

一种改进的领域本体语义相似度计算方法

曹 叡¹, 吴玲达^{1,2}

(1 装备学院 复杂电子系统仿真重点实验室, 北京 101416; 2 国防科技大学 信息系统与管理学院, 湖南 长沙 410073)

摘 要: 在深入研究语义相似度计算方法的基础上, 提出了一种改进的领域本体概念语义相似度算法。算法考虑了概念间层次顺序、节点深度、语义重合度对基本语义相似度的影响, 描述了决定概念语义相似度的各种因素, 分析并建立了这些因素的数学模型。通过实验对比与分析, 验证了此算法具有较好的精确性和准确性。

关键词: 领域本体; 语义相似度; 语义距离; 信息量; 概念属性; 兼容度

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2014)08-0109-06

An Improved Semantic Similarity Algorithm Based on Domain Ontology

CAO Rui¹, WU Ling-da^{1,2}

(1 Science and Technology on Complex Electronic System Simulation Laboratory, Academy of Equipment, Beijing 101416, China; 2 The Academy of Information System and Management, National University of Defense Technology, Changsha 410073, China)

Abstract: Based on the deep research of semantic similarity algorithm, an improved semantic similarity algorithm based on domain ontology is proposed. Hierarchical sequential, depth information and semantic coincidence degree between concepts are considered, which affect the basic semantic similarity. Semantic similarity determinant factors are described and analyzed; the factors' mathematical models are established. Experiment comparison and analysis verify that, the precision and accuracy of proposed semantic similarity algorithm are good.

Key words: domain ontology; semantic similarity; semantic distance; information content; concept attribute; compatibility

1 引言

作为新一代互连网络, 语义网通过对现有万维网的扩展, 为人与机器之间更好的交互与合作提供了新的前景, 也带来了新的技术挑战^[1]。本体论作为支撑语义网的三大核心技术之一^[2], 能够将领域中各种概念和关系有效地表达, 因此应用本体论知识解决概念间语义相似度计算的问题已经成为当今研究的热点。目前主要有基于语义距离、基于信息论、基于概念属性三种语义相似度计算模型^[3-5]。随着研究的逐渐深入, 黄果, 等^[5]将概念信息量与属性作为语义距离计算的决策因子, 提供了一种量化语义关系的手段; 赵永金, 等^[6]通过分析节点密度与概念

属性对语义相似度的影响, 对文献[2]的算法进行了扩展; 张兰芳^[7]在此基础上, 考虑概念间的语义距离、属性、层次顺序等因素, 设计了一种语义相似度算法, 但却忽略了概念间信息量对语义相似度的影响。

本文在文献[7]的语义相似度计算思想基础上, 进一步考虑概念信息量、节点深度、语义重合度等因素对语义相似度的影响, 提出了一种改进的领域本体概念语义相似度计算方法。该方法综合了目前三种主要的语义相似度计算模型, 补充完善了语义相似度计算需要考虑的因素, 使概念语义相似度计算方法更加合理。

收稿日期: 2013-10-26; 修回日期: 2013-12-27

基金项目: 国家自然科学基金项目(61202129)

2 语义相似度计算模型

2.1 基于语义距离的计算模型

这种模型的基本思想是利用两个概念节点在领域本体层次网络图中的最短路径来衡量概念间的语义相似度^[3,5-7]. 语义距离不仅要计算最短路径中边的数量,还要考虑到有向边的差异;因此,一般先确定有向边的权重,再计算有向边的长度,最终得到概念间的语义距离.

语义相似度的数值区间通常是 $[0,1]$,当语义距离为 0 时,语义相似度为 1,可以认为两个概念的意思是相同的. 概念之间的语义距离越大,语义相似度越小,即二者是成反比关系,可以根据需要选用适当的转化模型^[2].

2.2 基于信息论的计算模型

信息论认为一个概念 a 的信息量 $I(a) = -\log P(a)$, $P(a)$ 为概念 a 在领域本体 G 中出现的频率,有

$$P(a) = \frac{n(a) + \sum_{P(a,a')} n(a')}{n(G)} \quad (1)$$

式中, $n(a)$ 为概念 a 在该领域本体中出现的次数; $P(a,a')$ 表示 a 所有子概念 a' 组成的集合; $n(G)$ 为该领域本体所有概念的数量. 基于信息论的计算模型基本思想是通过领域本体中两个概念节点共享的信息量来衡量概念间的语义相似度^[4].

