基于数据包络分析的环境保护支出绩效评价软件V1.0

说明书

# 软件描述

环境保护支出的绩效评价是一个“多投入、多产出”的问题：环境保护机构在不同项目上投入不同数量的资金，环境保护成果也由多个指标（不同的有害气体排放量、不同的垃圾无害化处理量）构成，而没有一个固定指标（例如商业上的利润率）能够评价环境保护的绩效。

《基于数据包络分析的环境保护支出绩效评价软件》是一个Web应用软件，通过数学最优化模型和统计预测模型，评价和预测环境保护机构的财政绩效。本软件通过分析多个投入和产出指标的数据，使用数据包络分析算法建立绩效指标，评价决策单元[[1]](#footnote-1)的绩效；使用BHT-ARIMA算法实现短时间序列预测，预测环境保护机构下一年的绩效。

本软件能够帮助非统计学专业的用户快速实现数据包络分析算法和AAAI 2020学术会议提出的BHT-ARIMA算法，使用当前最优的模型计算环境保护支出的绩效，使用便捷且结果科学。

# 软件功能

软件分析数据的过程分为绩效评价和绩效预测：第一步，用户上传数据集，设置投入和产出变量，软件使用数据包络分析算法建立绩效评价指标。第二步，软件从数据集中提取年份和环境保护机构名称，根据计算得到的绩效，使用BHT-ARIMA算法预测每个环境保护机构在下一年的绩效。

## 绩效评价

本软件使用数据包络分析算法分析环境保护机构的绩效。输入数据是1个二维表格，其中每行代表1个样本，是一个环境保护机构在某一年的数据。每列代表1个原始指标，用户需指定一个或多个投入指标（例如环境保护支出金额）、一个或多个产出指标（例如工业二氧化硫去除量、工业烟尘去除量、生活垃圾无害化处理率）。设共有个样本，某个样本的投入指标的取值是，产出指标的取值是，样本的绩效是，则的值是优化模型的结果，如公式所示。

模型线性化：利用性质，该模型可以被转化为公式.

其中是模型的内部参数，代表每个样本上各投入和产出指标的权重。根据本模型的线性规划结果，可以求得每个样本的绩效，绩效值在0~1之间，与投入和产出指标的量纲无关。

## 绩效预测

本文使用BHT-ARIMA算法[[2]](#footnote-2)实现绩效预测，该算法适合预测短时间序列的下一时刻值。本软件根据数据集中，用户设定的时间和决策单元变量，从数据集中提取环境保护机构历年绩效的时间序列，每个环境保护机构对应一个时间序列，长度可以是不同的。BHT-ARIMA算法需要用户输入4个参数：ARIMA模型的历史项数p、移动平均长度q、差分阶数d，BHT-ARIMA算法在MDT过程中的时间滞后参数. 详细的算法内容参考脚注所列的学术论文。

在绩效预测中，软件输入每个环境保护机构的历史绩效序列，软件预测下一年的绩效，预测流程如图 1所示。



图 1 绩效预测流程图（示意图）

# 软件运行环境

|  |  |
| --- | --- |
| 操作系统 | 服务端：Ubuntu 20.04 LTS / Windows Server 2016  客户端：可联网操作系统不限 |
| 软件开发工具 | JetBrains PyCharm 2021.2  MobaXTerm Personal 20.6 |
| 依赖软件 | Python 3.9和源码中列出的函数库  JS Deliver  Google Chrome / Microsoft Edge 89以上版本浏览器 |
| 开发语言 | Python 3.9 |

# 软件特点

## 本软件的引导流程易于使用

传统地，数据包络分析算法和BHT-ARIMA算法需要有统计背景的分析人员，建立优化模型并求解才能实现。本软件自动化实现了这个流程，并且可以推广到其他绩效评价场景中，使得非专业用户也能根据本软件的引导，完成分析。软件具有网页形式的图形界面，每一步作为一个页面，绘制了作为分析依据的统计图，提示用户如何设定最优的参数，如何解读分析结果。另外，软件也具有一定的自由度，用户可以根据偏好调整参数。

## 本软件使用先进的数学模型

本软件使用的BHT-ARIMA算法是AAAI 2020学术会议的研究成果，短时间序列预测的准确度处于领先地位，使得本软件可以更准确地实现绩效预测。本软件使用的数据包络算法用成熟的函数库实现，运行结果稳定。

## 本软件易于部署和维护

本软件以云计算为基础，不需要用户安装软件，所需部署和维护全部是服务端的工作。软件使用Django函数库实现自动化管理数据库和网站：

数据库方面，本软件源码中已包含自动生成SQLite3数据库的脚本，不需要人工创建数据库。网络传输方面，本软件自身提供解析静态文件和前后端交互的程序，不需要特定的运行环境。算法方面，本软件由模块化算法组装，容易添加或删除组件，方便用户修改功能或仅安装选定的功能。

软件内部也有清晰的模块化结构，主要使用自动化脚本，方便在尽量不中断服务的情况下更新升级。

# 软件设计与实现

## 软件设计

本软件使用Django函数库解析网络请求，对用户的每个操作，基本流程如下：

1. 软件收到用户的HTTPS请求，与路由表匹配，根据请求的路径分配到不同的视图函数执行。
2. 视图函数大致分为5个部分：验证输入的格式是否正确，从数据库获取运算所需数据，执行计算过程，将数据写入数据库，将人类可读的结果返回到上下文字典（本质是一个Python字典类型的变量）。
3. HTML渲染器读取静态网页模板，将上下文字典中的内容渲染到网页上，作为HTTPS的返回请求发送。

