



(12) 发明专利申请

(10) 申请公布号 CN 105354737 A

(43) 申请公布日 2016. 02. 24

(21) 申请号 201510873584. 8

(74) 专利代理机构 南京苏高专利商标事务所
(普通合伙) 32204

(22) 申请日 2015. 12. 02

代理人 张弛

(71) 申请人 国家电网公司

地址 210061 江苏省南京市高新区高新路
20 号

(51) Int. Cl.
G06Q 30/06(2012. 01)

申请人 国电南瑞科技股份有限公司
国网天津市电力公司
南京南瑞集团公司
国电南瑞南京控制系统有限公司

(72) 发明人 杨永标 蒋菱 项添春 于建成
宋杰 王峥 王天昊 王旭东
吴磊 姚程 刘建宇 田娜
刘金华 杨伟光 黄莉 陈璐
王金明 颜盛军 朱庆 谢敏
周静 王冬 李奕杰

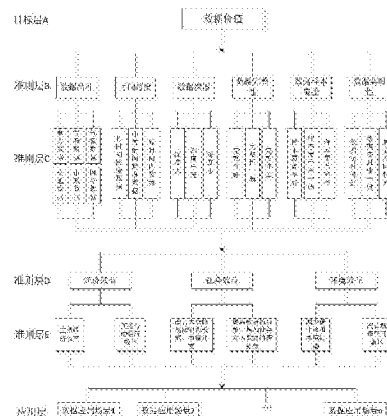
权利要求书2页 说明书4页 附图1页

(54) 发明名称

一种适合大数据价值评价的计算方法

(57) 摘要

本发明提出了一种适合大数据价值评价的计算方法,将数据品种、时间跨度、数据深度、数据完整性、数据样本覆盖和数据实时性与应用场景相结合,利用层次分析法既可进行定性分析又可定量分析的特点,进行数据价值的评价。并利用大数据技术,在不同的数据应用场景下又将产生不同的经济效益、社会效益和环境效益。所以提出将数据价值评价的六个维度和应用场景下的三类效益进行细分。首先确定某一类数据应用场景下的数据价值的基值量;然后由下往上进行权重计算,计算时先进行同层的各类细分环节两两之间的重要性比较的定性分析,给出它们之间的标度,然后进行定量计算,并乘上数据价值的基值量;最后得出该应用场景下的数据价值。



1. 一种适合大数据价值评价的计算方法,其特征在於,包括以下步骤:

(1)、将数据价值的评价过程所涉及的影响因素分成若干层次,建立多级递阶的层次结构模型,包括目标层、准则层和应用层,目标层即为数据价值;

(2)、确定应用层的数据价值基值量;

(3)、将同一层次中的任意两个影响因素进行重要性比较,对该两个影响因素重要性之比做出判断,给予量化;

为什么计算相对复杂,目的是什么?

(4)、对同一层次上的各个因素,以上一级的因素为依据,进行两两比较,根据评价尺度对它们的重要性给予一定的标度,确定其相对重要度,并据此构建判断矩阵;

(5)、计算判断矩阵的特征向量,以此确定各层因素的相对重要度,该相对重要度以权重量化;

(6)、通过上述权重的计算,再乘上数据价值基值量,即得到某一应用场景下的数据价值。

准则层的3个子层

2. 根据权利要求1所述的适合大数据价值评价的计算方法,其特征在於:步骤(2)中,准则层分为若干层,其中最靠近目标层的第一准则层包括数据品种、时间跨度、数据深度、数据完整性、数据样本覆盖和数据实时性;将数据品种、时间跨度、数据深度、数据完整性、数据样本覆盖和数据实时性每个再细分后的细分数据得到第二准则层;所述第三准则层包括经济效益、社会效益及环境效益。

3. 根据权利要求2所述的适合大数据价值评价的计算方法,其特征在於:所述步骤(3)中,第二基准层中除数据品种之外的各细分数据先定性分析,即按照自定的尺度,将细分数据归类,然后量化;该第二基准层中的细分数据之间不进行两两比较。

