申请人：华东师范大学

发明人：文涛、陈小红

第一申请人身份证号：510322199401051112

联系人：陈小红 手机：13916934760 [xhchen@sei.ecnu.edu.cn](mailto:xhchen@sei.ecnu.edu.cn)

地址：200062 上海市普陀区中山北路3663号

**发明名称：**一种基于差异比较对分析类图进行评价的方法

摘要：本发明公开了一种基于差异比较对分析类图进行评价的方法，用于比较不同开发团队针对同一领域的分析类图和标准分析类图的差异，完成对不同分析类图的评价。本发明在对分析类图进行评价时涉及到的操作包括：（1）根据分析类图的元模型，定义评价指标、外部输入和评分公式；（2）以标准分析类图和待评价分析类图对应的XML文档作为输入，根据评价指标从XML文档中抽取用于比较的数据；（3）根据同义词对标准分析类图和待评价分析类图的数据进行差异比较；（4）将差异比较的结果带入评分公式，计算得分，分值越高则待评价分析类图设计得越好。本发明通过对分析类图的比较，在分析阶段就可以发现所设计软件在内部模块的问题，度量了软件的质量。

**权利要求**

1、一种基于差异比较对分析类图进行评价的方法，其特征在于，该方法包括以下具体步骤：

步骤1：定义分析类图的评价指标、外部输入以及评分公式；

步骤2：以标准分析类图和待评价分析类图的XML文档作为输入，根据评价指标从XML文档中抽取用于比较的数据；

步骤3：对抽取的数据进行差异比较；

步骤4：将差异比较的结果带入评分公式，计算得分，完成对待评价分析类图的评价。

2、根据权利要求1所述的方法，其特征在于，步骤1中，所述分析类图的评价指标、外部输入以及评分公式具体为：

评价指标：为四个指标，即：完整性、核心类、关系复杂度、关系正确性；其中，所述关系正确性包括：普通关系正确性、多重性关系正确性指标；

外部输入：给出核心类、类相似度阈值、权重；所述权重为0到1之间的数；其中，权重具体包括：

（i）完整性的权重、核心类的权重、关系复杂度的权重、关系正确性的权重，其中；

（ii）普通关系正确性的权重、多重性关系正确性的权重，其中；

评分公式为：



其中分别代表完整性、核心类、关系复杂度、关系正确性的得分。

3、根据权利要求1所述的方法，其特征在于，步骤2中，所述抽取用于比较的数据为：类、属性、类与类之间的关系。

4、根据权利要求1所述的方法，其特征在于，步骤3中，所述对抽取的数据进行差异比较，当参与比较的两个数据是同义词或者包含相同的关键字时，就认为这两个数据是相似的；如果两个数据是相似的，就认为它们是相同的；具体过程如下：

（ⅰ）比较标准分析类图的类与待评价分析类图的类的差异，当待比较的两个类是同义词或者包含相同的关键字时，就认为它们是相似的；当待比较的两个类不是同义词并且没有包含相同的关键字时，如果这两个类中包含的属性和方法很多都是相同的，并且这两个类在各自的类图中和其它已知的相似类之间的关系结构也很多都是相同的，通过量化这种相似性，只要这个相似性的值不小于外部输入中给定的类相似度阈值，就认为这两个类也是相似的；当待评价分析类图中的某个类与标准分析类图中的某个类是相似的，那就认为这两个类是相同的，同时也认为待评价分析类图中的这个类是正确的；

（ⅱ）比较标准分析类图的属性与待评价分析类图的属性的差异，属性的比较是以类比较为基础的，只有在两个类是相似的时，才能比较其属性是否相似。当两个属性是同义词或者包含相同的关键字时，就认为它们是相似的；当待评价分析类图中的某个属性与标准分析类图中的某个属性是相似的，那就认为这两个属性是相同的，同时也就认为待评价分析类图中的这个属性是正确的；

