

· 应用研究 ·

基于 X-12-ARIMA-BP 模型的某省铁路企业工伤发生人数预测研究*

韩雪梅^{1#} 易济琨^{1#} 张笑琳² 司苗苗¹ 康盼¹ 张雨佳¹ 马桂妮¹ 闫欢^{1△} 丁明峰^{1△}

【提要】目的 基于某省历年工伤发生人数,构建 X-12-ARIMA-BP 模型对某省铁路企业工伤发生人数进行预测,提高其工伤事故管理能力。方法 基于某省铁路企业 2011 年第 1 季度至 2020 年第三季度的工伤发生人数数据,利用 X-12-ARIMA 模型以及 BP 模型对该企业工伤发生人数进行拟合预测。根据 X-12-ARIMA 模型对铁路企业工伤发生数的预测值和 BP 对误差的预测值,建立 X-12-ARIMA-BP 模型,对比三种模型的均方根误差(RMES)、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)评价指标,选择最优模型,最后运用该模型对某省铁路企业工伤人数进行预测。结果 X-12-ARIMA-BP模型的 RMSE、MAE 和 MAPE 值分别为 0.4205、0.2648、1.8719%,优于 X-12-ARIMA 模型的 2.2627、1.8361、11.9141%和 BP 神经网络模型的 1.8590、1.5194、11.9709%,因此 X-12-ARIMA-BP 模型的预测精度最高,是预测该铁路企业工伤发生人数的最佳模型。通过该模型对铁路企业 2021—2023 年每季度的工伤发生人数进行预测,结果显示,2021—2023 年该企业工伤发生趋势总体平稳。结论 X-12-ARIMA-BP 模型是预测该省铁路企业工伤发生人数的最优模型。

【关键词】 铁路企业 工伤 X-12-ARIMA-BP 模型

【中图分类号】 R195

【文献标识码】 A

DOI 10.11783/j.issn.1002-3674.2025.01.013

工伤是指劳动者在进行生产劳动的过程中或者进行与生产活动有关的社会活动中发生的事故伤害或者职业病伤害^[1]。工伤是职业人群伤残甚至死亡的重要原因,严重威胁着企业的生产经营和劳动者的生命安全^[2]。根据我国人力资源社会保障部最新的统计数据^[3],截至 2021 年末,全年认定(视同)工伤 129.9 万人,评定伤残等级 77.1 万人。中国劳动统计年鉴中显示^[4],2020 年某省认定(视同)工伤人数为 8092 人。由此可见,工伤事故是影响我国和该省职业人群生命健康的重要因素。作为该省重要产业,2020 年该铁路企业职工人数为 76151 人,职工众多。其工作环境普遍是需要野外、高空、井下等高危环境下作业,易受高温、严寒等极端天气影响,并且部分职工容易暴露于噪声、振动、机械伤害、车辆伤害、火灾、爆炸等职业风险中,多种危险因素交织易导致其工伤事故的发生。目前,组合模型在已有的工伤事故预测研究中并不多见,大部分研究都只是使用单一模型进行趋势预测。如果只使用一种模型,就会因其本身的不足而导致误差增大,而组合模型可以弱化单一模型的不足,从而提高预测结果的精确性。

本文先以季节性自回归积分滑动平均模型(X-12-autoregressive integrated moving average model,ARI-

MA)模型和误差逆传播神经网络(back-propagation network,BP)模型单独预测该企业的工伤发生人数,评价其拟合指标。之后为深度挖掘最优模型,合理预测该企业工伤发生人数,采用 BP 模型对 X-12-ARIMA 模型的误差进行修正,将该模型所得结果与 BP 模型修正结果相加得到改进后 X-12-ARIMA-BP 模型预测结果。比较 X-12-ARIMA 模型、BP 神经网络模型和改进后 X-12-ARIMA-BP 模型的评价指标选出最优预测模型,对该企业工伤发生人数进行预测,从而提高其工伤事故管理能力。

