

# 废旧电力电缆分级分类模型研究

李 旻<sup>1</sup>, 李 鑫<sup>1</sup>, 张雍斌<sup>1</sup>, 朱博闻<sup>1</sup>, 黄 华<sup>2</sup>, 田 阳<sup>2</sup>, 赵佳佳<sup>2</sup>

<sup>1</sup>上海恒能泰企业管理有限公司利永电力分公司, 上海

<sup>2</sup>上海程析智能科技有限公司, 上海

收稿日期: 2023年6月17日; 录用日期: 2023年7月19日; 发布日期: 2023年7月26日

## 摘 要

为提升废旧电缆处置效率与效益, 提高国网上海市电力公司废旧物资的资产回收水平, 本文通过建立废旧电缆评价体系, 提取对废旧电缆价值影响显著的特征指标; 通过二阶聚类法以提取出的特征指标为主要因素对历史数据进行分类; 通过决策树模型对废旧电缆各型号分类结果进行预测; 最后对决策树预测性能进行验证, 并产生风险统计表。结果表明, 预测正确率在90%以上, 风险均小于0.05, 可见分类结果优良。

## 关键词

废旧电缆, 分级分类, 分类模型, 物资处置

# Research on Classification Model of Waste Power Cables

Min Li<sup>1</sup>, Xin Li<sup>1</sup>, Yongbin Zhang<sup>1</sup>, Bowen Zhu<sup>1</sup>, Hua Huang<sup>2</sup>, Yang Tian<sup>2</sup>, Jiajia Zhao<sup>2</sup>

<sup>1</sup>Liyong Power Branch of Shanghai Hengnengtai Enterprise Management Co., Ltd., Shanghai

<sup>2</sup>Shanghai Chengxi Intelligent Technology Co., Ltd., Shanghai

Received: Jun. 17<sup>th</sup>, 2023; accepted: Jul. 19<sup>th</sup>, 2023; published: Jul. 26<sup>th</sup>, 2023

## Abstract

In order to improve the efficiency and benefit of waste cable disposal and improve the asset recovery level of waste materials of State Grid Shanghai Electric Power Company, this paper extracts the characteristic indicators that have a significant impact on the value of waste cables by establishing a waste cable evaluation system; through the second-order clustering method, the historical data is classified with the extracted feature index as the main factor; the classification results of various types of waste cables are predicted through the decision tree model; finally, the predic-

tion performance of the decision tree is verified and a risk statistics table is generated. The results show that the prediction accuracy rate is above 90%, and the risk is less than 0.05, which shows that the classification results are excellent.

## Keywords

Waste and Old Cables, Grading and Classification, Classification Model, Material Disposal

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

废旧电力电缆是电网企业非常重要的可再生资源。通过对废旧电力电缆科学高效的回收和再利用,可以达到节能减排、保护环境和提高资源利用率等多重效益。然而,目前电力公司对废旧电缆的回收处置仍存在许多问题,严重制约了废旧物资的资产回收水平。

据前期调研,上海电力公司废旧仓库中的废旧电缆种类多、总量大,不同价值电缆均混合堆放,分级颗粒度粗且缺乏统一的物资分级标准。同时由于废旧电缆的质量参差不齐,种类繁多且市场需求也各不相同,故废旧电缆混合堆放将导致后续废旧电缆回收估价困难,难以与回收商需求匹配,从而导致回收成交率低的问题发生。因此对废旧电缆进行分级处置变得尤为重要。

现阶段国内外关于物资分类的相关研究主要集中在煤矿物资分类、应急物资分类和库存物资分类等方面,针对废旧电力电缆的研究主要集中在废旧电缆回收处理技术理论研究、回收处理工艺设备方法研究和废旧电缆中材质回收再利用研究等,而针对废旧电力电缆的分类模型研究相对较少。在这种背景下,如何对废旧电缆进行有效的分级分类研究已经成为一个重要的课题。

本文主要探讨废旧电缆分级分类的相关问题,包括分类模型的选择,特征指标的提取以及模型的构建,旨在提高上海电力公司废旧电力电缆资产回收水平,为推动废旧电缆回收产业的发展提供有力支持。

## 2. 分类模型研究现状

最早开始广泛应用的物资分类方法是 ABC 分类法,但随着市场环境的不不断变化,简单的 ABC 分类法已经很难满足分类的复杂性要求,于是,国内外众多学者对分类方法及模型的理论和应用进行更深入的探究。

