

ICS
CCS

T/HNSBSXH

海南省博士协会团体标准

T/HNSBSXH 01-2025

农业科学数据在线分析挖掘技术规范

Technical Specification for Online Analysis and Mining of Agricultural Scientific Data

2025年5月16日发布

2025年6月16日实施

海南省博士协会 发布

目录

目录	I
前 言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 数据管理	2
4.1 数据领域	2
4.2 数据格式	2
4.3 数据大小	2
4.4 数据描述	2
4.5 数据管控	2
5 算子组件管理	2
5.1 算子组件类别	2
5.2 算子组件格式	2
5.3 算子组件大小	2
5.4 算子组件描述	3
5.5 算子组件封装	3
5.6 算子组件管控	3
6 分析挖掘	3
6.1 在线分析挖掘引擎	3
6.2 算子组件分析挖掘	3
6.3 场景分析挖掘	3
7 农业科学数据在线分析挖掘平台能力测试要求	3
7.1 数据上传、下载、修改、删除	3
7.2 算子组件上传、下载、修改和删除	4
8 农业科学数据在线分析挖掘技术核心能力测试标准	4
8.1 通用型算子组件	4
8.2 专业型算子组件	4
9 农业科学数据在线分析挖掘技术平台性能测试方法	4
9.1 算子组件执行速度	4
9.2 算子组件准确率	5
9.3 算子组件鲁棒性	5
10 农业科学数据在线分析挖掘安全管理	5
10.1 数据加密	5
10.2 权限管理	5
10.3 审计管理	5

参 考 文 献 6

全国团体标准信息平台

前 言

本文件按照《团体标准结构和编写指南》T/CAS 1.1-2017要求并参照GB/T 1.1-2020《标准化工作导则第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国农业科学院农业信息研究所、三亚中国农业科学院国家南繁研究院提出。

本文件由海南省博士协会归口。

本文件起草单位：中国农业科学院农业信息研究所、三亚中国农业科学院国家南繁研究院。

本文件主要起草人：周国民、张建华、闫燊、王健、赵晓燕、贺子康、李佳乐、樊景超、丘耘、王晓丽、满芮、刘婷婷、申家炜、陈晓静。

农业科学数据在线分析挖掘技术规范

1 范围

本文件规定了农业科学数据在线分析挖掘技术的数据管理、算子组件管理、分析挖掘、能力和性能测试的要求。

本文件适用于农业科学数据在线分析挖掘。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

YD/T 3762-2020 大数据 数据挖掘平台技术要求与测试方法

DB37/T 3431-2018 农业大数据 标准体系

3 术语和定义

下列术语和定义适用于本文件。

3.1 农业科学数据 Agricultural scientific data

是指在农业科学研究过程中产生的数据。

[来源：GB/T DB37/T 3431-2018 3.3，有修改]

3.2 在线分析挖掘 Online analysis mining

是指基于大数据和机器学习技术的数据处理和分析方法，可在线处理和分析挖掘数据，从中挖掘出提取有价值的信息。

注：数据挖掘通常与计算机科学有关，并通过统计、在线分析处理、情报检索、机器学习、专家系统(依靠过去的经验法则)和模式识别等诸多方法来实现上述目标。

[来源：YD/T 3762-2020 3.1，有修改]

3.3 算子组件 Operator component

是指用于执行特定数据操作的已封装算法模块或函数。

3.4 算子组件分析挖掘 Operator Component Analysis Mining

算子组件分析挖掘是指执行单个算子组件，并对数据进行分析挖掘。

3.5 应用场景 Scenario application

是指将多个算子组件按顺序或并行地连接起来，形成一个完整的数据处理流程，以实现具体场景的数据分析与应用。

[来源：YD/T 6267-2024 4.3，有修改]

3.6 场景分析挖掘 Scenario Analysis Mining

场景分析挖掘是根据具体的应用场景通过流水线式的作业方式将多个算子组合在一起完成分析和挖掘。

3.7 农业科学数据在线分析挖掘平台 An online analysis and mining platform for agricultural scientific data