2.3 基于概念属性的计算模型

在领域本体中,属性是对概念详尽的描述和补充,说明了概念的重要性质^[7]. 由于在现实生活中,人们一般通过比较事物之间相同属性的多少以判断两个事物是否相似,所以基于概念属性的语义相似度计算模型,其基本思想是通过领域本体中两个概念的属性差异来衡量概念间的语义相似度. 其中一种经典的计算模型^[8] 是

$$\text{sim}(a, b) = \varphi f(A \cap B) - \lambda(A - B) - \mu(B - A) \quad (2)$$

式中, A, B 分别是领域本体中概念 a 和 b 的属性集合; f 是一个非负函数; φ, λ, μ 为权重, λ 是比较对象, μ 是参照对象,描述了人们对这两种对象的关注程度.

3 一种改进的语义相似度计算方法

上节前两种计算模型对语义相似度的认识具有对称性,但是许多心理学实验提出了反例,并建立了

非对称语义相似度计算模型^[2,7-8];本文在这些模型思想的基础上,考虑概念间层次、深度、重合度等因素的影响,提出了一种改进的领域本体概念语义相似度计算方法. 计算分为四个部分:基本语义相似度、层次顺序、节点深度、语义重合度.

3.1 基本语义相似度

由基于语义距离、基于信息论和基于概念属性三种计算模型,可以得到概念间的距离、信息、属性相似度,它们构成了概念间的基本语义相似度.

3.1.1 概念距离相似度

在领域本体中,各概念之间有向边的意义与作用存在差别,所以基于语义距离的计算模型首先需要确定有向边的权重. 决定权重的主要因素有:

(1) 概念间的关系

本体概念之间存在多种关系,其中同义关系、继承关系、整体部分关系占很大的比例,一般在领域本体层次网络图中主要考虑这三种关系类型^[7]. 同义关系说明两个概念含义是相同的;继承关系说明概念之间是包含关系,与面向对象中子类与父类的关系类似,子概念包含父概念的信息,也包含自身特有的信息;整体部分关系说明概念之间是整体和组成部分的关系. 若 $\text{weight}_t(c, d)$ 表示概念 c 与 d 之间的关系类型对有向边权重的影响,依据文献^[7],当概念间是同义关系时, $\text{weight}_t(c, d)$ 值为 1;当概念间是继承关系时, $\text{weight}_t(c, d)$ 值为 1/2;当概念间是整体部分关系时, $\text{weight}_t(c, d)$ 值为 1/3.

(2) 有向边的层次

在层次网络图中,唯一的入口节点对应领域本体的根概念,第二层节点是对根概念的细化,随着各层节点的逐渐划分,节点对应概念的含义越来越具体,概念间有向边的权重也越来越大. 若 $\text{weight}_v(c, d)$ 表示概念 c 与 d 之间有向边层次对权重的影响,有

$$\text{weight}_v(c, d) = \sum_{i=1}^{\text{level}} \frac{1}{2^i} \quad i \in \mathbb{Z}^+ \quad (3)$$

以上公式能够较好反映出有向边所在层次 level 对该有向边权重的影响程度:在领域本体层次网络图中,概念间有向边所在层次越高,则该有向边的权重越小.

(3) 节点的密度

领域本体中各概念的细化程度可能存在差别,与其相对应的层次网络图中,各节点的密度存在差别. 若某一局部节点的密度较大,则说明在领域本体中,该节点所对应的概念被描述的越详细、分类越具

体、细化程度越大,以该节点作为起始点的有向边权重也越大。

定义 1 对于领域本体层次网络图 $G = \langle V, E \rangle$, 其中 V 是节点集合, E 是边的集合; $\deg^+(v)$ 为节点 v 的出度, $\Delta^+(G) = \max\{\deg^+(v) \mid v \in V\}$ 为图 $G = \langle V, E \rangle$ 最大出度。