每个视图函数关联一种操作，操作数据库的一个或多个数据表，是实现软件功能的基本单位。

基于数据包络分析的环境保护支出绩效评价软件运行后，用户通过浏览器访问<https://example.com/main>地址（根据部署地址确定），可以与软件交互：

1. 提供登录组件，供输入用户名和密码登录；提供“注册”页面，供登记用户名和密码，软件发送确认注册邮件，用户点击邮件上的链接注册。
2. 在首页上放置“采集数据”和“分析数据”按钮，“采集数据”的下拉列表包含了样例数据和数据预处理方法。“分析数据”按钮链接到任务列表，用户可查看分析任务的列表，执行删除和访问操作，查看错误日志。
3. 用户从“采集数据”按钮的下拉列表中选择下载样例数据，可以先分析软件内置的样例数据，熟悉软件使用方法。
4. 专业的用户从“采集数据”按钮的下拉列表中，选择CSMAR数据库的预处理方法，可以了解从CSMAR数据库中获得地区财政的环境保护支出的数据，并且按照网页所示的预处理程序自行处理数据，使得数据符合本软件要求的格式。
5. 用户点击“分析数据”按钮，并在任务列表点击“新建”按钮新建一个任务，进入引导流程的第一步；如果用户中途退出任务，再次进入时，点击任务名称，回到已完成的最后一个步骤。
6. 新建完任务后，用户指定时间、机构名称、投入、产出变量的名字，软件执行数据包络分析算法，并计算环境保护机构在统计时段每年的绩效。
7. 完成绩效评价后，用户可以选择一个环境保护机构，查看绩效变化的折线图，并且下载全部数据。
8. 完成绩效评价后，用户可以设置BHT-ARIMA模型的参数，软件执行BHT-ARIMA算法，预测环境保护机构下一年的绩效。
9. 用户可以在网页上预览绩效的预测值，并且下载结果。
10. 用户点击导航栏上的“我的分析”回到任务列表，点击Clixove徽标回到软件首页。

## 工作模块

本软件的工作模块如图 2所示。

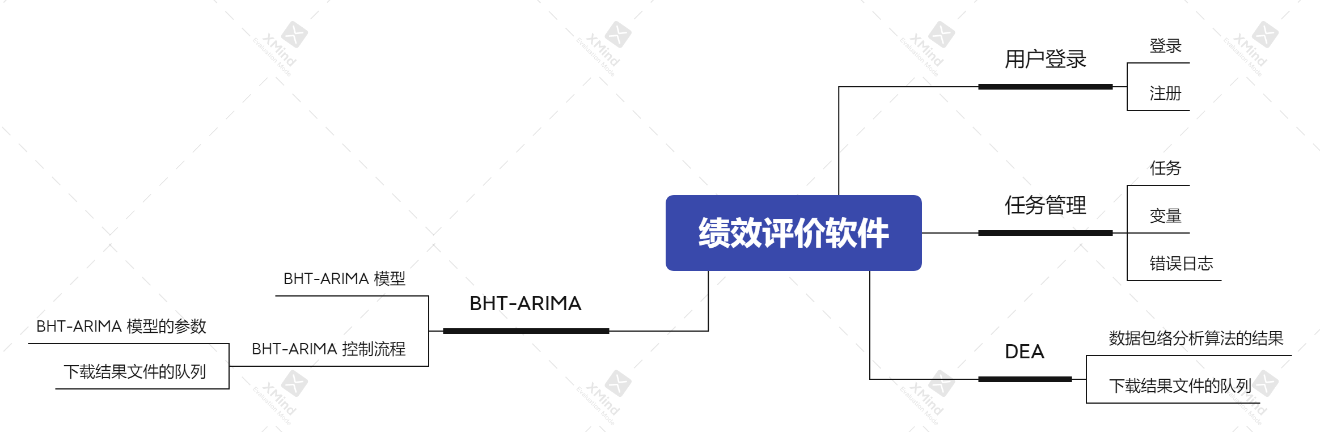


图 2 工作模块

**用户登录：**本模块负责用户的登录和注册功能，继承自Django函数库的Authorize模块，具有向用户发送注册确认邮件的功能。

**任务管理：**本模块负责管理所有的分析任务，每个任务具有独立的数据、模型、结果；本模块管理每个分析任务的繁忙状态，收集错误提示，并保持数据集及相关信息。本模块在分析流程中，具有创建任务、接收数据集、解析变量的功能。

**DEA：**本模块负责数据包络分析算法，负责从任务中读取数据，然后运行数据包络分析算法，并保存结果。数据库中有结果和下载文件队列，2个数据表，其中结果文件以二进制变量形式保存了绩效评价的结果，以JSON形式保存了决策单元的列表；下载文件队列是下载文件的缓存，当用户点击“下载全部数据”按钮时，软件将绩效评价的结果生成为Microsoft Excel表格，保存为文件，然后给用户提供文件下载链接。