资料与方法

1. 资料来源

本研究资料来源于某省人力资源与社会保障厅 2011—2020 年第三季度工伤鉴定报告、铁路企业安全事故报告、相关企业内部资料。

2. 伤害标准界定

根据工伤鉴定报告及《工伤保险条例》第十四条和第十五条规定界定职工所受伤害是否符合工伤认定情形。

3. 质量控制

本研究使用 Epidata 3.1 软件对某省铁路企业 2011—2020 年第三季度工伤鉴定报告的数据进行双录入核验,确保录入信息准确。导出 excel 格式数据并整理,剔除数据中的异常值和缺失数据。

4. 统计方法

(1) X-12-ARIMA 季节调整模型

ARIMA 模型近年来被广泛应用于工伤事故分析

* 基金项目:甘肃省哲学社会科学规划项目阶段性研究成果(2023YB056)

1. 兰州大学公共卫生学院(730030)

2. 英国帝国理工学院公共健康专业

#共同第一作者

△通信作者:闫欢, E-mail: yanh19@lzu.edu.cn; 丁明峰, E-mail: dingmf20@lzu.edu.cn

和预测等相关领域^[5-8]。X-12-ARIMA 就是在 ARIMA 的基础上对时间序列中包含的季节性变化加以调整,以提升预测精度的模型。

本研究使用 SPSS 22.0 软件的周期性分解对原始序列进行周期性分解。一个时间序列(Y_t)受多种因素影响。通常可以把这些因素分解为趋势成分(T_t)、循环成分(C_t)、季节性因素(S_t)和不规则成分(I_t)。经典的时间序列季节调整模型有两种:

$$\text{加法模型: } Y_t = T_t + C_t + S_t + I_t \quad (1)$$

$$\text{乘法模型: } Y_t = T_t \times C_t \times S_t \times I_t \quad (2)$$

一般来说,如果四种成分相互独立采用加法模型,相互影响采用乘法模型。若季节成分非常稳定,并不随原始序列水平成比例变化,就使用加法模型,反之则使用乘法模型^[9]。

由于趋势成分和循环成分不容易分离,以上季节调整方法将趋势成分和循环成分视为一种成分进行分解和调整。

本研究构建 X-12-ARIMA 季节调整模型分为三步:第一,对铁路企业 2011 年 1 季度至 2019 年 2 季度的时间序列数据进行季节性分解,分解后得到趋势-循环成分(TC_t)、季节性成分(S_t)和不规则成分(I_t)。第二,使用 SPSS 22.0 软件“Expert Modeler”功能,分别对各分解成分进行建模,并使用 2019 年 3 季度至 2020 年 3 季度的工伤发生数据验证模型的预测精度。通过贝叶斯信息准则(bayesian information criterions, BIC)值最小为最佳模型选择标准,同时采用决定系数(R^2)来评价模型拟合精度。Ljung-Box 检验用于检查模型是否符合白噪声,当模型显著性 $P > 0.05$ 时,说明该模型符合白噪声。第三,采用乘法或加法模型,计算各成分的预测值,得到每季度的工伤发生人数预测值。

(2)BP 模型

BP 是一种单向传播的多层前馈神经网络。与 X-12-ARIMA 模型相比,BP 具有很强的非线性映射能力,在求解影响因素较为复杂的问题时具有明显优势。同时 BP 还具有较强的容错能力,即使系统局部受损,也不会对最终的训练结果造成较大影响。

BP 由输入层(input layer)、隐含层(hidden layer)和输出层(output layer)构成,同一层的神经元之间无联结,前后层依次相连。其工作原理是信息的网络前馈(即正向传播)和误差的反馈(即反向传播)。

BP 建模及预测的具体实施步骤如下:

①数据预处理。数据量过大或者数据差异过大均会影响 BP 的学习效果,为了提高学习效率,需要对样本数据做归一化处理,从而提高模型的收敛速度和预测的准确性。

②参数设置。需要分别设定 BP 的输入层和输出

层的参数值。具体设置因研究数据的不同而异。

③隐含层层数及其节点数设定。一般拥有 1 个隐含层是最为普遍的设计。但是根据数据本身的特点,若在含有 1 个隐含层的情况下无法改善神经网络的学习效果,可以尝试增加隐含层的层数。隐含层节点数的设置是影响 BP 学习效果的关键,设置较大会造成网络复杂,增加训练时间和学习难度,还可能产生过度拟合等问题。隐含层节点数目过少时可能出现前馈和反馈不充分的现象,从而减低预测精度。隐含层数目的确定并没有统一的理论作为指导,因此需要经多次试验调试来确定,确定隐含层节点数常用参考公式如下^[10]:

$$n = \sqrt{n_i + n_o} + a \quad (3)$$

其中, n 为隐含层节点数的初始值, n_i 为输入节点的数值, n_o 为输出节点的数值, a 为整数,自 1 开始取值, n 自 3 开始取值。

④学习率、目标误差以及迭代次数设定。学习率、目标误差和迭代次数的高低均会影响网络模型的稳定性、收敛程度和训练时间的长短,因此在 BP 预测过程中需要不断调试,以达到最佳预测效果。

⑤代入预处理后的数据,模型自行划分训练集和验证集。训练集与验证集误差差异较小时预测效果较好,本研究 BP 的预测通过 Matlab 2021b 软件实现。

(3)X-12-ARIMA-BP 模型

X-12-ARIMA 模型善于处理线性数据,在拟合时间序列中的线性趋势方面具有优势,而 BP 模型络对非线性数据具有很强的捕捉能力,能够捕捉时间序列中的非线性成分,为了进一步探索更加优化的模型,本研究基于 X-12-ARIMA 模型与 BP 络模型,构建了 X-12-ARIMA-BP 模型对某省铁路企业工伤发生趋势进行预测。使用 X-12-ARIMA 模型对铁路企业工伤发生人数历史数据 $N(t)$ 进行拟合,得到 X-12-ARIMA 模型的预测结果 $L'(t)$,计算误差 $M(t)$,使用 BP 模型络对误差进行建模,得到预测结果 $M'(t)$,最后把两个预测结果加起来,得到 X-12-ARIMA-BP 模型的预测 $N'(t)$ 。

X-12-ARIMA-BP 模型的表达式为:

$$L'(t) + M'(t) = N'(t) \quad (4)$$

5.模型预测精度指标

本研究模型预测输出为 $\hat{y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n]^T$,评估模型预测性能的指标分别为 MAE(平均绝对误差)、RMSE(均方根误差)、MAPE(平均绝对百分比误差)。其中,RMSE、MAE 和 MAPE 的值越小,表示模型预测精度越高,建模效果越好。

各指标的计算公式如下:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (7)$$

结 果

1. 数据划分

本研究中将 2011 年第 1 季度至 2019 年第 2 季度的工伤数据作为训练集, 2019 年 3 季度—2020 年 3 季度的工伤数据作为验证集。铁路企业 2011 年第 1 季度—2020 年第 3 季度的工伤发生人数历史数据, 见表 1。

表 1 某省铁路企业工伤发生人数的季度数据(人)

年份	第一季度	第二季度	第三季度	第四季度
2011	25	17	14	23
2012	20	22	21	27
2013	17	28	36	28
2014	15	27	18	32
2015	15	19	16	24
2016	16	13	16	14
2017	12	16	13	13
2018	6	5	16	14
2019	14	21	24	11
2020	12	20	18	—

注: “—”表示缺失值。

2. X-12-ARIMA 模型的建立和预测

(1) 季节性分解

首先对 2011 年 1 季度—2019 年 2 季度该企业的工伤发生人数的历史数据进行季节性分解, 见图 1。由图 1 可知, 该序列具有明显的趋势性和季节性特征。

(2) 趋势-循环成分的预测、季节成分提取和不规则成分的预测

采用 ARIMA 对趋势-循环成分和不规则成分进行建模。通过 SPSS 22.0 软件的 Expert Modeler 功能进行辅助建模, 并利用自相关函数图和偏自相关函数图进行定阶, 最终选择 ARIMA(0, 1, 1) 模型和 ARIMA(2, 0, 1) 模型分别对趋势-循环成分和不规则成分进行拟合和预测。结果显示趋势-循环成分 R^2 为 0.943, Ljung-Box 检验的统计量为 26.243, 检验的 $P=0.070(P>0.05)$, 差异无统计学意义, 认为 ARIMA(0, 1, 1) 模型的残差序列为白噪声序列。不规则成分 R^2 为 0.428, 拟合效果一般。Ljung-Box 检验的统计量为 18.785, $P=0.224(P>0.05)$, 差异无统计学意义, 认为该模型的残差序列为白噪声序列。

