



(12) 发明专利申请

(10) 申请公布号 CN 103247008 A

(43) 申请公布日 2013.08.14

(21) 申请号 201310163339.9

(22) 申请日 2013.05.07

(71) 申请人 国家电网公司

地址 100031 北京市西城区西长安街 86 号

申请人 国网北京经济技术研究院

长沙理工大学

(72) 发明人 王熙亮 马瑞 秦璇 程鹏

徐慧明

(74) 专利代理机构 北京纪凯知识产权代理有限公司

11245

代理人 徐宁

(51) Int. Cl.

G06Q 50/06 (2012.01)

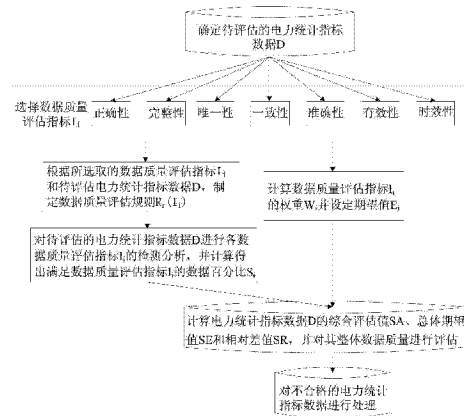
权利要求书2页 说明书10页 附图1页

(54) 发明名称

一种电力统计指标数据的质量评估方法

(57) 摘要

本发明涉及一种电力统计指标数据的质量评估方法,包括以下步骤:根据实际评估目的的需要,获取待评估的电力统计指标数据;选择相应的数据质量评估指标;制定与各质量评估指标对应的数据质量评估规则;计算质量评估指标 I_i 的权重 W_i 并设定期望值 E_i ;对统计指标数据进行各质量评估指标的检测分析,并计算出满足质量评估指标 I_i 的数据百分比 S_i ;根据权重 W_i 、合格百分比 S_i 和设定的期望值 E_i ,分别计算出统计指标数据的综合评估值 SA 、总体期望值 SE 和相对差值 SR ,并根据设定的评价等级对统计指标数据总体质量进行评估;对评估出的不合格的统计指标数据进行处理。本发明可以应用在电力行业的电力统计指标数据的质量评估中。



1. 一种电力统计指标数据的质量评估方法,其包括以下步骤:

1) 根据实际评估目的的需要,对电力统计指标数据进行整理,获取待评估的电力统计指标数据;

2) 根据所述待评估的电力统计指标数据,选择相应的数据质量评估指标 I_i ($i=1\cdots n$, n 为数据质量评估指标的个数);

3) 根据待评估的电力统计指标数据和所选取的数据质量评估指标 I_i , 制定各数据质量评估指标相对应的数据质量评估规则 $R_r(I_i)$;

4) 计算各数据质量评估指标 I_i 的权重 W_i 并相应设定期望值 E_i ;

5) 根据步骤 3) 中制定的数据质量评估规则 $R_r(I_i)$, 对待评估的电力统计指标数据进行检测分析, 并计算得出满足各数据质量评估指标 I_i 的数据百分比 S_i , S_i 介于 0 到 100 之间;

6) 根据步骤 4) 计算得到的权重 W_i 、合格百分比 S_i 和设定的期望值 E_i , 分别计算出所述待评估的电力统计指标数据的综合评估值 SA、总体期望值 SE 和相对差值 SR, 并根据设定的评价等级对所述待评估的电力统计指标数据总体质量进行评估;

7) 对评估出的不合格电力统计指标数据进行处理。

2. 如权利要求 1 所述的一种电力统计指标数据的质量评估方法,其特征在於:还包括步骤 8) 将相对差值 SR 与总体期望值 SE 进行比较, 得到待评估的电力统计指标数据 D 总体相对于总体期望值 SE 的质量状况, 具体过程为:若 SR 符号为正, 其数值比总体期望值 SE 越大, 则说明所述待评估的电力统计指标数据总体的数据质量比预期的越好;若 SR 符号为负, 其数值比总体期望值 SE 越大, 则说明所述待评估的电力统计指标数据总体的数据质量比预期的越差。

3. 如权利要求 1 所述的一种电力统计指标数据的质量评估方法,其特征在於:所述步骤 2) 中的数据质量评估指标为正确性、完整性、唯一性、一致性、准确性、有效性和时效性中的某一个或某几个。

4. 如权利要求 2 所述的一种电力统计指标数据的质量评估方法,其特征在於:所述步骤 2) 中的数据质量评估指标为正确性、完整性、唯一性、一致性、准确性、有效性和时效性中的某一个或某几个。

5. 如权利要求 1~4 任一项所述的一种电力统计指标数据的质量评估方法,其特征在於:所述步骤 3) 中每个数据质量评估规则 $R_r(I_i)$ 根据待评估的电力统计指标数据的特征和属性以及所选用的数据质量评估指标 I_i 的定义进行制定。

