

本体映射中相似度计算的改进

郑晓洁 张 琳

(上海海事大学信息工程学院 上海 201306)

摘 要 本体映射是一种常用的解决本体异构的方法,在本体映射过程中,概念相似度计算是重要的环节。针对目前概念相似度计算方法存在的不足,采用一种综合的方法进一步改善计算的过程,从本体概念的语义相似度、属性、实例、结构等方面计算概念的相似度。通过引入概念相关度以及属性论方法,寻找更加有效的概念相似度计算方法。最后的实验证明,此方法能适应不同规模的本体,并能提高概念对相似度的准确性。

关键词 本体,本体映射,相似度,属性论

中图法分类号 TM301 **文献标识码** A

Modification of Similarity Computation in Ontology Mapping

ZHENG Xiao-jie ZHANG Lin

(College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China)

Abstract Ontology mapping is one of common solutions of the ontology heterogeneous. Similarity computation among ontologies is the critical step of mapping process. Aiming at the current problems, and for making a further improvement in similarity computation, this paper put forward a comprehensive approach. This approach involves several factors of concept, semantic, properties, instances and structure of the concept. For finding more useful similarity computation method, the paper introduced relativity and attribute theory. At the end, an experiment was used to prove this method can adapt the different scale ontologies, and also can improve the accuracy of the similarity.

Keywords Ontology, Ontology mapping, Similarity, Attribute theory

1 引言

本体(Ontology)作为一种能在语义和知识层次上描述信息系统的概念模型建模工具,自提出以来就引起了国内外众多科研人员的关注,并在计算机领域如知识工程、数字图书馆、软件复用、信息检索以及语义 Web 等^[1]得到了广泛的应用。

随着本体技术的不断发展,以及本体应用领域的逐渐增多,产生了很多本体。但这些本体之间会因为构造目的、构造角度、构造方法等的不同,而不能直接进行互操作,这样也就违背了本体实现知识共享与重用的原则。目前解决异构本体之间互操作问题的比较可行和实用的方法是在异构本体的概念间建立映射关系。

本体概念相似度计算是本体映射的关键环节。目前在一些本体映射系统中采用单一的基于模式或基于本体的某个元素的方法来进行相似度计算^[2],例如文献^[3]中,华盛顿大学开发的 GLUE 系统只基于概念的实例来计算相似度。然而这些方法过于片面,因为本体概念的实例、关系、属性、结构等信息对本体概念也具有重要的描述作用,因此在计算概念相似度时也需要考虑这些因素。另外,目前在计算两个本体概念的相似度时,本体中的每一对概念都会被考虑在内,计算量

很大。但有的概念对根本不相似,计算它们的相似度是没有必要的。因此,计算时可以对概念对的数量进行限制,以减少计算量。

本文针对相似度计算过程中所存在的这些问题,提出了一种改进的计算概念相似度的方法。首先计算概念对之间的相关度,通过一个提前设定的阈值,选择具有较大相关性的概念对来加入一个候选概念对集;然后,针对候选概念对集中的概念对进行相似度的计算,这样可以减少计算量。在进行相似度计算时,分别计算概念的语义相似度以及基于概念的属性、实例、结构等几个方面的相似度,最后将这些相似度根据一定的策略进行合并,得到最终的本体的综合相似度。

在基于概念的属性相似度进行计算时,本文采用了一种新的方法——属性论,利用该理论中的重心剖分模型,计算得到概念对的基于属性的相似度。

2 本体映射

所谓本体映射,是指有两个本体 O_1 、 O_2 ,对于本体 O_1 中的每个概念试图在本体 O_2 中找到一个或几个语义相同或相近的对应概念,对于本体 O_2 中的概念亦是如此。本体映射通过在异构本体间建立映射规则,使得所要交互的信息借助这些规则能够在不同本体间进行传递。

到稿日期:2013-05-21 返修日期:2013-07-08 本文受国家自然科学基金、存储安全中介系统理论、仿真和实现技术研究(61070154)资助。

郑晓洁(1988—),女,硕士生,主要研究方向为模式识别与智能系统,E-mail:slam.dunk123@163.com;张琳(1973—),女,博士,副教授,硕士生导师,主要研究方向为港航信息化技术、智能信息处理、信息检索、本体与知识工程等。