**BHT-ARIMA：**本模块包含了BHT-ARIMA算法的源代码（**不包括为本软件的源码**，但属于本软件的运行依赖项），和本软件对该算法的运行控制模块。本软件负责保存用户对BHT-ARIMA算法的参数设置，处理数据并输入算法，然后运行该算法和整理结果。数据库中有参数表和下载文件的缓存，其中参数表保存了用户设置的模型的参数，以二进制变量的形式保存了预测结果。当用户访问预测结果页面时，软件自动解析预测结果，生成HTML文本并渲染在网页上；当用户点击“下载结果”按钮时，软件从预测结果生成Microsoft Excel表格，保存为文件，然后给用户提供文件下载链接。

## 主要源码文件的作用

**manage.py**: 软件运行的主函数，是启动和管理网络服务的主体。

**govt\_env\_protection\_eval\_dj/urls.py**: 路由表，负责分流网页请求。

**govt\_env\_protection\_eval\_dj/asgi.py**:由Django函数库自动生成，包含在ASGI服务器上运行的兼容性程序源码。

**govt\_env\_protection\_eval\_dj/wsgi.py**: 由Django函数库自动生成，包含在WSGI服务器上运行的兼容性程序源码。

**govt\_env\_protection\_eval\_dj/settings.py**: 全局设置，包含时区、语言、静态文件存储路径、媒体文件存储路径、密码规则、软件部署密码、模块列表等所有全局变量。

**templates**: 静态网页模板，包含软件界面的静态部分。动态部分由视图函数返回的上下文字典渲染。

**BHT\_ARIMA**: BHT-ARIMA算法的程序源码，**不包括为本软件的源码**，但属于本软件的运行依赖项。

其余的my\_login, task\_manager, DEA, BHT\_ARIMA\_RUNTIME结构相似，分别对应用户登录、任务管理、DEA、BHT-ARIMA模块，内部文件的功能是：

**migrations**: 数据库迁移记录，用以比较源码和数据库的差异；如果因为版本更新，软件检测到源码被修改，会自动和migrations文件夹中的记录比较，确认自身的版本位置。如果管理员执行更新，软件会按序号生成新的数据库迁移记录，使得源码定义的数据库结构和事实上的数据库结构一致。

**views.py**: 包含了视图函数，函数名称链接到路由表。

**models.py**: 定义了数据库结构，包括表名、字段名、字段类型、限制条件（例如文本字段的最大长度，整数字段的最大和最小值，文件储存地址，别名，日期字段的自动写入现在时间）。

**admin.py**: 数据库表格在管理员后台的显示格式，包括显示哪些字段，修改时使用哪种表单结构，是否分组筛选等。

**apps.py**: 供govt\_env\_protection\_eval\_dj/settings.py全局设置中，模块列表读取模块的入口。

# 软件使用说明

软件登录界面如图 3所示。



图 3 登录页面

## 管理员安装过程

（1）创建部署令牌

管理员获得程序源码后，在软件根目录下创建token文件夹，在文件夹中创建如下文件：

* django\_secret\_key: 文本文件，写入一个52位的随机字符串（包含英文大小写和数字），作为网络通信密钥。
* smtp.json: JSON文件（如果需要使用邮件注册功能），填入发送注册邮件的SMTP邮件服务信息，格式为：

{

"host": "example.com",

"port": 465,

"username": "registration@example.com",

"password": "anypassword"

}

（2）创建Python环境

pip install -r requirements.txt

导航到本软件的根目录，创建数据库和超级用户：

python manage.py migrate

python manage.py createsuperuser

按照命令行提示填写信息；超级用户拥有访问本软件数据库的最高权限。

（3）创建静态文件

在govt\_env\_protection\_eval\_dj/settings.py文件中执行替换：

# 将

STATICFILES\_DIRS = ['templates/static']

# 替换为

STATIC\_ROOT = 'templates/static'

然后运行：

python manage.py collectstatic

反向执行govt\_env\_protection\_eval\_dj/settings.py文件的替换操作，然后将文件中的DEBUG=True调整为DEBUG=False

（4）管理员设置

运行：

python manage.py 0.0.0.0:$port –insecure

其中8000是端口号，可以自定义；但是IP地址必须填写0.0.0.0或127.0.0.1，不能修改为其它值。如果软件是联网供用户使用的，IP地址必须填写0.0.0.0

管理员后台的首页如图 4所示。导航栏的右侧是返回软件前端和修改密码的按钮；页面主体部分的左栏是数据库中的每个数据表（在每个模块的admin.py文件中注册的），带有底纹的是模块名称，模块名称下的是数据表名；页面主体部分的右栏是当前管理员账号的操作记录。