6. 如权利要求 1~4 任一项所述的一种电力统计指标数据的质量评估方法,其特征在於:所述步骤 4) 中计算各质量评估指标 I_i 的权重 W_i 采用层次分析法,其步骤为:

①根据数据质量评估指标 I_i 的重要性标度关系含义表,通过列表方式确定各数据质量评估指标 I_i 之间的标度关系,进而得到判断矩阵;

②对步骤①得到的判断矩阵中的所有重要性程度值分别进行列归一化处理,得到列归一化矩阵;

③对列归一化矩阵的每一行进行求和运算,得到行相加值;

④对得到所有的行相加值进行求和运算,得到行相加总和值;

⑤计算各数据质量评估指标 I_i 的权重,即权重 = 行相加值 / 行相加总和值。

7. 如权利要求 5 所述的一种电力统计指标数据的质量评估方法,其特征在於:所述步

骤 4) 中计算各质量评估指标 I_i 的权重 W_i 采用层次分析法,其步骤为:

①根据数据质量评估指标 I_i 的重要性标度关系含义表,通过列表方式确定各数据质量评估指标 I_i 之间的标度关系,进而得到判断矩阵;

②对步骤①得到的判断矩阵中的所有重要性程度值分别进行列归一化处理,得到列归一化矩阵;

③对列归一化矩阵的每一行进行求和运算,得到行相加值;

④对得到所有的行相加值进行求和运算,得到行相加总和值;

⑤计算各数据质量评估指标 I_i 的权重,即权重 = 行相加值 / 行相加总和值。

8. 如权利要求 1 ~ 7 任一项所述的一种电力统计指标数据的质量评估方法,其特征在于:所述步骤 6) 中评价等级为:若 $SA \in (95, 100]$ 则评定待评估的电力统计指标数据总体的数据质量水平为“优”,若 $SA \in (90, 95]$ 则评定待评估的电力统计指标数据总体的数据质量水平为“良”,若 $SA \in (85, 90]$ 则评定待评估的电力统计指标数据总体的数据质量水平为“中”,若 $SA \in (0, 85]$ 则评定待评估的电力统计指标数据总体的数据质量水平为“差”。

9. 如权利要求 1 ~ 8 任一项所述的一种电力统计指标数据的质量评估方法,其特征在于:所述步骤 7) 中对评估出的不合格的电力统计指标数据进行处理,其具体处理过程包括:对异常的电力统计指标数据进行修正;对缺失的电力统计指标数据进行填补;对重复的电力统计指标数据进行删除;对表达格式不统一、数值无效的电力统计指标数据进行修正;对数值精度不合格的电力统计指标数据进行修改;对对现有研究可用性不强的电力统计指标数据进行删除。

一种电力统计指标数据的质量评估方法

技术领域

[0001] 本发明涉及一种数据质量评估方法,特别是关于一种适用于电力行业的电力统计指标数据的质量评估方法。

背景技术

[0002] 近年来,随着电力事业的快速发展,电力公司的各个部门都积累了越来越多、越来越复杂的数据,同时对电力统计信息的需求也越来越大,要求越来越高。统计信息在公司决策、经营管理和社会责任承担中发挥的作用也越来越重要。然而,由于电力公司数据采集系统人为和非人为的扰动,会造成各个数据源的数据质量参差不齐,加之数据库本身的某些问题,导致在数据集成时发生数据缺失和错误,从而造成数据整体质量不高,对公司决策规划及发展产生影响,因此对统计数据进行质量评估必将成为公司发展进程中一个必要的环节。

[0003] 现有技术中已经有较为完善的电力统计指标数据体系,但是很少有针对统计指标数据体系中的数据进行合理性、规范性和真实性检测分析的质量评估方法,数据质量评估往往只是零散地针对数据质量统计指标中比较重要的指标,如一致性、唯一性、完整性等来进行,尚未形成系统化的数据质量评估方法。