本体映射过程一般分为 5 个步骤^[2]:信息本体化、相似度提取、语义映射、映射执行和映射后处理过程。其中,相似度提取是本体映射的一个重要的步骤,它主要是进行相似度的计算,产生一个相似矩阵,之后的映射过程就是基于这个矩阵来进行的。

关于两个概念的相似度,有以下定义:

定义 1 $\text{Sim}: \omega_1 \times \omega_2 \times O_1 \times O_2 \rightarrow [0, 1]$, 相似值在 0 和 1 之间。 $\text{Sim}(M, N)$ 表示概念 M 和概念 N 之间的相似度。 O_1 和 O_2 是需要映射的两个本体, ω_1 和 ω_2 是 O_1 和 O_2 所基于的术语集。 $\text{Sim}(M, N) = 1$: M 和 N 是相同的两个概念; $\text{Sim}(M, N) = 0$: M 和 N 是两个完全不同的概念。

3 本体相似度计算过程

3.1 相关度以及候选概念对集

在本体中,概念的相似度和相关度有不同的意义。可以认为概念相似度反映概念之间的聚合特点,而概念相关度则反映概念之间的组合特点。但是,相似性与相关性并不是互斥的关系。Resnik 认为^[4]相似性可以被视为一种特殊的相关性,即对象间基于蕴含关系的相关性,对象间的蕴含关系体现了对象的共同性,因而蕴含关系的相关性等同于相似性。所以,通常情况下也认定相似的概念具有相关性,也就是说相似的概念一般是相关的。但反过来,相关的概念却不一定是相似的,如电视机与遥控器有很强的相关性但其相似性很低。可以说相似只是相关的一个特殊的方面。总之,相关度指概念之间相关的程度,是一个比相似性更普遍的概念^[5]。

在计算相关度时,本文采用 Hist-St-Onge 语义相关度算法,其基本思想是^[6]:若两个词 s_1 与 s_2 在 WordNet 同义词集 (*synset*) 中有一条较短的路径相连,则它们在语义上就有较大的相关度。若此路径不存在,则两者之间的相关度为 0。在 Hist-St-Onge 算法中,语义相关度的计算公式如下所示:

$$\text{Rel}_{\text{HS}}(s_1, s_2) = c - \text{len}(s_1, s_2) - k \times \text{turns}(s_1, s_2) \quad (1)$$

WordNet 是 Princeton 大学开发的一个知识库系统,其实质就是一部英语语义词典。

在式(1)中, c 和 k 是两个常量参数; $\text{turns}(s_1, s_2)$ 代表在 WordNet 中 s_1 与 s_2 之间的路径转向次数; $\text{len}(s_1, s_2)$ 是路径长度。 $\text{Rel}_{\text{HS}}(s_1, s_2)$ 的取值范围为 $[0, 1]$ 。给定一个阈值,只有当 $\text{Rel}_{\text{HS}}(s_1, s_2)$ 的值大于这个阈值时才认定两个概念具有概念相关性,从而判定它们之间也可能具有概念相似性,并将其加入到候选概念对集中。

3.2 概念的语义相似度

语义相似度是指概念间自身语义的相似程度,也就是通过概念的意思来计算相似度。在计算语义相似度时,可依据 WordNet 进行计算。

利用 WordNet 计算语义关系的方法很多,由于 WordNet 是树状结构的,因此这些方法中最为简单的就是利用路径长度的方法。在 WordNet 中通过识别两个词之间的最短路径来发现语义关系上的相似性。基本思想是:从一个词结点到另一个词结点的路径越短,则表明它们相似的程度越高;反之路径越长,相似度越低。本文中使用的 Leacock-Chodorow 语义相似度算法,考虑到当路径长度相同时,越靠近树根,语义相似度越小,引入了深度的制约条件^[6]。计算公式如式(2)所示:

$$\text{Sim}(M, N) = -\log \frac{\text{len}(M, N)}{2\text{Depth}} \quad (2)$$

式中, $\text{len}(M, N)$ 表示的是概念 M 与 N 的语义在 WordNet 之间的距离, Depth 表示概念 M 与 N 的语义在 WordNet 中的最大深度。由于概念 M, N 在 WordNet 中可能存在多个语义,需要将 M 概念的每一个语义都与 N 概念的每一个语义进行计算,最后选择最大的 $\text{Sim}(M, N)$ 作为两个概念的语义相似度,即 $\text{Sim}_{\text{max}} = \max(\text{Sim}(M, N))$ 。

