

# 计算机类高职数学建模教学改革实践与探索

苑占江 桂改花 林泽楷 黄丽婷 游家仁  
广东科学技术职业学院 广东珠海 519090

**摘要:** 数学建模在生活中的应用非常广泛, 很多行业遇到一些问题都会通过设置条件, 转换为数学问题进行解决。本文针对2022年全国大学生数学建模E题, 对物料需求量预测以及生产计划的问题进行分析, 运用熵权法—TOPSIS、XGBoost算法对附件中的历史数据进行处理分析, 利用其数据使用监督学习进行模型训练, 将得到的模型利用指标权重对物料进行分类筛选, 从而可以得到物料的相关数据以及生产计划。利用Python和Excel对附件中的数据运用熵权法—TOPSIS进行权重分析, 将得分进行归一化处理后得到的6种物料, 再利用XGBoost法依据6种指标的历史数据进行训练, 得到6种物料中有5种物料的都大于0.925, 拟合较好, 模型拟合度很高。采用滑动窗口法对问题一的模型进行优化改进, 6种物料的平均服务水平在93%以上。利用安全库存的计算公式使得库存量在安全范围内保存较低水平。利用问题一的模型得到随着周数的增加, 生产计划的平均服务水平呈现上升趋势, 而库存变化率呈下降趋势, 因此该方法可以推广到一般情况。

**关键词:** 小批量物料生产安排; XGBoost算法; 熵权法—TOPSIS; 时间序列

## Practice and Exploration on teaching Reform of mathematical modeling in computer higher vocational colleges

Zhanjiang Yuan, Gaihua Gui, Zekai Lin, Liting Huang, Jiaren You  
Guangdong Vocational College of Science and Technology, Zhuhai, Guangdong, 519090

**Abstract:** The application of mathematical modeling in life is very wide, many industries will encounter some problems through setting conditions, converted into mathematical problems to solve. Aiming at the mathematical modeling problem E of national college students in 2022, this paper analyzes the material demand prediction and production plan, uses entropy weight method -- TOPSIS and XGBoost algorithm to process and analyze the historical data in the attachment, and uses its data to conduct model training using supervised learning. The obtained model is used to classify and screen materials by index weight, so that the relevant data and production plan of materials can be obtained. Using Python and Excel, the entropy weight method -- TOPSIS was used to carry out weight analysis on the attached data, and six kinds of materials were obtained after normalized score processing. Then XGBoost method was used to train according to the historical data of six kinds of indicators, and the results showed that 5 kinds of materials in the six kinds of materials were greater than 0.925, which was a good fit. The model has a high degree of fitting. The sliding window method was adopted to optimize the model of problem 1, and the average service level of 6 kinds of materials was above 93%. The calculation formula of safety stock is used to keep

**基金项目:** 广东省高等职业教育教学改革研究与实践项目人工智能技术应用专业“产教融合、精准育人”人才培养体系的构建与实践(项目编号: GDJG2021155); 2019年广东科学技术职业学院校级教改项目以CCF CSP认证为导向的程序设计基础类课程改革与实践(项目编号: JG201949); 广东省智慧职教工程技术研究中心(项目编号: 2021A118); 2022年《应用数学(1)》校级课程思政示范课; 2021年《应用数学(1)》校级金课。

### 作者简介:

苑占江(1981.2—), 男, 汉族, 河北邢台, 博士研究生, 讲师, 研究方向: 人工智能, 算法, 数学建模。

桂改花(1981.9—), 女, 汉族, 山东聊城, 硕士研究生, 讲师; 研究方向: 应用数学, 数学建模, 数学教育。

the stock at a low level within the safety range. Using the model of problem 1, it is found that with the increase of weeks, the average service level of production plan shows an upward trend, while the change rate of inventory shows a downward trend. Therefore, this method can be extended to the general situation.

