



(12) 发明专利申请

(10) 申请公布号 CN 114897382 A

(43) 申请公布日 2022. 08. 12

(21) 申请号 202210553788.3

G06F 1/18 (2006.01)

(22) 申请日 2022.05.19

G06F 11/30 (2006.01)

(71) 申请人 重庆数宜信信用管理有限公司

地址 400056 重庆市巴南区南泉白鹤林16号52幢第三层

申请人 重庆工程学院

(72) 发明人 田茂洪 王万钧 李华林 黄婧

梁健 郭琳 廖宁 李杰

(74) 专利代理机构 重庆乐泰知识产权代理事务所(普通合伙) 50221

专利代理师 郭泽培

(51) Int. Cl.

G06Q 10/06 (2012.01)

G06K 9/62 (2022.01)

G06F 1/20 (2006.01)

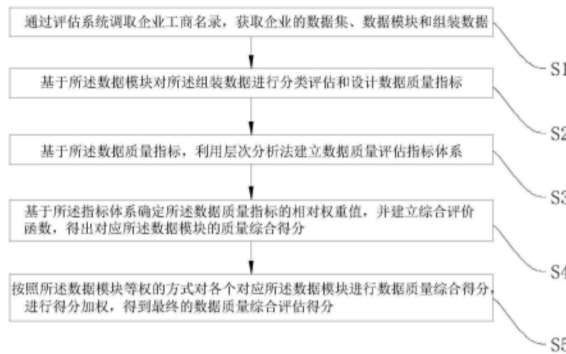
权利要求书1页 说明书10页 附图2页

(54) 发明名称

基于层次分析法的多元异构数据的质量评估方法及系统

(57) 摘要

本发明涉及企业数据管理技术领域,具体涉及基于层次分析法的多元异构数据的质量评估方法及系统;评估系统调取企业工商名录,获取数据集、数据模块和组装数据;基于所述数据模块对所述组装数据进行分类评估和设计数据质量指标;利用层次分析法建立数据质量评估指标体系;基于所述指标体系确定所述数据质量指标的相对权重值,并建立综合评价函数,得出对应所述数据模块的质量综合得分;按照所述数据模块等权的方式对各个对应的所述数据模块进行数据质量综合得分加权,得到最终的数据质量综合评估得分,该方法从定性和定量的角度出发,有效地评估企业多源异构数据的质量水平,供使用者根据实际情况选用。



1. 基于层次分析法的多元异构数据的质量评估方法及系统,其特征在于,包括以下步骤:

通过评估系统调取企业工商名录,获取企业的数据集、数据模块和组装数据;

基于所述数据模块对所述组装数据进行分类评估和设计数据质量指标;

基于所述数据质量指标,利用层次分析法建立数据质量评估指标体系;

基于所述指标体系确定所述数据质量指标的相对权重值,并建立综合评价函数,得出对应所述数据模块的质量综合得分;

按照所述数据模块等权的方式对各个对应的所述数据模块进行数据质量综合得分,进行得分加权,得到最终的数据质量综合评估得分。

2. 如权利要求1所述的基于层次分析法的多元异构数据的质量评估方法,其特征在于,所述通过评估系统调取企业工商名录,获取企业的数据集、数据模块和组装数据的具体方式为:

评估系统调取企业工商名录,并通过条件抽样的方式相结合,生成相应的目标客户名单的数据集;

基于所述数据集调用目标公司的相关数据,并将所述相关数据分类,得到数据模块;

通过服务接口对所述数据集进行调取,并对调取的所述数据集进行多源异构数据的组装,得到组装数据。

3. 如权利要求1所述的基于层次分析法的多元异构数据的质量评估方法,其特征在于,所述数据模块包括企业的工商基本信息、企业司法信息、企业行政处罚信息、企业违法违规信息、企业知识产权信息和企业实际经营信息。

4. 如权利要求2所述的基于层次分析法的多元异构数据的质量评估方法及系统,其特征在于,所述条件抽样筛选条件包括企业的数据的具有一定的丰富性,且相应字段的为空的比列小于第一预设值、企业区域等分布的集中度小于第二预设值和主体名称包含个体工商户和企业。

5. 如权利要求1所述的基于层次分析法的多元异构数据的质量评估方法及系统,其特征在于,所述基于所述数据质量指标,利用层次分析法建立数据质量评估指标体系的具体方式为:

基于所述数据质量指标构建一个四层的层次结构模型;

基于所述数据质量指标构建判断矩阵;

结合所述结构模型和所述判断矩阵,得到数据质量评估指标体系。

6. 如权利要求5所述的于层次分析法的多元异构数据的质量评估方法及系统,其特征在于,所述评估系统包括壳体、评估器和散热组件,所述评估器设置于所述壳体内,所述散热组件设置于所述壳体一侧。

7. 如权利要求6所述的于层次分析法的多元异构数据的质量评估方法及系统,其特征在于,所述散热组件包括风扇、防尘罩、冷风管和出风板,所述出风板设置于所述壳体内,所述冷风管与所述出风板连通,并贯穿所述壳体,所述风扇设置于所述冷却管内,所述防尘罩与所述壳体拆卸连接,且位于靠近所述风扇一侧。

