

基于概念主干 H 的知识联通及其联通测度^{*}

雷玉霞¹ 陆跃飞² 曹宝香¹ 王廷蔚¹(曲阜师范大学计算机学院 山东曲阜273165)¹ (中国科学院计算技术研究所 北京100080)²

Concept-Interconnectivity and Interconnection-Measure Based on Concept-Backbone

LEI Yu-Xia¹ SUI Yue-Fei² CAO Bao-Xiang¹ WANG Ting-Wei¹(Department of Computer Science, Qufu Teachers University, Qufu 273165)¹(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080)²

E-mail: yx-ww0188@sina.com.cn

Abstract Knowledge interconnection has proved to be more necessary in order to improve the degree of intelligence of computer. In this paper, we firstly introduce a method in taxonomy in the light of information entropy of the tree, and then concept-interconnectivity and interconnection analysis are mainly discussed. At last, we provide some interconnection-measures satisfying some intuition properties.

Keywords Information entropy, Taxonomy, Conceptual interconnectivity, Interconnection-measures

1 引言

领域专业知识在决策系统、智能教学系统(ITS)、知识查询系统等高技术产品中具有重要的作用。知识库是机器智能的重要来源,而大量实践证明计算机真正具有智能不仅需要大型知识库,而且更需要知识间广泛的互连^[1-3]。

在国外, Richardson 等研究人员创建的 MindNet 探讨了词汇之间的语义联通(Semantic Connection)。在 MindNet 中,词汇之间的联通方法就是通过充分约束的、扩展的语义关系路径(Semantic Relation Paths)而实现的,并且这种语义关系路径已经证明,词跟词之间的关系不被连接起来,就没有意义^[4,5]。

类似于 MindNet 的语义联通,我们以提高知识服务的质量为背景提高了概念联通。所谓概念联通是指通过关系链或属性或属性值将两个概念联系起来的过程。关系链是指形如“ $C_1, R_1, C_2, R_2, \dots, R_{n-1}, C_n$ ”的序列,其中 $C_1, \dots, C_n$ 表示概念, $R_1, \dots, R_{n-1}$ 表示关系。陈述的联通就是通过陈述的泛化或特化共同得到的有意义的陈述。本文首先讨论基于概念主干的概念联通以及联通测度,然后进一步讨论基于概念主干的陈述联通。

2 基于概念主干 H 的概念联通以及联通测度

2.1 概念主干 H

为了讨论基于固定概念分类的概念联通,我们首先介绍一下我们的概念分类方法。对于某个领域概念有多种分类形式。我们称由 Is-a(记作 $<$)关系和 Instance-of 关系(见表1)组成的概念分类称为概念主干 H 。概念主干上的概念集合 $C = GC \cup IC$, 其中 GC 表示不是个体的概念的集合, IC 表示个体的集合。

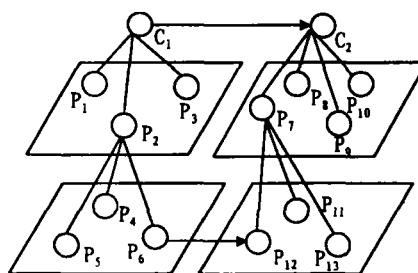
给定两个概念 C_1 和 C_2 , 如果 $C_1 < C_2$, 则 C_1 Is-a C_2 , 或者 C_1 是 C_2 的实例(Instance), 或者 C_1 是 C_2 的子概念(Sub-

concept), 或 C_2 是 C_1 的超概念(super-concept)。每个陈述 $S \in S$ 的形式为: $S = CRC'$, 其中 $C, C' \in C, R$ 是一个二元关系。其中 S 是陈述的集合。

注意, 概念主干上的每个节点都对应一个部分树(见图1)。这里的部分关系是指概念上的部分关系, 即一个概念是另一个概念的概念上的部分。例如说发动机是舰艇的部分, 是指对任意一个舰艇的个体都存在一个发动机的个体, 使得这个发动机是这个舰艇的部分。另外, 不同的部分树的节点之间也可能具有 Is-a 关系。

