

基于相似性计算的领域模型语义查询

余敦辉^{1,2}, 何克清¹, 李 兵¹

(1 武汉大学 计算机学院, 湖北 武汉 430070; 2 湖北大学 数学与计算机科学学院, 湖北 武汉 430062)

摘 要: 为提高领域资产库的复用能力, 解决领域模型与用户的个性化需求不能完全匹配时仍能为目标应用系统的构建提供支持的问题, 提出了领域模型语义查询算法. 通过服务模型相似度计算结果, 建立同一问题域中同类模型之间的语义关联, 并以此为基础, 提出领域模型的语义查询算法. 最后介绍了该算法在领域建模平台中的应用.

关键词: 语义查询; 领域模型; 服务模型; 语义相似性

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2011)09-0176-03

Semantic Query for Domain Models Based on Similarity Calculation

YU Dun-hui^{1,2}, HE Ke-qing¹, LI Bing¹

(1 School of Computer, Wuhan University, Wuhan 430070, China;

2 School of Mathematics and Computer Science, Hubei University, Wuhan 430062, China)

Abstract: To improve support for the development of target application system, the algorithm of semantic query for domain models is proposed. In particular, semantic similarities among service models are calculated by use of domain Ontology, then the semantic relationships between similar models among the same problem domain, based on which the semantic query algorithm of domain model is proposed. Finally, the application of this algorithm in the domain modeling platform is to introduced.

Key words: semantic query; domain model; service model; semantic similarity

1 引言

随着计算机网络尤其是 Internet 的日益普及和广泛应用, 网络式软运应运而生, 其设计和开发呈现了“用户主导、面向领域、柔性生产”的特点^[1-2]. 其中, 领域的可复用的共性资产是当前网络式软件敏捷开发和满足用户个性化和多元化目标的重要基础^[3]. 为此, 领域建模这一手段被采用, 而网络化软件的领域建模是在需求元建模框架 RGPS 的指导下完成的^[4-5]. 其基本思想是借助领域实体本体和操作本体的支持, 尝试利用角色、目标、过程和服务刻画网络式软件用户需求的真实意图, 针对特定应用领域中用户共性的需求建立相应的领域模型, 为目标应用系统构建时的复用奠定基础.

在此框架下, 大量的角色模型、目标模型、过程模型以及服务模型被创建. 并根据功能进行聚类, 将其中一些在业务上存在因果联系、或者在功能上相互依赖、相互制约的模型组织在一起, 形成一个问题域. 并且从隶属于同一个问题域的领域模型中抽取其基本信息, 生成领域问题本体, 使得针对同一问题域内的领域模型融合为一个整体. 并借助于领域问题本体的指导, 实现对领域模型的复用^[6].

然而, 在同一问题域中, 往往存在很多语义上非常相似甚至完全相同的模型, 另一方面, 面对用户某一个个性化需求, 领域资产库中可能没有与其完全匹配的领域模型. 为此, 网络化软件中更为强调领域模型对于用户需求的可满足性. 即面对用户的某些个性化需求, 待构建应用系统可能需要基于语义, 提

收稿日期: 2011-01-13; 修回日期: 2011-02-15

基金项目: 国家自然科学基金项目(60873083, 60803025, 60970017, 60903034); 国家“八六三”计划项目(2006AA04Z156); 国家“九七三”计划项目(2006CB708302, 2007CB310801); 湖北省青年杰出人才基金(2008CDB351)

供给用户多个可行的领域模型,供用户进行选择。

文中提出了领域资产的语义查询解决方法,其基本思路是首先计算领域资产中所有服务模型两两之间的语义相似度,进入建立起服务模型之间的关联,在此基础上,根据服务模型与过程模型之间的对应关系,计算出所有过程模型之间的语义关联,进而计算出所有目标模型之间的语义关联。同时,按模型之间语义关联的紧密程度设定语义关联等级。这样,在构建目标应用系统时,针对用户提出的某一个个性化需求,系统根据用户对模型匹配准确度的要求,按照不同的语义关联等级返回匹配的领域模型供用户选用。通过这种方法,能够有效地提高领域资产库的复用能力,为目标系统构建提供更好的支持。

2 服务模型语义相似度计算

服务模型之间的语义相似性可以从服务的操作、属性、输入输出、及其内部结构等四个层面来进行综合计算。服务的操作和输入输出联合起来体现了服务所能完成的功能,而服务的属性主要是指1类似于响应时间、花费等质量属性,它是服务可用性可靠性评价的重要指标。服务的内部结构则体现了实现某个业务功能的内部流程。

