



(12)发明专利

(10)授权公告号 CN 106021228 B
(45)授权公告日 2019.08.20

(21)申请号 201610329206.8
(22)申请日 2016.05.18
(65)同一申请的已公布的文献号
 申请公布号 CN 106021228 A
(43)申请公布日 2016.10.12
(73)专利权人 德稻全球创新网络(北京)有限公司
 地址 100080 北京市海淀区海淀北二街8号
 中关村soho1510室
(72)发明人 刘玉琴 李军 柳岸 王金秋
 李韦 朱东华 李维
(74)专利代理机构 北京天奇智新知识产权代理有限公司 11340
 代理人 刘黎明

(51)Int.Cl.
 G06F 17/27(2006.01)
(56)对比文件
 US 2010/0153094 A1,2010.06.17,
 US 2011/0191290 A1,2011.08.04,
 CN 103617219 A,2014.03.05,
 CN 104317897 A,2015.01.28,
 胡娟 等.基于主题图的学术博客知识组织模型研究.《图书情报工作》.2012,第56卷(第24期),第127-132页.
 李清茂.基于主题图的旅游文献组织方法研究.《应用实践》.2009,(第4期),第82-86页.
 审查员 王爽

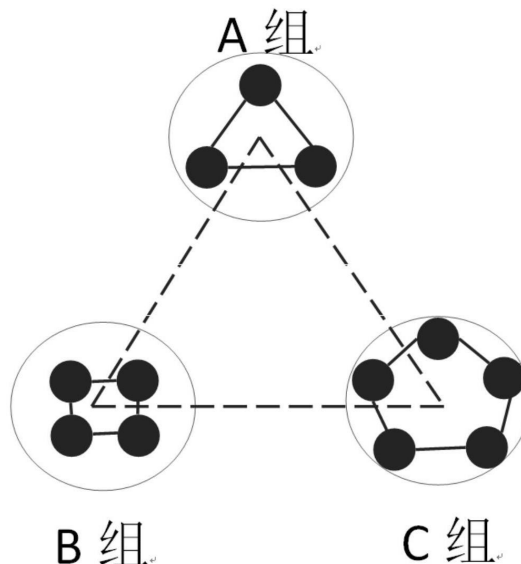
权利要求书3页 说明书7页 附图2页

(54)发明名称

一种利用知识地形图进行文本分析及系统

(57)摘要

本发明涉及一种文本分析的方法及系统,属于信息处理领域,具体涉及一种利用知识地形图进行文本分析的方法及系统。知识地形图的建立包括:坐标映射步骤,根据预设的规则将主题词划分为m组,将m组映射至平面中得到m个组坐标;计算各组内节点的坐标,并且将组内的中心点移至组坐标上;图形渲染步骤,用于根据主题词的关系强度建立反映像素点颜色值的密度函数;建立调色板以及所述调色板与密度函数的映射关系,根据所述映射关系渲染图形。利用本发明构建的知识地形图不仅简单易于实现,而且直观,可实现对大规模文本数据的快速浏览,挖掘出文本数据中的关键信息,可扩展性强。



1. 一种利用知识地形图进行文本分析的方法,其特征在于,知识地形图的建立包括:

坐标映射步骤,根据预设的规则将主题词划分为m组,将m组映射至平面中得到m个组坐标;计算各组内节点的坐标,并且将组内的中心点移至组坐标上;

图形渲染步骤,用于根据主题词的关系强度建立反映像素点颜色值的密度函数;建立调色板以及所述调色板与密度函数的映射关系,根据所述映射关系渲染图形。

2. 根据权利要求1所述的一种利用知识地形图进行文本分析的方法,其特征在于,所述坐标映射步骤进一步包括:

强度预处理子步骤,用于设定关系强度阈值,其中:大于该阈值的节点强度保留原值,小于等于该阈值的强度重新设定为0;

矩阵变换子步骤,将n个主题词分为m组,每个组内部的主题词关系强度均大于设定阈值;每个组之间存在至少一个节点与其它组的所有节点关系强度都小于等于设定阈值;通过行列变换将主题词的关系强度矩阵变换为 $Corr'_{n \times n}$:

$$Corr'_{n \times n} = \begin{bmatrix} & Group_1 & Group_2 & \cdots & Group_i & \cdots & Group_m \\ Group_1 & R_1 & 0 & \cdots & 0 & \cdots & 0 \\ Group_2 & 0 & R_2 & \cdots & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Group_i & 0 & 0 & \cdots & R_i & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Group_m & 0 & 0 & \cdots & 0 & \cdots & R_m \end{bmatrix}$$

