(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115186107 A
(43)申请公布日 2022.10.14

(21)申请号 202210928186.1

(22)申请日 2022.08.03

(71)申请人 北京印刷学院
地址 102600 北京市大兴区兴华大街二段1号

(72)发明人 刘玉琴 任慧超 门川琨 汪雪锋

(74)专利代理机构 北京智行阳光知识产权代理事务所(普通合伙) 11738
专利代理师 刘颖

(51)Int.Cl.
G06F 16/36(2019.01)
G06F 16/35(2019.01)
G06F 17/10(2006.01)
G06F 17/16(2006.01)
G06F 40/289(2020.01)

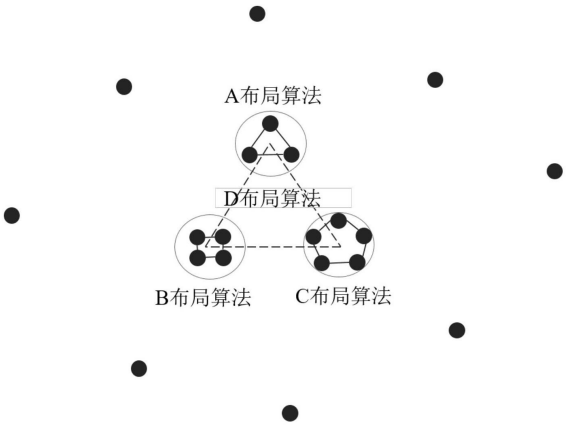
权利要求书3页 说明书7页 附图1页

(54)发明名称

一种基于技术主题图进行技术竞争力测度的方法

(57)摘要

本发明公开了一种基于技术主题图进行技术竞争力测度的方法,基于技术主题图,构建技术竞争力测度模型,用具体数字对企业或研发机构之间的技术竞争力进行度量,使决策者在观测技术主题图的同时不仅有视觉上的直观感知,也有更加直接的、更加精细化的数值参考。



CN 115186107 A

1.一种基于技术主题图进行技术竞争力测度的方法,其特征在于,具体过程为:

S1、对科技文本数据集进行分词处理,计算各个科技文本与主题词之间的隶属关系矩阵;

S2、基于各个科技文本与主题词之间的隶属关系矩阵,计算所有主题词之间的关系强度矩阵;

S3、根据步骤S2得到的关系强度矩阵,按照各个主题词之间的关系强度,应用聚类算法对所有主题词进行聚类,根据聚类结果为每个主题词加上类别标签;记聚类后的类别数为C,即主题词分为C组;之后,通过不同的布局算法将主题词映射到空间平面中的点;

S4、构建平面像素点类密度函数进行可视化:

S4.1、假设n个主题词的坐标分别为(x_i,y_i),i=1…n,主题词之间的二维欧氏距离平均值为$\overline{Distance}$;$Number_i$,i=1…n,表示出现了主题词i的科技文本数量;经过聚类后共有C个类别,每个类别下分别有n_c个主题词;$f(Number_i)$为主题词i的标准化值;像素点P的坐标(x,y);其中,密度函数α,β为非负数;

定义像素点的密度函数和类密度函数为:

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

类密度函数:$$Density(x,y,c)=\sum_{i=0}^{n_c} f(Numer_i)e^{-\alpha\left(\frac{\sqrt{(x-x_i)^2+(y-y_i)^2}}{\overline{Distance}}\right)^{\beta}},\alpha>0,\beta>0$$;

c表示聚类后的某一具体类别;

S4.2、融合聚类信息后,用$Density_{max}$表示最大的密度值,$Color_i$表示类别i=1…C的RGB模式颜色;像素点P(x,y)的RGB模式颜色计算如下:

$$Color(x,y)=\sum_{i=0}^{n_c} Density(x,y,c_i)/Density_{max}\times Color_i$$

其中,$Color_i$是RGB模式颜色的各通道取值;

S4.3、为了实现类似地形图等高线的可视化效果,同时等高线既能对同一类别下的主题词进行区分,又能对不同类别下的主题词进行区分,构建色彩强度函数:

f(Density(x,y)/Density(max));

