

文章编号: 1006-8309(2000)04-0032-03

• 技术与方法 •

人误的相关分析方法

李永娟, 王二平

(中国科学院心理研究所, 北京 100101)

摘要: 文章主要介绍通过相关分析的方法进行人误的研究, 寻找完成作业所需能力的高低和人误概率之间的关系。Fleishman 等人利用相关分析的方法研究发现完成任务所需要的能力越高, 人误的概率越高, 并且认知能力比知觉-运动能力更容易影响人误。研究还得到了核电领域人误概率对 5 种不同职务所需能力的回归方程, 结果具有较高的信度和效度。

关键词: 人误; 相关分析方法

中图分类号: X913 **文献标识码:** A

科学技术发展到今天, 直接的机械故障导致事故的比例已经下降到很低, 甚至可以达到 10^{-18} 的水平。但是 20 多年来, 人类接连的灾害事故表明, 在技术密集、包括技术设备、人以及组织三类元素的社会-技术系统中, 只靠发展技术设备是不可能保证系统安全的。Meister 提出, 20% ~ 50% 的机械故障与人的失误(Human error, 简称人误)有关。另有研究表明, 核电领域 20% ~ 70% 的系统失败归因于人误; Helmreich 提出航空领域 2/3 以上的事故都是因为机组成员的失误导致的。^[1] 因此, 人的可靠性对系统安全起到至关重要的作用。

1 人误和人的可靠性的定义

Reason 将人误定义为在执行预定目标过程中计划行为的失败;^[2] 或指人的意向性计划或动作, 在没有外力干预的前提下, 没有取得他所期望的结果或没有达到预期的目标。^[5] Senders 和 Moray 在分析总结不同定义的基础上, 将人误定义为操作者没有意向(intention)、规则或外部观察者没有期望却导致任务或系统超过可接受阈限的操作者行为。^[1] ASSET 认为当人没有按照预期执行任务时, 任何事情都可能成为人误。^[4]

在人机系统中, 人的可靠性是指人在规定的条件下和规定的时间内完成规定功能的能力。^[3] 它与人误虽然不是同一概念, 但两者关系非常密切, 许多研究中都不区分这两个词。但人误与违章(violation)是两个完全不同的概念, 违章是指故

意的违反某种规则或某种社会认可的行为规范。^[4] 违章是组织错误的表征, 需要通过改善组织的安全管理制度杜绝违章。^[5]

根据不同的标准, 可以将人误分成不同的类型。被研究者们广泛接受的一种分类方式是由丹麦心理学家 Jens. Rasmussen 提出的。Rasmussen 1983 年根据认知心理学的理论, 将人的认知活动分为技能基(skill based)、规则基(rule based)、知识基(knowledge based)三类, 并根据这个分类标准将人误分为相应的三种类型。技能基的行为一般是在无意识状态下发生, 是一种自动化、高度整和模式的行为; 规则基的行为是指在一个熟悉的工作环境下, 操作者按照记忆中存贮的规则进行操作; 知识基的行为是在任务操作过程中出现的程序和规则中没有的情况或事件, 操作者依靠自己的知识经验进行分析决策。^[6] 不同的行为过程中可能产生相应类型的人误。

2 概率估计

人类对复杂系统安全的研究是从研究技术设备开始的。技术设备的可靠性在二战后期被提出来并取得了长足的发展, 它主要通过经验概率评估得到。随后, 系统安全工程的概念提出, 开始从整个技术系统角度评估技术系统集成的可靠性。60 年代后期, 人们对复杂系统安全性研究不仅从技术上寻求突破, 也开始考虑到人的可靠性。1971 年 IEEE 发表的文集是人误(Human error)和人的可靠性(Human reliability)研究的里程碑。

人误和人的可靠性研究也成为近年来相关专家研究的重点。据不完全统计, 仅从 1994 年到 1997 年间, 在 PSYCHLIT 中收录的与人误有关的文献就近千篇, 这些文献主要是对航天、航空、核电、化工、石油化工、铁路、电力等领域安全与人误的研究结果。可见, 人误与人的可靠性是各种工业组织中共同关心的问题。

