



北京高等教育精品教材

BEIJING GAODENG JIAOYU JINGPIN JIAOCAI

教育部新世纪优秀人才支持计划 (NCET-07-0056)

事故预测 理论与方法

PREDICT ACCIDENT



郑小平 高金吉 刘梦婷 著

清华大学出版社



北京高等教育精品教材

BEIJING GAODENG JIAOYU JINGPIN JIAOCAI

教育部新世纪优秀人才支持计划 (NCET-07-0056)

事故预测 理论与方法

郑小平 高金吉 刘梦婷 著

清华大学出版社

北京

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

事故预测理论与方法/郑小平,高金吉,刘梦婷著. —北京:清华大学出版社,2009.6
ISBN 978-7-302-19730-0

I. 事… II. ①郑… ②高… ③刘… III. 事故—预测 IV. X928.03

中国版本图书馆 CIP 数据核字(2009)第 038122 号

责任编辑:周菁

责任校对:王凤芝

责任印制:王秀菊

出版发行:清华大学出版社

地址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮编:100084

社总机:010-62770175

邮购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印刷者:北京嘉实印刷有限公司

装订者:三河市兴旺装订有限公司

经销:全国新华书店

开本:185×260 印张:17.75 字数:397 千字

版次:2009 年 6 月第 1 版 印次:2009 年 6 月第 1 次印刷

印数:1~4000

定价:30.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770177 转 3103 产品编号:028825-01

任何工程活动都有一定风险。工程活动的本性决定了工程是一把双刃剑,它可能给社会带来福祉,也可能给社会带来危害,甚至带来灾难。我国每年由安全问题造成的经济损失约占 GDP 总量的 5%,伤亡逾 100 万人。近年来冰冻和地震等自然灾害也使我国蒙受重大损失。面对重大自然灾害、技术灾害和环境恶化,以事故控制与预防为主的防灾减灾工程必然成为人类所面临的重要课题。在科学发展观的指导下,“安全发展和可持续发展”催生了人类对事故预测的研究动力。绝不要“带血的 GDP”,已经成为《政府工作报告》体现“和谐社会”和“以人为本”的民生需求。国家“十一五”规划的“复杂系统灾变形成及其预测控制”,国家自然科学基金委 2009 年重大研究计划“非常规突发事件应急管理研究”,科技部 2009 年 973 项目“应对重大灾难、突发事件的关键科学问题”均体现了事故预测研究的重要性和必要性。

中国古代思想家老子曾就防患于未然做过精辟的论述:“其安也,易持也。其未兆也,易谋也。其脆易判,其微易散。为之乎其未有,治之乎其未乱。”在科技高度发展的今天我们也应坚持“预防为主”的方针,防技术灾难于未然。事故预测是防灾减灾工程和安全管理的基石,然而相关的研究还处于起步阶段。通过研究事故致灾机理,分析事故发生发展和形成的规律,谋求事故预测的有效性,为防灾减灾提供科学依据,是本书作者研究的目的。以自然灾害,生产事故、社会性群体事件为研究对象,在参考、凝练了国内外学者关于预测理论研究成果的基础上,本书汇集了作者多年潜心研究成果和经验总结,期望能在一定程度上推动事故预测理论与方法的研究和应用。本书主要贡献体现为以下几点。

一是系统梳理了当前事故预测的主要理论,同时还详细刻画了一些代表性的预测方法,比如:趋势和因果预测方法,数值和概率预测方法,一元和多元预测方法,线性和非线性预测方法,确定和不确定性方法,白色、灰色和黑箱预测方法。

二是对现有事故预测成果进行了比较分析,解析了相关适用场景和未来研究趋势。因为不同预测方法的复杂性、数据要求、准确程度均不相同,应用中选取合适的预测方法是很关键的,也是比较难的。

三是对部分预测理论和方法进行了修补和扩展。作者基于大量事故源资料信息,通过数据挖掘等手段,提取和凝练事故预测的方法扩展和模型改进。

四是重点研究了若干组合预测模型及其应用案例。每种预测方法都具有描绘预测对象内部结构特征的独特信息特征,通过组合预测扬长避短,实现系统复杂性预测的有效性和完整性。

事故预测,特别是预测的准确性是工程科学最难的领域之一,也是永恒的课题,书中如有不当之处在所难免。本书是我所见到的第一部专门研究事故预测理论与方法的学术性著作,是非常有益的探索,相信能给安全科学与工程管理领域的学者和同仁提供参考和帮助。