S5、构建企业或研发机构i在技术主题j下的竞争力测度模型:

w为参与竞争力测度的企业或研发机构的数量;s为技术领域涵盖的技术主题数量;$n_{i,j}$为企业或研发机构i在技术主题j的文献数量,$Position(x_k)$为科技文本k的技术地位,$Ability(x_k)$为科技文本k的质量;$Position(x_k)$取值为步骤S4.4中计算所得的色彩强度;因此,企业或研发机构i在技术主题j下的竞争力测度模型如下:

$$Competive(Corpration_i,Techno\log y_j)=\sum_{k=1}^{i=n_{i,j}} Position(D_k)\times Ability(D_k):$$

$$i=1\cdots w,j=1\cdots s$$

$Corpration_i$表示企业或研发机构i,$Technology_j$表示技术主题j;

S6、形成如下形式的机构、技术主题之间的竞争力矩阵。

2.根据权利要求1所述的方法，其特征在于，步骤S1中，所述隶属关系矩阵表示如下：

$$\begin{bmatrix} & Keyword_1 & Keyword_2 & \cdots & Keyword_j & \cdots & Keyword_n \\ Document_1 & b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1n} \\ Document_2 & b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Document_i & b_{i1} & b_{i2} & \cdots & b_{ij} & \cdots & b_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Document_m & b_{m1} & b_{m2} & \cdots & b_{mj} & \cdots & b_{mn} \end{bmatrix}$$

其中，m表示科技文本数量，n表示主题词数量，$Document_i$表示第i个科技文本，$Keyword_j$表示第j个主题词，b_{ij}表示第i个科技文本第j主题词出现的数量。

3.根据权利要求1所述的方法，其特征在于，步骤S2中，所述关系强度矩阵表示如下：

$$\begin{bmatrix} & Keyword_1 & Keyword_2 & \cdots & Keyword_j & \cdots & Keyword_n \\ Keyword_1 & r_{11} & r_{12} & \cdots & r_{1j} & \cdots & r_{1n} \\ Keyword_2 & r_{21} & r_{22} & \cdots & r_{2j} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Keyword_i & r_{i1} & r_{i2} & \cdots & r_{ij} & \cdots & r_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Keyword_n & r_{n1} & r_{n2} & \cdots & r_{nj} & \cdots & r_{nn} \end{bmatrix}$$

其中，n表示主题词数量，$Keyword_i$、$Keyword_j$表示第i、j个主题词，r_{ij}表示第i个主题词与第j主题词的共同出现的科技文本数量。

4.根据权利要求1所述的方法，其特征在于，步骤S4.3中，色彩强度函数具体为：

$$Strength(x,y)=\lfloor (Density(x,y)/Density(\max))\times N \rfloor / N$$

其中，$\lfloor\ \rfloor$为向下取整，N为强度级别数量。

5.根据权利要求4所述的方法，其特征在于，步骤S5中的竞争力测度模型可以转化为：

$$Competive(Corpration_i, Technology_j)$$

$$=\sum_{k=1}^{i=n_{i,j}} Position(D_k)\times Ability(D_k)$$

$$=\sum_{k=1}^{i=n_{i,j}} \lfloor (Density(D_k(x,y))/Density(\max))\times N \rfloor / N \times Ability(D_k)$$

其中$D_k(x,y)$为科技文本k在技术主题图中的坐标。

6.根据权利要求5所述的方法，其特征在于，$Ability(x_k)$采用科技文本被引用数量与引用该科技文本的机构数量比进行表示，即以科技文本被引用数量和被引用机构覆盖情况评价单个科技文本的质量，由此进一步将竞争力测度模型转化如下：

$$Competive(Corpration_i, Techno\log y_j)$$

$$= \sum_{k=1}^{i=n_{i,j}} Position(D_k) \times Ability(D_k)$$

$$= \sum_{k=1}^{i=n_{i,j}} \lfloor (Density(D_k(x,y)/Density(max)) \times N \rfloor / N$$

$$\times ReferencedNumber(D_k)/CorprationNumber(D_k)$$

其中，ReferencedNumber(D_k)为科技文本k的被引用数量，CorprationNumber(D_k)为引用了科技文本k的机构数量。

7.根据权利要求1所述的方法，其特征在于，对竞争力测度进行性归一化处理使其介于0-1之间，便于比较；计算公式如下：

$$\frac{Competive(Corpration_i, Technology_j)}{\underset{i=1\cdots w, j=1\cdots s}{Max}(Competive(Corpration_i, Technology_j))}$$。