基于层次分析法的多元异构数据的质量评估方法及系统

技术领域

[0001] 本发明涉及企业数据管理技术领域,尤其涉及基于层次分析法的多元异构数据的质量评估方法及系统。

背景技术

[0002] 企业经营过程中产生的各类数据支撑着企业征信服务的运营、监测和管理等重要活动,而企业数据的采集质量决定着征信服务的质量,低质量的企业信用信息对数据准确性、可靠性均得不到有效的保证,将影响企业真实信用的评估。在信用评估的实践中,企业信息的数据质量往往与企业实际的评级有着直接的关系,也受到企业实际在经营期间的经营活动等多种因素的影响。

[0003] 当前对于企业征信领域的的数据质量评估实现方法分大概有:一是在数据库的基础上,通过脚本的方式进行统计采样风险,评估相关数据字段的质量,一是通过数据库脚本统计分析的手段来评估历史数据质量水平,在技术和实现上有一定的局限性;二是采用传统机器学习技术,结合各类算法来评估数据质量。这种方法需要准备的样本进行人工标签,形成数据质量评估模型,当数据的清洗规则和约束发生变更时,则需要重新训练模型且过程比较繁杂。

[0004] 但以上方法均基于静态数据,在面对动态的对多源异构接口数据时评估能力是非常有限的,因此,需要一种数据质量评估方法与系统。

发明内容

[0005] 本发明的目的在于提供基于层次分析法的多元异构数据的质量评估方法及系统,旨在解决传统静态评估方法无法对动态的数据进行评估的问题。

[0006] 为实现上述目的,本发明提供了基于层次分析法的多元异构数据的质量评估方法,包括以下步骤:

[0007] 通过评估系统调取企业工商名录,获取企业的数据集、数据模块和组装数据;

[0008] 基于所述数据模块对所述组装数据进行分类评估和设计数据质量指标;

[0009] 基于所述数据质量指标,利用层次分析法建立数据质量评估指标体系;

[0010] 基于所述指标体系确定所述数据质量指标的相对权重值,并建立综合评价函数,得出对应所述数据模块的质量综合得分;

[0011] 按照所述数据模块等权的方式对各个对应的所述数据模块进行数据质量综合得分,进行得分加权,得到最终的数据质量综合评估得分。

[0012] 其中,所述通过评估系统调取企业工商名录,获取企业的数据集、数据模块和组装数据的具体方式为:

[0013] 评估系统调取企业工商名录,并通过条件抽样的方式相结合,生成相应的目标客户名单的数据集;

[0014] 基于所述数据集调用目标公司的相关数据,并将所述相关数据分类,得到数据模

块；

[0015] 通过服务接口对所述数据集进行调取，并对调取的所述数据集进行多源异构数据的组装，得到组装数据。

[0016] 其中，所述数据模块包括企业的工商基本信息、企业司法信息、企业行政处罚信息、企业违法违规信息、企业知识产权信息和企业实际经营信息。

[0017] 其中，所述条件抽样筛选条件包括企业的数据的具有一定的丰富性，且相应字段的为空的比列小于第一预设值、企业区域等分布的集中度小于第二预设值和主体名称包含个体工商户和企业。

[0018] 其中，所述基于所述数据质量指标，利用层次分析法建立数据质量评估指标体系的具体方式为：

[0019] 基于所述数据质量指标构建一个四层的层次结构模型；

[0020] 基于所述数据质量指标构建判断矩阵；

[0021] 结合所述结构模型和所述判断矩阵，得到数据质量评估指标体系。

[0022] 其中，所述评估系统包括壳体、评估器和散热组件，所述评估器设置于所述壳体内，所述散热组件设置于所述壳体一侧。

[0023] 其中，所述散热组件包括风扇、防尘罩、冷风管和出风板，所述出风板设置于所述壳体内，所述冷风管与所述出风板连通，并贯穿所述壳体，所述风扇设置于所述冷却管内，所述防尘罩与所述壳体拆卸连接，且位于靠近所述风扇一侧。

[0024] 本发明的基于层次分析法的多元异构数据的质量评估方法及系统，通过评估系统调取企业工商名录，获取企业的数据集、数据模块和组装数据；基于所述数据模块对所述组装数据进行分类评估和设计数据质量指标；基于所述数据质量指标，利用层次分析法建立数据质量评估指标体系；基于所述指标体系确定所述数据质量指标的相对权重值，并建立综合评价函数，得出对应所述数据模块的质量综合得分；按照所述数据模块等权的方式对各个对应所述数据模块进行数据质量综合得分，进行得分加权，得到最终的数据质量综合评估得分，该方法按照分数据版块建模的思路，从唯一性、完整性、正确性、一致性和有效性五个方面对围绕企业相关数据的使用质量进行检验评估，再通过建立的企业数据使用质量的评价指标体系，利用层次分析法计算各评价指标的相对权重值，经综合评价函数得出对于模块的数据质量综合得分，最后采用综合加权的方式得到其数据的使用质量综合得分。本发明从定性、定量地两个角度出发评估企业相关数据的使用质量水平，以供其使用者根据实际情况选用，解决传统静态评估方法无法对动态的数据进行评估的问题。

附图说明

[0025] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术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。