**Keywords:** Small batch material production arrangement; XGBoost algorithm; Entropy weight Method -- TOPSIS; Time series

从全国建模大赛分析, 随着数学建模在全世界越来越热, 一个学校在全国建模大赛中成绩的好坏已成为评定该学校生源好坏及教学质量优劣的标准, 很多同学缺少的不是实力, 而是锻炼的机会, 因而为了让更多的人在全国建模比赛前有一个练兵的机会, 全院、全校建模大赛的举办势在必行。随着全国建模大赛越来越热, 参赛选手每年剧增, 而如何挑选选手也成为老师一个头痛的问题。全院、全校建模大赛的举行很好地解决了这一难题, 通过预选, 选手们可以提前进入全国建模大赛的名单。

从同学们的角度分析, 同学们大部分都想学好数学, 以便参加全国大学生数学建模竞赛做准备, 把学到的知识应用到实际生活中, 竞赛正好为同学们提供了一个途径, 一个平台, 所以大部分同学都支持举办数学建模活动。从每年报名的情况说明了数学建模在同学们心目中还是有一定影响力的。

从老师们的角度分析, 计算机学院有建模方面的专业老师, 在近几年的建模竞赛中作为专业辅导老师带领参赛者取得优异的成绩, 同时也积累了许多宝贵的竞赛经验。我们在暑假期间对参加全国数学建模的会员进行集中培训。这为我们举办这次活动更是积累了实际的经验。在此基础上, 我相信我们有足够的实力和经验来举办这次竞赛。

由于数学建模在创新人才培养中的地位和作用所在, 数学建模受到了越来越受到重视和关注, 另一方面, 也是因为数学建模竞赛有很强的可比性和竞争性, 竞赛成绩是反映能力和水平的一个实力型指标, 也是高校评估的一个重要指标。

### 引言:

问题提出在当前经济全球化以及区域一体化的新发展格局下, 下游需求带动电子产品制造产业发展, 目前已成为我国重要的战略产业和经济快速增长的支柱产业之一, 然而经济全球化给我国企业带来发展机遇的同时, 也加剧了企业之间的竞争, 其实质是资源的竞争。

目前在制造行业的信息化进程中, 大多数技术都停留在分销和财务系统的应用层面, 很难深入到生产管理层面的应用。而生产方面也面临着许多问题, 比如在无

法知道实际需求量的情况下生产多品种的小批量物料会出现物料浪费、人力浪费等问题。生产管理是企业经营过程最复杂的一点, 对灵活性、成熟度有较高要求。而物料计划的编排是企业经营过程中的一个重要的增值管理点, 企业对物料生产计划的制定以及对生产能力的控制还不成熟, 与之密切相关的成本控制也成为制造企业内部最为关注的核心问题。

本文通过理论与实践相结合的途径, 研究生产计划和服务水平两大部分内容。文章对附件1的数据进行预处理, 问题一要求从中选择6种物料利用其历史数据建立周预测模型并进行评价。问题二要求在平均服务水平不低于85%的要求下制定从第101周后每周的物料生产计划。问题三在求出的6种物料当中选择一种物料计算其相关数据。计算6种物料的相关数据和各种综合结果。问题四要求调整周生产计划使得库存量和服务水平之间达到平衡。假设某周生产的物料只能在 $k$  ( $\geq 2$ )周或以后使用, 重新考虑问题二和问题三制定生产计划。问题层层递进, 最终结果证明所用方法可以推广应用。

## 一、XGBoost算法模型

### 1.XGBoost算法

本文主要研究物料生产计划预测, 选择Xgboost作为基础检测算法, 通过引入滑动窗口法、安全库存调节等技术来进一步改进算法, 提高Xgboost算法物料生产计划的预测能力, 降低生产多品种小批量物料过程中造成物料浪费损失和人力浪费的影响。

Xgboost (Extreme Gradient Boosting) 是2015年提出的一种集成学习算法, 在GBDT (Gradient Boosting Decision Tree) 的基础上, 通过改进算法拟合目标并且在目标函数中添加正则项, 来进一步优化模型性能。GBDT算法在优化损失函数时只使用了一阶导数信息, Xgboost算法则是对损失函数进行了二阶泰勒展开, 利用了损失函数一阶导数和二阶导数信息。因此该方法可以更好地拟合损失函数, 减少模型优化过程中存在的误差。具体定义如下所示。