8.根据权利要求1所述的方法，其特征在于，步骤S6中，所述竞争力矩阵表示如下：

$$\begin{bmatrix} & Technology_1 & Technology_2 & \cdots & Technology_j & \cdots & Technology_s \\ Corpration_1 & q_{11} & q_{12} & \cdots & q_{1j} & \cdots & q_{1s} \\ Corpration_2 & q_{21} & q_{22} & \cdots & q_{2j} & \cdots & q_{2s} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Corpration_i & q_{i1} & q_{i2} & \cdots & q_{ij} & \cdots & q_{is} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Corpration_w & q_{w1} & q_{w2} & \cdots & q_{wj} & \cdots & q_{ws} \end{bmatrix}$$

其中，q_{ij}为企业或研发机构i在技术主题j下技术竞争力测度值。

一种基于技术主题图进行技术竞争力测度的方法

技术领域

[0001] 本发明涉及文本数据处理技术领域，具体涉及一种基于技术主题图进行技术竞争力测度的方法。

背景技术

[0002] 随着文本挖掘和信息可视化技术的不断发展，出现了众多针对专利、论文、技术标准、研究报告等科技文本数据的分析方法。技术主题图融合了文本挖掘和信息可视化技术，从科技文本数据中挖掘技术信息，用直观的图形表现技术布局，被广泛应用于技术布局分析、技术竞争分析、技术结构分析等分析场景。基于技术主题图的科技文本数据分析从宏观角度为创新决策者提供视觉上的支撑，但视觉感知是全局的、粗粒度的、非精细化的，决策者在决策过程中往往需要更加准确的数值度量辅助决策。

[0003] 现有技术竞争力测度方法单独构建竞争力模型，没有与技术主题图相融合，具体可分为基于数据包络分析法(DEA)，如运用序列DEA方法测算中国工业行业与主要由OECD国家工业行业所构成的世界技术前沿的技术差距[1]。利用共同前沿理论，构建并联网络DEA模型测算2007年至2015年中国高技术制造业创新效率和区域间的技术差距等[2]。基于函数模型方法，如引入双国超越对数生产模型估算中美制造业分行业的全要素生产率差距[3]。引入超技术函数分析我国地区间的农业技术差距[4]。这些方法通过单一数值测度技术竞争力，没有与技术主题图进行融合，缺乏视觉的全局直观感知。

[0004] 采用技术主题图进行科技文本数据的技术分析，视觉呈现上一目了然、通俗易懂，但缺乏精细化地、细粒度地决策数据参考。技术主题图的可视化呈现上，主题密度图的呈现可辨识度和美观程度不足；可视化内容呈现多从宏观态势上分析技术竞争态势，缺乏细粒度地、精细化地、准确地数值度量参考。

[0005] 参考文献：

[0006] [1]陆剑,柳剑平,程时雄.中国与OECD主要国家工业行业技术差距的动态测度[J].世界经济,2014,37(09):25-52.

[0007] [2]肖仁桥,陈忠卫,钱丽.异质性技术视角下中国高技术制造业创新效率研究[J].管理科学,2018,31(01):48-68.

[0008] [3]黄勇峰,任若恩.中美两国制造业全要素生产率比较研究[J].经济学(季刊),2002(04):161-180.

[0009] [4]杨国涛.我国地区间农业技术差距的一种超函数测算方法[J].安徽农业科学,2007(22):6693-6694.DOI:10.13989/j.cnki.0517-6611.2007.22.030.