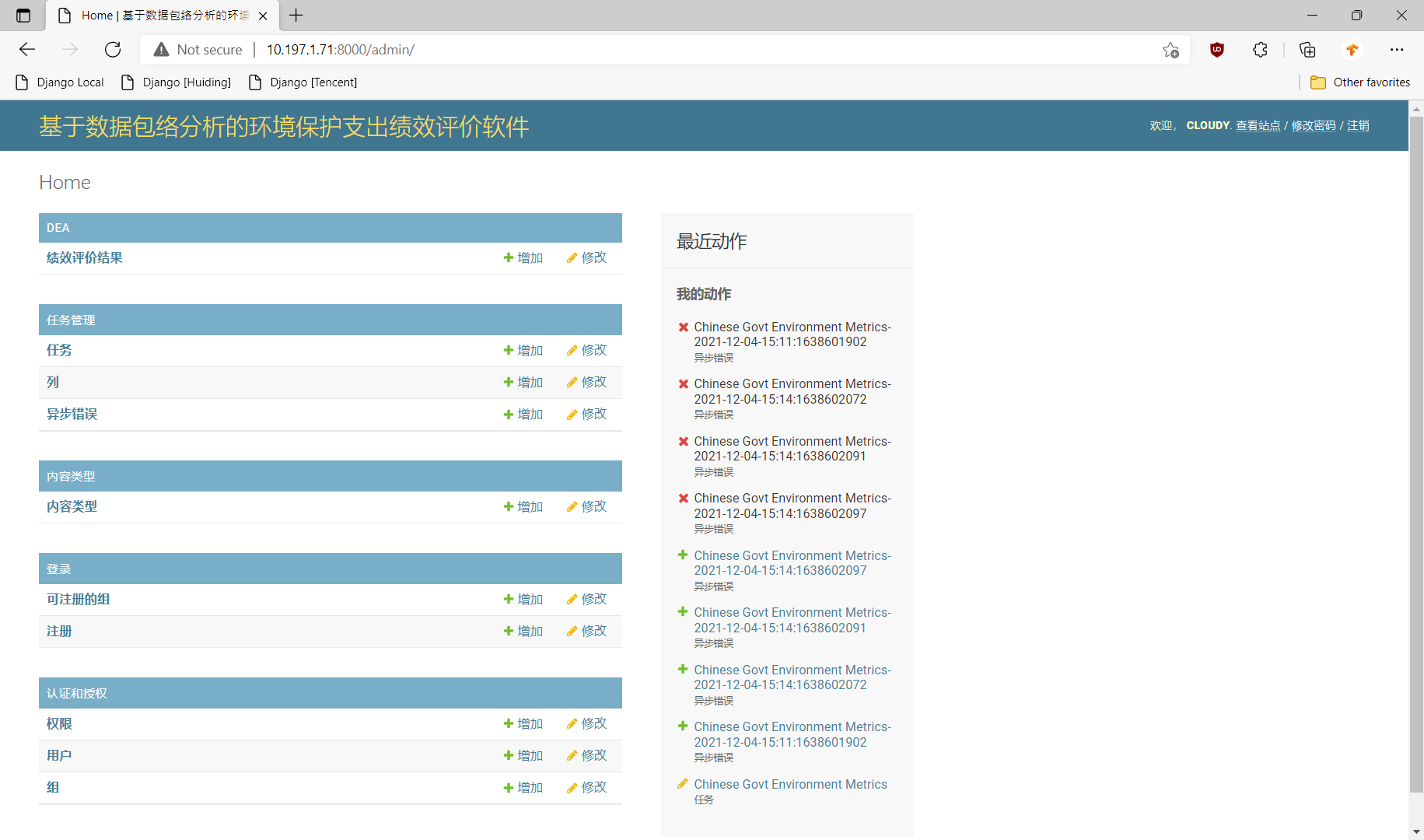


图 4 管理员的主页

在Groups表中添加用户组，为用户组设置权限，用户组的权限在Permissions表中预览。在Register groups中添加可自由注册的用户组（一般来说，添加一个用户注册时默认赋予的用户组）。