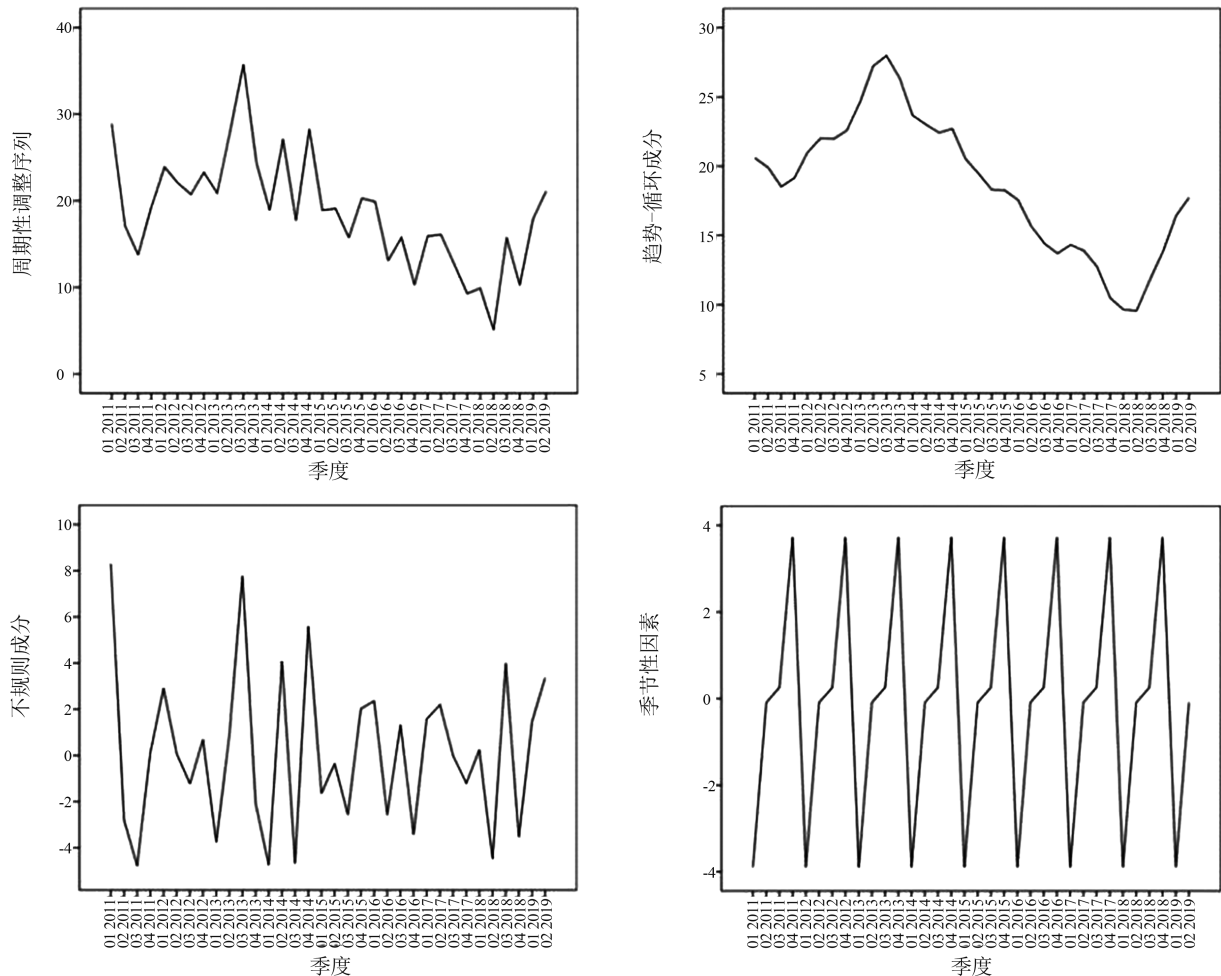


图 1 某省铁路企业时间序列季节性分解图

从图 1 可以看出来,季节成分是非常稳定的,并不随原始数据规模的增长而变化,所以本研究假设季节成分不变,截取 2017 年第 3 季度至 2018 年第 3 季度的季节成分作为 2019 年 3 季度至 2020 年 3 季度的季节成分。