发明内容

[0010] 针对现有技术的不足，本发明旨在提供一种基于技术主题图进行技术竞争力测度的方法。

[0011] 为了实现上述目的，本发明采用如下技术方案：

[0012] 一种基于技术主题图进行技术竞争力测度的方法，具体过程为：

[0013] S1、对科技文本数据集进行分词处理，计算各个科技文本与主题词之间的隶属关系矩阵；

[0014] S2、基于各个科技文本与主题词之间的隶属关系矩阵，计算所有主题词之间的关系强度矩阵；

[0015] S3、根据步骤S2得到的关系强度矩阵，按照各个主题词之间的关系强度，应用聚类算法对所有主题词进行聚类，根据聚类结果为每个主题词加上类别标签；记聚类后的类别数为C，即主题词分为C组；之后，通过不同的布局算法将主题词映射到空间平面中的点；

[0016] S4、构建平面像素点类密度函数进行可视化：

[0017] S4.1、假设n个主题词的坐标分别为(x_i,y_i)，i＝1…n，主题词之间的二维欧氏距离平均值为$\overline{Distance}$；$Number_i$，i＝1…n，表示出现了主题词i的科技文本数量；经过聚类后共有C个类别，每个类别下分别有n_c个主题词；$f(Number_i)$为主题词i的标准化值；像素点P的坐标(x,y)；其中，密度函数α,β为非负数；

[0018] 定义像素点的密度函数和类密度函数为：

[0019] 密度函数：$Density(x,y)=\sum_{i=0}^{n} f(Numer_i)e^{-\alpha\left(\frac{\sqrt{(x-x_i)^2+(y-y_i)^2}}{Distance}\right)^{\beta}},\alpha>0,\beta>0$；

[0020] 类密度函数：$Density(x,y,c)=\sum_{i=0}^{n_c} f(Numer_i)e^{-\alpha\left(\frac{\sqrt{(x-x_i)^2+(y-y_i)^2}}{Distance}\right)^{\beta}},\alpha>0,\beta>0$；

[0021] c表示聚类后的某一具体类别；

[0022] S4.2、融合聚类信息后，用$Density_{max}$表示最大的密度值，$Color_i$表示类别i＝1…C的RGB模式颜色；像素点P(x,y)的RGB模式颜色计算如下：

[0023] $$Color(x,y)=\sum_{i=0}^{n_c} Density(x,y,c_i)/Density_{max}\times Color_i$$

[0024] 其中，$Color_i$是RGB模式颜色的各通道取值；

[0025] S4.3、为了实现类似地形图等高线的可视化效果，同时等高线既能对同一类别下的主题词进行区分，又能对不同类别下的主题词进行区分，构建色彩强度函数：

[0026] f(Density(x,y)/Density(max))；

[0027] S5、构建企业或研发机构i在技术主题j下的竞争力测度模型：

[0028] w为参与竞争力测度的企业或研发机构的数量；s为技术领域涵盖的技术主题数量；$n_{i,j}$为企业或研发机构i在技术主题j的文献数量，$Position(x_k)$为科技文本k的技术地位，$Ability(x_k)$为科技文本k的质量；$Position(x_k)$取值为步骤S4.4中计算所得的色彩强度；因此，企业或研发机构i在技术主题j下的竞争力测度模型如下：

[0029] $$Competive(Corpration_i,Techno\log y_j)=\sum_{k=1}^{i=n_{i,j}} Position(D_k)\times Ability(D_k)$$；

$$i=1\cdots w,j=1\cdots s$$

[0030] $Corpration_i$表示企业或研发机构i，$Technology_j$表示技术主题j。

[0031] S6、形成如下形式的机构、技术主题之间的竞争力矩阵。

[0032] 进一步地，步骤S1中，所述隶属关系矩阵表示如下：

[0033]
$$\begin{bmatrix} & Keyword_1 & Keyword_2 & \cdots & Keyword_j & \cdots & Keyword_n \\ Document_1 & b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1n} \\ Document_2 & b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Document_i & b_{i1} & b_{i2} & \cdots & b_{ij} & \cdots & b_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Document_m & b_{m1} & b_{m2} & \cdots & b_{mj} & \cdots & b_{mn} \end{bmatrix}$$

[0034] 其中，m表示科技文本数量，n表示主题词数量，$Document_i$表示第i个科技文本，$Keyword_j$表示第j个主题词，b_{ij}表示第i个科技文本第j主题词出现的数量。

