

基于 LOF 算法的碳排放因子确认模型研究

赵吉祥 刘玉昌 杜永腾 迟立洋

山东电力工程咨询院有限公司

DOI:10.12238/eep.v6i6.1854

[摘要] 本文提出了一种基于LOF算法的碳排放因子确认模型构建方法,该方法包括提取产品生命周期各环节的特征要素,构建各环节特征要素数据集;从各环节特征要素数据集中调取一个领域内所有环节特征要素,计算当前领域内所有环节特征要素的局部异常因子,判断各个环节特征要素是否属于异常数据;如果某一环节特征要素在当前领域属于异常数据,判断该环节是否存在其他领域中,若是,则计算其他领域内所有环节特征要素的局部异常因子;如果该环节不属于任何领域,则新建该环节所属的领域;合并领域内所有环节特征要素,形成领域所属的特征要素数据集,完成碳排放因子确认模型构建。该模型能够有效地识别和排除碳排放数据中的异常值,提高碳排放数据的准确性和可靠性,为碳排放权交易市场和其他相关政策提供科学依据。

[关键词] LOF算法; 排放因子法; 数据异常诊断; 碳排放因子确认模型

中图分类号: X-652 **文献标识码:** A

Research on the construction method of carbon emission factor confirmation model based on LOF algorithm

Jixiang Zhao Yuchang Liu Yongteng Du Liyang Chi

Shandong Electric Power Engineering Consulting Institute Corp., Ltd

[Abstract] This paper proposes a construction method for the Carbon Emission Factor Confirmation model based on LOF (Local Outlier Factor) algorithm. The method includes extracting feature elements from each stage of the product life cycle and building a dataset of these feature elements. From the dataset, a set of feature elements from each stage within a domain is selected and the local outlier factor of all feature elements in the domain is calculated. The feature elements are then identified as normal or abnormal based on this calculation. If a feature element from a stage within the domain is identified as abnormal, the system checks if this stage exists in other domains. If it does, the local outlier factor of all feature elements in those other domains is calculated. If the stage does not exist in any domain, a new domain is created for it. Finally, all feature elements within a domain are combined to form a dataset for the domain, completing the construction of the Carbon Emission Factor Confirmation model. This model can effectively identify and exclude abnormal values in carbon emission data, improving the accuracy and reliability of carbon emission data and providing scientific evidence for carbon emissions trading markets and other relevant policies.

[Key words] LOF algorithm; emission factor method; data anomaly diagnosis; carbon emission factor confirmation model

引言

在碳排放核算和报告领域,许多国内外计算产品碳排放量的方法仍以碳排放因子为主。但这种方法存在以下一些问题:许多行业的碳排放数据主要来源于国家和地区统计系统的报告和碳排放交易所的监测数据,缺乏全面、准确、可靠的碳排放因子数据,这可能导致碳排放数据的误导和不准确;碳排放因子的计算需要进行大量数据的处理和分析,包括活动数据的收集、处理、分析以及排放因子的计算等,但是许多行业缺乏相应的数据

处理和分析技术,无法准确计算碳排放因子;同种燃料由于应用于产品各环节的特征影响因素不同导致碳排放因子的多样性;不同地区和行业采用的碳排放因子计算方法和标准不统一,这导致不同地区和行业之间的碳排放数据无法进行比较和分析。

1 具体实现

本文旨在分析现有碳排放数据的来源和可靠性问题,探讨如何获取全面、准确、可靠的碳排放因子数据,研究碳排放因子的计算方法和流程,针对不同行业的特点提出相应的数据处理

和分析技术。通过提出基于LOF算法的碳排放因子确认模型构建方法，能够为碳排放核算和报告领域提供更加准确可靠的数据计算和分析，促进全球碳排放管理和减排工作的有效开展。主要研究内容如下图所示：