在得到趋势-循环成分(TC_t)、季节性因素(S_t)和不规则因素(I_t)三种成分的预测值后,直接采用公式 $Y_t = TC_t + S_t + I_t$ 计算该省铁路企业 2019 年 3 季度至 2020 年 3 季度的工伤发生人数。X-12-ARIMA 模型拟合结果的 R^2 值为 0.8809,可见模型拟合效果良好。

3. BP 模型的建立与预测

本研究在建立 BP 模型时,网络创建函数选用 newcf,传递函数选用 tansig 和 purelin,训练函数选用 train。设置输入层神经节点数为 3,输出层节点数为 3,即用前 3 季度的工伤发生人数数据预测未来 3 季度的工伤发生人数。经多次反复测试,确定隐含层为 2 层,隐含层的神经节点数分别为 8 和 4,最终取学习率为 0.01,训练精度为 0.01 时预测效果最佳。

随着迭代次数的增加,训练误差逐渐减小,经过 112 次迭代之后模型得到最佳验证效果,其 MSE 为 0.0661,113 次迭代之后验证集误差开始增加,因此选择 112 次迭代得到的模型为最优模型,见图 2。

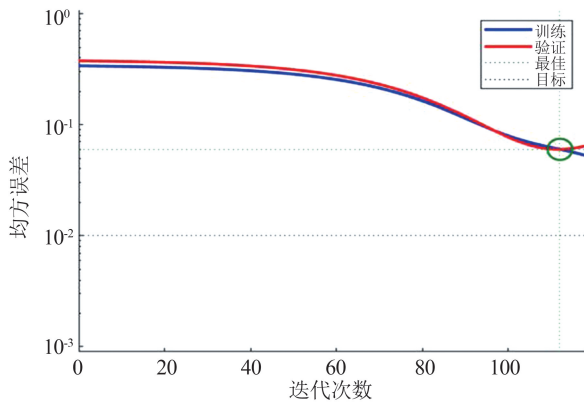


图 2 BP 模型训练误差曲线图

4. X-12-ARIMA-BP 模型的建立与预测

计算 X-12-ARIMA 模型对铁路企业工伤发生人数历史数据的拟合值和对应的实际值之间的误差,其中误差=拟合值-预测值,由此得到 2011 年 2 季度至 2020 年 3 季度的误差,依据误差建立误差值的 BP 模型。

网络创建方法与 BP 模型相同。设置输入层神经节点数为 3,输出层节点数为 1。经多次反复测试,确定隐含层为 2 层,隐含层的神经节点数分别为 8 和 2,最终取学习率为 0.01,训练精度为 0.01 时预测效果最佳。

随着迭代次数的增加,训练误差逐渐减小,经过 121 次迭代之后模型得到最佳验证效果,其 MSE 为

0.0452,122 次迭代之后验证集误差开始增加,因此选择 121 次迭代得到的模型为最优模型,见图 3。

计算出误差值的 BP 模型的 RMSE、MAE、MAPE 值分别为 0.0011、0.0213、5.4970%,可知误差值的 BP 模型的预测效果良好。

根据 X-12-ARIMA 对铁路企业工伤发生数的预测值和 BP 对误差的预测值,建立 X-12-ARIMA-BP 预测模型,依据公式 $L'(t) + M'(t) = N'(t)$,得到 X-12-ARIMA-BP 模型的预测结果。