其中,服务模型的结构相似度 $\text{Sim}_s(\text{SM}_1, \text{SM}_2)$ 的计算可参考文献[7]实现,在此不作重点阐述。而前三个方面的相似度都可以基本领域本体,利用词语的相似度来进行计算,为此,先给出两个词语 w_1, w_2 的相似度算法。

定义 1 (词语相似度) 两个词语 w_1, w_2 的相似度可以参考文献[8] 如下进行计算:

$$\text{Sim}_w(w_1, w_2) = \frac{C(w_1) \cap C(w_2)}{\text{Distant}(w_1, w_2) * (\rho * |h_{w_1} - h_{w_2}| + 1)}$$

其中, $C(w_1) \cap C(w_2)$ 表示两个词 w_1, w_2 语在领域本体中的语义重合度, $\text{Distant}(w_1, w_2)$ 表示 w_1, w_2 之间的语义距离; ρ 为调节因子; h_{w_1}, h_{w_2} 为 w_1, w_2 所分别对应的层次深度。

定义 2 (服务操作相似度) 两个服务模型 SM_1, SM_2 的操作 op_1 与 op_2 之间的相似度计算: $\text{Sim}_{op}(\text{SM}_1, \text{SM}_2) = \text{Sim}_w(op_1, op_2)$

定义 3 (服务属性相似度) 两个服务模型 SM_1, SM_2 的所有属性之间的相似度可如下计算:

$$\text{Sim}_p(\text{SM}_1, \text{SM}_2) = \sum_{i=1}^m (\max \sum_{j=1}^n (\text{Sim}_w(pr_{1i}, pr_{2j}))) / \min(m, n)$$

定义 4 (服务输入输出相似度) 两个服务模型 SM_1, SM_2 的输入输出参数之间的相似度计算如下:

$$\begin{aligned} & \text{Sim}_i(\text{SM}_1, \text{SM}_2) \\ &= (\max \sum_{i=1}^u \sum_{j=1}^v \text{sim}_w(in_{1i}, in_{2j}) * x_{ij}) / \min(u, v) \\ & \text{Sim}_o(\text{SM}_1, \text{SM}_2) \\ &= (\max \sum_{i=1}^s \sum_{j=1}^t \text{sim}_w(out_{1i}, out_{2j}) * x_{ij}) / \min(s, t) \\ (AP) \text{ s. t. } & \sum_{i=1}^u x_{ij} = 1 \quad (i = 1, 2, \dots, u) \\ & \sum_{j=1}^v x_{ij} = 1 \quad (j = 1, 2, \dots, v) \end{aligned}$$

其中, AP 中第一个约束表示服务模型 SM_1 中第 i 个参数如 in_{1i} 只能和 SM_2 中的参数匹配一次, 第二个约束则表示 SM_2 中第 j 个参数如 in_{2j} 只能和 SM_1 中的参数匹配一次。

定义 5 (服务相似度) 两个服务模型 SM_1, SM_2 的相似度可以通过如下计算:

$$\text{Sim}(\text{SM}_1, \text{SM}_2) = ((\text{Sim}_{op}(\text{SM}_1, \text{SM}_2) + \text{Sim}_p(\text{SM}_1, \text{SM}_2) + \text{Sim}_i(\text{SM}_1, \text{SM}_2) + \text{Sim}_o(\text{SM}_1, \text{SM}_2) + \text{Sim}_s(\text{SM}_1, \text{SM}_2)) / 5)$$

3 领域模型语义关联的建立

领域模型之间的语义关联关系一定存在于同一问题域中, 因此, 应该以问题域为考虑范围, 对隶属于同一问题域的同类模型建立它们之间的相似性关联。对于问题域中的四类领域模型, 建立其关联的方法存在很大的不同。

(1) 服务模型

基于服务模型两两之间的语义相似度计算结果, 若存在 $\text{Sim}(\text{SM}_i, \text{SM}_j)$ 大于设定的阈值, 就找到模型 SM_i 与 SM_j 在领域问题本体 DPO_k 中的节点 $\text{Node}_i, \text{Node}_j$, 给这两个节点各自增加一个属性, 用于记录其具有语义相似性关联的模型的基本信息, 然后分别找到与 SM_i, SM_j 对应的过程, 建立它们的相似关系。

(2) 过程模型

遍历 DPO_k 中的所有过程模型, 若发现存在于两个过程模型 PM_i, PM_j 中的很多过程彼此之间具有相似关系, 且过程之间的控制关系基本相同, 则认为这两个过程模型 PM_i, PM_j 相似, 找到模型 PM_i 与 PM_j 在领域问题本体 DPO_k 中的节点 Node_i 与 Node_j , 给这两个节点各自增加一个属性, 用于记录其具有语义相似性关联的模型的基本信息。