是指集成了农业科学数据挖掘应用场景以及算子组件的平台化产品,用于便捷创建农业科学数据挖掘业务,从农业科学数据中提取有价值的信息,从而支持科学决策、优化农业生产、提高资源利用效率。

[来源: YD/T 3760-2020 3.1, 有修改]

4 数据管理

农业科学数据在线分析挖掘技术具备数据源管理的功能。

应符合YD/T 3762-2020中4.1的规定。

4.1 数据领域

数据领域包括但不限于:作物科学类、动物科学与动物医学类、渔业科学类、草地与草业科学类、园艺科学类、农业微生物科学类、植物保护科学类、农业资源与环境科学类、食品营养与加工科学类、农业装备工程与信息类、农业农村经济科学类以及包含农业科技基础信息等领域。

4.2 数据格式

数据格式包括但不限于:

文本格式、数值格式、图像格式、视频格式以及其他数据文件格式。

4.3 数据大小

农业科学数据在线分析挖掘技术对数据大小有以下限制:

单个数据文件大小不超过100GB。

4.4 数据描述

数据需详细描述,包括但不限于:

A数据名称:数据的直观命名。

B数据类型:数据的格式。

C数据来源:数据的原始出处或生成途径。

D数据创建时间:数据在平台上的创建时间。

E数据更新时间:数据在平台上的更新时间。

F数据版本:数据在平台上的版本状态。

4.5 数据管控

需具备数据导入、数据下载、数据修改、数据删除等功能。

5 算子组件管理

需具备算子组件管理的功能,包括以下几个方面:

5.1 算子组件类别

算子组件类别可分为通用型和专业型。通用型包括:统计分析类、机器学习类、深度学习类等。专业型包括:作物育种类、耕地测评类、农业绿色发展类等。

5.2 算子组件格式

算子组件格式包括但不限于以下格式:

代码格式: Python 编程语言编写的代码文件。

配置文件: JSON、XML、YAML 等配置文件,用于描述算子的参数和依赖关系。

模型文件: TensorFlow、PyTorch、ONNX 等框架的模型文件。

文档文件: Markdown、PDF、HTML 等格式的文档文件,用于描述算子的功能和使用方法。

5.3 算子组件大小

单个算子组件大小不超过 1GB。

5.4 算子组件描述

算子组件描述包括但不限于：

算子组件元数据：包括名称、类型、版本、作者、创建时间、更新时间等。

算子组件标签：用于算子组件的分类和检索。

算子组件摘要：简要描述算子组件的功能和用途，帮助用户快速了解组件的基本信息。

算子组件使用说明：详细描述算子组件的使用方法、输入输出格式、参数配置等。

5.5 算子组件封装

通过 python、C++等引擎封装算子组件。

5.6 算子组件管控

需具备算子组件的导入、下载、修改、删除等功能。

6 分析挖掘

6.1 在线分析挖掘引擎

在线分析挖掘引擎需嵌入计算机编程语言编译器，例如：python编译器。

支持通过API接口、Web界面等方式调用在线分析挖掘引擎。

6.2 算子组件分析挖掘

算子组件分析挖掘流程分为以下 4 步：

- 1.加载数据；
- 2.调用算子组件；
- 3.执行算子组件；
- 4.返回结果。

6.3 场景分析挖掘

场景分析挖掘的要求为：

1.需明确分析任务的具体场景，包括但不限于如作物育种类场景、耕地评价类场景以及农业绿色发展类场景。

2.根据场景需求，准备相应的数据集。

场景分析挖掘的流程为：

- 1.构建算子组件流程，如数据预处理、特征提取、模型训练、预测和推理等。
- 2.设计完整的分析流程。
- 3.运行流程。
- 3.输出各种场景分析的结果数据。

7 农业科学数据在线分析挖掘平台能力测试要求

7.1 数据上传、下载、修改、删除

测试编号：7.1
测试项目：数据下载、上传、修改、删除
测试目的：实现多种形式数据下载、上传、修改、删除
预制条件：1. 部署测试环境 2. 预制多种形式数据
测试步骤：1. 对多种形式的数据进行上传、下载、修改、删除 2. 查看数据上传、下载、修改、删除结果
预期结果：平台下载、上传、修改、删除数据成功