完成操作后，退出管理员账号，完成部署。

## 用户使用过程

（1）软件的主页如图 5所示。页面分为导航栏和主体部分，导航栏上的Clixove徽标链接到首页，“我的分析”链接到分析任务的列表，点击“登入”弹出登录表单。



图 5 软件主页

（2）用户首次访问软件，在导航栏的右侧点击“登录”按钮；如果没有账号，点击“不是用户？注册”按钮，进入注册页面，如图 6所示。

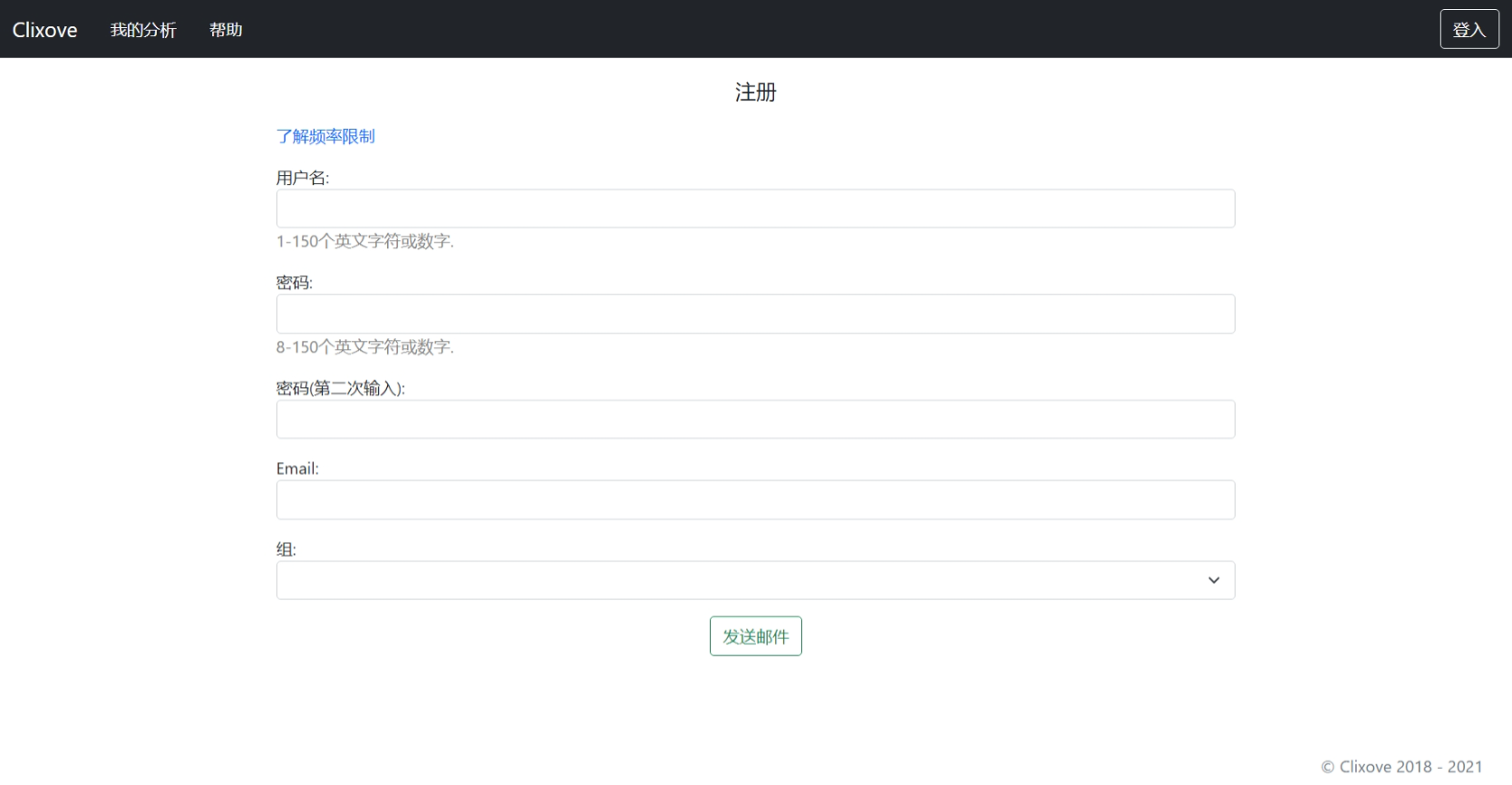


图 6 用户的注册页面

（3） 点击“发送邮件”按钮，软件向用户发送一封验证邮件，同时跳转到激活页面，如图 7所示。用户在收到的邮件中看到邀请码，需在激活页面填写，提交后用户注册成功，跳转到软件主页。