其中 R_i 为第i组中主题词的关系强度矩阵, $Group_i$ 为第i组;

组坐标映射子步骤,将m个组作为平面中的m个节点,组与组之间的关系强度设定为同一个固定值;采用Fruchterman-Reingold layout算法对m个节点进行坐标计算得到组坐标;

组内坐标映射子步骤,对每个组内的节点采用VosMapping算法进行坐标计算,通过坐标平移将每个组的中心位置设定为组坐标。

3. 根据权利要求1所述的一种利用知识地形图进行文本分析的方法,其特征在于,所述密度函数利用以下公式:

$$Density(x, y) = \sum_{i=0}^n f(Numer_i) e^{-\alpha \left(\frac{\sqrt{(x-x_i)^2 + (y-y_i)^2}}{Distance} \right)^\beta}, \alpha > 0, \beta > 0;$$

其中, (x_i, y_i) , $i=1 \cdots n$ 是n个主题词的坐标, $Distance$ 是主题词之间的二维欧氏距离平均值;每个主题词的数量为 $Numer_i$, $i=1 \cdots n$,像素点Point的坐标 (x, y) ; $f(Numer_i)$ 为主题词数量的标准化值; α, β 为非负数。

4. 根据权利要求1所述的一种利用知识地形图进行文本分析的方法,其特征在于,所述图形渲染步骤包括:

函数标准化子步骤,用于将密度函数标准化,使其取值为0-255之间的整数;

调色板建立子步骤,用于建立一个256色的调色板,存储为256个元素的颜色向量;

关系映射子步骤,用于建立像素点密度值与调色板的一一映射关系;并根据映射关系进行图形渲染。

5. 根据权利要求4所述的一种利用知识地形图进行文本分析的方法,其特征在于,所述函数标准化子步骤采用如下变换方式:

$$(\text{int})\left(\gamma \frac{\text{Density}(x, y)}{\text{Density}_{\max}} \times 255\right);$$

或

$$(\text{int})\left(255 - \gamma \frac{\text{Density}(x, y)}{\text{Density}_{\max}} \times 255\right), \quad 0 < \gamma \leq 1;$$

其中, (int) 运算表示取整操作; Density_{max} 表示密度最大值。

6. 一种利用知识地形图进行文本分析的系统,其特征在于,通过下述模块建立知识地形图:

坐标映射模块,根据预设的规则将主题词划分为m组,将m组映射至平面中得到m个组坐标;计算各组内节点的坐标,并且将组内的中心点移至组坐标上;

图形渲染模块,用于根据主题词的关系强度建立反映像素点颜色值的密度函数;建立调色板以及所述调色板与密度函数的映射关系,根据所述映射关系渲染图形。

7. 根据权利要求6所述的一种利用知识地形图进行文本分析的系统,其特征在于,所述坐标映射模块进一步包括:

强度预处理单元,用于设定关系强度阈值,其中:大于该阈值的节点强度保留原值,小于等于该阈值的强度重新设定为0;

矩阵变换单元,将n个主题词分为m组,每个组内部的主题词关系强度均大于设定阈值;每个组之间存在至少一个节点与其它组的所有节点关系强度都小于等于设定阈值;通过行列变换将主题词的关系强度矩阵变换为Corr'_{n×n}:

$$\text{Corr}'_{n \times n} = \begin{bmatrix} & \text{Group}_1 & \text{Group}_2 & \cdots & \text{Group}_i & \cdots & \text{Group}_m \\ \text{Group}_1 & R_1 & 0 & \cdots & 0 & \cdots & 0 \\ \text{Group}_2 & 0 & R_2 & \cdots & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \text{Group}_i & 0 & 0 & \cdots & R_i & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \text{Group}_m & 0 & 0 & \cdots & 0 & \cdots & R_m \end{bmatrix}$$

其中R_i为第i组中主题词的关系强度矩阵,Group_i为第i组;

组坐标映射单元,将m个组作为平面中的m个节点,组与组之间的关系强度设定为同一个固定值;采用Fruchterman-Reingold layout算法对m个节点进行坐标计算得到组坐标;