(3) 目标模型

基于模型聚类时领域目标模型语义相似度计算结果,将问题域 $PD[k]$ 中的所有与指定的目标模型 GM_i 的语义相似度大于设定的阈值的目标模型找到,然后找到模型 GM_i 的顶层目标在领域问题本体 DPO_k 中的节点 $Node_i$,给该节点增加一个属性,用于记录其具有语义相似性关联的模型的基本信息。

(4) 角色模型

一般一个问题域中可能只有一个或少数几个角色模型,这些模型中所涉及的角色基本不同,所以无需建立它们之间的语义相似性关联。

4 领域模型语义查询算法

有了领域模型之间的这种语义相似性关联,待构建应用系统复用领域资产时,将会对和其对应的领域问题本体进行语义查询。

领域资产的语义查询算法:

输入:需求目标 $Goal_i$;

输出:和需求目标 $Goal_i$ 相匹配的角色模型 RM_i 、目标模型 $GM_1, \dots, GM_j$ 、过程模型 $PM_1, \dots, PM_k$ 、服务模型 $SM_1, \dots, SM_m$;

{
基于需求目标 $Goal_i$,参照目标模型语义相似性算法找到与该目标最匹配的领域问题 $PD[k]$;

在 DPO_k 中找到与需求目标 $Goal_i$ 最匹配的领域目标模型 GM_i ;

根据 GM_i 的语义相似性关联属性找到所有可能匹配的目标模型 $GM_1, \dots, GM_j$,供用户选择;

for (用户选取的每一个目标模型 GM_i) {

找到 GM_i 所关联的角色 $Role$,找到 $Role$ 所在的角色模型 RM_i ;

for (该 GM_i 中的每个可操作目标 OG_{ij}) {

在 DPO_k 中找到对应的过程模型 PM_{ij} ;

根据 PM_{ij} 的语义相似性关联属性找到相关联的 $PM_1, \dots, PM_k$;

}

将查询得到的所有过程模型 $PM_1, \dots, PM_k$ 提供给用户去选择;

for (用户选取的每一个过程模型 PM_i) {

依次找出 PM_i 中的每一个过程 $Process_{ij}$;

for (每个过程 $Process_{ij}$) {

在 DPO_k 中找到对应的服务模型 SM_{ij} ;

找到和 SM_{ij} 相关联的其他服务模型 $SM_1, \dots, SM_m$;

}

将查询得到的所有服务模型 $SM_1, \dots, SM_m$ 提供给用户去选择。

}

}

}

5 结束语

文中针对领域资产不能完全满足应用系统构建时用户的个性化需求的问题,提出了对领域模型的语义查询的解决方案.即面对用户的个性化需求,待构建应用系统可能需要基于语义,提供给用户多个可供复用的领域模型,供用户进行选择.通过这种方法,有效地提高了领域资产的可复用程度,增强了领域资产库对目标应用系统开发的支持力度。

参考文献:

- [1] 何克清,彭蓉,王健,等.网络式软件[M].北京:科学出版社,2008.
- [2] Turner M, Budgen D, Brereton P. Turning software into a service[J]. IEEE Computer, 2003(36):38-44.
- [3] 王健.网络式软件的需求元建模框架及关键技术研究[D].湖北:武汉大学,2008.
- [4] Wang J, Kq He, Gong P, et al. RGPS: a unified requirements meta-modeling frame for networked software[C]// Proc of Third International Workshop on Advances and Applications of Problem Frames (IWAAPF08) at 30th International Conference on Software Engineering (ICSE '08), Leipzig, Germany: IEEE, 2008.
- [5] Wang J, Ke He, Li B, et al. Meta-models of Domain Modeling Framework for Networked Software[C]// Proc of the Sixth International Conference on Grid and Cooperative Computing. Urumchi, China: IEEE, 2007: 878-885.
- [6] 余敦辉,何克清,李兵,等.基于问题域的领域模型库度量[J].微电子学与计算机,2010,8(27):41-43.
- [7] Dh Yu, Ke He, Li B, et al. Research on clustering algorithm for domain model based on semantic similarity calculation[C]// Proc 2010 International Conference on Future Industrial Engineering and Application (ICFIEA 2010). Shenzhen, China: IEEE, 2010:416-419.

作者简介:

余敦辉 男,(1974-),博士后.研究方向为软件工程、面向服务的需求工程。