4. 根据权利要求2所述的适合大数据价值评价的计算方法,其特征在於:所述步骤(4)中,在确定了某一应用场景后,从下往上开始计算权重;对同属一层的各要素,以上一级的要素为准则进行两两比较,根据评价尺度确定其相对重要性,并据此构建判断矩阵A,并用Santy的1-9标度方法给出判断矩阵A的元素 a_{ij} 。

5. 根据权利要求2所述的适合大数据价值评价的计算方法,其特征在於:所述步骤(5)中,采用几何平均法计算各层相对于上一层的特征向量,归一化后得到的元素即为其权重系数W:

$$W = \frac{(\prod_{j=1}^n a_{ij})^{\frac{1}{n}}}{\sum_{i=1}^n (\prod_{j=1}^n a_{ij})^{\frac{1}{n}}}, \quad i=1,2,3,\dots,n, \quad W=(w_1, w_2, \dots, w_n)$$

权重系数采用这样的方式计算出目的和思路是什么?

其中n的值为各层的因素个数;

计算步骤如下:

判断矩阵A的元素按行相乘得一新向量;

将新向量的每个分量开n次方;

将所得的向量归一化即为权重向量;其对应的数值也就是下一层中各元素自身相对于上一层要素所占的权重;

根据层次法原理,利用判断矩阵A的理论最大特征值 λ_{\max} 与n之差检验一致性,一致

性指标 CI :

$$CI = \frac{\lambda_{\max} - n}{n - 1}; \quad \lambda_{\max} = \frac{1}{n} \sum_i \frac{(AW)_i}{w_i}$$

其中当 $CI < 0.1$, $CR = \frac{CI}{RI} < 0.1$ 时, 判断矩阵的一致性可以接受, 否则重新进行两两比较。

6. 根据权利要求 2 所述的适合大数据价值评价的计算方法, 其特征在于: 所述步骤 (6) 中, 获得同一层次各因素之间的相对重要度后, 就可以自上而下得计算各级因素对总体的综合重要度; 最后乘上该应用场景下的价值量基值就得到了某一数据相对于该应用场景的数据价值。

采用上述较为负载的计算方法的原因是因为这是层次分析法的一般解决问题分析问题的思路和套路

一种适合大数据价值评价的计算方法

技术领域

[0001] 本发明涉及的是用于数据价值评价的计算方法,属于大数据交易过程中的数据价值评价算法领域。

背景技术

[0002] 随着大数据时代的来临,服务计算、云计算、物联网等信息技术的快速发展为新兴交易模式的发展提供了新思路。交易的商品越来越“数据化”。电力、市政、交通、气象、经济等行业大数据彼此交融,形式多种多样,包括结构化数据、半结构化数据和非结构化数据等,应用数据挖掘、机器学习、并行计算、流计算等“大数据”技术,可以在体量巨大、类型繁多、低价值密度的数据中挖掘出高附加值的信

息,具有巨大的经济价值和社会价值。
[0003] 大数据交易中的主体是数据,如何给进行交易的数据一个合理的价值评价从而确定数据的交易价格就显得尤为重要。目前国内认可并已投入使用的评价大数据价值的六大维度分别是:数据品种、时间跨度、数据深度、数据完整性、数据样本覆盖和数据实时性。这六个数据价值的评价维度在不同的数据应用场景下所占的权重是不一样的,如果仅仅依靠数据买方或卖方的定性分析和逻辑判断,缺乏定量分析依据来确定数据的价值量,显然是不合理的。

[0004] 层次分析法 (Analytical Hierarchy Process) 综合了人们的主观判断,是一种简明、实用的定性分析与定量分析相结合的系统分析与评价的方法,适用于多目标、多判据的系统综合评价。考虑到数据价值评价的维度多、数据价值效益的维度多,以及不同应用场景下数据价值的不同,数据的价值是不同的,本发明提出了基于不同应用场景下,数据价值计算的层次分析法,以定性分析和定量分析相结合的方式来确定数据的价值量,给数据的价值计算提供了一个较好的研究方向。

发明内容

[0005] 发明目的:针对多元的海量大数据,提出了一种基于层次分析法的适合大数据的数据价值计算方法。

[0006] 技术方案:为达到上述目的,本发明可采用如下技术方案:

[0007] 一种适合大数据价值评价的计算方法,包括以下步骤:

[0008] (1)、将数据价值的评价过程所涉及的影响因素分成若干层次,建立多级递阶的层次结构模型,包括目标层、准则层和应用层;目标层即为数据价值;

[0009] (2)、确定应用层的数据价值基值量;

[0010] (3)、将同一层次的任意两个影响因素进行重要性比较,对该两个影响因素重要性之比做出判断,给予量化;

[0011] (4)、对同一层次上的各个因素,以上一级的因素为依据,进行两两比较,根据评价尺度对它们的重要性给予一定的标度,确定其相对重要度,并据此构建判断矩阵;

[0012] (5)、计算判断矩阵的特征向量,以此确定各层因素的相对重要度,该相对重要度

以权重量化；

[0013] (6)、通过上述权重的计算，再乘上数据价值基值量，即得到某一应用场景下的数据价值。

[0014] 有益效果：使用本发明的适合大数据的数据价值计算方法，可以获得各个数据应用场景下数据的价值量，从而有利于确定使得数据买卖双方都愿意接收的数据价格定位区间，实现大数据交易过程中买卖双方的“双赢”，有益于大数据交易市场的良性发展。

附图说明

[0015] 图 1 是是本发明的基于层次分析法的数据价值评价分层模型示意图。

具体实施方式

[0016] 下面结合附图对本发明进行详细的描述。

[0017] 本发明公开一种适合大数据价值评价的计算方法，包括以下步骤：

[0018] (1)、如图 1 所示，将数据价值的评价过程所涉及的影响因素分成若干层次，建立多级递阶的层次结构模型，包括目标层、准则层和应用层；目标层即为数据价值；

[0019] (2)、确定应用层的数据价值基值量；

[0020] (3)、将同一层次的任意两个影响因素进行重要性比较，对该两个影响因素重要性之比做出判断，给予量化；

[0021] (4)、对同一层次上的各个因素，以上一级的因素为依据，进行两两比较，根据评价尺度对它们的重要性给予一定的标度，确定其相对重要度，并据此构建判断矩阵；

[0022] (5)、计算判断矩阵的特征向量，以此确定各层因素的相对重要度，该相对重要度以权重量化；

[0023] (6)、通过上述权重的计算，再乘上数据价值基值量，即得到某一应用场景下的数据价值。

[0024] 进一步地，根据所述(1)中，大数据交易中的主体是数据，数据本身的质量将直接影响到数据的价值。目前评价数据价值的是数据品种、时间跨度、数据深度、数据完整性、数据样本覆盖和数据实时性这六个维度，由于其是与数据本身的价值最直接相关的，故将其放在最靠近目标层的第一准则层中。因为多元大数据的来源各异，各个种类的数据、各个时间跨度的数据、不同完整性的数据等对六大数据评价维度的影响也不尽相同，所以将六大维度细分后得到第二准则层。不同的数据应用场景下有不同的效益，所以对数据的价值评价中还要考虑到各类效益，故将效益也纳入到准则层中。所有的数据价值都是在应用层即数据的应用场景下进行价值计算的。因为本次发明的目的是求解数据价值量，故将数据价值放在目标层。以目前通用的数据价值六大评价维度（数据品种、时间跨度、数据深度、数据完整性、数据样本覆盖、数据实时性）为第一准则层，然后将数据品种依据其来源分为六大类，而其余的五类维度只做定性分析确定归类到三个子层之中，三个子层相对于其上的五个维度的权重系数按最好的为 0.95，次之的是 0.7，差的为 0.5 做计算。以完整性为例，完整性好相对权重取为 0.95，完整性一般的相对权重取为 0.7，完整性差的取为 0.5

[0025] 所述(2)中，数据价值的基值量的确定，主要是考虑到相同的数据在不同的数据应用场景下，数据价值自身就会有差异，而且对于不同的应用，数据的六大维度所占权重