[0026] 图1是本发明提供的基于层次分析法的多元异构数据的质量评估方法流程图。

[0027] 图2是所述通过评估系统调取企业工商名录，获取企业的数据集、数据模块和组装数据的具体方式流程图。

[0028] 图3是所述基于所述数据质量指标,利用层次分析法建立数据质量评估指标体系的具体方式流程图。

[0029] 图4是本发明提供的基于层次分析法的多元异构数据的质量评估系统的纵向剖视图。

[0030] 图中:1-壳体、2-评估器、3-散热组件、4-风扇、5-防尘罩、6-冷风管、7-出风板、8-导管、9-制冷片、10-固定杆、11-出风板本体、12-气嘴、13-连接杆。

具体实施方式

[0031] 下面详细描述本发明的实施例,所述实施例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。下面通过参考附图描述的实施例是示例性的,旨在用于解释本发明,而不能理解为对本发明的限制。

[0032] 请参阅图1至图4,本发明提供基于层次分析法的多元异构数据的质量评估方法,包括以下步骤:

[0033] S1:通过评估系统调取企业工商名录,获取企业的数据集、数据模块和组装数据;

[0034] 具体方式:

[0035] S11:评估系统调取企业工商名录,并通过条件抽样的方式相结合,生成相应的目标客户名的单数据集;

[0036] 其中,所述条件抽样筛选条件包括企业的数据的具有一定的丰富性,且相应字段的为空的比列小于第一预设值 ε_1 、企业区域等分布的集中度小于第二预设值 ε_2 和主体名称包含个体工商户和企业。

[0037] S12:基于所述数据集调用目标公司的相关数据,并将所述相关数据分类,得到数据模块。

[0038] 其中,所述数据模块包括企业的工商基本信息、企业司法信息、企业行政处罚信息、企业违法违规信息、企业知识产权信息和企业实际经营信息,具体如下表所示:

数据版块	数据字段
[0039] 工商基本信息	企业名称、企业 id、统一社会信用代码、注册号、组织机构代码、纳税人识别号、电话、邮箱、官网、企业名称简称、企业英文名称、经营状态、法定代表人、法定代表人 id、法定代表人住址、法定代表人电话、法定代表人邮箱、企业类型、企业类型编码、成立日期、注销原因、注销日期、吊销日期、登记机关、所属行业代码、所属行业名称、核准日期、员工人数、参保人数、经营地址、注册地址、营业开始日期、营业结束日期、营业期限、所在省份、所在城市、所在区县、所属地区、经营地址行政区划、注册地址行政编号、最后年检年度、注册资本、注册资本币种、实收资本、认缴出资币种、认缴出资额、经营范围、企业规模、上级机构名称、上级机构类型、简介、曾用名、logo 图片地址、企业名称、统一社会信用代码、企业 id、变更事项、变更日期、变更前内容、变更后内容等。
企业司法信息	企业名称、统一社会信用代码企、业 id、文书标题、被执行人名称、被执行人证件类型、被执行人证件号码、立案日期、案号、执行法院、执行金额(元)、执行标的(元)、执行内容、执行依据文号、终本日期、未履行金额(元)、申请

	执行人、案件状态、异议备注、诉讼地位、企业名称、统一社会信用代码、企业 id、文书标题、审结日期、被执行人名称、被执行人证件种类、被执行人证件号码、案号、诉讼地位、执行法院、正负面、文书类型、案由等。
企业行政处罚信息	企业名称、统一社会信用代码、企业 id、行政处罚记录编号、违法行为类型、违法事实、违法行为、处罚事由、处罚依据、处罚种类、处罚种类中文、处罚内容、处罚机关、处罚有效期、处罚决定文书号、处罚决定文书内容、决定机关、处罚日期、处罚金额、处罚决定书签发日期、公示日期、公示截止期、处罚执行情况、行政复议结果、是否历史信息等。
企业违法违章信息	企业名称、统一社会信用代码、企业 id、列入日期、列入机关名称、经营异常列入原因、移出日期、移出机关名称、退出异常名录原因、企业名称、统一社会信用代码、企业 id、行政处罚记录编号、违法行为类型、违法事实、违法行为、处罚事由、处罚依据、处罚种类、处罚种类中文、处罚内容、处罚机关、处罚有效期、处罚决定文书号、处罚决定文书内容、决定机关、处罚日期、处罚金额、处罚决定书签发日期、公示日期、公示截止期、处罚执行情况、行政复议结果、是否历史信息等
[0040] 企业知识产权信息	企业名称、统一社会信用代码、企业 id、专利名称、专利类型、专利申请日期、专利申请号、专利号、授予日期、专利有效期、专利授权公开日、专利授权公开号、简单法律状态、优先权号、优先权日、申请人、发明人、代理机构、代理人、IPC 分类号、CPC 分类号、申请人地址、申请人邮编、摘要、摘要附图、法律状态、权利要求、说明书、企业名称、统一社会信用代码、企业 id、著作 id、著作登记号、著作持有企业、著作版本号、著作登记类别、著作登记类别编号、著作作品名称、著作类型名称、著作创作完成日期、著作批准时间、著作相关公司、著作简称、著作最后更新日期、著作首次发表日期、著作登记日期、作品图片链接、代理机构、代理人、流程状态、服务项目/商品、企业名称、统一社会信用代码、企业 id、著作 id、著作登记号、著作登记类别、著作持有企业、著作版本号、著作类型名称、著作软件名称、著作完成时间、著作批准日期、著作相关公司、著作软件简称、著作最后更新日期、著作分类号、著作发布日期等
企业实际经营信息	企业名称、统一社会信用代码、企业 id、进销项标识、发票校验状态、价税合计、平均税率、是否蓝票、购买方的地址和电话、购买方的银行地址、购买方的名称、购买方的纳税人识别号、验码、开票日期、开票时间(时分秒)、发票代码、发票号码、发票状态、发票类型、发票类型编号、是否清单、设备编号、机器码、主代码、主标题、不含税金额合计、蓝票代码、蓝票号码、收款人、备注、复核人、开票人、销方地址电话、销方开户行及账号、销售方的名称、销售方的纳税人识别号、关税表、密码、合计税额、税收编码版本号、上传税局标志、原发票号码、含税税率标识、特殊票种、发票详细信息、商品编码、商品税目、税率、优惠政策标识、含税标志、税率标识、增值税特殊管理等