发明内容

[0004] 针对上述问题,本发明的目的是提供一种电力统计指标数据的质量评估方法,能够合理、规范、真实地对电力统计指标数据质量进行评估。

[0005] 为实现上述目的,本发明采取以下技术方案:一种电力统计指标数据的质量评估方法,其包括以下步骤:1) 根据实际评估目的的需要,对电力统计指标数据进行整理,获取待评估的电力统计指标数据;2) 根据所述待评估的电力统计指标数据,选择相应的数据质量评估指标 I_i ($i=1\cdots n$, n 为数据质量评估指标的个数);3) 根据待评估的电力统计指标数据和所选取的数据质量评估指标 I_i , 制定各数据质量评估指标相对应的数据质量评估规则 $R_r(I_i)$;4) 计算各数据质量评估指标 I_i 的权重 W_i 并相应设定期望值 E_i ;5) 根据步骤 3) 中制定的数据质量评估规则 $R_r(I_i)$, 对待评估的电力统计指标数据进行检测分析,并计算得出满足各数据质量评估指标 I_i 的数据百分比 S_i , S_i 介于 0 到 100 之间;6) 根据步骤 4) 计算得到的权重 W_i 、合格百分比 S_i 和设定的期望值 E_i , 分别计算出所述待评估的电力统计指标数据的综合评估值 SA、总体期望值 SE 和相对差值 SR, 并根据设定的评价等级对所述待评估的电力统计指标数据总体质量进行评估;7) 对评估出的不合格电力统计指标数据进行处理;8) 将相对差值 SR 与总体期望值 SE 进行比较,得到待评估的电力统计指标数据 D 总体相对于总体期望值 SE 的质量状况,具体过程为:若 SR 符号为正,其数值比总体期望值 SE 越大,则说明所述待评估的电力统计指标数据总体的数据质量比预期的越好;若 SR 符号为负,其数值比总体期望值 SE 越大,则说明所述待评估的电力统计指标数据总体的数据质量比预期的越差。

[0006] 所述步骤 2) 中的数据质量评估指标为正确性、完整性、唯一性、一致性、准确性、有效性和时效性中的某一个或某几个。

[0007] 所述步骤 3) 中每个数据质量评估规则 $R_r(I_i)$ 根据待评估的电力统计指标数据的特征和属性以及所选用的数据质量评估指标 I_i 的定义进行制定。

[0008] 所述步骤 4) 中计算各质量评估指标 I_i 的权重 W_i 采用层次分析法, 其步骤为: ①根据数据质量评估指标 I_i 的重要性标度关系含义表, 通过列表方式确定各数据质量评估指标 I_i 之间的标度关系, 进而得到判断矩阵; ②对步骤①得到的判断矩阵中的所有重要性程度值分别进行列归一化处理, 得到列归一化矩阵; ③对列归一化矩阵的每一行进行求和运算, 得到行相加值; ④对得到所有的行相加值进行求和运算, 得到行相加总和值; ⑤计算各数据质量评估指标 I_i 的权重, 即权重 = 行相加值 / 行相加总和值。

[0009] 所述步骤 6) 中评价等级为: 若 $SA \in (95, 100]$ 则评定待评估的电力统计指标数据总体的数据质量水平为“优”, 若 $SA \in (90, 95]$ 则评定待评估的电力统计指标数据总体的数据质量水平为“良”, 若 $SA \in (85, 90]$ 则评定待评估的电力统计指标数据总体的数据质量水平为“中”, 若 $SA \in (0, 85]$ 则评定待评估的电力统计指标数据总体的数据质量水平为“差”。

[0010] 所述步骤 7) 中对评估出的不合格的电力统计指标数据进行处理, 其具体处理过程包括: 对异常的电力统计指标数据进行修正; 对缺失的电力统计指标数据进行填补; 对重复的电力统计指标数据进行删除; 对表达格式不统一、数值无效的电力统计指标数据进行修正; 对数值精度不合格的电力统计指标数据进行修改; 对对现有研究可用性不强的电力统计指标数据进行删除。

[0011] 本发明由于采取以上技术方案, 其具有以下优点: 1、本发明根据获取的待评估的电力统计指标数据, 选择相应的数据质量评估指标, 可以包括正确性、完整性、唯一性、一致性、准确性、有效性和时效性, 并根据待评估的电力统计指标数据和所选取的数据质量评估指标, 制定相应的数据质量评估规则, 通过制定的数据质量评估规则对电力统计指标数据进行检测分析, 计算出电力统计指标数据的综合评估值 SA , 并根据设定的评价等级对电力统计指标数据的质量进行评估, 因此本发明可以对电力行业的统计指标数据进行合理、规范和真实地评估。2、本发明将相对差值 SR 与总体期望值 SE 进行比较, 得到待评估的电力统计指标数据 D 总体相对于总体期望值 SE 的质量状况, 具体过程为: 若 SR 符号为正, 其数值比总体期望值 SE 越大, 则说明所述待评估的电力统计指标数据总体的数据质量比预期的越好; 若 SR 符号为负, 其数值比总体期望值 SE 越大, 则说明所述待评估的电力统计指标数据总体的数据质量比预期的越差, 因此本发明可以提供对待评估电力统计指标数据的整体质量状况进行预测分析, 进一步深度挖掘电力统计指标数据的内在规律, 有效提高公司对电力统计指标数据的应用深度和监督作用, 帮助公司做出科学合理的决策。3、本发明由于可以从评价等级和总体质量状况两个方面对待评估的电力统计数据进行评估, 因此可以帮助电力行业相关公司了解统计指标数据总体的质量水平, 及时发现数据质量问题, 并采取相应的措施来修复数据质量问题, 提高数据质量。本发明可以应用在电力行业的电力统计指标数据的质量评估中。

