

范例推理结构相似性研究^{*}

郑浩森¹ 汤胤² 彭宏² 郑启伦²

(湛江教育学院 湛江524037)¹ (华南理工大学计算机科学与工程学院 广州510640)²

摘要 本文将范例的结构相似性问题提炼为结构图的相似问题,改变了将范例属性线性地映射到区间的方式,建立以关键节点为特征的分层模型,来进行范例结构图比较,并得到一个可以计算的合理的相似度数值,可以较为方便地解决结构相似性度量的问题,为复杂范例的多种类型的带偏序的属性的综合比较提供了基础。最后本文提供了一个例子,对相似性模型作了说明。

关键词 范例推理,范例表示,结构相似性

Research on Structural Similarity of Case-Based Reasoning

ZHENG Hao-Sen¹ TANY Yin² PENG Hong² ZHENG Qi-Lun²

(Zhangjiang Education College, Zhangjiang 524037)¹

(Institute of Computer Science, South China University of Tech., Guangzhou 510640)²

Abstract Abstracting the problem of structural similarity of cases to be the similarity of graph, this paper proposes the similarity model, based on maximum isomorphic sub-graph, characterized with several levels of key nodes, which enables the comparison among cases of complex structures with simple numerical result. The model has provided the basis of comprehensive comparison among the partially-ordered attributes of different types. A example is employed to illustrate the whole similarity model.

Keywords CBR, Case representation, Structural similarity

1 引言

基于范例推理(Case-Based Reasoning, CBR)在人工智能领域是一种非常有用的方法,引起了人工智能研究者的广泛关注,它已经应用到辅助复杂的工程设计如建筑业、工程制造等行业,显示出广泛的应用前景。

Friedrich Gebhardt 认为当范例比较除了单纯属性的比较外,有时其相似性主要基于结构图的比较,结构图的比较又基于子图同构^[4]。Holyoak 和 Thagard 1989年的约束满足理论认为范例相似性主要是基于谓词的语义相似性,基于结构的结构相似性,基于关键词的语用相似性等等^[2]。Guus Schreiber 的知识模板将知识分为分类、评估、诊断、监控、综合、配置设计、分配、规划、调度等等,并给出了各类模板的结构特性^[3]。

在 CBR 循环中,相似性度量问题可以说贯穿几乎所有的范例推理的环节如范例库组织、范例搜索获取、类比转换等等。遗憾的是,对于相似性的度量基本上都是各自将范例的属性映射到一个实数区间,然后以曼哈顿距离^[2]等作为范例相似性的判断标准。这些方法并没有考虑范例的知识结构问题,因为范例往往是比较复杂的知识,具有比较复杂的知识结构,其属性往往并不是简单的线性结构或者单层的子图可以描述的。

本文研究范例的结构相似性问题,建立以关键节点为特征的分层模型来进行范例结构图比较,并得到一个可以计算的合理相似度数值,可以较为方便地解决结构相似性度量的问题。

2 模型定义

我们的核心思想,首先是范例的知识利用图来表示,主要

是将范例中的概念、特征、行为或其它关系表示成问题空间中的点,即概念获取和性质描述过程,使其内外部变为内部形式,并努力保证该信息的完整,并形成范例结构图。而其结构图通常在一个具体问题中,有些关键步骤是不可省略的,有些则是可有可无的,对范例结构图中的节点按重要性依次分级,如1级节点、2级节点等等。将范例结构图的节点按重要性分级以后,通过子图同构来定义相似度,如*i*级相似度定义为两个范例结构图*i*级节点及弧的所构成的子结构图的最大同构子图所包含的节点数的百分比。

值得注意的是,我们在这里不但关注范例的结构相似性问题,还将节点类型的各自不同纳入考虑范围。

定义1(范例知识结构图) 对范例通过特征抽取等抽象手段,得到的图 $G(V, R)$,其中 V 是基于知识模板^[3]的一定类型的节点的集合; R 是集合 V 中的节点间的有向弧集合。这里,节点描述知识中的实体概念、行为过程、属性特征等等,节点类型由知识模板给出;节点间的有向弧是这些节点间的优先顺序、时间顺序、空间顺序、约束、依赖等关系。

如烹饪是一个典型的规划过程,烹饪过程中的某些步骤是关键性的并且次序不可以更改的。以番茄炒蛋为例,其知识图可以简略表示为:

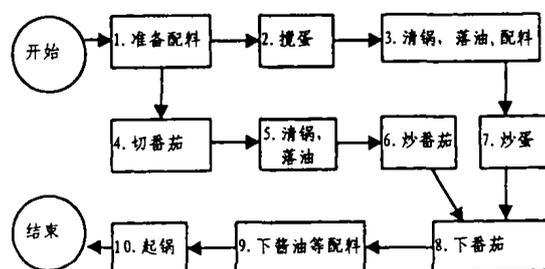


图1 一个典型的烹饪知识结构图

^{*}广东省科技攻关项目(A1020103)。郑浩森 硕士,研究方向:人工智能与数据挖掘;汤胤 博士生,研究方向:人工智能与数据挖掘;彭宏 教授,博导,研究方向:智能计算技术;郑启伦 教授,博导,研究方向:智能计算技术,神经网络理论及应用。

请注意图1中,我们省却了一些诸如对煮熟程度的判断,对时间的描述。图中,节点8必须在6,7之前,节点8以前的工作是可以并行的。另外,有些步骤如炒蛋,下番茄等显然是其中关键的节点。

范例结构图不是本文讨论的重点,这里从略。

定义2(节点分级与图的分级) 设范例A与范例B的结构图,分别为结构图a与结构图b,对它们指定的节点进行分k级,即从1,2,...,k。

将结构图a中 $j \leq i$ 级的所有节点和这些节点之间的弧,形成的子结构图称为 a_i 。这样我们分别得到了结构图a与结构图b的各级子图 $(a_1, b_1), (a_2, b_2) \dots (a_k, b_k)$, 相应的节点数分别为 $(m_1, n_1), (m_2, n_2) \dots (m_k, n_k)$ 。经过分级的范例结构图是我们讨论的基础。

定义3(同构子图) 对结构图 $a(V_a, R_a)$ 与结构图 $b(V_b, R_b)$, 其中 V_a, R_a, V_b, R_b 分别为结构图a和结构图b的节点集和有向弧集,则结构图a和结构图b的同构子图定义为在合理的节点对应情况下的图

$$c = c(V_c, R_c) \quad (2.1)$$

其中,

$$V_c = V_a \cap V_b \quad (2.2)$$

$$R_c = R_a \cap R_b \quad (2.3)$$

这里提到的合理的节点对应是指对应的节点必须是同一类型并且相同级别,以满足知识类型对应的需要。

定义4(最大同构子图) 在同构子图的集合中,包含最多节点的同构子图称为最大同构子图。有时,我们可能求出多个最大同构子图,则取所有节点出入度的总和最大者,若还有多个,则取第一个匹配成功者。

定义5(范例图的结构相似性) 结构图a与结构图b的i级相似定义为,对第i级结构图 a_i 和图 b_i , 求得它们的最大同构子图 $G_{a_i} \cong G_{b_i}$, 使得对应节点对中含有i级节点对最多, 设i级节点对共有 k_i 个, 则结构图a和结构图b相似或结构图a与结构图b的i级相似度为:

$$S_i = \frac{k_i}{\max\{m_i, n_i\}} \quad (2.4)$$

例如,结构图A与结构图B的2级相似就是,对结构图A和结构图B的2级结构图 a_2 和 b_2 , 求得最大的同构子图 $G_{a_2} \cong G_{b_2}$, 使得子结构图中含有的节点数最多, 设2级节点对共有 k_2 个, 则结构图a与结构图b的2级相似度为:

$$S_2 = \frac{k_2}{\max\{m_2, n_2\}} \quad (2.5)$$

现定义结构图a与结构图b的相似度为:

$$S = \sum_{i=1}^k S_i \cdot w_i \quad (2.6)$$

这里, $\sum_{i=1}^k w_i = 1$, 一般来说, $i < j$ 时, $w_i > w_j$ 。可以看到结构图a与结构图b的相似度是由各级相似度加权得到的,在知识的结构中,往往有一些步骤是要比其他步骤重要,这些节点构成的子结构图的相似性自然比其他节点构成的子结构图重要。

算法1 最大同构子图算法

输入 结构图 $a(V_a, R_a)$, 结构图 $b(V_b, R_b)$ 以及它们节点类型和级别信息。

输出 子图 $c(V_c, R_c)$

1. 对于结构图 $a(V_a, R_a)$, 求其1级子图 $a_1(V_{a1}, R_{a1})$;
2. 对于结构图 $b(V_b, R_b)$, 求其1级子图 $b_1(V_{b1}, R_{b1})$;
3. 求 $c_1 = (V_{c1}, R_{c1}) = (V_{a1} \cap V_{b1}, R_{a1} \cap R_{b1})$;
4. $\max \leq$ 节点分层数;
5. $i = 1$;
6. 在 c_i 的基础上, 添加 $i+1$ 级节点, 以及 $i+1$ 节点间的弧和 $i+1$ 节点与 $j < i$ 级节点间的弧, 根据节点类型进行匹配, 得到 $c_{i+1}(V_{a+i+1}, R_{a+i+1})$;