[0041] S13:通过服务接口对所述数据集进行调取,并对调取的所述数据集进行多源异构数据的组装,得到组装数据。

[0042] 其中,导致所述企业数据产生质量问题的原因包括网络问题、设备问题、系统自身局限性、不同来源数据进行融合时出现的问题以及人为因素。

[0043] S2:基于所述数据模块对所述组装数据进行分类评估和数据质量指标的设计;

[0044] 其中,假设模块数据集 D_j 中共有 m 条相关的数据,将该数据集定义为 $D_j = \{C_1, C_2, \dots, C_m\}$,其中 C_i 为每企业对应应该模块的数据, C_i 数据共有 n 个特征,表示为 $C_j = \{Q_1,$

Q_2, \dots, Q_n },为了更好的度量数据质量,建立数据质量评估标准,标准如下:

[0045] 从唯一性、完整性、正确性、一致性和有效性五个大方面对数据进行评估,具体数据质量评估标准和指标评估模型如下:

[0046] 唯一性指标包括以各个模块为对象的企业数据的每条数据唯一(若存在相同的,则判断为重复)、企业统一社会信用代码唯一(若存在相同,则判断为重复)、相应字段的码值唯一(若存在相同,则判断为重复)、数据结构唯一(不同的接口表,数据的外层结构应该保持一致)上述的每条数据唯一的指标计算如下:

$$[0047] \quad U_{data} = 1 - \frac{\text{count}(D^*)}{\sum_{l=1}^m \text{count}(D_j)},$$

[0048] 其中 D^* 为数据 D_j 中以企业为对象的相同的数据集,编码唯一的指标计算为:

$$[0049] \quad U_{isc} = 1 - \frac{\text{count}(D^*)}{\sum_{l=1}^m \text{count}(D_j)},$$

[0050] 其中 D^* 为数据 D_j 中以企业为唯一标识(统一社会信用代码)的相同的数据集,码值唯一的指标计算为:

$$[0051] \quad U_{code} = 1 - \frac{\text{count}(D^*)}{\sum_{l=1}^m \text{count}(D_j)},$$

[0052] 其中 D^* 为数据 D_j 中码值的相同的数据集,数据结构唯一的指标计算为:

$$[0053] \quad U_{st} = \frac{\text{count}(D^*)}{\sum_{l=1}^m \text{count}(D_j)},$$

[0054] 其中 D^* 为数据 D_j 中码值的相同的数据集,对于数据完整性的评估主要从接口的完整性、数据字段完整性、表(信息)结构的完整性等角度出发进行评估。具体的指标值计算如下:

[0055] 接口完整,数据的实时查询和清洗,对数据接口的完整有较高的要求,数据接口查询条件是否唯一、接口的状态码值设计是否合理等,对于该部分质量的评估可以通过接口实时调用以及查看接口设计文档,

$$[0056] \quad U_{jk} = \frac{\text{count}(K^*)}{\sum_{l=1}^m \text{count}(K_j)},$$

[0057] 其中 K^* 满足条件的数量, K_j 为每模块往下的单一接口,对于数据正确性的评估主要从动态(列表)信息正确、静态(字典)信息正确、映射关系正确等,对于该部分的评估主要通过接口请求回来的数据与自身格式(协议)一致,与三方接口对比字段信息正确,

$$[0058] \quad U_{jz} = \frac{\text{count}(K^*)}{\sum_{l=1}^m \text{count}(K_j)},$$

[0059] 其中 K^* 满足条件的数量, K_j 为每模块往下的单一接口,对于数据一致性的评估主要从协议一致(在不同的表中,相同语义的字段,相应的协议是否一致)、字段内容(类型)一致(指报文格式与传输通信协议的一致性情况,区别于判断数据的正确性)、与三方接口协议的一致(相同含义的字段,两接口保持一致)、接口响应时间的一致(与所提供的资料以及三方接口调用的时间进行综合比较)等,

$$[0060] \quad U_{jy} = \frac{\text{count}(K^*)}{\sum_{l=1}^m \text{count}(K_j)},$$