若 $weight_d(c, d)$ 表示概念 c 与 d 之间节点密度对有向边权重的影响;不失一般性,若概念 c 为有向边起始点,则认为节点密度对权重的影响为

$$weight_d(c, d) = \frac{\deg^+(c)}{\Delta^+(G)} \quad (4)$$

若 $weight(c, d)$ 表示概念 c 与 d 之间有向边的权重,综合以上因素可以得到有向边权重为

$$weight(c, d) = \alpha weight_t(c, d) + \beta weight_v(c, d) + \gamma weight_d(c, d) \quad (5)$$

式中, α, β, γ 是取值区间为 $[0, 1]$ 的因子权重,且 $\alpha + \beta + \gamma = 1$ 。确定有向边的权重之后,参考文献[5]的方法,能够由有向边的权重进一步得到有向边的长度以及概念间的语义距离;从而对于领域本体中的两个概念 a, b , 概念距离相似度为

$$sim(a, b)_s = \frac{\theta}{L_s(a, b)^2 + \theta} \quad (6)$$

式中, $L_s(a, b)$ 代表概念 a 到 b 的语义距离, θ 为大于 0 的调节因子[5]。

3.1.2 概念信息相似度

基于信息论的计算模型认为领域本体中概念间共享的信息量越多则相似度越大。在领域本体层次网络图中,节点在细化其上层节点的同时,其对应的概念也得到了祖先的信息,所以可以认为概念间共同的祖先节点是它们之间共享信息的一种表现。这里采用适用于同一领域本体的相似度计算模型,利用概念间最近共同祖先节点的信息量与概念信息量的比值来衡量概念信息相似度[4]。对于领域本体中的两个概念 a, b , 概念信息相似度为

$$sim(a, b)_I = \frac{2 \times I(e)}{I(a) + I(b)} \quad (7)$$

式中, e 为领域本体中概念 a 与 b 的最近共同祖先节点, $I(e), I(a), I(b)$ 为概念 a, b, e 的信息量。

3.1.3 概念属性相似度

领域本体一方面准确定义了领域中的各个概念,另一方面也详尽描述了概念的属性以及概念之间的各种关系,基于概念属性的计算模型以概念属性集合之间的相近程度作为判断概念属性相似度的标准。

定义 2 在领域本体中,概念之间的属性距离是指这两个概念所有的属性差异的大小。

参考基于多维属性的相似度计算模型,对于领域本体中的两个概念 a, b , 若它们的属性集合 $A(a), A(b)$ 分别为 $A(a) = \{A_0, A_1, A_2, \dots, A_m\}$ 和 $A(b) = \{B_0, B_1, B_2, \dots, B_n\}$ 则他们之间的属性距离 $L_A(a, b)$ 为

$$L_A(a, b) = \left\{ \sum_{i=0}^{\min(m, n)} \varphi(A_i, B_i) + |m - n| \right\}^{\frac{1}{2}} \quad (8)$$

$$\varphi(A_i, B_i) = \begin{cases} 0, & \text{tag}(A_i) = \text{tag}(B_i) \& \\ & \text{value}(A_i) = \text{value}(B_i); \\ 1, & \text{tag}(A_i) = \text{tag}(B_i) \& \\ & \text{value}(A_i) \neq \text{value}(B_i); \\ 1, & \text{tag}(A_i) \neq \text{tag}(B_i); \end{cases}$$

式中, i 为自然数, $\text{tag}(A_i), \text{tag}(B_i)$ 和 $\text{value}(A_i), \text{value}(B_i)$ 分别代表属性 A_i, B_i 的名称和值。对于领域本体中的两个概念 a, b , 概念属性相似度为

$$sim(a, b)_A = \frac{\tau}{L_A(a, b)^2 + \tau} \quad (9)$$

式中, τ 为大于 0 的调节因子。综合概念距离相似度、概念信息相似度、概念属性相似度,对于领域本体中的两个概念 a, b , 可以得到基本语义相似度为

$$sim_{Bsc}(a, b) = \delta sim_s(a, b) + \epsilon sim_I(a, b) + \eta sim_A(a, b) \quad (10)$$