表1 H 中的语义关系

语义类型	含义
Is-a	表示概念之间的上下位关系
Instance-of	表示个体和概念之间的语义关系



→表示 is-a 关系, — 表示部分关系

图1 H 中节点相应的部分树(部分)

图1主要说明两点:一是在 H 上的每个节点对应一个部分树。例如 P_1, P_2 和 P_3 是概念 C_1 的部分, P_6 又是 P_2 的部分, 包含 P_1, P_2 和 P_3 的方框表示 P_1, P_2 和 P_3 在树的同一层上;二是下位概念的部分可能是上位概念的部分的子概念。例如 P_6 是 P_{12} 的子概念, 其中 P_6 是 C_1 的部分, P_{12} 是 C_2 的部分, C_1 是 C_2 的子概念。

2.2 基于概念主干 H 的概念联通

^{*} 本文思想形成于中科院计算所大规模知识处理课题组。雷玉霞 硕士, 陆跃飞 博士导师, 研究员, 曹宝香 硕士导师, 教授, 王廷蔚 硕士生。

首先,给出几个定义,然后讨论概念联通的基本性质。

定义2.1 设 C_1 和 C_2 是 H 上的任意两个概念(不是 T)。则:1)若 $\exists C \in H (C_1 < C \wedge C_2 < C)$, 则称 C_1 和 C_2 是上位联通的。记为 $\text{hei}(C_1, C_2, C)$; 2)若 $\exists C \in H (C < C_1 \wedge C < C_2)$, 则称 C_1 和 C_2 是下位联通的。记为 $\text{hoi}(C_1, C_2, C)$; 3)若存在概念 $X_1 (=C_1), X_2, \dots, X_n (=C_2), Y_1, \dots, Y_{n-1}$, 使得 $\text{hxi}(X_1, X_2, Y_1), \text{hxi}(X_2, X_3, Y_2), \dots, \text{hxi}(X_{n-1}, X_n, Y_{n-1})$, 则称 C_1 和 C_2 是混合位联通的。记为 $\text{hhi}(C_1, C_2)$ 。其中, hxi 可以是 hei 或 hoi 。

我们把上述联通统称为结构可联通的,记为 $\text{itc}(C_1, C_2)$, 表示 C_1 和 C_2 是结构上可联通的。下面给出一些基本的性质:

命题2.1 “hei”, “hoi”和“hhi”作为关系是对称的,非自反的;“itc”作为关系是对称的、自反的;“是实例”作为联通关系是非自反的、非对称的和反传递的。

命题2.2 结构联通具有如下基本性质:

- 1) $\text{hei}(C_1, C_2, C) \rightarrow \text{hhi}(C_1, C_2)$, 和 $\text{hoi}(C_1, C_2, C) \rightarrow \text{hhi}(C_1, C_2)$;
- 2) $\text{hei}(C_1, C_2, C) \wedge \text{hei}(C_2, C_3, C') \rightarrow \text{hei}(C_1, C_3, \text{lub}(C, C'))$;
- 3) $\text{hei}(C_1, C_2, C) \wedge \text{hei}(C_2, C_3, C') \rightarrow \text{hoi}(C, C', C_2)$;
- 4) $\text{hei}(C_1, C_2, C) \wedge \text{hei}(C_2, C_3, C') \rightarrow \text{hhi}(C_1, C_3)$;
- 5) $\text{hoi}(C_1, C_2, C) \wedge \text{hoi}(C_2, C_3, C') \rightarrow \text{hei}(C, C', C_2)$;
- 6) $\text{hoi}(C_1, C_2, C) \wedge \text{hoi}(C_2, C_3, C') \rightarrow \text{hhi}(C_1, C_3)$;
- 7) $\text{hei}(C_1, C_2, C) \wedge \text{hoi}(C_2, C_3, C') \rightarrow \text{hhi}(C_1, C_3)$;
- 8) $\text{instance}(C_1, C_3) \wedge \text{instance}(C_2, C_3) \rightarrow \text{hei}(C_1, C_2, C_3)$;
- 9) $\text{instance}(C_1, C_2) \wedge \text{instance}(C_1, C_3) \rightarrow \text{hoi}(C_2, C_3, C_1)$;
- 10) $\text{itc}(C_1, C_2) \rightarrow \forall (C_3 (\text{instance}(C_3, C_1) \rightarrow \text{itc}(C_3, C_2)))$;
- 11) $\text{itc}(C_1, C_2) \rightarrow \forall C_3 (\text{instance}(C_3, C_2) \rightarrow \text{itc}(C_1, C_3))$ 。