对于废旧物资分类问题,现有文献主要以神经网络、SVM、聚类模型等分类方法为工具进行研究。在国内,韩赛等[1]针对煤炭企业物资分类,建立概率神经网络物资分类模型,通过 MATLAB 仿真分类结果,实现了煤炭物资重要程度的分类,提高煤炭物资库存管理效率。魏伟等[2]基于 SVM 模型,设计了人工智能电力线路故障检测分类系统。聚类分析也常被用于研究物资分类,王庆荣等[3]基于改进的 K 均值聚类分类算法,对应急物资分类问题进行探究。江玮璠[4]针对库存物资分类问题,使用模糊聚类分析来进行物资的多准则 ABC 分类。在国外,Havens 等[5]对比 3 种扩展的模糊 c 均值(FCM)聚类算法对于大数据的执行效率。Ferreira 等[6]给出一种利用 MapReduce 开展大规模数据聚类的方法。

综上所述,对于物资分类模型,国内外学者展开了丰富的研究,本文根据废旧电缆特性以及数据特点,采用聚类分析中的二阶聚类法对废旧电缆进行分类。其优点为可同时分析连续变量和多个分类变量、

可自动分析输出最优的聚类数目、并且可处理大型数据集。

### 3. 废旧电缆评价体系构建

根据废旧电缆价值的不同对废旧电缆进行分级分类，首先需要确定物资的评价体系，通过对上海电力公司废旧物资处置平台所收集的数据进行汇总，获取废旧电缆的数据信息；其次根据已有数据设计指标；最后根据指标数据进一步进行特征筛选，遴选出重要指标进行模型构建。

#### 3.1. 指标梳理

根据国网上海市电力公司 2022 年废旧物资处置历史数据，将废旧电力电缆评价体系分为三个部分：分别为基本信息、物理属性和电缆参数。具体指标梳理结果如下表 1 所示：

**Table 1.** Preliminary development of dimensional indicators

**表 1.** 维度指标初步拟定

维度	指标	指标类型
基本信息	业务类型	[工程废料；废弃材料；固定资产]
	项目类型	[营销；生产性更改；基建]
物理属性	电压等级	[AC 10 KV; AC 35 KV; AC 110 KV; AC 220 KV...]
	截面积	[16; 35; 70; 120; 185; 240...]
	线芯	[1; 3; 4]
电缆参数	电缆保护套	[YJLW; YJV; YJQ]
	铠装	有无铠装
	阻燃	[ZA; ZB; ZC; 无阻燃]
	阻水	[无阻水；普通]

#### 3.2. 特征提取

由于废旧电缆有价值的主要是金属部分，其中占比最大的是铜含量。在拍卖过程中，不同型号规格的废旧电缆的价格主要由其提炼所得铜的重量决定，并且不同型号、不同规格的废旧电缆具有不同的含铜率。因此，本文将基于含铜率对废旧电缆的特征指标进行提取。

依据上海电力 2022 年废旧电缆实际含铜量以及组成材料分析抽样报告处置明细中的数据，汇总得到各类型电力电缆中导体纯铜的含量，组成成分重量，经计算得到电缆含铜率。将含铜率作为因变量，对上述梳理出的指标进行影响程度分析，旨在探索各指标对电力电缆含铜率是否有显著性影响，以提取影响电力电缆拍卖价格的特征指标。

##### 3.2.1. 主体间效应检验

此处采用的是方差分析法，方差分析本质上研究的是分类型自变量对数值型因变量的影响。方差分析分为单因素和多因素，本文采取的是多因素方差分析，用于研究一个因变量是否受到多个因素的影响，它检验多个因素取值水平的不同组合之间，因变量的均值之间是否存在显著的差异。模型结果如下表 2 所示：

**Table 2.** Inter subjective effect test  
**表 2.** 主体间效应检验

因变量		含铜率(%)	
源	自由度	F 值	显著性
电压等级	4	506.33	0.000
截面积	8	329.83	0.000
线芯	2	1012.88	0.000
项目类型	0	.	.
铠装	2	527.81	0.000
业务类型	2	30.972	0.134
电缆保护套	0	.	.
阻燃	3	120.498	0.000
阻水	0	.	.