式中, δ, ϵ, η 是取值区间为 $[0, 1]$ 的权重,且 $\delta + \epsilon + \eta = 1$ 。另外,还需要考虑到领域本体的完善程度,若领域本体中未对某概念的属性进行描述,则有 $\eta = 0$, 且 $\delta + \epsilon = 1$ 。

3.2 层次顺序

非对称语义相似度计算模型,能够表达人们对比较对象的关注度,认为相似度具有方向性[2,7-8]。如文献[7]中的简单农作物本体,人们对“小麦”与“麦类”相似的认知,和“麦类”与“小麦”相似的认知是不同的, $sim(\text{小麦}, \text{麦类})$ 相当于弹性匹配中的精确匹配[9],即用户以“麦类”为参照对象[8]时,作为比较对象“小麦”仅能满足部分需求; $sim(\text{麦类}, \text{小麦})$ 相当于弹性匹配中的包含匹配,即用户以“小麦”为参照对象时,作为比较对象“麦类”包含关于参照对象“小麦”的信息。因此,人们对 $sim(\text{小麦}, \text{麦类})$ 大于 $sim(\text{麦类}, \text{小麦})$ 的认知,符合了非几何模型[8]所认为的语义相似度具有方向性的观点。

在领域本体层次网络图中,对于非同义关系的

两个概念,节点间的层次顺序是影响语义相似度的直接因素.若两个节点不变,处于高层的节点作为比较对象,处于低层的节点作为参照对象,它们之间的相似度记为 $sim(\text{高层}, \text{低层})$;处于低层的节点作为比较对象,处于高层的节点作为参照对象,它们之间的相似度记为 $sim(\text{低层}, \text{高层})$,一般有 $sim(\text{高层}, \text{低层}) < sim(\text{低层}, \text{高层})$;于是对领域本体中的两个概念 a, b ,可以得到它们之间的层次顺序对概念语义相似度的影响^[7]:

$$corr_L(a, b) = \frac{lyr(b) - lyr(a)}{lyr(G)} \quad (11)$$

式中, $lyr(a), lyr(b)$ 为概念 a, b 的所在的层次;入口节点所在层次最高,值为 1; $lyr(G)$ 为领域本体层次网络图的最底层.该影响因素的数值区间为 $[-1, 1]$,当概念节点 a 比 b 层次低时,式(11)能够加强概念间的语义相似度;当概念节点 a 比 b 层次高时,式(11)能够削弱概念间的语义相似度.

3.3 节点深度

在领域本体层次网络图中,若两个节点之间的路径长度不变,处于深层的两个节点,它们之间的概念相似度要大于处于低层节点之间的概念相似度;同样若节点之间的路径长度不变,节点所在层次越接近,它们之间的概念相似度越大.

某一节点到领域本体层次网络图入口节点的距离称为该节点的深度^[7],对于领域本体中的两个概念 a, b ,可以得到它们之间的深度信息对概念语义相似度的影响:

$$corr_D(a, b) = 1 - \frac{2|dep(a) - dep(b)|}{dep(a) + dep(b)} \quad (12)$$

式中, $dep(a), dep(b)$ 为概念 a, b 的深度信息.该影响因素的数值区间为 $[-1, 1]$,同时,当两个节点的深度总和增大时,式(12)能够加强概念间的语义相似度;当节点间的深度之差增大时,式(12)能够削弱概念间的语义相似度.

3.4 语义重合度

上位概念是指在领域本体层次网络图中,从概念节点到入口节点经过的所有祖先节点对应的概念集合.领域本体概念间上位概念的重合程度称为语义重合度^[10],它通过祖先节点的重合程度来反映概念间的相似程度.对于层次网络图中的两个概念节点 a, b ,可以通过它们共同祖先节点的个数来衡量语义重合度:

$$corr_H(a, b) = \frac{f_n(H_a \cap H_b)}{\max(f_n(H_a), f_n(H_b))} \quad (13)$$

式中, H_a, H_b 分别代表从节点 a, b 到入口节点经过

的所有节点组成的集合; f_n 为取集合个数函数.由于语义重合度属于概念之间相似程度的一种描述,所以与语义相似度的数值区间一样为 $[0, 1]$,语义重合度从一定程度上完善了语义相似度需要考虑的因素.