中国工程院院士

高金吉

2009年4月15日

第 1 章 绪论	1
1.1 事故	1
1.1.1 事故的定义	1
1.1.2 事故的指标	2
1.1.3 事故的特征	3
1.2 事故预测	3
1.2.1 预测原理	3
1.2.2 事故预测过程	3
1.3 事故预测方法	4
1.3.1 回归预测法	5
1.3.2 时间序列预测法	6
1.3.3 马尔可夫预测法	7
1.3.4 灰色预测法	7
1.3.5 贝叶斯网络预测法	8
1.3.6 神经网络预测法	9
第 2 章 回归预测法	10
2.1 概述	10
2.1.1 回归分析概述	10
2.1.2 回归预测法概述	10
2.1.3 回归分析的事故预测概述	12
2.2 一元回归模型	12
2.2.1 线性化	13
2.2.2 参数估计	14
2.2.3 模型检验	18
2.2.4 预测模型	22
2.3 多元回归模型	24
2.3.1 参数估计	25

2.3.2	模型检验	27
2.3.3	自变量选择	29
2.3.4	预测模型	35
2.3.5	事故预测举例	37
2.4	线性回归注意的问题	37
2.4.1	残差分析	38
2.4.2	异方差问题	38
2.4.3	自相关问题	41
2.5	离散预测模型	44
2.5.1	泊松回归模型	44
2.5.2	负二项回归模型	47
2.5.3	logit 模型	48
2.6	事故预测实例	52
第3章	时间序列预测法	55
3.1	概述	55
3.1.1	时间序列简介	55
3.1.2	时间序列预测法概述	55
3.1.3	事故预测的时间序列预测法	57
3.2	基础知识	57
3.2.1	基本概念	57
3.2.2	平稳性和可逆性	58
3.2.3	滑动平均过程(moving average processes)	59
3.2.4	自回归模型(autoregressive model)	60
3.2.5	自回归滑动平均模型(autoregressive moving average model)	63
3.3	ARMA 建模	65
3.3.1	模型的识别和定阶	65
3.3.2	模型的参数估计	67
3.3.3	模型的检验	70
3.3.4	模型的预测	71
3.4	ARIMA 建模	75
3.4.1	平稳性的检验	76
3.4.2	非平稳数据的处理	80
3.4.3	ARIMA 模型	80
3.4.4	ARIMA 建模	83
3.4.5	ARIMA 季节模型	84
3.5	指数平滑模型	85

3.5.1	Brown 指数平滑模型	85
3.5.2	Holt 指数平滑法	89
3.5.3	Holt-Winters 指数平滑法	90
3.6	预测实例	92
3.6.1	ARIMA 预测	92
3.6.2	指数平滑模型	95
第 4 章	马尔可夫链预测法	98
4.1	概述	98
4.1.1	马尔可夫链简介	98
4.1.2	马尔可夫链预测简介	98
4.1.3	马尔可夫链事故预测	99
4.2	马氏链的基础知识	99
4.2.1	马氏链的基本概念	99
4.2.2	马尔可夫链的状态分类	100
4.2.3	平稳分布和遍历性	102
4.2.4	隐马尔可夫链	104
4.2.5	吸收态马尔可夫链预测模型	110
4.3	状态空间的划分	111
4.3.1	经验分组法	111
4.3.2	样本均值均方差分级法	112
4.3.3	有序样本聚类法	113
4.3.4	基于目标函数的模糊 ISODATA 聚类分析法	116
4.4	转移概率的计算和检验	119
4.4.1	马氏链转移概率的计算	119
4.4.2	马氏性的检验	121
4.4.3	齐次性的检验	121
4.5	马尔可夫链预测法	122
4.5.1	传统马氏链预测法及其改进	122
4.5.2	高阶马尔可夫链预测	124
4.5.3	隐马尔可夫链的预测算法	125
4.6	预测实例	125
第 5 章	灰色预测法	129
5.1	概述	129
5.1.1	灰色系统理论	129
5.1.2	灰色预测简介	129
5.1.3	五步建模思想	130