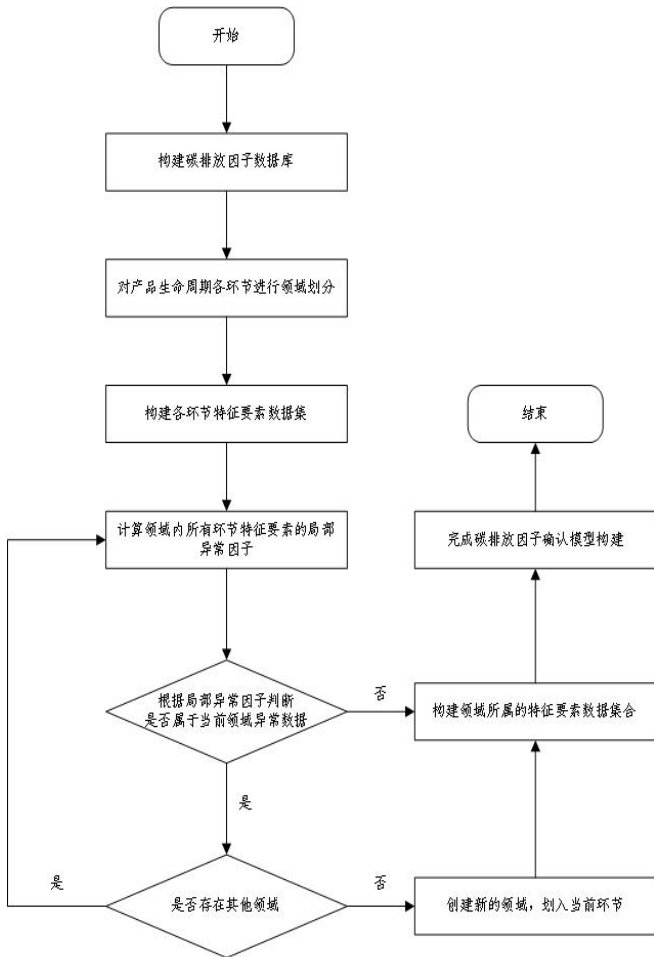


图1 基于LOF算法的碳排放因子确认模型构建方法流程

1.1 领域划分

从碳排放因子库中调取产品生命周期各个环节的碳排放因子数值，将碳排放因子值相等的环节划归为同一领域。其中，所述产品生命周期各环节包括产品原料开采、运输、制造、使用和废弃全生命周期的各个环节。

例如，碳排放因子数据库中的碳排放因子

$$EF = \{ef_1, ef_2, \dots, ef_n\}$$

其中，代表不同碳排放因子值的数量。

煤制品-焦炭-燃烧：碳排放因子值1.92；

石油及石油制品-天然气凝析液-钻井：碳排放因子值1.92；

按照碳排放因子的数值进行分类，碳排放因子数值相等的各产品生命周期环节划分为同一领域。

煤制品-焦炭-燃烧和石油及石油制品-天然气凝析液-钻井的碳排放因子值相同，划分至同一个领域。

$$\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$$

其中， n 代表当前碳排放因子值包含的领域数量。

$$\alpha_i = \{lnk_1, lnk_2, \dots, lnk_n\}$$

其中， lnk_n 代表领域内多个产品生命周期环节。

1.2 构建特征要素数据集

从预设产品生命周期各环节碳排放数据库中，提取产品生命周期各环节 lnk_i 的特征要素 p_i ，构建各环节特征要素数据集：

其中，每个特征要素由特征要素属性以及特征要素属性的数值构成。

例如，煤制品-焦炭-燃烧的特征要素：剧烈性、发光性、放热性等。

石油及石油制品-天然气凝析液-钻井的特征要素：地质条件、钻井液类型、设备效率、燃烧剧烈性、燃烧发光性、燃烧放热性等。

$$lnk_i = \{p_1, p_2, \dots, p_n\}$$

其中， n 代表特征要素的数量。

每个特征因素 p_i 包括两方面内容：

(1) 特征记录 N ，表示在任意时刻所有特征要素属性集合，公式如下：

$$N = \{n_1, n_2, \dots, n_n\}$$

其中 $n_i, i = 1, 2, \dots, n$ 表示特征要素属性。

例如，煤制品-焦炭-燃烧的特征要素剧烈性的 $N = \{\text{热量, 强度, 分解}\}$ ；

煤制品-焦炭-燃烧的特征要素放热性的 $N = \{\text{热量, 速度, 效率}\}$ ；

(2) 特征要素数据记录 v_t ，表示时刻 t 当前特征要素包含所有 n 个特征要素属性的数值记录，公式如下：

$$V_t = \{v_1, v_2, \dots, v_n\}$$

其中 $v_i, i = 1, 2, \dots, n$ 表示特征要素属性数值。

煤制品-焦炭-燃烧的特征要素放热性的数值：

$N = \{\text{热量} = \{12.12, 13.43, 43.12, 32.34\}, \text{速度} = \{1.25, 8.35, 6.24, 9.22\}, \text{效率} = \{15.36, 35.36, 63.12, 66.14\}\}$ 。

