文章编号:1674-7607(2026)10-0009-09 DOI:10.19805/j.cnki.jcspe.2026.250284

基于 Apriori 算法优化神经网络输入特征的煤电机组集可靠性预测

揭新世[1]， 张晓璐[2]， 陈晓峰[2]， 乔登科[3]， 肖显斌[1,4]，
李建锋[5]， 李媛园[2]， 田 宇[6]
(1. 华北电力大学 新能源学院，北京 102206；2. 华北电力科学研究院有限责任公司，
北京 100045；3. 国家电网有限公司华北分部，北京 100031；
4. 怀柔实验室山西研究院，山西太原 030032；5. 中国电力企业联合会，北京 100761；
6. 山西国峰煤电有限责任公司，山西吕梁 032299)

摘 要：随着煤电转型的发展，电网需要进行大规模调度等宏观决策的场合不断增加，以往针对单机组的可靠性预测方法无法满足电网对机组集可靠性信息的需求。为了准确高效地进行煤电机组集可靠性预测，提出了一种基于 Apriori 算法优化 BP 神经网络输入特征(Apriori-BP)的机组集可靠性预测模型。Apriori 算法根据影响因素对预测目标的影响衡量它们之间的关联，能更准确地衡量复合型影响因素整体与预测目标的关联度，从而更好地优化模型输入。将 Apriori-BP 预测模型应用于实际电网，对多个不同机组集的可靠性进行预测，并与灰色关联分析-BP 模型、主成分分析-BP 模型的预测误差进行对比。结果表明：Apriori-BP 模型具有最高的预测精度，均方根误差和平均绝对误差分别为 0.021 4 和 0.009 7，且泛化能力和鲁棒性更好。

关键词：可靠性预测；模型输入优化；Apriori 算法；BP 神经网络；机组集特征提取
中图分类号：TM621 **文献标志码**：A **学科分类号**：470.30

Reliability Prediction of Coal-fired Power Unit Set Based on Apriori Algorithm Optimized Neural Network Input Feature

JIE Xinshi[1]， *ZHANG Xiaolu*[2]， *CHEN Xiaofeng*[2]， *QIAO Dengke*[3]， *XIAO Xianbin*[1,4]，
LI Jianfeng[5]， *LI Yuanyuan*[2]， *TIAN Yu*[6]
(1. School of New Energy, North China Electric Power University, Beijing 102206, China;
2. North China Electric Power Research Institute Co., Ltd, Beijing 100045, China;
3. North China Branch of State Grid Corporation of China, Beijing 100031, China;
4. Shanxi Research Institute of Huairou Laboratory, Taiyuan 030032, Shanxi Province, China;
5. China Electricity Council, Beijing 100761, China; 6. Shanxi Guofeng Coal Power Co., Ltd,
Lvliang 032299, Shanxi Province, China)

Abstract: With the development of coal-fired power transformation, there is an increasing need for large-

收稿日期:2025-05-20 **修订日期**:2025-06-30
基金项目:国家重点研发计划项目(2022YFB4100302)
作者简介:揭新世(2000—)，男，广西防城港人，硕士研究生，研究方向为煤电机组可靠性分析及预测。
肖显斌(通信作者)，男，教授，博士，E-mail：xiaoxianbin@ncepu.edu.cn。

scale scheduling and other macro decision-making in the power grid. Previous reliability prediction methods for single units cannot meet the power grid's demand for reliability information of unit sets. In order to accurately and efficiently predict the reliability of coal-fired power unit sets, a reliability prediction model of unit sets based on Apriori algorithm optimized back propagation (BP) neural network input (Apriori-BP) was proposed. The Apriori algorithm measured the correlation between the influencing factors and the prediction target based on their impact, which could more accurately measure the overall correlation between the composite influencing factors and the prediction target, thereby better optimizing the model input. Apply the Apriori-BP prediction model to the actual power grid, predict the reliability of multiple different sets of units, and compare the prediction errors with the grey relational analysis (GRA)-BP model and principal components analysis (PCA)-BP model. Results indicate that the Apriori-BP model has the highest prediction accuracy, with root mean square error and mean absolute error of 0.021 4 and 0.009 7, and better generalization ability and robustness.

Key words: reliability prediction; model input optimization; Apriori algorithm; BP neural network; unit set feature extraction

煤电转型的发展给煤电机组可靠运行带来了新的挑战,更对机组可靠性提出了更高的要求。“十三五”以来,我国积极推进电力清洁能源转型,可再生能源占比不断提高。由于风、光等可再生能源具有随机性、波动性和间歇性,其大规模并网会提高源、荷两端不确定性[1]。为保障供电稳定性,煤电正从传统的主力电源向承担"兜底保障"功能转变,成为提供可靠容量及调峰调频服务的“兜底”电源[2],这一转型使得煤电机组需要更频繁地启停、更宽范围地调节负荷,不仅加剧了设备损耗,更对机组的快速响应能力、运行灵活性和供电可靠性提出了前所未有的高标准。