传统的人的可靠性分析(Human Reliability Analysis, HRA) 很多都是借用技术设备可靠性的思路, 通过概率估计方法进行定量分析。人的可靠性分析技术的代表方法有人失误率预测技术(technique for human error rate prediction, THERP)、操作员动作树分析方法(operator action trees, OAT)、人的认知可靠性模型(human cognitive reliability, HCR)等。^[7]

传统概率估计的方法, 能够直观地描述人误和人的可靠性, 但没有对人本身的特性和人误机制进行深层探讨。并且知识基的错误是一种开放的人误, 传统的经验概率是不可能估计的。同时, 由于组织错误尤其是管理失误也是一种知识基的错误^[5], 也不能由传统的经验概率估计。而当违章成为事故的贡献因素时, 这种概率就成为一个不可重复的值, 因为违章的出现是没有规律可循的。因此研究者们一直在探索新的评判人误和人的可靠性的方法。

3 相关分析和回归分析

Fleishman 等研究者试图根据执行职务任务所需的能力建立模型来预测人误。针对核领域在任务执行过程中, 很少能够收集到可能出现错误的客观绩效数据的现实, 开发出一种新的评价人误概率(Human Error Probabilities, HEPs)方法。他们利用美国空军开发的职务绩效测量系统(Job Performance Measurement System, JPMS)提供的数据进行分析, JPMS 是美国空军对职业专家评价职务熟练程度的效标。首先利用行为描述(behavior description)的方法进行作业分析, 对这些具体作业用行为动词(例: 计算、校对、安装等)进行分类。对专家被试通过职务样本测试进行 JPMS 测定, 得到各种具体作业(task)中人误概率的平均值和标准差。测试结果表明不同行为动词出现人误概率的高低不同, 并且出现人误概率高的作业与人的认知能力有关; 而出现人误概率低的行为动词与人的知觉-速度定向能力有关。因此, 可以将完成这些任务所需要的能力分成两类:

认知能力(cognitive abilities, 例, 计算)和知觉速度定向的能力(perceptual motor oriented abilities, 例, 安装)。

为了寻求核电领域与空军作业的一致性, 他们利用 Fleishman 和 Quaintance 在 1984 年开发的能力需求方法寻求两个领域内共同的比较标准(yardstick)。这里的能力(ability)是指个体具有的普遍的能够完成不同作业的能限(capacity), 而不是指技能(skill), 即完成某一或某一组具体任务的熟练水平。从 Fleishman Job Analysis Survey (F-JAS) 7 点量表中选择 16 种相关的能力作为评价指标, 同时选择核电领域 4 种职务共 34 种作业, 空军领域选择 4 种职务共 32 种作业。分别由各领域专家被试对完成这些作业所需要的各种能力高低进行评价。结果得到各种作业所需不同能力的平均数和标准差。通过对 32 种空军作业的能力需求和 JPMS 测得的人误概率结果进行相关分析, 发现需要相似能力的作业出现的人误概率也相似; 需要某一较高峰值(peak)能力的作业趋向于较高的人误概率; 需要 2 种以上中等以上能力的作业比需要 2 种或以下中等能力的作业有较高的人误概率; 需要较高认知能力的作业比需要知觉-运动能力的作业有高的人误概率。这种方法具有较高的评定者间信度。因此 F-JAS 利用能力需求问卷提供了一种比较可靠的作业分类方法。

尽管核电领域与空军领域有一些差异(例如空军一般单独作业, 而核电领域需要多人合作等), 但它们的作业相对等级的人误概率可能是一致的。因此, 根据空军研究的思路, Fleishman 等人开始在核电领域进行研究。同空军研究相似, 选择几种特定职务的具体作业, 根据空军研究结果, 从 16 种能力中选择 8 种表现出与人误概率有很强独立相关的能力, 其中包括 5 种认知能力(书面理解, 对问题敏感性, 口语表达, 演绎推理和形象化)和 3 种知觉-运动能力(控制精确性, 近距离视觉和手动灵活性)。并且鉴于核电领域客观的人误概率数据比较难以得到, 重新发展了人误概率评定量表。根据文献、专家评定以及模拟器数据, 得到了各职务通用核电作业的客观错误比率, 并将这些作业作为标杆锚定(benchmark anchors), 根据它们可能出现人误的概率分别将它们放置在 9 点人误可能性评定量表的不同位置。这样评定其它作业的专家可以根据这个标准确定