3.3 基于概念属性的相似度计算

属性对于概念具有重要的描述作用,如果可以用一个概念的属性集近似地描述另一个概念,那么两个概念之间可能具有相似性。

3.3.1 属性论简介

事物只有通过属性才能反映自身以及与其他事物之间的关系,所以属性是人们区分不同事物的标志或基准^[7]。同理,在本体中也可通过属性来区分不同的概念。

在人工智能领域,属性论是以神经科学、哲学和逻辑学为基础的,旨在对人类认识与思维的各种行为加以数学表达的理论。该理论提出了属性是人类感觉信息处理之单元的观点和基于属性计算的感知与认知模拟的数学理论的方法,简称属性论^[8]。

本文采用属性论中的属性重心剖分模型(又称属性坐标系)对本体概念相似度进行计算。

3.3.2 属性重心剖分模型

人们在认识事物过程中感觉所检测出的和编码的各种简单属性被合成一个复杂的整合属性,并且以整合属性的方式被记忆和储存起来。把属性的整合过程定义为属性合取^[9]。

定义 1 设 $M(X), N(X)$ 分别为事物 X 的不同属性,用 $\wedge$ 表示合取算子,则属性的合取过程表示为: $S(X) = M(X) \wedge N(X)$, 称 $S(X)$ 为合属性,而 $M(X)$ 和 $N(X)$ 称为素属性。可用合属性 $S(X)$ 代表事物 X 。

定义 2 设 n 维单纯形 $K = \{e_0, e_1, \dots, e_n\}$, 其顶点为属性集 P_x 中的 $n+1$ 个属性 $e_j(x)$, 则 K 为属性多面体。在 K 的第一次重心剖分 $K^{(1)}$ 中, $r+1$ 个属性的整合属性 $e_{0r} \wedge e_{1r} \wedge \dots \wedge e_{rr}$ 置放在由这 $r+1$ 个属性所构成的 r 维单纯形的重心剖分点上,记为 $p(s_r)$, 且 $p(s_r) = e_{0r} \wedge e_{1r} \wedge \dots \wedge e_{rr}$ 。以此类推,得到 K 的 m 次剖分 $K^{(m)}(x)$, 称为 x 的属性重心坐标系。这样的模型被称为属性的重心剖分模型。

3.3.3 属性重心剖分模型在相似度计算中的应用

一个概念类似于一个事物,可以用概念的一组特征属性共同表示这个概念。

1. 构建概念的重心剖分模型

假设 M, N 分别为本体 O_1 和 O_2 的概念。现基于属性论对概念 M 进行表示。设概念 M 的属性集为 (a_1, a_2, a_3) , 分别为属性集中的各个属性设定一个数值 w_{1i} ($0 \leq w_{1i} \leq 1$), 其表示各属性对于 M 的描述程度,相当于各个属性对于 M 的重要性的权值。由此可以得到属性集 (a_1, a_2, a_3) 针对于概念 M 的合属性向量 $d = (w_{11}, w_{12}, w_{13})$ 。设 $s = (w_1, w_2, w_3)$ 是以合属性向量 (w_{11}, w_{12}, w_{13}) 的分量坐标 w_{1i} 为顶点的单纯形, 则 d 与 s 的交点 G_1 恰好是单纯形 s 的重心剖分点。重心剖分点 G_1 表达了各属性之间权值大小的关系, G_1 距某个顶点越远,表示该顶点所在坐标轴代表的属性对于概念的描述程