也不相同,所以需要先确定某场景下,数据的价值基值量,其值的确定是在结合数据增值量(即通过这个数据的研究分析,应用后可能获得的价值)后给出的。

[0026] 而所述(3)中,第二基准层中的时间跨度、数据深度、数据完整性、数据样本覆盖和数据实时性各自的三类细分环节先定性分析,即按照一定的尺度,将数据归类到三个层次中的一个中(如长时间跨度数据、中等时间跨度数据和短时间内数据),然后量化,所以三个层次之间不再做两两比较。

[0027] 进一步地,所述(4)中,在确定了某一应用场景后,从下往上开始计算权重。对同属一层的各要素,以上一级的要素为准则进行两两比较,根据评价尺度确定其相对重要性,并据此构建判断矩阵A。

[0028] 首先用Santy的1-9标度方法给出判断矩阵A的元素 a_{ij} 。其遵循两个原则:1.不是所有因素放在一起比较,而是两两相互比较;2.采用相对尺度,以尽可能减少性质不同的诸因素相互比较的困难,以提高准确度。判断矩阵元素 a_{ij} 的标度方法如下表所示:

[0029]

标度	含义
----	----

[0030]

1	表示两个因素相比,具有同样重要性
3	表示两个因素相比,一个因素比另一个因数稍微重要
5	表示两个因素相比,一个因素比另一个因数明显重要
7	表示两个因素相比,一个因素比另一个因数强烈重要
9	表示两个因素相比,一个因素比另一个因数机端重要
2,4,6,8	上述两相邻判断的中值
倒数	因素 <i>i</i> 与 <i>j</i> 比较的判断 a_{ij} ,则因素 <i>j</i> 与 <i>i</i> 比较的判断 $a_{ji} = 1/a_{ij}$

[0031] 再进一步说明所述(5)中,采用几何平均法(方根法)计算各层相对于上一层的特征向量,归一化后得到的元素即为其权重系数:

[0032]
$$W = \frac{(\prod_{j=1}^n a_{ij})^{\frac{1}{n}}}{\sum_{i=1}^n (\prod_{j=1}^n a_{ij})^{\frac{1}{n}}}, i=1,2,3,\dots,n, W = (w_1, w_2, \dots, w_n)$$

[0033] 其中n的值由该层的因素个数决定。

[0034] 计算步骤如下:

[0035] ①判断矩阵A的元素按行相乘得一新向量;

[0036] ②将新向量的每个分量开n次方;

[0037] ③将所得的向量归一化即为权重向量。其对应的数值也就是下一层中各元素自身

相对于上一层要素所占的权重。

[0038] 需要注意的是,实际评价中,评价者由于只能进行粗略判断,这样有时就会犯不一致的错误。如,已判断 B_1 比 B_2 重要, B_2 比 B_3 较重要,那么 B_1 就应该比 B_3 更重要,但是若判断 B_1 比 B_3 较重要或同等重要,这就犯了逻辑错误,所以需要一致性检验。

[0039] 根据层次法原理,利用判断矩阵 A 的理论最大特征值 λ_{\max} 与 n 之差检验一致性。一致性指标:

$$[0040] \quad CI = \frac{\lambda_{\max} - n}{n - 1}; \quad \lambda_{\max} = \frac{1}{n} \sum_i \frac{(AW)_i}{w_i}$$

[0041] 一般认为 $CI < 0.1$, $CR = \frac{CI}{RI} < 0.1$ 时,判断矩阵的一致性可以接受,否则重新进行两两比较。其中平均随机一致性指标如下表所示:

[0042]

阶数	3	4	5	6	7	8	9	10	11
RI	0.58	0.89	1.12	1.26	1.36	1.41	1.46	1.49	1.52

[0043] 最后所述 (6) 中,获得同一层次各因素之间的相对重要度后,就可以自上而下得计算各级因素对总体的综合重要度。最后乘上该应用场景下的价值量基值就得到了某一数据相对于该应用场景的数据价值

[0044] 以上显示和描述了本发明的基本原理、主要特征、有益效果和实施方案。本发明可以计算不同数据应用场景下数据的价值量,是一种通用的方法。以上重点阐述的是以层次分析法,定性的给各层因素两两之间的重要性一个标度,然后通过计算获得相对权重,最后计算得到数据的价值量。其具体的各个维度、因素的权重以及数据价值的基值量要根据实际的应用场景获取。通过本发明计算所得的数据价值,结合了定性、定量两种分析方式,并具体结合了实际的应用场景,更适用于数据交易中对数据价值的计算评价。