2.3 H 上的联通测度

因为 Is-a 具有传递性,我们规定在一条同向的 Is-a 链上的任两个概念的联通度都相等。另外,我们认为联通测度应该满足下面一些直观性质:

- 1)两个概念的极小上界(极大下界)越多,其联通强度就越大;
- 2)同时是上位联通和下位联通的两个概念的联通强度要大于路径相同的仅是上位联通(下位联通)的联通强度;
- 3)若两对概念的联通路经关于某个节点是对称的,则它们各自的联通强度应该相同。在图2中,战斗舰艇与消防艇的联通强度应相同于勤务舰船与航空母舰的;
- 4)若两对概念的路径距离相同,则概念层次差别小的联通强度应大于层次差别大的。在图2中,潜艇与侦察船的联通度要大于航空母舰与勤务舰船的。

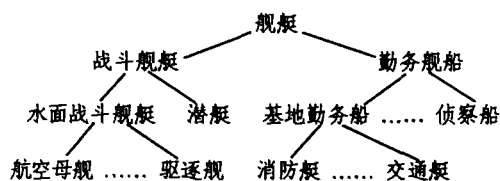


图2 《中国大百科全书》军事卷关于舰艇的分类(部分)

下面先给出两个重要定义,然后具体给出上联通测度、下联通测度和混合联通测度:

定义2.2 称概念 C 是 C_1 和 C_2 的极小上界,如果 C_1 和 C_2 同时满足: $C_1 < C, C_2 < C; \forall D (C_1 < D \wedge C_2 < D \rightarrow C < D)$ 。记作 $\text{lub}(C_1, C_2)$ 。

定义2.3 称概念 C 是 C_1 和 C_2 的极大下界,如果 C_1 和 C_2 同时满足: $C < C_1, C < C_2$;

$\forall D (D < C_1 \wedge D < C_2 \rightarrow D < C)$ 。记作 $\text{glb}(C_1, C_2)$ 。

命题2.3 概念的极小上界和极大下界具有如下基本性质: 1) $C_1 < C_2 \rightarrow \text{lub}(C_1, C_2) = C_2$; 2) $C_1 < C_2 \rightarrow \text{glb}(C_1, C_2) = C_1$ 。

设 $L(C)$ 表示概念 C 在 H 中的层数; $\text{Lub}(C, C')$ 和 $\text{Glb}(C, C')$ 分别表示 C 和 C' 的极小上界集和极大下界集。定义概念 C 和 C' 之间的上联通度如下:

$$D^+(C, C') = \sum_{C'' \in \text{Lub}(C, C')} \left\{ \frac{1}{|L(C) - L(C'')|} + \frac{1}{|L(C') - L(C'')|} + \frac{1}{|L(C) - L(C')| + 1} \right\} \quad (1)$$

注意,当 C 和 C' 的最小上界是 C 或 C' 时,我们规定 $D^+(C, C') = \infty$ 。当两个概念有公共的子节点时,即使上联通度很小,其直观上的联通度也同样是较大的。为了反映这种情况,定义下联通度如下:

$$D^-(C, C') = \sum_{C'' \in \text{Glb}(C, C')} \left\{ \frac{1}{|L(C) - L(C'')|} + \frac{1}{|L(C') - L(C'')|} \right\} \quad (2)$$

