

系统建模与仿真应用的校验、确认与验收

王仁春,李昊,戴金海

(国防科技大学航天与材料工程学院,湖南长沙410073)

摘要:以美国国防部(DoD)近年来对建模与仿真(M&S)应用所做的校验、确认与验收(VV&A)研究工作为基础,介绍了关于M&S应用的校验、确认、验收、鉴定等概念;结合软件工程的3P指标提出一种通用的形式化表示的M&S应用VV&A框架;讨论了M&S应用可信度指标的层次化体系的建立与评价方法,引入了Kiviat图;最后介绍了一个用于M&S可信度评价的计算机软件平台——EE(Evaluation Environment,评价环境)。

关键词:建模与仿真应用;可信度;校验、确认与验收

中图分类号:TP391.9 **文献标识码:**A

Verification, Validation and Accreditation of System Modeling and Simulation Applications

WANG Ren - chun, LI Hao, DAI Jin - hai

(College of Astronautics and Material Engineering, NUDT., Changsha, Hunan410073, China)

ABSTRACT:Based on works in Verification, Validation and Accreditation (VV&A) of Modeling and Simulation (M&S) applications recently conducted by the Department of Defense (DoD), some concepts about M&S applications such as verification, validation, accreditation, and certification are presented. Then by integrating with the 3P indicators of software engineering, a universal formalized VV&A architecture for M&S applications is built up. The creation and evaluation of the hierarchy of the credibility indicators for M&S applications are discussed, especially Kiviat graphs are presented. At last a software platform, EE (Evaluation Environment), which can support the evaluation of the M&S credibility is introduced.

KEYWORDS: Modeling and simulation applications; Credibility; Verification, validation and accreditation (VV&V)

1 引言

系统建模与仿真(M&S)是现代科学技术研究与系统分析、设计的有效手段,其有效性建立在模型有效、可信的基础上。为保证M&S应用的有效性和可信度,必须在M&S应用的全生命周期内独立并行地开展校验、确认与验收(VV&A)工作,这对M&S应用的成功与否具有决定性意义^{[1][2]}。

很多国家的研究部门和职能机构在M&S应用的VV&A领域开展了大量研究工作,其中以美国国防部(DoD)最具代表性,其下属的国防建模与仿真办公室(DMSO)专门编撰了VV&A推荐实践指南(RPG)^[3],其中对VV&A的概念、理论、方法和相关技术作了一系列深刻论述。其他国家也相继建立了与VV&A有关的认证组织,如:英国的验收服务组织(<http://www.ukas.com>),日本的一致性评价验收理事会(<http://www.jab.or.jp>),德国的验收委员会(<http://www.dar.bam.de/indexe.html>)等。在学术界和工程界也开展了大量

的研究工作^{[4][5][6][7][8]}。

国内许多学者也开展了与VV&A有关的研究。但总的看来,有如下不足:

- 1)概念不统一,不完备;
- 2)应用对象不明确;
- 3)实施过程缺乏目的明确、完备的阶段划分;
- 4)缺少评价指标和评价方法;
- 5)没有可供使用的软件工具。

针对上述问题,本文借鉴当前国际上在此领域的最新研究成果和成功经验,从理论、方法和软件工具等方面系统介绍了M&S应用的VV&A,内容上具有权威性、规范性、通用性、完整性和可行性,对VV&A的实施具有指导意义。

2 M&S应用的VV&A

总地说来,VV&A是确保M&S应用总体有效性和可信度的过程,这一过程需要贯穿M&S应用的全生命周期。与VV&A相关的活动包括:校验(Verification)、确认(Validation)、验收(Accreditation)与鉴定(Certification),前三者的缩

写合在一起即人们常说的 VV&A。

校验和确认是 VV&A 的基础,这两项活动体现了 M&S 应用的 VV&A 过程的理论和方法水平,因此有时也将校验和确认合称 V&V。令 a 表示在 M&S 应用的全生命周期内生成的人工物,如需求规范、概念模型、设计规范、可执行模块或仿真数据等。则 V&V 的一般性定义如下^[2]:

● a 校验 (Verification) 是指:“对模型 a 的变换准确度 (Accuracy) 的评估,即由模型 a' 变换成模型 a 时,这种变换准确吗?”回答的问题是:“正确地由 a' 生成了 a 吗?” a 校验的参照物是 a 的前驱 a' 。

● a 确认 (Validation) 是指:“对模型 a 的行为准确度 (Accuracy) 的评估,即模型 a 的行为足够准确地表达了对象系统(即源系统)的行为吗?”回答的问题是:“生成了正确的 a 吗?” a 确认的参照物是对象系统的行为。

这里的变换准确度,是指在将描述模型变换为计算机程序时, a 所具有的准确度;行为准确度则是指与对象系统中的对应成分相比较, a 所具有的准确度。

国内学者大多把 Validation 译为验证,其意义是不完备的。一个模型的确认手段不只是“验证”,有时也需要通过分析来确认其行为准确度。特别是对概念模型或描述模型的确认,不是在仿真结果产生以后,而是在该模型完成以后立即确认,这时采用的就是分析确认,而非验证。此外,当仿真的对象系统是拟议中的系统时,它的概念模型、描述模型,甚至整个仿真模型的确认,都没法通过验证来完成,而只能是分析确认。因此,确认一词包含了验证与分析确认两个方面。

验收由美国国防部 (DoD) 的 M&S 小组定义为“官方鉴定一个模型、仿真或二者共同对用户的特定用途是可接受的”^[9]。另一方面,国际标准化组织 (ISO) 将验收与鉴定定义如下^[10]:

● 验收是一个“过程,通过该过程一个权威机构给出正式承认:一个团体或个人完全实现了所指定的任务。”

● 鉴定是一个“过程,通过该过程第三方给出书面担保:一个产品、过程或服务符合所规定的特征。”

上述 ISO 定义同 DoD 的 M&S 小组普遍使用的定义是不相容的。

在 ISO 制定的验收和鉴定概念的基础上,Balci 提出了美国政府所用的一个验收与鉴定的综合性框架,如图 1 所示^[6]。图中的验收权威机构一般是国家级的,如美国科学技术协会 (NIST)。愿意充当 M&S 鉴定代理的公司或组织向验收权威机构提出申请,验收权威机构通过对申请者进行审查,审查对象包括申请单位和单位中执行鉴定任务的关键人员,对合格者颁发正式的资质认证。当前许多公司充当 ISO 9000 的认证代理,可参阅 <http://www.praxiom.com/accreditors.htm>。如 ISO 定义所指,认证必须由独立的第三方执行,另外的两方分别指 M&S 应用资助者和 M&S 应用开发者。为了保证真正的鉴定或认证独立,IEEE 标准 1012^[11] 分别从技术、管理和财政上对第三方的独立性作了规定。鉴定和验收是指对 M&S 的开发过程和结果进行认证,并把表示认可

的形式物,如证书、盖有批准印章的文件等授予 M&S 应用资助者而非 M&S 应用开发者,只有当 M&S 应用资助者接受授予后,才可使用 M&S 应用的结果。由此可见,这些表示认可的形式物的错误授予的结果将十分严重。

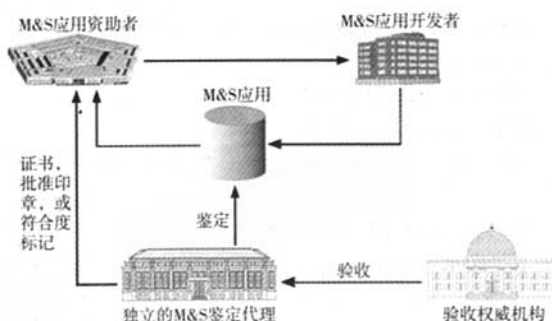


图1 验收与鉴定的综合性框架

由于鉴定和验收工作的内容主要是管理上的而非技术上的,而且根据各国的情况也有所不同,因此,本文从技术角度将此二者归并在一起,表示为 VV&A 工作中除 V&V 之外的所有工作,仍用验收 (Accreditation) 表示。这样, VV&A 仍然是一个表示对 M&S 应用可信度进行确定的完整体系。