引入正则化项的逻辑回归与线性回归, 防止过拟合, 对代价函数做二阶Taylor展开, 能让其自动学习分裂方向

且支持列抽样，模型表示为：

$$\hat{y}_i = \sum_{k=1}^k f_k(x_i), f_k \in F (i=1, 2, \dots, n) \quad (1)$$

其中  $\hat{y}_i$  为样本  $x_i$  的预测结果； $k$  表示决策树中树的总和； $f_k$  表示决策树第  $k$  棵树。模型表示为：

$$F = \{f(x) = \omega_{q(x)}\} (q: R^m \rightarrow \{1, 2, \dots, T\}, \omega \in R^T) \quad (2)$$

上式表示 CART 决策树结构集合， $q$  表示样本映射到叶子节点的树结构； $T$  表示叶子节点树； $\omega$  表示叶子节点的实数分数。

建立 XGBoost 目标函数最小化寻找最优参数，建立最优模型。目标函数为：

$$Obj = L + \Omega \quad (3)$$

其中， $L$  表示误差函数项， $\Omega$  表示模型复杂度函数项。误差函数项运算式：

$$L = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

## 2. 模型评价指标

题目要求在建立周预测评价模型后再利用历史数据对其进行评价分析，本题采取较为常用的五个评价指标系数进行模型评价。

(1) *MSE* (均方误差)

$$MSE = \frac{SSE}{n} = \frac{1}{n} \sum_{i=1}^m \omega_i (y_i - \hat{y}_i)^2 \quad (5)$$

其中  $\frac{SSE}{n}$  与 *SSE* 之间区别并不大 ( $n$  为样本的个数)。

(2) *RMSE* (均方根误差)

在统计参数时，均方根误差被称为回归系统拟合的标准误差，是 *MSE* 的平方根，计算公式如下：

$$RMSE = \sqrt{MSE} = \sqrt{\frac{SSE}{n}} = \sqrt{\frac{1}{n} \sum_{i=1}^m \omega_i (y_i - \hat{y}_i)^2} \quad (6)$$

(3) *MAE* (平均绝对误差)

计算所有单个观察值与算术平均值的偏差的绝对值的平均，可以准确反映实际预测误差的大小，计算公式如下：

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (7)$$

(4)  $R^2$  (决定系数)

$R^2$  表示一个随机变量与多个随机变量关系的数字特征，反映回归模式中因变量变化可靠程度的一个统计指标， $R^2$  越接近于 1 证明拟合程度越好；反之，如果越接近于 0，说明拟合度越差。计算公式如下：

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \quad (8)$$

(5) *MAPE* (平均绝对百分比误差)

利用绝对值避免正误差和反误差相互抵消，从而比较各种时间序列模型预测的相对误差的准确性，计算公式如下：

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (9)$$

## 二、实验仿真

### 1. 问题分析

问题一 首先对附件 1 中的数据按照时间的周次和出现的次数进行频数求和，筛选出频数之和大于 100 的物料进行评价，后选取物料的平均销售单价、需求数量总和、平均销售总价以及频数作为评价指标采用熵权法—TOPSIS 来分析，利用欧氏距离法计算其权重在不同物料之间的占比，根据排序可得 6 种应当重点关注的物料，然后再根据其数据以周为单位建立预测模型并进行评价。

问题二 要求在平均服务水平不低于 85% 的情况下对物料需求量的预测值、需求特征、库存量和缺货量等方面进行考虑制定新的生产计划。本题首先利用滑动窗口法对时序数据进行处理使得模型可以得到更准确的数据，然后再将需要考虑的指标作为评价指标计算其服务水平得出相关数据。

问题三 要求在库存量和服务水平之间达到平衡的情况下调整现有的生产计划，并重新计算 6 种物料的相关数据，当想要提升服务水平时，就需增加库存量。但库存量并不是越高越好，库存量越大，物料所需的占用资金就越多，这会使企业的总利润减小。库存量是一个动态的值，依照问题二的模型，提高库存量的唯一方法就是增加安全库存量。本题在此基础上，优先考虑服务水平来计算最小安全库存量，尽可能减小物料占用资金，利用安全库存量的计算公式使得库存量在合适范围内。

### 2. 数据预处理

首先运用 Python 把附件中的数据按照时间的周次和出现的次数进行统计汇总，再利用 Excel 将同一周内同一类物料进行一个频数求和。例如：第五周中有同样的物料与周次，但出现的频数不同，我们将其频数作为同类进行频数求和，从而得到一个新的频数之和。根据频数之和利用 Excel 将附件中的 284 种物料频数低于 100 的数据筛选剔除掉，最后剩下的物料为 58 种。