5.1.4	灰色事故预测	131
5.2	基础知识	132
5.2.1	灰生成	132
5.2.2	灰关联分析	134
5.3	GM(1,1)建模	135
5.3.1	GM(1,1)模型	136
5.3.2	残差修正模型	140
5.3.3	灰色灾变预测	141
5.3.4	GM(1,1)模型的适用范围	144
5.4	几种典型的 GM 模型	144
5.4.1	GM(M,N)模型	144
5.4.2	GM(1,N)模型	145
5.4.3	GM(0,N)模型	146
5.4.4	Verhulst 模型	147
5.4.5	灰色模型的适用场合	148
5.5	模型的改进	149
5.5.1	基于残差修正的改进模型	149
5.5.2	基于初始条件和信息更新的改进模型	150
5.5.3	基于数据变换的改进模型	152
5.5.4	针对内部建模机制的改进模型	156
5.6	事故预测实例分析	158
5.6.1	模型的建立	158
5.6.2	模型的检验	159
5.6.3	预测结果和分析	161
第 6 章	贝叶斯网络预测法	162
6.1	概述	162
6.1.1	贝叶斯网络简介	162
6.1.2	事故预测应用	163
6.1.3	贝叶斯网络的概念	164
6.2	预备知识	165
6.2.1	先验概率	165
6.2.2	信息论基础	167
6.2.3	势函数理论	169
6.3	贝叶斯网络学习	169
6.3.1	参数学习	170
6.3.2	结构学习	177

6.4	贝叶斯网络推理	183
6.4.1	变量消元法(VE)	183
6.4.2	超树推理法(polytree)	187
6.4.3	连接树推理法(junction tree)	188
6.5	扩展模型	192
6.5.1	定性贝叶斯网	192
6.5.2	高斯贝叶斯网络	195
6.5.3	动态贝叶斯网	199
6.5.4	多模块贝叶斯网	204
6.5.5	面向对象的贝叶斯网	207
第7章	神经网络预测法	210
7.1	概述	210
7.1.1	人工神经网络概述	210
7.1.2	神经网络预测概述	211
7.1.3	神经网络事故预测概述	211
7.2	神经网络的基础知识	212
7.2.1	人工神经网络的基本结构和模型	212
7.2.2	BP神经网络	218
7.2.3	RBF神经网络	223
7.2.4	神经网络建模的注意事项	227
7.3	神经网络预测模型	229
7.3.1	神经网络趋势预测	229
7.3.2	神经网络回归预测	231
7.4	预测实例	233
7.4.1	神经网络趋势预测实例	233
7.4.2	神经网络回归预测实例	235
第8章	组合预测	239
8.1	概述	239
8.1.1	组合预测法简介	239
8.1.2	组合预测的事故预测简介	239
8.1.3	预测的评价	240
8.2	模型组合法	241
8.2.1	灰色马尔可夫预测模型	241
8.2.2	灰色线性回归预测模型	247
8.2.3	ARIMA神经网络混合预测模型	250

8.3 结果组合法	254
8.3.1 非最优组合模型方法	254
8.3.2 最优组合预测方法	257
8.3.3 神经网络组合法	265
参考文献	268

事故预测,是基于可知的信息和情报,应用一定的预测技术和手段,对预测对象的安全状况和趋势进行预报和预测,从而达到事故优生的目的。

探索灾变传导过程中孕灾体、致灾体、受灾体的脆弱性、传染性和稳定性,提取和凝练其机理机制,是安全科学最核心的问题。当安全科学研究的重点从探索事故致因转移到探索事故优生,标志着事故预测正日益成为当前安全研究最基本的科学问题之一。事故预测的精确性必须基于两个最基本的前提:一是可知的信息,目前世界各国政府和组织已经意识到建立灾害数据库的重要性,也正积极开展工作;二是正确的事故预测方法,目前关于事故预测理论与方法的研究成果较少。理论界对于事故预测理论和方法缺乏系统性,而业界在进行事故预测时也往往难以选择合适的预测方法。预测是现代安全管理的重要组成部分,要做好安全预测工作,首先必须学习事故预测的基本理论和方法。