3 通用的 VV&A 体系

对一个具体的 M&S 应用来说, VV&A 应贯穿于其全生命周期,过去那种认为对仿真结果准确度的验证要等到 M&S 研究结束时再进行的想法是完全错误的^[1]。因为 M&S 应用的规模越来越大,而测试的原理之一告诉我们对大型复杂的 M&S 应用的完全测试是不可能的。也就是说,我们无法通过验证来保证 M&S 应用结果 100% 的准确。但从另一方面看,当前 M&S 应用有着多方面深层次的意义,又必须力争保证 M&S 应用具有一定的可信度。

有三类主要指标影响 M&S 应用的可信度,即软件工程中的 3P^[12]:

● 产品 (Product): 主要是指 M&S 应用中所生成的一切人工物,包括模型、规范等;

● 过程 (Process): 生成并输出一系列人工物的过程;

● 项目 (Project): 包括人员、文档、方案制定、质量保证、性能成熟度等。

将 M&S 应用可信度形式化表示为:

$$Cer = \{P1, P2, P3\}$$

其中,

Cer 表示 M&S 应用可信度指标集;

P1 表示产品指标;

P2 表示过程指标;

P3 表示项目指标。

由于 VV&A 贯穿于 M&S 应用开发全生命周期的始终,因此需要对 M&S 应用开发周期进行阶段划分,可将 M&S 应用的全生命周期归纳为一个八元式:

$$LP = \{UoD, CM, RSp, DSp, EM, IA, Re, ReP\}$$

其中,

LP 表示 M&S 应用的全生命周期。

UoD 表示 M&S 应用的论域,也称仿真研究的问题域,即 M&S 应用研究中所涉及的所有实体的一个范畴。

CM 表示 M&S 应用对象系统的概念模型。

RSp 表示 M&S 需求规范。

DSp 表示 M&S 设计规范。

EM 表示可执行的 M&S 模块。

IA 表示将所有 EM 合在一起以体现 M&S 应用的总体属性。

Re 表示 M&S 应用获得的结果。

ReP 表示对 M&S 应用结果的陈述。

说明:这里的 EM 和 IA 体现了对大型复杂 M&S 应用的策略。对大型复杂 M&S 应用,无法一次建立起完整的仿真模型,必须分而治之,将其分解为若干模块来分别进行研究。但这就就会产生新的疑问:通过所有模块能否体现出总体的 M&S 应用的属性?从整体论的角度分析所得到的答案是否定的。Balci 制定的规则之一^[5]说明了对每个模块的成功测试并不能表明 M&S 应用的总体可信度。因此,除了单元测试,还要进行系统测试,以分析各模块在一起共同运行时所涌现出的特性。另外,由于 M&S 应用过于复杂,内容繁多,对其结果的陈述也是必须考虑的问题,以避免获得了正确的结果,却因表述上出了问题而被曲解,最终导致 M&S 应用失败的情况发生^[3]。

这样,由 M&S 应用的可信度指标集和全生命周期,可定义 M&S 应用的 VV&A 为:

$$\langle VV\&A \rangle = Cer^T \times LP$$

得到的 VV&A 体系是一个 3×8 的矩阵,其中每个元素都以 V&V 为技术基础,最终对可信度的评价结果进行验收。

4 M&S 应用可信度指标的建立与评价

从前面所给出的通用 VV&A 体系可以看出,M&S 可信度指标也是一个层次化的体系,然后使用一定的方法,从这个指标体系中得出 M&S 应用的总体可信度(可接受性)。

4.1 指标划分

采用计算机科学中“树”的概念,可对 M&S 应用的总体可信度进行层次划分。例如,与前面 M&S 应用开发全生命周期相对应,可将 M&S 应用的总体可信度在第 1 层分解为 M&S 需求可信度、概念模型、M&S 设计、可执行的 M&S 模块、集成的 M&S 应用、M&S 结果和结果陈述七个指标。但这七个指标与 M&S 应用总体可信度一样,都是复杂的定性概念,无法直接测量,必须向下逐级分解。这样就可得到一个完整的指标层次体系,此体系内有三类指标:

- 根指标,位于层次体系的顶点,无父指标,表示定性概念“M&S 可接受性”。

- 枝指标,至少有一个父指标和一个子指标。

- 叶指标,至少有一个父指标,无子指标。

此层次体系建立的标准为:

- 完备性,包含仿真问题域内所有本质要素。

- 深度,提供了度量所需的细节程度。

- 可测性,叶指标可直接度量或评估。

有时,一个子指标可能有多于一个的父指标,为保证树状层次结构不出现交叉,可使此子指标多次出现,这样就可保证整个层次指标体系为一有向无环图。

4.2 量化打分

有了指标层次体系,还需确定对其评价的方法。由于叶指标可直接度量,因此从对叶指标的打分开始。

在 VV&A 中所用的分数有三种:

- 直接分数,为一个 0 到 100 间的实数值。

- 模糊分数,为一个 0 到 100 间的实数闭区间。

- 标称分数,对前两类分数的定性表示,一般通过映射表量化为前两类分数,见表 1。

表 1 标称分数集举例

标称分数	优 +	优	优 -
数字分数	[97, 100]	[94, 97)	[90, 94)
标称分数	良 +	良	良 -
数字分数	[87, 90)	[84, 87)	[80, 84)
标称分数	中 +	中	中 -
数字分数	[77, 80)	[74, 77)	[70, 74)
标称分数	差 -	差	差 +
数字分数	[67, 70)	[64, 67)	[60, 64)
标称分数	拒绝 -	拒绝	拒绝 +
数字分数	[40, 60)	[20, 40)	[0, 20)

出于应用上的方便,一般将直接分数也化成模糊分数,如 75 表示为 [75, 75]。在项目上对多个标称分数集进行定义,并设置一个标称分数集作为项目缺省,然后为每个叶指标选择一个标称分数集。

为计算和表示上的方便,又将模糊分数表示为三个参数,即区间的最小值(区间下界)、区间的最大值(区间上界)和区间的平均值(区间中点)。

从低层指标分数获得高层指标分数通过加权求和的方法实现。设第 m 层的某指标有 n 个子指标,分数值为 $S_i, i = 1, 2, \dots, n$,则制定一组权值 $w_i, i = 1, 2, \dots, n$,则此指标的分数 $S = \sum_{i=1}^n S_i \times w_i$ 。这样逐级对指标加权求和,最终得到根指标的分数。这里,对权值的制定是由领域专家完成的,但当 $n > 5$ 时会变得困难,可采用层次分析法(AHP)等方法辅助加权。

另外,为了更直观地分析加权求和的情况,可使用 Kivi-at 图^[7]。图 2 给出了一种具有七个子指标的情况。具体做法是:用圆心表示 0 分,圆周表示 100 分;按权值不同对半径作划分;在半径上作出每个子指标分数的最小值、最大值和均值;将不同指标的同类分值联结成闭合多边形区域,则:由分数值小值点所围成的最小区域表示最坏的情况,由分数值最大值点所围成的区域表示最好的情况。由平均分点围成的

区域表示了最可能的情况。

4.3 M&S 可信度评价的计算
机软件平台——EE

EE (Evaluation Environment, 评价环境)是由美国 Orca 计算机公司开发的一种基于 Web 的客户机/服务器型软件系统^[7], 见图 3。EE 通过 128 位加密的 SSL 运行在 Internet 上, 它可使地理上分布的人员以一种协作的方式开展复杂的评价项目。有关 EE 更为详细的介绍可在 <https://www.orca-computer.com/ee> 获得。EE 可用于各种类型的评价项目, 包括 M&S 可信度评价。EE 对本文中所介绍的 M&S 可信度指标建立与评价方法大部分作了实现。同时 EE 还使用了许多开放的技术标准, 如 XML、XSL、SVG、DHTML 和 PDF。EE 具有其自己的 XML 标记语言, 称为 EEML, 用作项目数据的输入/输出、获得/存储和报告生成。EE 通过使用 XML 技术提供了 business-to-business 的数据交换功能。