目前的预测方法往往只针对单台机组的可靠性进行预测,无法满足电网进行宏观决策时的需求。虽然现有的可靠性预测研究已取得显著进展,有助于电厂优化运行策略和检修方案,并为电网调度提供参考,但这些方法仍存在局限性。《国家能源局关于加强电力可靠性数据治理 深化可靠性数据应用发展的通知》提出,要深化可靠性数据应用,推动电力可靠性动态评价,充分发挥数据的支撑作用 。数据驱动模型在处理复杂系统和大规模数据方面具有显著优势,相较于传统方法在故障预测准确性等方面有大幅提升效果[3]。当前煤电行业基于数据驱动的可靠性预测研究可分为以下两类。一为基于运行参数的预测研究。文献[4]基于长短期记忆神经网络构建运行参数的预测模型,并提出了 2 种可靠性评估方法,将两者结合可实现设备的可靠性预测,使用图论法建立机组和设备的可靠性关系后,可通过预测设备的运行参数实现机组的可靠性预测。文献[5]建立了基于多变量状态估计技术的故障预警模型,在此基础上提出了一种多监测参数的机组短期可靠性加权评估模型,并采用基于动态寻优的灰色预测法对异常监测参数进行短期预测,根据预测值计算机组未来的可靠性序列。二为基于可靠性统计数据的预测研究。文献[6]对 14 台 600 MW 机组 12 a 的可靠性数据进行统计分析,基于最小二乘法进行数学建模,考虑了运行时间对该模型的影响,用威布尔分布对非计划停运小时数进行拟合,得到可靠度函数后再对原数学模型进行修正,修正后的预测最小误差达 0.8%。文献[7]提出用检修系数表征机组可靠性变化规律,发现该系数符合幂函数的数学模型,可以较好地对运行可靠性数据进行拟合。文献[8]在此基础上改进了符合机组可靠性变化规律的数学模型,根据该模型和计划停运系数可求出等效可用系数。用该方法分别对不同容量的 8 台机组的可靠性数据进行预测,所得预测值相对误差在 −1.48%～2.69%。这些方法虽然预测精度较高,但均局限于单台机组的预测范畴。随着煤电向基础保障性和系统调节性电源转型发展,电网需要进行大规模调度、产能规划等宏观决策的场合不断增加[9],需要可靠性预测技术来辅助决策。

相比单机组可靠性预测技术,基于机组集历史信息的可靠性预测技术更符合宏观决策需求。本文提出了一种基于 Apriori 算法优化 BP 神经网络输入特征的机组集可靠性预测方法。首先利用 Apriori 算法对可靠性数据进行挖掘,得到各影响因素及

其对机组可靠性的影响程度，依据影响程度大小筛选模型输入特征，再将其代入 BP 神经网络进行训练，根据预测结果找出最优输入特征，构建最优机组集可靠性预测模型。所提基于 Apriori 算法的输入特征优化方法为输入特征筛选提供了一种思路；同时，基于历史信息建立集群可靠性预测框架，可为宏观决策提供可靠依据。

1　关联规则与 Apriori 算法

1.1　关联规则

关联规则挖掘是数据挖掘的一种重要方法，最早由 Agrawal 等[10]于 1993 年提出。关联规则涉及的某一事物称为项，包含了 0 个或者多个项的集合则为项集，项集的长度为项的个数，若某项集中含有 k 个项，则称该项集为 k-项集。关联规则是项集与项集之间的关系，假设有两个相互独立的项集 X 和 Y，它们之间的关联规则可表示为 $X\Rightarrow Y$，前后项交换后表达的关联规则不同。关联规则中有 2 种基本评价指标：支持度、置信度。其中，支持度为某个规则或项集出现的频繁程度，置信度为前项出现时后项也出现的频繁程度，可理解为在前项出现条件下后项出现的条件概率[11]。本文除了以上 2 种基本指标外，还引入了提升度作为评价指标。

支持度对项集的表达式为

$$\mathrm{Support}(X)=\frac{\mathrm{count}(X)}{N} \tag{1}$$

支持度对关联规则的表达式为

$$\mathrm{Support}(X\Rightarrow Y)=\frac{\mathrm{count}(X\bigcap Y)}{N} \tag{2}$$

式中：$\mathrm{Support}(X)$、$\mathrm{Support}(X\Rightarrow Y)$ 分别为项集 X 的支持度以及关联规则 $X\Rightarrow Y$ 的支持度；$\mathrm{count}(X)$ 为项集 X 在数据库中出现的次数；$\mathrm{count}(X\bigcap Y)$ 为 $X\bigcap Y$ 在数据库中出现的次数；N 为数据库中的数据总数。

置信度为

$$\mathrm{Confidence}(X\Rightarrow Y)=\frac{\mathrm{count}(X\bigcap Y)}{\mathrm{count}(X)} \tag{3}$$

提升度为

$$\mathrm{Lift}(X\Rightarrow Y)=\frac{\mathrm{Confidence}(X\Rightarrow Y)}{\mathrm{Support}(Y)} \tag{4}$$

式中：$\mathrm{Confidence}(X\Rightarrow Y)$、$\mathrm{Lift}(X\Rightarrow Y)$ 分别为关联规则 $X\Rightarrow Y$ 的置信度和提升度。

提升度是前项出现对后项出现的影响程度，反映了前后项的相关程度，以 1 为界，提升度越大，正相关性越强，反之负相关性越强，当提升度等于 1 时，认为前后项间没有相关性[12]。

1.2　Apriori 算法

Apriori 算法是关联规则挖掘中较为常用的一种基于广度优先的算法，它的核心思想是逐层搜索，先生成频繁项集，然后再从频繁项集中寻找关联规则。它的具体步骤可分为：(1)扫描数据集，找出所有 1-项集和并求出它们的支持度，然后根据最小支持度筛选频繁 1-项集；(2)在频繁 1-项集的基础上拼接产生 2-项集，根据最小支持度筛选频繁 2-项集；(3)在之前产生频繁项集中拼接产生 3-项集，然后筛选频繁 3-项集，循环直到找不到符合要求的项集；(4)根据求得频繁项集生成关联规则，并计算相关指标[13]。算法流程图如图 1 所示。