3.5 语义相似度计算

综合基本语义相似度、层次顺序、节点深度、语义重合度,对于领域本体中的两个概念 a, b ,语义相似度为

$$sim(a, b) = \sigma sim_{Bsc}(a, b) + \xi corr_L(a, b) + \zeta corr_D(a, b) + \varsigma corr_H(a, b) \quad (14)$$

式中, $\sigma, \xi, \zeta, \varsigma$ 为取值区间 $[0, 1]$ 的权重,且 $\sigma + \xi + \zeta + \varsigma = 1$.特别地,当概念 a, b 为同义关系,则有值为 0,且 $\sigma + \xi + \zeta = 1$.

4 实验结果与分析

领域本体概念语义相似度计算原型系统以 Java 语言编程实现.实验软件环境为:操作系统 Windows XP;开发平台 Eclipse;本体建模工具 Protégé 4.0;本体库接口 Protégé API.实验数据:为方便与原始算法进行对比,采用文献[7]中的简单农作物本体及其属性描述作为实验对象.

综合考虑基本语义相似度、层次顺序、节点深度、语义重合度四个因素在领域本体概念语义相似度计算中的地位 and 作用,权重的取值如下: $\sigma = 0.8$, $\varsigma = \xi = 0.05$, $\zeta = 0.1$.本实验认为概念间的距离、信息、属性相似度对基本语义相似度的贡献是相同的,有 $\delta = \epsilon = \eta = 1/3$;根据各因素对有向边的权重影响程度,有 $\alpha = 0.7, \beta = 0.2, \gamma = 0.1$;实验中调节因子 $\theta = 5, \tau = 1$.

在上述条件下,对实验数据进行概念语义相似度计算,并与两种原始经典算法得到的结果^[7]进行对比;实验结果如表 1 所示.

其中“原始 1”为赵永金等^[6]提出的概念语义相似度计算方法,“原始 2”为张兰芳^[7]的语义相似度改进算法.通过分析实验结果可知:

(1)本文算法具备“原始 2”算法在语义相似度计算中考虑的所有影响因素.例如,由于节点密度不同,导致“稻米”与“玉米”的相似度小于“燕麦”与“小麦”的概念语义相似度;有向边的层次不同,导致“萝卜”与“甜玉米”的相似度小于“萝卜”和“玉米”的相似度;由于本文算法为非对称语义相似度模型,所以得到的“玉米”与“甜玉米”相似度小于“甜玉米”与

“玉米”的相似度,以上例子都与文献[7]的实验分析一致。

表 1 农作物本体概念之间的语义相似度

序号	比较概念	参照概念	原始 1	原始 2	本文	专家
1	小麦	甜玉米	0.19	0.46	0.23	0.21
2	燕麦	小麦	0.40	0.57	0.55	0.55
3	燕麦	甜玉米	0.19	0.46	0.23	0.21
4	大麦	小麦	0.33	0.53	0.49	0.50
5	大麦	燕麦	0.33	0.53	0.49	0.50
6	麦	甜玉米	0.16	0.44	0.22	0.19
7	稻米	小麦	0.26	0.52	0.30	0.33
8	稻米	大麦	0.19	0.47	0.26	0.31
9	稻米	燕麦	0.26	0.52	0.30	0.33
10	稻米	甜玉米	0.27	0.49	0.38	0.36
11	玉米	小麦	0.21	0.50	0.29	0.31
12	玉米	大麦	0.17	0.47	0.28	0.30
13	玉米	燕麦	0.21	0.50	0.29	0.31
14	玉米	稻米	0.33	0.54	0.50	0.45
15	玉米	甜玉米	0.54	0.52	0.56	0.58
16	萝卜	小麦	0.08	0.41	0.14	0.10
17	萝卜	大麦	0.08	0.41	0.14	0.10
18	萝卜	燕麦	0.08	0.41	0.14	0.10
19	萝卜	稻米	0.08	0.41	0.13	0.13
20	萝卜	玉米	0.08	0.41	0.13	0.15
21	萝卜	甜玉米	0.08	0.40	0.11	0.12
22	甜玉米	小麦	0.19	0.51	0.25	0.23
23	甜玉米	大麦	0.16	0.49	0.24	0.21
24	甜玉米	燕麦	0.19	0.51	0.25	0.23
25	甜玉米	稻米	0.27	0.54	0.40	0.40
26	甜玉米	玉米	0.54	0.57	0.58	0.60
27	甜玉米	萝卜	0.08	0.45	0.13	0.14
28	甜玉米	糯玉米	0.37	0.58	0.57	0.58