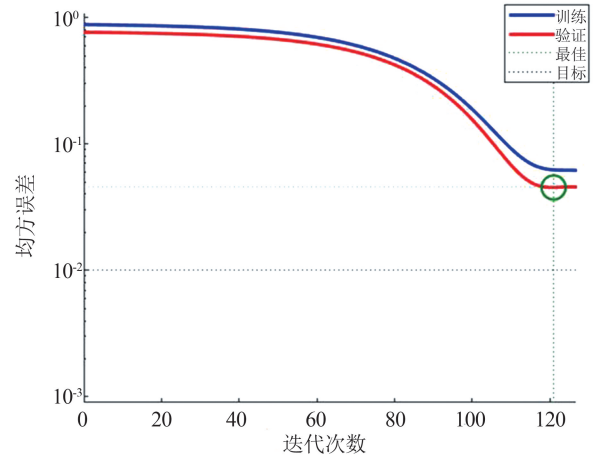


图 3 误差值的 BP 模型的训练误差曲线图

5. 三种模型的预测精度指标比较

X-12-ARIMA 模型拟合预测的 RMSE、MAE、MAPE 分别为 2.2627、1.8361、11.9141,BP 神经网络模型拟合预测的 RMSE、MAE、MAPE 分别为 1.8590、1.5194、11.9709。改进后 X-12-ARIMA-BP 模型的 RMSE、MAE、MAPE 分别为 0.4205、0.2648、1.8719。可以看出改进后 X-12-ARIMA-BP 模型预测精度最高,是预测该铁路企业工伤发生人数的最佳模型,见表 2。

表 2 三种模型的预测精度指标

模型	RMSE	MAE	MAPE(%)
X-12-ARIMA	2.2627	1.8361	11.9141
BP	1.8590	1.5194	11.9709
X-12-ARIMA-BP	0.4205	0.2648	1.8719

6. X-12-ARIMA-BP 模型的应用分析

使用 X-12-ARIMA-BP 模型预测该企业 2021—2023 年的工伤发生人数。最终预测结果见表 3,铁路企业 2021—2023 年工伤发生人数趋势变化图,见图 4。

表 3 铁路企业 2021—2023 年每季度工伤发生人数预测结果(人)

年份	第一季度	第二季度	第三季度	第四季度
2021	10.4574	20.9011	20.3779	23.8796
2022	13.3336	21.1250	16.2977	19.9078
2023	15.2866	20.2079	17.1986	19.1434

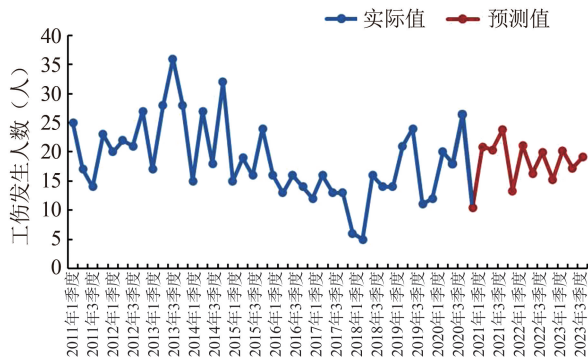


图 4 铁路企业 2021—2023 年工伤发生人数趋势变化图

讨论

一般来说,单一预测模型由于自身条件限制,可能由于无法全面掌握信息导致预测精度不高,结果不够可靠^[11],而通过模型的组合可以发挥单一模型的优势,克服短板。翟静等^[12]结合 ARIMA 模型和 BP 模型构建了一种新的组合预测模型应用于我国粮食产量的预测,并验证了组合预测模型在预测精度方面总体上优于单项预测模型。本文中 X-12-ARIMA 模型在预测时间序列中的非线性成分时的优势并不突出,而 BP 模型更善于捕捉非线性成分,它可以在指标间相关度较高、呈非线性变化,或数据缺漏不全等情况下得到比较满意的结果^[13]。虽然单一的 BP 模型具有很强的非线性映射能力,但是依然无法有效降低预测时一些偶然因素的影响。因此,本文结合两种模型优势,基于 X-12-ARIMA 模型对铁路企业工伤发生数的预测值和 BP 对误差的预测值,建立 X-12-ARIMA-BP 预测模型,探索对某省铁路企业工伤发生人数最优的预测模型,提高预测结果可靠性。