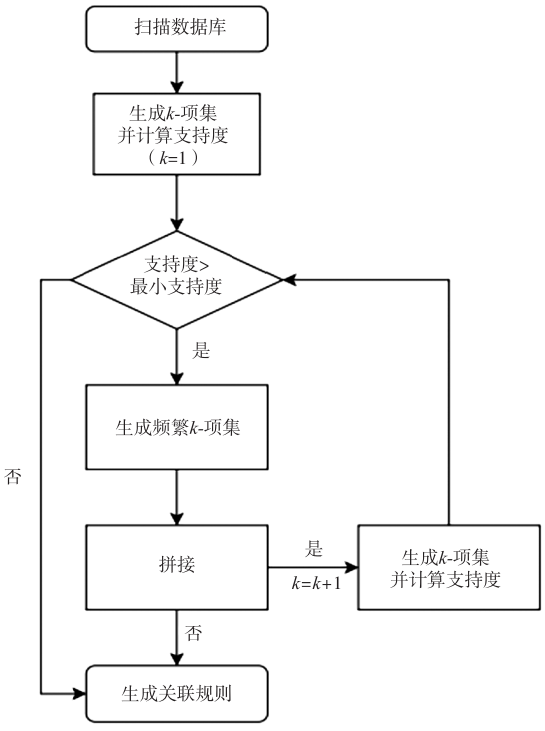


图 1　Apriori 算法流程

Fig. 1　Apriori algorithm process

1.3　基于 Apriori 算法的机组集特征提取

1.1 节中提升度可以反映关联规则中前项对后项的影响程度，当关联规则中前项为机组可靠性影响因素，后项为机组可靠性指标时，该关联规则计算出的提升度可以代表影响因素对机组可靠性的影响。由于影响有正向和负向，且某一影响因素的不同项对可靠性的影响不一定相同(不都是正/负向)，本文提出相对提升度来量化影响因素对可靠性的影

响程度。相对提升度的计算公式如下。

相对提升度(对某一项)为

$$\text{Relative Lift}(X\Rightarrow Y)=\left|\text{Lift}(X\Rightarrow Y)-1\right| \tag{5}$$

相对提升度(对某一影响因素)为

$$\text{Relative Lift}(X\Rightarrow Y)=\frac{\sum_{i=1}^{n}\text{Relative Lift}(X_i\Rightarrow Y)}{n} \tag{6}$$

式中:此处 X 代表影响因素;X_i 为影响因素 X 中的某一项;n 为包含在影响因素 X 中的项的总数;此处 Y 代表可靠性指标。

相对提升度越大,说明提升度相对于 1 的差值越大,表明对应的关联规则中,前项对后项的影响程度越大。而对于某个影响因素,只要求出所有以该影响因素为前项,可靠性指标为后项的关联规则的相对提升度,取平均值,就可以得出该影响因素对可靠性指标的相对提升度,实现某因素影响程度的量化,从而找出强影响因素。

2 BP 神经网络

BP 神经网络是一种正向传播与误差逆向传播的多层前馈神经网络,由输入层、隐含层与输出层构成。在训练过程中,在网络输入层输入输入值,经过隐含层后在输出层输出预测值,若预测值与实际值出现误差则进行逆向传播,在迭代步数内不断调整阈值和权值,使误差尽可能变小[14]。

3 基于 Apriori-BP 的机组集可靠性预测模型

3.1 Apriori 算法优化 BP 神经网络输入特征的可靠性预测模型

在建立可靠性预测模型时,使用 Apriori 算法优化模型输入特征,可降低模型训练的复杂度,减少模型训练时间,删去影响程度较小的输入特征还能防止输入冗余,提高模型的预测精度。模型构建流程如下:①对数据进行预处理,删去空值、奇异值等,离散化数据,方便 Apriori 算法挖掘有用信息;②使用 Apriori 算法挖掘数据,计算出不同影响因素对机组可靠性的影响程度(相对提升度)并进行排序;③整理输入数据后将训练集部分代入 BP 神经网络进行训练;④用测试集检验训练完成的模型并记录预测误差;⑤对比预测误差,若误差小于等于上一次预测的误差,则转到⑥,否则跳转到⑦(第一次预测直接到⑥);⑥删去影响程度最小的输入特征,返回③;⑦模型构建完成。具体的构建流程图如图 2 所示。

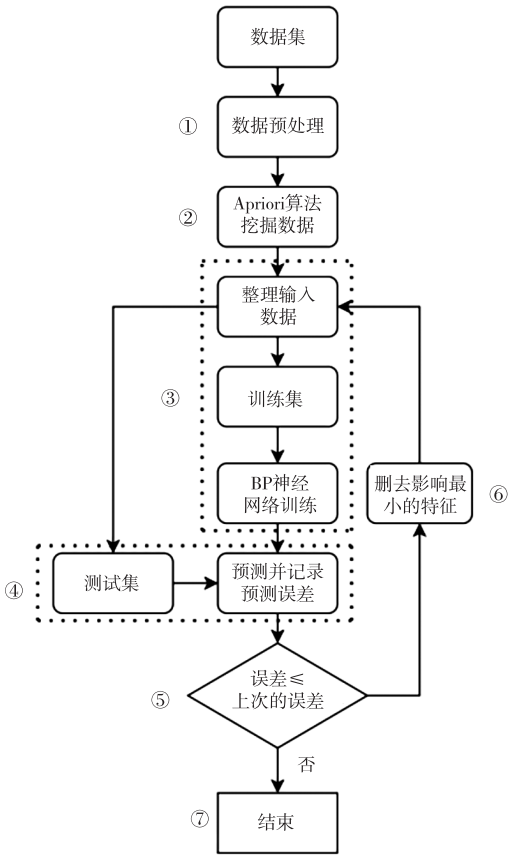


图 2 Apriori-BP 预测模型构建流程

Fig. 2 Apriori-BP prediction model construction process

3.2 模型评价指标

为比较模型性能,选用均方根误差 e_{RMSE}(root mean square error,RMSE)和平均绝对误差 e_{MAE}(mean absolute error,MAE)作为模型评价指标,它们的计算公式分别为

$$e_{\text{RMSE}}=\frac{\sqrt{\sum_{i=1}^{S}(\hat{z}_i-z_i)^2}}{S} \tag{7}$$

$$e_{\text{MAE}}=\frac{\sum_{i=1}^{S}\left|\hat{z}_i-z_i\right|}{S} \tag{8}$$

式中:$\hat{z}$ 为预测值;z 为实际值;S 为测试集中数据样本的总数。

这两个指标越小,表明模型预测的精度越高。

4 案例分析

实验数据取自某区域 2017—2021 年机组可靠性统计数据,可分为机组台账和机组运行事件统计两部分,其中机组台账包含统计范围内 1 972 台机组的注册信息,机组运行事件统计则包含了 90 491 条运行事件记录。由于非计划停运对机组正常运行的影响较大,且不受人为调控,是电厂最为关心的事

