

基于石油领域本体的概念相似度级联模型^①

赵国梁, 宫法明

(中国石油大学(华东)计算机与通信工程学院, 青岛 266580)

通讯作者: 赵国梁, E-mail: wf_zhaoguoliang@163.com

摘要: 提出了一种利用级联模型来计算本体中概念间相似度的新方法. 在模型的第一阶段, 采用了基于距离的语义相似度计算方法, 计算出概念对在本体中的路径得分; 第二阶段, 采用 IC (Information Content) 算法精确计算概念对间相似度得分, 并利用概念的公共子代集合对算法进行了扩展; 第三阶段我们采用了特征整合策略, 将所有的相似性得分构建成特征向量来描述概念对, 并且使用权重来平衡第一阶段与第二阶段的相似度结算得分. 最后使用 BP 神经网络确定两个概念的相似性. 我们对新提出的语义相似度算法进行了评估, 并与现有的方法相比. 实验结果表明, 该方法有效提高相似度算法的准确性和科学性.

关键词: 级联模型; 石油本体; BP 神经网络; 路径得分; IC 得分

引用格式: 赵国梁, 宫法明. 基于石油领域本体的概念相似度级联模型. 计算机系统应用, 2018, 27(7): 182-187. <http://www.c-s-a.org.cn/1003-3254/6462.html>

Cascade Model for Semantic Similarity of Concept Based on Petroleum Ontology

ZHAO Guo-Liang, GONG Fa-Ming

(College of Computer & Communication Engineering, China University of Petroleum, Qingdao 266580, China)

Abstract: This paper presents a new cascade architecture to calculate the similarity between concepts in the ontology. In the first stage of proposed model, we use path-based methods to calculate the concept of path score in the ontology. In the second stage, we use Information Content (IC)-based methods to obtain similarity scores of two extended concepts with further consideration of the two concepts of public parent set and public subset. In the third stage, we adopt a feature integration strategy to combine all the similarity scores derived from the ontology to construct various kinds of features to characterize each concept pair and using weights to balance the concept of first-stage-scores with the third-stage-scores. In the end, BP neural network is used to obtain the similarity of two concepts. This model has been evaluated and compared with existing methods when applied to the task of semantic similarity estimation. Experimental results show that the proposed method effectively improves the accuracy and scientificity of similar calculation.

Key words: cascaded model; petroleum ontology; BP neural network; path-score; IC-score

1 概述

概念的语义相似度计算已经成为数据挖掘和信息检索领域的基本问题, 而且是自然语言处理的核心问题^[1]. 例如, 语义相似性度量已应用于词义消歧^[2]、信

息提取^[3]、文本相似度计算^[4]以及文本聚类^[5].

目前, 语义相似度的计算方法大致可分为两类: (1) 依赖于分类中术语间的层次关系的基于路径的度量; (2) 根据概念的特异性分类, 基于语料库信息内容

① 基金项目: 科技部创新方法工作专项资助 (2015IM010300); 北京市重点实验室开放课题 (BKBD-2017KF07)

Foundation item: Special Fundation for Innovation Methods of Ministry of Science and Technology (2015IM010300); Open Project of Beijing Key Laboratory (BKBD-2017KF07)

收稿时间: 2017-11-30; 修改时间: 2017-12-21; 采用时间: 2018-01-04; csa 在线出版时间: 2018-06-27

(IC) 的测度. 许多相似度计算是建立在本体的内在结构上, 在本文中, 我们首先考虑在查询概念对在本体中的路径信息, 然后我们不仅考虑了相关概念的父类集合, 并考虑了子代集合. 这种方法可以获取进行相似性度量的概念对间更多、更具体的信息. 事实上, 不同的度量方法从不同的角度刻画了两个对象的相似性或相异性, 这可能有助于整合各种相似性度量以获得更好的结果.

