



(12)发明专利申请

(10)申请公布号 CN 109284420 A

(43)申请公布日 2019.01.29

(21)申请号 201811010713.0

(22)申请日 2018.08.31

(71)申请人 国科赛思(北京)科技有限公司

地址 100085 北京市海淀区安宁庄西路9号
院29号楼5层507室

(72)发明人 李自豪

(74)专利代理机构 北京市商泰律师事务所

11255

代理人 黄晓军

(51)Int.Cl.

G06F 16/901(2019.01)

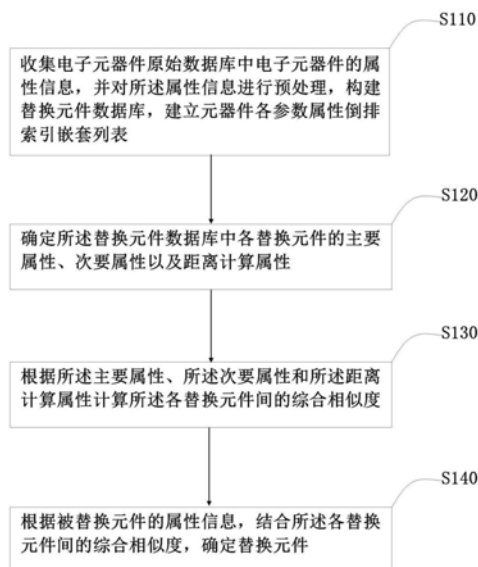
权利要求书3页 说明书14页 附图4页

(54)发明名称

电子元器件替代选型系统及替代选型方法

(57)摘要

本发明公开了一种电子元器件替代选型系统及方法,属于电子元件替代选择方法技术领域,收集电子元器件原始数据库中电子元器件的属性信息,并对属性信息进行预处理,构建替换元件数据库;确定替换元件数据库中各替换元件的主要属性、次要属性以及距离计算属性;根据主要属性、次要属性和距离计算属性计算各替换元件间的综合相似度;根据被替换元件的属性信息,结合各替换元件间的综合相似度,确定替换元件。本发明通过建立元器件综合相似度模型寻找相似元器件,有效减少了计算复杂度,效率更高;同时结合替代元器件实际使用经历和用户的反馈信息,不断迭代更新替代方案,与现有的技术方案相比其替代结果经过实际检验,可信度更高。



1. 一种电子元器件替代选型系统,其特征在于,包括:

数据预处理模块,用于收集电子元器件原始数据库中电子元器件的属性信息,并对所述属性信息进行预处理,构建替换元件数据库,建立元器件各参数属性倒排索引嵌套列表;

属性索引分类模块,用于确定所述替换元件数据库中各替换元件的主要属性、次要属性以及距离计算属性;

综合相似度计算模块,用于根据所述主要属性、所述次要属性和所述距离计算属性计算所述各替换元件间的综合相似度;

替换元件选择模块,用于根据被替换元件的属性信息,结合所述各替换元件间的综合相似度,确定替换元件。

2. 一种电子元器件替代选型方法,其特征在于,包括如下流程步骤:

步骤S110:收集电子元器件原始数据库中电子元器件的属性信息,并对所述属性信息进行预处理,构建替换元件数据库,建立元器件各参数属性倒排索引嵌套列表;

步骤S120:确定所述替换元件数据库中各替换元件的主要属性、次要属性以及距离计算属性;

步骤S130:根据所述主要属性、所述次要属性和所述距离计算属性计算所述各替换元件间的综合相似度;

步骤S140:根据被替换元件的属性信息,结合所述各替换元件间的综合相似度,确定替换元件。

3. 根据权利要求2所述的电子元器件替代选型方法,其特征在于,所述步骤S110具体包括:合并相同属性参数、统一属性参数单位格式、删除错误数据、补全缺失数据。

4. 根据权利要求3所述的电子元器件替代选型方法,其特征在于,所述步骤S120具体包括:

根据相应属性下不同种类数目来判段主要属性和距离计算属性:

$$\eta = \frac{x_i}{N}, \quad \begin{cases} \eta \leq \theta, & \text{主要属性} \\ \eta > \theta, & \text{距离计算属性} \end{cases}, \quad \eta \in (0,1]$$

其中, x_i 表示数值型属性*i*的种类数目, N 为子类元器件数据库中元器件数据总数, η 为属性种类丰富度, θ 为判别阈值。

5. 根据权利要求4所述的电子元器件替代选型方法,其特征在于,所述步骤S130具体包括:

分别计算主要属性相似度、次要属性相似度和距离计算属性相似度,结合所述主要属性相似度和所述距离计算属性相似度确定主要相似度,结合所述主要相似度和所述次要属性相似度确定所述综合相似度。

6. 根据权利要求5所述的电子元器件替代选型方法,其特征在于,所述主要属性相似度的计算方法包括:

建立主要属性广义相似度计算公式:

$$Sm_{i,j} = \frac{\sum_{k=0}^n X_{(k_i,k_j)}}{n} \times 100\%, \quad X_{(k_i,k_j)} = \begin{cases} 1 & k_i = k_j \\ 0 & k_i \neq k_j \end{cases}$$

其中, $S_{m_{i,j}}$ 表示元器件 i 、 j 间主要属性的相似度, k_i 为元器件 i 在主要属性为 k 时的种类, n 为替换元件数据库中所有主要属性的数目, $X(k_i, k_j)$ 为判断元器件 i 、 j 的主要属性 k 异同符号变量。

7. 根据权利要求6所述的电子元器件替代选型方法, 其特征在于, 所述距离计算属性相似度的计算方法包括:

对所述距离计算属性的数据进行Z-Score标准化处理, 假设有 n 个样本有 m 个指标, 则每个变量可表示为 x_{ij} , 经标准化处理后其值为 x_{ij}^* :

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (s_j \neq 0)$$

其中, \bar{x}_j 表示样本均值, s_j 表示样本标准差,

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2};$$

经Z-Score标准化处理后, 替换元件数据库中各个元器件计算属性间的欧氏距离:

$$D_{i,j} = \sqrt{\sum_{k=1}^n (i_k - j_k)^2}, \quad i_k, j_k \neq \text{nan}$$

$$D_{i,j} = 0, i_k = \text{nan}$$

$$D_{i,j} = 0.01, j_k = \text{nan}$$

其中, $D_{i,j}$ 为元器件 i 、 j 距离计算属性间的欧氏距离, i 、 j 分别为被替代元器件和经过筛选后的相似元器件, i_k 、 j_k 分别表示元器件 i 、 j 在第 k 个计算属性的取值, n 为计算属性的总数;

将元器件距离计算属性间的欧氏距离转化为相似度:

$$Sc_{i,j} = \left(1 - \frac{D_{i,j} - \text{Min}(D_i)}{\text{Max}(D_i) - \text{Min}(D_i)} \right) \times 100\%$$

其中, $S_{c_{i,j}}$ 表示元器件 i 、 j 间距离计算属性相似度, $\text{Max}(D_i)$ 表示元器件 i 与相似元器件距离计算属性的最大距离, $\text{Min}(D_i)$ 为元器件 i 与相似元器件距离计算属性的最小距离。

8. 根据权利要求7所述的电子元器件替代选型方法, 其特征在于, 所述次要属性相似度的计算方法具体包括:

建立次要属性广义相似度的计算模型为:

$$Ss_{i,j} = \frac{\sum_{k=0}^n X_{(k_i, k_j)}}{n} \times 100\%, \quad X_{(k_i, k_j)} = \begin{cases} 1 & k_i = k_j \\ 0 & k_i \neq k_j \end{cases}$$