[0061] 其中 K^* 满足条件的数量, K_j 为每模块往下的单一接口,对于数据有效性的评估主要从数据内容的有效(相同类别的数据,返回数据字段保持一致)、数据更新时间的有效性(与所提供的资料以及三方接口调用的时间进行综合比较);

$$[0062] \quad U_{jx} = \frac{\text{sum}(K^*)}{\sum_{l=1}^m \text{count}(K_j)},$$

[0063] 其中,

$$[0064] \quad K^* = \begin{cases} 1(T_d \geq \theta) \\ 0(T_d < \theta) \end{cases}, T_d = T_{end} - T_{start};$$

[0065] 满足条件的数量, K_j 为每模块往下的单一接口,灾备机制的有效性(是否存在相应的灾备机制和硬件环境,并有相应试运行,若试运行成功的次数大于 λ ,则有效为1,否则为0),

$$[0066] \quad U_{jb} = \frac{\text{sum}(K_{jb}^*)}{\sum_{l=1}^m \text{count}(K_j)};$$

$$[0067] \quad \text{其中: } K_{jb}^* = \begin{cases} 1(\text{if } T_d \leq \lambda) \\ 0(\text{if } T_d > \lambda) \end{cases}, T_d = T_{end} - T_{start};$$

[0068] 在规定的时间内启动相应预案方案;数据管理制度的有效性(是否存在相应的数据管理制度,若 n 抽查次数,在规定的时间内 λ ,则为1;否则为0),

$$[0069] \quad U_{js} = \frac{\text{sum}(K_{js}^*)}{\sum_{l=1}^m \text{count}(K_j)},$$

$$[0070] \quad \text{其中: } K_{js}^* = \begin{cases} 1(\text{if } T_d \leq \lambda) \\ 0(\text{if } T_d > \lambda) \end{cases}, T_d = T_{end} - T_{start}.$$

[0071] 接口响应时间的一致(与所提供的资料以及三方接口调用的时间进行综合比较)等。具体的指标如下:

大类	小类	指标	说明	
[0072]	使用质量	正确性	动态(列表)信息正确	自身格式(协议)一致,与三方接口对比字段信息保持一致;
	使用质量	正确性	静态(字典)信息正确	自身格式(协议)一致,与三方接口对比字段信息保持一致;
	使用质量	正确性	映射关系正确	自身格式(协议)一致,与三方接口对比字段信息保持一致;
	使用质量	一致性	协议一致	在不同的表中,相同语义的字段,相应的协议是否一致;
	使用质量	一致性	字段内容(类型)一致	指报文格式与传输通信协议的一致性情况,区别于判断数据的正确性;
	使用质量	一致性	与三方接口协议的一致	相同含义的字段,两接口保持一致;
	使用质量	一致性	接口响应时间的一致	与所提供的资料以及三方接口调用的时间进行综合比较;
	使用质量	有效性	数据内容的有效	相同类别的数据,返回数据字段保持一致;
	使用质量	有效性	数据更新时间的有效性	与所提供的资料以及三方接口调用的时间进行综合比较;
	使用质量	有效性	灾备机制的有效性	是否存在相应的灾备机制和硬件环境,并有相应试运行;
	使用质量	有效性	数据管理制度的有效性	是否存在相应的数据管理制度(如:生产数据的变更流程及机制);
	[0073]	使用质量	唯一性	每条数据唯一
使用质量		唯一性	编码唯一	主体的标识唯一,且该标识为行业通用且接口调用的查询条件唯一;
使用质量		唯一性	码值唯一	字段码值唯一,且不重复;
使用质量		唯一性	数据结构唯一	不同接口的外层表结构唯一;
使用质量		完整性	数据接口完整	数据接口中各个字段的内容都应该是完整的,且接口入参查询条件是唯一的;
使用质量		完整性	数据字段完整	具有完整的字段、码值、异常校验等协议说明;
使用质量		完整性	表(信息)结构完整	若表内若存在英文、码值,则有对应的中文名称;

[0074] S3:基于所述数据质量指标,利用层次分析法建立数据质量评估指标体系;

[0075] 具体方式:

[0076] S31:基于所述数据质量指标构建一个四层的层次结构模型;

[0077] 其中,第一层是目标层,是数据质量评估的总体目标;第二层和第三层是中间层,代表数据质量评估的准则,是对总目标的分解;第四层是准则层,代表影响目标实现的具体因素。

[0078] S32:基于所述数据质量指标构建判断矩阵;

[0079] 其中,对所述层次结构模型中每一层指标之间按照下述的标度方法,形成判断矩阵;所述判断矩阵包括原始质量与使用质量的判断矩阵、原始质量唯一性、完整性和正确性的判断矩阵、使用质量一致性和有效性的判断矩阵、唯一性的判断矩阵、完整性的判断矩

阵、正确性的判断矩阵、一致性的判断矩阵以及有效性的判断矩阵。标度方法如下：

序号	重要性程度	标度
1	因素相比较有相同重要性	1
2	一个比另一个稍微重要	3
3	一个比另一个明显重要	5
4	一个比另一个强烈重要	7
5	一个比另一个极度重要	9
6	其他	2/4/6/8