附图说明

[0012] 图 1 是本发明质量评估方法的流程示意图

具体实施方式

[0013] 下面结合附图和实施例对本发明进行详细的描述。

[0014] 如图 1 所示,本发明的电力统计指标数据的质量评估方法,包括以下步骤:

[0015] 1) 根据实际评估目的的需要,对电力统计指标数据进行整理,获取待评估的电力统计指标数据 D;

[0016] 电力统计指标数据多而复杂,涉及多个业务范围,如:电网运行、设备生产、人力资源和资产动态等方面。由于每次评估的目的不一定相同,因此所选取的电力统计指标数据也不一定相同。电力统计指标数据可以是同一张电力统计报表中的数据,也可以是不同报表中的多个统计指标数据。因此,在评估前需要根据实际评估目的对多而复杂的电力统计指标数据进行整理,获取待评估的电力统计指标数据 D。

[0017] 本发明以某年部分电力统计指标数据为实施例具体说明,如表 1 所示,待评估的电力统计指标数据 D 包括线损率、供电量、线损电量、平均用电负荷率、本网最高发电负荷和全社会用户个数六个电力统计指标数据。

[0018] 表 1 某年部分电力统计指标数据表

[0019]

月份	线损率(%)	供电量(万千瓦时)	线损电(万千瓦时)	平均用电负荷率(%)	本网最高发电负荷(万千瓦)	全社会用户个数(个)
1	6.58	813703	53538.73	85.00	1300.2	760165
2	-2.91	611117	-17785.32	82.00	1243.1	628899
3	11.61	746465	86640.48	87.00	1129	659823
4	0.89	665139	5902.62	89.00	1067.9	659234

[0020]

5	7.33	664401	48674.79	88.00	1023.1	615727
6	7.17	647311	46421.67		1003.7	600887.89
7	21.65	860178	186204.16	88.00	1282	673973
8	8.49	903929	76746.2	88.00	1524.4	827182
9	-2.27	745886	-16905.02	87.47	1310.4	762788.88
10	5.98	941834	58569	87.60	3405	871897
11	7.74	937985	72575	86.25	1568	865410
12	13.75	1016003	139679	85.20	1734.7	876324

[0021] 2) 根据待评估的电力统计指标数据 D, 选择相应的数据质量评估指标 I_i ;

[0022] 由于电力统计指标数据的质量在不同的背景下有不同的定义, 不同的用户对电力统计指标数据的关注点不同, 因此所选取的数据质量评估指标也不同。目前, 电力统计指标数据正呈现爆炸性增长的态势, 且存在着数据异常、数据缺失、属性冗余和格式不规范等问题, 因此, 结合待评估的电力统计指标数据的实际意义, 可以从现有的数据质量评估指标中选取某一个或某几个数据质量评估指标 I_i ($i = 1, 2, \dots, n$, 其中 n 为数据质量评估指标的个数) 对待评估的电力统计指标数据进行评估, 在此不作限制。本发明的实施例选取正确性、完整性、唯一性、一致性、准确性和有效性六个数据质量评估指标对待评估的电力统计指标数据 D 的质量进行评估。正确性用于对待评估的电力统计指标数据 D 的事实符合性和输入正确性进行评估; 完整性用于对待评估的电力统计指标数据 D 是否存在缺失记录或缺失字段进行评估; 唯一性用于对待评估的电力统计指标数据 D 是否存在重复的记录进行评估; 一致性用于对待评估的电力统计指标数据 D 的表达格式是否一致进行评估; 准确性用于对待评估的电力统计指标数据 D 的精度是否准确进行评估; 有效性用于对待评估的电力统计指标数据 D 的格式与数值是否有效进行评估。

[0023] 3) 根据待评估的电力统计指标数据 D 和所选取的数据质量评估指标 I_i , 制定数据质量评估规则 $R_r(I_i)$;

[0024] 根据待评估的电力统计指标数据 D 的特征和属性以及所选取的数据质量评估指标 I_i 的定义制定与数据质量评估指标 I_i 相对应的数据质量评估规则 $R_r(I_i)$ ($i = 1, 2, \dots, n$, 其中 n 为数据质量评估指标的个数)。一般, 正确性的数据质量评估规则指电力统计指标数据无异常值; 完整性的数据质量评估规则指电力统计指标数据无空值; 唯一性的数据质量评估规则指电力统计指标数据唯一、无重复; 一致性的数据质量评估规则指电力统计指标数据的表达形式一致; 准确性的数据质量评估规则指电力统计指标数据的精度统一; 有效性的数据质量评估规则指电力统计指标数据的属性和格式有效。