我们引入了一个新的级联模型进行高效的概念相似度排名. 与以前的方法不同, 级联模型能够逐步修剪本体和细化排名顺序. 通过级联, 可以产生更高质量的

结果和更快的查询执行时间. (1) 在粗计算阶段, 我们的方法是针对整个本体使用基于路径的训练措施来获得计算过两个概念间的相似度得分; (2) 精确计算及扩展阶段, 我们使用不同的 IC 算法从不同的方面来计算每一个概念的相似度得分. 为了充分考虑概念中的可用信息, 将目标概念使用其子代和父代概念集进行扩展, 然后用于下一个阶段的训练; (3) 利用权重来平衡粗计算和精确计算的相似度得分; (4) 最后通过训练 BP 神经网络得到概念的相似性排名. 我们在石油本体模型和 Babel-net 上对本文的模型进行了测试, 实验结果表明我们的方法提高了相似度计算的准确度.

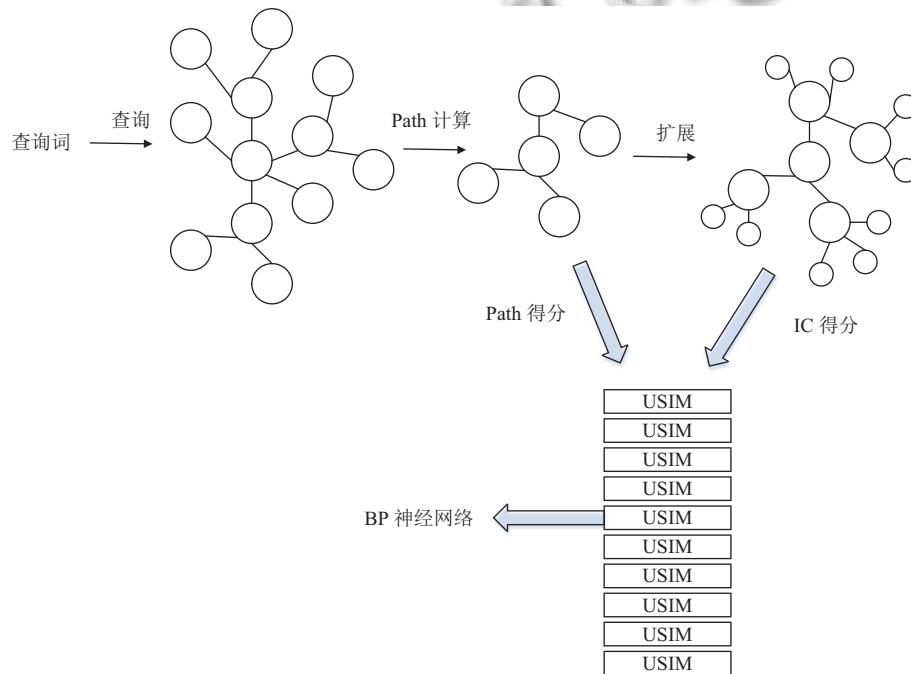


图1 概念相似度计算的模型体系结构

2 相关工作

语义相似度估计是自然语言处理 (NLP) 的一个重要组成部分, 在许多 NLP 应用中得到了成功的应用. 基于路径的方法主要考虑概念在本体中的路径距离来确定它们的语义相似度^[6]. 基于 IC 的方法主要比较所涉及的概念及其父代或子代的属性^[7]. 与基于路径的方法相比, 基于 IC 的度量对可变语义距离的问题不敏感^[8]. 此外, 基于 IC 的方法强调了本体的同一层次上的术语并不总是等价的思想, 因为它们在本体中的重要性或特异性是由它们的信息量来衡量的^[9]. Alexopoulou^[10] 提出了“Closest Sense”的方法, 该方法计算了可能的概

念语义类型与目标词周围的语义类型之间的平均最短距离. Garla^[11] 使用 Patwardhan^[12] 提出的语义关系算法对基于路径和基于分类的相似性度量进行评价. Resnik^[8] 和 Jiang^[13] 提出的语义相似度算法被广泛的使用. Rada^[14] 提出了一种基于两个目标词间最短路径的相似度计算方法. Wu^[9] 定义一个相似性度量由三部分组成: 概念在本体中的公共子集, 概念间的公共属性以及它们的最近公共父代距离.

