, 水资源)	1	1	1	0	0
v_4 (淡水湖, 河流, 线状制图, 地质)	1	0	0	1	1
v_5 (河流, 线状制图, 地质)	1	0	0	1	1

c

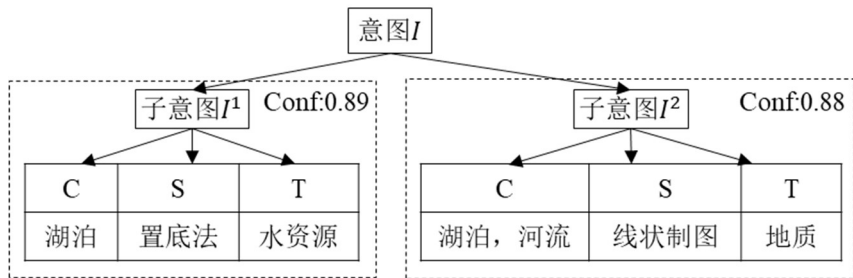
点边关联矩阵	e_1 (C, 湖泊, w_1)	e_2 (S, 置底法, w_2)	e_3 (T, 水资源, w_3)	e_4 (S, 线状制图, w_4)	e_5 (T, 地质, w_5)	e_6 (C, 河流, w_6)
v_1 (淡水湖, 置底法, 水资源)	1	1	1	0	0	0
v_2 (湖泊, 置底法, 水资源)	1	1	1	0	0	0
v_3 (湾, 置底法, 水资源)	0	1	1	0	0	0
v_4 (淡水湖, 河流, 线状制图, 地质)	1	0	0	1	1	1
v_5 (河流, 线状制图, 地质)	0	0	0	1	1	1

d

图 4



a



b

图 5

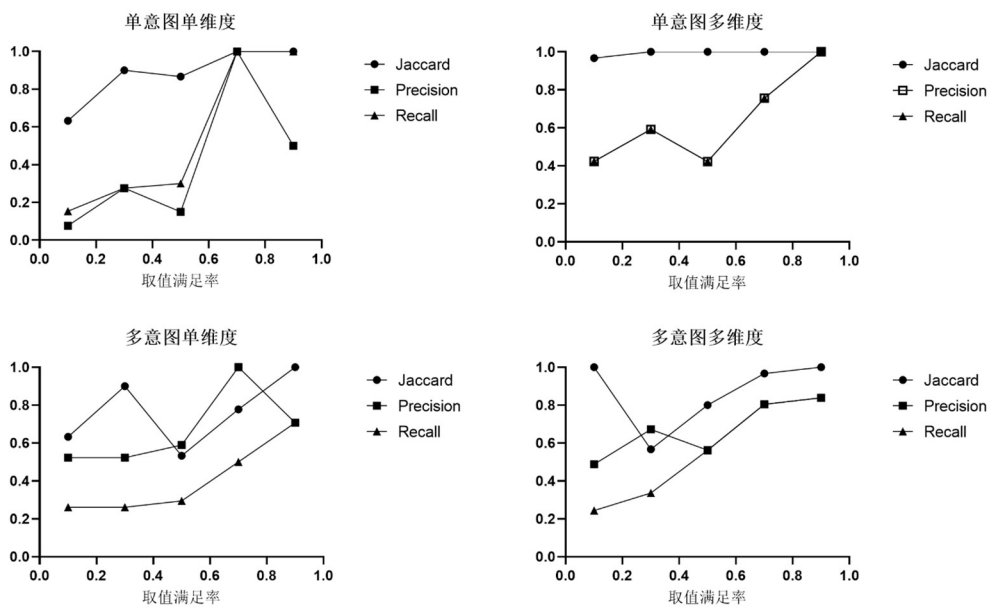


图 6