

# 基于 CMMI 的资源模型

王 慧, 周伯生

(北京航空航天大学计算机学院, 北京 100083)

**摘要:** 在 CMMI 实施过程中, 以有效、真实、及时、合理的度量数据为基础建立了项目的资源模型。随着组织成熟度的提高和项目数据的积累, 在已经完成的项目基础上, 按照聚类的思想选择相似的项目组, 执行资源模型合并的算法, 建立组织的资源模型。正确地建立并使用资源模型对有效策划、预测、监督和控制项目的资源有显著的作用。

**关键词:** 资源模型; 能力成熟度模型集成; GQM

## CMMI-based Resource Model

WANG Hui, ZHOU Bo-sheng

(School of Computer Science and Technology, Beihang University, Beijing 100083)

**【Abstract】** Based on the useful, true, timely and reasonable software metrics, this paper builds the project resource model during CMMI performance. When plenty of resources data are collected during the improvement of organization's level of capability maturity, based on the resource model of completed projects, projects resources data can be incorporated and organization resource model can be built by using the equivalent class partition technique based on clustering. Correct resource model can be used to plan, predict, monitor and control project resources quantitatively and efficiently.

**【Key words】** resource model; capability maturity model integration(CMMI); GQM

### 1 概述

由 SEI 提出的能力成熟度模型集成(capability maturity model integration, CMMI)已经广泛应用于现代软件组织/企业的过程改进和评估中<sup>[1]</sup>。以北京为例, 在实施过程改进的企业中, 64.94% 采用了 CMMI 作为过程改进的指导框架, 远远高于其他质量标准, 如 ISO9000(50.65%)、PMBOK(16.98%) 等<sup>[2]</sup>。2006 年 8 月, CMMI 1.2 版本发布, 其应用领域进一步扩展到了硬件和传统工业<sup>[3]</sup>。作为过程改进框架, CMMI 定义了目标和最佳实践, 但是没有说明如何进行实施。因此, 基于 CMMI 的过程实施模型的研究具有重要而现实的意义。

资源模型是过程实施模型中的一个重要组成部分, 用于有效地估计、分析和预测过程实施; 寻找过程改进机会以及策划、监督和控制项目的资源消耗。目前, 组织/企业在构建自己的资源模型时, 对其构建方法、表达方式、度量基础以及项目分类等问题缺乏科学的指导, 因此, 本文提出了资源模型构建与合并算法, 并且应用 GQM 模型、聚类成熟的工程和数学方法解决资源模型构建与合并过程中度量的建立、数据的收集以及项目的分类等问题。

资源模型是对项目人员工作量花费建立的模型, 具体可以定义为 3 种比例的集合: (1) 生命周期某阶段花费的时间与项目总周期的比例; (2) 生命周期某阶段花费的工作量与项目总工作量的比例; (3) 生命周期某阶段某种任务类型花费的工作量与该阶段总工作量的比例。

由于成熟度等级不同, CMMI 实施过程中对资源模型的要求有所区别。在能力成熟度等级为 2 级的企业中, 资源模型通过单个项目的度量数据建立, 作为该项目进行资源管理的基础。在能力成熟度等级为 3 级或者以上的企业中, 资源模型通过对同类项目资源数据的合并获得, 并且作为企业过

程资产(organizational process assets)的一部分进行管理<sup>[4]</sup>, 是今后对同类项目进行资源管理的参考依据。

资源模型的构建基础是组织/企业自身项目的实施数据, 为了保证数据的有效性、真实性、及时性与合理性, 项目需要合理的度量定义和易于实施的数据收集机制。随着企业/组织项目实施数据的增加和能力成熟度等级的提高, 有必要根据项目的不同属性对其进行科学分类, 进而针对同类项目应用构建算法建立组织级的资源模型。

软件项目的成本主要是人员成本, 一般情况下, 人员成本可以通过工作量换算获得。除了人员成本之外, 企业/组织在构建成本模型(cost model)的时候, 还要综合考虑硬件成本、设施成本等因素。

### 2 度量基础

显然, 资源模型的质量取决于度量的质量, 有效、真实、及时、合理的度量数据是建立高质量资源模型的基础。因此, 在资源模型这个已定义的上下文环境中, 如何科学地选择度量以及如何有效地实施度量数据收集是度量需要解决的 2 个基本问题。