度即重要性的权值越小。

现构建概念 M 的属性坐标系^[9]。属性 a_1, a_2, a_3 表示坐标系的 3 个轴, 1 为坐标单位长度, 长度的值表示各属性的权值。设概念 M 的合属性向量 $d = (w_{11}, w_{12}, w_{13}) = (0.8, 0.5, 0.7)$ 在属性坐标系中对应一个二维单纯形, 如图 1 所示, $\triangle ABC$ 所在的平面方程为: $\frac{a_1}{0.8} + \frac{a_2}{0.5} + \frac{a_3}{0.7} = 1$ 。

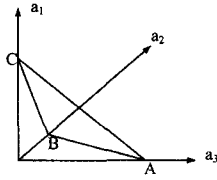


图 1 属性坐标表示

把 $\triangle ABC$ 从坐标系中抽出来, 如图 2 所示, 这时包含任一点 X 的 $\triangle ABC$ 的方程为:

$$\begin{cases} X = \lambda_1 A + \lambda_2 B + \lambda_3 C \\ \lambda_1 + \lambda_2 + \lambda_3 = 1 \end{cases}$$

式中, $\lambda_1, \lambda_2, \lambda_3$ 均大于 0。则概念 M 的合属性向量 d 所确定单纯形的重心为:

$$G_1 = (g_1, g_2, g_3) = (\frac{w_{11}}{3}, \frac{w_{12}}{3}, \frac{w_{13}}{3})$$

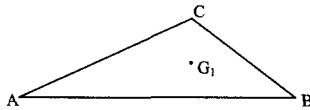


图 2 属性单纯形

重复上述过程, 得到概念 M 的属性集针对于概念 N 的合属性向量的单纯形重心点 G_1' 以及概念 N 的属性集分别针对于概念 M 和概念 N 的合属性向量的单纯形重心点 G_2 和 G_2' 。

2. 相似度计算

由于重心剖分点 G_1 表达了各属性之间权值大小的关系, 因此在属性坐标系中, 坐标原点 O 与重心剖分点 G_1 的连线所构成的向量 OG_1 可以表示属性集对于概念的描述, 将 OG_1 向量称为 M 的属性集针对于概念 M 的重心线, OG_1' 、 OG_2 和 OG_2' 也是如此。

如果两个概念相同, 则它们的属性对于两个概念的描述程度应该是相同的, 即对应的合属性向量是相同的, 此时两个重心线是重合的, 两者之间的夹角为 0; 如果两个概念相似, 则两个概念的属性也应该相似, 那么可以用一个概念的属性来大致描述另一个概念, 两个重心线之间存在着一个夹角, 并且此夹角随着两个概念之间相似度的减小而增大。因此, 我们可以采用两个重心线之间的夹角余弦值来表示两个概念之间的相似度, 值越大则表示两个概念越相似, 当值为 1 时, 表示两个概念完全相同; 反之, 值越小表示两个概念相似度越小, 当值为 0 时, 表示两个概念完全不同。

计算时, 首先计算概念 M 与概念 N 基于概念 M 的属性的相似度, 即 OG_1 与 OG_1' 的夹角; 然后, 计算两个概念基于概念 N 的属性的相似度, 即 OG_2 与 OG_2' 的夹角。

通过计算得到 $\cos(OG_1, OG_1')$ 和 $\cos(OG_2, OG_2')$, 最后根据式(3)计算得到最终的两个概念基于属性的相似度:

$$\text{Sim}_{pre}(M, N) = 0.5 \times \cos(OG_1, OG_1') + 0.5 \times \cos(OG_2, OG_2') \quad (3)$$

3.4 基于概念实例的相似度计算

基于实例计算概念相似度的理论依据是^[2]: 如果两个概念所具有的实例全部相同, 那么这两个概念是相同的; 如果两个概念具有相同实例的比重是相同的, 那么这两个概念是相似的。需要注意的是, 一个概念的实例也是它祖先概念的实例。

将概念 M 和 N 基于实例的相似度记为 $\text{Sim}_{instance}(M, N)$, 利用一定量的实例在两个概念中出现的联合分布概率来计算两个概念的相似度。其计算公式为:

$$\text{Sim}_{instance}(M, N) = \frac{P(M \cap N)}{P(M \cup N)} = \frac{P(M, N)}{P(M, N) + P(M, \bar{N}) + P(\bar{M}, N)} \quad (4)$$