件之一，可以简单、直接地反映出机组的可靠性，本文选取非计划停运的相关指标作为可靠性指标。统计范围内所有机组构成的机组集称为“全部”机组集，根据数据集计算出“全部”机组集近 5 年的可靠性指标，各可靠性指标的计算公式如式(9)～式(11)所示，计算结果如表 1 所示。

表 1　“全部”机组集近 5 年的可靠性指标

Tab. 1　Reliability indicators of the "all" unit set in the past five years

参数	2017 年	2018 年	2019 年	2020 年	2021 年
统计台数	1936	1917	1922	1940	1972
平均非计划停运次数/(次·台$^{-1}$·a^{-1})	0.60	0.68	0.46	0.45	0.42
平均非计划停运时长/(h·台$^{-1}$·a^{-1})	54.11	96.04	37.51	39.18	94.54
发电能力损失/(MW·h)	4.18×10^7	7.25×10^7	3.01×10^7	3.07×10^7	8.23×10^7

平均非计划停运次数 N_o 为

$$N_o=\frac{\sum_{i=1}^{S_u}n_{oi}}{S_u} \tag{9}$$

平均非计划停运时长 T_o 为

$$T_o=\frac{\sum_{i=1}^{S_u}t_{oi}}{S_u} \tag{10}$$

发电能力损失 E_{loss} 为

$$E_{loss}=\sum_{i=1}^{S_u}W_i\times t_{oi} \tag{11}$$

式中：S_u 为机组集包含的机组数；n_{oi} 为第 i 台机组在一定时期内的总非计划停运次数；t_{oi} 第 i 台机组在一定时期内的非计划停运事件的总持续时间；W_i 为第 i 台机组的额定容量。

由表 1 可知，2017—2021 年间“全部”机组集的平均非计划停运次数总体上呈下降趋势，但平均非计划停运时长以及发电能力损失在 2019—2021 年呈现上升趋势，并在 2021 年出现显著回升，该现象表明该机组集整体可靠性有所下降，应进一步研究该机组集的可靠性。此外，平均非计划停运时长和发电能力损失的最大变化幅度分别达 141% 和 168%，表现出明显的突变特征，预测的不确定性较大，而非计划停运次数的的最大变化幅度仅为 32%。鉴于当前针对机组集可靠性事件的预测研究较少，本文后续主要针对非计划停运次数的相关数据进行预测。

4.1　数据预处理

原始数据中存在含空值、奇异值、无效值的问题数据，除此之外还有一些与机组可靠性没有直接关联的信息，包括一些已经停止统计的机组相关信息，在使用前应删去这部分信息。由于 Apriori 算法属于布尔型关联规则算法，它只关心数据存在与否，不考虑数据本身的意义，只能处理布尔型变量，而且原始数据中的数据较为分散，每项对应的样本量过小，计算后所得结果的随机性较大，没有实际意义，需要将数据离散化。本文将机组容量分为大、中、小型；行政归属分为集团 A、集团 B 等 8 类；所在地分为区域 A、区域 B 等 8 类；非计划停运次数分为 0 次、1 次、2 次及以上；服役年数分为青年、中年、老年等，具体分类如表 2 所示。

表 2　数据离散化分类表

Tab. 2　Discretization classification table of data

数据类型	分类依据	具体分类
机组容量	按机组容量大小：<300 MW 为小型；[300，600) MW 为中型；≥600 MW 为大型	大型、中型、小型
行政归属	按机组所属发电集团：如民企、地方国企、5 大发电集团等央企，共 8 类	集团 A、集团 B、集团 C、集团 D、集团 E、集团 F、集团 G、集团 H
所在地	按机组所在地区：如东北、华北、华中等地区，共 8 类	区域 A、区域 B、区域 C、区域 D、区域 E、区域 F、区域 G、区域 H
服役年数	按机组服役年数：<10 a 为青年；[10，20]a 为中年；>20 a 为老年	青年、中年、老年
发电机厂家	按机组发电机厂家：如三大发电机厂、外国发电机厂、国内其他发电机厂，共 5 类	发电机厂家 A、发电机厂家 B、发电机厂家 C、发电机厂家 D、发电机厂家 E
锅炉厂家	按机组锅炉厂家：如三大锅炉厂、外国锅炉厂、国内其他锅炉厂，共 5 类	锅炉厂家 A、锅炉厂家 B、锅炉厂家 C、锅炉厂家 D、锅炉厂家 E

续表

数据类型	分类依据	具体分类
汽轮机厂家	按机组汽轮机厂家：如三大汽轮机厂、外国汽轮机厂、国内其他汽轮机厂，共 5 类	汽轮机厂家 A、汽轮机厂家 B、汽轮机厂家 C、汽轮机厂家 D、汽轮机厂家 E
变压器厂家	按机组变压器厂家：如三大变压器厂、外国变压器厂、国内其他变压器厂，共 5 类	变压器厂家 A、变压器厂家 B、变压器厂家 C、变压器厂家 D、变压器厂家 E
一年内的检修次数	按机组一年内检修次数：0 为 0 次；1 为 1 次；≥2 为 2 次及以上	0 次、1 次、2 次及以上
一年内的非计划停运次数	按机组一年内非计划停运次数：0 为 0 次；1 为 1 次；≥2 为 2 次及以上	0 次、1 次、2 次及以上

