

文章编号: 1003-207(2010)04-0183-10

安全事故现状与趋势分析方法研究

刘卓军¹, 柳刚^{1,2}

(1. 中国科学院数学与系统科学研究院, 北京 100190; 2. 中国科学院研究生院, 北京 100190)

摘要: 包括安全生产事故预防、控制在内的安全生产监督管理工作需要客观把握安全生产的现状并准确地判断其总体发展趋势。本文基于我国安全生产事故快报数据, 对近年安全生产现状做出季节性等分析, 并为短期安全生产形势变化的预测与判断不仅提出了三个可行的 ARIMA-BP、ARIMA-RBF 以及 ARIMA-GRNN 非线性组合模型而且进一步基于 RBF, 对前述三个模型再次进行非线性组合, 给出了一种新的双重非线性组合趋势分析方法。实证结果表明, 双重非线性组合能够较为精确地预测安全生产事故的发展趋势, 可以为安全生产事故的预防、控制和应对提供管理和决策支持。

关键词: 安全生产; ARIMA 模型; BP 模型; RBF 模型; GRNN 模型; 非线性组合
中图分类号: C931 **文献标识码:** A

1 引言

目前, 安全生产仍然是世界各国都必须面对和应对的重大问题。对此, 我国坚持“安全第一, 预防为主, 综合治理”的方针。显然, 通过对各类安全生产事故现状和特征分析有助于认识安全生产事故的规律, 通过对安全生产事故发展趋势的准确判断有助于对安全生产事故的预防、控制和应对做出有效的决策。

国内外关于安全生产事故研究的报告^[1]和文献很多, 主要集中在以下方面: (1) 某一类安全生产事故的预测预警, 如李书全和窦艳杰(2008)^[2]针对建筑安全事故, 提出了 RS-SVM 预测模型; 钟铭(2009)^[3]采用模糊数据融合算法, 对煤矿瓦斯进行预警。(2) 安全事故发生的规律分析, 主要包括: 安全事故的成因分析, 安全事故与相关经济指标之间的关系。如吕海燕(2004)^[4]对我国工矿企业、消防火灾、交通等进行总量统计分析, 提出了相应的事故统计分析指标体系框架; Kopits 和 Cropper(2005)^[5]专门对交通死亡事故与经济增长的关系作了深入研究; 任荣森(2009)^[6]等人建立了油库的多层递阶解释结构

模型(ISM), 将油库安全事故成因分为根本原因、直接原因和间接原因三个层次。(3) 对安全生产管理的研究, 主要包括: 对事故管理绩效的评估、城市生产安全事故应急管理评价、安全事故损失的计算、安全事故中的风险管理。如牛跃林等人(2008)^[7]建立了城市生产安全事故应急管理评价指标体系, 为城市的生产安全事故应急体系建设提供决策依据; Karen(2009)^[8]和 Shi(2009)^[9]对煤矿安全事故风险管理及造成的损失从政府、行业规模和运营等方面作了探讨。

上述已有研究主要围绕特定行业领域展开, 而且主要集中在煤矿、建筑工程、交通安全生产事故。固然, 安全事故规律需要分不同行业领域进行研究, 但也需要整体认识和把握, 这样才能制定更科学更完善的国家安全生产战略规划。基于此, 本文重点针对安全生产事故的起数及其造成的死亡人数开展方法建模和实证研究。首先把安全生产事故按其造成的死亡人数进行分类, 具体地将安全生产事故分成 10 类: 死亡 i 个人的事故为第 i 类 ($i = 1, 2, \dots, 9$), 死亡 10 人及以上的事故为第 10 类。这样一种分类方式符合国家安全生产按不同伤亡和损失规模做整体统计的原则。因此, 相应的分析结果有助于国家安全生产事故预防、控制和应对的监管工作。

显然, 对不同类型的事故所采用的处置和应对方式以及相应的资源配置也不同。本文对从 2001 年 1 月到 2008 年 12 月国家安全生产快报包含的 1 万 7 千 6 百多条数据进行了分类型、分时段、分区域

收稿日期: 2009-05-05; 修订日期: 2010-07-05

基金项目: 国家科技支撑计划资助项目(2006BAK04A23)

作者简介: 刘卓军(1958-), 男(汉族), 黑龙江人, 中国科学院数学与系统科学研究院研究员, 研究方向: 管理科学与工程。

的整理分析和比较,特别结合各类安全生产月平均事故量序列的季节指数图,指出了需要重点关注的月份。由于安全生产管理决策离不开预测,而预测的精度又决定了预测结果的可用性和有效性。所以基于时间序列,建立了10类安全生产事故的数量以及安全生产事故总量的ARIMA-BP、ARIMA-RBF以及ARIMA-GRNN非线性组合预测模型,并结合RBF,构建了基于前述三个模型的双重非线性组合预测方法,提高了预测精度。

2 安全生产事故现状分析

根据国家安全生产监督管理总局官方网站发布

表1 2001-2008年各类安全生产事故总量统计

事故起数 年份	死亡 1人	死亡 2人	死亡 3人	死亡 4人	死亡 5人	死亡 6人	死亡 7人	死亡 8人	死亡 9人	死亡 10人及以上	合计
2001年	533	111	209	123	98	52	37	33	25	90	1311
2002年	1445	230	412	200	121	87	40	45	27	109	2716
2003年	1793	265	547	227	141	81	51	46	32	113	3296
2004年	1382	197	640	284	159	96	59	34	18	117	2986
2005年	1339	169	720	296	182	97	45	34	22	126	3030
2006年	134	22	731	258	152	119	55	49	19	94	1633
2007年	3	4	762	283	142	91	55	46	32	78	1496
2008年	4	4	567	215	123	66	55	29	25	88	1176
合计	6633	1002	4588	1886	1118	689	397	316	200	815	

的安全生产事故快报数据,对2001-2008年的安全生产状况进行了分类型、分时段、分区域的整理分析和比较,并结合各类安全生产月平均事故量序列季节指数图,指出了需重点关注的月份。

2.1 年度状况

为了分析全国安全生产事故近年发展趋势及其变化,对2001-2008年间的安全生产事故进行了整理和10个类别的分类统计。整体年度状况见表1。基于表1,把2001-2008年各类安全生产事故的发生变化趋势制成如下的图1。