最近, Dang^[15] 提出了一种新的两阶段学习模型. Dang 使用一组有限的特征集合, 包括加权词语、相邻性和扩展项来训练整个检索的排名函数. Wang^[16] 提出

了一种级联模型,通过逐步细化和精炼候选文档集来尽可能减少检索中的不利因素,提高检索的质量。朱新华等^[17]提出了一种综合的词语语义相似度计算方法,算法通过特殊的单调递减曲线的边权重策略,并且采用以词语距离为主要因素、分支节点数和分支间隔为微调节参数的方法,改进了现有的词林词语相似度算法。李阳等^[18]提出一种通用的实体相似度计算方法,通过清洗噪声数据,对数值、列表以及文本等不同数据类型进行预处理,使用 SVM、随机森林等集成学习模型以及排序学习模型进行建模。Pesquita^[19]考虑了影响相似度的内在和外在问题,以及如何处理这些问题,强调了不同环境下的最佳措施,并比较了不同的实施策略及其使用效果。

3 模型

在这部分,我们介绍计算概念相似度的级联模型的具体细节。模型的体系结构如图 1 所示。

3.1 粗计算阶段:基于路径的相似度算法

粗计算阶段是模型的第一阶段,我们使用基于路径的相似计算方法来计算概念间的相似度得分,然后将该得分以及路径集合作为后续阶段输入数据,以生成最终的查询结果。

Leacock^[20]最先提出了一种基于路径的本体相似度计算方法,并且被广泛的使用。假设 c_1 和 c_2 是两个概念,他们考虑了概念的最大深度,定义了公式 (1):

$$PSIM_{lch}(c_1, c_2) = -\log \frac{depth(c_1) + depth(c_2) - depth(LCS(c_1, c_2))}{2 \times max_depth} \quad (1)$$

其中 max_depth 代表本体的最大深度值。

本文考虑到不同本体领域存在许多差异性,采用 Batet^[21]提出的特定领域的相似性度量模型,它充分考虑了本体的特性以及多重继承的关系,与石油领域本体基本相似,如公式 (2) 所示:

$$PSIM_{batet}(c_1, c_2) = -\log \frac{|T(c_1) \cup T(c_2)| - |T(c_1) \cap T(c_2)|}{|T(c_1) \cup T(c_2)|} \quad (2)$$

其中 $T(c)$ 是所有本体概念 c 的集合,公共概念集 $|T(c_1) \cap T(c_2)|$ 代表两个概念的最小公共集合 LCS (Least Common Subsumer) 的深度的公共部分,非公共概念集 $|T(c_1) \cup T(c_2)| - |T(c_1) \cap T(c_2)|$ 代表 c_1 概念与 c_2 概念各自独立的最小公共集合的总和。

本文使用公式 (1) 对公式 (2) 进行了改进,如公式 (3) 所示:

$$PSIM_{batet}(c_1, c_2) = -\log \frac{depth(c_1) + depth(c_2) - 2 \times depth(LCS(c_1, c_2))}{depth(c_1) + depth(c_2) - depth(LCS(c_1, c_2))} \quad (3)$$

3.2 精确计算阶段:基于 IC 相似度计算

在精确计算阶段,本文使用基于 IC 的相似度算法去计算从扩建阶段获得的所有结果集中每个概念的得分。每个概念的得分将被作为构建成特征向量。本阶段将采用 5 种目前使用最为广泛的基于 IC 的相似度算法作为精确计算阶段的方法,先使用这 5 种去获取概念在本体上部(概念与其父代集合)的相似度得分,并且通过扩展阶段重新定义公式,来获取概念在本体下部(概念与其子代集合)的相似度得分。以下五种算法都是基于概念在本体上部的相似度计算方法。