在正常操作条件下,完成其它作业所需要的能力,以及可能出现人误的在量表上的等级。

对评定结果进行信度分析,发现评定者间的信度比较高。力量表的评定者间信度为 0.83, HEP 量表的为 0.64,属于可以接受的水平。相关分析的结果发现,完成任务所以序的能力与人误概率都有较高的相关,并且大部分相关都有显著性;认知能力与人误概率的相关普遍高于知觉-运动能力与其相关,尤其是对问题敏感性、书面理解和演绎推理表现出非常强的相关,相关系数分别为 0.68, 0.65, 0.65。同时,对两种不同核电站(压水堆和沸水堆)和各种不同职务进行分析,发现尽管沸水堆核电站与人误概率的相关相对较低,两种核电站中任务相对等级在人误概率上是非常相似的,并且不同领域的相同职务作业与人误概率的相关程度比较相似(例都是控制室监控任务)。同时对影响人误概率的能力进行进一步

的回归分析,得到各种能力与人误概率之间的回归系数。表 1 中,共得到 5 种职务的人误概率对完成作业所需的各种不同能力的回归方程。其中,各种不同能力下的各行值分别代表该能力在相应职务中参与回归方程的回归系数;R 值分别是不同职务中各种能力与人误概率的复相关系数;R² 是决定系数,它是指回归平方和在总离差平方和中所占的比例,R² 值的大小反应了表 1 中各种能力对人误概率值的解释能力。并且这个结果具有较高的交叉效度,表明在正常操作条件下,根据能力需求评定,多重回归模型能够有效地预测不同职务间和不同核电站间的人误概率。^[8]

这种相关和回归分析的方法,将对人误概率的研究转化成对完成作业所需要的能力高低的研究。而只要找到有效的参照标准,完成作业所需能力的相对高低就不难求。这样,通过能力来预测人误概率的目的也就可以顺利达到了。

表 1 各种职务的多重回归方程

职务分类	R	R ²	书面理解	控制精确性	问题敏感性	近距离视觉	口语表达	手动灵活性	演绎推理	形象化	数字能力	阻止能力	标准误
A	0.841**	0.707	0.663	0.037	0.123	-0.777	-0.094	0.153	0.219	0.416	0.055	10.572	0.437
B	0.848**	0.719	0.337	-0.127	0.601	0.251	0.041	-0.067	-0.413	0.243	0.170	-0.179	0.408
C	0.787**	0.619	0.193	0.133	0.328	-0.075	0.136	-0.349	0.301	0.129	-0.138	0.301	0.502
D	0.748**	0.560	0.121	-0.022	0.128	-0.240	-0.064	0.129	0.210	0.071	0.002	0.902	0.208
E	0.734**	0.539	-0.168	0.174	0.150	-0.014	0.298	-0.128	-0.019	0.275	0.082	0.511	0.414

注:所有的 R 都在 P < 0.0005 水平上显著

4 讨论

Fleishman 等人的研究中存在很多缺陷,首先评定者的数量和作业相对都比较少,因此受无关因素的影响比较大。并且这个研究只选择了 F-JAS 量表的 50 多种能力中的 8 种,也许完成某项作业需要更多的能力,而且这项研究只考虑了能力与人误概率之间的关系,也许还有更多的影响人误概率的因素需要考虑,例职务类型,核电站类型,影响绩效的工作环境:噪音,高温,工作轮换,培训等;这些因素都是研究者意识到并且正在努力解决过程中。但这项研究至少提供给我们一个新的研究思路,探索出完成作业所需能力和人误概率之间的关系,并且不同能力对人误概率的影响程度也可能不同。同时,这项研究也为核电站的人员选拔和培训提供了一个很好的思路,可以针对不同的职务所需的不同能力选择合适的人才,同时,对不同职务的职员进行不同重点的培训。

