

基于范例库推理的软件成本估算模型研究

方海光^{1,2,3}, 陈 澎^{2,3}, 余 莉^{2,3}

(1. 四川师范大学计算机软件实验室, 成都 610068; 2. 中国科学院成都计算机应用研究所, 成都 610041;

3. 中国科学院研究生院, 北京 100080)

摘 要: 用传统的经验函数估算软件成本有很多局限性, 采用基于范例库推理的估算方法可以很好地弥补其中的问题。讨论了软件成本估算和基于 CBR 推理研究的特点, 从总体上阐述了 COSCBR 系统结构, 描述了系统重要的研究方面: 影响软件成本因素; 层次推理; COSCBR 系统的范例表示方法; 相似度的基本计算算法。

关键词: 范例库推理; 软件成本; 相似度; 层次推理

Research on Software Cost Evaluation Model Based on Case-based Reasoning

FANG Haiguang^{1,2,3}, CHEN Peng^{2,3}, SHE Li^{2,3}

(1. Laboratory of Computer Software, Sichuan Normal University, Chengdu 610068; 2. Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu 610041; 3. Graduate School of Chinese Academy of Sciences, Beijing 100080)

【Abstract】 Software cost estimate with the traditional experience function has the shortcoming, and applying CBR estimate method can make up these problems. This paper describes the research on software cost estimate and the CBR characteristic, introduces the COSCBR system architecture. The essential aspects are presented: the structure of the software cost factors; the hierarchy reasoning process; the COSCBR case expression; and the calculating similarity arithmetic.

【Key words】 Case-based reasoning(CBR); Software cost; Similarity; Hierarchy reasoning

软件开发成本估算主要是对软件项目规模和工作量进行估算^[1]。这类估算常常具有经验性和时间相关性。软件项目进行成本估算时一般步骤为^[3,4]: (1)软件大小估算, 估算问题大小(如功能点FP)和估算解决方案的大小(如源代码行数SLOC); (2)工作量(如人月PM)和工作进度估算; (3)估算反馈。如果软件成本估算超出实际发生, 则造成组织资源浪费, 反之则使其失控而造成预算超标和交付拖延^[2]。显然, 作出更接近实际发生的软件成本估算是必要的。传统的软件成本估算的方法主要通过经验公式进行估算^[5], 通过软件开发人员和项目规模驱动因素之间的函数关系进行的^[3,4]。这种方法在实际应用中通常受到以下因素的制约: (1)较难确定成本驱动因素, 以及成本驱动因素间相对的程度; (2)因素之间的相关性在函数公式估算时很难完全考虑; (3)不正确数据很难识别; (4)由于成本驱动因素的复杂性, 需要协调估算模型的复杂性和估算效果间的关系。

范例推理(Case Based Reasoning, CBR)是通过访问范例库中过去同类问题(源范例)的求解从而获得当前问题(目标范例)解决方案的重要推理方法。CBR方法能有效地解决知识表达困难或者无法表达的领域问题, 准确地分析所要解决的具体问题, 非常适用于需要重新和大部分借鉴以往经验的情况。

构建基于 CBR 的软件成本估算模型的优势在于: (1)软件产品成本驱动因素的相互关系被作为实例存储, 而非作为数学公式和相应的约束表达存储; (2)判断和决策建立在解决相关问题的经验的基础上, 而非形式化规则或公式; (3)不需要在现有的范例和因素间建立确定的属性匹配关系; (4)允许非数值化的软件产品特征作为成本驱动因素。通过对软件成本的推理过程

的分析, 可以得知基于 CBR 是模仿软件成本估算过程的合适方法。以此为基础, 将 CBR 应用于软件成本估算领域, 建立基于 CBR 的软件成本评估系统模型(Software Cost Evaluation System Based on CBR, COSCBR)。