[0035] 进一步地，步骤S2中，所述关系强度矩阵表示如下：

[0036]
$$\begin{bmatrix} & Keyword_1 & Keyword_2 & \cdots & Keyword_j & \cdots & Keyword_n \\ Keyword_1 & r_{11} & r_{12} & \cdots & r_{1j} & \cdots & r_{1n} \\ Keyword_2 & r_{21} & r_{22} & \cdots & r_{2j} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Keyword_i & r_{i1} & r_{i2} & \cdots & r_{ij} & \cdots & r_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Keyword_n & r_{n1} & r_{n2} & \cdots & r_{nj} & \cdots & r_{nn} \end{bmatrix}$$

[0037] 其中，n表示主题词数量，$Keyword_i$、$Keyword_j$表示第i、j个主题词，r_{ij}表示第i个主题词与第j主题词的共同出现的科技文本数量。

[0038] 进一步地，步骤S4.3中，色彩强度函数具体为：

[0039]
$$Strength(x,y)=\lfloor (Density(x,y)/Density(\max))\times N \rfloor / N$$

[0040] 其中，$\lfloor\ \rfloor$为向下取整，N为强度级别数量。

[0041] 更进一步地，步骤S5中的竞争力测度模型可以转化为：

[0042]
$$Competive(Corpration_i, Technology_j)$$
$$=\sum_{k=1}^{i=n_{i,j}} Position(D_k)\times Ability(D_k)$$
$$=\sum_{k=1}^{i=n_{i,j}} \lfloor (Density(D_k(x,y))/Density(\max))\times N \rfloor / N \times Ability(D_k)$$

[0043] 其中$D_k(x,y)$为科技文本k在技术主题图中的坐标。

[0044] 再进一步地，$Ability(x_k)$采用科技文本被引用数量与引用该科技文本的机构数量比进行表示，即以科技文本被引用数量和被引用机构覆盖情况评价单个科技文本的质量，由此进一步将竞争力测度模型转化如下：

[0045]
$$Competive(Corpration_i, Techno\log y_j)$$
$$= \sum_{k=1}^{i=n_{i,j}} Position(D_k) \times Ability(D_k)$$
$$= \sum_{k=1}^{i=n_{i,j}} \lfloor (Density(D_k(x,y)) / Density(max)) \times N \rfloor / N$$
$$\times ReferencedNumber(D_k) / CorprationNumber(D_k)$$

[0046] 其中，ReferencedNumber(D_k)为科技文本k的被引用数量，CorprationNumber(D_k)为引用了科技文本k的机构数量。

[0047] 进一步地，对竞争力测度进行性归一化处理使其介于0-1之间，便于比较；计算公式如下：

[0048]
$$\frac{Competive(Corpration_i, Technology_j)}{\underset{i=1\cdots w, j=1\cdots s}{Max} (Competive(Corpration_i, Technology_j))}。$$

[0049] 进一步地，步骤S6中，所述竞争力矩阵表示如下：

[0050]
$$\begin{bmatrix} & Technology_1 & Technology_2 & \cdots & Technology_j & \cdots & Technology_s \\ Corpration_1 & q_{11} & q_{12} & \cdots & q_{1j} & \cdots & q_{1s} \\ Corpration_2 & q_{21} & q_{22} & \cdots & q_{2j} & \cdots & q_{2s} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Corpration_i & q_{i1} & q_{i2} & \cdots & q_{ij} & \cdots & q_{is} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Corpration_w & q_{w1} & q_{w2} & \cdots & q_{wj} & \cdots & q_{ws} \end{bmatrix}$$

[0051] 其中，q_{ij}为企业或研发机构i在技术主题j下技术竞争力测度值。

[0052] 本发明的有益效果在于：本发明提供的基于技术主题图分析测度技术竞争力的方法，基于技术主题图，构建技术竞争力测度模型，用具体数字对企业或研发机构之间的技术竞争力进行度量，使决策者在观测技术主题图的同时不仅有视觉上的直观感知，也有更加直接的、更加精细化的数值参考。

附图说明

[0053] 图1为本发明实施例中技术主题图布局算法融合示意图；

[0054] 图2为本发明实施例中基于技术主题图的技术竞争力测度模型构建思路图。

具体实施方式

[0055] 以下将结合附图对本发明作进一步的描述，需要说明的是，本实施例以本技术方案为前提，给出了详细的实施方式和具体的操作过程，但本发明的保护范围并不限于本实施例。