4. R 方= 0.959 (调整后 R 方 = 0.959)

根据主体间效应检验发现，以显著性水平 0.05 为基准，表明分类型自变量对电力电缆含铜率具有显著影响效果的因素为：电压等级、截面积、线芯、铠装、阻燃。

### 3.2.2. 特征指标提取

经主体间效应检验得到的显著指标可能仅仅是统计学意义上的显著，且各种统计分析结果都不能脱离实际。对于本文的研究对象，废旧电缆中有价值的主要为其中的金属材料铜，铠装则主要包含金属钢，但在本次研究的废旧电缆中，铠装在入库前均已拆卸，因此不将铠装纳入考量。阻燃为电缆的性能指标，对电缆含铜率并无影响。剔除无关指标后，故将最后的特征指标定为电压等级、线芯、截面积。具体框架如下表 3 所示：

**Table 3.** Characteristic indicators  
**表 3.** 特征指标

维度	指标	指标类型
物理属性	电压等级	[AC 10 KV; AC 35 KV; AC 110 KV; AC 220 KV...]
	截面积	[16; 35; 70; 120; 185; 240...]
	线芯	[1 芯; 3 芯; 4 芯]

### 3.2.3. 关系强度检验

确定特征指标后还需判断各因素与含铜率之间的关系强度。由于废旧电缆的含铜率是连续变量且影响废旧电缆价值的物理属性等都为分类变量，Eta 系数主要是判断分类变量与连续变量之间的关系强度，因此本文采用 Eta 系数来计算诸多因素与含铜率之间的关系强度。有关结果如下表 4 所示：

**Table 4.** Correlation measurement  
**表 4.** 相关性测量

	Eta	Eta 平方
含铜率*电压等级	0.608	0.370
含铜率*截面积	0.659	0.434
含铜率*线芯	0.608	0.370

根据 Eta 系数的判定准则，当 Eta 平方低于 0.06 则是低相关，处于 0.06~0.16 之间说明两变量中强度相关，大于 0.16 说明两变量强相关。所以从上表可以看出，各因素与铜重量都有着高强度的相关关系。

## 4. 模型构建

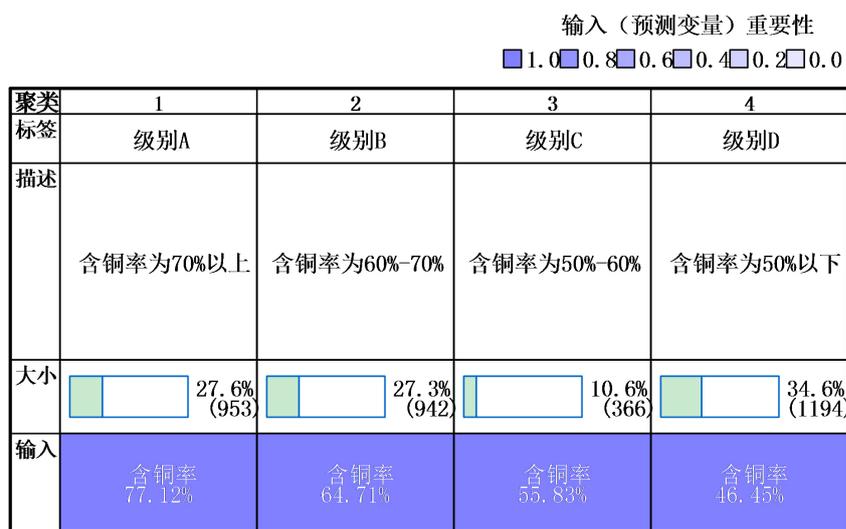
### 4.1. 分类模型构建

考虑到废旧电力电缆的价值会受到电压等级、导体截面积等多种因素的影响，且每种因素又有众多水平，因此本文在众多分类法中选择了聚类方法，该方法无论是定性变量亦或者是定量变量都可以将其很好的进行分类，同时还能有效地分析大样本数据。

根据上文对历史数据以及不同规格型号的废旧电缆含铜率的梳理，以及各指标对含铜率影响程度的分析，通过聚类法对模型进行构建。通过根据过往经验和对分类类别个数的试验，最终选择将类别个数定在 4 类，经过二阶聚类的生成过程，产出如图 1 所示四个类别。分布情况如图 2 所示。

根据聚类过程的判别输出预测变量重要性图，如图 3 所示，可以看出截面积、电压等级、线芯三个特征指标对废旧电缆含铜率尤为重要，这也从另一角度说明本文特征提取的精确性。