[0081] S33:结合所述结构模型和所述判断矩阵,得到数据质量评估指标体系。

[0082] 综上所述,正确性相对于唯一性和完整性都是明显重要,而完整性对于唯一性是稍微重要的,从而得出判断矩阵如下:

	L11(正确性)	L12(一致性)	L13(有效性)	L14(唯一性)	L15(完整性)
L11(正确性)	1	1/2	1/6	1/2	1/9
L12(一致性)	2	1	1/4	1/3	1/3
L13(有效性)	6	4	1	1/3	1/4
L14(唯一性)	7	3	3	1	1/2
L15(完整性)	9	2	4	2	1

[0084] 假设A层的指标 A_m 与下层指标构成元素 $\{Q_1, Q_2, \dots, Q_n\}$ 有关联,各元素的判断矩阵如下:

$$I = (a_{ij})_{n \times n} = \begin{pmatrix} a_{11} & L & a_{1n} \\ M & O & M \\ a_{n1} & L & a_{nn} \end{pmatrix};$$

[0086] 其中 $a_{ij} * a_{ji} = 1$ 。

[0087] S4:基于所述指标体系确定所述数据质量指标的相对权重值,并建立综合评价函数,得出对应所述数据模块的质量综合得分;

[0088] 其中,计算数据集 D_j 的特征向量和最大特征值得到权重系数,采用方根法求解判断矩阵的特征向量近似解和最大特征值,计算得到每层指标的权重,具体求解过程如下计算每一行元素的乘积,再对乘积求n次方根,其中n为矩阵阶数:

$$\bar{a}_i = \sqrt[n]{\prod_{j=1}^n a_{ij}}。$$

[0090] 计算每行乘积n次方根与方根和的商,构成特征向量 $w_i = (w_1, w_2, \dots, w_3)^T$,其中:

$$w_i = \frac{\bar{a}_i}{a}。$$

[0092] 数据集 D_j 最大特征值 λ 为:

[0093]
$$\lambda = \sum_{i=1}^n \frac{(IW)_i}{n\omega_i} ;$$

[0094] 一致性校验为:

[0095]
$$CR = \frac{\lambda - n}{RI(n-1)},$$

[0096] 若 $CR < 0.1$, 则认为有效, 通过相应的校验, 计算各个指标得分 r_i 。由于上述的指标的范围0到1之间, 现将指标得分按照其指标值从采用线性映射为0-100之间, 得到各个指标得分, 四级分层评估体系, 总分为各个子版块得分, 子版块得分由各个指标得分加权, 各个指标根据其自身的指标值得到指标得分, 即子版块得分:

[0097]
$$score(r, w) = \sum_{i=1}^n (r_i \times w_i) \quad \text{其中, } w_i \in [0, 1], \sum_{i=1}^n w_i = 1$$

[0098] w_i 为各个指标的权重, r_i 为各个指标的得分。

[0099] S5: 按照所述数据模块等权的方式对各个所述对应模块的数据质量综合得分, 进行得分加权, 得到最终的数据质量综合评估得分。

[0100] 其中, 本次评估的最终得分:

[0101]
$$score = \sum_{i=1}^6 (score_i \times w)$$

[0102] 其中 $w = \frac{1}{6}$,

[0103] 然后将本次评估的最终得分按照百分制将得分进行5挡划分, 具体如下:

[0104]

评估等级	差	一般	中等	良好	优秀
分数区间	$0 \leq score < 60$	$60 \leq score < 70$	$70 \leq score < 80$	$80 \leq score < 90$	$90 \leq score \leq 100$

[0105] 进一步的, 所述评估系统包括壳体1、评估器2和散热组件3, 所述评估器2设置于所述壳体1内, 所述散热组件3设置于所述壳体1一侧; 所述散热组件3包括风扇4、防尘罩5、冷风管6和出风板7, 所述出风板7设置于所述壳体1内, 所述冷风管6与所述出风板7连通, 并贯穿所述壳体1, 所述风扇4设置于所述冷却管内, 所述防尘罩5与所述壳体1拆卸连接, 且位于靠近所述风扇4一侧

[0106] 在本实施方式中, 所述壳体1用于保护所述评估器2, 并对所述散热组件3提供安装条件, 所述评估器2用于对企业数据质量进行评估, 所述散热组件3用于对所述评估器2散热, 防止所述评估器2长时间运作的热量较高, 加快所述评估器2的老化, 降低了所述评估器2的使用寿命, 具体的, 通过所述风扇4往所述冷风管6内吹风, 所述冷风管6将风制冷导入所述出风板7, 经所述出风板7均匀的吹向所述评估器2, 使所述评估器2均匀散热, 所述防尘罩5的设置, 防止所述风扇4往所述冷风管6内吹风时吸热空气中的灰尘。

[0107] 所述冷风管6包括导管8、制冷片9和固定杆10, 所述导管8与所述出风板7连通, 并贯穿所述壳体1, 所述固定杆10与所述导管8拆卸连接, 并与所述壳体1拆卸连接, 且位于所述壳体1内, 所述制冷片9设置于所述导管8远离所述出风板7一侧。