4.2 特征参数提取

本文采用 Apriori 算法对可靠性统计数据进行分析，并量化各个影响因素对机组可靠性的影响程度。以相对提升度表示影响程度大小，相对提升度越大的影响因素越适宜作为机组集可靠性预测的特征参数，各影响因素对非计划停运次数的相对提升度如图 3 所示。

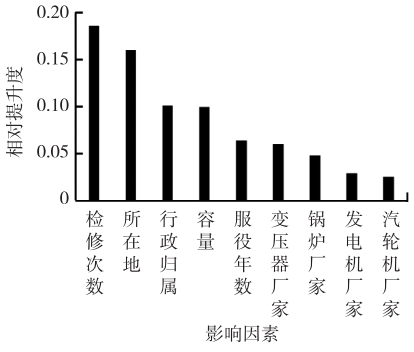


图 3 各影响因素相对提升度计算结果

Fig. 3 Calculation results of relative lift of various influencing factors

由图 3 可知，所有影响因素的相对提升度都大于 0，说明它们均对可靠性有一定影响，且影响程度为：检修次数＞所在地＞行政归属＞容量＞服役年数＞变压器厂家＞锅炉厂家＞发电机厂家＞汽轮机厂家。

4.3 神经网络输入数据库构建

由于本文选取的影响因素中包含所在地等非数据型特征，在将数据传入神经网络进行训练前需要将数据整理成模型可识别的形式。经整理后，输入层的输入特征为机组集内不同类别机组（如青年机组）的数量占机组总数的百分比和机组的平均年龄，输出层的预测值为机组集内机组一年内发生 n 次非计划停运的概率（$n=0,1,2,\ldots,10$）。将统计范围内的机组分成多个机组集，分别统计上述数据，构成神经网络输入数据库。并将 2017—2020 年的数据划为训练集，2021 年的数据划为测试集。

4.4 模型参数设置

BP 神经网络中部分参数的取值缺乏明确的理论依据，建模时通常参考前人经验或经反复试验确定[15]。本文采用控制变量法确定隐含层神经元和模型最大迭代步数的取值，在其他参数不变的情况下，每次只改变其中某一个参数的取值进行建模、训练，通过比较不同模型测试集预测误差找出性能最佳的模型参数取值。不同隐含层神经元数、不同迭代步数取值模型的预测误差如图 4、图 5 所示。

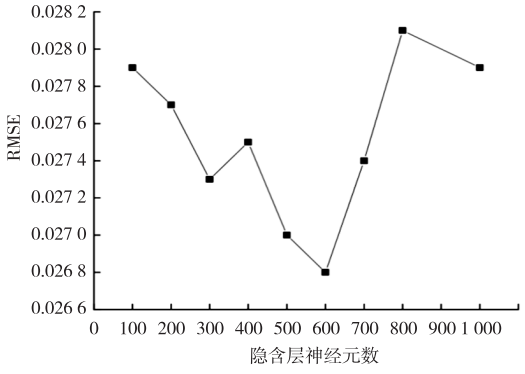


图 4 隐含层神经元数不同时模型预测误差

Fig. 4 Prediction error of models with different numbers of hidden layer neurons

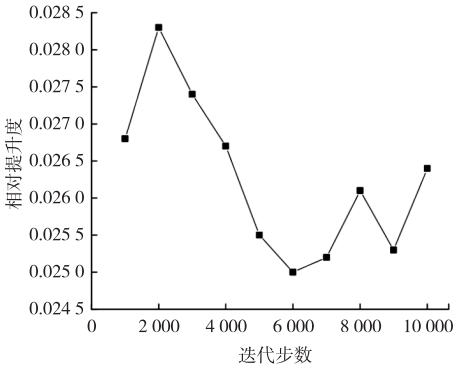


图 5 迭代步数不同时模型预测误差

Fig. 5 Model prediction error with different iteration steps

从图 4、图 5 可以看出，隐含层神经元数和迭代步数分别为 600、6 000 时，模型的预测误差最小，因

此后续 BP 神经网络隐含层神经元数统一设为 600，迭代步数设为 6 000。本文 BP 神经网络结构图如图 6 所示。

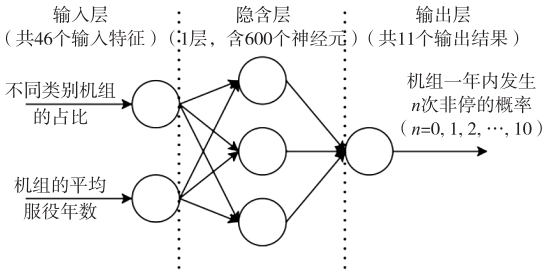


图 6　BP 神经网络结构图

Fig. 6　Structural diagram of BP neural network

在图 6 中，不同类别机组的占比主要包括不同机组容量机组的占比、不同所在地机组的占比等共 9 大类，具体类别如表 2 中前 9 个具体分类所示。

4.5　实验结果与分析

为了提高预测精度，减小预测误差，采用 Apriori 算法优化 BP 神经网络输入，首先将全部机组集特征输入 BP 神经网络进行训练、预测，然后逐个删去相对提升度最小的机组集特征。在不改变模型结构、模型参数条件下，比较预测误差，如表 3 所示。

表 3　机组集特征提取对预测误差的影响

Tab. 3　Impact of unit set feature extraction on prediction errors

输入数据集	e_{RMSE}	e_{MAE}
①：完整输入数据集	0.025 0	0.010 5
②：删去汽轮机厂家占比	0.021 4	0.009 7
③：在②的基础上再删去发电机厂家占比	0.022 7	0.010 0
④：在③的基础上再删去锅炉厂家占比	0.024 7	0.010 4

