

利用共用本体构造本体映射

郑 诚¹, 陈 涛², 查 琦²

(1. 安徽大学 计算智能与信号处理教育部重点实验室, 安徽 合肥 230039;

2. 安徽大学 计算机科学与技术学院, 安徽 合肥 230039)

摘 要:一些代理机语言使得代理机能够理解代理机之间传递的信息的内容和潜在含义。交互中用到的信息载体是本体, 由于本体是多种多样的, 而且不同的本体对于同一个事物的描述是不一样的, 这就阻碍了代理机之间的交互, 这个问题被称作交互问题。文中提出共用本体的概念, 建立通用的本体体系, 首先建立本体与通用本体之间的映射, 然后构建两个或两个以上异构本体之间的映射。实验表明, 此方法不仅能够解决结构异构问题, 也能解决语义异构问题。

关键词:本体; 本体映射; 结构异构; 语义异构; 交互问题; 共用本体

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2008)10-0020-03

Mapping Ontology Based on Common Ontology

ZHENG Cheng¹, CHEN Tao², ZHA Qi²

(1. Ministry of Education Key Laboratory of Intelligent Computing & Signal Processing,

Anhui University, Hefei 230039, China;

2. School of Computer Science and Technology, Anhui University, Hefei 230039, China)

Abstract: Some agent communication languages enable agents to specify the intention and the content of the information they exchanged. The information carrier in communication is ontology. Because ontologies are various and the descriptions of different ontologies to one thing are different, so it hampers communication and collaboration among agents, which is known as the interoperability problem. Proposed the concept of common ontology. Through establishing the system of common ontologies of the world, create mappings between different ontologies and common ontologies, and then figure out the mapping among them. According to the result of experiments, this method can solve the structural heterogeneity problems and also the semantic heterogeneity problems.

Key words: ontology; ontology mapping; structural heterogeneity; semantic heterogeneity; common ontology

0 引言

随着网络的飞速发展, Internet 提供了获得信息的新渠道, 语义 WEB 的出现使得 WEB 内容更加丰富, 同时也产生了更多的信息。为了适应飞速增长的信息供求, 需要新的技术, 需要智能工具。代理机就是在这样的环境下产生的, 代理机技术就是一种可靠的分发和收集信息的方法^[1]。

代理机交流语言, 如: ACL、KQML, 在开放的多代理机系统中提供了一种标准。应该将有效的协议和语言运用到代理机上(例如, FIPA 协议, KIF, 和 SP)。交互中使用到的本体^[2]取决于交流的主题。由于主题可

以是无穷多个, 而且不同的本体对于同一个主题的描述是不一样的, 要开发出一套公众接受的标准需要很长时间。在缺乏标准的情况下, 阻碍代理机之间交互的难题被称作交互问题^[3~5]。

交互问题也出现在异构数据库领域中^[2]。Internet 使得访问各自独立开发的数据库成为可能, 不管是属于不同组织的, 还是在网络时代之前处在不同的站点的。除非知道怎么样关联数据库间的信息, 否则需要访问几个数据库的查询将是不可能的。一种关联不同数据库信息的方法是使用本体描述数据库的基本的语义结构。

1 交互能力

为了能达到交互的目的, 首先要解决两个问题: 结构异构和语义异构^[3,6,7]。结构异构是信息的表示的异构。本体所表达的信息可以表示成多种形式。在一个多代理机系统中, 本体是交流的基础。一个代理机

收稿日期: 2008-01-02

基金项目: 安徽省自然科学基金资助项目(050420204); 安徽省高校自然科学基金项目(2006kj055B)

作者简介: 郑 诚(1964-), 男, 博士, 副教授, 主要研究方向为语义 Web、数据挖掘。

的信息实际存储方式以及形式对其所处的环境是屏蔽的。语义异构是信息所表示的潜在意思的不同。举个例子,一个人的信息可以被描述成不同的本体。下面的几点区分本体:(1)不同的语义结构——结构冲突^[2];(2)不同的名字对应相同的信息或是相同的名字对应不同的信息——命名冲突^[3];(3)不同的代号表示相同的数据——数据冲突。由于不同的单位,不同精度的不同表达,数据冲突可以更精确地定义。

下面的两个本体^[8]说明结构和语义异构的形式(一个名称跟上一对括号表示一个本体,如:本体名称(属性列表...)):

Person (christian name, family name, street, city, country, phone number, email address)