1 COSCBR 系统结构

COSCBR 是基于 CBR 的软件成本估算系统, 系统主要由 HCI、范例获取、推理匹配、时序更新和范例库等模块组成, 如图 1 所示。

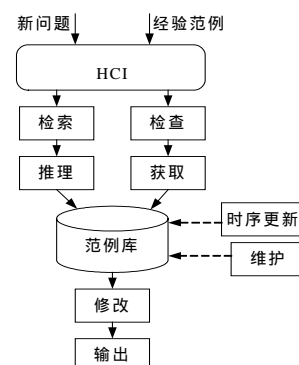


图1 COSCBR 的系统结构

HCI 模块是包括新问题推理和经验范例学习部分, 经验范例学习是为建立范例库、进行语法和语义检查提供接口; 新问

基金项目: 国家“973”计划基金资助项目(CB318003, CB030602)

作者简介: 方海光(1975-), 男, 博士生, 主研方向: 软件工程经济学与人工智能; 陈 澎、余 莉, 博士生

收稿日期: 2005-10-30 **E-mail:** fanghaiguang@hotmail.com

题推理提供推理解释输入输出的接口。范例获取模块是提供完善的范例输入接口,并将事实描述和经验知识整合;知识检查模块对各类知识进行语法语义检查,以保证知识库的一致性。推理匹配是根据用户提交的待求解的新问题事实信息,对范例进行检索,找到最相近的范例并将它提交给修改模块进行修正,从而得到推理估算信息。时序更新模块是根据范例的时间信息对范例的时序影响因子进行更新。范例库是推理的数据基础,用于存放范例的特征属性。COSCBR 中设计的 CBR 基本推理算法描述如下:

- S1: 根据用户提出的软件特征属性(表 1 所示属性)构造新范例;
- S2: 将此范例与范例库中的范例进行相似度计算,求出相似度;
- S3: 范例库检索是否结束? 完毕,则返回相似度最大的范例,转向 S4; 否则转向 S2;
- S4: 将返回的范例作为目标问题解决方案,并对目标方案进行适当的调整。

2 COSCBR 关键方面研究

2.1 影响软件成本的因素构造

影响软件成本的因素有很多种,根据软件成本估算主要模型的分析,归纳出当前主要影响软件成本的属性^[3-5],可以把各属性划分为相应的大类因子。在COSCBR采用了软件成本基本属性中的 26 项和 5 大类因子,其与COCOMO81 和COCOMO 分析比较如表 1 所示,其中Y表示模型中采用的属性。

表 1 影响软件成本的因素比较

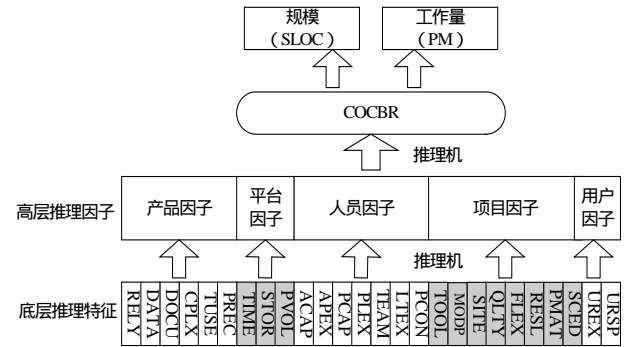
| 因子 | 属性 | COCOMO81(详细模型): 15 项 | COCOMO(后体系结构): 24 项 | COSCBR(详细结构): 26 项 |
|----|-----------------|----------------------|---------------------|--------------------|
| 产品 | 要求的软件可靠性(RELY) | Y | Y(乘数) | Y |
| | 数据库规模(DATA) | Y | Y(乘数) | Y |
| | 匹配生命周期文档(DOCU) | | Y(乘数) | Y |
| | 产品复杂性(CPLX) | Y | Y(乘数) | Y |
| | 要求的可复用性(TUSE) | | Y(乘数) | Y |
| 平台 | 先例性(PREC) | | Y(比例) | Y(先例/再工程) |
| | 执行时间约束(TIME) | Y | Y(乘数) | Y |
| | 主存储约束(STOR) | Y | Y(乘数) | Y |
| | 平台易变性(PVOL) | Y(VIRT) | Y(乘数) | Y |
| | 计算机周转时间(TURN) | Y | | |
| 人员 | 分析员能力(ACAP) | Y | Y(乘数) | Y |
| | 应用经验(APEX) | Y | Y(乘数) | Y |
| | 程序员能力(PCAP) | Y | Y(乘数) | Y |
| | 平台经验(PLEX) | Y(VEPX) | Y(乘数) | Y |
| | 团队凝聚力(TEAM) | | Y(比例) | Y |
| 项目 | 编程语言经验(LTEX) | Y | Y(乘数) | Y |
| | 人员连续性(PCON) | | Y(乘数) | Y |
| | 软件工具的使用(TOOL) | Y | Y(乘数) | Y |
| | 现代编程规范应用(MODP) | Y | | Y |
| | 多点开发(SITE) | | Y(乘数) | Y |
| 用户 | 质量要求(QLTY) | | | Y |
| | 开发灵活性(FLEX) | | Y(比例) | Y |
| | 体系结构/风险化解(RESL) | | Y(比例) | Y(风险化解) |
| | 过程成熟度(PMAT) | | Y(比例) | Y |
| | 要求的开发进度(SCED) | Y | Y(乘数) | Y |
| 用户 | 用户项目经验(UREX) | | | Y |
| | 用户支持性(URSP) | | | Y |
| | 用户 1(USR1) | | Y(乘数) | |
| | 用户 2(USR2) | | Y(乘数) | |