Resnik 首先把概念信息量 (IC) 应用到相似度计算中。在 Resnik 的模型中,相似度通过两个概念的最小公共集合的 IC 进行计算,公式定义如下:

$$USIM_{res}(c_1, c_2) = IC(LCS(c_1, c_2)) \quad (4)$$

Lin^[22]在 Resnik 相似度算法的基础上,额外考虑了概念 c_1 和概念 c_2 的 IC 信息量值。模型 (Lin) 定义如下:

$$USIM_{lin} = \frac{2 \times USIM_{res}(c_1, c_2)}{IC(c_1) + IC(c_2)} \quad (5)$$

Jiang 和 Conrath 提出的模型根据公式 (6) 进行量化:

$$USIM_{jcn}(c_1, c_2) = -(IC(c_1) + IC(c_2) - 2 \times USIM_{res}(c_1, c_2)) \quad (6)$$

Sánchez 和 Batet 提出了一种新的语义相似度度量方法,定义如公式 (7) 所示:

$$USIM_{ic-lch}(c_1, c_2) = -\log \frac{IC(c_1) + IC(c_2) - USIM_{res}(c_1, c_2)}{2 \times max_IC} \quad (7)$$

通过考虑到本体中概念对之间的语义距离以及概念在本体中的深度, Wu 提出的相似度计算模型定义如公式 (8):

$$USIM_{wu}(t_1, t_2) = \frac{1}{1 + \gamma} \times \frac{\alpha_{IC}}{\alpha_{IC} + \beta_{IC}} \quad (8)$$

3.3 扩展阶段

为了能够获取概念的更多属性,以便获得更好的数据进行训练,使最终计算结果的准确性更高、更具潜力。因此,我们扩展本体概念集合以获得更多信息。

Zhang^[23]发现两个概念的共同子集同样影响相似

性得分.本文中,我们认为两个概念的下一代子集影响两个概念间的相似度得分,如图2所示.

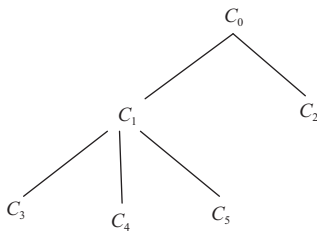


图2 概念子集图

这意味着两个概念的子代集合也可以用来描述它们的相似性度量.我们在下面的公式中重新定义子代公共部分的IC值:

$$\overline{IC}(t) = \max\left(\sum IC(x_i) + \sum IC(y_i) - IC_t\right) \quad (9)$$

其中, x_i 和 y_i 分别代表两个概念的下一级子代, $\overline{IC}(t)$ 是他们公共子代集合 t 的初始IC值.

最终, Resnik 提出的相似的计算模型就被重新定义为如下:

$$SIM(t) = \max\{\overline{IC}(t) | t \in CD(t_1, t_2)\} \quad (10)$$

其中, $CD(t_1, t_2)$ 代表 t_1 和 t_2 在本体结构下公共子集.

其余四种相似度算法以同样的方式重新定义.

3.4 相似度特征

本文把两个概念之间的相似度得分作为是每一对概念的特征值,通过构造基于本体的不同相似度模型获取相似性得分值来进一步构造特征向量来表示概念对,相似度特征表示为公式(11):

$$F_{u+l} = [USIM, LSIM] \quad (11)$$

其中, $USIM$ 和 $LSIM$ 分别表示概念对的公共父代相似度得分与公共子代相似度得分.为了研究不同相似性特征的整合对于概念对相似性计算的影响,我们构建了两种综合特征向量;第一个综合特征向量 F_u 只考虑概念对在本体中的的公共父代集合的相似性得分(上级IC相似度得分)组成相似度特征向量,第二个综合特征向量,在特征向量 F_u 的基础上,增加了概念对在本体中的子代集合的相似度得分(下级IC相似度得分)来构建相似度特征向量,如公式(12)所示.

$$F_u = [USIM_r, USIM_l, USIM_j, USIM_i, USIM_w] \quad (12)$$

其中, $USIM_r, USIM_l, USIM_j, USIM_i$ 和 $USIM_w$ 分别

表示 Resnik、Lin、Jiang、Schlicker 和 Wu 提出的基于概念对在本体中公共父代的上级相似度得分.

$$F_{u+l} = \left[\begin{array}{c} USIM_r, USIM_l, USIM_j, USIM_i, USIM_w, \\ LSIM_r, LSIM_l, LSIM_j, LSIM_i, LSIM_w \end{array} \right] \quad (13)$$

其中, $LSIM_r, LSIM_l, LSIM_j, LSIM_i$ 和 $LSIM_w$ 分别表示 Resnik、Lin、Jiang、Schlicker 和 Wu 提出的扩展的基于概念对在本体中的公共子代的下级相似度得分.