本体 1

在本体 1 中, 'street' 包含了门牌号, 'phone number' 也包含了国家代码、区域码以及本地号码。

Person (Name (first name, family name), Address (street, number, City (name, area code), Country (name, country code)), phone number, email address)

本体 2

在本体 2 中 'phone number' 只描述了本地号码,而区域码和国家代码都在不同的属性中。

每个本体都有不同的结构。本体 1 结构是平板的,而本体 2 的结构是层次的。这个结构冲突很容易解决,因为本体 2 或多或少扩展了本体 1。当两个本体有着不同的层次结构,那么结构上的差异就比较大。

两个本体间的命名冲突是一个难题。不同概念的名字用于同一类型的数据,例如 'first name' 和 'christian name'。更甚的是,同一概念的名称用于相差甚微的数据。因此,为了交互性,必须将数据分割以及合并。为了与本体 2 中的 street 映射,本体 1 中的 street 需要拆分。反向映射时需要将本体 2 的某些字段合并。必须找到一个本体间的概念映射,才能达到交互的能力。

2 当前方法的缺点

目前,已经有一些解决语义异构的方法了。但是许多解决方法需要用自动或半自动的方法获得一个共同的本体,参见文献[5,9~11]。这些方法都很依赖下面的假设和条件:

- * 概念是由一些共享的概念组成的;
- * 不同的本体是由一个初始本体分化的结果;
- * 由人工来关联不同本体之间的概念定义,以及解决可能的冲突。

在近几年出现的方法中,Wache H 等指出在两个本体间建立映射仍然是一个开放的难题^[11]。文中介绍的方法可以解决部分的问题。本方法很好解决了本体的结构、命名等一些问题。只需要利用共用本体构建本体,除此之外无需其他额外的信息。

3 映射过程

文中阐述的是利用共用本体构造本体映射。参考一些面向对象语言里的类的概念,一般的面向对象语言会提供一些已经写好的类以供程序员去使用,当然也可以自己建立。这里的共用本体,即是构造一些共用性的本体,以供其他本体做映射的时候进行参考。下面说明一下共用本体的构建(本文引用了文献[3]中的两个例子本体,也就是上述本体 1 和本体 2)。

在 Java 语言中,Sun 公司已经给程序员提供了各种功能的类库,并以包的形式进行分类存放的。所有的类都是从一个叫 Object 的终极类派生出来的。那么在本体构建的时候能不能应用此概念呢?这样可以使不同领域里用到的相同的本体分开。这样就可以避免使用不同领域内的本体。

可以将面向对象的思想应用于共用本体的构建。

下面举个共用本体的例子,共用本体的根本体以及从其派生的本体:

Object

→Creature (name, class, ...)

→Human (Name (firstName, familyName), age, sex, Address (street, number, City (name, code), Country (name, code)), phoneNo, emailAddress ...)

假设 Object 是所有本体的根, Creature 是从其派生出来的一个分支, Human 又是 Creature 的子类,其中 Human 中的 Name, Address, City 以及 Country 都是本体,其余的以小写字母开头的都是本体的属性。

在构建本体 1 和本体 2 时参照上述的本体进行构建。本体 1 与共用本体之间的映射,如表 1 所示。

表 1 本体 1 与共用本体间的映射关系

本体 1	共用本体
Person. christian name	Human. Name. firstName
Person. family name	Human. Name. familyName
Person. street	Human. Address. street
Person. city	Human. City. name Human. City. code
Person. country	Human. Country. name Human. Country. code
Person. phone number	Human. phoneNo
Person. email address	Human. emailAddress

其中 "||" 表示联合前后两个属性,而且是无序联合,即不分次序。用 "|" 表示按次序联合。

本体 2 与共用本体之间的映射,如表 2 所示。

表 2 本体 2 与共用本体间的映射关系

本体 2	共用本体
Person. name. first name	Human. Name. firstName
Person. name. family name	Human. Name. familyName
Person. address. street	Human. Address. street
Person. address. number	Human. Address. number
Person. address. city. name	Human. Address. City. name
Person. address. city. area code	Human. Address. City. code
Person. address. country. name	Human. Address. Country. name
Person. address. country. country code	Human. Address. Country. code
Person. phone number	Human. phoneNo
Person. email address	Human. emailAddress

有了上面本体 1 和本体 2 与共用本体之间的映射,那么本体 1 与本体 2 之间的映射就很容易了。通过共用本体就可以很快建立本体 1 和本体 2 之间的映射关系。