1.1 事 故

1.1.1 事故的定义

Berckhoff 在《生产和防止事故》一书中,定义事故为:人(个人或集体)在为实现某种意图而进行的活动过程中,突然发生的、违反人的意志的、迫使活动暂时或永久停止的事件。一般意义上的事故是生产过程中的事故,就是在工程建设、生产活动以及交通运输等社会活动中发生的可能带来物质损失和人身伤害的意外事件。本书所讨论的事故是一个更广义的概念,可以理解为:一项人们主观上不愿意出现的、导致人员伤亡、健康损害、环境及经济损失的意外事件,主要包括自然灾害事故,生产安全事故和社会性安全事故。

1. 自然灾害事故

自然灾害是地球上的自然变异,包括人类活动诱发的自然变异,这种灾害无时无刻不在发生,当这种变异给人类社会带来危害时,即构成自然灾害。主要包括旱涝灾害、气象灾害、地震灾害、地质灾害、海洋灾害、生物灾害和森林草原火灾等。例如我国四川汶川大地震、缅甸水灾、美国“卡特里娜”风灾、澳大利亚雹灾等。

2. 安全生产事故

安全生产事故是指在与生产经营有关的活动中突发的,由于违章作业、维护不周、操

作失误、设计缺陷等原因,导致原生产经营活动暂时或永远停止的事件。主要包括:交通事故、火灾事故、矿业安全事故、建筑业安全事故以及不合格产品等。但不包括没有正确使用产品以及并非由产品缺陷所导致的事故。例如“4·28”胶济铁路事故、吉林“2·15”火灾事故、重庆开县天然气井喷事件、“欣弗”药品事件、日本水饺事件、“松下”暖风机事件等。

3. 社会性事故

社会性事故是指危害公共安全的事件。根据影响公共安全的因素,社会性事故主要包括以下一些类别:群体性事件、暴力事件、经济危机、公共卫生事故。例如四川广安群体事件、拉萨暴力事件、东南亚金融危机、非典事件以及近期美国次贷危机引起的全球金融海啸等。

1.1.2 事故的指标

事故指标是反映事故发生和伤害情况的一系列特征量,对于不同事故的统计有不同的事故指标。最常见的是绝对指标和相对指标。绝对指标是描述事故发展的规模、水平的统计指标,包括事故发生次数、死亡人数、受伤人数、损失的工时数和经济损失。相对指标是反映事故的发展程度或比例关系的指标,是相对某一模式的事故发生的比率。例如相对人员模式的相对指标有10万人死亡率,相对生产产量模式的指标有百万车次事故率等。本书主要讨论基于事故指标的预测,如某行业事故的发生次数、伤亡率、事故经济损失等。根据我国国家安全生产监督管理局的资料,安全生产领域的事故指标体系包括五大绝对指标和四大相对指标^①,如图1-1所示。

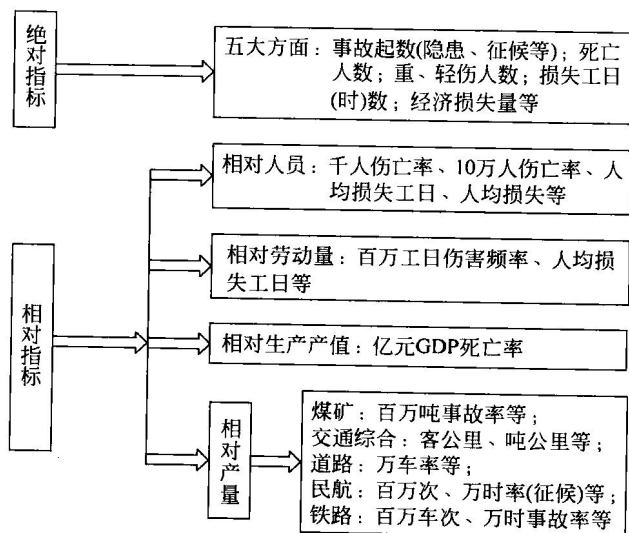


图 1-1 事故指标体系

^① 罗云,吕海燕,白福利. 事故分析预测与事故管理[M]. 北京:化学工业出版社,2006.