### 3. 指标选取和模型结果

采用熵权法—TOPSIS选取评价指标进行权重分析从而精准地反映各评价方案之间的差距。首先选择指标：平均销售单价、需求数量总和和平均销售总价和频数，然后对此指标进行正向化和标准化，再运用欧氏距离法对指标进行加权，计算不同物料与最大值最小值之间的距离，进行归一化得分排序，最终得到评价结果，6种物料编号分别是：6004021055、6004021155、6004020918、6004010174、6004020900、6004021030。本题以出现频数、需求量总和、平均销售单价以及平均销售总价作为指标进行综合得分指数评价，综合得分指数越高代表其研究对象越好，因此选取上述6种物料作为重点关注。

将预处理后的物料需求量数据分为训练集和测试集两部分，使用神经网络中的XGBoost模型对物料进行预测，预测结果该模型可以对物料的需求量进行有效预测，图中橙色曲线表示物料的实际真实需求量，蓝色曲线表示模型的预测量，橙色曲线能够有效覆盖蓝色曲线，可以得知该模型的拟合效果较好。

### 4. 模型评价

选取较为常用的五种评价指标，按照5种评价指标系数对模型进行预测，通过评价指标系数得知6种物料中有5种物料的 $R^2$ 都大于0.925，拟合较好，模型拟合度很高。基于XGBoost算法的预测模型，因其不错的学习效果 and 训练速度，应用数据挖掘，有很好的理论意义。

### 三、模型改进和生产计划预测

增加了对物料需求量的预测值、需求特征、库存量和缺货量等方面的考虑，能够有效地预测未来多个数据，为了让XGBoost模型能够有效地预测未来多个数据，因此采用滑动窗口法对模型进行改进。滑动窗口法是用于时间序列预测的一种特殊数据处理方法，第一种预测方法主要是采用前 $n$ 个数据预测未来1个数据，即使用时间序列 $X_n$ 预测 $X_{n+1}$ 。

选定6种物料中的一种物料中的6004020918物料，对其进行第101~110周的生产计划数、实际需求量、库

存量、缺货量和服务水平统计表第101~110周的生产计划、实际需求、库存、缺货量及服务水平，再将选定的6种物料按照生产计划数、实际需求量、库存量、缺货量和服务水平统计，在上述相关数据得基础上计算其均值，平均服务水平分别为93%、97.35%、94.81%、93.43%、99.17%和96.79%。

对于物料生产计划的安排随着 $k$ 值的增加，生产计划的平均服务水平呈上升趋势，而库存变化率呈下降趋势，所以根据题目判断该方法能推广到一般情况。

### 四、结束语

针对在生产多品种小批量物料过程中造成物料浪费损失和人力浪费等问题，本文提出了一种改进的XGBoost算法。使得改进之后的算法既解决了企业生产过程中小批量物料的实际需求量预测问题，同时使模型更专注于少数类别训练比较困难的样本，运用的XGBoost算法采用了二阶泰勒展开，不仅支持CART还支持线性分类器，且相对于GBDT用到的一阶泰勒展开精度更高，提高模型的预测性能。对于同样具有时序特点的其他工业物料生产需求量的预测同样适用，具有一定的泛化能力。

### 参考文献：

- [1]李福，徐良杰，朱然博，罗浩顺，陈国俊.基于XGBoost算法的共享单车借车需求量预测[J].武汉理工大学学报（交通科学与工程版），2021，45（05）：880-884.
- [2]崔佳杉，年梅，张俊.基于XGBoost算法的学业成绩预警模型研究[J].计算机与数字工程，2022，8（50）.
- [3]薛嘉怡.基于决策树模型改进的商场销售额预测[D].厦门大学，2019.
- [4]刘美凤.高校本科生学业预警制度研究[D].西安：陕西师范大学，2017.
- [5]尹茂竹.基于大数据的高校学生学业成绩预警分析[D].天津：天津商业大学，2018.
- [6]陈子健，朱晓亮.基于教育数据挖掘的在线学习者学业成绩预测建模研究[J].中国电化教育，2017（12）：75-81.