[0025] 本发明针对上述实施例制定的数据质量评估规则如表 2 所示:

[0026] 表 2 数据质量评估规则表

[0027]

I \ D	线损率	供电量	线损电量	平均用电负荷率	本网最高发电负荷	全社会用户个数
正确性 I1	无基于统计的异常值, 无基于关联规则的异常值					
完整性 I2	非空					
唯一性 I3	唯一					
一致性 I4	以“%”形式表示	——	——	以“%”形式表示	——	——
准确性 I5	精度为 2	精度为 0	精度为 2	精度为 2	精度为 1	精度为 2
有效性 I6	属性值介于 0 到 100, 且格式有效	属性值大于 0, 且格式有效	大于 0, 且表达格式有效	介于 0 到 100, 且格式有效	大于 0, 且格式有效	大于 0, 且格式有效

[0028] 4) 计算数据质量评估指标 I_i 的权重 W_i 并根据实际评估目的的需要设定期望值 E_i ;

[0029] 本发明采用层次分析法计算数据质量评估指标的权重 W_i , 其包括以下步骤:

[0030] ①根据数据质量评估指标 I_i 的重要性标度关系含义表, 通过列表方式确定各数据质量评估指标 I_i 之间的标度关系, 进而得到判断矩阵;

[0031] 标度关系的确定就是将各数据质量评估指标 I_i 进行两两比较, 确定重要性程度, 并对重要性程度按 1 ~ 9 赋值。

[0032] 假设进行两两比较的两个数据质量评估指标为 I_k 和 I_h , 其中 I_k 和 I_h 分别为 I_i ($i=1, 2 \dots n$, 其中 n 为数据质量评估指标的个数) 中的某一个, 数据质量评估指标的重要性标度关系含义表如表 3 所示, 假设标度关系确定后的判断矩阵为 $A=(a_{kh})_{n \times n}$, a_{kh} 表示数据质量评估指标 I_k 与数据质量评估指标 I_h 的重要性相比较的结果, 判断矩阵 A 具有如下性质: $a_{kh} > 0, a_{kh} = 1/a_{hk}, a_{kk} = 1$ 。

[0033] 表 3 数据质量评估指标的重要性标度关系含义表

[0034]

含义	重要性程度
I_k 与 I_h 具有同等重要性	1
I_k 比 I_h 稍微重要	3
I_k 比 I_h 明显重要	5
I_k 比 I_h 强烈重要	7
I_k 比 I_h 极端重要	9
I_k 比 I_h 的重要性程度分别介于上述判断的中间程度	2, 4, 6, 8
I_h 比 I_k 的重要性程度分别与上述对应	1, 1/2, 1/3...1/9

[0035] ②对步骤①得到的判断矩阵中的所有重要性程度值分别进行列归一化处理, 得到

列归一化矩阵；

[0036] ③对列归一化矩阵的每一行进行求和运算，得到行相加值；

[0037] ④对得到的所有行相加值进行求和运算，得到行相加总和值；

[0038] ⑤计算各数据质量评估指标的权重，即权重 = 行相加值 / 行相加总和值。

[0039] 本发明的实施例中，假设数据质量评估指标中正确性 I_1 比完整性 I_2 稍微重要，那么 $a_{12}=3$ ，本发明的实施例形成的判断矩阵如表 4 所示：

[0040] 表 4 判断矩阵

[0041]

评估指标	正确性	完整性	唯一性	一致性	准确性	有效性
正确性	1	3	6	4	4	3
完整性	1/3	1	4	2	2	1
唯一性	1/6	1/4	1	1/3	1/3	1/4
一致性	1/4	1/2	3	1	1	1/2
准确性	1/4	1/2	3	1	1	1/2
有效性	1/3	1	4	2	2	1

[0042] 数据质量评估指标 I_i 的列归一化、行相加值和权重如表 5 所示。同时，根据实际评估目的的需要，人为设定每个数据质量评估指标所期望的结果 E_i (E_i 为介于 0 到 100% 之间的实数)。

[0043] 表 5 数据质量评估指标的列归一化、行相加值、权重表和期望值

[0044]

评估指标	列归一化						行相加值	权重	期望值 $E_i\%$
正确性	0.4286	0.4800	0.2857	0.3871	0.3871	0.4800	2.4485	0.4081	98
完整性	0.1429	0.1600	0.1905	0.1935	0.1935	0.1600	1.0404	0.1734	98
唯一性	0.0714	0.0400	0.0476	0.0323	0.0323	0.0400	0.2636	0.0439	100
一致性	0.1071	0.0800	0.1429	0.0968	0.0968	0.0800	0.6036	0.1006	100
准确性	0.1071	0.0800	0.1429	0.0968	0.0968	0.0800	0.6036	0.1006	98
有效性	0.1429	0.1600	0.1905	0.1935	0.1935	0.1600	1.0404	0.1734	98