1.1.3 事故的特征

事故是人们在生产生活中不期望发生的各种事件,事故普遍具有如下的特征。

(1) 因果性。事故的因果性是指一切事故的发生都是有其原因的,这些原因就是潜伏的危险因素。这些危险因素有的来自于人的行为和管理缺陷,也有的来自于物和环境的不安全状态。在一定的时间和空间内危险因素相互作用,就会导致系统的隐患、偏差、故障、失效,以致发生事故。因果关系表现为继承性,原因的多层次性和演化性。

(2) 随机性和规律性。事故的随机性是指事故的发生是偶然的,同样的前因随时间的进程导致的后果不一定相同,但是在偶然的事故中孕育着必然性,必然性通过偶然性事件表现出来。对于个别事故具有不确定性,但对大样本则表现为统计规律性。概率论、数理统计与随机过程等数学工具,是研究具有统计规律性的随机现象的重要手段。找出事故发生、发展的规律,认识事故,为预防事故提供依据。

(3) 潜伏性和再现性。事故的潜伏性是指事故在尚未发生或还没有造成后果前,各种事故征兆是被掩盖的,系统表面看似处于安全状态,但实际上却存在危险因素。潜伏性使得人们对认识事故、弄清事故发生的可能性及预防事故成为一件非常困难的事情。这就要求人们总结已发生事故的经验,不断地探索和研究事故预测的模型和方法,提高预测的精度。如果没有真正了解事故发生的原因,并及时采取有效措施,类似事故就还会出现;而如果积极做好预测预防工作,则是可以避免事故的再现性的。

1.2 事故预测

1.2.1 预测原理

事故的发生往往具有随机性,而且导致事故的原因往往潜伏着多种复杂因素,这就给预测带来了很大的难度,一般来说,事故预测可以遵循以下基本原理。

(1) 可知性原理。根据科学试验和事故经验,人们可以获得关于预测对象发展规律的感性和理性认识,从中发现导致事故的影响因素,通过总结它的过去和现在来推测未来的变化趋势和可能出现的危害,这是一切预测活动的基础。

(2) 连续性原理。预测对象的发展是连续的过程,现在的安全状态是过去状态的演变结果,未来的安全状态是现在安全状态的演化。对于同一个事物,可以根据事物发展的惯性,来推断未来的发展趋势,这是预测中的时序关系预测法的理论基础。

(3) 可类推原理。预测的事件必然是具有某种结构的,如果已经知道两个不同事件之间的相互制约关系有着共同的发展规律,则可利用一个事件的发展规律来类推另一个事件的发展趋势,这是预测中因果关系预测法的理论基础。

1.2.2 事故预测过程

预测由预测目标、预测信息、预测模型和预测结果4个部分组成^①。根据预测对象的

^① 隋鹏程,陈宝智,隋旭.安全原理[M].北京:化学工业出版社,2007.

不同,预测程度也不一样。预测过程见图 1-2。

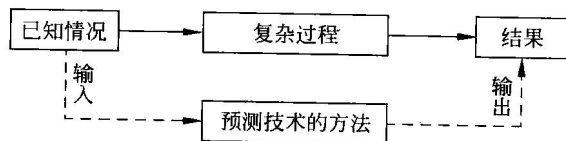


图 1-2 预测过程

(1) 确定目标。对于事故预测,预测的目标是为了探求事故发生的趋势和内在的规律,以便分析出未来事故发生的可能性,提前采取安全对策,做出预防工作,使事故的风险控制在可接受的水平。在明确事故预测的目标之后,就可以确定收集什么资料。除此之外还要根据预测对象、预测期限,明确预测的性质和内容,从而组织预测人员,做好搜集资料的工作。

(2) 收集信息。准确而全面的信息收集是预测的基础,它直接影响预测结果的精确性。需要收集的信息不但要有预测对象的现状信息,而且要有历史信息以及相关信息。其中反映事物发展的历史数据,例如 1995—2007 年某城市发生的火灾次数的历史统计资料,称为纵向统计资料(也称作时间序列数据);而某一特定时间内对于同一预测对象所需的各种相关的统计资料,例如 2007 年我国各大城市发生火灾的次数,称为横向统计资料(也称为截面数据)。对于收集到的信息,还需要注意观察,在处理数据时要进行客观而全面的分析,对已掌握的资料也要进行周密的检查和筛选,使资料、数据的误差降低到预测所允许的范围之内。