分类结果如表 5 所示，废旧电缆属性所对应的含铜率大部分在 50% 以下的被为第 4 类，可将其定义为 D 级别；含铜率在 50%~60% 的被分为第 3 类，可将其定义为 C 级别；含铜率 60%~70% 的被分为第 2 类，可将其定义为 B 级别；含铜率在 70% 以上的被分为第 1 类，可将其定义为 A 级别。



**Figure 1.** Cluster category results

**图 1.** 聚类类别结果

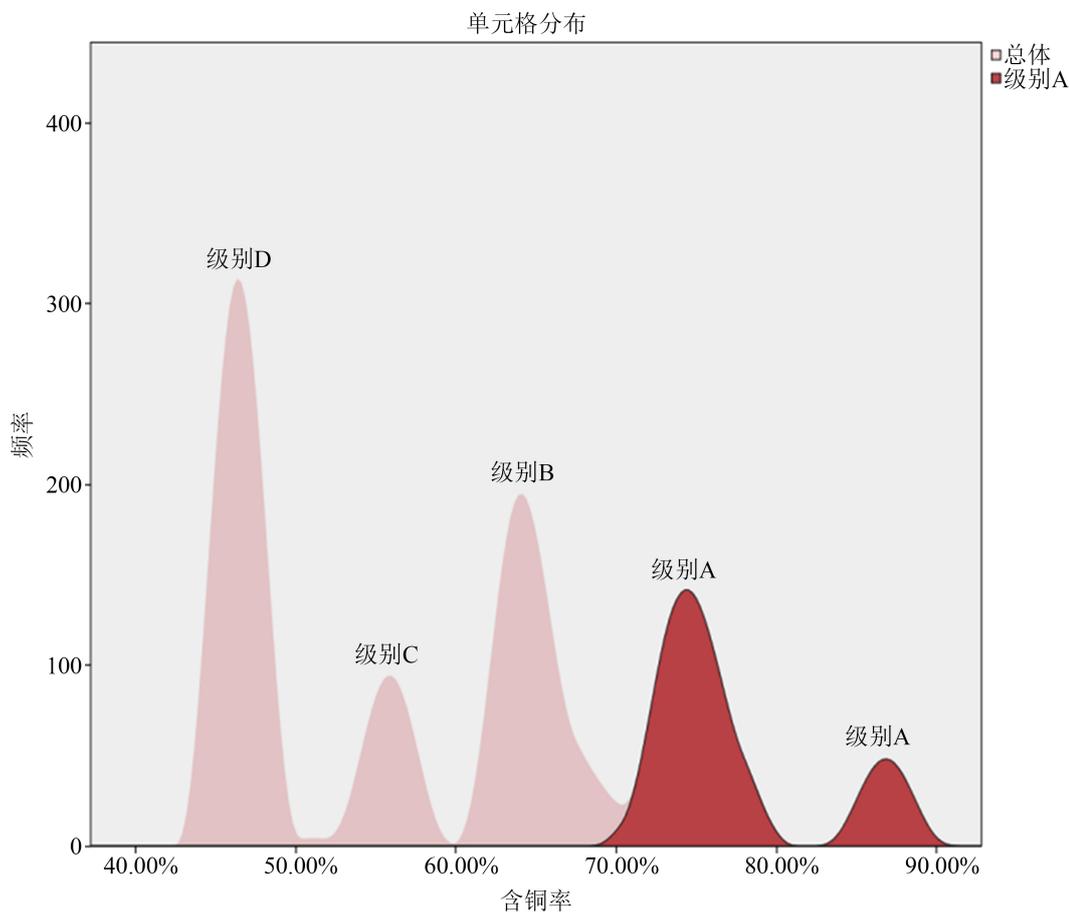


Figure 2. Cluster distribution map  
图 2. 聚类分布图

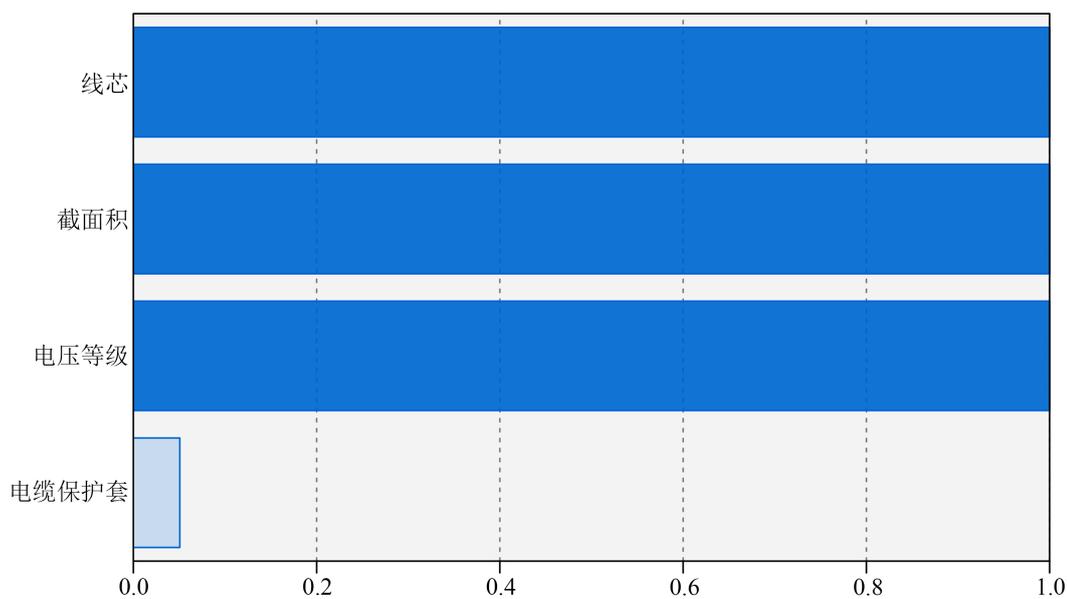


Figure 3. Predictive variable importance map  
图 3. 预测变量重要性图

**Table 5.** Classification table for waste power cables  
**表 5.** 废旧电力电缆分级表

电压等级	截面积	线芯	含铜率	分类
低压电力电缆 (0.6/1 kV)	16	4	59.77%	2
	35	4	69.10%	2
	70	4	73.45%	1
	120	4	66.20%	1
	185	4	79.90%	1
	240	4	81.80%	1