本体 1 与本体 2 之间的映射如表 3 所示。

表 3 本体 1 与本体 2 通过共用本体建立的映射关系

本体 1	共用本体	本体 2
Person. christian name	Human. Name. firstName	Person. name. first name
Person. family name	Human. Name. familyName	Person. name. family name
Person. street	Human. Address. street	Person. address. street
Person. city	Human. City. name Human. City. code	Person. address. city. name Person. address. city. area code
Person. country	Human. Country. name Human. Country. code	Person. address. country. name Person. address. country. country code
Person. phone number	Human. phoneNo	Person. phone number
Person. email address	Human. emailAddress	Person. email address

通过文献[3]中的两个本体实例,可以看出:

(1)本体 1 和本体 2 之间存在结构上的差异:本体 1 是平板式的,本体 2 是层次的;

(2)本体 1 和本体 2 之间存在语义上的差异:例如本体 1 中的 city 与本体 2 中的 City。

通过上面介绍的方法,能够很好解决上述结构和语义上的差异。

4 实验

为了验证这个方法的可行性,笔者做了一个小实验,因为不是演示代理机的工作原理,所以就没有加入代理机的交互部分。实验中建立了与实验相关的部分共用本体(当前只是为了演示本方法的可行性)。实验数据来自某银行总行和分行的数据字典。总行数据字典和分行的数据字典在结构和语义上有很多区别,在总行数据要下发到分行,或者分行的数据要汇总到总

行时会遇到先前提到的问题——交互问题。

用文中介绍的方法解决此问题:

首先,建立共用本体;

其次,根据两个数据库的表结构分别建立各行对应的本体;

再次,用文中的方法进行映射。

结果显示,本方法能够实现总行和分行数据之间的交换。

5 结束语

介绍了一种建立在代理机基础上的本体映射的方法,它可以推广到其他本体映射上。两个代理机之间就可以通过文中所介绍的方法进行交互。此方法只需要在建立本体时参考共用本体。而这种共用本体库的建立,可以借鉴 Java 的类库的建立。本方法尝试着从其他技术中找出可以利用的方面,通过本体的方法做映射。结果显示,本方法是有效的。

参考文献:

- [1] van Eijk R M, de Boer F S, van der Hoek W. On dynamically generated translators in agent communication[J]. International Journal of Intelligent Systems, 2001, 16: 587 - 607.
- [2] Gruber T R. Toward principles for the design of ontologies used for knowledge sharing[J]. International Journal of Human-Computer Studies, 1995, 43: 907 - 928.
- [3] Weinstein P C, Birmingham W P. Agent communication with differentiated ontologies: eight new measures of description compatibility[R]. Technical Report CSETR - 383 - 99 Michigan: Artificial Intelligence Laboratory, University of Michigan, 1999.
- [4] Sheth A P, Larson J A. Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Computing Surveys, 1990, 22: 183 - 236.
- [5] Wiesman F, Roos N, Vogt P. Automatic ontology mapping for agent communication[J]. International Journal of Intelligent Systems, 2002, 17: 235 - 251.
- [6] Steels L, Vogt P. Grounding adaptive language games in robotic agents[C]//In Husband C, Harvey I. Proceedings of the Fourth European Conference on Artificial Life. Cambridge, MA: MIT Press, 1997.
- [7] Steels L. Emergent adaptive lexicons[C]//In Maes P. From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulating Adaptive Behavior. Cambridge, MA: MIT Press, 1996.
- [8] 李善平,尹奇,胡玉杰,等.本体论研究综述[J].计算机研究与发展, 2004, 41: 24 - 25.
- [9] Hammer J, McLeod D. An approach to resolving semantic

~16Hz, k3b 和 k6b 则在 8~12Hz。但是所有的受试者在使用整个 8~24Hz 频带的信号时的分类正确率普遍都比较低。因此应该考虑对于不同的受试者选取不同的特征频带信号。