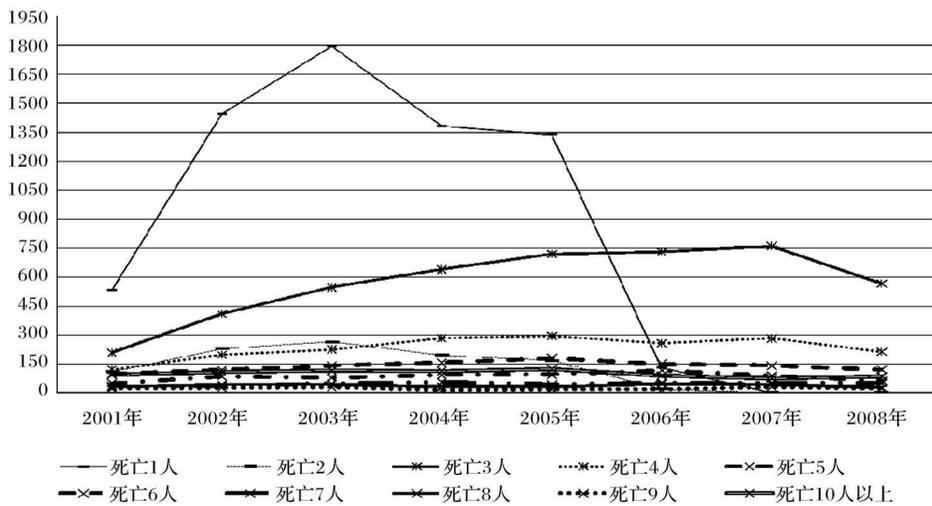


图1 2001-2008年各类安全生产事故每年事故量趋势变化图

2.2 月度状况

采用季节指数对2001-2008年的安全生产事故进行月度趋势分析。假设观测值序列的数据结构以C期为一个周期,共有T个周期,本文中C=12, T=8。用 X_{ijk} 表示第i类生产事故200j年第k月的事件起数,则第i类事故在第k月的季节指数 $S_k =$

$$\frac{C \sum_{j=1}^T X_{ijk}}{\sum_{j=1}^T \sum_{h=1}^C X_{ijh}}$$

。如果季节指数值大于1,那么第i类事故该月的事件平均值高于其总的事故平均值;如果季节指数值小于1,那么第i类事故该月的事件平均值低于其总的事故平均值。各类事故的季节指数图汇总在下面的图2。

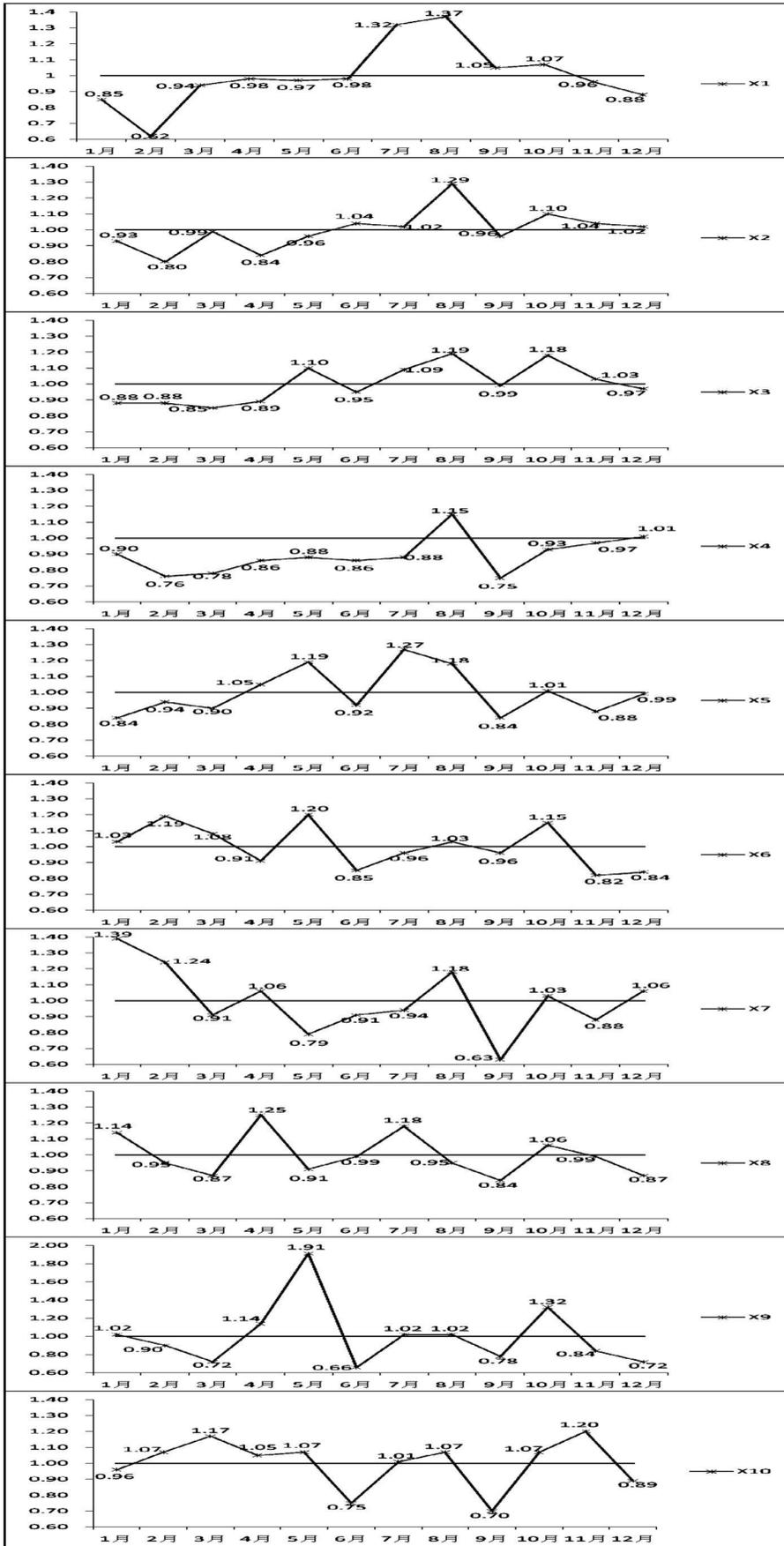


图 2 2001-2008 年各类安全事故月平均事故量序列季节指数图

从季节指数上看,第 1 类至第 5 类事故每年的 8 月份是事故高发期,需要加强监管。特别地,第 1 类和第 5 类每年的 7 月份也需要高度关注。第 6 类至第 10 类事故呈现与前面 5 类不同的特点,第 6 类需要加强监管的月份是每年的 2 月、5 月和 10 月;第 7 类需要加强监管的月份是每年的 1 月、2 月和 8 月;第 8 类需要加强监管的月份是每年的 1 月、4 月和 7 月;第 9 类需要加强监管的月份是每年的 4 月、5 月和 10 月;第 10 类需要加强监管的月份是每年的 3 月、8 月和 11 月。