AC 10 kV	70	3	47.80%	4
	120	3	53.50%	3
	240	3	62.70%	2
	400	3	67.60%	2
AC 35 kV	240	1	53.32%	3
	240	3	54.06%	3
	400	3	54.30%	3
	630	1	66.60%	2
AC 110 kV	400	1	41.06%	4
	630	1	43.90%	4
	800	1	54.82%	3
	1000	1	56.88%	3
	1200	1	59.34%	2
AC 220 kV	800	1	41.59%	4
	1000	1	46.69%	4
	1200	1	50.04%	4

#### 4.2. 预测模型构建

上方的分类结果仅仅是对历史数据进行的分类，也许并不能很好的对新加入的数据进行一个预测分类。基于此本文还需对新样本进行预测，本文采用决策树进行预测模型的构建，其优势在于可以根据每个特征的不同取值对废旧电缆进行精准分类，并且可视化程度较高。

决策树一般分为两大类型，分类决策树主要用于对离散因变量的分类，回归决策树主要用于对连续因变量的预测，由于废旧电缆的特征指标均为定性变量，故而选用决策树中的分类树方法完成对新样本的预测工作。

模型的算法本文采用 CHAID (卡方自动交互检测)生长法,其原理是在每一步都选择与因变量有最强交互作用的自变量,如果依据这一自变量划分的类别与因变量实际类别不存在显著不同,那么将合并这些类别。最终输出如下图 4 所示的决策树结构。