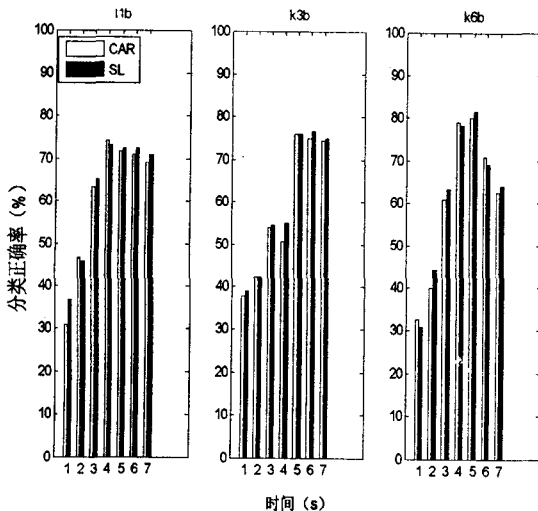


图 3 做 SL 滤波和 CAR 滤波的分类正确率比较

表 1 三名受试者在进行运动想像阶段(3s 后)选取不同频带信号获得的分类正确率

		3~4(s)	4~5(s)	5~6(s)	6~7(s)
l1b	8~12(Hz)	73.33	72.50	72.50	70.83
	12~16(Hz)	75.00	82.50	80.83	73.33
	16~24(Hz)	65.00	62.50	79.17	79.17
	8~24(Hz)	67.00	80.83	70.00	64.17
k3b	8~12(Hz)	55.00	76.11	76.67	75.00
	12~16(Hz)	60.00	60.56	63.33	66.11
	16~24(Hz)	65.00	67.78	68.33	68.89
	8~24(Hz)	63.33	64.44	58.33	64.44
k6b	8~12(Hz)	78.33	81.67	69.17	64.17
	12~16(Hz)	70.83	77.50	75.00	75.83
	16~24(Hz)	79.17	52.50	74.17	74.17
	8~24(Hz)	75.00	75.83	70.83	71.67

4 结束语

通过对实验的研究结果表明:大量的电极不一定带来好的分类效果,应尽量利用电极之间的先验知识。选取电极 C3, C4, Cz 周围的部分电极上获得脑电信号,进行特征提取,可以大大减少计算量,也能得到较好的效果。至于具体的电极位置和数量的选取,还应进行进一步的研究,从而可以用最少的脑电数据获得

最好的分类效果。另外,出现提示信号(想象)后,真正开始想象运动前,分类正确率便开始上升,提示在准备运动阶段已经发生了幅值的变化,而在提示发出后,ERD/ERS 特征更明显。分类正确率得到大幅度提高,最高可达到 82.5%。总之,文中采取的算法对运动想象脑电信号进行特征提取和分类可以获得较好的分类正确率,可以将这些特征转化为某种控制信号处理控制光标的移动或辅助运动设备,就可以帮助严重瘫痪病人通过计算机与外界环境进行交流和沟通或者帮助他们实现自己想做的动作,因此在脑机接口的应用中有较高的实用价值。

参考文献:

- [1] Pfurtscheller G, Neuper C. Motor imagery activates primary sensorimotor area in humans[J]. Neuroscience Letters, 1997, 239:65-68.
- [2] Currana E A, Stokesa B M J. Learning to control brain activity: A review of the production and control of EEG components for driving brain - computer interface (BCI) systems[J]. Brain and cognition, 2003, 51(3):326-336.
- [3] Pfurtscheller G, Aranibar A. Event - related cortical desynchronization detected by power measurements of scalp EEG [J]. Electroencephalography and Clinical Neurophysiology, 1977, 42(8):817-826.
- [4] Pfurtscheller G, Neuper C. Motor Imagery and Direct Brain - Computer Communication[J]. Proceedings of IEEE, 2001, 89(7):1123-1134.
- [5] Peters B O, Pfurtscheller G, Flyvbjerg H. Automatic differentiation of multichannel EEG signals[J]. IEEE Trans BME, 2001, 48(1):111-116.
- [6] Domhege G, Blankertz B, Curio G, et al. Boosting bit rates in noninvasive EEG single - trial classification by feature combination and paradigm[J]. IEEE Tran. Biomed. Eng., 2004, 51:993-1002.
- [7] Fukunaga K. Introduction to statistical pattern recognition [M]. second ed. Boston: Academic Press, 1990.
- [8] Müller - Gerking J, Pfurtscheller G, Flyvbjerg H. Designing optimal spatial filters for single - trial EEG classification in a movement task[J]. Electroenc. Clin. Neurophys, 1999(5): 787-798.

(上接第 22 页)

heterogeneity in a federation of autonomous, heterogeneous database systems[J]. Journal for Intelligent and Cooperative Information Systems, 1993, 2(1):51-83.

- [10] Milo T, Zohar S. Using schema matching to simplify heterogeneous data translation [R]. Michigan: Artificial Intelligence

Laboratory, University of Michigan, 1998.

- [11] Wache H, Voegelé T, Visser U, et al. Ontology - based integration of information - a survey of existing approaches[R]. Michigan: Artificial Intelligence Laboratory, University of Michigan, 2001.