注意式(2)不包含 $C, C' \in \text{Glb}(C, C')$ 的情况,这种情况用式(1)。上面分别刻画的上、下联通测度,综合上两式我们给出一个联通测度如下:

$$D(C, C') = D^+(C, C') + D^-(C, C') \quad (3)$$

式(3)的直观意义是:两个概念的联通测度值等于上联通度和下联通度之和,且值越大说明其联通强度越大。例如, D^+ (战斗舰艇,勤务舰船) = 3; D^+ (战斗舰艇,消防艇) = D^+ (勤务舰船,航空母舰) = 5/3。 D^+ (潜艇,侦察船) = 2 > D^+ (航空母舰,勤务舰船) = 5/3。

命题2.4 式(3)满足上述1)~4)的直观性质。

3 基于概念主干 H 的陈述联通

为了便于讨论,我们假设 R 是左右区间封闭的。一个关系 R 是左区间封闭的,是指对于任何概念 $C, D, E, F \in C$, 如果 $(C < D < E) \wedge (C, F) \in R \wedge (E, F) \in R$, 则 $(D, F) \in R$ 。类似地,一个关系是右区间封闭的,是指对于任何概念 $C, D, E, F \in C$, 如果 $(E < D < F) \wedge (C, E) \in R \wedge (C, F) \in R$, 则有 $(C, D) \in R$ 。显然,有下面的命题:

命题3.1 若二元关系 R_1 和 R_2 是左(右)区间封闭的,则 $R = R_1 \cap R_2$ 也是左(右)区间封闭的。

给定两个陈述 $S = C_1 R_1 C_2$, 和 $S' = C_1' R_2 C_2'$, 下面讨论两个陈述之间的联通。对于 $R_1 = R_2$ 的讨论情况,请参阅文[1]。下面,我们讨论 $R_1 \neq R_2$ 的情况。先令 $R = R_1 \cap R_2$ 。首先,考虑简单的情况,即 $C_2 = C_2'$ 。想找到联通 T , 我们首先要找到 C_1 和 C_1' 的 $\text{Lub}(C_1, C_1')$, 记作 C 。如果 $(C, C_2) \in R$, 则令 $T = C R C_2$ 。其次,我们找到 C_1 和 C_1' 的 $\text{Glb}(C_1, C_1')$, 记作 D 。如果 $(D, C_2) \in R$, 则令 $T = D R C_2$ 。如果 C 和 D 都不存在,则 S 和 S' 之间不能进行有意义的联通。

现在,我们考虑 $C_2 \neq C_2'$ 的情况。因为 $\text{Lub}(C_1, C_1')$ 是一定存在的,所以我们先研究 $\text{Lub}(C_1, C_1')$ 的情况。我们定义一个概念:关系 R 是和-封闭的,如果对于任何 $C, D, E \in C$, 有

$$(C, E) \in R \wedge (D, E) \in R \rightarrow (\text{Lub}(C, D), E) \in R \quad (4)$$

或

(下转第179页)

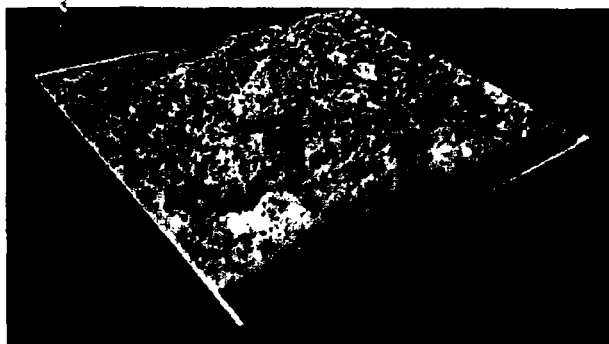


图8 原始图像的三维重建整体效果图

结论 本文通过采样、三角化、色彩控制的方法实现了灰度图像的三维重建。实际应用表明,该方法能有效地生成三维效果较好的图像。将图4与图6和图8进行比较,图6和图8更加