5 总结与展望

M&S 以其独有的优势被用于许多学科的研究中, 其中不乏非常重要的领域, 如国防领域的 SBA(基于仿真的采办)、军事训练、政治领域的决策制定、商业领域的成本-效益评估、科技领域的航天系统等。这些研究有许多共同特点: 成本高、周期长、政治影响力大、牵涉人员多、对象复杂等, 因此需要高可信度。VV&A 是保证 M&S 应用可信度的活动, 它应与 M&S 应用的开发同步进行并贯穿于 M&S 应用的始终。由于当前的 M&S 应用的规模庞大、结构复杂, 从 M&S 角度看, 无法建立与实际系统完全一致的仿真模型, 因此, 要保证 M&S 应用 100% 的准确度是不可能的, 也是不现实的。为尽可能保证 M&S 应用的可信度, 必须严格进行 VV&A。

本文详细介绍了对系统 M&S 应用的 VV&A 过程, 并提供了对 M&S 应用可信度指标的建立与评价的思路, 其目的就在于提高人们对 M&S 应用可信度的重视, 在 M&S 应用开发实践中严格执行 VV&A。同时, 本文给出的通用框架, 为实现一个完整的 M&S 应用的 VV&A 体系提供了理论方法基础。

参考文献:

[1] Averill M Law, W David Kelton. Simulation Modeling and Simulation, 3rd [M]. McGraw-Hill, 2000.

[2] Christopher A. Chung. Simulation modeling handbook: a practical approach [M]. Library of Congress Cataloging-in-Publication Data, 2003.

[3] Defense Modeling and Simulation Office (DMSO), Verification, Validation, and Accreditation Recommended Practices Guide [M]. U.S. Department of Defense. Office of the Director of Defense Research and Engineering, November 1996.

[4] Osman Balci. Quality assessment, verification, and validation of modeling and simulation applications [J]. WSC'04, 2004.

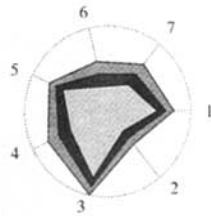


图2 Kiviat 图



图3 基于 Web 的 EE 软件系统

[5] Osman Balci. Verification, Validation, and certification of modeling and simulation applications [J]. WSC'03, 2003.

[6] Osman Balci and Said D. Saadi. Proposed standard processes for certification of modeling and simulation applications [J]. WSC'02, 2002.

[7] Osman Balci et al. A collaborative evaluation environment for credibility assessment of modeling and simulation applications [J]. WSC'02, 2002.

[8] Peter Clark, Peter Ryan and Lucien Zalcmann. Advanced Distributed Simulation for the Australian Defence Force [R]. Air Operations Division Aeronautical & Maritime Research Laboratory, DSTO-GD-0255, 2000.

[9] DoDI. DoD modeling and simulation verification, validation, and accreditation [S]. Department of Defense Instruction 5000.61, Apr., 1996.

[10] A Rae, P Robert and H L Hausen. Software Evaluation for Certification: Principles, Practice, and Legal Liability [J]. London, UK: McGraw-Hill, 1995.

[11] IEEE. IEEE Standards for Software Verification and Validation [S]. IEEE Standard 1012. Washington, DC: IEEE Computer Society, 1998.

[12] J M Voas. Guest editor's introduction: certification: reducing the hidden costs of poor quality [J]. IEEE Software, 1999, 16 (4): 22-25.

[作者简介]



王仁春(1963-),男(汉族),浙江杭州人,高工,博士生,研究方向:系统 M&S 应用的 VV&A,武器装备系统论证与型号管理;

李昊(1978-),男(汉族),辽宁盖县人,博士生,研究方向:复杂系统建模与仿真,基于 Agent 的建模

与仿真;

戴金海(1943-),男(汉族),河北昌黎人,教授,博导,中国计算机用户协会仿真应用分会副理事长,《计算机仿真》编委会副主任,研究方向:飞行器系统建模与仿真,飞行器动力学、制导与控制,武器系统及运用工程。