2.3 区域安全生产事故状况

为了掌握不同地区安全生产事故的情况,通过数据分析整理,给出了 2001-2008 年各类安全生产事故在全国各地区的分布情况,见表 2:

表 2 我国 31 个省市区 2001-2008 年各季度安全生产事故平均总量统计表

省市区	第一季度	第二季度	第三季度	第四季度
01 北京市	4 000	6 375	6 000	4 750
02 天津市	0 750	1 250	1 375	1 250
03 上海市	1 125	2 375	2 125	2 375
04 重庆市	42 875	44 875	51 125	43 500
05 四川省	71 625	74 875	84 875	83 250
06 黑龙江省	13 875	15 000	25 750	19 125
07 吉林省	12 875	15 750	18 250	18 750
08 辽宁省	9 750	14 000	20 625	14 375
09 河北省	16 500	25 500	20 500	19 625
10 山西省	11 625	14 250	17 125	14 500
11 山东省	7 875	11 375	12 875	12 750
12 陕西省	24 375	26 875	32 875	29 500
13 河南省	10 125	13 750	14 500	7 625
14 湖北省	10 500	10 125	8 375	7 625
15 湖南省	41 875	48 500	61 000	47 000
16 江西省	4 500	8 000	9 500	6 375
17 安徽省	19 250	21 875	21 750	21 625
18 江苏省	6 625	7 375	5 875	7 875
19 浙江省	12 750	12 000	10 500	17 000
20 福建省	21 000	19 875	23 125	17 500
21 广东省	22 750	22 125	16 000	15 750
22 海南省	2 3750	2 000	1 625	1 875
23 贵州省	36 625	42 375	59 750	49 875
24 云南省	24 250	25 750	22 000	21 000
25 甘肃省	15 250	18 500	21 875	21 000
26 青海省	2 250	2 000	1 375	1 875
27 广西壮族自治区	14 875	8 875	14 375	18 125
28 内蒙古自治区	7 875	8 875	10 750	9 375
29 宁夏回族自治区	2 875	5 500	4 000	5 375
30 新疆维吾尔自治区	5 875	6 875	13 000	10 875
31 西藏自治区	4 625	3 625	8 125	7 125

需要说明的是,对于一个地区的安全生产状况的分析除关注事故量和死亡人数外,还要与地区的

地理环境、经济规模、产业情况及人口数量等因素结合起来进行关联分析。

3 各类安全生产事故总量预测

第 2 节从不同角度分析和考察了 2001-2008 年的安全生产事故现状,本节基于时间序列模型及神经网络,通过三阶段方式给出了双重非线性组合预测方法。

安全生产事故量随着时间的变化而变化,由于 ARIMA 模型在各类安全生产事故量预测过程中既考虑了这些数量在时间序列上的相关性,又考虑了随机变化的干扰性,对短期趋势预测具有较高的准确率。所以,第一阶段对各类安全生产事故量采用 ARIMA 建模。

因为安全生产事故涉及多种复杂的影响因素,应用单一的预测方法难于满足预测的有效性,所以有必要采用多个预测方法进行组合预测,综合利用各种预测方法提供的信息,以便减少预测误差而获得更好的预测结果。Bates^[10]和 Clemen^[11]都曾提出,可以将各种单一的预测结果进行组合得到一种组合预测结果,以达到改善预测效果的目的。20 世纪 90 年代初,Granger 和 Terasvirta^[12]等人研究发现,借助非线性函数关系的非线性组合预测比线性组合预测所得的结果误差要小。又由于非线性函数关系确定非常困难,而用神经网络来实现这种非线性函数映射相对容易并且神经网络具有自组织、自学习能力、非线性逼近能力强的特点。所以,第二阶段在第一阶段结论的基础上应用神经网络技术构建了三种可行的安全生产事故量非线性组合预测模型。

最后,第三阶段在第二阶段基础上给出了双重非线性组合预测方法。

对以下各阶段得出的预测模型,我们都利用 2001 年到 2009 年 9 月的安全生产事故快报数据作为已知信息,对 2009 年 10 月和 11 月的情况做出预测,并与实际值做出比较以验证模型的精度。

3.1 第一阶段预测

3.1.1 ARIMA 模型^[13]

ARIMA 模型又叫 Box-Jenkins 法,是以美国统计学家 George E. P. Box 和英国统计学家 Gwilym M. Jenkins 命名的时间序列分析方法。其结构如下:

$$\begin{cases} \Phi(B) \nabla^d X_t = \Theta(B) \varepsilon \\ E(\varepsilon) = 0, Var(\varepsilon) = \sigma_\varepsilon^2, E(\varepsilon_s \varepsilon_t) = 0, s \neq t \\ E(X_s \varepsilon_t) = 0, \forall s < t \end{cases}$$

简记为 ARIMA(p, d, q)。当在 d 阶差分后又作了步长为 L 的周期差分, 则此时模型结构为:

$\nabla_L \nabla^d X_t = \frac{\Theta(B)}{\Phi(B)} \varepsilon_t$ 。式中, B 为后移算子, $\nabla^d = (1 - B)^d$; $\Phi(B) = 1 - \alpha_1 B - \dots - \alpha_p B^p$; $\Theta(B) = 1 - \beta_1 B - \dots - \beta_q B^q$; 其中, $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ 是需要估计的自回归系数和滑动平均系数, $\alpha_p \neq 0, \beta_q \neq 0$; ε 为零均值白噪声序列。