其中, $Ss_{i,j}$ 为元器件 i 、 j 间的次要属性相似度, k_i 为元器件 i 在次要属性 k 中的种类, n 为该子类中元器件所有次要属性的数目, $X_{(k_i, k_j)}$ 为判断元器件 i 、 j 的次要属性 k 异同符号变量。

9. 根据权利要求8所述的电子元器件替代选型方法, 其特征在于, 所述主要相似度计算模型为:

$$Ps_{i,j} = \begin{cases} \omega_1 Sm_{i,j} + \omega_2 Sc_{i,j} \\ Sm_{i,j}, & Sc_{i,j} = 0 \end{cases}$$

$$\text{s.t. } \omega_1 + \omega_2 = 1$$

$$\omega_1 \geq 0$$

$$\omega_2 \geq 0$$

其中, $Ps_{i,j}$ 为元器件 i, j 间的主要相似度, ω_1, ω_2 分别为主要属性相似度和距离计算属性相似度权值。

10. 根据权利要求9所述的电子元器件替代选型方法, 其特征在于, 所述综合相似度计算模型为:

结合主要相似度模型式和次要相似度模型, 构建元器件综合相似度模型:

$$Sc_{i,j} = \begin{cases} \omega_a Ps_{i,j} + \omega_b Ss_{i,j} \\ Ps_{i,j}, & Ss_{i,j} = 0 \end{cases}$$

$$\text{s.t. } \omega_a + \omega_b = 1$$

$$\omega_a \geq \omega_b \geq 0$$

其中, $Sc_{i,j}$ 为元器件 i, j 间的综合相似度, ω_a, ω_b 分别表示元器件主要相似度和次要相似度的权值。

电子元器件替代选型系统及替代选型方法

技术领域

[0001] 本发明涉及电子元器件替代选择方法技术领域,具体涉及一种计算速度快、成本低、效率高的电子元器件替代选型系统及选择方法。

背景技术

[0002] 电子元器件,尤其是集成电路作为工业发展的基础和核心零部件广泛应用于电力、金融、交通、通信、计算机等公共设施以及航空、航天、电子、船舶、兵器、核工业等各种武器装备和军事设备中。

[0003] 我国电子元器件行业已经形成多种所有制并存、门类较为齐全、拥有上万家企业的庞大产业。然而对于关键、核心元器件如集成电路、光电子器件、电真空器件仍大量依赖进口。

[0004] 大量依赖进口元器件主要存在以下风险:停产断档与出口限制导致货源保障困难,假冒伪劣和筛选试验困难导致进口原器件存在质量风险,进口器件缺少相应的技术手册等相关资料,其性能参数、可靠性水平等往往无法预先开展检测和筛选工作,这将导致在设备使用过程中可能存在参数异常、早期失效等情况,进口元器件植入后门存在安全隐患,此外,我国进口的元器件大多为塑封工业级和商业级,应用于航空航天和武器装备等高可靠领域,其可靠性和环境适应性难以满足要求。

[0005] 因此,元器件的国产化替代显得尤为重要。近年来,我国已在元器件国产化替代方面加大投入并取得了很大进展,但由于各单位在元器件国产化替代方面执行力度不一,元器件国产化替代仍然存在一些问题。缺乏顶层指导文件和对应的国家标准,元器件参数体系不完整、命名规则不统一,国内外元器件质量等级、分类体系不同,元器件参数名称、格式不统一,元器件的替代缺少实际的应用检验,同时,部分装备研制单位虽然自行制定了内部的元器件国产化替代规则,但由于设计、加工等因素,国产化替代验证只能结合实际的模块、整机统筹开展,难以全面验证元器件的性能。

[0006] 目前国际上公认的几种替代类型主要包括:直接替代、基本替代、功能替代和相似替代。上述这些替代方式均需要从其性能参数、封装尺寸及外形、基本功能和可靠性与环境适应性这几个方面综合考虑。因此,现有的技术方案一般是比较元器件的固有参数指标,同时结合工程实际以及元器件应用经历判断替代是否可行。其中对于元器件性能参数的比较一般是通过计算相似度量化反映。

[0007] 元器件,尤其是以集成电路为代表的各类器件,其固有的参数指标种类繁多,指标类型主要包括确定数字类型、确定符号类型和确定区间类型。因此其属性参数为混合型数据类型。确定数字类型:该类型元器件参数值可以是连续或离散的数值,如电阻 30Ω ,电压 $40V$ 等。确定符号类型:该类型元器件参数值通常是以明确的专业术语枚举表示,如质量等级:军品级、工业级等;封装类型:引线、直插等。确定区间类型:该类型元器件参数值一般为某一数值区间,如温度范围: $-65^{\circ}\text{C}\sim 175^{\circ}\text{C}$, $-40^{\circ}\text{C}\sim 85^{\circ}\text{C}$ 等。

[0008] 数值型数据其相似度的计算方法一般为欧式距离、余弦相似度或皮尔森相关系数

等,而符号型数据其相似度的计算方法一般为杰卡德系数和简单匹配系数等。而对于混合型数据类型相似度的计算现有的技术方案主要是分别计算不同类型数据间的相似度,然后进行加权融合。如基于简单匹配方法。该方法是混合数据相似度计算的一种简单、高效的方式,对于同一属性参数,属性值相同数据距离标记为0,不同标记为1。D.Randall和Tony R提出的基于概率的相似度度量方法。该方法通过定义HVD距离,计算属性值在一个类中出现的概率或条件概率来度量相似度。基于粗糙集方法,如滕书华、鲁敏等提出的基于一般二元关系的粗糙集加权不确定性度量。该方法利用粗糙集的相关理论,构造距离度量函数。Yogalaskshmi J、Chandrashekar R提出的混合数据的聚类算法,该方法对数值属性,利用欧式距离进行度量,对于符号属性则采用一种粗糙集方法的度量函数。基于信息熵方法,如Li X和YeN在混合数据相似性度量函数中引入信息熵的思想提高处理不确定信息的能力。基于概念层次树方法。该方法是层次距离的延伸,将两个相连结点之间赋予权值,以表示两个节点距离。

[0009] 上述方法对混合数据的相似度计算时均将进行距离加权融合的操作,而部分方法对权值的选择有较大的主观性,其说服力和可信度较差。此外,在实际的元器件替代选型任务中,元器件种类繁多且数量庞大,若通过上述方式在全局解空间中计算搜索相似元器件,其时间成本和空间成本均较高。

发明内容

[0010] 本发明的目的在于提供一种计算快、效率高、可信度高的电子元器件替代选型系统及选型方法,以解决上述背景技术中存在的技术问题。

[0011] 为了实现上述目的,本发明采取了如下技术方案:

[0012] 一方面,本发明提供一种电子元器件替代选型系统,该系统包括:

[0013] 数据预处理模块,用于收集电子元器件原始数据库中电子元器件的属性信息,并对所述属性信息进行预处理,构建替换元件数据库,建立元器件各参数属性倒排索引嵌套列表;

[0014] 属性索引分类模块,用于确定所述替换元件数据库中各替换元件的主要属性、次要属性以及距离计算属性;

[0015] 综合相似度计算模块,用于根据所述主要属性、所述次要属性和所述距离计算属性计算所述各替换元件间的综合相似度;

[0016] 替换元件选择模块,用于根据被替换元件的属性信息,结合所述各替换元件间的综合相似度,确定替换元件。

[0017] 另一方面,本发明提供一种电子元器件替代选型方法,该方法包括如下流程步骤:

[0018] 步骤S110:收集电子元器件原始数据库中电子元器件的属性信息,并对所述属性信息进行预处理,构建替换元件数据库,建立元器件各参数属性倒排索引嵌套列表;

[0019] 步骤S120:确定所述替换元件数据库中各替换元件的主要属性、次要属性以及距离计算属性;

[0020] 步骤S130:根据所述主要属性、所述次要属性和所述距离计算属性计算所述各替换元件间的综合相似度;