组内坐标映射单元,对每个组内的节点采用VosMapping算法进行坐标计算,通过坐标平移将每个组的中心位置设定为组坐标。

8. 根据权利要求6所述的一种利用知识地形图进行文本分析的系统,其特征在于,所述密度函数利用以下公式:

$$\text{Density}(x, y) = \sum_{i=0}^n f(\text{Numer}_i) e^{-\alpha \left(\frac{\sqrt{(x-x_i)^2 + (y-y_i)^2}}{\text{Distance}} \right)^\beta}, \quad \alpha > 0, \beta > 0;$$

其中, (x_i, y_i) , $i=1 \cdots n$ 是 n 个主题词的坐标, $\overline{Distance}$ 是主题词之间的二维欧氏距离平均值; 每个主题词的数量为 $Numer_i$, $i=1 \cdots n$, 像素点 Point 的坐标 (x, y) ; $f(Numer_i)$ 为主题词数量的标准化值; α, β 为非负数。

9. 根据权利要求6所述的一种利用知识地形图进行文本分析的系统, 其特征在于, 所述图形渲染模块包括:

函数标准化单元, 用于将密度函数标准化, 使其取值为0-255之间的整数;

调色板建立单元, 用于建立一个256色的调色板, 存储为256个元素的颜色向量;

关系映射单元, 用于建立像素点密度值与调色板的一一映射关系; 并根据映射关系进行图形渲染。

10. 根据权利要求9所述的一种利用知识地形图进行文本分析的系统, 其特征在于, 所述函数标准化单元采用如下变换方式:

$$(\text{int})\left(\gamma \frac{Density(x, y)}{Density_{\max}} \times 255\right);$$

或

$$(\text{int})\left(255 - \gamma \frac{Density(x, y)}{Density_{\max}} \times 255\right), \quad 0 < \gamma \leq 1;$$

其中, (int) 运算表示取整操作; $Density_{\max}$ 表示密度最大值。

一种利用知识地形图进行文本分析的方法及系统

技术领域

[0001] 本发明涉及一种文本分析的方法及系统,属于信息处理领域,具体涉及一种利用知识地形图进行文本分析的方法及系统。

背景技术

[0002] 知识地形图通过类似于地理信息系统中的等高线图实现对文本数据的可视化,通过颜色的深浅区别数据的多少以及数据之间的关系。有些文献中也将其称为景观图或主题图,尽管名称和表现形式不完全相同,其基本思想是一致的。

[0003] 热力图是知识地形图的一种简单的变换形式,是对自然界的热力成像原理的计算机模拟,通过红黄蓝三种颜色的深浅来区别数据的多少,颜色块区别数据的密集程度。热力图的技术实现相对简单,但与知识地形图相比表现力略显不足。

[0004] 知识地形图主要应用在文本数据分析中。美国Sandia国家实验室开发的复杂网络分析工具VxInsight是一个影响比较大的知识地形图绘制工具。汤姆森·路透科技信息集团的Aureka专利地图,可视化效果比VxInsight更加精细、美观,其应用更加偏向于商业化用户。

[0005] 在主题词较多的情况下,知识地形图可视化效果仍然保持清晰。就知识地形图的绘制来说,有典型的等高线地形图、热力图、彩虹图等。等高线图实现简单,但不是所有的等高线绘制方法都能有效的表现分析结果。应用该方法最好的是汤姆森·路透科技的Aureka专利地图。其应用的神经网络算法+渲染技术较为复杂,普通的等高线地形图很难实现同类的表现效果。热力图、彩虹图实现手段简单。但与Aureka地形图相比,其可读性和表现力有所降低。

[0006] 本发明设计一种知识地形图的绘制方法,既能够达到Aureka地形图的可视化表现力,又能以相对简单的技术加以实现。

发明内容

[0007] 本发明主要是解决现有技术所存在的表现效果差、实现手段复杂的技术问题,提供了一种利用知识地形图进行文本分析的方法及系统。该方法及系统应用复杂网络Fruchterman-Reingold layout与VosMapping算法进行主题词的平面布局,建立以主题词数量为参数的平面像素点密度函数,映射平面上点的颜色,最后进行知识地形图的渲染,可视化表现力好,并且实现技术简单。

[0008] 本发明的上述技术问题主要是通过下述技术方案得以解决的:

[0009] 一种利用知识地形图进行文本分析的方法,知识地形图的建立包括:

[0010] 坐标映射步骤,根据预设的规则将主题词划分为m组,将m组映射至平面中得到m个组坐标;计算各组内节点的坐标,并且将组内的中心点移至组坐标上;

[0011] 图形渲染步骤,用于根据主题词的关系强度建立反映像素点颜色值的密度函数;建立调色板以及所述调色板与密度函数的映射关系,根据所述映射关系渲染图形。

[0012] 优化的,上述的一种利用知识地形图进行文本分析的方法,所述坐标映射步骤进一步包括:

[0013] 强度预处理子步骤,用于设定关系强度阈值,其中:大于该阈值的节点强度保留原值,小于等于该阈值的强度重新设定为0;

[0014] 矩阵变换子步骤,将n个主题词分为m组,每个组内部的主题词关系强度均大于设定阈值;每个组之间存在至少一个节点与其它组的所有节点关系强度都小于等于设定阈值;通过行列变换将主题词的关系强度矩阵变换为 $Corr'_{n \times n}$:

$$[0015] \quad Corr'_{n \times n} = \begin{bmatrix} & Group_1 & Group_2 & \cdots & Group_i & \cdots & Group_m \\ Group_1 & R_1 & 0 & \cdots & 0 & \cdots & 0 \\ Group_2 & 0 & R_2 & \cdots & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Group_i & 0 & 0 & \cdots & R_i & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Group_m & 0 & 0 & \cdots & 0 & \cdots & R_m \end{bmatrix}$$

[0016] 其中 R_i 为第i组中主题词的关系强度矩阵, $Group_i$ 为第i组;

[0017] 组坐标映射子步骤,将m个组作为平面中的m个节点,组与组之间的关系强度设定为同一个固定值;采用Fruchterman-Reingold layout算法对m个节点进行坐标计算得到组坐标;

[0018] 组内坐标映射子步骤,对每个组内的节点采用VosMapping算法进行坐标计算,通过坐标平移将每个组的中心位置设定为组坐标。

[0019] 优化的,上述的一种利用知识地形图进行文本分析的方法,所述密度函数基于以下公式:

$$[0020] \quad Density(x, y) = \sum_{i=0}^n f(Numer_i) e^{-\alpha \left(\frac{\sqrt{(x-x_i)^2 + (y-y_i)^2}}{Distance} \right)^\beta}, \alpha > 0, \beta > 0;$$

[0021] 其中, (x_i, y_i) , $i=1 \cdots n$ 是n个主题词的坐标, $\overline{Distance}$ 是主题词之间的二维欧氏距离平均值;每个主题词的数量为Numer, $i=1 \cdots n$, 像素点Point的坐标 (x, y) ; $f(Numer_i)$ 为主题词数量的标准化值; α, β 为非负数。

[0022] 优化的,上述的一种利用知识地形图进行文本分析的方法,所述图形渲染步骤包括:

[0023] 函数标准化子步骤,用于将密度函数标准化,使其取值为0-255之间的整数;

[0024] 调色板建立子步骤,用于建立一个256色的调色板,存储为256个元素的颜色向量;

[0025] 关系映射子步骤,用于建立像素点密度值与调色板的一一映射关系;并根据映射关系进行图像渲染。

[0026] 优化的,上述的一种利用知识地形图进行文本分析的方法,所述函数标准化子步骤采用如下变换方式:

$$[0027] \quad (\text{int}) \left(\gamma \frac{Density(x, y)}{Density_{\max}} \times 255 \right);$$

[0028] 或

$$[0029] \quad (\text{int})(255 - \gamma \frac{\text{Density}(x, y)}{\text{Density}_{\max}} \times 255), \quad 0 < \gamma \leq 1;$$

[0030] 其中, (int) 运算表示取整操作; Density_{\max} 表示密度最大值。

[0031] 一种利用知识地形图进行文本分析的系统, 通过下述模块建立知识地形图:

[0032] 坐标映射模块, 根据预设的规则将主题词划分为 m 组, 将 m 组映射至平面中得到 m 个组坐标; 计算各组内节点的坐标, 并且将组内的中心点移至组坐标上;

[0033] 图形渲染模块, 用于根据主题词的关系强度建立反映像素点颜色值的密度函数; 建立调色板以及所述调色板与密度函数的映射关系, 根据所述映射关系渲染图形。

[0034] 优化的, 上述的述的一种利用知识地形图进行文本分析的系统, 所述坐标映射模块进一步包括:

[0035] 强度预处理单元, 用于设定关系强度阈值, 其中: 大于该阈值的节点强度保留原值, 小于等于该阈值的强度重新设定为 0;

[0036] 矩阵变换单元, 将 n 个主题词分为 m 组, 每个组内部的主题词关系强度均大于设定阈值; 每个组之间存在至少一个节点与其它组的所有节点关系强度都小于等于设定阈值; 通过行列变换将主题词的关系强度矩阵变换为 $\text{Corr}'_{n \times n}$:

$$[0037] \quad \text{Corr}'_{n \times n} = \begin{bmatrix} & \text{Group}_1 & \text{Group}_2 & \cdots & \text{Group}_i & \cdots & \text{Group}_m \\ \text{Group}_1 & R_1 & 0 & \cdots & 0 & \cdots & 0 \\ \text{Group}_2 & 0 & R_2 & \cdots & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \text{Group}_i & 0 & 0 & \cdots & R_i & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \text{Group}_m & 0 & 0 & \cdots & 0 & \cdots & R_m \end{bmatrix}$$

[0038] 其中 R_i 为第 i 组中主题词的关系强度矩阵, Group_i 为第 i 组;

[0039] 组坐标映射单元, 将 m 个组作为平面中的 m 个节点, 组与组之间的关系强度设定为同一个固定值; 采用 Fruchterman-Reingold layout 算法对 m 个节点进行坐标计算得到组坐标;

[0040] 组内坐标映射单元, 对每个组内的节点采用 VosMapping 算法进行坐标计算, 通过坐标平移将每个组的中心位置设定为组坐标。

[0041] 优化的, 上述的述的一种利用知识地形图进行文本分析的系统, 所述密度函数基于以下公式:

$$[0042] \quad \text{Density}(x, y) = \sum_{i=0}^n f(\text{Numer}_i) e^{-\alpha \left(\frac{\sqrt{(x-x_i)^2 + (y-y_i)^2}}{\text{Distance}} \right)^\beta}, \quad \alpha > 0, \beta > 0;$$

[0043] 其中, $(x_i, y_i), i = 1 \cdots n$ 是 n 个主题词的坐标, $\overline{\text{Distance}}$ 是主题词之间的二维欧氏距离平均值; 每个主题词的数量为 $\text{Numer}, i = 1 \cdots n$, 像素点 Point 的坐标 (x, y) ; $f(\text{Numer}_i)$ 为主题词数量的标准化值; α, β 为非负数。

[0044] 优化的, 上述的述的一种利用知识地形图进行文本分析的系统, 所述图形渲染模块包括:

[0045] 函数标准化单元, 用于将密度函数标准化, 使其取值为 0-255 之间的整数;

[0046] 调色板建立单元,用于建立一个256色的调色板,存储为256个元素的颜色向量;

[0047] 关系映射单元,用于建立像素点密度值与调色板的一一映射关系;并根据映射关系进行图像渲染。

[0048] 优化的,上述的述的一种利用知识地形图进行文本分析的系统,所述函数标准化单元采用如下变换方式:

$$[0049] \quad (\text{int})\left(\gamma \frac{\text{Density}(x, y)}{\text{Density}_{\max}} \times 255\right);$$

[0050] 或

$$[0051] \quad (\text{int})\left(255 - \gamma \frac{\text{Density}(x, y)}{\text{Density}_{\max}} \times 255\right), \quad 0 < \gamma \leq 1;$$

[0052] 其中, (int) 运算表示取整操作; Density_{\max} 表示密度最大值。

[0053] 因此,本发明具有如下优点:利用本发明构建的知识地形图不仅简单易于实现,而且直观,可实现对大规模文本数据的快速浏览,挖掘出文本数据中的关键信息,可扩展性强。

附图说明

[0054] 图1:主题词平面布局算法应用示意图

[0055] 图2:电脑屏幕像素单元格划分示意图

[0056] 图3:采用本发明绘制的知识地形图效果图

具体实施方式

[0057] 下面通过实施例,并结合附图,对本发明的技术方案作进一步具体的说明。

[0058] 实施例:

[0059] 实施例1:

[0060] 步骤一:采用文本挖掘方法抽取文本数据中的主题词,获取主题词列表。

[0061] 步骤二:主题词关系强度计算。计算主题词的同现关键矩阵,依据同现矩阵计算各主题词之间的关系强度矩阵,计算方法可以采取倒排文档频率、信息熵、互信息等,也可以直接采用同现数量度量关系强度。假设n个主题词之间的关系强度矩阵为 $\text{Corr}_{n \times n}$ 。

$$[0062] \quad \text{Corr}_{n \times n} = \begin{bmatrix} & \text{Keyword}_1 & \text{Keyword}_2 & \cdots & \text{Keyword}_i & \cdots & \text{Keyword}_n \\ \text{Keyword}_1 & r_{11} & r_{12} & \cdots & r_{1i} & \cdots & r_{1n} \\ \text{Keyword}_2 & r_{21} & r_{22} & \cdots & r_{2i} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \text{Keyword}_i & r_{i1} & r_{i2} & \cdots & r_{ii} & \cdots & r_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \text{Keyword}_n & r_{n1} & r_{n2} & \cdots & r_{ni} & \cdots & r_{nn} \end{bmatrix}$$

[0063] 步骤三:主题词平面布局

[0064] 为了绘制知识地形图,需要确定主题词在平面图中的位置坐标,操作过程如下:

[0065] a. 设定关系强度阈值,大于该阈值的节点强度保留原值,小于等于该阈值的强度

重新设定为0。

[0066] b. 将n个主题词分为m组,满足每个组内部的主题词关系强度均大于设定阈值;每个组之间存在至少一个节点与其它组的所有节点关系强度都小于等于设定阈值。通过行列变换将主题词的关系强度矩阵变换为 $Corr'_{n \times n}$

$$[0067] \quad Corr'_{n \times n} = \begin{bmatrix} & Group_1 & Group_2 & \cdots & Group_i & \cdots & Group_m \\ Group_1 & R_1 & 0 & \cdots & 0 & \cdots & 0 \\ Group_2 & 0 & R_2 & \cdots & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Group_i & 0 & 0 & \cdots & R_i & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Group_m & 0 & 0 & \cdots & 0 & \cdots & R_m \end{bmatrix}$$

[0068] 其中 R_i 为第i组中主题词的关系强度矩阵。

[0069] c. 将m个组作为平面中的m个节点,组与组之间的关系强度设定为同一个固定值。采用Fruchterman-Reingold layout算法对m个节点进行坐标计算。

[0070] d. 对每个组内的节点采用VosMapping算法进行坐标计算。

[0071] 步骤四:基于主题词数量与布局坐标构建平面像素点的密度函数

[0072] 主题词的坐标确定后,将其绘制到计算机屏幕,需要确定每个象素点的颜色。为此,建立一个密度函数,用于影射每个象素点的颜色值。

[0073] 假设:n个主题词的坐标分别为 (x_i, y_i) , $i=1 \cdots n$,主题词之间的二维欧氏距离平均值为 $\overline{Distance}$,每个主题词的数量为Numer, $i=1 \cdots n$,像素点Point的坐标 (x, y) 。

[0074] 定义像素点的密度函数公式为:

$$[0075] \quad Density(x, y) = \sum_{i=0}^n f(Numer_i) e^{-\alpha \left(\frac{\sqrt{(x-x_i)^2 + (y-y_i)^2}}{Distance} \right)^\beta}, \alpha > 0, \beta > 0$$

[0076] 其中, $f(Numer_i)$ 为主题词数量的标准化值; α, β 为非负数, 其取值不同, 地形图效果不同。

[0077] 步骤五:计算像素点的色彩进行地形图渲染

[0078] a. 将密度函数标准化,使其取值为0-255之间的整数,可以采用如下变换方式:

$$[0079] \quad (\text{int})\left(\gamma \frac{Density(x, y)}{Density_{\max}} \times 255\right) \text{ 或 } (\text{int})\left(255 - \gamma \frac{Density(x, y)}{Density_{\max}} \times 255\right), \quad 0 < \gamma \leq 1$$

[0080] b. 建立一个256色的调色板,存储为256个元素的颜色向量。

[0081] c. 建立像素点密度值与调色板的一一映射关系。举例:如果像素点标准化后的密度值为200,该像素点对应的颜色为调色板中第200个颜色的色彩。

[0082] 实施例2:

[0083] 步骤一:采用文本分词方法获取文本数据集中的主题词。

[0084] 步骤二:采用主题词同现频数作为主题词关系强度值。

[0085] 步骤三:应用Fruchterman-Reingold layout算法和VOSMapping算法对主题词平面坐标进行计算。

[0086] 如图示1,假设有12个主题词,分为A、B、C三组,关系强度矩阵分别为 R_1, R_2, R_3 。在对

12个主题词进行布局的过程中,首先把A、B、C三个组看作三个节点,节点距离(图中虚线)相等。采用Fruchterman-Reingold layout算法对这三个节点进行布局,记录每个节点的中心位置。对每个分组内部的节点,如A组内的三个节点、B组内的四个节点、C组内的5个节点分别采用VosMapping算法进行布局,然后通过坐标平移将每个组的中心位置设定为通过Fruchterman-Reingold layout算法得到的三个点的坐标。

[0087] 步骤四:构建平面像素点的密度函数, $\alpha=0.5, \beta=2$

$$[0088] \quad Density(x, y) = \sum_{i=0}^n Numer_i e^{-0.5 \left(\frac{\sqrt{(x-x_i)^2 + (y-y_i)^2}}{Distance} \right)^2}$$

[0089] 为减少计算量,降低地形图的渲染时间延迟,并不是每个像素的密度函数都需要进行计算,而是将整个屏幕划分为若干格子,每个格子作为一个像素点对待,计算每个格子的密度函数,其示意图如2所示。最后,通过图形拉伸使图形与电脑屏幕重合。

[0090] 步骤五:计算像素点的色彩进行地形图渲染。建立256色的RGB调色板{-392960,-392448,-391680,-391168,-390656,-389888,-389376,-388608,-388096,-387584,-321280,-321024,-320256,-319744,-319232,-318464,-317952,-317184,-316672,-316160,-315392,-314880,-314368,-313856,-313088,-312576,-312064,-311296,-310784,-310272,-309504,-308992,-242944,-242176,-241664,-241152,-240640,-239872,-239360,-238848,-238080,-237568,-236800,-236288,-235776,-235008,-234752,-233984,-233472,-232960,-232192,-231680,-230912,-164864,-164352,-163584,-163072,-162560,-162048,-161280,-160768,-160256,-159488,-158976,-158208,-157696,-157184,-156416,-155904,-155392,-154880,-154368,-153600,-153088,-152320,-86272,-85760,-84992,-84480,-83968,-83456,-82944,-82176,-81664,-80896,-80384,-79872,-79104,-78592,-78080,-77312,-77056,-76288,-75776,-75008,-74496,-8448,-7680,-7168,-6400,-5888,-5376,-4608,-4352,-3584,-3072,-2560,-133376,-264448,-461056,-592128,-788736,-919808,-1116160,-1181696,-1378304,-1509376,-1705984,-1837056,-2033664,-2164736,-2361344,-2492416,-2623488,-2820096,-2951168,-3082240,-3278592,-3409664,-3540736,-3737344,-3868416,-4065024,-4196096,-4392704,-4523776,-4720384,-4785920,-4982528,-5113600,-5309952,-5441024,-5637632,-5768704,-5965312,-6096384,-6227456,-6424064,-6555136,-6686208,-6882816,-7013888,-7144960,-7341312,-7472384,-7668992,-7800064,-7996672,-8127744,-8324352,-8389888,-8586496,-8717568,-8914176,-9045248,-9241856,-9372928,-9569280,-9700352,-9831424,-10028032,-10159104,-10290176,-10486784,-10617856,-10748928,-10945536,-11076608,-11273216,-11404288,-11600640,-11731712,-11928320,-11993856,-12190464,-12321536,-12518144,-12649216,-12845824,-12976896,-13173504,-13304576,-13435648,-13632000,-13763072,-13894144,-14090752,-14221824,-14352896,-14549504,-14680576,-14877184,-15008256,-15204864,-15335936,-15532544,-15598080,-15794432,-15925504,-16122112,-16253184,-16449792,-16580864,-16712191,-16713211,-16714486,-16715761,-16717037,-16718312,-16719587,-

16720606,-16721882,-16723157,-16724432,-16725707,-16726727,-16728002,-16729278,-16730553,-16731572,-16732847,-16734123,-16735143,-16736418,-16737693,-16738968,-16740243,-16741263,-16742538,-16743813,-16745089,-16746364,-16747639,-16748659,-16749934,-16751209,-16752484,-16753759,-16754779,-16756055,-16757330,-16758349,-16759624,-16760899,-16762175,-16762175} ;