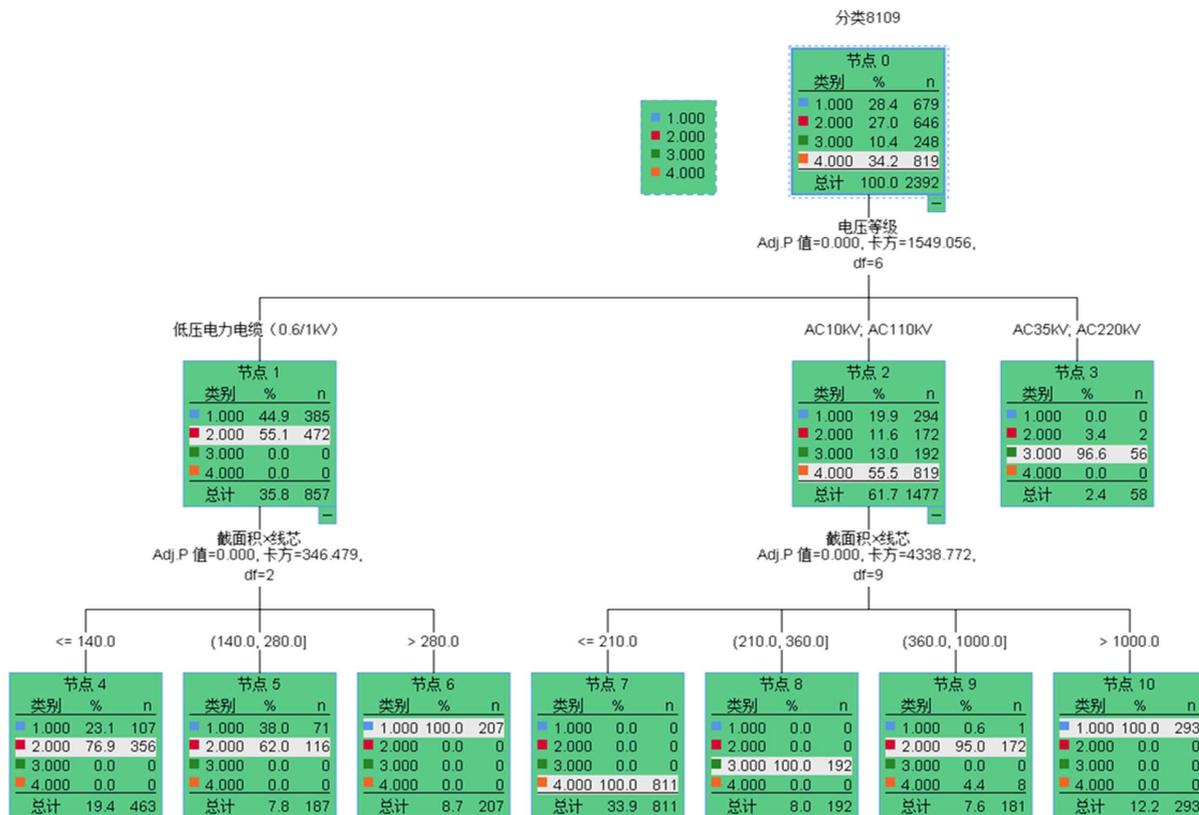


Figure 4. Decision tree structure

图 4. 决策树结构

由上图可见:根节点处的因素为电压等级,若电压等级为低压电缆则走向左边的节点 1,根据截面与线芯之积大小继续分类,若截面小于等于 140,则走向节点 4,分为第 2 类;截面在 140 至 280 之间则走向节点 5,若截面大于 280 则走向节点 6,分为第 1 类;若电压等级为 10 kV、110 kV,则走向节点 2,根据截面积大小继续分类,直至所有节点的类别都为占比百分之百的类别,则停止生长,实现对电力电缆的完全分类。

## 5. 模型检验

通过模型对物资进行分类以及预测后,需要对模型的准确度进行检验。模型检验可以帮助我们确定模型是否准确地反映了数据的结构,以及是否存在任何误分类以及帮助我们评估模型的效果。因此需要确定模型的分类结果是否符合我们的实际需求,并且是否能够为我们提供有用的信息。如果模型的效果不佳,那么我们需要对模型进行改进,以提高其分类准确度和预测能力。

首先对二阶聚类模型进行质量检验,本文对聚类结果的评估选用模型轮廓测量系数。轮廓系数结合了聚类的凝聚度和分离度,其可用来评估聚类的效果。该系数取值区间为 $[-1, 1]$ ,值越大,则表明该聚类结果越良好。

由 SPSS 计算出聚类模型的轮廓系数并生成可视化图，如图 5 所示，可见，轮廓系数的数值较大，接近 1，聚类模型分类结果较好。

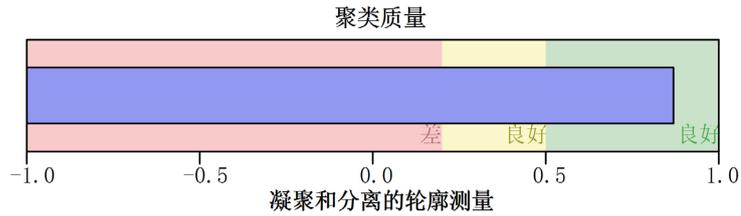


Figure 5. Model quality inspection chart  
图 5. 模型质量检验图

下面对决策树模型的预测性能进行检验，决策树的预测性能是指在生成树的过程中，对样本实际分类与生成树中的分类进行对比，从而计算出模型的正确百分比。为保证训练样本充分性，本文选取 70% 的样本数据作为模型构建的训练样本，再对剩余 30% 的样本数据进行模型检验。以此来对决策树预测性能进行梳理，得出预测性能表如表 6 所示。