在此公式中涉及 3 个概率: $P(M, N)$ 、 $P(M, \bar{N})$ 、 $P(\bar{M}, N)$ 。其中 $P(M, \bar{N})$ 是从一个本体的实例空间中随机选取的一个实例属于 M 但不属于 N 的概率, $P(M, N)$ 和 $P(\bar{M}, N)$ 也是如此。

用 U_i 表示本体 O_i 中的实例集, $K(U_i)$ 表示实例集中的实例个数, 用 $K(U_i^{M, N})$ 表示在 U_i 中既属于 M 又属于 N 的实例个数。以 $P(M, N)$ 的计算为例, 其计算过程如下^[10]:

(1) 对于本体 O_1 中的实例集 U_1 , 把它分成属于概念 M 的实例集 U_1^M 和不属于概念 M 的实例集 $U_1^{\bar{M}}$;

(2) 把这两个实例集中的实例分别作为正反样本, 用机器学习方法来训练对于概念 M 的学习器 L ;

(3) 对于本体 O_2 的实例集 U_2 , 把它分成属于概念 N 的实例集 U_2^N 和不属于概念 N 的实例集 $U_2^{\bar{N}}$;

(4) 使用学习器 L 对实例集 U_2^N 中的实例进行分类, 分成两个实例集 $U_2^{M, N}$ 和 $U_2^{\bar{M}, N}$ 。同样把实例集 $U_2^{\bar{N}}$ 分成两个实例集 $U_2^{M, \bar{N}}$ 和 $U_2^{\bar{M}, \bar{N}}$ 。这样可以获得实例集 $U_2^{M, N}$ 、 $U_2^{\bar{M}, N}$ 、 $U_2^{M, \bar{N}}$ 和 $U_2^{\bar{M}, \bar{N}}$ 。

(5) 把本体 O_1 和本体 O_2 的位置调换, 重复以上各步, 这样最终可以获得实例集 $U_2^{M, N}$ 、 $U_2^{\bar{M}, N}$ 、 $U_2^{M, \bar{N}}$ 和 $U_2^{\bar{M}, \bar{N}}$ 。从各步计算中得到 $K(U_1)$ 、 $K(U_2)$ 、 $K(U_1^{M, N})$ 、 $K(U_2^{M, N})$, 利用下式来计算 $P(M, N)$:

$$P(M, N) = \frac{K(U_1^{M, N}) + K(U_2^{M, N})}{K(U_1) + K(U_2)} \quad (5)$$

采用同样的步骤方法来计算 $P(A, \bar{B})$ 和 $P(\bar{A}, B)$, 计算公式分别为:

$$P(M, \bar{N}) = \frac{K(U_1^{M, \bar{N}}) + K(U_2^{\bar{M}, \bar{N}})}{K(U_1) + K(U_2)} \quad (6)$$

$$P(\bar{M}, N) = \frac{K(U_1^{\bar{M}, N}) + K(U_2^{M, N})}{K(U_1) + K(U_2)} \quad (7)$$

最后根据式(4)计算得到概念 M 和 N 基于实例的相似度 $\text{Sim}_{instance}(M, N)$ 。

3.5 基于概念结构的相似度计算

本体概念的关系描述的是概念与概念之间的交互作用。从语义上讲, 本体中的基本关系有 4 种: kind-of、part-of、instance-of、attribute-of。其中 part-of 表示概念之间部分与整体的关系; kind-of 表示概念之间的继承关系。

在本体中可以依据概念的 kind-of 和 part-of 关系将所有的概念连成一个树状结构, 即将概念分层次。根据本体论的启发规则的一些规定, 如果两个概念的父概念相似, 那么这两个概念也可能相似; 如果概念对中的概念 M 和概念 N 都有

多个子概念,设定一个阈值 δ ,若概念 M 中有大于 δ 个子概念与概念 N 中的子概念相似,则可认为概念 M 与概念 N 是相似的;如果某概念的兄弟概念都与另一概念 X 相似,那么此概念也与 X 相似,再参考文献[11],计算概念结构相似度可以采用公式:

$$\text{Sim}_{\text{struc}}(M, N) = \frac{\alpha \text{Sim}_p(M, N) + \beta \text{Sim}_s(M, N) + \gamma \text{Sim}_b(M, N)}{\alpha + \beta + \gamma} \quad (8)$$