图 7 用户的激活页面

（4）在软件主页点击“采集数据”按钮，软件弹出预处理程序的介绍页面，如图 8所示。

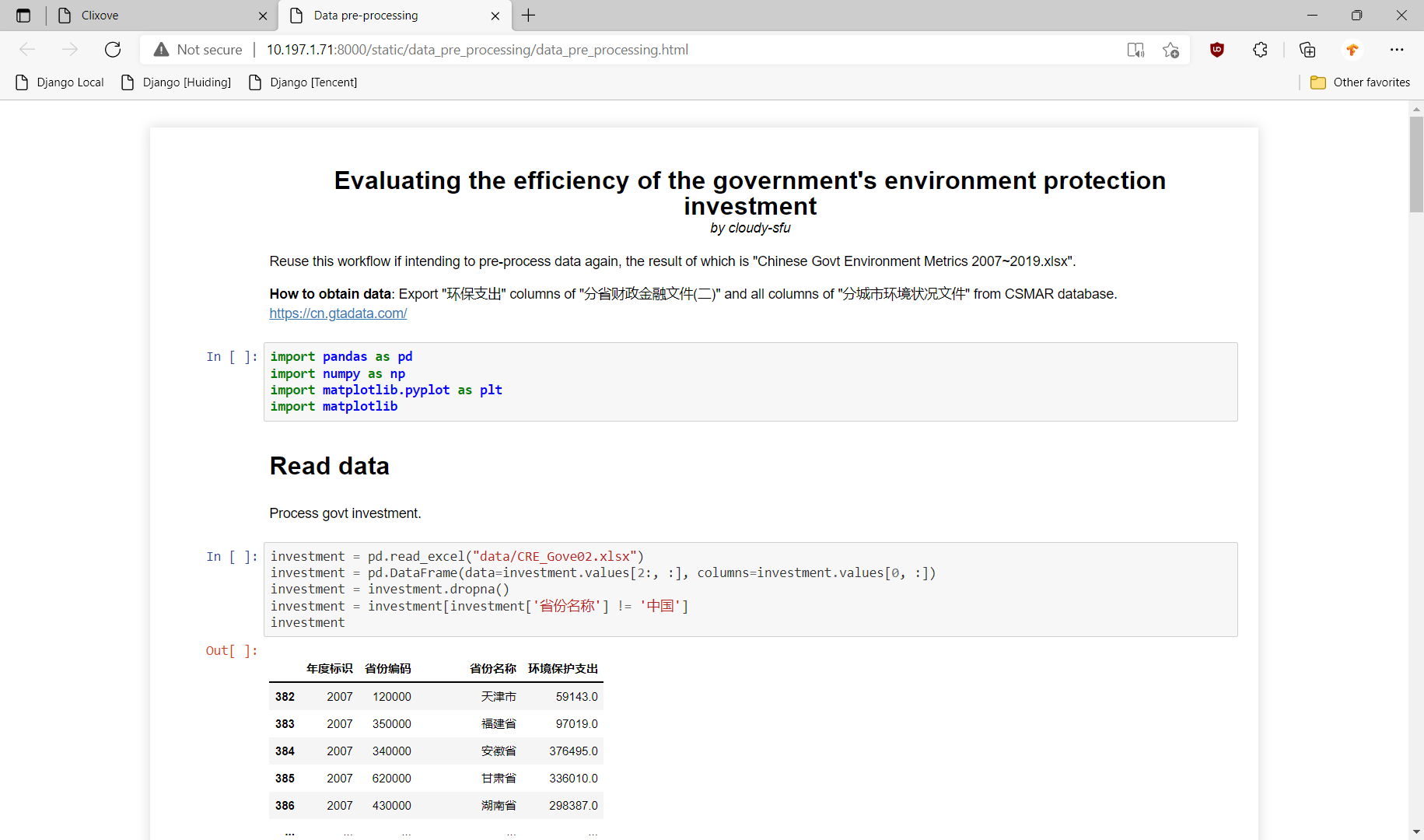


图 8 预处理程序的介绍

（5）点击“分析数据”，进入分析任务的列表，如图 9所示。



图 9 任务列表页面

（6）点击“新建分析”按钮，新建一个任务，如图 10所示。用户填写任务名称，并上传数据集。数据集格式参考软件主页“采集数据”部分，点击“下载样例数据”按钮可获得样例数据。

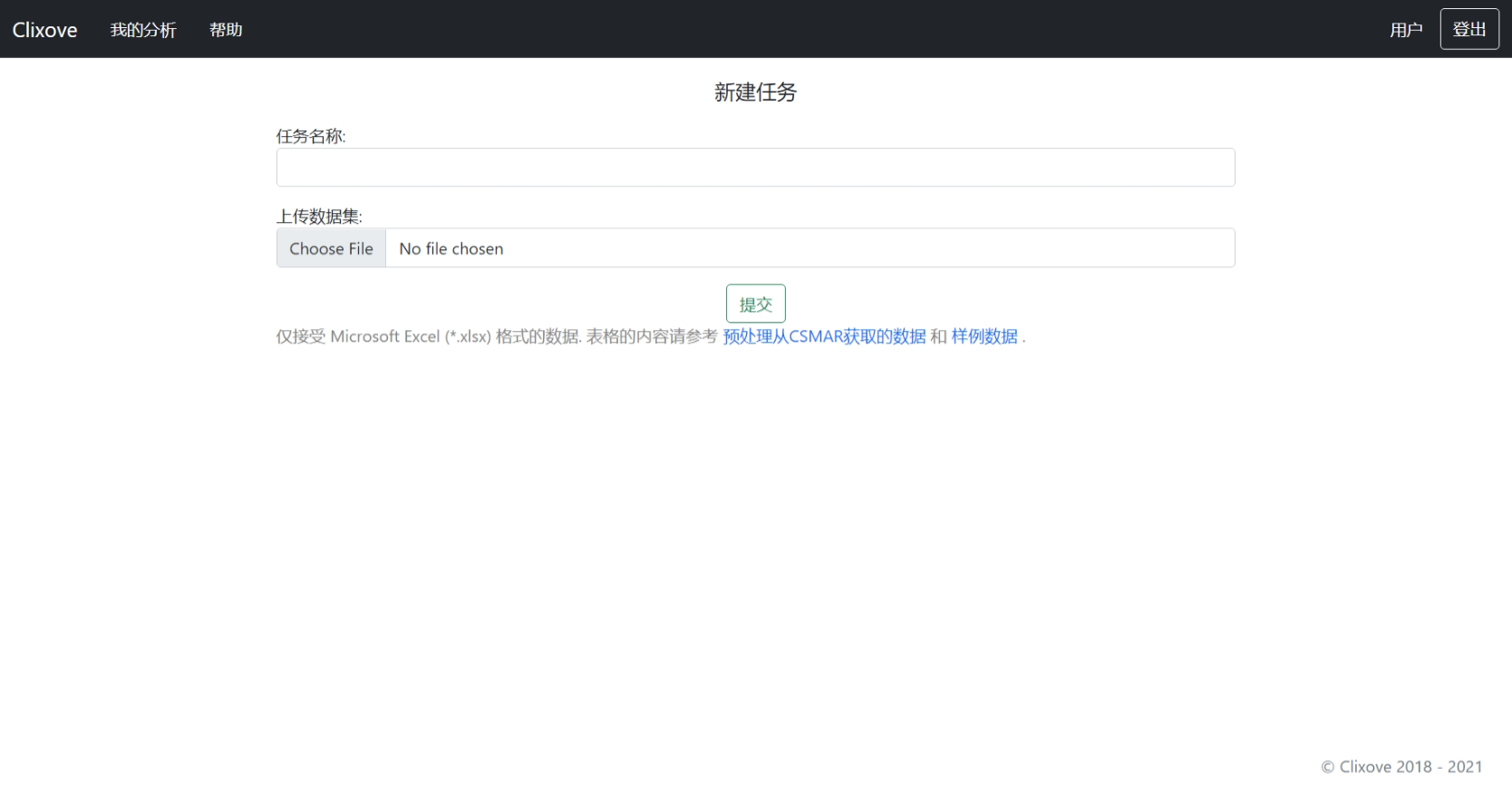


图 10 新建任务页面

（7）完成上一步后，用户进入设置变量页面，如图 11所示。用户在本页填写一个代表年份（或其他时间尺度）的变量，一个代表环境保护机构名称的变量，多个投入变量，多个产出变量。在左侧的漂浮卡片上，可以点击按钮，以紧凑或展开的方式查看投入、产出变量的候选项，也可以全选变量。