突出了土壤颗粒的空间关系。从图8中可以明显地看出,右上部整体高于左部,中部有一个比较深的凹陷。同时,通过三维处理的图像可以进一步放大,还可以对其进行旋转,工程技术人员对于原土壤图像的理解就能够更加深刻,更接近实际土壤空间状况,对于判断地基下沉依据将更符合实际。

参考文献

- 1 姜志国. 颗粒显微图像分析方法. 粉体技术, 1996, 2(2): 40~43
- 2 杨铁笙, 等. 图像处理技术用于泥沙颗粒分析的研究. 1999, 5(6): 13~19
- 3 Richard S, Wright Jr., Sweet M, 潇湘工作室译. OpenGL 超级宝典. 人民邮电出版社, 2001, 6
- 4 Woo M, Davis J, Shreiner D 著. The Official Guide to Learning OpenGL, Version 1.2 Publisher: Addison Wesley Professional, 1999, 8

(上接第173页)

$$(C, D) \in R \wedge (C, E) \in R \rightarrow (C, \text{Lub}(D, E)) \in R \quad (5)$$

如果 C, D, E 满足(4), 则我们记作 $C, D; E$ 。如果 C, D, E 满足(5), 则我们记作 $C; D, E$ 。

命题3.2 若二元关系 R_1 和 R_2 是和-封闭的, 则 $R = R_1 \cap R_2$ 是和-封闭的。

命题3.3 假设在 $\langle C \rangle$ 构成一个概念树。对于任意给定的两个陈述 $S = C_1 R_1 C_2$ 和 $S' = C_1' R_2 C_2'$ 。如果 R_1 和 R_2 是和-封闭的, $C_1 R_1 C_2 \in S$ 和 $C_1' R_2 C_2' \in S$, $S_1 = C_1' R_1 C_2 \in S$, $S_2 = C_1 R_2 C_2' \in S$, 则 S 和 S' 存在一个联通 T 。其中, $R = R_1 \cap R_2$ 。

证明: 由命题3.2可知 R 是和-封闭的。因此 S 的特化 $C_1 R_1 C_2$ 和 $S_1 = C_1' R_1 C_2$ 有一个共同的泛化, 例如 $T_1 = \text{Lub}(C_1, C_1') R_1 C_2$, 同理 $S_2 = C_1 R_2 C_2'$ 和 S' 的特化 $C_1' R_2 C_2'$ 有一个共同的泛化, 例如 $T_2 = \text{Lub}(C_1, C_1') R_2 C_2'$, 进而可知 T_1 和 T_2 存在一个共同的泛化, 例如 $\text{Lub}(C_1, C_1') R \text{Lub}(C_2, C_2')$ 。设 $T = \text{CRD}$, 则 T 是 S 和 S' 的一个联通。

这个命题表明给定两个陈述 $S = C_1 R_1 C_2$ 和 $S' = C_1' R_2 C_2'$ 。如果 R_1 和 R_2 是和-封闭的, 并且 $C_1, C_1'; C_2, C_1, C_1'; C_2'$ 和 $\text{Lub}(C_1, C_1'); C_2, C_2'$, 则 S 和 S' 一定存在有意义的联通; 否则不一定存在联通。

再考虑更特殊的情况, 即如果 $\text{Glb}(C_1, C_1')$ 和 $\text{Glb}(C_2, C_2')$ 是存在的。为了讨论这种情况, 我们定义: 一个关系 R 是交-封闭的, 即如果对于任何 $C, D, E \in C$, 有

$$(C, E) \in R \wedge (D, E) \in R \rightarrow (\text{Glb}(C, D), E) \in R \quad (6)$$

或

$$(C, D) \in R \wedge (C, E) \in R \rightarrow (C, \text{Glb}(D, E)) \in R \quad (7)$$

如果 C, D, E 满足(6), 则我们记作 $C, D; E$ 。如果 C, D, E 满足(7), 则我们记作 $C; D, E$ 。对于这种特殊的情况, 也有类似命题:

命题3.4 若二元关系 R_1 和 R_2 是交-封闭的, 则 $R = R_1 \cap R_2$ 是交-封闭的。

命题3.5 假设在 $\langle C \rangle$ 构成一个层状概念图。对于给定的两个陈述 $S = C_1 R_1 C_2$ 和 $S' = C_1' R_2 C_2'$ 。如果 R_1 和 R_2 是和-封闭的, $C_1 R_1 C_2 \in S$ 和 $C_1' R_2 C_2' \in S$, $S_1 = C_1' R_1 C_2 \in S$, $S_2 = C_1 R_2 C_2' \in S$, $\text{Glb}(C_1, C_1')$ 和 $\text{Glb}(C_2, C_2')$ 是存在的, 则 S 和 S' 存在一个联通 T 。

证明: 类似于命题3.2的证明。

这个命题表明给定两个陈述 $S = C_1 R_1 C_2$ 和 $S' = C_1'$

$R_2 C_2'$ 。如果 R_1 和 R_2 是和-封闭的, $C_1 R_1 C_2 \in S$ 和 $C_1' R_2 C_2' \in S$, 并且 $C_1, C_1'; C_2, C_1, C_1'; C_2'$, $\text{Glb}(C_1, C_1')$ 和 $\text{Glb}(C_2, C_2')$ 是存在的, 另外 $\text{Glb}(C_1, C_1'); C_2, C_2'$, 则 S 和 S' 一定存在联通; 否则不一定存在联通。

现在, 令 $R = R_1 \cup R_2$ 的。有下面的类似命题:

命题3.6 假设在 $\langle C \rangle$ 构成一个概念树。对于任意给定的两个陈述 $S = C_1 R_1 C_2$ 和 $S' = C_1' R_2 C_2'$ 。如果 R 是和-封闭的, $S_1 = C_1' R_1 C_2 \in S$, $S_2 = C_1 R_2 C_2' \in S$, 则 S 和 S' 存在一个联通 T 。其中, $R = R_1 \cap R_2$ 。

命题3.7 假设在 $\langle C \rangle$ 构成一个层状概念图。对于给定的两个陈述 $S = C_1 R_1 C_2$ 和 $S' = C_1' R_2 C_2'$ 。如果 R 是交-封闭的, $S_1 = C_1' R_1 C_2 \in S$, $S_2 = C_1 R_2 C_2' \in S$, $\text{Glb}(C_1, C_1')$ 和 $\text{Glb}(C_2, C_2')$ 是存在的, 则 S 和 S' 存在一个联通 T 。

结束语 本文首先讨论基于概念主干 H 的概念联通以及联通分析; 然后, 给出了联通测度所应满足的直观性质, 并给出了上联通测度、下联通测度和混合联通测度, 这些测度满足所提出的直观性质。我们主要从理论上分析了基于概念主干的陈述联通, 给出了判定存在有意义联通的条件, 即命题3.2、3.5、3.6和命题3.7。本文仅讨论了基于概念主干 H 的概念联通以及联通测度, 并在此基础上进一步探讨了基于概念主干的陈述联通, 没有探讨基于部分关系的联通分析。下一步, 将继续讨论这类联通以及联通测度。

致谢 中科院计算机大规模知识处理课题组的曹存根研究员首先提出知识联通, 并给于作者有益的指导。在此表示感谢!

参考文献

- 1 雷玉霞, 陆跃飞. 基于 Concept-Relation 模型的知识联通. 计算机科学, 2002, 29(11): 53~58
- 2 Lei Yuxia, Cao Cungen, Sui Yuefei. Based-CR Knowledge Interconnection in National Knowledge Infrastructure. ICII'2001, C: 134~139
- 3 Gallant S I. Connectionist expert systems. Communications of the ACM, 1988, 31: 152~169
- 4 Richardson S D, Dolan W B, Vanderwende L. MindNet: acquiring and structuring semantic information from text. Microsoft Research Technical Publications, 1998. 网址: ftp://ftp.research.microsoft.com/pub/tr/tr-98-23.doc
- 5 Richardson S. Determining similarity and inferring relations in a lexical knowledge base: [PhD. dissertation]. City University of New York, 1997