[0045] 5) 对待评估的电力统计指标数据 D 进行各数据质量评估指标的检测分析，并计算得出满足数据质量评估指标 I_i 的数据百分比 S_i (S_i 介于 0 到 100% 之间， $i=1, 2, \dots, n$ ，其中 n

为数据质量评估指标的个数)；

[0046] 对待评估的电力统计指标数据 D 进行各数据质量评估指标 I_i 的检测分析时,若待评估的电力统计指标数据 D 满足各数据质量评估指标 I_i 所对应的数据质量评估规则 $R_r(I_i)$,则说明该电力统计指标数 D 全部合格;若有电力统计指标数据 D 不满足某个数据质量评估指标 I_i 下的任意一条数据质量评估规则 $R_r(I_i)$,则说明这个待评估的电力统计指标数据 D 不满足这个数据质量评估指标 I_i ,不满足该数据质量评估指标 I_i 的电力统计指标数据的个数即为不合格数。

[0047] 对待评估的电力统计指标数据 D 进行各数据质量评估指标 I_i 的检测分析的具体过程为：

[0048] (1) 对待评估的电力统计指标数据 D 进行正确性检测；

[0049] 因为电力统计指标数据可能包括单一统计指标数据、有直接逻辑关系的多个统计指标数据和无直接逻辑关系的多个统计指标数中的某一种或者某几种,因此本发明分别对这三种数据情况进行分析,其中：

[0050] ①对单一电力统计指标数据进行正确性检测时直接采用箱线图识别出不合格的数据。

[0051] ②对有直接逻辑关系的电力统计指标数据进行正确性检测时根据电力统计指标数据之间存在的逻辑关系衡量这些电力统计指标数据是否合格,识别出不合格数据。

[0052] ③无直接逻辑关系的多个电力统计指标数据之间虽然没有直接的逻辑关系,但是它们大部分之间会具有一定的相关性,因此,可以通过回归分析,得出它们之间的近似函数关系,通过分析预测值与实际值之间的差距,识别出不合格数据。

[0053] (2) 对待评估的电力统计指标数据 D 进行完整性检测时可以通过对空缺值的检测实现,如果待评估的电力统计指标数据存在空缺值,则认为其不满足完整性指标,空缺值的个数即为不满足完整性指标的个数。

[0054] (3) 对待评估的电力统计指标数据 D 进行唯一性检测时,可以从以下三方面进行分析：

[0055] ①判断电力统计报表中是否存在两个相同的时间变量。

[0056] ②判断电力统计报表中是否存在两个相同的统计标名。

[0057] ③判断电力统计报表中不同行或列对应的数据是否完全相同或相同个数是否超过某一阈值 N (N 根据待评估的电力统计指标数据的个数确定)。

[0058] 若待评估的电力统计指标数据 D 满足上述三项中的任意一项,则认为其存在重复数据或有重复嫌疑。对于初步检测出的重复数据,进行分析判断,最终确定其是否为“真”重复,重复的数据个数即为不满足唯一性指标的数据个数。

[0059] (4) 电力统计指标数据都是数值型数据,采用阿拉伯数字的形式描述,对待评估的电力统计指标数据 D 进行一致性检测时若对其全部数据都进行检测,则会增加不必要的工作量,因此对待评估的电力统计指标数据 D 进行一致性的检测时缩减为对比率类数据进行检测。对于比率类数据,有小数、“%”和“/”三种表达格式,例如,电力统计指标数据线损率可以用 0.9、90% 或 9/10 这三种格式中的任意一种表示。对电力统计指标数据一致性的检测按以下步骤进行：

[0060] ①预先设定一种参考格式；

[0061] ②将待评估的电力统计指标数据 D 与参考格式进行对比分析,考察两者是否一致,若存在差异,则认为待评估的电力统计指标数据 D 的格式不符合要求,不符合要求的数据个数即为不满足一致性指标的数据个数。

[0062] (5) 对待评估的电力统计指标数据 D 进行准确性检测时,主要是考察待评估的电力统计指标数据 D 的精度是否满足要求,其检测原理与一致性指标类似,其检测过程为:

[0063] ①预先定义待评估的电力统计指标数据 D 的参考精度值;

[0064] ②计算每个数据小数点“.”后的字符个数,得到该数据的精度,而对于不存在小数点“.”的数据,则直接将精度置为 0;

[0065] ③考察待评估的电力统计指标数据 D 的精度是否满足预先定义的参考精度值,不满足参考精度值的待评估的电力统计指标数据 D 的个数即为不满足准确性指标的数据个数。