[0108] 在本实施方式中, 所述导管8用于将风导入所述出风板7内, 所述制冷片9为半导体制冷片, 用于制冷所述导管8内的风, 提高对所述评估器2的散热效果, 所述固定杆10的设

置,防止所述导管8往所述出风板7导风时发生晃动。

[0109] 所述出风板7包括出风板本体11、气嘴12和连接杆13,所述连接杆13设置于所述壳体1顶部,所述出风板本体11与所述连接杆13拆卸连接,且位于靠近所述评估器2一侧,所述气嘴12设置于所述出风板本体11靠近所述评估器2一侧。

[0110] 在本实施方式中,所述连接杆13用于将所述出风板本体11安装在所述壳体1内,所述气嘴12具有多个,均匀的设置于所述出风板本体11上,所述出风板本体11将冷风分别导入所述气嘴12,通过所述气嘴12均匀的吹向所述评估器2,从而使得所述评估器2得以均匀散热。

[0111] 以上所揭露的仅为本发明一种较佳实施例而已,当然不能以此来限定本发明之权利范围,本领域普通技术人员可以理解实现上述实施例的全部或部分流程,并依本发明权利要求所作的等同变化,仍属于发明所涵盖的范围。

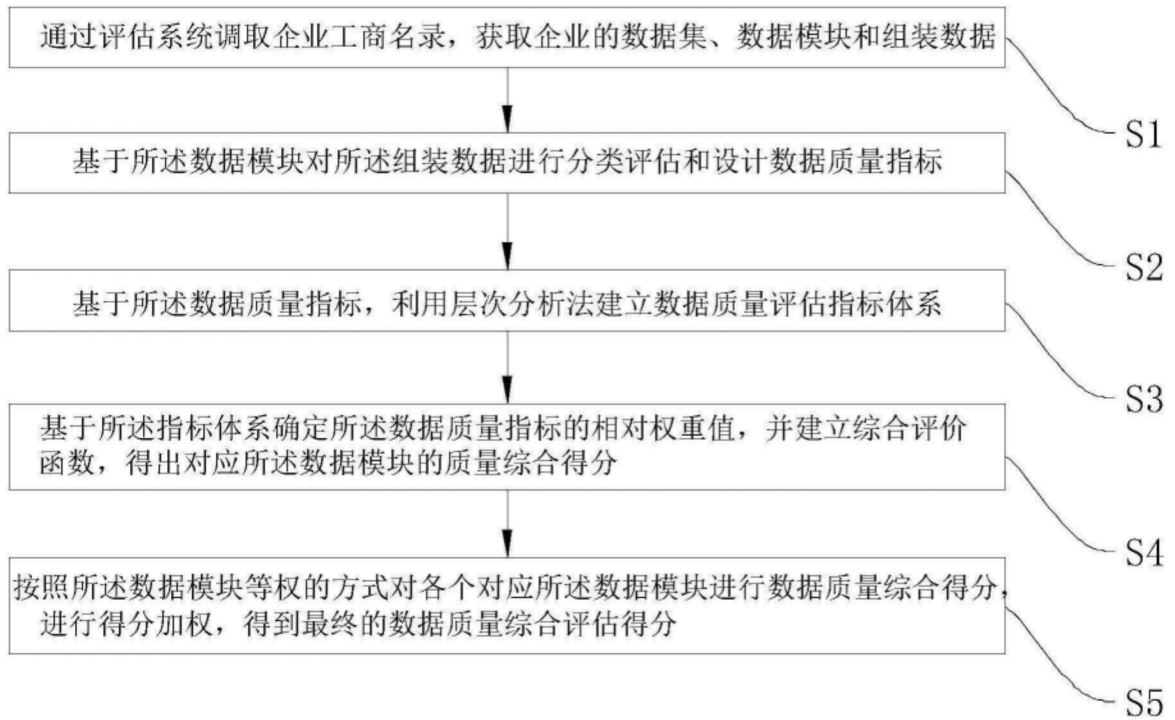


图1

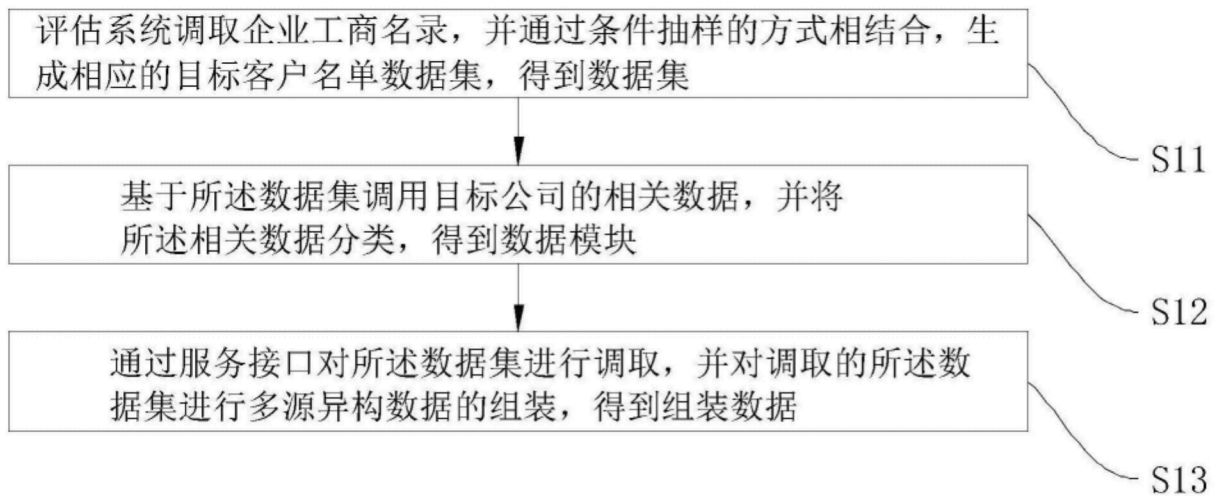


图2

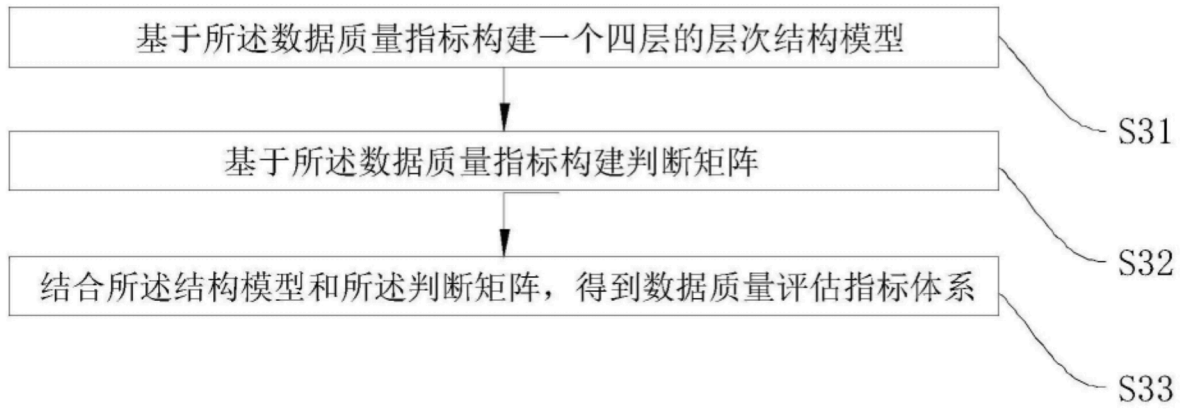


图3

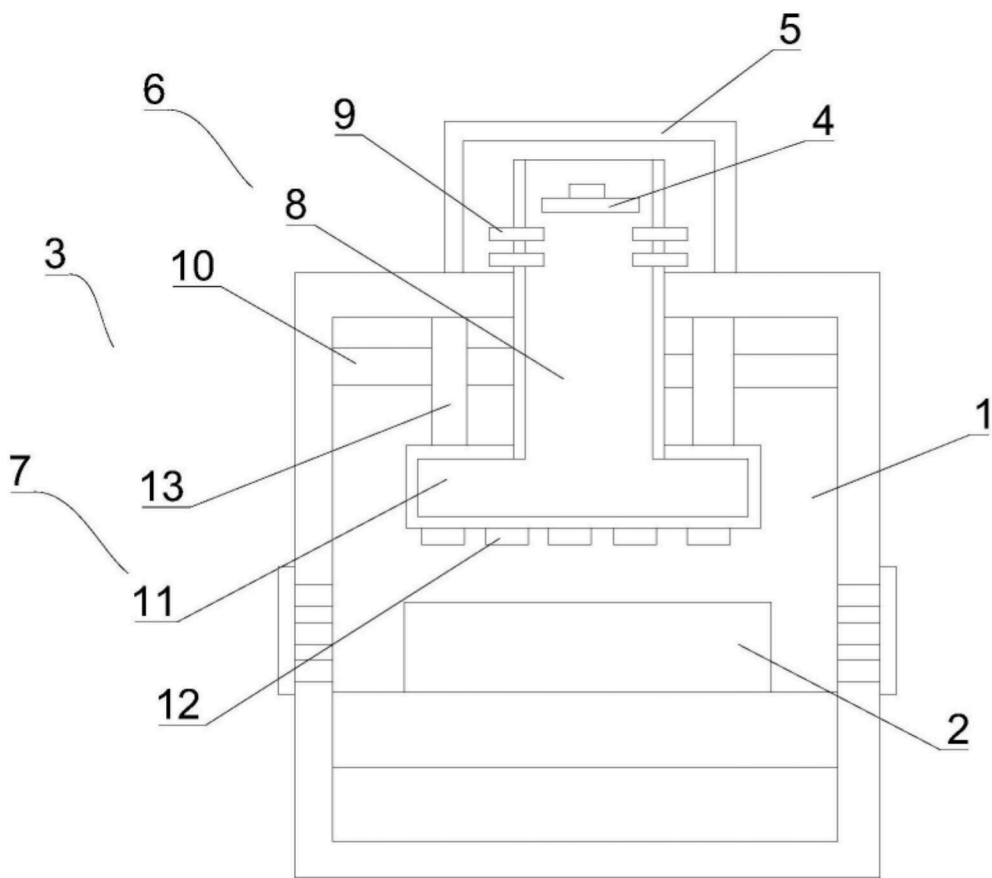


图4