[0021] 步骤S140:根据被替换元件的属性信息,结合所述各替换元件间的综合相似度,确

定替换元件。

[0022] 进一步的,所述步骤S110具体包括:合并相同属性参数、统一属性参数单位格式、删除错误数据、补全缺失数据。

[0023] 进一步的,所述步骤S120具体包括:

[0024] 根据相应属性下不同种类数目来判段主要属性和距离计算属性:

$$[0025] \quad \eta = \frac{x_i}{N}, \quad \begin{cases} \eta \leq \theta, & \text{主要属性} \\ \eta > \theta, & \text{距离计算属性} \end{cases}, \eta \in (0, 1]$$

[0026] 其中, x_i 表示数值型属性i的种类数目,N为子类元器件数据库中元器件数据总数, η 为属性种类丰富度, θ 为判别阈值。

[0027] 进一步的,所述步骤S130具体包括:

[0028] 分别计算主要属性相似度、次要属性相似度和距离计算属性相似度,结合所述主要属性相似度和所述距离计算属性相似度确定主要相似度,结合所述主要相似度和所述次要属性相似度确定所述综合相似度。

[0029] 进一步的,所述主要属性相似度的计算方法包括:

[0030] 建立主要属性广义相似度计算公式:

$$[0031] \quad Sm_{i,j} = \frac{\sum_{k=0}^n X_{(k_i,k_j)}}{n} \times 100\%, \quad X_{(k_i,k_j)} = \begin{cases} 1 & k_i = k_j \\ 0 & k_i \neq k_j \end{cases}$$

[0032] 其中, $S_{m_{i,j}}$ 表示元器件i、j间主要属性的相似度, k_i 为元器件i在主要属性为k时的种类,n为替换元件数据库中所有主要属性的数目, $X(k_i,k_j)$ 为判断元器件i、j的主要属性k异同符号变量。

[0033] 进一步的,所述距离计算属性相似度的计算方法包括:

[0034] 对所述距离计算属性的数据进行Z-Score标准化处理,假设有n个样本有m个指标,则每个变量可表示为 x_{ij} ,经标准化处理后其值为 x_{ij}^* ;

$$[0035] \quad x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (s_j \neq 0)$$

[0036] 其中, \bar{x}_j 表示样本均值, s_j 表示样本标准差,

$$[0037] \quad \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2};$$

[0038] 经Z-Score标准化处理后,替换元件数据库中各个元器件计算属性间的欧氏距离:

$$[0039] \quad D_{i,j} = \sqrt{\sum_{k=1}^n (i_k - j_k)^2}, \quad i_k, j_k \neq \text{nan}$$

[0040] $D_{i,j} = 0, \quad i_k = \text{nan}$

[0041] $D_{i,j} = 0.01, \quad j_k = \text{nan}$

[0042] 其中, $D_{i,j}$ 为元器件i、j距离计算属性间的欧氏距离,i、j分别为被替代元器件和经

过筛选后的相似元器件, i_k 、 j_k 分别表示元器件 i 、 j 在第 k 个计算属性的取值, n 为计算属性的总数;

[0043] 将元器件距离计算属性间的欧氏距离转化为相似度:

$$[0044] \quad Sc_{i,j} = \left(1 - \frac{D_{i,j} - \text{Min}(D_i)}{\text{Max}(D_i) - \text{Min}(D_i)} \right) \times 100\%$$

[0045] 其中, $Sc_{i,j}$ 表示元器件 i 、 j 间距离计算属性相似度, $\text{Max}(D_i)$ 表示元器件 i 与相似元器件距离计算属性的最大距离, $\text{Min}(D_i)$ 为元器件 i 与相似元器件距离计算属性的最小距离。

[0046] 进一步的,所述次要属性相似度的计算方法具体包括:

[0047] 建立次要属性广义相似度的计算模型为:

$$[0048] \quad Ss_{i,j} = \frac{\sum_{k=0}^n X_{(k_i, k_j)}}{n} \times 100\%, \quad X_{(k_i, k_j)} = \begin{cases} 1 & k_i = k_j \\ 0 & k_i \neq k_j \end{cases}$$

[0049] 其中, $Ss_{i,j}$ 为元器件 i 、 j 间的次要属性相似度, k_i 为元器件 i 在次要属性 k 中的种类, n 为该子类中元器件所有次要属性的数目, $X_{(k_i, k_j)}$ 为判断元器件 i 、 j 的次要属性 k 异同符号变量。

[0050] 进一步的,所述主要相似度计算模型为:

$$[0051] \quad Ps_{i,j} = \begin{cases} \omega_1 Sm_{i,j} + \omega_2 Sc_{i,j} \\ Sm_{i,j}, & Sc_{i,j} = 0 \end{cases}$$

$$[0052] \quad \text{s.t. } \omega_1 + \omega_2 = 1$$

$$[0053] \quad \omega_1 \geq 0$$

$$[0054] \quad \omega_2 \geq 0$$

[0055] 其中, $Ps_{i,j}$ 为元器件 i 、 j 间的主要相似度, ω_1 、 ω_2 分别为主要属性相似度和距离计算属性相似度权值。

[0056] 进一步的,所述综合相似度计算模型为:

[0057] 结合主要相似度模型式和次要相似度模型,构建元器件综合相似度模型:

$$[0058] \quad Sc_{i,j} = \begin{cases} \omega_a Ps_{i,j} + \omega_b Ss_{i,j} \\ Ps_{i,j}, & Ss_{i,j} = 0 \end{cases}$$

$$[0059] \quad \text{s.t. } \omega_a + \omega_b = 1$$

$$[0060] \quad \omega_a \geq 0, \omega_b \geq 0$$

[0061] 其中, $Sc_{i,j}$ 为元器件 i 、 j 间的综合相似度, ω_a 、 ω_b 分别表示元器件主要相似度和次要相似度的权值。

[0062] 本发明有益效果:通过建立元器件综合相似度模型寻找相似元器件,该方法与现有的技术方法相比有效的减少了计算的时间复杂度和空间复杂度,效率更高;同时结合替代元器件实际使用经历和用户的反馈信息,不断迭代更新替代方案,与现有的技术方案相

比其替代结果经过实际检验,可信度更高,具有广阔的应用前景。

[0063] 本发明附加的方面和优点将在下面的描述中部分给出,这些将从下面的描述中变得明显,或通过本发明的实践了解到。

附图说明

[0064] 为了更清楚地说明本发明实施例的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0065] 图1为本发明实施例一所述的电子元器件替代选型系统原理框图。

[0066] 图2为本发明实施例一所述的电子元器件替代选型方法流程图。

[0067] 图3为本发明实施例二所述的电子元器件替代选型系统原理框图。

[0068] 图4为本发明实施例二所述的主要属性数目与距离计算属性数目的比值和主要属性的相似度权值的拟合函数示意图。

具体实施方式

[0069] 下面详细叙述本发明的实施方式,所述实施方式的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过附图描述的实施方式是示例性的,仅用于解释本发明,而不能解释为对本发明的限制。

[0070] 本技术领域技术人员可以理解,除非特意声明,这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是,本发明的说明书中使用的措辞“包括”是指存在所述特征、整数、步骤、操作、元件和/或组件,但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件和/或它们的组。应该理解,这里使用的“连接”或“耦接”可以包括无线连接或耦接,使用的措辞“和/或”包括一个或多个相关联的列出项的任一单元和全部组合。

[0071] 本技术领域技术人员可以理解,除非另外定义,这里使用的所有术语(包括技术术语和科学术语)具有与本发明所属领域中的普通技术人员的一般理解相同的意义。还应该理解的是,诸如通用字典中定义的那些术语应该被理解为具有与现有技术的上下文中的意义一致的意义,并且除非像这里一样定义,不会用理想化或过于正式的含义来解释。