（iii）比较标准分析类图的类与类之间的关系与带评价分析类图的类与类之间的关系的差异，类与类之间的关系的差异比较是以类比较为基础的，只有与关系相关的两个类是相似的时，才能比较它们之间的关系是否相似；当两个关系是相同类型的时，就认为它们是相似的。当待评价分析类图中的某个关系和标准分析类图中的某个关系是相似的，那就认为这两个关系是相同的，同时也就认为待评价分析类图中的这个关系是正确的。

5、根据权利要求1所述的方法，其特征在于，所述步骤4具体包括以下步骤：

（i）根据差异比较结果，计算完整性得分；得分公式为：



其中分别为相似属性个数、标准分析类图中所有属性个数；

（ii）根据差异比较结果，计算核心类得分；得分公式为：



其中分别为待评价分析类图中核心类的个数、外部输入中核心类的个数；

（iii）根据差异比较结果，计算关系复杂度得分；得分公式为：



假设待评价分析类图中有个类能在标准分析类图中找到相似类，代表任意两个相似类进行组合得到组合个数；将标准分析类图和待评价分析类图分别看成两个无向图，其中类为图中的节点，类与类之间的关系为图中的边；表示待评价分析类图中的第两个类在中分别作为起点和终点的路径长度，与它们的相似类在分别作为起点和终点的路径长度的比值；

（iv）根据差异比较结果，计算关系正确性得分；得分公式为：



其中表示标准分析类图中所有关系的个数，表示相似关系的个数，表示待评价分析类图中在普通关系层面上正确但在多重性关系层面错误的关系的个数；

（v）将（i）~（iv）的结果带入评分公式，计算待评价分析类图的总得分；得分公式为：



其中分别代表完整性、核心类、关系复杂度、关系正确性的得分。

**说明书**

一种基于差异比较对分析类图进行评价的方法

**技术领域**

本发明属于软件质量度量技术领域，具体涉及一种使用差异比较对软件开发的分析阶段的中间产物-分析类图进行评价打分，通过分数的高低度量软件设计的好坏的方法。

**背景技术**

随着软件开发方法的成熟和多样化，软件在各行各业都得到了广泛的应用。行业软件越来越多，质量参差不齐，因此在成本利益最大化的条件下，如何度量一个软件的质量，为使用方选择最适合的软件，为开发方指导进一步的开发，便显得尤为重要。

目前对软件度量的方法很多，比如有从面向对象的角度出发，定义大量度量指标从各方面对软件的质量进行度量；有从软件源代码角度出发，分析代码中的类之间的调用关系、依赖关系、耦合度等完成对软件质量的度量；有从测试角度出发，通过各种测试方法对软件的质量进行度量。但这些度量方法基本都使用在软件开发阶段或者软件完成阶段，在软件的生命周期中，其质量缺陷发现得越早，所付出的代价也就越小，因此有必要在更早的阶段就对其质量进行度量。

分析阶段是软件生命周期的早期阶段，分析类图是该阶段的一个中间产物。它可以反映出软件内部的模块设计、模块之间的调用关系等情况。对于一个具体领域的软件而言，往往存在一个标准的分析类图，当开发人员需要设计开发某款软件时，可以将自己设计的分析类图与标准的分析类图进行比较，从而发现自己设计的软件的模块划分是否合理，模块之间的调用关系是否合理等。软件内部的模块设计是软件质量度量的重要内容，在相同条件下，软件模块的设计越好，其质量也就越好，因此可以使用分析类图来度量软件的质量。

**发明内容**

本发明的目的是提供一种基于差异比较对分析类图进行评价的方法，对于一个待评价的分析类图，该方法首先定义它的评价指标、外部输入和评分公式；然后以标准分析类图和待评价分析了图的XML文档作为输入，根据评价指标分别抽取用于比较的数据；接着根据同义词对抽取的数据进行比较，找出它们的差异；最后将差异比较结果带入评分公式，计算得分；通过得分的高低反映软件内部模块设计的合理性和质量的好坏。