Table 6. Predictive performance table  
表 6. 预测性能表

	实测	分类				正确百分比
		预测				
		1	2	3	4	
样本训练	1	500	179	0	0	73.6%
	2	0	644	2	0	99.7%
	3	0	0	248	0	100.0%
	4	0	8	0	811	99.0%
	总体百分比		20.9%	34.7%	10.5%	33.9%
检验	1	187	87	0	0	68.2%
	2	0	296	0	0	100.0%
	3	0	0	118	0	100.0%
	4	0	6	0	369	98.4%
	总体百分比		17.6%	36.6%	11.1%	34.7%

生长法: CHAID  
因变量: 分类 8109

检验结果显示：训练样本预测正确率为 92.1%，检验样本预测正确率为 91.3%，正确率均在 90% 以上，可见模型预测效果较好。

本文对决策树模型进行风险检测，产出风险统计表如表 7 所示。从风险表格中可以看出数据的估算

值以及标准误差，估算值是依据输入的因变量求和得到，而标准误差则是具体的因变量数据与估算值之间的差值。一般来说，标准误差值越小，数据风险也就越小。结果表明，训练样本和检验样本风险均小于 0.05，可见模型构建效果较为优良。

**Table 7.** Risk statistics table  
**表 7.** 风险统计表

样本	风险	
	估算	标准误差
训练	0.079	0.006
检验	0.087	0.009

生长法: CHAID  
因变量: 二阶聚类编号

## 6. 结语

本文对废旧电缆的分级分类进行了深入的研究，主要工作及成效如下：

1) 通过梳理废旧电力电缆各维度的属性信息，建立废旧电缆基于含铜率的评价体系，通过方差分析法提取对含铜率具有影响的特征指标，结果为电压等级、截面积、线芯。

2) 通过二阶聚类法将现有废旧电缆种类分为四个级别，含铜率在 70% 以上的为 A 级别；含铜率 60%~70% 为 B 级别；含铜率在 50%~60% 为 C 级别；含铜率在 50% 以下的为 D 级别。并生成不同种类废旧电力电缆分级表。

3) 通过决策树模型对所有型号废旧电缆进行分级结果预测，产生具有高可视化程度的决策树结果图，以实现废旧电缆的完全分类。

4) 通过对分类结果进行模型质量检验，生成模型质量检验图与预测性能及风险统计表，表明二阶聚类模型质量良好，决策树模型预测性能正确率在 90% 以上，风险小于 0.05。

需要注意的是，分级后的废旧电缆若要达到处置效率及效益的提升，必须依据含铜率的不同建立合理的价值评价体系，这也将是下一步的研究重点。

## 参考文献

- [1] 韩赛, 卢建军, 卫晨, 刘志鹏. 基于概率神经网络的煤炭企业物资分类方法研究[J]. 工矿自动化, 2014, 40(4): 38-41.
- [2] 魏伟, 高赐威, 宋梦, 明昊. 基于 HHO-SVM 的电力线路故障检测分类系统[J]. 综合智慧能源, 2023, 45(3): 1-8.
- [3] 王庆荣, 赵小柠, 杨景玉. 基于改进 K 均值聚类 POS 算法的应急物资分类研究[J]. 兰州交通大学学报, 2014, 33(3): 213-216+222.
- [4] 江玮璠. 基于模糊聚类分析的多准则 ABC 库存管理[J]. 物流技术, 2009, 28(1): 97-98+139.
- [5] Havens, T., Chitta, R., *et al.* (2011) Speedup of Fuzzy and Possibilistic Kernel c-Means for Large-Scale Clustering. 2011 *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, Taipei, 27-30 June 2011, 463-470. <https://doi.org/10.1109/FUZZY.2011.6007618>
- [6] Cordeiro, R.L.F., Traina, C., *et al.* (2011) Clustering Very Large Multi-Dimensional Datasets with MapReduce. In *KDD 11 Conference Committee, Ed., Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '11)*, Association for Computing Machinery, New York, 690-698.