[0072] 为便于理解本发明,下面结合附图以具体实施例对本发明作进一步解释说明,且具体实施例并不构成对本发明实施例的限定。

[0073] 本领域技术人员应该理解,附图只是实施例的示意图,附图中的部件并不一定是实施本发明所必须的。

[0074] 实施例一

[0075] 如图1所示,本发明实施例一提供一种电子元器件替代选型系统,该系统包括有如下功能模块:

[0076] 数据预处理模块,用于收集电子元器件原始数据库中电子元器件的属性信息,并对所述属性信息进行预处理,构建替换元件数据库,建立元器件各参数属性倒排索引嵌套列表;

[0077] 属性索引分类模块,用于确定所述替换元件数据库中各替换元件的主要属性、次要属性以及距离计算属性;

[0078] 综合相似度计算模块,用于根据所述主要属性、所述次要属性和所述距离计算属性计算所述各替换元件间的综合相似度;

[0079] 替换元件选择模块,用于根据被替换元件的属性信息,结合所述各替换元件间的综合相似度,确定替换元件。

[0080] 如图2所示,本发明实施例一提供的一种利用上述电子元器件替代选型系统进行电子元器件替代选型的方法,该方法主要包括如下流程步骤:

[0081] 步骤S110:收集电子元器件原始数据库中电子元器件的属性信息,并对所述属性信息进行预处理,构建替换元件数据库,建立元器件各参数属性倒排索引嵌套列表;

[0082] 步骤S120:确定所述替换元件数据库中各替换元件的主要属性、次要属性以及距离计算属性;

[0083] 步骤S130:根据所述主要属性、所述次要属性和所述距离计算属性计算所述各替换元件间的综合相似度;

[0084] 步骤S140:根据被替换元件的属性信息,结合所述各替换元件间的综合相似度,确定替换元件。

[0085] 在本发明实施例一所述的方法中,所述步骤S110具体包括:合并相同属性参数、统一属性参数单位格式、删除错误数据、补全缺失数据。

[0086] 在本发明实施例一所述的方法中,所述步骤S120具体包括:

[0087] 根据相应属性下不同种类数目来判段主要属性和距离计算属性:

$$[0088] \quad \eta = \frac{x_i}{N}, \quad \begin{cases} \eta \leq \theta, & \text{主要属性} \\ \eta > \theta, & \text{距离计算属性} \end{cases}, \eta \in (0, 1]$$

[0089] 其中, x_i 表示数值型属性i的种类数目,N为子类元器件数据库中元器件数据总数, η 为属性种类丰富度, θ 为判别阈值。

[0090] 在本发明实施例一所述的方法中,所述步骤S130具体包括:

[0091] 分别计算主要属性相似度、次要属性相似度和距离计算属性相似度,结合所述主要属性相似度和所述距离计算属性相似度确定主要相似度,结合所述主要相似度和所述次要属性相似度确定所述综合相似度。

[0092] 在本发明实施例一所述的方法中,所述主要属性相似度的计算方法包括:

[0093] 建立主要属性广义相似度计算公式:

$$[0094] \quad Sm_{i,j} = \frac{\sum_{k=0}^n X_{(k_i,k_j)}}{n} \times 100\%, \quad X_{(k_i,k_j)} = \begin{cases} 1 & k_i = k_j \\ 0, & k_i \neq k_j \end{cases}$$

[0095] 其中, $Sm_{i,j}$ 表示元器件i、j间主要属性的相似度, k_i 为元器件i在主要属性为k时的种类,n为替换元件数据库中所有主要属性的数目, $X(k_i,k_j)$ 为判断元器件i、j的主要属性k异同符号变量。

[0096] 在本发明实施例一所述的方法中,所述距离计算属性相似度的计算方法包括:

[0097] 对所述距离计算属性的数据进行Z-Score标准化处理,假设有n个样本有m个指标,则每个变量可表示为 x_{ij} ,经标准化处理后其值为 x_{ij}^* ;

$$[0098] \quad x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (s_j \neq 0)$$

[0099] 其中, \bar{x}_j 表示样本均值, s_j 表示样本标准差,

$$[0100] \quad \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2};$$

[0101] 经Z-Score标准化处理后,替换元件数据库中各个元器件计算属性间的欧氏距离:

$$[0102] \quad D_{i,j} = \sqrt{\sum_{k=1}^n (i_k - j_k)^2}, \quad i_k, j_k \neq \text{nan}$$

$$[0103] \quad D_{i,j} = 0, \quad i_k = \text{nan}$$

$$[0104] \quad D_{i,j} = 0.01, \quad j_k = \text{nan}$$

[0105] 其中, $D_{i,j}$ 为元器件i、j距离计算属性间的欧氏距离,i、j分别为被替代元器件和经过筛选后的相似元器件, i_k 、 j_k 分别表示元器件i、j在第k个计算属性的取值,n为计算属性的总数;

[0106] 将元器件距离计算属性间的欧氏距离转化为相似度:

$$[0107] \quad Sc_{i,j} = \left(1 - \frac{D_{i,j} - \text{Min}(D_i)}{\text{Max}(D_i) - \text{Min}(D_i)} \right) \times 100\%$$

[0108] 其中, $Sc_{i,j}$ 表示元器件i、j间距离计算属性相似度, $\text{Max}(D_i)$ 表示元器件i与相似元器件距离计算属性的最大距离, $\text{Min}(D_i)$ 为元器件i与相似元器件距离计算属性的最小距离。

[0109] 在本发明实施例一所述的方法中,所述次要属性相似度的计算方法具体包括:

[0110] 建立次要属性广义相似度的计算模型为:

$$[0111] \quad Ss_{i,j} = \frac{\sum_{k=0}^n X_{(k_i, k_j)}}{n} \times 100\%, \quad X_{(k_i, k_j)} = \begin{cases} 1 & k_i = k_j \\ 0, & k_i \neq k_j \end{cases}$$

[0112] 其中, $Ss_{i,j}$ 为元器件i、j间的次要属性相似度, k_i 为元器件i在次要属性k中的种类,n为该子类中元器件所有次要属性的数目, $X_{(k_i, k_j)}$ 为判断元器件i、j的次要属性k异同符号变量。

[0113] 在本发明实施例一所述的方法中,所述主要相似度计算模型为:

$$[0114] \quad Ps_{i,j} = \begin{cases} \omega_1 Sm_{i,j} + \omega_2 Sc_{i,j} \\ Sm_{i,j}, & Sc_{i,j} = 0 \end{cases}$$

$$[0115] \quad \text{s.t. } \omega_1 + \omega_2 = 1$$

[0116] $\omega_1 \geq 0$

[0117] $\omega_2 \geq 0$

[0118] 其中, $Ps_{i,j}$ 为元器件 i, j 间的主要相似度, ω_1, ω_2 分别为主要属性相似度和距离计算属性相似度权值。

[0119] 在本发明实施例一所述的方法中, 所述综合相似度计算模型为:

[0120] 结合主要相似度模型式和次要相似度模型, 构建元器件综合相似度模型:

$$[0121] \quad Sc_{i,j} = \begin{cases} \omega_a Ps_{i,j} + \omega_b Ss_{i,j} \\ Ps_{i,j}, & Ss_{i,j} = 0 \end{cases}$$