实现本发明目的的具体技术方案是：

一种基于差异比较对分析类图进行评价的方法，该方法包括以下具体步骤：

步骤1：定义分析类图的评价指标、外部输入以及评分公式；具体为：

评价指标：为四个指标，即：完整性、核心类、关系复杂度、关系正确性；其中，所述关系正确性包括：普通关系正确性、多重性关系正确性指标。每个指标的具体含义如下：

（i）完整性：指的是属性完整性。

（ii）核心类：为完成某一业务必须存在的类。

（iii）关系复杂度：用于衡量类图中所有类与类之间关系的复杂性。

（iv）关系正确性：分为普通关系正确性和多重性关系正确性，用于衡量待评价分析类图中的有用类之间的关系的正确性。当待评价类图中的某个类能在标准分析类图中找到相似的类时，就称这个类为有用类。

外部输入：给出核心类、类相似度阈值、权重；所述权重为0到1之间的数；其中，权重具体包括：

（i）完整性的权重、核心类的权重、关系复杂度的权重、关系正确性的权重，其中；

（ii）普通关系正确性的权重、多重性关系正确性的权重，其中；

评分公式为：



其中分别代表完整性、核心类、关系复杂度、关系正确性的得分。

步骤2：以标准分析类图和待评价分析类图的XML文档作为输入，根据评价指标从XML文档中抽取用于比较的数据；其中，所述抽取用于比较的数据为：类、属性、类与类之间的关系。

步骤3：对抽取的数据进行差异比较，当参与比较的两个数据是同义词或者包含相同的关键字时，就认为这两个数据是相似的；如果两个数据是相似的，就认为它们是相同的；具体过程如下：

（ⅰ）比较标准分析类图的类与待评价分析类图的类的差异，当待比较的两个类是同义词或者包含相同的关键字时，就认为它们是相似的；当待比较的两个类不是同义词并且没有包含相同的关键字时，如果这两个类中包含的属性和方法很多都是相同的，并且这两个类在各自的类图中和其它已知的相似类之间的关系结构也很多都是相同的，通过量化这种相似性，只要这个相似性的值不小于外部输入中给定的类相似度阈值，就认为这两个类也是相似的。当待评价分析类图中的某个类与标准分析类图中的某个类是相似的，那就认为这两个类是相同的，同时也认为待评价分析类图中的这个类是正确的。其中相似性的量化过程具体如下：

定义上述相似性的量化计算公式为：。

上述需要比较相似性的两个类，属于标准分析类图的记为，属于待评价分析类图的记为。中的所有属性和方法构成集合，中的所有属性和方法构成集合，集合为和的交集，中的元素为和中相同名字的属性或者方法。在标准分析类图中和其它所有已知相似类的关系构成集合，在待评价分析类图中和其它所有已知相似类的关系构成集合。对于，，若和中除了和的名字不同以外，其它内容全部相同，则将放入集合中。上述公式中、、、。

（ⅱ）比较标准分析类图的属性与待评价分析类图的属性的差异，属性的比较是以类比较为基础的，只有在两个类是相似的时，才能比较其属性是否相似。当两个属性是同义词或者包含相同的关键字时，就认为它们是相似的。当待评价分析类图中的某个属性与标准分析类图中的某个属性是相似的，那就认为这两个属性是相同的，同时也就认为待评价分析类图中的这个属性是正确的。

（iii）比较标准分析类图的类与类之间的关系与带评价分析类图的类与类之间的关系的差异，类与类之间的关系的差异比较是以类比较为基础的，只有与关系相关的两个类是相似的时，才能比较它们之间的关系是否相似。当两个关系是相同类型的时，就认为它们是相似的。当待评价分析类图中的某个关系和标准分析类图中的某个关系是相似的，那就认为这两个关系是相同的，同时也就认为待评价分析类图中的这个关系是正确的。

步骤4：将差异比较的结果带入评分公式，计算得分，完成对待评价分析类图的评价；具体包括以下步骤：

（i）根据差异比较结果，统计待评价分析类图中正确的属性的个数和标准分析类图中所有属性的个数，计算完整性得分。具体过程为：

定义完整性的得分为：，记标准分析类图中所有属性构成的集合为，待评价分析类图中所有属性构成的结合为，两个分析类图中相似的属性构成的集合为，因此有。因此上式中、，分别代表标准分析类图中所有属性的个数、两个分析类图中相似的属性的个数。