#### 2.1 基于 GQM 的度量确定

GQM 模型是 20 世纪 80 年代中期由美国马里兰大学的教授 Victor R. Basili 提出的一种面向目标的、自上而下由目标逐步细化到度量的度量定义方法, 目前在国际上得到了广泛的应用<sup>[5]</sup>。该模型有 3 个层次: (1) 概念层, 即目标层, 它描述了通过度量期望达到的度量目标; (2) 可操作层, 即问题层,

**作者简介:** 王 慧(1978 - ), 女, 博士研究生, 主研方向: 软件工程; 周伯生, 教授、博士生导师

**收稿日期:** 2007-03-09 **E-mail:** wanghui@buaa.edu.cn

该层针对上一层的目标定义了一组问题，这些问题从各个角度对度量的目标进行描述；(3)量化层，即度量层，对于每个问题，都定义了一组数据与之相关联，通过这些数据可以对问题有个量化的回答。GQM模型提供的方法是根据问题确定度量和数据元素，然后将度量以可视化的形式表示出来。

应用 GQM 方法，可以建立关于资源模型的度量如下：

(1)目标：建立资源模型。

(2)问题：1)影响项目分类的项目属性如何确定？2)项目工作量的属性如何确定？

(3)度量：见表 1。

表 1 度量目标“建立资源模型”的问题和度量

问题	度量
如何确定影响项目分类的项目属性	系统类型
	系统复杂度
	项目生命周期类型
	项目工作语言
	产品规模
	项目周期
	项目团队大小
如何确定项目工作量的属性	人员技能程度
	项目当前阶段
	工作量花费
	任务类型

其中，属于问题 1)的度量是建立企业组织级资源模型时进行项目科学分类的度量基础；属于问题 2)的度量是构建项目级和组织级资源模型的度量基础。

## 2.2 数据收集方法

在建立资源模型所必需的度量确定之后，需要对项目的这些度量数据进行收集。在项目实际的实施过程中，每天填写日志是一种可操作性很强的数据收集方法，可以保证度量数据的及时、准确。

在日志的设计和填写中，需要注意的问题包括：

(1)组织/企业统一日志格式，日志内容要求覆盖资源模型的度量。

(2)日志中的系统类型、产品规模、生命周期类型等度量的具体含义和数据表示单位需要在组织级进行统一定义。

(3)日志中的任务类型、工作量花费等度量的具体含义和数据表示单位需要在组织级进行统一定义。

(4)日志填写应用自动化工具，它们有助于数据的收集，能使分析更一致、更及时、更准确。

(5)日志填写、总结和分析的频率建议分别为每天、每周和每月。

表 2 是一个关于日志的例子。其中，工作包与项目的 WBS(work breakdown structure)分解一致；任务类型由企业统一定义为立项论证、方案论证、需求分析、硬件设计、软件编码、集成、联试、测试、维护、考核、纠错、QA、CM、PP、PMC、MA、RM、技术试验、会议、其他；工作量度量单位为 h。

表 2 某公司的日志模版

工作日志表									
项目名称	项目规模			项目阶段					
姓名	本周起始时间	x年x月x日-x年x月x日			产品类型				
任务类型	工作包	工作时间/h							合计
		星期 1	星期 2	星期 3	星期 4	星期 5	星期 6	星期 7	
工作时间小计									

在应用度量数据构建资源模型之前，还要对所收集的度量数据的异常性进行判断和处理。关于异常数据的识别和处

理可以参考统计过程控制的方法，通过设置阈值来识别和排除过大或者过小的不合理度量值。

## 3 资源模型的构建与表达

结合度量基础和资源模型的定义，资源模型的构建可以形式化表述如下。

记：

$n$  为项目生命周期阶段总数；

$m$  为项目中任务类型总数；

$R$  为资源模型；

$T_{iBegin}$  为生命周期阶段  $i$  的开始时间；

$T_{iEnd}$  为生命周期阶段  $i$  的结束时间；

$E_{ij}$  为生命周期阶段  $i$  中在任务类型  $j$  上花费的工作量，

可以通过将日志中该任务类型在该阶段每天花费的工作量的度量数据相加获得。

则可构造： $R = \{R_{Time}, R_{Effort}, R_{Type}\}$ ，其为项目资源模型  $R$  的 3 个组成部分。其中， $R_{Time} = \{r_{Time1}, r_{Time2}, \dots, r_{Timei}, \dots, r_{Timen}\}$ ； $r_{Timei}$  为生命周期阶段  $i$  花费的时间与项目总周期的比例；

$$r_{Timei} = (T_{iEnd} - T_{iBegin}) / (T_{nEnd} - T_{1Begin}) \times 100\%$$

$$R_{Effort} = \{r_{Effort1}, r_{Effort2}, \dots, r_{Efforti}, \dots, r_{Effortn}\}$$