(3) 构建模型。建模首先是要选择预测方法,然后再设计预测模型,进行预测。据不完全统计,现有的各类预测方法达 300 种之多,预测方法的选取应根据预测对象、信息资料、预测目标等来确定,预测方法的确定是一个渐进的过程,也是最关键的环节。基于确定的预测方法建立起多参数的预测模型,通过对信息数据的处理,来选取和识别模型参数,再通过推理判断,揭示出事故优生的内在规律性。在预测的结构模型中,不同参数决定了模型的性质。预测的实质就在于把所获得的数据资料输入到预测模型中,确定参数的取值,然后通过运算和分析,求出初步的预测结果。在事故预测中,根据预测值和历史事件的比对,预测检验是必要的,甚至还必须对预测模型进行修正和对误差原因进行分析。

(4) 输出结果。预测结果就是在预测分析的基础上最后提出的事物发展的趋势、程度、特点以及各种可能性结论。对于事故预测,通过对结果的鉴定来考察预测的实际意义,找出预测与现实之间的误差大小。根据预测结果,实行动态跟踪,及时制定可能发生事故的防范措施。

1.3 事故预测方法

预测可分为定性预测法和定量预测法,其中定量预测法还分为时序关系预测法和因果关系预测法两类。时序关系预测法,是把客观事物发展的惯性趋势在时间轴上用变量随时间的变化规律表现出来,用变量把以往的纵向统计资料进行外推的预测方法。这种

预测方法主要是用于预测对象的内在发展趋势明确的情况,而且需要收集纵向统计资料。因果关系预测法,则是根据观测对象的依存关系,找出预测所需要的因果信息,以事物变化的因果关系为基础,用一种近似的函数关系表示出来,并依靠历史数据,构建相应的因果模型。这种预测方法比较适用于事物之间的因果关系清晰且具备比较全面的横向统计资料的情况。

由于定性的方法主观因素较强,多用于定性的分析,往往不能得到精确的预测结果。本书以事故发生的指标为预测对象,重点介绍了事故预测的定量方法,它们分别是回归分析预测法、时间序列预测法、马尔可夫链预测法、灰色预测法、贝叶斯网络预测法和神经网络预测法。其中时间序列预测法、马尔可夫链预测法、灰色预测法和神经网络预测法都属于时序关系预测法;回归分析预测法和贝叶斯网络预测法是因果关系预测法。一般而言,时间序列的纵向统计资料比较容易收集,所以会发现时间预测法的应用较多。但是也有些情形,往往需要对事故发生的因素及其相互间的关系作更深入的分析,这就需要运用因果关系预测法。实际上,各种预测方法都有自身的优势和缺陷,往往需要针对不同事故的模式和特征选择同自身相适应的预测方法。然而建模者对变量的选择和数学模型的选择往往有一定程度上的主观性和经验性,若要保证事故预测的有效性,目前比较好的方法是利用不同方法的组合预测,因为集成分散的单个预测方法的优势,可以减少总体的不确定性,有效提高预测的精度,这一方法在最后一章中也将给予介绍。

1.3.1 回归预测法

回归预测是根据历史数据的变化规律,寻找自变量与因变量之间的关系,确定模型参数,据此作出预测的方法。根据自变量的个数可将回归问题分为一元回归和多元回归。按照回归方程的类型又可分为线性回归和非线性回归。

多元线性回归在事故分析时可能会带来一些不必要的统计特征,因此为了弥补多元线性回归的缺陷,提出了用泊松分布来对事故发生的概率建模,而当样本数据过度离散(over-dispersion)时,泊松模型可能无法准确描述其概率分布,或过高估计事故发生的可能性,这时可以选择负二项分布。泊松分布和负二项分布均是常用的离散型分布(如 Abdel-Aty 和 Radwan,2000^①; Evans,2003^②)。Maher 和 Summersgill(1996)^③以交通事故为例,将广义线性回归(GLMs)、泊松和负二项模型的适用性进行了比较,发现基于最小二乘法,具有泊松误差结构的线性回归比传统的多元线性回归更适用。考虑到观察样本中出现较多零事件的情况,此时泊松和负二项式模型不能解释这种分布,从而在此基础上提出了 ZIP 回归预测模型(zero-inflated Poisson)和 ZINB 预测模型(zero-inflated negative binomial),这种模型具有两个状态,也就是将零状态从原有的数据状态中分离出来(Shankar

① Mohamed A. Abdel-Aty, A. Essam Radwan. Modeling traffic accident occurrence and involvement[J]. Accident Analysis and Prevention,2000,32: 633~642.