7.2 算子组件上传、下载、修改和删除

测试编号：7.2
测试项目：算子组件上传、下载、修改和删除
测试目的：支持包括代码格式、配置文件、模型文件、文档文件等多种形式的算子组件上传、下载、修改和删除
预置条件：已训练好的算子组件
测试步骤：1. 对算子组件进行上传、下载、修改和删除 2. 查看算子组件上传、下载、修改、删除结果
预期结果：算子组件上传、下载、修改、删除成功

8 农业科学数据在线分析挖掘技术核心能力测试标准

8.1 通用型算子组件

测试编号：8.1
测试项目：通用型算子组件的调用、运行、反馈
测试目的：实现通用型算子组件的调用、运行、反馈
预置条件：1. 农业科学数据在线分析挖掘平台测试环境正常运行 2. 有通用性算子和初始数据
测试步骤：1. 调用通用型算子组件 2. 选择对应的初始数据 3. 运行通用型算子组件
预期结果：通用型算子组件成功调用、运行、反馈

8.2 专业型算子组件

测试编号：8.2
测试项目：专业型算子组件的调用、运行、反馈
测试目的：实现专业型算子组件的调用、运行、反馈
预置条件：1. 农业科学数据在线分析挖掘平台测试环境正常运行 2. 有专业型算子和初始数据
测试步骤：1. 调用专业型算子组件 2. 选择对应的初始数据 3. 运行专业型算子组件
预期结果：专业型算子组件成功调用、运行、反馈

9 农业科学数据在线分析挖掘技术平台性能测试方法

9.1 算子组件执行速度

测试编号：9.1
测试项目：算子组件执行速度
测试目的：验证算子组件的执行速度
预置条件：1. 农业科学数据在线分析挖掘平台测试环境 2. 有算子组件和初始数据
测试步骤：1. 调用算子组件 2. 选择对应的数据 3. 运行算子组件 4. 记录算子组件的执行速度
预期结果：反馈算子组件的执行速度，算子组件能够快速、正常的运行

9.2 算子组件准确率

测试编号：9.2
测试项目：算子组件准确率
测试目的：验证算子组件的准确率
预制条件：1. 农业科学数据在线分析挖掘平台测试环境 2. 有算子组件和初始数据
测试步骤：1. 调用算子组件 2. 选择对应的数据 3. 运行算子组件 4. 记录算子组件运行的准确率
预期结果：反馈算子组件运行的准确率，算子组件能够精准计算出结果

9.3 算子组件鲁棒性

测试编号：9.3
测试项目：算子组件鲁棒性
测试目的：验证算子组件多次执行中运行速率和准确率的稳定性
预制条件：1. 农业科学数据在线分析挖掘平台测试环境 2. 有算子组件和初始数据
测试步骤：1. 调用算子组件 2. 选择对应的数据 3. 运行算子组件 4. 反复运行共计5次算子组件，分别记录每一次的运行速率、结果
预期结果：1. 每一次的计算结果均一致 2. 每一次的运行速率呈现均态分布，不会出现大幅波动

10 农业科学数据在线分析挖掘安全管理

10.1 数据加密

数据需采用符合行业标准的加密算法对原始农业数据进行加密处理。数据在传输过程中，必须使用符合行业标准的加密协议进行通道加密。

10.2 权限管理

平台用户必须通过身份认证机制登录平台。并基于最小权限原则，为不同用户分配不同的访问权限。

10.3 审计管理

平台应建立全面的操作审计日志系统，记录所有用户对数据的操作行为，包括登录时间、IP 地址、操作内容、操作结果等信息。

参 考 文 献

- [1] GB/T 1.1—2020 标准化工作导则 第1部分：标准化文件的结构和起草规则
- [2] YD/T 3762-2020 大数据 数据挖掘平台技术要求与测试方法
- [3] YD/T 3760-2020 大数据 数据管理平台技术要求与测试方法
- [4] NY/T 4061-2021 农业大数据核心元数据
- [5] DB37/T 4473—2021 农业大数据分类与编码规范
- [6] DB37/T 3431-2018 农业大数据 标准体系
- [7] 20230715-T-469（征求意见稿）人工智能 算子接口 第1部分：基础数学类