

附件 5

ICS 点击此处添加 ICS 号

CCS 点击此处添加 CCS 号

NY

中华人民共和国农业行业标准

NY/T XXXXX—XXXX

## 农业数据与模型智能匹配技术规范

Technical specification for intelligent recommendation of agricultural models

(点击此处添加与国际标准一致性程度的标识)

征求意见稿

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

2024 - XX - XX 发布

2024 - XX - XX 实施

中华人民共和国农业农村部 发布



# 目 录

前 言 .....	II
引 言 .....	III
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 农业数据与模型匹配度评价方法 .....	1
4.1 数据类型识别 .....	1
4.2 数据领域识别 .....	2
4.3 数据地域识别 .....	2
4.4 数据时间识别 .....	2
4.5 数据维度识别 .....	2
4.6 数据质量识别 .....	3
4.7 农业模型分类 .....	3
5 农业数据与模型匹配度评价标准 .....	3
6 农业数据与模型的智能匹配 .....	4
参 考 文 献 .....	5

## 前 言

本文件按照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。

本文件由农业农村部市场与信息化司提出。

本文件由农业农村部数据标准化技术委员会归口。

本文件起草单位：三亚中国农业科学院国家南繁研究院

本文件主要起草人：

（本文件历次版本发布情况）

# 引 言

农业模型作为连接数据与决策的关键工具，对于提高农业生产效率、优化资源管理和保障食品安全具有重要意义。然而，面对种类繁多的农业数据和模型，如何匹配最合适的模型以充分利用数据，成为研究人员和实践者面临的一大挑战。

本标准建立了一套农业数据与模型智能匹配的技术规范，划分了数据与模型匹配程度的三个等级，为研究人员提供清晰的指导，确保选用的数据模型能够最有效地支持农业科学研究和应用，从而提高农业科学研究的准确性和可靠性。

本文件按照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。

本文件由农业农村部市场与信息化司提出并组织实施。

本文件由农业农村部数据标准化技术委员会归口。

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：



# 农业数据与模型智能匹配技术规范

## 1 范围

本文件界定了农业数据与模型的匹配度评价标准，根据数据特性及模型类别进行智能匹配。本文件适用于农业数据的分析、挖掘及应用，有利于合理利用农业模型。

## 2 规范性引用文件

本文件没有规范性引用文件。

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**农业数据** Agricultural data

以指导农业生产经营为目的，由农业科研人员在科学实验、观察等科研活动过程中产生的数据。

### 3.2

**农业模型** Agricultural model

以农业系统要素为研究对象，为研究农业问题的定量规律所建立数学模型的总称。可用于不同的农业应用场景，包括定量预测、监测、预警与决策控制等。

### 3.3

**智能匹配** Smart matching

根据农业数据与模型匹配程度，为农业数据智能匹配最佳模型。

## 4 农业数据与模型匹配度评价方法

### 4.1 数据类型识别

数据类型识别重点在于确定农业数据的文件格式、编码方式等与模型兼容的程度，确保数据能够顺利导入和处理，从而有效支持模型的运行和分析。数据类型包括文本数据、数值数据、图像数据、视频数据、分类数据等。

#### a) 文本数据

文本数据是指由字符组成的、用于表达信息的数据形式，这些信息可以是自然语言文字（如文章、评论、报告等）、代码、符号等。

#### b) 数值数据

数值数据是指表示数量或测量结果的数据类型，可以用数字来表示，并且支持数学运算。

#### c) 图像数据

图像数据是以数字形式存储的视觉信息，由表示颜色和亮度的像素矩阵组成。

d) 视频数据

视频数据是由一系列连续的图像帧组成，每帧代表一个时间点上的视觉信息，通常伴有音频轨道。

e) 分类数据

分类数据是将信息按照特定类别或标签进行划分的数据类型，常用于表示属性或特征的非数值信息。

#### 4.2 数据领域识别

数据所属农业领域的识别，是确保数据与模型匹配的基础。可以通过数据的来源、用途、研究对象等信息来确定数据领域。包含农业、林业、畜牧业、渔业、农业经济、农业环境、农业生物技术等领域。

a) 农业：涵盖了作物种植、土壤管理、灌溉技术、肥料使用等领域。包括谷物、豆类、油料作物、蔬菜、水果、纤维作物、药用作物等数据。

b) 林业：涵盖了森林资源管理、树木生长监测、森林病虫害防治等领域。包括针叶树、阔叶树、经济林木等数据。

c) 畜牧业：涵盖了畜禽养殖、饲料管理、疫病防控等领域。包括家禽、家畜、特种养殖等数据。

d) 渔业：涵盖了水产养殖、渔业资源管理、水质监测等领域。包括淡水鱼类、海水鱼类、甲壳类、贝类等数据。

e) 农业经济：涵盖了农产品市场分析、农业政策评估、农业风险管理等领域。包括贸易数据、供需关系、补贴政策、灾害风险等数据。

f) 农业环境：涵盖了土壤环境、水资源管理、大气环境等领域。包括土壤质量、水质监测、温室气体排放等数据。

g) 农业生物技术：包括转基因作物、生物肥料、生物农药等数据。

#### 4.3 数据地域识别

数据地域识别能够确保数据与模型的适用性和准确性，通过分析数据的地理来源和适用范围，可以优化资源配置和模型参数。具体内容包括确定数据的地理来源、分析地域特征。

a) 确定数据的地理来源：利用GIS、遥感技术获取具有地理信息的数据，记录数据的经纬度、行政区划名称，确保数据的来源地与研究区域一致，避免因地域差异导致的数据不适用。