3.5 得分权重

基于路径的相似度算法只考虑到概念对在本体中的路径长度,忽略了每个概念的信息量.相反,基于IC的相似度算法只考虑到每个概念的信息量,忽略了每个概念对在本体中的路径长度.为了充分利用这两种信息,本文通过使用得分门网络来实现的,该网络为每个概念生成聚合权重,控制两种相似度算法计算的相似度得分对最终相似度得分的贡献.综合特征向量被改写为公式(14).

$$\begin{aligned} F &= [NewUSIM, NewLSIM] \\ NewUSIM &= \delta_P USIM + \delta_U USIM \\ NewLSIM &= \delta_P USIM + \delta_L LSIM \end{aligned} \quad (14)$$

其中, $\delta_P, \delta_U, \delta_L$ 分别代表基于路径得分、基于上级IC信息量得分、基于下级IC信息量得分的权重.我们对比了几组不同权重下的实验结果,发现当 δ_P 为0.3, δ_U 和 δ_L 为0.7时,实现结果最好.

3.6 BP神经网络

BP神经网络是误差反向传播算法训练的多层前馈神经网络,是目前应用最广泛的神经网络模型之一. BP网络可以学习和存储大量的输入输出映射关系,而不需要揭示描述映射关系的数学方程. BP神经网络的结构由输入层、隐层和输出层,在神经元的连接权值和阈值,输入层和输出层只有一个,它的单位数量与实际输入输出参数一致;隐层可以是一个或多个神经元参与,数字必须重复计算.由于三层神经网络具有很好的函数逼近功能,结构设计简单,运算能力强.在本文中,我们是用三层BP神经网络,如图3所示.

隐藏层节点的数目由公式(15)确定:

$$m = \sqrt{l+n+a} \quad (15)$$

其中, n 和 l 分别代表输入和输出单元的数目, m 代表隐藏节点的数目,通常情况下的取值在1到10的范围内.我们使用 Sigmoid 函数作为激活函数:

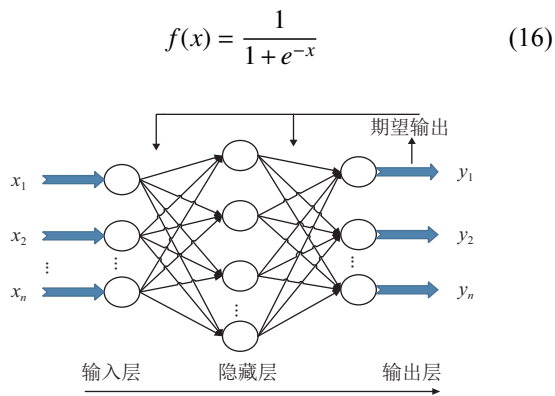


图3 三层 BP 神经网络

训练 BP 神经网络时,我们将最终的相似度特征向量作为 BP 神经网络的输入,将两个概念对应的相似度分类作为输出。

对于输出结果,我们期望 BP 神经网络将两个概念的相似性分数划分为 10 个类别,0 个代表不相似,9 个代表极其相似(相同),值越高,相似度越高。

4 实验评估

4.1 数据集

在本文中,石油本体和 BabelNet^[24]用来作为数据集测试本文的方法。我们收集了石油领域的数据,建立了自己的石油领域本体。BabelNet 是一个多语言语义网络,其概念和关系是利用自动映射算法从英语中最大的有效语义词典 WordNet^[25]和最大的覆盖面的百科全书 Wikipedia 中获得的。