[0056] 本实施例提供一种基于技术主题图进行技术竞争力测度的方法，具体过程为：

[0057] S1、对科技文本数据集进行分词处理，计算各个科技文本与主题词之间的隶属关系矩阵。所述隶属关系矩阵表示如下：

[0058]
$$\begin{bmatrix} & Keyword_1 & Keyword_2 & \cdots & Keyword_j & \cdots & Keyword_n \\ Document_1 & b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1n} \\ Document_2 & b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Document_i & b_{i1} & b_{i2} & \cdots & b_{ij} & \cdots & b_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Document_m & b_{m1} & b_{m2} & \cdots & b_{mj} & \cdots & b_{mn} \end{bmatrix}$$

[0059] 其中，m表示科技文本数量，n表示主题词数量，$Document_i$表示第i个科技文本，$Keyword_j$表示第j个主题词，b_{ij}表示第i个科技文本第j主题词出现的数量。

[0060] S2、基于各个科技文本与主题词之间的隶属关系矩阵，计算所有主题词之间的关系强度矩阵。所述关系强度矩阵表示如下：

[0061]
$$\begin{bmatrix} & Keyword_1 & Keyword_2 & \cdots & Keyword_j & \cdots & Keyword_n \\ Keyword_1 & r_{11} & r_{12} & \cdots & r_{1j} & \cdots & r_{1n} \\ Keyword_2 & r_{21} & r_{22} & \cdots & r_{2j} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Keyword_i & r_{i1} & r_{i2} & \cdots & r_{ij} & \cdots & r_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Keyword_n & r_{n1} & r_{n2} & \cdots & r_{nj} & \cdots & r_{nn} \end{bmatrix}$$

[0062] 所述关系强度矩阵具体由隶属关系矩阵的转置与隶属关系矩阵的乘积计算得到，或者基于隶属关系矩阵，采用倒排文档频率、信息熵、互信息等方式计算得到。其中，n表示主题词数量，$Keyword_i$、$Keyword_j$表示第i、j个主题词，r_{ij}表示第i个主题词与第j主题词的关系强度。

[0063] S3、根据步骤S2得到的关系强度矩阵，按照各个主题词之间的关系强度，应用聚类算法对所有主题词进行聚类，根据聚类结果为每个主题词加上类别标签。采用K-Means聚类算法，假定聚类后的类别数为C，即主题词分为C组。之后，通过不同的布局算法将主题词映射到空间平面中的点，整体的布局算法融合示意图如图1所示。

[0064] S4、构建平面像素点类密度函数进行可视化：

[0065] S4.1、假设n个主题词的坐标分别为(x_i,y_i)，i＝1…n，主题词之间的二维欧氏距离平均值为$\overline{Distance}$，$Number_i$，i＝1…n，表示出现了主题词i的科技文本数量，用以揭示技术主题下的文本内容；经过聚类后共有C个类别，每个类别下分别有n_c个主题词；$f(Number_i)$为主题词i的标准化值；像素点P的坐标(x,y)。其中，密度函数α,β为非负数，其取值不同，主题图效果不同。

[0066] 定义像素点的密度函数和类密度函数为：

[0067] 密度函数：$Density(x,y)=\sum_{i=0}^{n} f(Numer_i)e^{-\alpha\left(\frac{\sqrt{(x-x_i)^2+(y-y_i)^2}}{Distance}\right)^{\beta}},\alpha>0,\beta>0$

[0068] 类密度函数：$Density(x,y,c)=\sum_{i=0}^{n_c} f(Numer_i)e^{-\alpha\left(\frac{\sqrt{(x-x_i)^2+(y-y_i)^2}}{Distance}\right)^{\beta}},\alpha>0,\beta>0$

[0069] 其中c表示聚类后某一个具体类别。

[0070] S4.2、融合聚类信息后，用$Density_{max}$表示最大的密度值，$Color_i$表示类别i＝1…C的RGB模式颜色。

[0071] 像素点P(x,y)的RGB模式颜色计算如下：

[0072] $$Color(x,y)=\sum_{i=0}^{n_c} Density(x,y,c_i)/Density_{max}\times Color_i$$