[0122] s. t. $\omega_a + \omega_b = 1$

[0123] $\omega_a \geq 0, \omega_b \geq 0$

[0124] 其中, $Sc_{i,j}$ 为元器件 i, j 间的综合相似度, ω_a, ω_b 分别表示元器件主要相似度和次要相似度的权值。

[0125] 实施例二

[0126] 如图3所示, 本发明实施例提供了一种电子元器件替代选型系统及方法, 首先利用数据预处理模块对元器件原始数据进行预处理, 合并相同属性参数、统一属性参数单位格式、删除错误数据、补全缺失数据。然后参考相关标准手册并结合技术人员的知识经验, 确定被替代元器件的主要属性, 次要属性和距离计算属性。(例如对于片式膜电阻器, 其主要属性为功率、温度特性、精度、功率等, 次要属性为质量等级、封装类型、介质类型等, 距离计算属性为阻值。) 其中, 主要属性和次要属性是指其该属性包括的种类一般为枚举型数据或者种类数目较少的属性(对于片式膜电阻器的质量等级这次要属性, 其种类有“G”, “J”, “国军标”等等, 为枚举型数据, 该属性的种类数目仅为片式膜电阻器这一子类元器件数据数目的0.002%左右。) 其次, 根据替代原器件与被替代元器件主要属性中种类相同的数目计算广义相似度, 筛选出相似度较高的部分替代元器件(一般为100组左右)。同样, 对次要属性依据属性中种类相同的数目计算广义相似度。接下来, 对距离计算属性(阻值)中的数据进行Z-Score标准化处理, 并对筛选出的替代元器件距离计算属性计算欧氏距离, 并将欧氏距离转换为相似度。将替代元器件所对应的各个相似度加权融合得到综合相似度, 对综合相似度从大到小排序。最后, 结合相关的元器件替代型号资料、元器件的实际使用经历数据和用户的反馈信息为用户推荐元器件替代型号, 并不断根据用户反馈迭代、更新替代方案。

[0127] 具体的, 本发明实施例二所述的电子元器件替代选型系统具体功能如下:

[0128] 元器件数据清洗模块(数据预处理模块)实现功能主要包括: ①元器件属性参数的拆分。如稳压二极管中某一属性参数名为“CURRENT-REVERSELEAKAGE@VR”, 其值为1.1 μ A@3.5V, 如要获得该属性中的反向漏电流我们则需依据分隔符“@”拆分并提取数据。②元器件属性合并功能。如电阻中“电压”和“Voltage”这两属性参数实际为同一属性, 对此需要将这两属性数据进行合并、统一。③错误数据删除功能。如电阻元器件其关键属性电阻数据缺失, 或者对于电压属性某一数据记录为45KMV, 数据单位错误, 则认为其均为错误数据需要进行删除。④数据单位统一功能。如电阻属性单位包括 Ω 、K Ω 、M Ω 等, 电压单位包括V、KV、mV等, 需要将其单位统一。

[0129] 元器件数据清洗模块由Python语言编写,其运行环境为Python 3.6.4.该模块所依赖的外部函数库包括:numpy 1.14.0、pandas 0.22.0、csv。该模块从数据库中读取元器件原始数据并对数据进行清洗,最后将处理所得数据保存至数据库中。

[0130] 元器件相似度计算模块的主要功能:为衡量不同元器件间相似程度这一模糊指标,以寻找相似或相同元器件,通过建立数学模型,计算量化各个元器件间的相似程度,实现不同元器件的相互替代。具体建模过程如下:

[0131] 1.1主要属性相似度的计算

[0132] 主要属性是指在元器件属性参数中十分重要的那些属性。主要属性应需满足该属性下所包含的数据类型为枚举型字符串,或者其数据类型为数值型,但种类数目远小于数据库中这一子类元器件数目的要求。主要属性一般在该子类元器件中被大部分元器件所共有,通过确定主要属性可以将元器件数据划分为较多的小块,以较大的区分不同元器件。在实际计算中,本专利通过主要属性对相似元器件进行初步筛选,经过筛选后的相似元器件搜索范围能大大缩小至100组左右,以减少计算时间和存储空间,避免计算和存储资源不必要的浪费。

[0133] 1.1.1主要属性的确定

[0134] 举例来说,对于瓷介电容器这一元器件子类,参考相关的手册、标准,该元器件的关键属性包括容值、额定电压、温度特性、精度、介质类型等。对于介质类型这一属性,其种类主要包括1类瓷、2类瓷、2R1等等,为枚举数据类型。对于额定电压这一属性,其种类主要包括 $\pm 10\%$ 、 $\pm 5\%$ 、 $\pm 20\%$ 、 $\pm 2\%$ 等等,虽然该属性为数值数据类型,但是其种类数目为远小于该元器件子类数据总数,故将其视为主要属性而非距离计算属性。对于容值这一属性,其数据类型为数值型,且种类数目较多,故将其视为距离计算属性。对于同为数值型的关键属性。

[0135] 本发明的具体实施例二中,主要依据该属性下不同种类数目来判段主要属性和距离计算属性,判别规则如下式(2-1)所示。

$$[0136] \quad \eta = \frac{x_i}{N}, \quad \begin{cases} \eta \leq \theta, & \text{主要属性} \\ \eta > \theta, & \text{距离计算属性} \end{cases}, \quad \eta \in (0,1] \quad (2-1)$$

[0137] 上式中, x_i 为数值型属性i的种类数目; N 为子类元器件数据库中元器件数据总数; η 为属性种类丰富度; θ 为判别阈值。

[0138] η 为属性种类丰富度,反映该属性下所包含种类的丰富程度,该值越大表示该属性包含的不同种类数目越多,当 $\eta=1$ 时,表示该属性下每一元器件的种类均不相同,均为独自的一类。 θ 为判别阈值,若 $\eta \leq \theta$ 则表示该属性为主要属性,否则为距离计算属性。其中, θ 值的确定需考虑相似度距离计算的时间复杂度、空间复杂度、精度和搜索出的相似元器件数目, θ 值一般取0.001,同时根据不同情况下,可对 θ 值做些许调整。

[0139] 在实际应用中,如果 θ 值设置较小则大部分属性将被判别为距离计算属性,此时需通过计算每一元器件间的欧氏距离寻找相似元器件,其空间和时间成本虽然较高,但是该方法计算精度较高且能搜索出较多不同类型的相似元器件。如果 θ 值设定较大,则大部分属性将被判别为主要属性,主要属性的相似度计算较为简单,时间成本大大降低,但是其精度较低,搜索出的元器件种类数目较少。因此,合理设置 θ 值较为重要。

[0140] 1.1.2主要属性相似度的计算

[0141] 主要属性均为枚举型或种类数目较少的那些属性,对于主要属性相似度的计算,为降低计算成本,本专利并不采用现有的距离和相似度计算公式,而是建立广义相似度计算公式,如式(2-2)所示。

$$[0142] \quad Sm_{i,j} = \frac{\sum_{k=0}^n X_{(k_i,k_j)}}{n} \times 100\%, \quad X_{(k_i,k_j)} = \begin{cases} 1 & k_i = k_j \\ 0 & k_i \neq k_j \end{cases} \quad (2-2)$$

[0143] 上式中, $Sm_{i,j}$ 为元器件*i,j*间主要属性的相似度; k_i 为元器件*i*在主要属性为*k*时的种类; n 为该元器件数据库中所有主要属性的数目; $X_{(k_i,k_j)}$ 为判断元器件*i,j*的主要属性*k*异同符号变量。

[0144] 如果元器件*i,j*的主要属性中的类别取值均相同,则两元器件的相似度 $Sm_{i,j}$ 为1,如果元器件*i,j*的主要属性的类别取值均不相同,则两元器件的相似度 $Sm_{i,j}$ 为0。显然,对于两元器件,如果每一主要属性下种类相同的数目越多,则相似度的值越大。

[0145] 在本发明的具体实施例二中,在编程过程中为减少时间复杂度,通过建立倒排索引,将该子类元器件下各个属性中不同种类的元器件索引保存为嵌套列表,在搜索元器件*i*的相似元器件时只需在列表中查找出每个属性中相同种类的元器件索引,再对每一属性的索引集合求并集,并依据相同索引出现次数从大到小排序,计算主要参数相似度。通过主要属性和式(2-3)对元器件数目进行筛选,可以缩小搜索范围。