4.2 评判基准

为了得到一个比较基准,我们的实验标准类似于 Miller 和 Charlest 提出的基准,给予五十名受试者(四十名石油工程学生和十名石油领域教授)相同的 330 个名词对(30 个名词对与 Miller 和 Charles 相同,300 个名词对是关于石油的)。这些操作和 Miller 和 Charles 的用法完全一样。一半的受试者以随机顺序收到单词对的列表,另一半接受固定顺序的列表。对于同一概念对,Miller 和 Charles 基准的平均评分和本文中评价基准的平均评分之间的相关程度为 95%,说明我们的基准是有效的。

4.3 实验结果

为了验证所提出方法的有效性,本研究采用十倍交叉验证,并以精确度作为验证指标。在十次交叉验证中,数据集被随机划分为十个相等的子集,验证过程重

复十次。每次保留其中一个子集作为验证数据,其余四个子集作为训练数据。交叉验证的十个结果的平均值产生一个总体估计。使用公式 (17) 表示的准确率和召回率来判断实验结果,

$$Precision = \frac{TP}{TP + FP} \quad (17)$$

表 1 和表 2 列出了实验的预测结果特点,在石油本体和 babelnet 上采用了不同的相似性度量策略。从表格中,我们可以看出,采用基于路径、基于概念父代、子代扩展集合的信息量相似度计算方法相结合的特征通常比只采用其中一种策略对应的综合特征具有更高的精度。

表 1 不同相似性算法在石油本体中的精度得分

方法	出处	类型	精确度
Resnik	Resnik ^[8]	IC-based	0.7887
Lin	Lin ^[22]	IC-based	0.8021
Jiang and Conrath	Jiang ^[13]	IC-based	0.8267
Sánchez and Batet	Sánchez and Batet ^[21]	IC-based	0.8312
Wu	Wu ^[9]	IC-based	0.8511
F_{p+u}	本文	Vector-based	0.8837
F_{p+u+l}	本文	Vector-based	0.9044

表 2 不同相似性算法在 BabelNet 中的精度得分

方法	出处	类型	精确度
Resnik	Resnik ^[8]	IC-based	0.8391
Lin	Lin ^[22]	IC-based	0.8435
Jiang and Conrath	Jiang ^[13]	IC-based	0.8675
Sánchez and Batet	Sánchez and Batet ^[21]	IC-based	0.8742
Wu	Wu ^[9]	IC-based	0.9081
F_{p+u}	本文	Vector-based	0.9191
F_{p+u+l}	本文	Vector-based	0.9233

从表 2 中的结果可以得出几个结论。本文中的相似度计算方法在石油本体和 BabelNet 获得最高的精确度得分,分别为 90.44% 和 92.33%。这表明我们从本体中获得的信息越多,分类器的预测能力就越好。本文中的算法在不同的本体上表现出良好的性能。

5 结论

在本文的研究中,我们提出了一种基于石油本体的概念对语义相似性的计算方法。将不同相似性测度得到的信息作为 BP 神经网络的输入。我们相信,基于路径的、基于 IC 和扩展的基于 IC 的测度的更多信息可以提高预测性能,我们的方法的缺点是计算量大,预