从表 3 可以看出，删去汽轮机厂家占比后的输入数据集为最优输入数据集，使用 Apriori 算法优化模型输入可明显降低预测误差，优化后模型预测误差中 e_{RMSE} 相对优化前降低了 14.4%，e_{MAE} 相对降低了 7.6%。

为比较不同方法优化输入的差异，分别采用灰色关联分析（grey relational analysis，GRA）和主成分分析法（principal components analysis，PCA）对输入进行优化，并对同一机组集进行预测。基于 GRA 计算数据集中各影响因素与机组可靠性的关联度，结果如表 4 所示。

表 4　机组可靠性影响因素关联度排序

Tab. 4　Correlation degree ranking of factors affecting unit reliability

影响因素	关联度
检修次数	0.985 6
容量	0.980 8
服役年数	0.978 2
变压器厂家	0.973 4
行政归属	0.971 2
锅炉厂家	0.969 8
发电机厂家	0.965 6
汽轮机厂家	0.965 6
所在地	0.965 5

其中，关联度大于 0.7 的影响因素为强关联因素，小于 0.3 的影响因素为弱相关因素[16]。选择关联度大于 0.7 的影响因素作为输入，根据表 4 中关联度的计算结果进行筛选，不剔除任一影响因素，输入数据集为完整输入数据集。

对数据集中各影响因素与机组可靠性进行主成分分析，结果如表 5 所示。

表 5　主成分对应的特征值及其方差贡献率

Tab. 5　Eigenvalues and variance contribution rates of principal components

各主成分	特征值	贡献率/%	累积贡献率/%
锅炉厂家	4.723	51.148	51.148
行政归属	1.201	20.805	71.954
服役年数	1.913	12.428	84.381
所在地	0.665	11.529	95.910

当样本的累积贡献率达到 85%以上时，认为这些主成分信息可以替代原始信息，从而完成数据降维[17]。从表 5 可以得出，锅炉厂家、行政归属、服役年数、所在地这四类影响因素对原始数据的累积贡献率达 95.910%，因此选择这四类影响因素作为模型输入数据集。

为评估不同优化方法对模型性能的影响，采用 3 组经不同方法优化的输入数据集为输入特征，基于 BP 神经网络对同一机组集进行预测。3 种输入优化方法对应的模型预测误差如表 6 所示。

表 6　同一机组集预测误差对比

Tab. 6　Prediction error comparison of the same unit set

预测模型	e_{RMSE}	e_{MAE}
Apriori-BP	0.021 4	0.009 7
GRA-BP	0.025 0	0.010 5
PCA-BP	0.029 9	0.011 9

从表 6 可以看出，相较于 GRA 和 PCA 优化输入后的模型，Apriori-BP 模型展现出显著的性能优势，其 e_{RMSE} 降低了 14.4%、28.43%，e_{MAE} 降低了 7.6%、18.5%，表明使用 Apriori 算法优化输入的效果最好。这是因为本文所选的影响因素主要为所在地等机组属性，属于复合型影响因素，其中包含多个项，如所在地 A、所在地 B 等，影响因素整体与可靠性指标间的数学关系较为复杂，而 PCA 和 GRA 这两种传统优化输入方法主要从数学角度对影响因素进行分析，根据特征与预测目标的某些数学关系来衡量它们之间的关联，如 PCA 倾向于保留数据集中对方差贡献大的特征[18]，GRA 则是根据特征与预测目标发展趋势的相似或相异程度衡量它们之间的关联[19]。采用这两种方法进行输入优化可能导致特征误选或信息损失。相比之下，Apriori 算法主要从统计学的角度对影响因素进行分析，根据影响因素出现对可靠性指标的影响衡量它们之间的关联，相对提升度可以准确地描述复合型影响因素整体与可靠性指标的关联，从而更好地实现输入优化。

为了验证模型的泛化能力，采用 3 组经不同方法优化的输入数据集为输入特征，基于 BP 神经网络对 30 个不同的机组集进行预测。3 种输入优化方法对应的模型预测误差如图 7 所示，误差变动范围如表 7 所示。

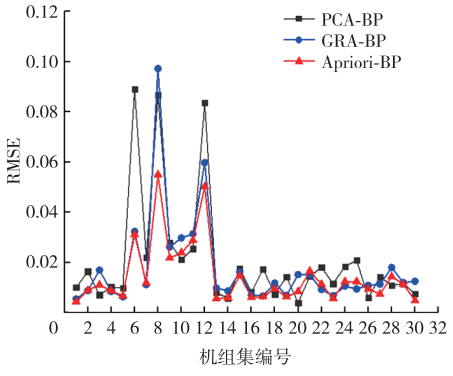


图 7 不同机组集预测误差

Fig. 7 Prediction errors of various unit sets

表 7 不同机组集预测误差变化幅度对比

Tab. 7 Variation range comparison of prediction errors by different models of various unit sets

预测模型	e_{RMSE} 变化幅度[1)]	e_{MAE} 变化幅度
Apriori-BP	0.050 6	0.025 8
GRA-BP	0.091 7	0.045 0
PCA-BP	0.085 1	0.048 2

注：1)变化幅度＝最大误差－最小误差。

采用同一模型对不同机组集进行可靠性预测时，误差有所差异。预测误差的最大值减去最小值得到误差变化幅度。变化幅度越小，表明模型的鲁棒性越强。从图 7 和表 7 可以看出，不同方法优化输入后的模型对不同机组集的预测误差差距较大，其中 Apriori-BP 模型对预测范围内 70% 的机组集的 RMSE 要低于 PCA-BP 模型，对 73.3% 的机组集的 RMSE 低于 GRA-BP 模型，且误差波动更小，表明 Apriori-BP 模型的泛化能力最好。在对 8 号和 12 号机组集进行预测时，模型误差显著增大，此时 PCA-BP 模型的 e_{RMSE} 达 0.086 7、0.083 7，GRA-BP 模型的 e_{RMSE} 达 0.097 2、0.057 8，而 Apriori-BP 模型的 e_{RMSE} 仅为 0.054 9、0.050 2，表明 Apriori-BP 模型在面对异常值时的预测性能更好，鲁棒性更强。