本文结果显示 X-12-ARIMA-BP 模型的 RMSE、MAE 和 MAPE 值分别为 0.4205、0.2648、1.8719%,优于 X-12-ARIMA 模型和 BP 模型的预测精度指标。这与杜子学^[14]的研究结果一致,其采用 X-12-ARIMA-BP 模型预测未来短期跨座式单轨空调系统故障,并验证了该模型相对于 X-12-ARIMA 模型和 BP 模型来说具有更好的预测精度。显然,X-12-ARIMA-BP 模型在考虑问题时更加的系统全面,在减少预测误差方面也更具优势。

从现实情况考虑,铁路工伤事故发生原因复杂,多是人、机器、环境、管理、时间等多种因素交织叠加的综合结果,除了具有明显的周期性和季节性特点外,还易受到突发事件的随机扰动影响。如果只使用单一预测模型预测,可能会由于模型自身的不足和偶然因素的影响而产生较大的误差,从而降低预测的准确性和科学性,同时也无法在现实中得到较好的应用。相比之下,X-12-ARIMA-BP 模型的预测结果是可靠的。

通过 X-12-ARIMA-BP 模型对铁路企业 2021—2023 年每季度的工伤发生人数进行预测,结果显示,2021—2023 年该企业工伤发生趋势总体平稳。这表明未来该企业工伤发生人数增加的趋势可能会得到控制,也说明当前该企业的职业安全管理方法行之有效,安全管理体系发挥了积极作用。然而,维持平稳趋势并不是国家和企业的最终目标,根据国家出台的《工伤预防五年行动计划(2021—2025)》^[15]的目标要求,要把重点行业的工伤发生率降低 20%,对照计划目标,该企业需要进一步探索新的管理方法,优化安全管理体系,除了应积极采取加强岗前安全培训,强化安全监督机制,多设立安全警示标志等措施外,本研究还建议企业能够搭建员工职业安全管理和宣传大数据平台,利用数字技术突破人力和物力对管理效果的限制,把有关职业技能的理论知识和安全教育的线下培训转移到线上,方便员工多渠道参与,打通员工接受职业安全教育的“最后一公里”。

虽然本研究构建的 X-12-ARIMA-BP 模型精度较高,但受到全国疫情形势的影响,致使 2020 年该企业部分线路停运,职工停工,后因疫情得到控制,又陆续复工复产,这可能会使该模型的预测精度受到一定影响。因此,铁路企业还应充分考虑突然停工可能带来的工伤发生人数减少,更应关注停产后企业加紧复工复产可能带来的工伤事故增多的情况。合理有序进行复工复产,及时帮助职工调整工作状态,不断提升企业工伤事故管理能力。

参考文献

- [1] 龚怡琳,冯宇彤,王慧卿.北京市某三甲医院员工工伤情况调查[J].医院管理论坛,2018,35(12):12-15.
- [2] 李雪.2000—2018 年某铁路企业职工工伤事故调查[J].疾病预防控制通报,2021,36(6):70-73.
- [3] 中华人民共和国人力资源和社会保障部.2021 年度人力资源和社会保障事业发展统计公报[EB/OL].https://www.mohrss.gov.cn/SYrlzyhshbzb/zwgk/szrs/tjgb/202206/t20220607_452104.html.
- [4] 国家统计局人口和就业统计司.中国劳动统计年鉴[M].北京:中国统计出版社,2021,396-397.
- [5] 张磊.基于 BP 的哈尔滨地铁小半径曲线钢轨磨耗预测[D].哈尔滨:哈尔滨工业大学,2021.
- [6] Huang SC, ZX, Zhang BC. Multiple regression analysis on occupational safety and economic and social development[J]. Journal of China Coal Society, 2005, 11(5): 580-584.
- [7] Wang XZ. Development report of production safety and economic society[R]. Beijing: Coal Industry Press, 2006.
- [8] Pransky GS, Benjamin KL, Savageau JA, et al. Outcomes in work-related injuries: A comparison of older and younger workers[J]. American Journal of Industrial Medicine, 2005, 47(2): 104-112.
- [9] Tiesman HM, Gwilliam M, Rojek J, et al. The impact of a crash prevention program in a large law enforcement agency[J]. American Journal of Industrial Medicine, 2019, 62(10): 847-858.

(下转第 84 页)