完成设置后，用户需提交表单。

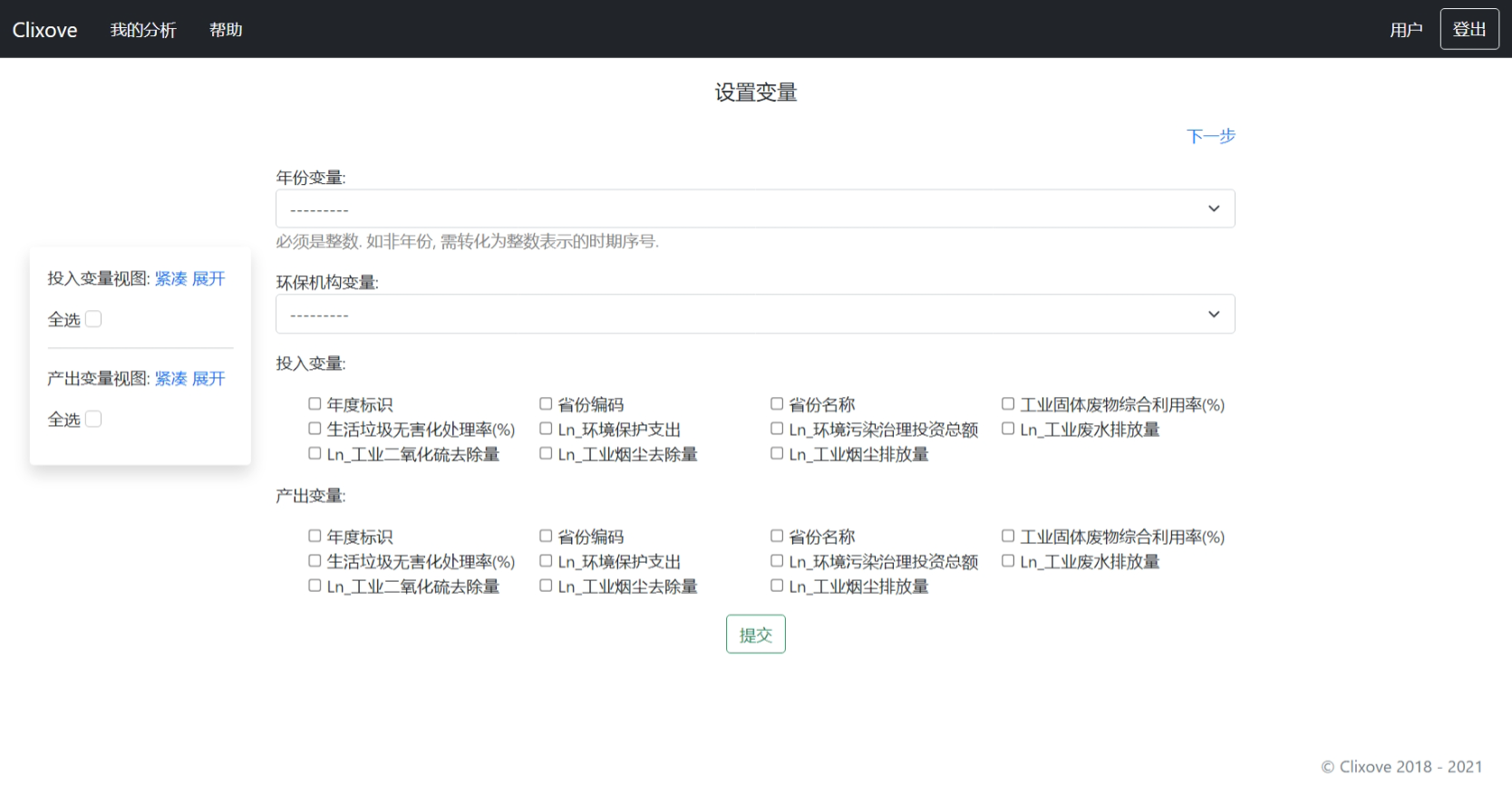


图 11 设置变量页面

提交表单后，数据包络分析算法开始运行，软件提示用户需等待一段时间，既可留在本页，也可退出软件，如图 12所示。



图 12 设置变量页面（运行提示）

（8）数据包络分析算法运行后，任务的状态变为空闲。在上一步中离开页面的用户，可在任务列表中点击任务的名称进入“查看绩效”页面；对在上一步中没有离开页面的用户，软件会重定向到“查看绩效”页面并以横幅提示分析完成。

查看绩效页面如图 13所示，用户选择环境保护机构，提交表单，软件在本页上显示环境保护机构的绩效历史折线图。其中，因为本软件绘制折线图的工具默认不支持中文，所以通过“环境保护机构名称是否包含中日韩文字”选项选择字体；如果该选项填写“是”则使用思源宋体显示图片的标题，否则使用默认字体。