（ii）根据差异比较结果，统计待评价分析类图中正确的类的个数和标准分析类图中所有类的个数，计算核心类得分。具体过程为：

定义核心类的得分为：。记外部输入中的核心类构成的集合为，一般情况下标准分析类图应该包含所有核心类，待评价分析类图中包含的核心类构成的集合为，因此上式中、分别代表外部输入中提供的核心类的个数、待评价分析类图中存在的核心类的个数。

（iii）根据差异比较结果，计算关系复杂度得分。具体过程为：

A. 将类图映射为一个无向连通图，其中类图中的类映射为中的点，类与类之间的关系映射为中的边。记标准分析类图对应的无向连通图为，待评价分析类图对应的无向连通图为。

B. 将为完成某一项业务所需要的类、属性、方法分别定义为有用类、有用属性、有用方法，一般情况下这些类即为标准分析类图中的所有类、所有属性、所有方法，在待评价分析类图中的类、属性、方法，如果能在标准分析类图中找到相似类的类、属性、方法，那么这些类、属性、方法就为待评价分析类图中的有用类、有用属性、有用方法。

C. 记待评价分析类图中所有有用类组成的集合为，这些有用类在标准分析类图中的相似类组成的集合为（和是相似类）。对于，在中找到这两个类之间存在的连通路径，记路径的数量为，路径构成的集合为，其中，，；对于，在中找到这两个类之间存在的连通路径，记路径的数量为，路径构成的集合为，其中，，。

当时，表明在标准分析类图中的和之间能进行通信的路径更多，待评价分析类图中的和之间路径的设计不合理，缺少部分路径。当时，则表明在待评价分析类图中的和之间设计了过多的通信路径，和之间存在的关系比和之间存在的关系更复杂。

D. 记和之间的复杂度得分为，令



其中表示的复杂度得分。对于，有：



其中表示路径中包含的边数。当时，表明待评价分析类图中的两个类之间进行通信时经过的边数比标准分析类图中与它们相似的两个类之间进行通信时经过的边数还少，这是不合理的。

记待评价分析类图的复杂度得分为，考虑到共存在中组合，因此令：



（iv）根据差异比较结果，计算关系正确性得分。具体过程为：

关系正确性是相对于标准分析类图而言的，比如待评价分析类图中有两个有用类，对应的在标准分析类图中有两个有用类。如果之间的关系为继承，同时之间的关系也为继承，此时就认为之间的关系是正确的，并且和之间的关系是相同的。

当需要比较的两个关系是关联关系关系（1对1,1对多，多对多）时，此时将关系的正确性分为普通关系正确性和多重性关系正确性。普通关系正确性指的是不考虑多重性的影响，比如之间是关联关系关系，之间也是关联关系关系，那么在普通关系层面上，这个关系就是正确的。多重性关系正确性指的在普通关系正确的基础上，两个关系的多重性也必须相同，比如之间是1对多关系，但是之间却是1对1关系，那么在多重性关系层面上，这个关系就是错的。

记。对于，，若均为关联关系关系，并且与相关的两个类和与相关的两个类是相似的，那么即为一对在普通关系层面上正确但在多重性关系层面错误的关系，将作为元素放入集合中。

定义关系正确性的得分为：，其中表示标准分析类图中总的关系个数，表示两个分析类图中相似的关系；表示待评价分析类图中在普通关系层面上正确但在多重性关系层面错误的关系的个数。

（v）将（i）、（ii）、（iii）、（iv）的结果带入评分公式，计算待评价分析类图的总得分。具体过程为：



其中分别为完整性、核心类、关系复杂度、关系正确性的得分。

软件质量的度量可以从很多维度出发，软件内部的模块设计是其中很重要的一个维度，如果一个软件内部模块设计混乱，模块之间的调用关系复杂，不能很好的满足高内聚、低耦合等要求，那么它的其它维度做得再好，其质量也会大打折扣。分析类图是软件生命周期的早期阶段的产物，能反映软件内部模块的设计情况。