[0073] 其中，$Color_i$是RGB模式颜色的各通道取值。

[0074] S4.3、为了实现类似地形图等高线的可视化效果，同时等高线既能对同一类别下的主题词进行区分，又能对不同类别下的主题词进行区分，构建色彩强度函数：

[0075] f(Density(x,y)/Density(max))

[0076] 色彩强度函数应该是阶梯函数，才能达到等高线的效果，简单的色彩强度函数可以是：

[0077] $$Strength(x,y)=\lfloor(Density(x,y)/Density(\max))\times N\rfloor/N$$

[0078] 其中，$\lfloor\ \rfloor$为向下取整，N为强度级别数量，强度级别数量直接影响绘制效果，使得可视化结果区别于热力图、密度图、一般地形图形式技术主题图。

[0079] S5、构建企业或研发机构i在技术主题j下的竞争力测度模型。构建思路如图2所示。w为参与竞争力测度的企业或研发机构的数量；s为技术领域涵盖的技术主题数量，在技术主题中通过聚类或者社区发现算法确定；$n_{i,j}$为企业或研发机构i在技术主题j的文献数量，$Position(x_k)$为科技文本k的技术地位，$Ability(x_k)$为科技文本k的质量。对应到技术主题图上，$Position(x_k)$取值为步骤S4.4中计算所得的色彩强度，N＝10时，其取值分别范围为{0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9}。$Ability(x_k)$是单个科技文本质量的评价指标，可以用论文的被引用数量、专利的同族数量等衡量单个科技文本质量的数值指标。因此，企业或研发机构i在技术主题j下的竞争力测度模型如下：

[0080] $$Competive(Corpration_i,Techno\log y_j)=\sum_{k=1}^{i=n_{i,j}} Position(D_k)\times Ability(D_k)：$$

$$i=1\cdots w,j=1\cdots s$$

[0081] $Corpration_i$表示企业或研发机构i，$Technology_j$表示技术主题j。

[0082] 将上述公式转化为以下公式：

$$Competive(Corpration_i, Technology_j)$$

[0083]
$$= \sum_{k=1}^{i=n_{i,j}} Position(D_k) \times Ability(D_k)$$
$$= \sum_{k=1}^{i=n_{i,j}} \lfloor (Density(D_k(x,y)/Density(\max)) \times N \rfloor / N \times Ability(D_k)$$

[0084] 其中$D_k(x,y)$为科技文本k在技术主题图中的坐标；

[0085] $Ability(x_k)$采用科技文本被引用数量与引用该科技文本的机构数量比进行表示，即以科技文本被引用数量和被引用机构覆盖情况评价单个科技文本的质量，由此进一步将竞争力测度模型转化如下：

$$Competive(Corpration_i, Techno\log y_j)$$

[0086]
$$= \sum_{k=1}^{i=n_{i,j}} Position(D_k) \times Ability(D_k)$$
$$= \sum_{k=1}^{i=n_{i,j}} \lfloor (Density(D_k(x,y)/Density(max)) \times N \rfloor / N$$
$$\times ReferencedNumber(D_k) / CorprationNumber(D_k)$$

[0087] 其中，$ReferencedNumber(D_k)$为科技文本k的被引用数量，$CorprationNumber(D_k)$为引用了科技文本k的机构数量。

[0088] 对竞争力测度进行性归一化处理使其介于0-1之间，便于比较。计算公式如下：

[0089]
$$\frac{Competive(Corpration_i, Technology_j)}{\underset{i=1\cdots w, j=1\cdots s}{Max}(Competive(Corpration_i, Technology_j))}$$

[0090] S6、形成如下形式的机构、技术主题之间的竞争力矩阵。

[0091]
$$\begin{bmatrix} & Technology_1 & Technology_2 & \cdots & Technology_j & \cdots & Technology_s \\ Corpration_1 & q_{11} & q_{12} & \cdots & q_{1j} & \cdots & q_{1s} \\ Corpration_2 & q_{21} & q_{22} & \cdots & q_{2j} & \cdots & q_{2s} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Corpration_i & q_{i1} & q_{i2} & \cdots & q_{ij} & \cdots & q_{is} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Corpration_w & q_{w1} & q_{w2} & \cdots & q_{wj} & \cdots & q_{ws} \end{bmatrix}$$