$r_{Efforti}$  为生命周期阶段  $i$  中花费的工作量与项目总工作量的比例，

$$r_{Efforti} = \frac{\sum_{j=1}^m E_{ij}}{\sum_{i=1}^n \sum_{j=1}^m E_{ij}} \times 100\%$$

$$R_{Type} = \{r_{Type11}, r_{Type12}, \dots, r_{Typeij}, \dots, r_{Typenm}\}$$

$r_{Typeij}$  为生命周期阶段  $i$  中在任务类型  $j$  上花费的工作量与该阶段  $i$  总工作量的比例，

$$r_{Typeij} = E_{ij} / \sum_{j=1}^m E_{ij} \times 100\%$$

应用以上构建算法可以获得项目级的资源模型。为了更加直观地观察分布与趋势，资源模型还可以通过图形的方式进行表达。图 1 给出了一个项目资源模型图形表达的实例。

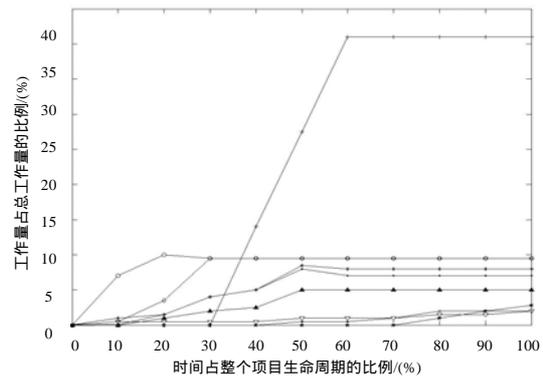


图 1 某公司项目资源模型实例

## 4 资源模型的合并

通过合并项目资源模型可以获得组织/企业级的资源模型。在进行资源模型合并的过程中需要解决以下 2 个问题：

(1)如何根据项目属性对项目进行科学分类；

(2)资源模型合并算法。

### 4.1 基于聚类方法的项目分类

聚类技术是研究“物以类聚”的一种分类方法，主要解决仅凭经验无法确切分类的问题，是数值分类学的一个分支<sup>[6]</sup>。采取这种方法根据项目属性度量对项目进行科学分类

的思路是：当遇到仅凭经验不能根据项目的某个属性A进行分类的时候，将该属性的度量值构造成属性度量值全集  $A_M$ 。在此基础上，利用类聚技术将属性度量值划分成若干等价类，从而获得项目在该属性上的分类。

聚类方法的基础是对“距离”的定义，这里距离表示 2 个度量值之间的近似程度。常用的距离有 2 类：(1)几何距离，例如：绝对值距离，欧式距离等；(2)相似性距离，例如：相似系数。因为项目某属性度量值为一个具体的数值，所以选择绝对值距离比较适宜。绝对值距离为  $d_{ij} = |x_i - x_j|$ 。

以项目规模为例，设  $i$  个项目规模的度量集为  $S_M = \{m_1, m_2, \dots, m_i\}$ ，并将每个度量值视为一类，即有  $M_i = \{m_i\}$ ，再依据规模度量数据求得各“类”之间的距离矩阵为

$$D(0) = \begin{bmatrix} 0 & & & & & \\ d_{21} & 0 & & & & \\ \vdots & \vdots & \ddots & & & \\ d_{i1} & d_{i2} & \dots & d_{ii} & & \end{bmatrix}$$

不失一般性，设  $d_{21}, d_{i,i-1}$  值为最大，从而可以将类  $M_1$  与  $M_2$ ， $M_{i-1}$  与  $M_i$  分别合并成为 2 个新类  $M_{i+1}$  与  $M_{i+2}$ ，重新计算距离矩阵  $D(1)$ 。重复这个过程，直至完成对所有类的处理。最终结果可以用类聚图表示，见图 2。