1.3 特征要素异常判断

从各环节特征要素数据集中调取一个领域内所有环节特征要素，基于LOF算法计算当前领域内所有环节特征要素的局部异常因子，进而判断各个环节特征要素是否属于异常数据；

根据距离度量来对一个领域内所有环节特征要素进行 K 邻近排序，计算各个环节特征要素在当前领域内的平均局部可达密。

(1) 对于每个特征要素 p_i ，找出其 k 个最近邻点，通过对所有数据点进行排序，选择距离最小的 k 个点。对于数据集中任意两个数据点 p 和 q ，计算可达距离 $D_r(p, q)$ ，公式如下：

$$D_r(p, q) = \max \{d(p, q), D_k(q)\}$$

其中， $d(p, q)$ 代表 p 和 q 之间的距离；

$D_k(q)$ 代表特征要素点 q 的 k -邻近距离。

计算领域 a 内局部可达密度 $LRD_\alpha(p)$, 其局部可达密度指的是 p 到其邻域内所有点的平均可达距离的倒数。公式如下:

$$LRD_\alpha(p) = \frac{1}{\sum_{q \in N_\alpha(p)} \frac{D_r(p, q)}{|N_\alpha(p)|}}$$

其中 $|N_\alpha(p)|$ 为 p 在 a 领域内数据值点的个数, 即数据点(即特征要素) p 的局部可达密度。

(2) 将任一环节特征要素在当前领域内的平均局部可达密度与该环节特征要素的局部可达密度求比值, 计算局部异常因子, 为的领域内的平均局部可达密度跟数据点(即特征要素)的局部可达密度的比值。公式如下:

$$LOF_\alpha(p) = \frac{\sum_{q \in N_\alpha(p)} \frac{LRD(q)}{LRD(p)}}{|N_\alpha(p)|}$$

(3) 根据局部异常因子与预设阈值比较来判断各个环节特征要素是否属于异常数据。

当 $LOF_\alpha(p) \leq \theta_\alpha$, 说明特征要素 p 归属于领域 α_i ; 当 $LOF_\alpha(p) > \theta_\alpha$, 说明特征要素 p 不应归属于领域 α_i ; θ_α 为设定阈值。

(4) 如果某一环节特征要素在当前领域属于异常数据, 判断该环节是否存在其他领域中, 若是, 则计算其他领域内所有环节特征要素的局部异常因子; 如果该环节不属于任何领域, 则新建该环节所属的领域;

计算领域 α_{i+1} 的局部异常因子, 直至出现下列两种情况之一:

- (1) $LOF_{\alpha+1}(p) \leq \theta_{\alpha+1}$, 则 $p \in \alpha+1$ 领域;
- (2) $LOF_{\alpha+1}(p) > \theta_{\alpha+1}$, 新增领域 $\alpha+2$, 则 $p \in \alpha+2$

领域。

(3) 合并领域内所有环节特征要素, 形成领域所属的特征要素数据集, 完成碳排放因子确认模型构建。

形成碳排放因子 EF 的领域集合: $EF_n^\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 。

其中, 碳排放因子确认模型为:

$$\alpha_n^p = p_1 + p_2, \dots + p_n$$

$$\alpha_n^{ef} = ef_n$$

其中, α_n^p 代表当前领域的碳排放因子特征要素;

p_1, p_2, \dots, p_n 代表特征要素;

α_n^{ef} 代表当前领域的碳排放因子数值。

2 结束语

本文通过将碳排放因子值相等的环节预先划归为同一领域, 再基于LOF算法计算当前领域内所有环节特征要素的局部异常因子, 来判断各个环节特征要素是否属于异常数据, 采用数据复杂分布异常检测技术, 合并和简化碳排放因子的所属特征要素, 形成了领域所属的特征要素数据集, 完成了碳排放因子确认模型构建, 提高了产品碳排放因子匹配度, 能够协助企业提升产品碳足迹精确度, 达到降低企业碳排放成本的目标。

[参考文献]

[1]董丝雨,刘温馨.贯彻新发展理念助力碳达峰碳中和:2023碳达峰碳中和绿色发展论坛综述[N].人民日报,2023-04-11(10).

[2]IPCC. Climate change 2013:The physical science basis: Working Group I contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change[M].New York,USA:Cambridge University Press,2014.

[3]BUDIYANTO M A,HUZAIFI M H,SIRAIT S J,et al.Evaluation of CO2 Emissions and Energy Use with Different Container Terminal Layouts[J].Scientific Reports,2021,11(1):5476-5476.

作者简介:

赵吉祥(1982--),男,汉族,山东省新泰市人,高级工程师,研究方向:综合智慧能源。