[0146] do $Sn = \sum \lambda_i (i = n, n-1, n-2 \dots)$, until $Sn \geq 100$

$$[0147] \quad Sn = \begin{cases} \text{Min}(Sn, 200), & i = n \\ Sn, & i < n \end{cases} \quad (2-3)$$

[0148] 上式中, Sn 为经主要属性筛选出的元器件数目; λ_i 为各主要属性中类别相同数目为*i*的元器件数目; n 为主要属性的个数。

[0149] 1.2距离计算属性相似度的计算

[0150] 距离计算属性一般是指数据类型为数值型且在式(2-1)中满足 $\eta > 0$ 条件的关键属性。此时,距离计算属性对应的数据已经经过单位换算等处理后的数据。同时在计算各元器件间的欧氏距离时为避免不同量纲的影响,本专利对距离计算属性的数据进行Z-Score标准化处理如式(2-4)。

[0151] 假设有*n*个样本有*m*个指标,则每个变量可表示为 x_{ij} ,经标准化处理后其值为 x_{ij}^* 。 x_{ij}^* 的计算公式如下。

$$[0152] \quad x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (s_j \neq 0) \quad (2-4)$$

[0153] 上式中, \bar{x}_j 为样本均值, $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$;

[0154] s_j 为样本标准差, $s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_{ij})^2}$ 。

[0155] 经Z-Score标准化处理后,各个元器件计算属性间的欧氏距离计算公式如(2-5)所示。

[0156] $D_{i,j} = \sqrt{\sum_{k=1}^n (i_k - j_k)^2}$, $i_k, j_k \neq \text{nan}$

[0157] $D_{i,j} = 0$, $i_k = \text{nan}$

[0158] $D_{i,j} = 0.01$, $j_k = \text{nan}$ (2-5)

[0159] 上式中, $D_{i,j}$ 为元器件*i*,*j*距离计算属性间的欧氏距离;*i*,*j*分别为被替代元器件和经过筛选后的相似元器件; i_k, j_k 为元器件*i*,*j*在第*k*个计算属性的取值;*n*为计算属性的总数。

[0160] 显然,元器件*i*,*j*计算属性间的距离 $D_{i,j}$ 越大表示两者差别越大,元器件*i*,*j*越不相似。如果被替代元器件*i*的属性种类数据*k*为空(nan),则定义元器件间的相似度为0,如果经过筛选后的相似元器件*j*的属性种类值*k*为空(nan),则定义元器件间的距离计算属性相似度为一个较小值0.01,用以区分当被替代元器件与相似元器件拥有相同属性种类值时其距离计算属性相似度为0的情形。为方便计算各元器件间的相似度,我们按式(2-6)将元器件计算属性间的欧氏距离转化为相似度。

[0161] $Sc_{i,j} = \left(1 - \frac{D_{i,j} - \text{Min}(D_i)}{\text{Max}(D_i) - \text{Min}(D_i)} \right) \times 100\%$ (2-6)

[0162] 上式中, $Sc_{i,j}$ 为元器件*i*,*j*间距离计算属性相似度; $\text{Max}(D_i)$ 为元器件*i*与相似元器件距离计算属性的最大距离; $\text{Min}(D_i)$ 为元器件*i*与相似元器件距离计算属性的最小距离。

[0163] 如果该子类元器件没有距离计算属性,则 $Sc_{i,j} = 0$ 。此时,只对经过主要属性筛选后的元器件计算属性间的相似度,以减少计算量,缩短计算时间。

[0164] 1.3主要相似度模型

[0165] 根据关键参数即主要属性和距离计算属性构建主要相似度模型。如式(2-7)所示。

[0166] $Ps_{i,j} = \begin{cases} \omega_1 Sm_{i,j} + \omega_2 Sc_{i,j} \\ Sm_{i,j}, & Sc_{i,j} = 0 \end{cases}$

[0167] s.t. $\omega_1 + \omega_2 = 1$

[0168] $\omega_1 \geq 0$

[0169] $\omega_2 \geq 0$ (2-7)

[0170] 上式中, $Ps_{i,j}$ 为元器件*i*,*j*间的主要相似度; ω_1, ω_2 分别为主要属性相似度和距离计算属性相似度权值。

[0171] 在本发明的具体实施例二中,在相似元器件的搜索过程中采取的策略主要是,首先保证两元器件间主要属性的相同数目最多,在主要属性相同数目一致的情况下再计算距离计算属性间的相似度,该相似度越大则两元器件越相似。

[0172] 主要属性相似度的计算,依赖元器件属性种类相同的数目,该相似度为离散值,且

计算出的相似度在数值上有明显区别(例如,两元器件主要属性的相同数目分别为5和4时计算所得的相似度有较大区别),因此主要属性相似度这一指标对两元器件间的相似程度有较大的区分,对元器件是否相似的判别能力更强。

[0173] 随着主要属性数目的增加,在主要相似度计算时赋予该项的权值将不断增大。而距离计算属性相似度为连续值,两元器件在此属性值相差较大的情况下仍有可能计算出较大的相似度。因此,该相似度对元器件相似程度的判别能力较弱,且随着主要属性数目的增加,赋予该项的权值将不断减小。

[0174] 综上所述,在权值的设计时,应更偏向于对相似元器件判别能力较强的主要属性。当主要属性数目和距离计算属性数目的比值变化时,两不同类型的相似度权值的取值也应随之变化。记主要属性数目与距离计算属性数目的比值为 ε 。当 ε 取不同值时,本专利设计主要属性的相似度权值 ω_1 的取值情况如表1所示。

[0175] 表1

[0176]

ε	50	25	6	4	3	2	1	1/2	1/3	1/4	1/8	1/10
ω_1	0.96	0.92	0.9	0.85	0.8	0.7	0.65	0.6	0.55	0.5	0.22	0.2

[0177] 由上表中 ε 与 ω_1 间一一对应的取值关系,分别选择指数函数、幂函数、构造函数、四次多项式函数和有理数逼近函数,利用最小二乘法拟合 ε 与 ω_1 间的函数关系,其拟合结果如表2所示。

[0178] 表2

[0179]

拟合函数	误差平方和 (SSE)	复相关系数 (RSquare)	调整自由度 (Ad-justed R-Square)	均方根误差 (RMSE)
$\omega_1 = \frac{-0.569e^{-0.293\varepsilon} + 1.317}{10.8e^{14.34\varepsilon} + 1.364}$	0.003	0.997	0.992	0.023
$\omega_1 = 0.776e^{0.004\varepsilon} - 0.804e^{-63.477\varepsilon}$	0.039	0.939	0.913	0.075
$\omega_1 = -0.334\varepsilon^{-0.396} + 1.025$	0.015	0.976	0.97	0.044
$\omega_1 = -0.013\varepsilon^4 + 0.163\varepsilon^3 - 0.666\varepsilon^2 + 1.067\varepsilon + 0.174$	0.044	0.916	0.848	0.094
$\omega_1 = \frac{0.902\varepsilon - 0.022}{\varepsilon + 0.222}$	0.031	0.957	0.947	0.058

[0180] ε 与 ω_1 的各个拟合函数表达式和拟合精度如表2所示,其中构造函数的拟合精度最高,但函数表达式较为复杂,幂函数拟合精度次之,四次多项式函数拟合精度最低,其误差平方和SSE为0.044。此外, ε - ω_1 各自拟合函数如图4所示,由图4可以明显看出多项式函数拟合效果较差,并且其图像并不符合本文的设计要求。因此,综合考虑拟合函数复杂度和拟合精度的要求,选择幂函数 $\omega_1 = -0.334\varepsilon^{-0.396} + 1.025$ 作为主要属性数目与距离计算属性数目的比值 ε 与主要属性相似度权值 ω_1 的关系函数,如式(2-8)。