在本页上点击“下载图片”按钮，可以下载\*.svg格式的折线图；点击“下载全部数据”，可以下载数据集中所有样本的年份、名称、绩效。

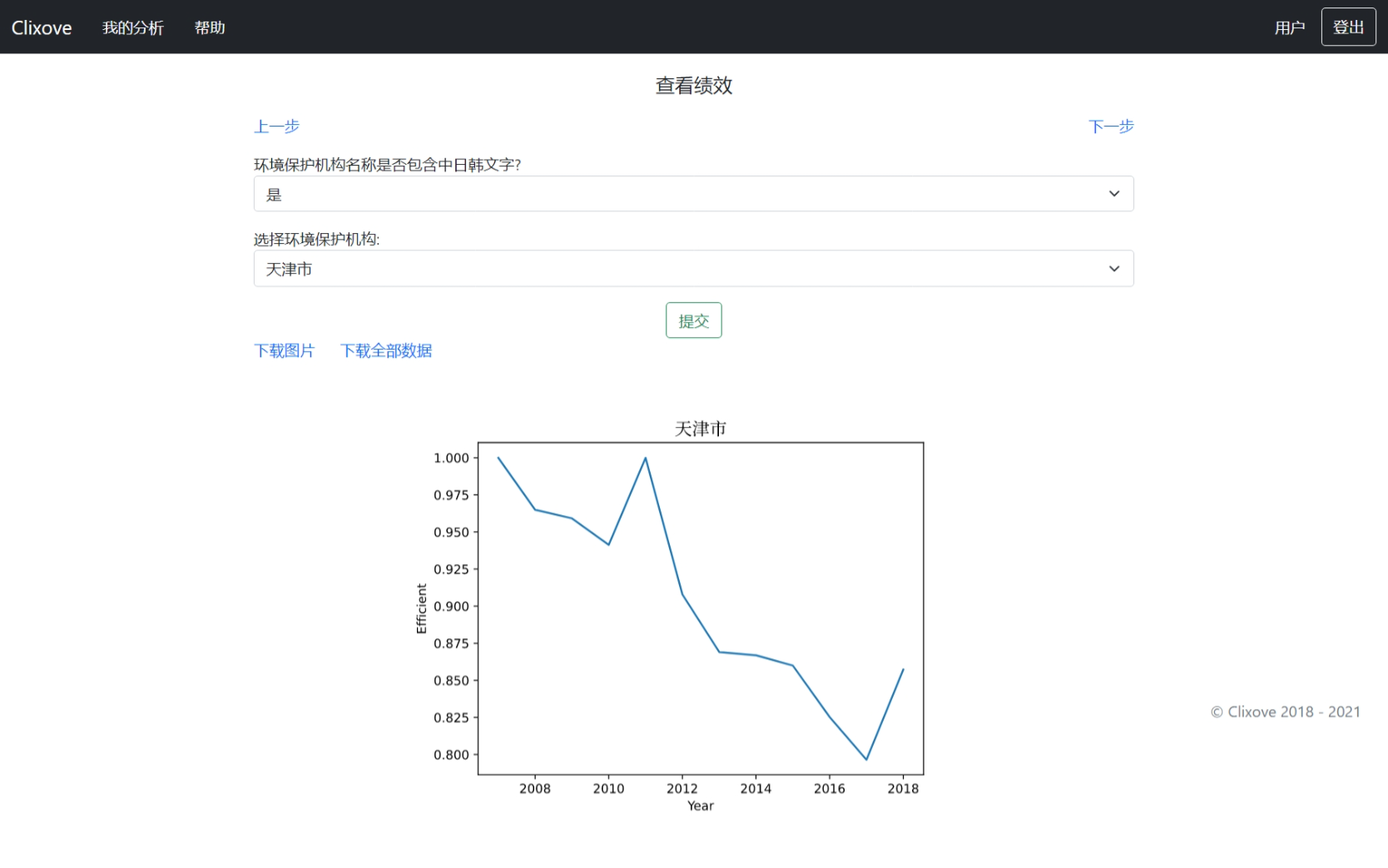


图 13 查看绩效页面

（9）点击“下一步”进入设置BHT-ARIMA模型页面，如图 14所示。



图 14 设置BHT-ARIMA模型页面

用户在本页上填写BHT-ARIMA模型所需参数，左侧的漂浮卡片对参数含义作了简单解释，软件主页的“绩效评价”有跳转向参考文献的链接，用户可参考这篇文献获得详细的指示。

用户提交表单后，需等待BHT-ARIMA算法运行，等待过程和（7）步相似。

（10）模型完成预测后，跳转到查看绩效预测页面，结果如图 15所示。

用户可在页面上直接预览结果，也可点击“下载结果”按钮下载表格。



图 15 查看绩效预测页面

到此为止，用户已实现本软件的主要功能。

1. 决策单元：调配多投入、多产出项目的实施主体，在环境保护支出的绩效评价案例中，指代每个省的环境保护机构。 [↑](#footnote-ref-1)
2. Shi, Qiquan, et al. "Block Hankel tensor ARIMA for multiple short time series forecasting." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 04. 2020. [↑](#footnote-ref-2)