[0066] (6) 对待评估的电力统计指标数据 D 进行有效性检测时,从待评估的电力统计指标数据 D 的格式有效性和数值有效性两方面分析。在对待评估的电力统计指标数据 D 进行格式有效性分析之前,必须先统计出每一个待评估的电力统计指标数据 D 的所有有效格式,然后再将该待评估的电力统计指标数据 D 下的所有数据与其有效格式逐一进行对比,若待评估的电力统计指标数据 D 的表达格式与有效格式相符合,则认为其满足格式有效性这一要求,否则认为该数据不满足格式有效性的要求。数值有效性的分析是判断每个待评估的电力统计指标数据的数值是否介于某一值域范围之内,例如供电量为正数,线损率在 0 到 100% 之间,但是对于整数类数据(例如用户个数),除了分析它的数值大小之外,还必须满足整数这一要求。

[0067] 本发明的实施例中,根据各数据质量评估指标 I_i 所对应的数据质量评估规则 $R_r(I_i)$,对表 1 中的电力统计指标数据 D 进行检测分析得到满足各数据质量评估指标 I_i 的电力统计指标数据的个数,即合格数,并计算出满足各数据质量评估指标 I_i 的电力统计指标数据的百分比 S_i ,即合格百分比,如表 6 所示。

[0068] 表 6 待评估的电力统计指标数据的合格百分比及设定的期望值

[0069]

评估指标 I_i	权重 W_i	合格数	合格百分比 S_i (%)	期望值 E_i %
正确性	0.4081	71	98.6	98
完整性	0.1734	71	98.6	98
唯一性	0.0439	72	100	100
一致性	0.1006	72	100	100
准确性	0.1006	66	91.7	98
有效性	0.1734	68	94.4	98

[0070]

[0071] 6) 根据权重 W_i 、合格百分比 S_i 和设定的期望值 E_i ,分别计算出电力统计指标数据

的综合评估值 SA、总体期望值 SE 和相对差值 SR,并根据设定的评价等级对待评估的电力统计指标数据 D 总体的质量进行评估;

[0072] 综合评估值 SA、总体期望值 SE 和相对差值 SR 分别由以下表达式确定。

$$[0073] \quad SA = \frac{\sum_{i=1}^n W_i \times S_i}{\sum_{i=1}^n W_i}, SE = \frac{\sum_{i=1}^n W_i \times E_i}{\sum_{i=1}^n W_i}, SR = SA - SE$$

[0074] 式中, SA 反映待评估的电力统计指标数据 D 总体的真实数据质量状况, SE 反映对待评估的电力统计指标数据 D 总体的期望值, SR 反映待评估的电力统计指标数据 D 相对于总体期望值 SE 的质量状况, n 为数据质量评估指标的个数。

[0075] 根据综合评估值 SA,对待评估的电力统计指标数据 D 的总体数据质量水平设定评价等级,数据质量评价等级表如表 7 所示。

[0076] 表 7 数据质量评价等级表

[0077]

综合评估值 SA	数据质量评价等级
SA ∈ (95, 100]	优
SA ∈ (90, 95]	良
SA ∈ (85, 90]	中
SA ∈ (0, 85]	差

[0078] 对于待评估的电力统计指标数据 D 的相对差值 SR,若 SR 符号为正,其数值比总体期望值 SE 越大,则说明待评估的电力统计指标数据 D 的数据质量比预期的越好;若 SR 符号为负,其数值比总体期望值 SE 越大,则说明待评估的电力统计指标数据 D 总体的数据质量比预期的越差。

[0079] 本发明的实施例中,根据表 6 的数据可以计算出电力统计指标数据的综合评估值为 97.37988,总体期望值为 98.289,相对差值为 -0.90912,得出如下结论:

[0080] ①待评估的电力统计指标数据的综合评估值为 97.37988,大于 95,因此待评估的电力统计指标数据的质量水平属于“优”。

[0081] ②相对差值为 -0.90912,说明待评估的电力统计指标数据 D 总体的质量水平比预期的要差,但差距不大。

[0082] 7) 对不合格的电力统计指标数据进行处理,以提高待评估的电力统计指标数据 D 的质量,具体处理过程为:对异常的电力统计指标数据进行修正;对缺失的电力统计指标数据进行填补;对重复的电力统计指标数据进行删除;对表达格式不统一、数值无效的电力统计指标数据进行修正;对数值精度不合格的电力统计指标数据进行修改;对对现有研究可用性不强的电力统计指标数据进行删除。

[0083] 上述实施例中,本发明的数据质量评估指标还可以包括时效性,可以对待评估的电力统计指标数据进行时效性检测,因相同的电力统计指标数据不同年份、不同月份的电