90 年代初开始,人误研究者们发现,只从个体角度出发研究人误只是预防人为事故的必要而非充分条件。事故的发生可能是由于个体失误,也有很多是由于组织的问题导致的,即组织错误(organizational failures)。组织错误是一个非常广泛的概念,它从本质上讲也是一种人误,但已不属于传统个体人误的范畴。可以将组织错误理解为个体在行为或决策时由于知识经验的缺乏而出现的人误没有及时发现、控制而转化成的管理制度、规则、程序、政策、战略决策的缺陷或错误。它们或者成为潜在错误,威胁着组织的防御机制,或者与个体人误或违章一起成为事故的“贡献因素(contributing factors)”。相对于个体人误来讲,组织错误的发现在时间上具有一定的滞后性。组织错误可以表现在管理错误、制度缺陷、沟通不良、组织文化薄弱、培训不足、协调不充分等各个方面^[5]。如果单纯用概率方法来分析这些失误或者事故是远远不够的。借鉴 Fleishman 等人的方法,通过相

文章编号: 1006-8309(2000)04-0035-04

诊查型用户界面可用性评价方法(IM)(下)

——比较与建议

吴昌旭, 张侃

(中国科学院心理研究所, 北京 100101)

摘要: 首先对文(上)评述的7种诊查型方法(Inspection Methods, IM)进行了多维度定性综合比较,使评价人员能在实际运用中,根据不同方法的特点与要求选择合适的评价方法;通过比较还发现其中的工效学检查表有较多优势;然后,本文对各种诊查型方法之间进行定量比较的方法作了探新,并在总体上对现有诊查型方法存在的问题提出了一些建议。

关键词: 人机界面; 用户界面; 可用性评价; 诊查型方法

中图分类号: TB18; TP31 **文献标识码:** A

1 诊查型各方法综合比较

7种诊查型方法虽各自均有存在优缺点,但它们之间缺少系统的比较。没有比较,评价人员就很难在这些方法中依据实际情况选择合适的方法;没有比较,就很难分析方法之间的功效差异。

1.1 各诊查型方法之间的多维度定性比较

我们综合了Dix, Virzi与Nielsen等的研究成果^[1,2,3],运用了Barry Kirwan比较多种人误预测法的比较方法^[4],对这7种方法进行了多维度定性比较,见表1。

比较结果: 能使人机界面评价人员根据不同方法的特点与要求,选择较为合适的可用性评价方法;通过计分计算发现,在无需可用性专家参与的各方法中,工效学检查表(累计得分25.5)具有较多的优点与较高的可操作性,尤其适合在我国目前缺少可行性评价专家等情况限制下运用。

1.2 各诊查型方法之间的定量比较的方法

从目前所收集到的资料看,尚未见到有关这些方法之间的定量比较的研究,因此我们认为有必要对定量比较的方法加以探索。

1.2.1 对方法的评分者间信度的计算 可以求两个或两组评价人员的所预测的可用性问题的类型与数量的相关系数,或者以预测的可用性问题的严重等级计算不同方法的肯德尔和谐系数;方法的评分者间信度越高,说明这一方法获得的评价结果就越不容易受到评价人员个体差异的影响。

1.2.2 对方法同时效度(Concurrent Validity)的计算 借用对不同人因错误分析技术同时效度的计算方法:^[6]

设 $a =$ 用IM预测到的和实际观察到的可用性一致的问题数量(即击中的问题数量)

关分析找出各种错误与事故之间的关系,对于通过前馈控制预防各种失误和事故会有积极的作用。

参考文献

- [1] Robert L, Helmreich. Managing human error in aviation[J]. Scientific American, 1997, 276(5): 62-68.
- [2] Reason J. Human Error[M]. UK, Cambridge: Cambridge University Press, 1990.
- [3] 朱祖祥. 人类工效学[M]. 杭州: 浙江教育出版社, [J].
- [4] Reason J. Driving errors, driving violations and accident involvement[J]. Ergonomics, 1995, 38(5): 1036-1048.

- [5] 王二平. 人误研究的组织定向[J]. 人类工效学, 1999, 5(1): 44-47.
- [6] Rasmussen J. Mental procedures in real-life tasks: A case study of electronic trouble shooting[J]. Ergonomics, 1974, (17): 293-307.
- [7] 高佳, 黄祥瑞. 人的可靠性分析研究的进展[J]. 人类工效学, 1996, 2(4): 52-57.
- [8] Fleishman ED, Buffardi LC, Monach RA, et al. Development of a Model to predict Human Error Rates From the Ability Requirement of Job Tasks[M]. NRG 04-91-361, 1994.

[收稿日期] 1999-12-13

[修回日期] 2000-10-27