b) 分析地域特征：分析数据所在地区的自然条件，例如气候条件、土壤类型、地形地貌、水资源等，确保选择的模型能够根据不同地区的环境条件进行智能决策。

#### 4.4 数据时间识别

识别数据时间能够确保数据时效性以及数据与模型的相关性，重点识别数据的时间段以及时间长度，例如数据包含几年的信息，哪一年的信息，连续的还是间断的。

#### 4.5 数据维度识别

将数据维度划分为一维数据、二维数据、多维数据，以全面理解数据的大小和结构，从而选择合适的模型进行分析和应用。

a) 一维数据：只有一个索引或变量，通常表示为一个有序的列表或数组，每个数据点都只包含一个值，形成一条线性的数据结构。数据量通常较小。

b) 二维数据：具有两个索引或变量的数据，通常表示为表格或矩阵，每个数据点由两个坐标（行和列）唯一确定，形成一个二维的表格结构。数据量通常较大。



c) 多维数据：具有两个以上索引或变量的数据，通常表示为高维数组或张量，每个数据点由多个坐标唯一确定，形成一个多维的结构，数据量通常非常大。

#### 4.6 数据质量识别

数据质量识别包括识别数据的完整性、准确性、一致性、精确性、唯一性等，从而提高数据分析和模型应用的准确性。

- a) 数据完整性：即数据中是否存在缺失值或空值，计算数据缺失率。
- b) 数据准确性：即数据是否真实反映了实际情况，通过与已知标准或参考数据进行比较，计算误差率。
- c) 数据精确性：即数据的分辨率和详细程度，检查数据的分辨率和粒度，识别数据的精确性。
- d) 数据唯一性：即数据中是否存在重复记录。

#### 4.7 农业模型分类

根据农业模型的特点，可将农业模型分为以下5类。具体如下：

- a) 统计分析类模型：基于统计学原理，通过分析历史数据来建立数学模型，用于预测和解释农业系统的变量关系。例如最小二乘法、线性回归、PCA主成分分析、单因素方差分析等。
- b) 机器学习类模型：通过训练算法从历史数据中学习规律和模式，用于预测和分类任务，能够处理复杂的非线性关系，常见的包括K\_Means、KNN、XGboost、SVR、Random\_Forest等。
- c) 深度学习类模型：一种基于神经网络的机器学习方法，通过多层神经网络结构学习数据的高级特征，特别适合处理大规模和高维度的数据，如图像、声音和文本数据。常见的包括CNN、RNN、DNN、Yolov7、Transformer等。
- d) 专业类模型：针对特定农业问题设计的专用模型，通常结合了农业科学的理论知识和实践经验。例如，全基因组预测模型DNNGP，功能基因预测模型DeepGOPlus、单细胞亚群预测模型NRTPredictor等。
- e) 其他类模型：不属于上述四类的模型，可能涉及新的技术和方法，或者跨学科的综合模型。

### 5 农业数据与模型匹配度评价标准

为了科学、合理地评估农业科学数据与模型之间的匹配度，本标准依据数据模型在以上数据类型、数据领域、数据地域、数据时间、数据维度、数据质量六个指标上的表现，将匹配度划分为“完全匹配”、“部分匹配”和“完全不匹配”三个等级。这一分级旨在为研究人员提供清晰的指导，确保选用的数据模型能够最有效地支持农业科学研究和应用。

表 1 农业数据与模型匹配度评价标准

匹配等级	规则描述
完全匹配	数据在类型、领域、地域、时间、维度和质量上均完全符合农业模型的需求时，视为完全匹配。
部分匹配	数据在类型、领域、地域、时间、维度和质量中有3~5个指标符合农业模型的需求，而其余指标存在一定程度的不匹配时，视为部分匹配。

匹配等级	规则描述
完全不匹配	数据在类型、领域、地域、时间、维度和质量中的大部分或全部指标与农业模型的要求严重不符时，视为完全不匹配。

## 6 农业数据与模型的智能匹配

在获取农业数据的基础上，详细分析不同农业数据的特点和农业应用场景的需求，精准地选择与之匹配的农业模型，从而提升农业研究和实践的效果。农业数据与模型的智能匹配流程如下：

- a) 首先，识别出农业数据的特性，包括数据类型、数据领域、数据地域、数据时间、数据维度、数据质量六大指标。
- b) 其次，识别出农业模型的类别，包括统计分析类模型、机器学习类模型、深度学习类模型、专业模型和其他模型。
- c) 最后，根据农业数据与模型匹配度评价标准对数据模型进行智能匹配，划分为完全匹配、部分匹配、完全不匹配三个等级。

### 参 考 文 献

- [1] 周国民.我国农业大数据应用进展综述[J].农业大数据学报,2019,1(01):16-23.
  - [2] 彭秀媛,王枫,周国民.农业科学数据共享技术系统研究[J].园艺与种苗,2019,39(02):53-57.
  - [3] 曹宏鑫,葛道阔,张文宇,等.农业模型发展分析及应用案例[J].智慧农业(中英文),2020,2(01):147-162.
-