式中, M, N 分别为本体 O_1 和 O_2 的概念, $\text{Sim}_p(M, N)$ 表示度量概念 M, N 的父概念集之间的相似度, $\text{Sim}_b(M, N)$ (兄弟概念集)和 $\text{Sim}_s(M, N)$ (子概念集)与之类似, α, β, γ 为权重因子。由于在概念的层次结构中,父、子、兄弟概念对概念相似度的影响是不同的,其中父概念占有绝对的权重,因此可预先设定 $\alpha \geq \beta \geq \gamma > 0$ 。

3.6 总的概念相似度

最后为基于概念的语义、属性、实例、结构等方面计算所得的相似度分别分配一个权值,综合这些相似度得到概念对的最终相似度 $\text{Sim}(M, N)$:

$$\text{Sim}(M, N) = w_1 \text{Sim}_{\text{sema}}(M, N) + w_2 \text{Sim}_{\text{pre}}(M, N) + w_3 \text{Sim}_{\text{instance}}(M, N) + w_4 \text{Sim}_{\text{struc}}(M, N) \quad (9)$$

其中,权值 $w_1 + w_2 + w_3 + w_4 = 1$ ($w_1, w_2, w_3, w_4 > 0$), w_1, w_2, w_3, w_4 可以根据一定的策略进行设定。

4 实验及分析

4.1 实验数据

为验证上述算法的有效性,本文依据文献[12]以及相关领域的知识构建了两个本体即 O_1 和 O_2 ,其中 O_1 是基于水路运输的货物的本体; O_2 是基于集装箱运输的货物本体。每个

本体中具体所使用的数据的统计如表 1 所列。

表 1 实验数据

本体	概念	属性	实例
O_1 水路运输货物本体	21	14	17
O_2 集装箱运输货物本体	18	12	21

4.2 结果与分析

通过结合 WordNet,计算得到概念的语义相似度;参考属性的重心剖分模型以及相应权值的设定,得到概念基于属性的相似度;采用名称学习器,得到概念基于实例的相似度;由于本体的结构关系是在语义基础上的,因此采用概念的父、兄、子概念的语义相似度来计算概念的结构相似度;最后采用等权值的方式,即 $w_1 = 0.25, w_2 = 0.25, w_3 = 0.25, w_4 = 0.25$,将这 4 个角度的相似度结合在一起,得到综合的相似度值。表 2 给出了两个本体中一些概念对的 4 个角度的相似度值以及综合的相似度值。

传统的 GLUE 系统只是基于实例的概率统计方法来进行相似度的计算,而当需要映射的两个本体的实例规模较小时,如本例中构建的本体,通过这种方法计算所得的相似度是比较粗略的、不精确的,这从表 2 的基于实例的相似度一行可以看出。

表 2 也显示,如果仅从概念的某个方面去计算两个概念的相似度,不同角度可能得到不同的映射关系。而本文所采用的综合的相似度计算方法,从概念的语义、属性、实例、结构等因素出发,得到概念的多个不同角度的相似度,然后为每一个相似度设定一个相等的权值,最终得到一个综合值。这一综合值能够较为全面地反映出概念对之间的相似性,使得之后的映射过程更为精确。

表 2 实验结果

相似度类型	1		2		3	
	O_1	O_2	O_1	O_2	O_1	O_2
	Explosives	Dangerous cargo	Flammable liquids	Bulk liquid cargo	Perishable cargo and reefer cargo	refrigerated cargo
语义相似度		0.222		0.651		0.699
基于特征属性的相似度		0.881		0.733		1
基于实例的相似度		0.182		0.286		0.429
基于结构的相似度		0.581		0.576		0.934
总的相似度		0.467		0.562		0.766