5 结论

(1) 所提出的基于 Apriori 算法的输入特征优化方法能更准确地衡量复合型影响因素整体与预测目标的关联度，从而更好地优化模型输入。

(2) 使用 Apriori 算法优化 BP 神经网络输入特征对模型性能的提升效果更显著，对同一机组集进行预测时，Apriori-BP 模型预测的 RMSE 较 GRA-BP 模型降低了 14.4%，较 PCA-BP 模型降低了 28.43%，且在对不同机组集进行预测时表现出更好的泛化能力和鲁棒性。

(3) 在对 30 个不同机组集进行预测时，Apriori-BP 模型预测结果的均方根误差变化幅度最小(0.050 6)，表明该模型的鲁棒性最强，适用于大量机组集的可靠性预测。

参考文献：

[1] 袁家海，苗若兰，张健. 碳中和背景下我国火电行业低碳发展展望[J]. 中国国情国力，2022(12)：9-13.
YUAN Jiahai，MIAO Ruolan，ZHANG Jian. Outlook for low carbon development of China's power industry under the background of carbon neutrality[J]. China National Conditions and Strength，2022(12)：9-13.

[2] 王放放，杨鹏威，赵光金，等. 新型电力系统下火电机组灵活性运行技术发展及挑战[J]. 发电技术，2024，45(2)：189-198.
WANG Fangfang，YANG Pengwei，ZHAO Guangjin，et al. Development and challenge of flexible operation technology of thermal power units under new power system[J]. Power Generation Technology，2024，45

(2)：189-198.

[3] 王晓华，张勤武，乔玉良，等. 电力设备数据驱动的可靠性分析与预测模型研究[J]. 电力设备管理，2024(24)：180-182.
WANG Xiaohua，ZHANG Qinwu，QIAO Yuliang，et al. Research on reliability analysis and prediction model of data driven power equipment[J]. Electric Power Equipment Management，2024(24)：180-182.

[4] 许杨旸. 火电机组实时可靠性评估技术研究[D]. 北京：华北电力大学，2019.
XU Yangyang. Research on the assessment of real-time reliability of thermal power plant[D]. Beijing：North China Electric Power University，2019.

[5] 张蓓. 基于故障预警的火电机组可靠性评估技术研究[D]. 北京：华北电力大学，2018.
ZHANG Bei. Research on reliability evaluation technology of thermal power unit based on fault early warning[D]. Beijing：North China Electric Power University，2018.

[6] 孙奉仲，王凯，黄新元，等. 600 MW 火电机组可靠性数据的分析及其预测模型[J]. 山东大学学报(工学版)，2005，35(5)：42-44，82.
SUN Fengzhong，WANG Kai，HUANG Xinyuan，et al. Reliability data analysis and its prediction model of the 600 MW units in power plants[J]. Journal of Shandong University (Engineering Science)，2005，35(5)：42-44，82.

[7] 史进渊. 发电设备可靠性增长的研究[J]. 动力工程，1991，11(5)：51-53，50-65.
SHI Jinyuan. Study on reliability boost of power plant equipment[J]. Power Engineering，1991，11(5)：51-53，50-65.

[8] 史进渊. 大型发电机组可靠性预测方法与应用[J]. 机械工程学报，2011，47(18)：165-172.
SHI Jinyuan. Reliability prediction methods and application of large capacity generating units[J]. Journal of Mechanical Engineering，2011，47(18)：165-172.

[9] 王森，李凤婷，张高航，等. 面向新能源高渗透系统调峰和调频的系统级储能容量需求分析[J]. 电力自动化设备，2024，44(1)：24-31.
WANG Sen，LI Fengting，ZHANG Gaohang，et al. Capacity demand analysis of system-level energy storage for peak regulation and frequency regulation of power system with high penetration of renewable energy[J]. Electric Power Automation Equipment，2024，44(1)：24-31.

[10] AGRAWAL R，IMIELINSKI T，SWAMI A. Mining association rules between sets of items in large databases[C]//Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data. Washington D C，USA：ACM Press，1993：207-216.

[11] 张维，刘吉臻，高明明. 基于数据挖掘的循环流化床锅炉辅机故障预警[J]. 动力工程学报，2019，39(10)：826-833.
ZHANG Wei，LIU Jizhen，GAO Mingming. Early fault warning of auxiliary equipment for circulating fluidized bed boilers based on data mining[J]. Journal of Chinese Society of Power Engineering，2019，39(10)：826-833.

[12] XIE Dongfan，WANG Meihong，ZHAO Xiaomei. A spatiotemporal Apriori approach to capture dynamic associations of regional traffic congestion[J]. IEEE Access，2020，8：3695-3709.

[13] 李建强，陈星旭，赵凯，等. 基于数据分析的燃煤机组制粉系统与配风方式优化[J]. 动力工程学报，2017，37(11)：876-882，911.
LI Jianqiang，CHEN Xingxu，ZHAO Kai，et al. Optimization on pulverizing system and air distribution method of a coal-fired unit based on data analysis[J]. Journal of Chinese Society of Power Engineering，2017，37(11)：876-882，911.

[14] 钟安德，吴自玉，谢宗效，等. 基于 ACO-BP 神经网络的光伏发电短期功率预测研究[J]. 电子设计工程，2024，32(18)：82-86.
ZHONG Ande，WU Ziyu，XIE Zongxiao，et al. Research on short-term power prediction of photovoltaic power generation based on ACO-BP neural network[J]. Electronic Design Engineering，2024，32(18)：82-86.