测性能会受到 BP 神经网络的影响。将来, 我们计划采用一种更全面的方法来预测两个概念的相似性。

参考文献

- 1 Lavanya S, Arya SS. An approach for measuring semantic similarity between words using SVM and LS-SVM. Proceedings of the 2012 International Conference on Computer Communication and Informatics. Coimbatore, India. 2012. 1–4.
- 2 McInnes BT, Pedersen T. Evaluating measures of semantic similarity and relatedness to disambiguate terms in biomedical text. Journal of Biomedical Informatics, 2013, 46(6): 1116–1124. [doi: 10.1016/j.jbi.2013.08.008]
- 3 Hassan H, Hassan A, Emam O. Unsupervised information extraction approach using graph mutual reinforcement. Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Sydney, Australia. 2006. 501–508.
- 4 Corley C, Mihalcea R. Measuring the semantic similarity of texts. Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment. Ann Arbor, MI, USA. 2005. 13–18.
- 5 Huang AN. Similarity measures for text document clustering. Proceedings of the New Zealand Computer Science Research Student Conference 2008. Christchurch, New Zealand. 2008.
- 6 Wu HW, Su ZC, Mao FL, *et al.* Prediction of functional modules based on comparative genome analysis and Gene Ontology application. Nucleic Acids Research, 2005, 33(9): 2822–2837. [doi: 10.1093/nar/gki573]
- 7 Sánchez D, Batet M, Isern D. Ontology-based information content computation. Knowledge-Based Systems, 2011, 24(2): 297–303.
- 8 Resnik P. Using information content to evaluate semantic similarity in a taxonomy. Proceedings of the 14th International Joint Conference on Artificial Intelligence. Montreal, Quebec, Canada. 1995, 1. 448–453.
- 9 Wu XM, Pang EL, Lin K, *et al.* Improving the measurement of semantic similarity between gene ontology terms and gene products: Insights from an edge- and IC-based hybrid method. PLoS One, 2013, 8(5): e66745. [doi: 10.1371/journal.pone.0066745]
- 10 Alexopoulou D, Andreopoulos B, Dietze H, *et al.* Biomedical word sense disambiguation with ontologies and metadata: Automation meets accuracy. BMC Bioinformatics, 2009, 10: 28. [doi: 10.1186/1471-2105-10-28]
- 11 Garla VN, Brandt C. Knowledge-based biomedical word sense disambiguation: An evaluation and application to clinical document classification. Journal of the American Medical Informatics Association, 2013, 20(5): 882–886. [doi: 10.1136/amiajnl-2012-001350]
- 12 Patwardhan S, Banerjee S, Pedersen T. Using measures of semantic relatedness for word sense disambiguation. Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Text Processing. Mexico City, Mexico. 2003. 241–257.
- 13 Jiang JJ, Conrath DW. Semantic similarity based on corpus statistics and lexical taxonomy. arXiv:cmp-lg/9709008, 1997. 11512–0.
- 14 Rada R, Mili H, Bicknell E, *et al.* Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man, and Cybernetics, 1989, 19(1): 17–30. [doi: 10.1109/21.24528]
- 15 Dang V, Bendersky M, Croft WB. Two-stage learning to rank for information retrieval. In: Serdyukov P, Braslavski P, Kuznetsov SO, *et al.*, eds. Advances in Information Retrieval. Berlin Heidelberg: Springer, 2013: 423–434.
- 16 Wang LD, Lin J, Metzler D. A cascade ranking model for efficient ranked retrieval. Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. Beijing, China. 2011. 105–114.
- 17 朱新华, 马润聪, 孙柳, 等. 基于知网与词林的词语语义相似度计算. 中文信息学报, 2016, 30(4): 29–36.
- 18 李阳, 高大启. 知识图谱中实体相似度计算研究. 中文信息学报, 2017, 31(1): 140–146.
- 19 Pesquita C. Semantic similarity in the gene ontology. In: Dessimoz C, Škunca N, eds. The Gene Ontology Handbook. New York: Humana Press, 2017. 161–173.
- 20 Leacock C, Chodorow M. Combining local context and wordnet similarity for word sense identification. In: Fellbaum C, ed. WordNet: An Electronic Lexical Database. Cambridge: MIT, 1998. 265–283.
- 21 Batet M, Sánchez D, Valls A. An ontology-based measure to compute semantic similarity in biomedicine. Journal of Biomedical Informatics, 2011, 44(1): 118.
- 22 Lin DK. An information-theoretic definition of similarity. Proceedings of the 15th International Conference on Machine Learning. San Francisco, CA, USA. 1998. 296–304.
- 23 Zhang SB, Lai JH. Semantic similarity measurement between gene ontology terms based on exclusively inherited shared information. Gene, 2014, 558(1): 108.
- 24 Navigli R, Ponzetto SP. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence, 2012, 193: 217–250. [doi: 10.1016/j.artint.2012.07.001]
- 25 Fellbaum C. WordNet: An electronic lexical database. Cambridge: MIT Press, 1998.