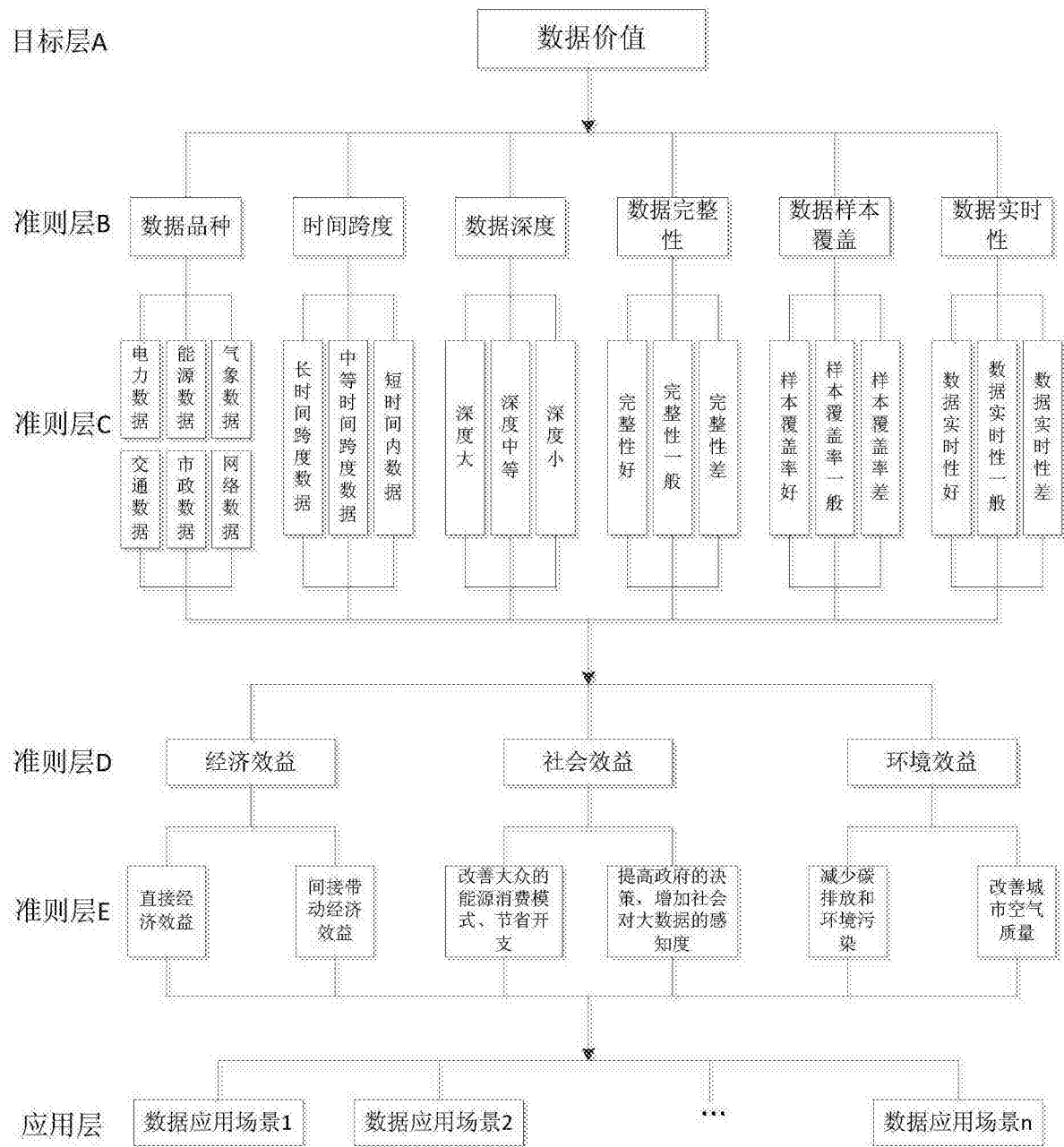


图 1