3.1.2 ARIMA 模型预测

ARIMA 建模过程包括数据检验、模型识别、参数估计、模型适应性诊断等。选择 2001-2009 年 9 月之间的数据拟合 ARIMA 模型。

记 $X_i (i = 1, 2, \dots, 9)$ 代表死亡 i 个人的事故总量, X_{10} 代表死亡 10 人及以上的事故总量, X_{11} 代表总的事故量。 $X_{it} (i = 1, 2, \dots, 9)$ 代表第 t 期死亡 i 个人的事故总量, X_{10t} 代表第 t 期死亡 10 人及以上的事故总量, X_{11t} 代表第 t 期总的事故量。得到如下 11 个预测模型。

$$(1A) (1 + 0.41189B - 0.17536B^4 + 0.39474B^6 + 0.24147B^7 - 0.20759B^9) \nabla X_{1t} = \varepsilon_t, \varepsilon_t \sim WN(0, 477.19);$$

$$(2A) (1 + 0.54802B + 0.21288B^2 + 0.18633B^5 + 0.25675B^6 + 0.20701B^8) \nabla X_{2t} = \varepsilon_t, \varepsilon_t \sim WN(0, 14.84);$$

$$(3A) (1 + 0.56396B) \nabla X_{3t} = (1 - 0.38228B^2) \varepsilon_t, \varepsilon_t \sim WN(0, 108.02);$$

$$(4A) (1 - 0.26609B - 0.33063B^3) (X_{4t} - 19.06625) = \varepsilon_t, \varepsilon_t \sim WN(0, 39.22);$$

$$(5A) \nabla_2 \nabla X_{5t} = (1 - B)(1 - B^2) X_{5t} = \frac{(1 - 1.52709B^2 + 0.6311B^4)}{(1 + 0.60906B)} \varepsilon_t, \varepsilon_t \sim WN(0, 19.68);$$

$$(6A) \nabla_2 \nabla_2 X_{6t} = (1 - B^2)(1 - B^2) X_{6t} = \frac{(1 - 0.86252B^2)}{(1 + 0.70046B^2 + 0.50054B^4 + 0.31398B^6)} \varepsilon_t, \varepsilon_t \sim WN(0, 13.62);$$

$$(7A) \nabla_4 \nabla_2 X_{7t} = \frac{1}{(1 + 0.61708B^2 + 1.06179B^4 + 0.46342B^6 + 0.45628B^8)} \varepsilon_t, \varepsilon_t \sim WN(0, 10.28);$$

$$(8A) (1 + 0.81153B + 0.61162B^2 + 0.51758B^3 + 0.20181B^4) \nabla X_{8t} = \varepsilon_t, \varepsilon_t \sim WN(0, 3.65);$$

$$(9A) \nabla^6 \nabla X_{9t} = (1 - B)(1 - B^6) X_{9t} =$$

$$\frac{(1 - 0.17689B^2 - 0.66637B^6)}{(1 + 0.69881B)} \varepsilon_t, \varepsilon_t \sim WN(0, 3.93);$$

$$(10A) (1 + 0.30446B^2) \nabla_2 X_{10t} = (1 - 0.79891B^2) \varepsilon_t, \varepsilon_t \sim WN(0, 12.49);$$

$$(11A) (1 - 0.4416B - 0.2876B^2 + 0.18249B^6 - 0.21841B^9) \nabla X_{11t} = (1 - 0.77191B) \varepsilon_t, \varepsilon_t \sim WN(0, 937.366)$$

根据上述模型, 得到 2009 年 10-11 月份的各类安全生产事故量的点预测值, 见表 7。同时也给出点预测值的置信水平为 95% 的置信限以及安全生产事故总量时间序列拟合图, 见表 8 和图 3。其中置信限计算公式如下:

$\Phi(B) \nabla^d X_t = \Theta(B) \varepsilon_t$,
记 $\Phi^*(B) = \Phi(B) \nabla^d = 1 - \alpha_1^* B - \alpha_2^* B^2 - \dots$,
又令 $X_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots$, 根据 B 的同幂次系数相等的原理, 可以确定 ψ_1, ψ_2, \dots , 即

$$\begin{cases} \psi_1 = \alpha_1^* - \beta_1 \\ \psi_2 = \alpha_1^* \psi_1 + \alpha_2^* - \beta_2 \\ \dots \dots \dots \\ \psi_j = \alpha_1^* \psi_{j-1} + \dots + \alpha_{p-d}^* \psi_{j-p-d} - \beta_j \end{cases}$$

其中, $\psi_j = \begin{cases} 0, j < 0 \\ 1, j = 1, \dots, q \end{cases}$, $\beta_j = 0, j > q$ 。在均方误差最小原则下, 容易求出

$$X_t(l) = \psi_1 \varepsilon_t + \psi_2 \varepsilon_{t-1} + \psi_3 \varepsilon_{t-2} + \dots, X_{t+l} = \varepsilon_{t+l} + \psi_1 \varepsilon_{t+l-1} + \dots + \psi_{l-1} \varepsilon_{t+1} + \psi_l \varepsilon_t + \dots, \text{所以 } e_t(l) = \varepsilon_{t+l} + \psi_1 \varepsilon_{t+l-1} + \dots + \psi_{l-1} \varepsilon_{t+1}, \text{Var}(e_t(l)) = (1 + \psi_1^2 + \dots + \psi_{l-1}^2) \alpha_\varepsilon^2, X_{t+l} \text{ 的置信水平为 } 1 - \alpha \text{ 的置信限为: } (X_t(l) \mp Z_{1-\frac{\alpha}{2}} \cdot (1 + \psi_1^2 + \dots + \psi_{l-1}^2)^{\frac{1}{2}} \cdot \alpha_\varepsilon)$$

其中, $Z_{1-\frac{\alpha}{2}}$ 为标准正态分布 $1 - \frac{\alpha}{2}$ 的分位数, 本文 α 取 5%。(注: $e_t(l)$ 为 t+l 时期的预测误差)

3.1.3 安全生产事故总量 ARIMA 拟合预测图

图 3 是安全生产事故总量 ARIMA 拟合图。图中星号为实际值, 中间实曲线为依据 3.1.2 中 (11A) 得到的预测值, 上下两条虚曲线为预测值置信水平为 95% 的置信上限和置信下限。

3.2 第二阶段预测

3.2.1 BP 模型^[14]

BP 神经网络由 Rumelhart 等人于 1986 年建立, 是一种多层前馈型神经网络, 由输入层、隐层和输出层构成。隐层神经元的传递函数通常用 log-sigmoid 型函数、tansigmoid 型函数、purelin 纯线性型函数。输出层的函数常用 purelin 纯线性型函