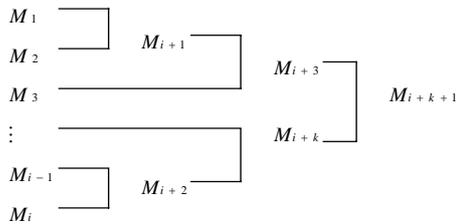


图 2 软件产品规模类聚

形成对集合  $S_M$  的划分，即根据产品规模属性对项目的划分，如图 3 所示。

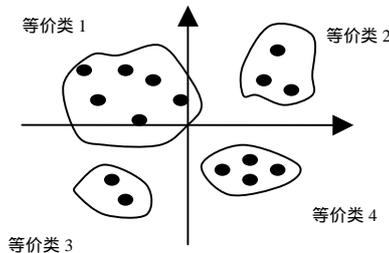


图 3 项目根据产品规模属性的划分

项目按照属性的分类是构造组织级资源模型，即进行资源模型合并的基础。

#### 4.2 资源合并算法

对于 4.1 节中分类完成的同一类项目，应用资源合并算法构建组织级的资源模型。以第 3 节中的形式化表述为基础，算法的形式化表述如下：

又记：

$s$  为同一类型项目中项目的总数；

$(T_{iEnd} - T_{iBegin})_k$  为项目  $k$  中生命周期阶段  $i$  花费的时间；

$E_{ijk}$  为项目  $k$  中生命周期阶段  $i$  在任务类型  $j$  上花费的工作量，可以通过将该项目日志中该任务类型在该阶段每天花费的工作量的度量数据相加获得；

$R_k$  为同一类型的  $s$  个项目平均的生命周期长度与项目  $k$  的生命周期长度的比例系数，

$$R_k = \frac{\sum_{k=1}^s (T_{nEnd} - T_{1Begin})_k}{s \times (T_{nEnd} - T_{1Begin})_k}$$

则可构造：

$$r_{Timei} = \frac{\sum_{k=1}^s ((T_{iEnd} - T_{iBegin})_k \times R_k)}{\sum_{k=1}^s (T_{nEnd} - T_{1Begin})_k} \times 100\%$$

$$r_{Effort} = \frac{\sum_{k=1}^s \left( \left( \sum_{j=1}^m E_{ijk} \right) \times R_k \right)}{\sum_{k=1}^s \left( \left( \sum_{j=1}^m E_{ijk} \right) \times R_k \right)} \times 100\%$$

$$r_{Typeij} = \frac{\sum_{k=1}^s (E_{ijk} \times R_k)}{\sum_{k=1}^s \left( \left( \sum_{j=1}^m E_{ijk} \right) \times R_k \right)} \times 100\%$$

## 5 结束语

应用 GQM 模型、类聚等成熟的工程和数学方法以及资源模型构建、合并算法，组织/企业可以根据自身的能力成熟度情况建立并应用资源模型对其资源，尤其是人力资源花费进行有效的管理。该模型与其他 4 种过程实施模型(生命周期模型、度量模型、质量模型、控制模型)结合，已经应用于 27 家采用 CMMI 过程改进框架进行过程改进和评估的企业中，目前获得的反馈表明它们可以有效帮助企业软件能力成熟度的提高。以过程实施模型为基础的“项目管理系统”工具提供了对资源模型建立的自动支持，包括日志中资源模型度量数据的自动收集、资源模型曲线的建立和资源模型的合并。

## 参考文献

- 1 Liu Xiaqing, Sun Yan, Kane G, et al. QFD Application in Software Process Management and Improvement Based on CMM[J]. ACM Sigsoft Software Engineering Notes, 2005, 30(4).
- 2 2006 年北京软件产业发展蓝皮书(投资分册)[Z]. 北京软件与信息服务业促进中心. 2006-12.
- 3 CMMI Product Team. CMMI® for Development[Z]. (2006-08). <http://www.sei.cmu.edu/publications/documents/06.reports/06tr008.html>.
- 4 Chrissis M B, Konrad M, Shrum S. CMMI Guidelines for Process Integration and Product Improvement[M]. [S.L.]: Addison-Wesley Longman Publishing Co., 2003.
- 5 Basili V R, Gianluigi C, Dieter R H. The Goal Question Metric Paradigm[M]. [S.L.]: John Wiley & Sons, 1994.
- 6 常兆光, 王清河, 宋岱才, 等. 随机数据处理方法[M]. 东营: 山东石油大学出版社, 1997-08.

(上接第 18 页)

## 参考文献

- 1 Gupta P, Kumar P R. The Capacity of Wireless Networks[J]. IEEE Transactions on Information Theory, 2000, 46(2): 388-404.
- 2 Reznik A, Verdu S. On the Transport Capacity of a Broadcast Gaussian Channel[J]. Communications in Information and Systems, 2002, 2(2): 183-201.
- 3 Bansal N, Liu Z. Capacity, Delay and Mobility in Wireless Ad-Hoc Networks[C]//Proc. of IEEE INFORM'03. 2003.
- 4 Hou T, Li V. Transmission Range Control in Multihop Packet Radio Networks[J]. IEEE Transactions on Communications, 1986, 34(1): 38-44.