7. $i = i + 1$;
8. if $i < \max$, then go to 6;
9. 输出结果
10. 结束

算法2 结构图相似度算法

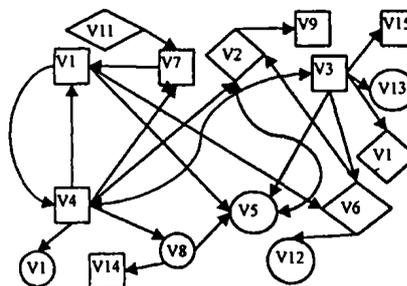
输入 结构图 $a(V_a, R_a)$, 结构图 $b(V_b, R_b)$ 以及它们节点类型和级别信息。

输出 每级相似度 S_i 和总相似度 S

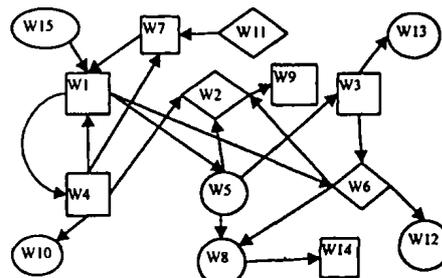
1. 对于结构图 $a(V_a, R_a)$, 求其1级子图 $a_1(V_{a1}, R_{a1})$;
2. 对于结构图 $b(V_b, R_b)$, 求其1级子图 $b_1(V_{b1}, R_{b1})$;
3. 求 $c_1 = (V_{c1}, R_{c1}) = (V_{a1} \cap V_{b1}, R_{a1} \cap R_{b1})$;
4. 取得 c_1 的1级节点数目, 计算 S_1 并输出;
5. $\max =$ 节点分层数;
6. 输入每级的权值 $\{w_i\}$;
7. $i = 1$;
8. 在 c_i 的基础上, 添加 $i+1$ 级节点, 以及 $i+1$ 节点间的弧和 $i+1$ 节点与 $j < i$ 级节点间的弧, 根据节点类型进行匹配, 得到 $c_{i+1}(V_{a+i+1}, R_{a+i+1})$;
9. 取得 c_i 的 i 级节点数目, 计算 S_i 并输出;
10. $i = i + 1$;
11. if $i < \max$, then go to 8;
12. 输出结果 S ;
13. 结束

3 一个简单的例子

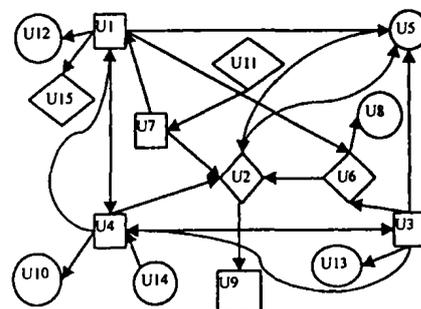
设范例A, B, C的结构图分别如下所示, 节点共分3级, 令 $w_1 = 0.70, w_2 = 0.20, w_3 = 0.10$, 我们现在计算范例B对A和范例C对A的相似度 S 。



(1) 范例A的结构图



(2) 范例B的结构图



(3) 范例C的结构图

• 这里 □ ○ ◇ 分别表示不同类型的节点

图2 范例A, B, C的结构图

三个结构图的节点分级情况如表1所示。

表1 结构图 A,B,C 的节点分级情况

	A	B	C
1级	v1,v2,v3,v4,v5	w1,w2,w4,w5	u1,u2,u3,u4,u5
2级	v6,v7,v8	w6,w7,w8	u6,u7
3级	v9,v10,v11,v12, v13,v14,v15,v16	w9,w10,w11,w12, w13,w14,w15	u8,u9,u10,u11, u12,u13,u14,u15

经过匹配算法,我们得到:

一级节点对

A 与 B: (v1,w1), (v2,w2), (v4,w4), (v5,w5)

A 与 C: (v1,u1), (v2,u2), (v3,u3), (v4,u4), (v5,u5)

二级节点对:

A 与 B: (v6,w6), (v7,w7)

A 与 C: (v6,u6), (v7,u7)

三级节点对:

A 与 B: (v9,w9), (v10,w10), (v11,w11), (v12,w12)

A 与 C: (v9,u9), (v10,u10), (v11,u11), (v12,u8), (v13,u13)

计算结果如表2所示。

表2 范例 A,B,C 的结构相似度计算表

	B 与 A 的相似	C 与 A 的相似
1级, w1=.70	4/5	5/5
2级, w2=.20	2/3	2/3
3级, w3=.10	4/8	5/8
加权总计	≈.74	≈.90

4 模型讨论

本文提出的知识结构相似模型,是基于节点分级的最大

同构子图。这样的模型好处在于:

(1)一个范例结构图中,并不是所有节点都是同等重要的,我们有区别地对待结构图中的节点,提出了这样的概念:知识结构的相似,首先要满足关键节点及其关系的相似,然后才是次要和再次要节点及其关系的相似。通俗地说,两件事情要相似肯定要保证关键问题相似,在次要问题相似而关键问题不相似的情况下,我们不可以认为两件事情是相似的。本文的模型避免了对节点一视同仁,粗放的角度而采用了相对精细的视角去看问题。

(2)一次性纳入知识结构图的所有节点和弧来计算相似性,显然计算量巨大,我们的分级方法使得结构图的比较的计算量大大减小,仅在关键节点和弧相似的情况下进入下一层更精细的比较,避免了无谓的计算和节点数量和弧过多带来的匹配错误问题,可以大大提高效率。

结论 本文针对范例的结构相似性问题,提出一个按关键节点分层的基于最大同构子图的相似性模型,使得复杂结构的范例知识可以进行比较,并得到简单的数值结果,为复杂范例的多种类型的带偏序的属性的综合比较提供了基础。

参考文献

- 1 Leake D B. CBR in Context; The Present and Future. AAAI Press/MIT Press, 1996
- 2 史忠植. 知识发现. 清华大学出版社, 2002
- 3 Schreiber G. 知识工程和知识管理. 史忠植, 梁永全, 吴斌译. 机械工业出版社, 2003
- 4 Gebhardt F. Structure oriented case retrieval. [Fabel-Report 39]. GMD, Sankt Augustin, Dec. 1995

(上接第140页)

业务流程运行。

5 利用 SynchroFLOW 实现企业应用集成

SynchroFLOW 降低业务逻辑、业务数据和业务操作实体三者间的耦合,在企业应用异构环境中可以实现功能构件级和流程级不同层次的集成。软件开发商开发的 EJB、应用程序、FORM 和与过程定义有关的附件通过系统管理进行注册和注销;用 FORM 开发工具开发的 FORM 通过应用程序管理提供的 API 进行注册;EJB 通过 EJB 部署工具部署到 EJB 服务器上;EJB、应用程序、FORM 和附件的索引信息保存到应用程序数据库中。利用 SynchroFLOW 的 FORM 开发工具开发的 FORM 或用第三方开发工具(如 FrontPage)开发的 FORM 以及以 JSP、HTML、EJB 封装的组件,通过注册,成为一个业务活动功能构件。

在企业应用中有很多应用程序是二次开发商和用户自己开发的,有的是第三方开发的 COST 组件,这些应用程序是 workflow 虚拟机所不认识的。SynchroFLOW 通过应用程序的注册和注销机制实现对它们的集成。应用程序的集成有两种方式,一种是将做好的应用程序作为活动一个功能构件,将应用程序索引信息登记在应用程序管理中,供过程定义时集成到活动中,当业务流程运行时它随活动一块执行;另一种是调用应用程序,由相应的 API 进行处理客户端注册应用程序的请求,还有响应调用应用程序的请求。

结束语 西安协同公司开发业务流程管理与集成系统

SynchroFLOW 的过程中采用并实现了面向业务的软件体系结构这一思想。该软件成功应用在电子政务、电讯等行业中,特别是在中国电信本地网资源管理工程选型中,选型专家组对 Exchange、Flowmark、WebLogic Integration 和 SynchroFLOW 从业务的表达能力、对异构系统的集成性和业务的柔性管理等方面进行比较,最后决定中国电信的19个省的本地网资源管理系统的工作流平台软件所需的167套 workflow 管理软件全部采用 SynchroFLOW。中国电信通过基于 SynchroFLOW 开发的资源管理系统提高面向网络资源的管理能力、面向业务的处理能力,提高面向市场的竞争能力。

参考文献

- 1 Kalakota R, Robinson M. E-Business 2.0 Roadmap for Success. Addison-Wesley, 2001
- 2 Bass L, Clements P, Kazman R. Software Architecture in Practice. Addison-Wesley, 1998
- 3 Lauesen S. Software Requirements: Styles and Techniques. Addison-Wesley, 2002
- 4 Workflow Management Coalition. Workflow Technology - White paper. <http://www.wfmc.org/standards/docs.htm>, Feb. 1998
- 5 杨斌,郝克刚. 基于 Web 的工作流管理系统的解决方案. 西北大学学报(自然科学版), 1999(6): 491~494
- 6 岳晓丽,杨斌,郝克刚. 令牌驱动式工作流计算模型. 计算机研究与发展, 2000(12): 1513~1519
- 7 郝克刚,王斌君,安贵. WPD 中的 JOIN 语义问题和分区解决方案. 计算机科学, 2003
- 8 董卫云,李伟宏,郝克刚. 一个企业应用集成的设计模式. 计算机科学, 2003, 30(10)
- 9 西安协同数码股份有限公司. SynchroFLOW 概要设计. 2001