[0091] 进行像素点密度函数与画板的映射关系,取 $\gamma = 1$,进行地形图的绘制,得到如图3形式的知识地形图。

[0092] 本文中所描述的具体实施例仅仅是对本发明精神作举例说明。本发明所属技术领域的技术人员可以对所描述的具体实施例做各种各样的修改或补充或采用类似的方式替代,但并不会偏离本发明的精神或者超越所附权利要求书所定义的范围。

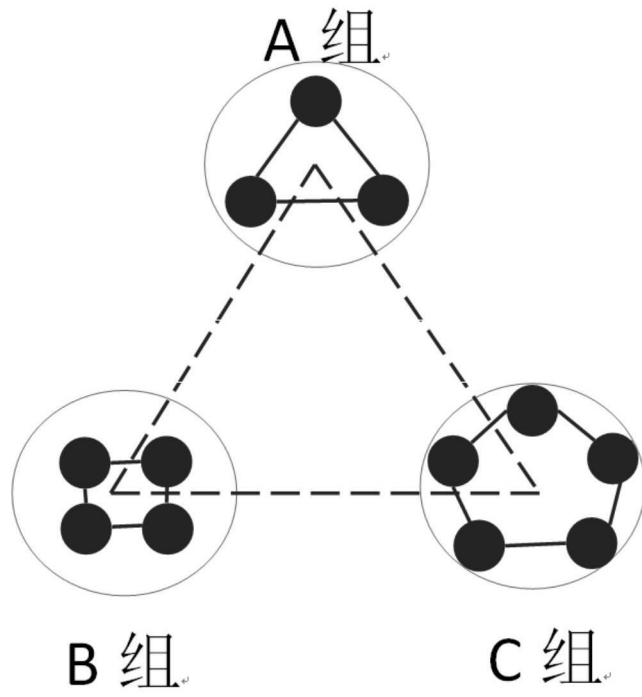


图1

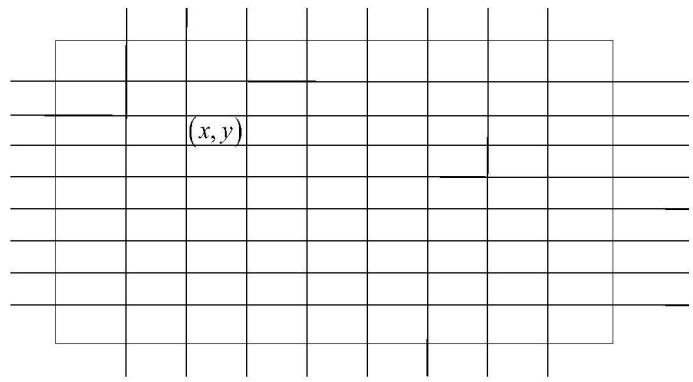


图2

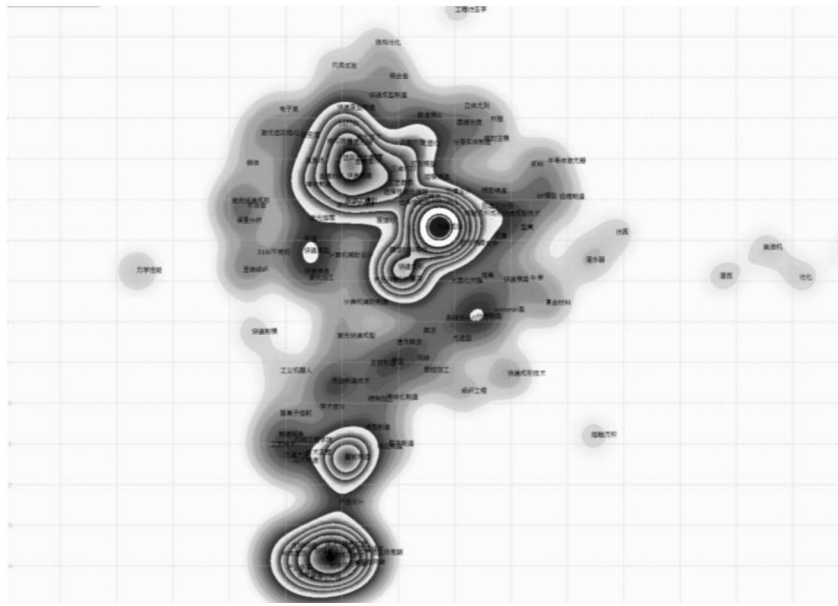


图3