[0181] $\omega_1 = \text{Max}[\text{Min}(-0.334\varepsilon^{-0.396} + 1.025, 1), 0]$ (2-8)

[0182] ε - ω_1 的函数表达式如上所示,由于权值要求 $0 \leq \omega_1 \leq 1$,故本专利通过Min-Max限制 ω_1 的范围。

[0183] 1.4次要属性相似度计算

[0184] 次要属性是指元器件中较为关键的那些属性,其重要程度略低于主要属性。次要属性也需满足该属性下所包含的种类为枚举型字符串,或者数据类型为数值型但其种类数目远小于此子类元器件数目这一要求。次要属性应被该子类中的大部分元器件所共有,其各属性包含的种类数目可以相对较少。

[0185] 次要属性相似度的计算与主要属性的方法相同。本发明一个具体实施例中,通过定义广义相似度来衡量两元器件间次要属性间的相似程度。参考(2-2)式,次要属性广义相似度的计算公式如下式(2-9)。

$$[0186] \quad S_{S_{i,j}} = \frac{\sum_{k=0}^n X_{(k_i,k_j)}}{n} \times 100\%, \quad X_{(k_i,k_j)} = \begin{cases} 1 & k_i = k_j \\ 0 & k_i \neq k_j \end{cases} \quad (2-9)$$

[0187] 上式中, $S_{S_{i,j}}$ 为元器件*i,j*间的次要属性相似度; k_i 为元器件*i*在次要属性*k*中的种类; n 为该子类中元器件所有次要属性的数目; $X_{(k_i,k_j)}$ 为判断元器件*i,j*的次要属性*k*异同符号变量。

[0188] 显然,如果两元器件每一次要属性下种类相同的数目越多,则两元器件间的次要属性相似度值 $S_{S_{i,j}}$ 将越大。将次要属性相似度作为次要相似度,同时构建元器件综合相似度模型。

[0189] 1.5综合相似度模型

[0190] 结合元器件间的主要相似度模型式(2-7)和次要相似度模型(2-9),构建元器件综合相似度模型,如式(2-10)所示。

$$[0191] \quad S_{C_{i,j}} = \begin{cases} \omega_a P_{S_{i,j}} + \omega_b S_{S_{i,j}} \\ P_{S_{i,j}}, & S_{S_{i,j}} = 0 \end{cases}$$

$$[0192] \quad \text{s.t. } \omega_a + \omega_b = 1$$

$$[0193] \quad \omega_a \geq \omega_b \geq 0 \quad (2-10)$$

[0194] 上式中, $S_{C_{i,j}}$ 为元器件*i,j*间的综合相似度; ω_a, ω_b 分别表示元器件主要相似度和次要相似度的权值。

[0195] 在本发明的具体实施例二中,一般情况下,在计算元器件综合相似度时,认为主要相似度较次要相似度更为重要,故 $\omega_a \geq \omega_b \geq 0$,这里取 $\omega_a = 0.65, \omega_b = 0.35$ 。若元器件无次要属性时,则将主要相似度作为元器件综合相似度。

[0196] 通过构建元器件综合相似度模型计算综合相似度,可以衡量元器件间的相似程度,实现相似元器件的查找和元器件替代选型功能。

[0197] 综上,本发明的具体实施例中,首先根据元器件的主要属性筛选出部分相似的元器件(约100组左右)并计算其主要属性相似度。然后在筛选出的元器件中计算各个元器件间距离计算属性的欧氏距离,并将其转换为相似度。接下来结合主要属性相似度和距离计算属性相似度构建主要相似度模型,计算各筛选器件的主要相似度。同时构建次要属性相似度模型,计算各筛选器件的次要属性相似度。最后根据元器件间的主要相似度和次要相似度构建综合相似度模型,计算元器件的综合相似度,并根据综合相似度由大到小对相似元器件进行排序,得到相似程度由大至小元器件列表。

[0198] 在本发明的具体实施例二中,元器件替代选型模块(替换元件选择模块)由Python语言编写,其运行环境为Python 3.6.4。该模块所依赖的外部函数模块包括:numpy 1.14.0、pandas 0.22.0、progressbar23.36.0、sklearn、csv、time、random。该模块可以根据被替代元器件的相关信息,查找与其相似的元器件,计算元器件间的综合相似度,同时保存相似元器件的索引和相似度。该模块所包含的主要函数如表3所示。

[0199] 表3

[0200]

函数名	函数功能描述
similar_parts_search	主函数,实现元器件替代选型的功能
z_score	对已经清洗后的元器件的距离计算属性数据Z-Score标准化处理
block_original	依据元器件各个属性中的不同种类对元器件数据集合进行分块,构建分块嵌套列表
part_distance	查找与被替代元器件相似的元器件,并保存相似元器件的索引和综合相似度
index_find	在嵌套列表中查找与被替代元器件属性值相同的元器件索引
sort_index	根据元器件中属性种类相同的数目由多到少对相似元器件的索引进行排序
resort_block_index	对拥有相同属性种类数目的元器件索引值由小到大排序
distance_calculate	计算元器件间距离计算属性相似度
distance_sort	根据距离计算属性相似度由大到小对相似元器件的索引进行排序
design_distance	计算元器件重要属性或次要属性间的广义相似度
find_ciss_id	根据经搜索得到的相似元器件的索引查找其对应的型号信息
test_results	在子类元器件数据库中随机选择多个元器件作为测试集,搜索与之对应的相似元器件并保存其索引和综合相似度

[0201] 本模块根据上述综合相似度模型计算所得相似元器件列表并结合元器件实际使用数据和元器件有关替代资料信息,获得元器件替代型号,并将其推荐给用户。同时,根据用户对替换元器件使用情况的反馈信息不断迭代、更新元器件替代方案。

[0202] 综上所述,本发明实施例所述的电子元器件替代选型系统及选型方法,针对元器件命名规则、性能参数格式、质量等级、分类体系不统一的问题,本专利通过对收集所得的大量元器件原始数据进行整理,制定统一的规则,编写数据清洗模块,对元器件的性能参数数据进行清洗,统一元器件的单位格式等。

[0203] 针对元器件相似性搜索时间成本、空间成本高的问题,本专利结合元器件数据自有的特点,首先根据元器件的核心参数排除大部分相似度较低或不能实现功能替代的元器件。然后结合元器件属性参数数目判断该属性所属类型(数值型或符号型),对于符号型数据计算广义相似度,对于数值型数据计算欧氏距离,缩短计算时间。

[0204] 通过以上的实施方式的描述可知,本领域的技术人员可以清楚地了解到本发明可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在存储介质中,如ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例或者实施例的某些部分所述的方法。

[0205] 以上所述,仅为本发明较佳的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应该以权利要求的保护范围为准。