本发明通过将待评价分析类图和标准分析类图进行比较，可以在早期阶段就发现自己设计的软件在模块设计上的问题，从而在一定程度上度量了软件的质量。

**附图说明**

图1是标准分析类图；

图2是待评价分析类图。

**具体实施方式**

步骤1：定义分析类图的评价指标、外部输入以及评分公式。

评价指标：为四个指标，即：完整性、核心类、关系复杂度、关系正确性；其中，所述关系正确性包括：普通关系正确性、多重性关系正确性指标；

外部输入：给出核心类、类相似度阈值、权重；所述权重为0到1之间的数；其中，权重具体包括：

（i）完整性的权重、核心类的权重、关系复杂度的权重、关系正确性的权重，其中；

（ii）普通关系正确性的权重、多重性关系正确性的权重，其中；

评分公式为：



其中分别代表完整性、核心类、关系复杂度、关系正确性的得分。

步骤2：以标准分析类图和待评价分析类图的XML文档作为输入，根据评价指标从XML文档中抽取用于比较的数据。

根据评价指标可知，进行差异比较时需要抽取的信息包括：类、属性、类与类之间的关系。

步骤3：对抽取的数据进行差异比较。

* 当待比较的两个类属于3种情况（类名是同义词、类名中包含相同的关键字、类名的相似度阈值不小于给定的相似度阈值）中的一种时就认为它们是相似。
* 当两个类是相似的，此时对于两个类中的属性，如果它们是同义词或者包含相同的关键字，那么就认为这两个属性是相似的。
* 当待比较的两个类与类之间的关系包含相似的类名，并且关系的类型相同时，就认为这两个关系是相似的。

步骤4：将差异比较的结果带入评分公式，计算得分，完成对待评价分析类图的评价。

（i）根据差异比较结果，统计待评价分析类图中正确的属性的个数和标准分析类图中所有属性的个数，计算完整性得分；完整性得分公式为：

（ii）根据差异比较结果，统计待评价分析类图中正确的类的个数和标准分析类图中所有类的个数，计算核心类得分；核心类得分公式为：

（iii）根据差异比较结果，计算关系复杂度得分；关系复杂度得分公式为：



（iv）根据差异比较结果，计算关系正确性得分；关系正确性得分公式为：



（v）将（i）、（ii）、（iii）、（iv）的结果带入评分公式，计算待评价分析类图的总得分。总得分公式为：



**实施例**

为了详细说明分析类图评价的各步骤，本实施例选择火车票订票系统作为示例进行描述。

下面结合附图对本发明实施例描述如下：

步骤1：定义分析类图的评价指标、外部输入和评分公式

定义总分为1.0分，类相似度阈值为0.8，核心类为订单、车票、付款、退票，四个大指之间的权重相等，两个小指标之间的权重相等。

步骤2：使用EA软件生成标准分析类图和待评价分析类图的XML文档，根据评价指标抽取用于比较的信息

表1 分析类图信息表

|  |  |  |
| --- | --- | --- |
|  | 标准分析类图 | 待评价分析类图 |
| 类 | 用户、订单管理、订单、付款  退票、车票、车票描述、车站 | 消费者、订单、付钱、退票  车票、描述1、描述2、车站 |
| 属性 | 用户.姓名、用户.证件号  订单.订单号、订单.时间  付款.金额  退票.退票金额  车票.车票编号、车票.日期  车票描述.车票编号、车票描述.价格  车站.名字、车站.车站编号 | 消费者.姓名、消费者.年龄  订单.时间  付款.价格  退票.金额  车票.编号、车票.日期  描述1.车票编号、描述1.价格  描述2.编号、描述2.价格  车站.名字、车站.车站编号 |

表2 分析类图关系信息表

|  |  |
| --- | --- |
|  | 关系 |
| 标准分析类图 | 关联关系：用户 -> 订单(1...\*)  聚合关系：订单 -> 订单管理  关联关系：订单 -> 付款(1...1)  关联关系：订单 -> 车票(1...\*)  关联关系：车票 -> 退票(1...1)  关联关系：车票 -> 车票描述(\*...1)  关联关系：车票描述-->车站(1...1)  关联关系：车站-->车票描述(1...1) |
| 待评价分析类图 | 关联关系：消费者 -> 订单(1...\*)  关联关系：订单 -> 付钱(1...\*)  关联关系：订单 -> 车票(1...\*)  关联关系：车票 -> 退票(1...1)  关联关系：车票 -> 描述1(\*...1)  关联关系：退票 -> 描述2(1...1)  关联关系：描述1 -> 车站(1...\*) |