数。它是前向网络的核心部分,体现了人工神经网络的精华。本文采用 BP 建立了第一种非线性组合预测方法,记作: ARIMA-BP。BP 模型表达式如

下: $O = f(WP + B)$, 其中 O 为节点输出; W 为节点连接权值; P 为节点输入; B 为神经元阈值; f 为输入/输出的传递函数。

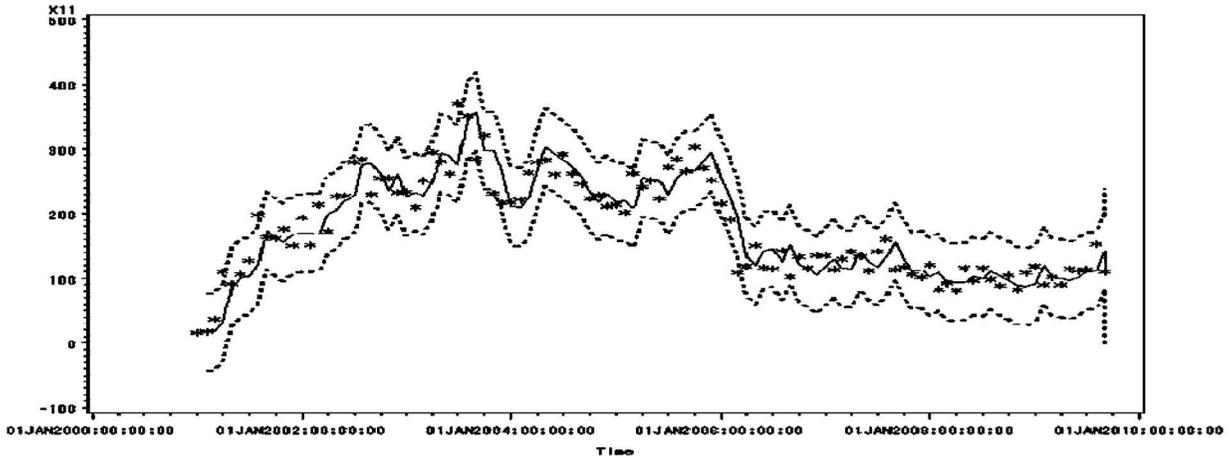


图3 安全生产事故总量时间序列拟合预测图

3.2.2 ARIMA-BP 模型预测

依据 3.1.2 的 11 个模型,对 10 类事故以及事故总量进行第二阶段建模。ARIMA-BP 模型的实现过程由模型输入、网络训练、网络测试、模型输出等四个部分组成。确定神经网络为 3 层,输入层、隐层及输出层,隐层采用 S 型函数作为传递函数,输出层取线性函数。网络训练采用基于数值最优化理论的 Levenberg Marquardt^[15-16] 方法进行,然后调用这些训练好的权值与阈值,经输出层输出即可得到预测值。依据 3.1.2,确定了安全生产事故量随时间变化的规律,因此构建的输入输出样本对以及网络结构各层神经元个数见表 3。

表3 ARIMA-BP 模型结构表

事故类型	网络输入	期望输出	网络结构
X1	$X_{1t-11}, X_{1t-10}, \dots, X_{1t-1}$	X_{1t}	11-10-1
X2	$X_{2t-10}, X_{2t-9}, \dots, X_{2t-1}$	X_{2t}	10-10-1
X3	$X_{3t-3}, X_{3t-2}, \dots, X_{3t-1}$	X_{3t}	3-6-1
X4	$X_{4t-4}, X_{4t-3}, \dots, X_{4t-1}$	X_{4t}	4-5-1
X5	$X_{5t-5}, X_{5t-4}, \dots, X_{5t-1}$	X_{5t}	5-5-1
X6	$X_{6t-11}, X_{6t-10}, \dots, X_{6t-1}$	X_{6t}	11-10-1
X7	$X_{7t-14}, X_{7t-12}, \dots, X_{7t-1}$	X_{7t}	14-12-1
X8	$X_{8t-6}, X_{8t-5}, \dots, X_{8t-1}$	X_{8t}	6-6-1
X9	$X_{9t-9}, X_{9t-8}, \dots, X_{9t-1}$	X_{9t}	9-10-1
X10	$X_{10t-5}, X_{10t-4}, \dots, X_{10t-1}$	X_{10t}	5-5-1
X11	$X_{11t-11}, X_{11t-10}, \dots, X_{11t-1}$	X_{11t}	11-10-1

应用 ARIMA-BP 模型关于 2009 年 10 月和 11 月事故起数的预测结果见表 7。

3.2.3 RBF 模型^[17-18]

径向基函数(RBF)是多维空间插值的传统技术,由 Powell 于 1985 年提出。1988 年, Broomhead 和 Lowe 根据生物神经元具有局部响应这一特点,将 RBF 引入神经网络设计中,产生了 RBF 神经网络。它属于前馈型神经网络,是一种三层的前向网络。第一层为输入层;第二层为隐层(径向基层);第三层为线性输出层。1989 年, Jackson 论证了 RBF 神经网络对非线性函数的一致逼近性能。因而将它应用于复杂的时间序列预测会取得较好的效果。本文采用 RBF 建立了第二种非线性组合预测方法,记作: ARIMA-RBF。RBF 模型表达式如下: 径向基层输入为 $\|W_1 - P\|$, 径向基层输出为 $O_1 = \text{radbas}(\|W_1 - P\|B_1)$, 线性输出层输出为 $O_2 = \text{purelin}(W_2^* O_1 + B_2)$ 。其中, P 为输入向量; W_1 和 W_2 分别为径向基层和线性输出层的节点连接权值; B_1 和 B_2 分别为径向基层和线性输出层的神经元阈值。

3.2.4 ARIMA-RBF 模型预测

ARIMA-RBF 模型构建的输入输出样本对与 ARIMA-BP 完全相同。隐层的径向基函数设计为高斯函数,输入为输入向量与权值向量的距离乘以阈值,输出即为神经元的输出: $O = \text{radbas}(\|W - P\|B)$ 。其中 P 为输入向量; W 为径向基层的节点连接权值; B 为径向基层的神经元阈值。另外,还需要给出径向基函数的散布常数(SPREAD)。本文综合以下三个原则确定:(1)通常,散布常数的选取取决于输入向量之间的距离,要求是大于最小距离,小

于最大距离; (2) 网络的预测误差和逼近误差; (3) 散布常数等于 1 作为参照。首先由输入向量间距离矩阵确定 SPREAD 的范围, 然后在此范围内结合预测误差、逼近误差以及参考散布常数 1 决定最终的 SPREAD 值。ARIMA-RBF 模型结构见表 4:

表 4 ARIMA-RBF 模型结构表

事故类型	网络输入	期望输出	SPREAD
X1	$X_{1t-11}, X_{1t-10}, \dots, X_{1t-1}$	X_{1t}	0.10
X2	$X_{2t-10}, X_{2t-9}, \dots, X_{2t-1}$	X_{2t}	0.10
X3	$X_{3t-3}, X_{3t-2}, \dots, X_{3t-1}$	X_{3t}	0.04
X4	$X_{4t-4}, X_{4t-3}, \dots, X_{4t-1}$	X_{4t}	0.15
X5	$X_{5t-5}, X_{5t-4}, \dots, X_{5t-1}$	X_{5t}	1.80
X6	$X_{6t-11}, X_{6t-10}, \dots, X_{6t-1}$	X_{6t}	1.50
X7	$X_{7t-14}, X_{7t-12}, \dots, X_{7t-1}$	X_{7t}	1.50
X8	$X_{8t-6}, X_{8t-5}, \dots, X_{8t-1}$	X_{8t}	0.70
X9	$X_{9t-9}, X_{9t-8}, \dots, X_{9t-1}$	X_{9t}	1.70
X10	$X_{10t-5}, X_{10t-4}, \dots, X_{10t-1}$	X_{10t}	0.50
X11	$X_{11t-11}, X_{11t-10}, \dots, X_{11t-1}$	X_{11t}	2.00

应用 ARIMA-RBF 模型关于 2009 年 10 月和 11 月事故起数的预测结果见表 7。

3.2.5 GRNN 模型^[19-20]

广义回归神经网络 (GRNN) 是美国学者 Donald F. Specht 在 1991 年提出的。在结构上与 RBF 网络较为相似。它由输入层、模式层、求和层和输出层四层构成。GRNN 也具有强烈的非线性映射能力, 适用于解决非线性问题。GRNN 网络最后收敛于样本量积聚较多的优化回归面, 并且在样本数据较少时, 预测效果也比较好。另外它还可以处理不稳定的数据。本文采用 GRNN 建立了第三种非线性组合预测方法, 记作: ARIMA-GRNN。GRNN

模型表达式如下: $E[y | x] = \frac{\int_{-\infty}^{\infty} y f(x, y) dy}{\int_{-\infty}^{\infty} f(x, y) dy}$, 其中

$E[y | x]$ 是在输入向量为 x 的条件下, 输出向量 y 的期望值。 $f(x, y)$ 是 x 和 y 的联合概率密度函数, 可由 Parzen 非参数估计出来。

3.2.6 ARIMA-GRNN 模型预测

ARIMA-GRNN 模型构建的输入输出样本对也与 ARIMA-BP 相同。通过 GRNN 模型不断学习归纳事故类型网络输入期望输出 SPREAD 输出和输出变量之间的关系。类似的, GRNN 的 SPREAD 值的确定原则与 3.2.4 一样, ARIMA-GRNN 模型结构见表 5:

表 5 ARIMA-GRNN 模型结构表

事故类型	网络输入	期望输出	SPREAD
X1	$X_{1t-11}, X_{1t-10}, \dots, X_{1t-1}$	X_{1t}	0.10
X2	$X_{2t-10}, X_{2t-9}, \dots, X_{2t-1}$	X_{2t}	0.10
X3	$X_{3t-3}, X_{3t-2}, \dots, X_{3t-1}$	X_{3t}	0.06
X4	$X_{4t-4}, X_{4t-3}, \dots, X_{4t-1}$	X_{4t}	0.15
X5	$X_{5t-5}, X_{5t-4}, \dots, X_{5t-1}$	X_{5t}	1.80
X6	$X_{6t-11}, X_{6t-10}, \dots, X_{6t-1}$	X_{6t}	1.50
X7	$X_{7t-14}, X_{7t-12}, \dots, X_{7t-1}$	X_{7t}	1.50
X8	$X_{8t-6}, X_{8t-5}, \dots, X_{8t-1}$	X_{8t}	0.70
X9	$X_{9t-9}, X_{9t-8}, \dots, X_{9t-1}$	X_{9t}	1.70
X10	$X_{10t-5}, X_{10t-4}, \dots, X_{10t-1}$	X_{10t}	0.50
X11	$X_{11t-11}, X_{11t-10}, \dots, X_{11t-1}$	X_{11t}	1.00

应用 ARIMA-GRNN 模型关于 2009 年 10 月和 11 月事故起数的预测结果见表 7。

3.3 第三阶段预测

3.3.1 双重非线性组合模型的构建

将第二阶段三种预测方法关于训练样本的预测结果作为 RBF 的输入向量, 输入节点数为 3, 而实际值作为期望输出向量训练网络, 输出节点数为 1。SPREAD 值的确定原则与 3.2.4 一样。最后, 把上述三种预测方法关于 2009 年 10 月和 11 月事故起数的预测值作为输入向量来预测 10 月和 11 月的事事故起数。上述基于 RBF 的 ARIMA-BP、ARIMA-RBF 以及 ARIMA-GRNN 的双重非线性组合模型记作: $f^{RBFCombination}(ARIMA-BP, ARIMA-RBF, ARIMA-GRNN)$ 。

3.3.2 双重非线性组合模型预测

由 3.3.1 得到的双重非线性组合模型结构见表 6, 应用双重非线性组合模型关于 2009 年 10 月和 11 月事故起数的预测结果见表 7。

3.4 模型预测效果的比较

我们把各种模型对 2009 年 10-11 月各类安全生产事故的数量和安全生产事故总量的预测结果总结在下面的表 7。同时用表 9 的误差评价表, 即平均绝对误差 (MAE) 和均方根误差 (RMSE) 来比较各种模型。

从表 7 和表 9 看, ARIMA-BP、ARIMA-RBF、ARIMA-GRNN 和 $f^{RBFCombination}$ 均优于 ARIMA。三个单一的非线性组合预测模型中, 最优的是 ARIMA-RBF, 其次是 ARIMA-BP, 最后是 ARIMA-GRNN。而本文最后给出的 $f^{RBFCombination}$ 不仅优于单一的 ARIMA 预测模型而且也优于三个单一的非线性组合预测模型。 $f^{RBFCombination}$ 最小 MAE 为 0, 最大 MAE 为 3.2319。最小 RMSE 为 0, 最大 RMSE 为 4.3221。 $f^{RBFCombination}$ 的最大 MAE 分别仅

为 ARIMA、ARIMA-BP、ARIMA-RBF、ARIMA-GRNN 最大 MAE 的 30.2%、41.82%、35.93% 和 30.03%。 $f^{RBF-Combinaon}$ 的最大 RMSE 分别仅为 ARIMA、ARIMA-BP、ARIMA-RBF、ARIMA-GRNN 最大 RMSE 的 40.26%、46.45%、44.98%