(2)本文算法得到的语义相似度数值覆盖区间较大.对于实验中的简单农作物本体,本文算法得到的概念间语义相似度最大值为 $sim(\text{甜玉米}, \text{玉米}) = 0.58$,语义相似度最小值为 $sim(\text{萝卜}, \text{甜玉米}) = 0.11$,语义相似度数值覆盖区间为 $[0.11, 0.58]$;同理,可得“原始 1”算法的语义相似度数值覆盖区间

为 $[0.08, 0.54]$;“原始 2”算法的语义相似度数值覆盖区间为 $[0.41, 0.57]$. 本文算法与“原始 1”算法均具备较大的语义相似度数值覆盖区间,而在应用语义相似度解决实际问题时,大的数值覆盖区间会带来较高的精度.例如,在应用概念语义相似度进行语义扩展时,若比较概念与参照概念之间的语义相似度大于给定的提取阈值,则比较概念将作为参照概念的扩展概念;对于同样阈值,若语义相似度数值覆盖区间较小,则属于提取区间内的概念相对较多,导致了语义扩展精度的降低.

(3)对于相似的概念,本文算法得到的语义相似度计算结果数值较高.例如,虽然本文算法与“原始 1”算法的相似度数值覆盖区间差异不大,但是对于“小麦”与“燕麦”这组相似的概念,后者的计算结果数值较低,不太符合专家经验^[7],而本文算法能够得到较为准确的计算结果.为进一步分析本文算法的合理性,将本文算法与两种原始的语义相似度算法的结果与专家经验进行对比,如图 1 所示.这里专家经验值由 10 位相关领域的专家给出,并采用加权平均法根据专家重要度赋予相应权重得到最终结果,本实验得到的具体数值见表 1.

由图 1 可见与两种原始概念语义相似度算法相比,本文算法的计算结果曲线与专家经验差距较小,吻合程度较高;可以认为本文算法要优于两种原始算法,所得结果更加符合人类的主观判断.这里引入偏离度与兼容度^[6]验证该结论,若有 N 对概念,对第 i 对概念采用概念语义相似度算法计算的结果 Sim_i 与专家经验 Emp_i 的偏离度为

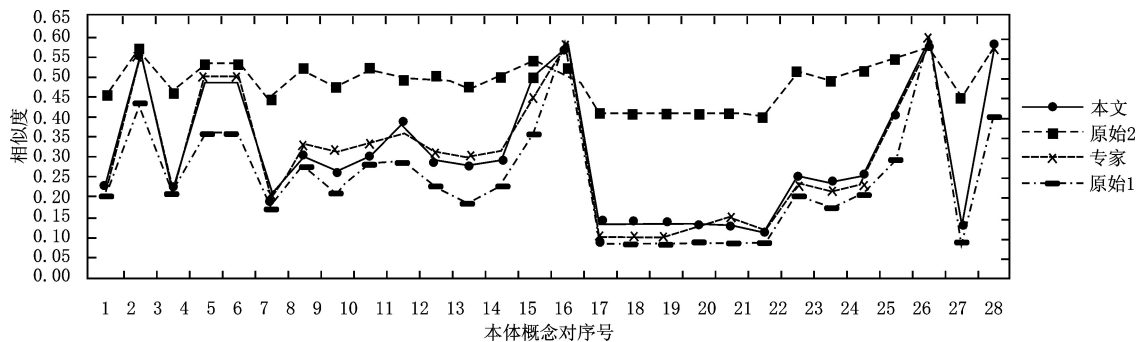


图 1 概念语义相似度计算结果对比分析图

$$Dev_i = \frac{|Sim_i - Emp_i|}{Emp_i} \quad i \in [1, N], N \in \mathbb{Z}^+ \quad (15)$$