2.2 层次推理过程

根据范例信息量的详细程度,可以构建不同级别的推理模式:基于软件成本属性的详细推理和在此之上建立的基于软件成本因子推理的概要推理。详细推理模式采用 26 项成本属性信息构建;概要推理采用 5 大类因子的评定信息构建。总体上,COSCBR 形成了层次推理过程。其基本想法是:几个主要的比

较高层的因子影响着软件成本估算的结果,其中包括产品因子,人员因子等。而这几个因子又是由一些比较低层的属性决定的,如数据库规模、分析员能力等,由此形成了递阶结构的二层推理体系,便于对软件成本各因子分别进行推理。属性比因子相对来说要易于精确度量,层次推理描述了因子和它们所依赖的属性之间的一致性。通常以度量因子所依赖的属性的方式来度量因子。COSCBR 推理模块由 2 个推理层组成(如图 2 所示):底层推理层进行量化、归并和划分组合,获得软件成本的因子数据;高层推理进行检索、排序和因子调整来得到最优范例信息。

图 2 模型层次推理过程



2.3 COSCBR 范例的表示

范例的表示方法不仅影响能否转化为范例的形式,而且还影响范例的推理和范例库的维护。鉴于软件成本类知识具有不确定性的特点,将范例 C 用多元式 $C = \{I, S, F, E\}$ 来表示,其中 I 表示问题的说明信息。S 表示底层推理特征,即估算范例的各种属性,格式为: {属性名 1: 模糊描述, 权值}。根据软件成本估算的范例特点,确定属性权重评价尺度为 5 级,如:非常准确(0.9),很准确(0.7),准确(0.5),一般(0.3),不准确(0.1)。系统将以范例数据为基础,根据模糊向量转化矩阵,用相应的程度系数来量化该信息,作为进一步推理的量化基础。F 表示高层推理因子,即由属性 S 通过底层推理计算出的因子值(针对不同范例数据的完整程度,可以部分从范例数据直接得到),这些因子作为高层概要推理的基础。E 表示软件成本的估算结果。

表 2 COSCBR 范例表示的例子

| | 描述 | 格式 | 例子 |
|---|------|---|---|
| I | 说明信息 | {Case 序号} {软件信息描述} | {Case 286} {项目开始时间: 2003 年 6 月 18 日; 项目名称: XX 公司 HRMIS 系统} |
| S | 特征 | {属性名 1: 模糊描述, 权值} {属性名 2: 模糊描述, 权值} ... {属性名 26: 模糊描述, 权值} | {RELY: 非常准确, 0.066} {TEAM: 成熟, 0.042} ... {PMAT: 很成熟, 0.051} |
| F | 项目情况 | {因子名 1: 因子值; } ... {因子名 5: 因子值; } {时序因子: 因子值; } | {产品因子: 0.208} {平台因子: 0.141} ... {时序因子: 0.951} |
| E | 成本估算 | {规模: 值} {工作量: 值} | {规模: 28} {工作量: 4} |

由于软件成本估算具有时间相关性,因此为了让既相似也是最有用的范例被检索到,需对传统的方法进行改进,以

(下转第 220 页)