和 36.34%。由此可见,双重非线性组合模型不仅提高了预测精度而且还能满足非线性的预测要求,这为解决因安全生产事故量随机变动而造成的预测困难及精度问题提供了新的思路和方法,是安全生产中可行的预测模型。

表 6 $f^{RBF-Combinaon}$ 模型结构表

事故类型	网络输入	期望输出	SPREAD
X1	$X1_{t(ARIMA-BP)}, X1_{t(ARIMA-RBF)}, X1_{t(ARIMA-GRNN)}$	$X1_t$	0.011
X2	$X2_{t(ARIMA-BP)}, X2_{t(ARIMA-RBF)}, X2_{t(ARIMA-GRNN)}$	$X2_t$	0.120
X3	$X3_{t(ARIMA-BP)}, X3_{t(ARIMA-RBF)}, X3_{t(ARIMA-GRNN)}$	$X3_t$	0.330
X4	$X4_{t(ARIMA-BP)}, X4_{t(ARIMA-RBF)}, X4_{t(ARIMA-GRNN)}$	$X4_t$	0.790
X5	$X5_{t(ARIMA-BP)}, X5_{t(ARIMA-RBF)}, X5_{t(ARIMA-GRNN)}$	$X5_t$	1.000
X6	$X6_{t(ARIMA-BP)}, X6_{t(ARIMA-RBF)}, X6_{t(ARIMA-GRNN)}$	$X6_t$	0.040
X7	$X7_{t(ARIMA-BP)}, X7_{t(ARIMA-RBF)}, X7_{t(ARIMA-GRNN)}$	$X7_t$	1.000
X8	$X8_{t(ARIMA-BP)}, X8_{t(ARIMA-RBF)}, X8_{t(ARIMA-GRNN)}$	$X8_t$	0.100
X9	$X9_{t(ARIMA-BP)}, X9_{t(ARIMA-RBF)}, X9_{t(ARIMA-GRNN)}$	$X9_t$	1.000
X10	$X10_{t(ARIMA-BP)}, X10_{t(ARIMA-RBF)}, X10_{t(ARIMA-GRNN)}$	$X10_t$	1.000
X11	$X11_{t(ARIMA-BP)}, X11_{t(ARIMA-RBF)}, X11_{t(ARIMA-GRNN)}$	$X11_t$	1.000

表 7 各模型预测结果与实际值对比表

事故类型	月份	ARIMA	ARIMA-BP	ARIMA-RBF	ARIMA-GRNN	双重非线性组合	实际值
X1	Oct-09	0.3458	0.2363	1.0000	1.0000	1.0000	1
	Nov-09	0.5981	1.1846	1.0000	1.0000	1.0000	1
X2	Oct-09	0.5401	1.2631	1.0000	1.0000	1.8623	1
	Nov-09	0.0000	0.7473	1.0000	1.0000	2.2960	3
X3	Oct-09	58.1669	70.5482	62.3831	61.6840	61.8984	68
	Nov-09	59.5683	60.9092	60.3751	45.0000	48.3621	48
X4	Oct-09	19.3007	20.5902	18.7207	14.9993	19.0416	19
	Nov-09	23.0658	20.2279	20.2533	20.2913	20.2533	21
X5	Oct-09	10.6704	14.7755	12.7571	11.1068	13.5266	17
	Nov-09	14.3278	8.6321	7.2730	12.0481	7.4949	7
X6	Oct-09	6.6129	7.8951	5.1808	6.2122	6.0000	6
	Nov-09	7.6094	5.9925	4.6983	6.4539	6.0000	6
X7	Oct-09	2.2773	1.9003	2.2836	3.8334	2.8889	2
	Nov-09	5.0013	3.0437	2.5596	3.7257	2.5505	2
X8	Oct-09	2.0034	2.4768	2.6678	2.9014	3.0023	3
	Nov-09	2.2442	3.6287	2.9089	3.0199	3.0035	3
X9	Oct-09	0.6359	2.0166	1.1437	1.7311	1.1572	1
	Nov-09	1.0435	3.9198	2.7678	1.9044	2.7752	3
X10	Oct-09	5.5889	9.1611	7.7147	9.9021	7.8815	8
	Nov-09	5.4890	7.4849	8.0418	10.8010	8.3702	6
X11	Oct-09	131.7766	130.4441	133.0933	116.1790	132.3547	132
	Nov-09	118.5314	115.6359	113.3715	110.2932	113.3678	116

表 8 2009 年 10-11 月 ARIMA 预测值置信水平为 95% 的置信区间

置信限	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
Oct-09	(0, 43.1606)	(0, 8.0898)	(37.7969, 78.5369)	(7.0259, 31.5755)	(1.9745, 19.3663)	(0, 13.8451)	(0, 8.5616)	(0, 5.7497)	(0, 4.5212)	(0, 12.5155)	(71.7695, 191.7837)
Nov-09	(0, 50.2683)	(0, 8.0448)	(37.3461, 81.7906)	(7.7273, 33.1312)	(4.9910, 23.6646)	(0.3773, 14.8416)	(0, 11.2856)	(0, 6.0564)	(0, 5.1012)	(0, 12.4156)	(46.3108, 190.7520)

表9 各模型误差评价表

事故类型	评价标准	ARIMA	ARIMA-BP	ARIMA-RBF	ARIMA-GRNN	双重非线性组合
X1	MAE	0.5281	0.4742	0.0000	0.0000	0.0000
	RMSE	0.5430	0.5556	0.0000	0.0000	0.0000
X2	MAE	1.7300	1.2579	1.0000	1.0000	0.7832
	RMSE	2.1461	1.6037	1.4142	1.4142	0.7871
X3	MAE	10.7007	7.7287	8.9960	4.6580	3.2319
	RMSE	10.7358	9.3043	9.6097	4.9443	4.3221
X4	MAE	1.1833	1.1815	0.5130	2.3547	0.3942
	RMSE	1.4761	1.2500	0.5637	2.8730	0.5288
X5	MAE	6.8287	1.9283	2.2580	5.4706	1.9842
	RMSE	6.8469	1.9509	3.0064	5.4869	2.4809
X6	MAE	1.1111	0.9513	1.0605	0.3331	0.0000
	RMSE	1.2177	1.3400	1.0876	0.3543	0.0000
X7	MAE	1.6393	0.5717	0.7116	1.7796	0.7197
	RMSE	2.1313	0.7414	0.7277	1.7804	0.7393
X8	MAE	0.8762	0.5760	0.2117	0.0593	0.0029
	RMSE	0.8844	0.5784	0.2435	0.0714	0.0028
X9	MAE	1.1603	0.9682	0.1880	0.9134	0.1910
	RMSE	1.4072	0.9694	0.1931	0.9313	0.1941
X10	MAE	1.4611	1.3230	1.1636	3.3516	1.2444
	RMSE	1.7428	1.3329	1.4578	3.6515	1.6781
X11	MAE	1.3774	0.9602	1.8609	10.7639	1.4935
	RMSE	1.7969	1.1299	2.8468	11.8921	1.8781
	总体 MAE	2.5997	1.6292	1.6330	2.7895	0.9132
	总体 RMSE	2.8117	1.8870	1.9228	3.0363	1.1465