[0092] 其中，q_{ij}为企业或研发机构i在技术主题j下技术竞争力测度值。

[0093] 对于本领域的技术人员来说，可以根据以上的技术方案和构思，给出各种相应的改变和变形，而所有的这些改变和变形，都应该包括在本发明权利要求的保护范围之内。

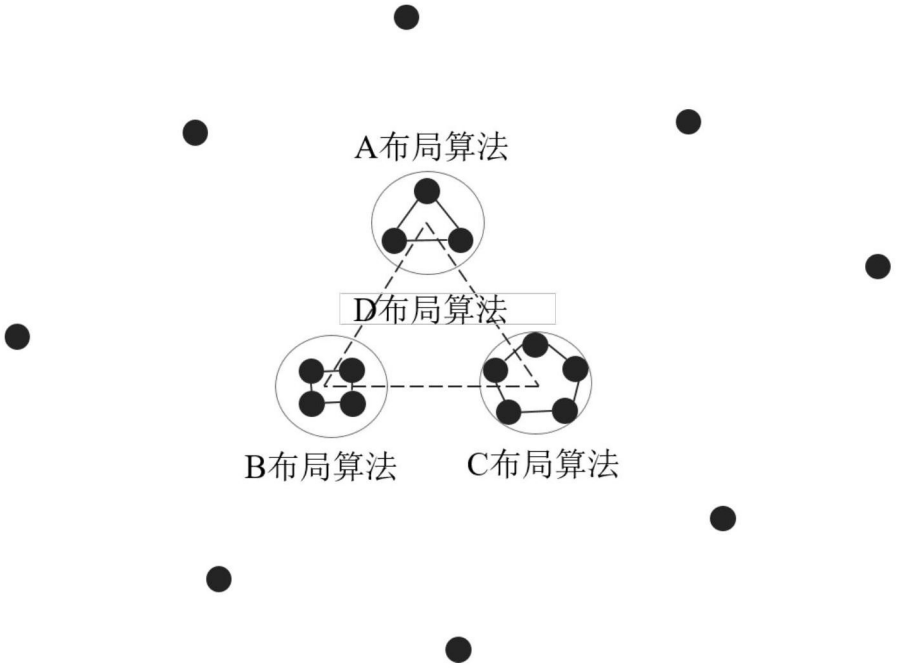

A布局算法
D布局算法
B布局算法
C布局算法

图1

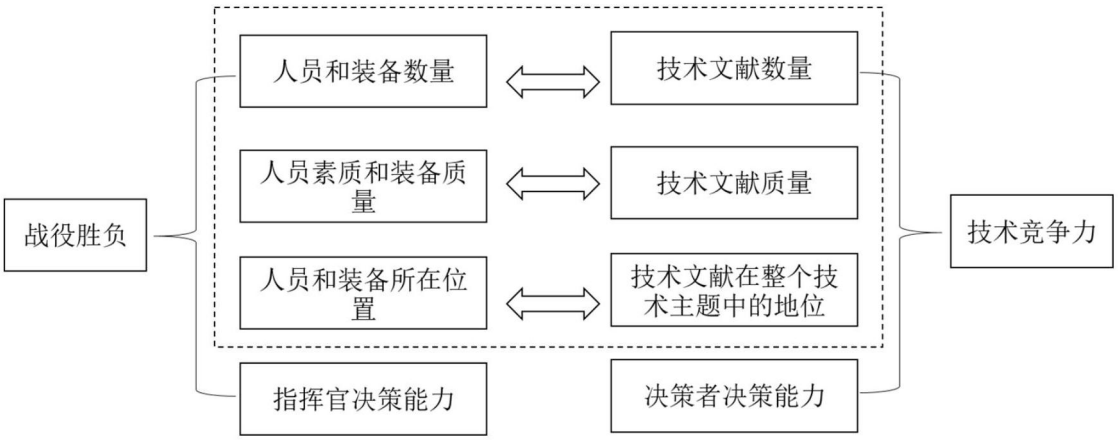

人员和装备数量
技术文献数量
人员素质和装备质量
技术文献质量
战役胜负
人员和装备所在位置
技术文献在整个技术主题中的地位
技术竞争力
指挥官决策能力
决策者决策能力

图2