步骤3：对抽取的数据进行差异比较

根据同义词信息，待评价分析类图和标准分析类图对比，其缺少的信息和多余的信息如下：

表3 分析类图差异比较结果信息表

|  |  |  |
| --- | --- | --- |
|  | 缺少的信息 | 多余的信息 |
| 类 | 订单管理 | 描述2 |
| 属性 | 用户.证件号、订单.订单号  付款.金额 | 消费者.年龄、付钱.价格  描述2.编号、描述2.价格 |

表4 待评价分析类图普通关系差异比较结果表

|  |  |
| --- | --- |
|  | 普通关系 |
| 缺少 | 聚合关系：订单 -> 订单管理  关联关系：车站-->车票描述(1...1)  关联关系：车票描述-->车站(1...1) |
| 多余 | 关联关系：描述1 -> 车站(1...\*) |

表5 待评价分析类图多重性关系差异比较结果表

|  |  |
| --- | --- |
|  | 多重性关系 |
| 缺少 | 关联关系：订单 -> 付钱(1...1) |
| 多余 | 关联关系：订单 -> 付钱(1...\*) |

步骤4：将差异比较结果带入评分公式，计算得分，完成对评价分析类图的评价

根据差异比较结果可知：待评价分析类图在完整性部分的得分为0.75分，在核心类部分的得分为1.0分，在关系复杂度部分的得分为1.0分，在关系正确性部分的得分为0.56分，总分为0.83分。

以上过程表明待评价分析类图中缺少部分属性，拥有外部输入中要求的全部核心类，有用类之间的关系也并不复杂，但有用类之间的关系正确性不高。说明该软件在内部划分模块时，保证了基本必要的模块，同时模块之间的调用关系并不复杂，但模块之间调用的方式存在一定的问题。最终得分为0.83分，说明软件质量较好。

**附图：**

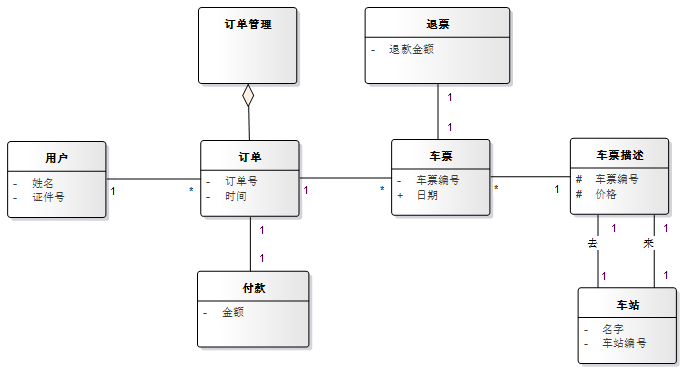


图1

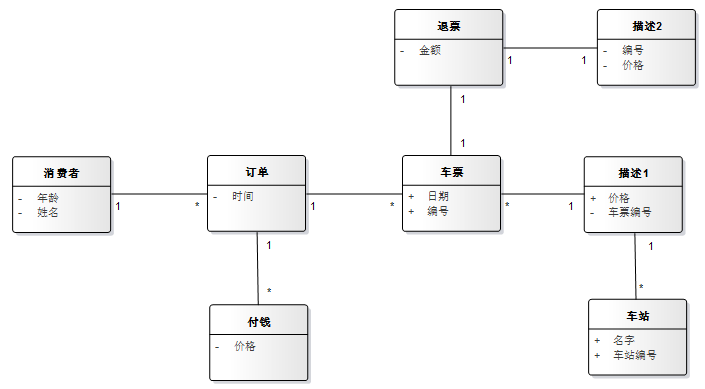


图2