则概念语义相似度算法的计算结果与专家经验的兼容度为

$$Cpt = \frac{1}{N} \sum_{i=1}^N (1 - Dev_i) \quad (16)$$

根据表 1 的数据,可以得到本文算法的兼容度是 89.72%;“原始 1”算法的兼容度是 73.70%;“原

始 2”算法的兼容度是 53.64%。由于兼容度表示了概念语义相似度算法得出的结果与专家经验的吻合程度,所以与两种原始算法相比,本文算法所得结果更接近专家经验,符合人类的主观判断。

以上分析说明:(1)本文算法考虑了原始经典算法在语义相似度计算时分析的各种因素;(2)通过对原始算法的改进,本文算法的结果在精确性和准确性上都有了提高;(3)本文算法得到的结果与专家经验较接近,符合人类的主观判断。另外,虽然本文所考虑的因素较多,在一定程度上提高了算法的复杂度,但是高速发展的计算机技术,已经能够解决这种数量级别的运算。

5 结束语

本文综合了语义相似度计算需考虑的多种因素,提出了一种改进的领域本体概念语义相似度计算方法;通过实验分析与验证,与原始算法相比,本文算法在精确性和准确性上都有了提高,为应用语义相似度解决实际问题提供了更好的支持。本文采用的实验数据本体结构较简单,会对算法计算结果产生一定的影响,下一步将开展提高算法兼容度的研究。

参考文献:

- [1] Shunmughavel V, Jaganathan P. Semantic enrichment in ontology mapping using concept similarity computing[C]// IEEE Fourth International Conference on Advanced Computing 2012. Chennai India: MIT, 2012.
- [2] 戴维民. 语义网信息组织技术与方法[M]. 上海: 学林出版社, 2008.

- [3] 韩欣, 樊永生, 马春森, 等. 基于树状结构的语义相似度计算方法分析[J]. 微电子学与计算机, 2012, 29(5): 38-41.
- [4] Holger F. The GOSim package. The GO Consortium, [EB/OL]. [2013-10-28]. <http://www.bioconductor.org/package/2.13/bioc/html/GOSim.html>.
- [5] 黄果, 周竹荣. 基于领域本体的概念语义相似度计算研究[J]. 计算机工程与设计, 2007, 28(10): 2460-2463.
- [6] 赵永金, 郑洪源, 丁秋林. 一种基于本体的语义相似度算法研究[J]. 计算机应用, 2009, 29(11): 3074-3076.
- [7] 张兰芳. 一种基于本体的自然语言语义相似度算法[J]. 桂林理工大学学报, 2012, 32(2): 253-258.
- [8] Tervsky. Features of similarity[J]. Psychological Review, 1977, 84(4): 327-352.
- [9] Paolucci M, Payne T, Sycara K, et al. Semantic matching of web services capabilities[C]// First International Semantic Web Conference. Sardinia, Italy: Springer, 2002: 333-347.
- [10] Li Zhaolong, DU Junping. A conceptual semantic similarity calculation model based On tourism domain ontology[C]// Proceedings of the 31st Chinese Control Conference. Hefei China: IEEE CPP, 2012: 3863-3868.

作者简介:

曹 飏 男, (1985-), 博士研究生. 研究方向为空间信息获取与处理技术, 多媒体信息系统与虚拟现实技术。

吴玲达 女, (1962-), 博士, 教授. 研究方向为空间信息获取与处理技术, 多媒体信息系统与虚拟现实技术。

(上接第 108 页)

作者简介:

鞠 豪 男, (1987-), 硕士研究生. 研究方向为模拟和混合集成电路设计。

万美琳 男, (1988-), 博士. 研究方向为射频集成电路设计。

马硝霞 女, (1988-), 硕士研究生. 研究方向为模拟集成电

路设计。

韩 爽 女, (1989-), 硕士研究生. 研究方向为数字集成电路设计。

戴 葵 男, (1968-), 教授. 研究方向为计算机体系结构、集成电路设计和信息安全。