结束语 本文提出了一种综合的计算异构本体概念对的相似度方法。该方法首先引入了相关度概念,通过计算本体间概念对的相关度筛选得到一个候选概念对集,降低了大规模本体映射时相似度计算的时间复杂度;然后又分别从概念的语义、属性、实例以及结构 4 个方面计算了相应的相似度;最后,通过为每一个基于单因素计算得到的相似度赋一个权值,得到了概念对的最终的综合相似度。在进行本体映射时,采用此相似度决定两个概念是否可以建立映射关系。

由于属性是人们区分不同事物的基准,因此在计算基于概念属性的相似度时,本文引入了一种新的方法——属性论,该理论提出了属性是人类感觉信息处理之单元的观点和基于属性计算的感知与认知模拟的数学理论的方法。通过构建属性集对概念对描述的重心剖分模型,得到两个概念基于属性的相似度。

本文所提出的综合方法尽管提高了概念对相似度的精

度,但由于涉及到概念的多个因素,计算量较大,而且在权值分配以及设定属性对于概念的描述程度等方面都是主观设定的,因此国际经济对于映射来说是不利的。在之后的研究中,主要将针对这些方面做进一步的改进,使得异构本体间的映射更加高效和精确。

参考文献

- [1] 邓志鸿,唐世渭,张铭,等. Ontology 研究综述[J]. 北京大学学报:自然版,2002,38(5):730-738
- [2] 郑丽萍. 本体映射的研究[D]. 济南:山东科技大学,2005
- [3] Doan A H, Madhavan J, Domingos P, et al. Learning to Map between Ontologies on the Semantic Web[C]//Proc World-Wide Web Conf. ACM Press, May 2002:1132-1151
- [4] Resnik P. Using Information Content to Evaluate Semantic Similarity in a Taxonomy[C]//Proceeding of the International Joint Conference on artificial Intelligence, Montreal Canada, 1995:

- [5] 张路. 本体映射中概念语义相似度计算方法研究[D]. 兰州: 兰州理工大学, 2010
- [6] 张承立, 陈剑波, 齐开悦. 基于语义网的语义相似度算法改进[J]. 计算机工程与应用, 2006, 17: 165-179
- [7] 冯嘉礼. 人工神经元的一种定性映射解释[J]. 计算机科学, 2001, 28(S1): 248
- [8] 王涛, 樊孝忠, 朱俭. 基于属性重心剖分模型和 LSA 的文本相似度计算[C]// 2010 3rd International Conference on Computational

Intelligence and Industrial Application, 2010

- [9] 潘谦红, 王炬, 史忠植. 基于属性论的文本相似度计算[J]. 计算机学报, 1999, 22(6): 651-655
- [10] 甘健侯, 姜跃, 夏幼明. 本体方法及应用[M]. 北京: 科学出版社, 2011: 73-74
- [11] 张忠平, 田淑霞, 刘洪强. 一种综合的本体相似度计算方法[J]. 计算机科学, 2008, 35(12): 142-182
- [12] 王学锋, 周晶洁, 殷明, 等. 货物学(第二版)[M]. 上海: 同济大学出版社, 2009: 232-242

(上接第 97 页)

5.4 时间复杂性分析

基于统计量模板的特征提取算法中, 模板移动至某一像素位置需要做 $L * L$ 次比较和累加, 而一幅图像包含了 WH 个像素, 因此时间复杂度为 $O(WHL^2)$ 。文献[15]自相关函数方法的时间复杂度是 $O(WHL^2)$, 但其需要做 WHL^2 次乘法, 因此本文提出的特征提取方法速度要更快一些。不同 K 值时的时间开销比较如表 3 所列。参数 L 越大, 则建模时间开销越大, 图 9 显示了建模时间开销与参数 L 的关系, 建模时间开销几乎成线性关系增加。

表 3 不同 K 值时的时间开销 (s)