[15] 李友坤. BP 神经网络的研究分析及改进应用[D]. 淮南：安徽理工大学，2012.
LI Youkun. Analysis and improvement applications of BP neural network[D]. Huainan：Anhui University of Science & Technology，2012.

[16] 万俊杰，任丽佳，单鸿涛，等. 基于灰色关联分析与 ISSA-LSSVM 的配电网可靠性预测[J]. 控制工程，2023，30(5)：856-864.
WAN Junjie，REN Lijia，SHAN Hongtao，et al. Reliability prediction of distribution network based on grey relational analysis and ISSA-LSSVM[J]. Control Engineering of China，2023，30(5)：856-864.

[17] YANG Linfeng，LI Wei，ZHANG Cuo，et al. A multistage semianticipativity model for UC via distributionally robust optimization combined with principal component analysis[J]. IEEE Transactions on Industrial Informatics，2024，20(2)：2632-2643.

(下转第 51 页)

学报，2023，30(7)：110-117.
LIU Houji，XIE Jian，LI Yiman，et al. Stress relaxation behavior test and creep constitutive equation of GH3536 nickel-based wrought superalloy[J]. Journal of Plasticity Engineering，2023，30(7)：110-117.

[8] 朱擎，董建新，张麦仓，等. GH4738 高温合金的应力松弛影响因素[J]. 稀有金属材料与工程，2017，46(9)：2546-2552.
ZHU Qing，DONG Jianxin，ZHANG Maicang，et al. Stress relaxation behavior of GH4738 superalloy[J]. Rare Metal Materials and Engineering，2017，46(9)：2546-2552.

[9] NECHACHE A，BOUZID A H. The effect of cylinder and hub creep on the load relaxation in bolted flanged joints[J]. Journal of Pressure Vessel Technology，2008，130(3)：031211.

[10] GJESTLAND H，NUSSBAUM G，REGAZZONI G，et al. Stress-relaxation and creep behaviour of some rapidly solidified magnesium alloys[J]. Materials Science and Engineering：A，1991，134：1197-1200.

[11] 江河，董建新，张麦仓，等. 服役条件下镍基高温合金应力松弛微观机制[J]. 金属学报，2019，55(9)：1211-1220.
JIANG He，DONG Jianxin，ZHANG Maicang，et al. Stress relaxation mechanism for typical nickel-based superalloys under service condition[J]. Acta Metallurgica Sinica，2019，55(9)：1211-1220.

[12] 帅词俊，段吉安，王炯. 关于黏弹性材料的广义 Maxwell 模型[J]. 力学学报，2006，38(4)：565-569.
SHUAI Cijun，DUAN Ji'an，WANG Jiong. A method of establishing generalized Maxwell model for viscoelastic material[J]. Chinese Journal of Theoretical and Applied Mechanics，2006，38(4)：565-569.

[13] 郭进全，轩福贞，何磊. 螺栓材料 1Cr10NiMoW2VNbN 的应力松弛行为及预测模型[J]. 核动力工程，2008，29(6)：119-124.
GUO Jinquan，XUAN Fuzhen，HE Lei. Stress relaxation performance and prediction models for bolt material of 1Cr10NiMoW2VNbN[J]. Nuclear Power Engineering，2008，29(6)：119-124.

[14] 常鹏鹏，孔永华，陈国胜，等. GH4169 合金的高温应力松弛行为及蠕变本构方程[J]. 机械工程材料，2015，39(5)：89-92.
CHANG Pengpeng，KONG Yonghua，CHEN Guosheng，et al. Stress relaxation behavior and creep constitutive equation of GH4169 superalloy at high temperature[J]. Materials for Mechanical Engineering，2015，39(5)：89-92.

[15] 朱智，张立文，顾森东. Hastelloy C-276 合金应力松弛试验及蠕变本构方程[J]. 中国有色金属学报，2012，22(4)：1063-1067.
ZHU Zhi，ZHANG Liwen，GU Sendong. Stress relaxation test of Hastelloy C-276 alloy and its creep constitutive equation[J]. The Chinese Journal of Nonferrous Metals，2012，22(4)：1063-1067.

[16] 毛剑峰，王炜哲，张军辉，等. 汽轮机螺栓松弛对汽缸蠕变强度的影响[J]. 动力工程学报，2013，33(2)：107-111.
MAO Jianfeng，WANG Weizhe，ZHANG Junhui，et al. Influence of bolt relaxation on creep strength of turbine casings[J]. Journal of Chinese Society of Power Engineering，2013，33(2)：107-111.

[17] 杨志远. GH4169 合金应力松弛行为及有限元模拟研究[D]. 哈尔滨：哈尔滨工业大学，2020.
YANG Zhiyuan. Research on stress relaxation behavior and finite element simulation of GH4169 alloy[D]. Harbin：Harbin Institute of Technology，2020.

（上接第 17 页）

[18] 黄伟，李常伟，席建忠. 基于 PCA-MEA-WNN 模型的燃烧不稳定性检测方法[J]. 动力工程学报，2021，41(4)：286-293.
HUANG Wei，LI Changwei，XI Jianzhong. Detection method of combustion instability based on PCA-MEA-WNN model[J]. Journal of Chinese Society of Power Engineering，2021，41(4)：286-293.

[19] 卢绪祥，苏一鸣，吴家腾，等. 基于 EMD 及灰色关联度的滑动轴承润滑状态故障诊断研究[J]. 动力工程学报，2016，36(1)：42-47.
LU Xuxiang，SU Yiming，WU Jiateng，et al. Fault diagnosis on lubrication state of journal bearings based on EMD and grey relational degree[J]. Journal of Chinese Society of Power Engineering，2016，36(1)：42-47.