力统计数据通常存在着相同的规律,而且同一月份在不同年份的多个电力统计指标数据也会存在着相同的规律,因此,对待评估的电力统计指标数据 D 进行时效性的检测时,可以通过绘制多线线图将同月份的数据在同一张统计图上描绘出来,对比各线图的升降、间距等变化规律,能够较为直观地揭示出数据质量统计指标与时间的关系。由此可以看出,时效性是和时间参数紧密相连的,所以不符合时效性的电力统计指标数据可能没有,也可能是一年或几年的数据,这样出现不符合时效性的电力统计指标数据的可能性就较大。一般,选取待评估的电力统计指标数据时会根据评估目的的需要选取相关年份的电力统计指标数据,因此,如果不是非常关心时效性,就可以不选取该数据质量评估指标,或赋予该数据质量评估指标较小的权重。

[0084] 上述各实施例仅用于说明本发明,其中方法的实施步骤都是可以有所变化的,凡是在本发明技术方案的基础上进行的等同变换和改进,均不应排除在本发明的保护范围之外。

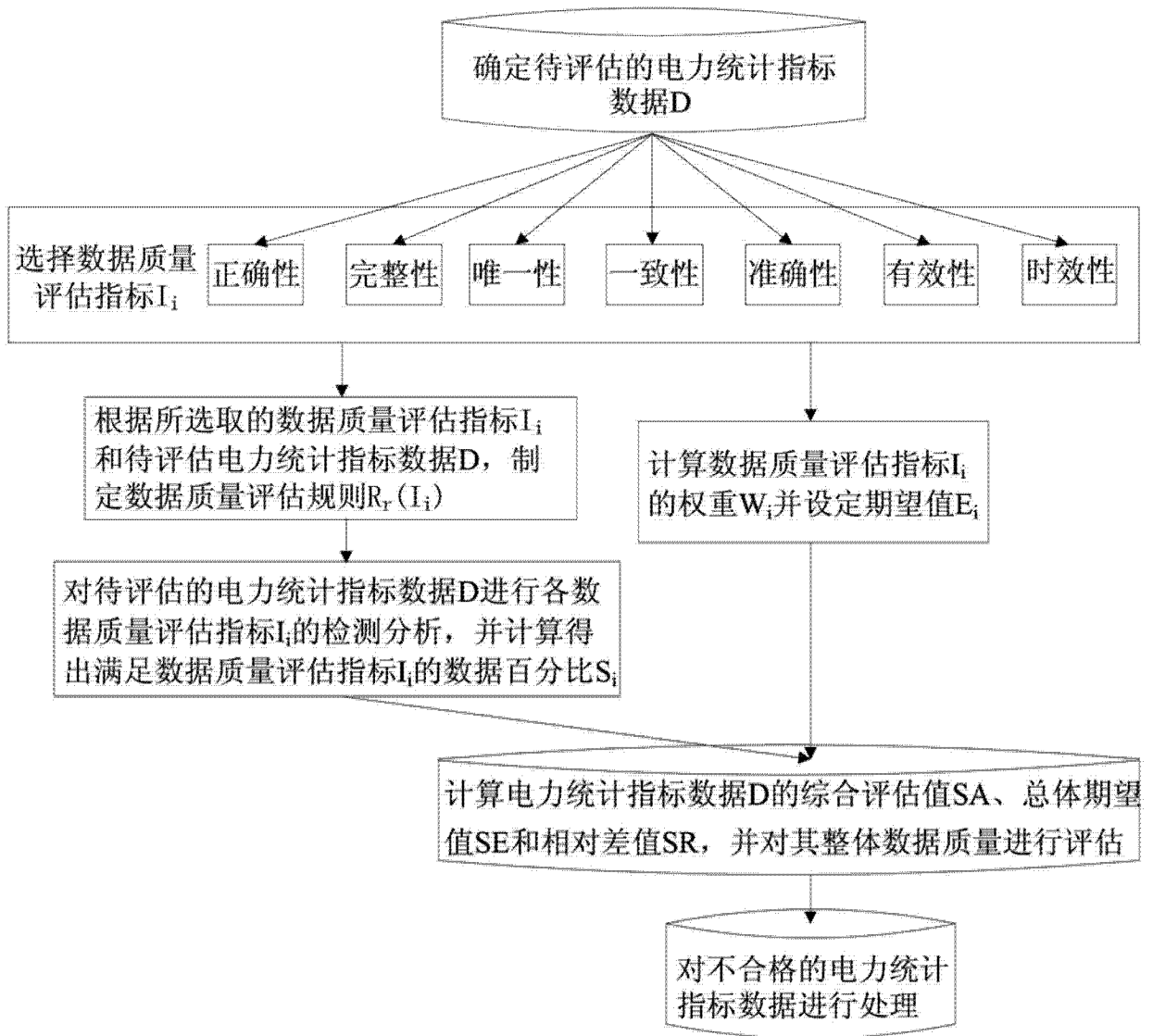


图 1