K	32	64	128	256
本文方法	0.013	0.051	0.199	0.813
自相关方法[15]	0.219	0.155	0.715	3.00

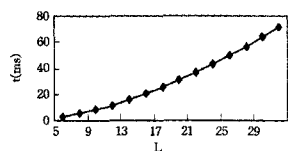


图 9 不同 L 对建模时间开销的影响

结束语 针对误差分散半调图像分类问题, 对半调图像特征提取方法进行了研究。根据误差分散半调图像的特点, 给出了像素对的概念和统计量模板描述模型, 提出了基于统计量模板的特征提取方法。其次, 建立了误差最小优化目标函数, 利用梯度下降法来求取最优类特征矩阵, 以此来描述类特征。分析了迭代求解的收敛性和类特征矩阵可分性。最后, 通过大量实验比较和分析, 验证了所提方法的有效性。

下一步将集中进行误差分散半调图像特点的深入分析、基于统计量模板的分类方法研究及统计量模板的扩展。

参考文献

- [1] Ulichney R. Digital halftoning and the physical reconstruction function[D]. Cambridge: MIT, 1986
- [2] Ulichney R. Dithering with blue noise[J]. Proceedings of the IEEE, 1988, 76(1): 56-79
- [3] Knuth D E. Digital halftones by dot diffusion[J]. ACM Transactions Graph, 1987, 6(4): 245-273
- [4] Mese M, Vaidyanathan P. Look up table Method for inverse halftoning[J]. IEEE Transactions on Image Processing, 2000, 10(10): 1566-1578
- [5] Miceli C M, Parker K J. Inverse halftoning[J]. Journal of Electron Imaging, 1992, 1: 143-151

- [6] Saika Y, Okamoto K, Matsubara F. Probabilistic Modeling to Inverse Halftoning based on Super Resolution[C]//Proceedings of International Conference on Control, Automation and Systems. New York: IEEE Press, 2010: 162-167
- [7] Stevenson R. Inverse Halftoning via MAP Estimation[J]. IEEE Transactions on Image Processing, 1997, 6(4): 574-583
- [8] Unal G B, Cetin A E. Restoration of error-diffused images using projection onto convex sets[J]. IEEE Transactions on Image Processing, 2001, 10(12): 1836-1841
- [9] Mese M, Vaidyanathan P. Optimized halftoning using dot diffusion and methods for inverse halftoning[J]. IEEE Transactions on Image Processing, 2000, 9(4): 691-709
- [10] Huang W B, Su A W, Kuo Y H. Neural network based method for image halftoning and inverse halftoning[J]. Expert Systems with Applications, 2008, 34: 2491-2501
- [11] Fung Y, Chan Y. A POCS-Based Restoration Algorithm for Restoring Halftoned Color-Quantized Images[J]. IEEE Transactions on Image Processing, 2006, 15(7): 1985-1992
- [12] Mese M, Vaidyanathan P. Tree-structured method for LUT inverse halftoning and for image halftone[J]. IEEE Transactions on Image Processing, 2002, 11(6): 644-655
- [13] Chung K, Wu S. Inverse halftoning algorithm using edge-based lookup table approach[J]. IEEE Transactions on Image Processing, 2005, 14(10): 1583-1589
- [14] Suetake N, Tanaka G, et al. Look-Up Table and Gaussian Filter-Based Inverse Halftoning Method Excellent in Gray-Scale Reproducibility of Details and Flat Regions[J]. OPTICAL REVIEW, 2009, 16(6): 594-600
- [15] Chang P C, Yu C S. Neural net classification and LMS reconstruction to halftone images[C]//Proceedings of SPIE-The International Society for Optical Engineering. San Jose: SPIE, 1997, 3309: 592-602
- [16] 孔月萍, 曾平, 张跃鹏. 一种半调图像分类识别算法[J]. 西安电子科技大学学报: 自然科学版, 2011, 38(5): 62-69
- [17] 孔月萍. 图像逆半调及其质量评价技术研究[D]. 西安: 西安电子科技大学, 2008
- [18] Liu Y F, Guo J M, Lee J D. Inverse Halftoning Based on the Bayesian Theorem[J]. IEEE Transactions on Image Processing, 2011, 20(4): 1077-1084
- [19] Liu Y F, Guo J M, Lee J D. Halftone Image Classification Using LMS Algorithm and Naive Bayes[J]. IEEE Transactions on Image Processing, 2011, 20(10): 2837-2847
- [20] 黄平, 孟永刚. 最优化理论与方法[M]. 北京: 清华大学出版社, 2009