4 结语

本文按生产事故造成伤亡和损失规模的原则重新对安全生产事故进行分类,并对各类安全生产事故进行了现状分析,借助季节指数分析法对各类安全生产事故指出了应重点预防的时段建议。最后为了精确地预测安全生产事故的发展趋势,对各类安全生产事故量以及安全生产事故总量先建立了三个可行的非线性组合预测模型,即 ARIMA-BP、ARIMA-RBF 以及 ARIMA-GRNN,然后在此基础上构建了一种新的双重非线性组合趋势分析方法。为提高预测的准确性提供了新的方法和思路,为管理和决策部门制定安全生产战略规划提供了理论依据。通过本文建立的模型及其实证研究结果,一定程度上说明就我国目前所处的经济发展阶段而言,各类安全生产事故发生的最严重程度及经过努力可以实现的改进程度是可以做出统计分析和预测的。显然,各类事故发生的频度也有很大差别,进一步弄清其原因是今后值得特别关注的问题。各类事故在一天 24 小时内的不同时间段发生的频度同样值得进一步关注,所有这些对控制和改进安全生产状况都具有重要意义。

参考文献:

- [1] 王显政. 安全生产与经济社会发展报告[M]. 北京: 煤炭工业出版社, 2006.
- [2] 李书全, 窦艳杰. 基于 RS-SVM 模型的建筑安全事故预测模型[J]. 统计与决策, 2008(19): 56-58.
- [3] 钟铭. 基于模糊数据融合的煤矿瓦斯预警研究[D]. 西安: 西安科技大学博士论文, 2009.
- [4] 吕海燕. 生产安全事故统计分析及预测理论方法研究[D]. 北京: 北京林业大学博士论文, 2004.
- [5] Kopits, E., Cropper, M.. Traffic fatalities and economic growth[J]. Accident Analysis and Prevention, 2005, 37(1): 169-178.
- [6] 任荣森, 雍歧东, 张宁. 基于 ISM 的油库安全事故成因分析[J]. 后勤工程学院学报, 2009, 25(5): 58-61.
- [7] 牛跃林, 王花平, 谢金亮. 基于模糊数学的城市生产安全事故应急管理综合能力综合评价[J]. 江西理工大学学报, 2008, 29(1): 53-56.
- [8] Karen, P.. Blood on the coal: The effect of organizational size and differentiation on coal mine accidents[J]. Journal of safety Research, 2009, 40(3): 85-95.
- [9] Shi, X. P.. Have government regulations improved workplace safety? A test of the asynchronous regulatory effects in China's coal industry, 1995-2006[J]. Journal of Safety Research, 2009, 40(3): 207-213.

- [10] Bates, J. M., Granger, C. W. J., Terasvirta, T.. The combination of Forecasts [J]. *Operations Research Quarterly*, 1969, 20(4): 451– 468.
- [11] Clemen, R. T. . Combination forecasts: A review and annotated bibliography [J]. *International Journal of Forecasting*, 1989, (5): 559– 584.
- [12] Deutch, M. , Granger, C. W. J., Terasvirta, T. The combination of forecasts using changing weights [J]. *International Journal of Forecasting*, 1994, 10: 47– 57.
- [13] George, E. P. B. , Gwilym, M. J. , Gregory, C. R. . *Time Series Analysis, Forecasting, Control*, [M]. *Erlewood Cliffs, NJ Prentice Hall*, 1994.
- [14] Rumelhart, D. E., Hinton, G. E., Williams, R. J. . Learning representations of backpropagation errors [J]. *Nature*, 1986(323) : 533– 536.
- [15] Levenberg, K.. A method for the solution of certain problem in least squares [J]. *Quart. Appl. Math.*, 1944, 2 : 164– 168.
- [16] Marquardt, D.. An algorithm for leastsquares estimation of nonlinear parameters [J]. *SIAM J. Appl. Math.*, 1963, 11: 431– 441.
- [17] Powell, M .J. D.. Radical basis functions for multivariable interpolation: A review [C]. *IMA conference on Algorithms for the Approximation of Functions and Data*, RM CS, Shrivenham, England: 1985: 143– 167.
- [18] Park, J. . Universal approximations using RBF networks [J]. *Neural computation*, 1991, 3: 246– 257.
- [19] Specht, D. F. A general regression neural network [J]. *IEEE Transactions on Neural Networks*, 1991, 2(6): 568– 576.
- [20] Specht, D. F. . Generalization accuracy of probabilistic neural networks compared with backpropagation networks [C]. *Seattle: International Joint Conference on Neural Networks IJCNN– 91*, 1991, 1: 887– 892.

The Analysis and Forecast of National Accident Trends in Work Safety

LIU Zhuo jun¹, LIU Gang^{1,2}

(1. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China;

2. Graduate School of Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Based on the data query system of accidents in work provided by the State Administration of Work Safety, not only does this paper analyzes the statistics of work safety accidents but also explores the development trend of work safety accidents from 2001 to 2008. (accidents in work safety are classified by the number of deaths). At the same time, the ARIMA-BP, ARIMA-RBF and ARIMA-GRNN nonlinear combining forecast model are constructed. At last, this paper constructs a double nonlinear combining forecast model by combining three single nonlinear combining forecast models with RBF network. The results indicate that the accuracy of the double nonlinear combining forecasting models are higher than the accuracy of the single ARIMA model and three single nonlinear combining forecast models. This study provides a new method for promoting the development and improvement of strategic planning of work safety.

Key words: work safety; ARIMA model; BP model; RBF model; GRNN model; nonlinear combination