② Andrew W. Evans. Estimating transport fatality risk from past accident data[J]. Accident Analysis and Prevention,2003,35: 459~472.

③ Michael J Maher, Ian Summersgill. A comprehensive methodology for the fitting of predictive accident models[J]. Accident Analysis and Prevention,1996,28(3): 281~296.

et al., 1997^①)。之后考虑到事故分布中的异质性(heterogeneity), Milton et al. (2007)^②将“混合 logit”(mixed logit)模型用到事故预测中,来评估整体而不是局部的事故分布。

当预测的长度大于占有的原始数据长度时,采用回归方法进行预测在理论上不能保证预测结果的精度;另外,可能出现量化结果与定性分析结果不符的现象,有时难以找到合适的回归方程类型。因此,针对诸多不同的事故因素和可能后果,要选择不同的模型,而且事故与各个因素之间的关系可能相当复杂,对建立的模型进行检验是很重要的,可以发现这些样本数据之间是否存在重要的区别,从而判断是否需要增加或合并数据,是否需要针对事故的不同情况分别建立模型(Abbas, 2005^③)。

1.3.2 时间序列预测法

时间序列的变化受许多因素的影响,概括的讲,可以将影响时间序列变化的因素分为四种,即长期趋势因素、季节变动因素、周期变动因素和不规则变动因素。在时间序列分解模型的基础上,对四种变动因素有侧重地进行预处理,从而派生出剔除季节变动法、移动平均法、自回归法、时间函数拟合法等具体预测方法。在事故预测中,常用的方法有指数平滑预测法和 ARIMA 预测法。

指数平滑预测法是在加权平均法的基础上发展起来的,其数据的重要程度按时间上的远近成非线性递减。这种方法一般用于实际数据序列以随机变动为主的场合,可以消除时间序列的偶然性变动,进而寻找预测对象的变化特征和趋势。Hyndman 等人(2002)^④对指数平滑法进行了分类和描述,总结出 15 种模型,每种模型包含一种趋势成分(none, additive, damped additive, multiplicative, damped multiplicative)和一种季节成分(none, additive, multiplicative)。ARIMA 预测法^⑤则主要试图解决以下两个问题:一是分析时间序列的随机性、平稳性和季节性;二是在对时间序列分析的基础上,选择恰当的模型进行预测。其预测模型分为:自回归模型(AR)、移动平均模型(MA)和自回归移动平均混合模型。对于某些不平稳的序列经过差分变换后的序列再运用 ARIMA 模型,习惯上称为 ARIMA 模型,即求和自回归移动平均模型。当 ARIMA(p, d, q)中“ d 为小数($d > 0$)的情况下”,也可以使用,即 ARFIMA 模型,它常常被用在假设检验中来决定这个序列是否具有长期记忆(long memory),也可以用来进行 k 步(k -step ahead)预测(Reisena et al., 1999^⑥)。

① Shankar V, Milton J and Mannering F. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry[J]. Accident Analysis and Prevention, 1997, 29(6): 829~837.

② Milton J C, Shankar V N, Mannering F L. Highway accident severities and the mixed logit model: An exploratory empirical analysis[J]. Accident Analysis and Prevention, 2007, 40(1): 260~266.

③ Khaled A. Abbas. Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt[J]. Accident Analysis and Prevention, 2004, 36: 149~163.

④ Hyndmana R J, Koehler A B, Snyder R D, Grose S. A state space framework for automatic forecasting using exponential smoothing methods[J]. International Journal of Forecasting, 2002, 18: 439~454.

⑤ Box G E P, Jenkins G M, Gregory C Reinsel, Gu Lan translation(顾岚译). Times series analysis forecasting and control (时间序列分析预测与控制)[M]. Beijing: the Statistics Press of China(中国统计出版社), 1997.

⑥ Reisen V A, Lopes S. Some simulations and applications of forecasting long-memory time-series models[J]. Journal of Statistical Planning and Inference, 1999, 80: 269~287.