图1

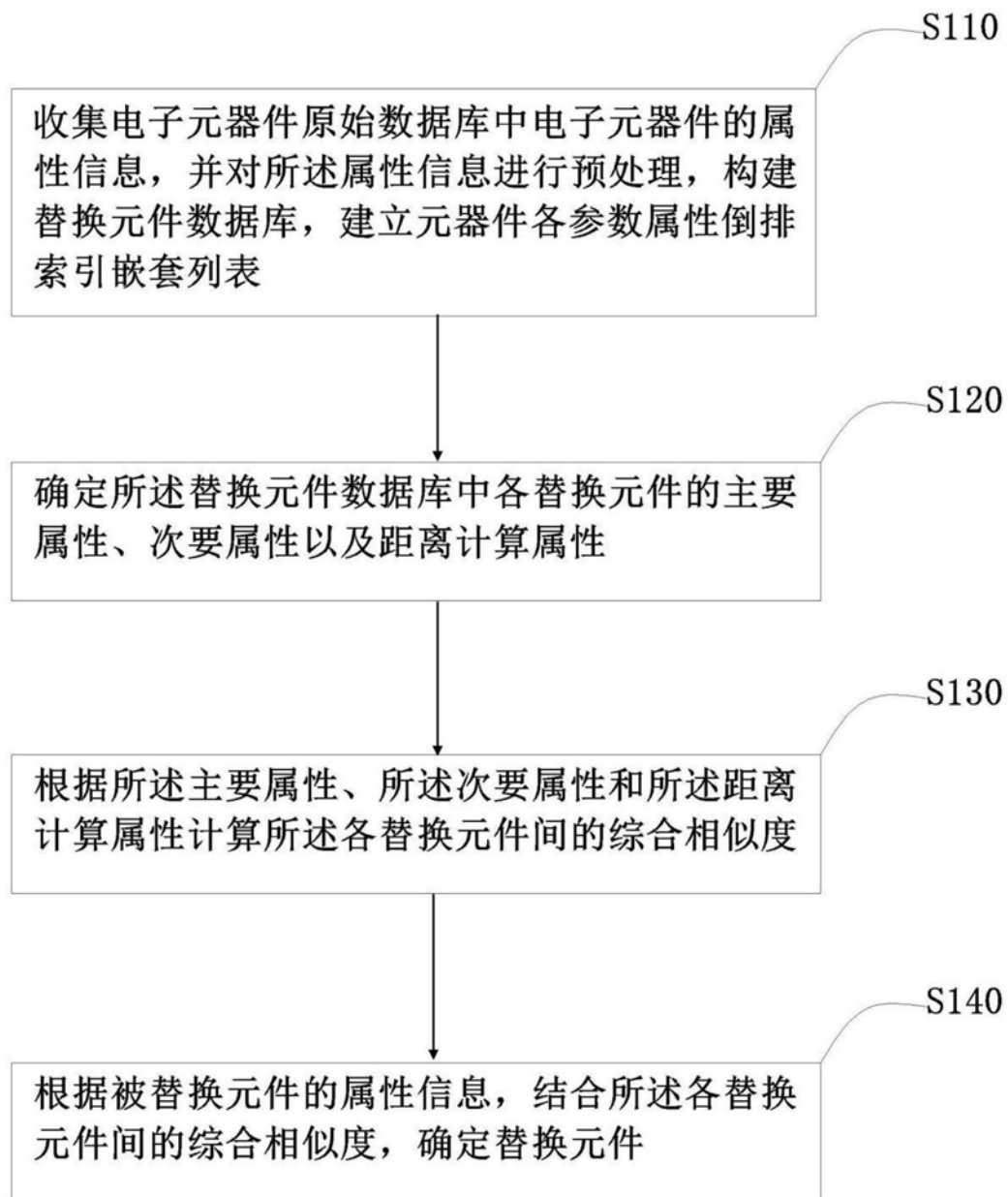


图2

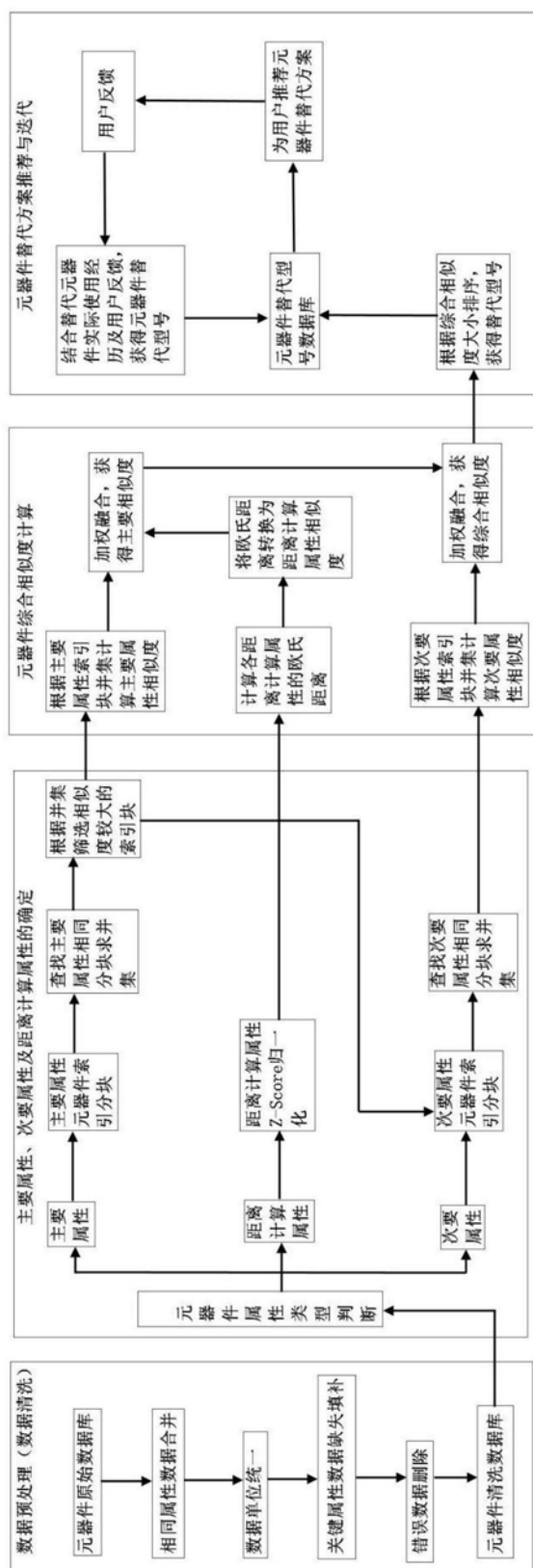


图3

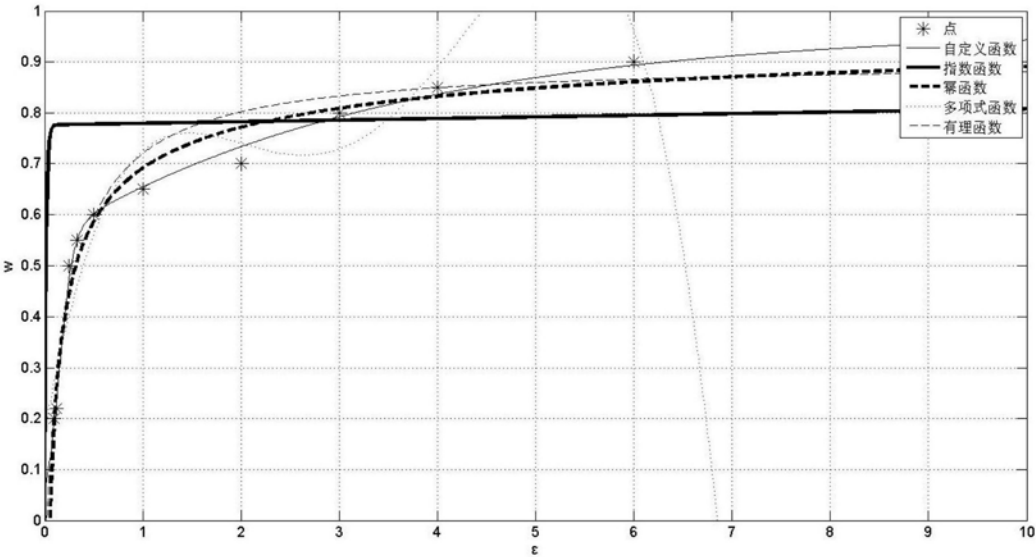


图4