《智能信息处理》课程考试

基于本体的信息内容计算

姜楠

考核	到课[10]	作业[20]	考试[70]	课程成绩[100]
得分				

2021年12月20日

基于本体的信息内容计算

姜楠

（大连海事大学信息科学技术学院 辽宁大连）

摘要: 一个概念的信息内容(IC)提供了其概括性/具体性程度的估计，这是一个能够更好地理解概念语义的维度。因此，IC已成功地应用于概念间语义相似度的自动评估。过去IC被估计为概念在语料库中出现的概率。然而，由于语料库的依赖性和数据的稀疏性，该方法的适用性和可扩展性受到了限制。最近，一些作者提出了基于IC的度量方法，使用从本体中提取的分类特征对特定概念进行度量，获得了很有前景的结果。在本文中，我们分析了这些基于本体的IC计算方法，并提出了一些改进，旨在更好地捕获特定概念在本体中建模的语义证据。在应用于语义相似度估计任务时，我们的方法已经与相关的工作(包括语料库和基于本体的工作)进行了评估和比较。一个广泛使用的基准测试的结果表明，我们的方法能够实现相似性估计，这比相关工作更好地与人类判断相关。

关键字:IC；概念语义；基于本体

Ontology-based information content computation

Jiang Nan

(School of Information Science and Engineering, Dalian Maritime University, Dalian China)

Abstract: The information content (IC) of a concept provides an estimation of its degree of generality/concreteness,a dimension which enables a better understanding of concept’s semantics. As a result, IC has been suc-cessfully applied to the automatic assessment of the semantic similarity between concepts. In the past,IC has been estimated as the probability of appearance of concepts in corpora. However, the applicabilityand scalability of this method are hampered due to corpora dependency and data sparseness. Morerecently, some authors proposed IC-based measures using taxonomical features extracted from an ontol-ogy for a particular concept, obtaining promising results. In this paper, we analyse these ontology-basedapproaches for IC computation and propose several improvements aimed to better capture the semanticevidence modelled in the ontology for the particular concept. Our approach has been evaluated and com-pared with related works (both corpora and ontology-based ones) when applied to the task of semanticsimilarity estimation. Results obtained for a widely used benchmark show that our method enables sim-ilarity estimations which are better correlated with human judgements than related works.

Keyword:IC；conceptual semantics；ontology-based

1 引言

概念的IC(information content)是计算语言学的一个基本维度。它表示概念出现在上下文中时所提供的信息量。其基本思想是，一般和抽象的实体在话语中比更具体和专业的实体呈现更少的集成。对概念的集成度进行适当的量化，可以通过评估与这些概念相关的词的语义概括性或具体性的程度来提高文本的理解。事实上，IC在过去已经被应用于语义相似度的计算[11,15,18,26,28]，它是人类组织和分类物体的基本原则。

概念之间的语义相似度，理解为它们的分类相似度[13]，有许多直接和相关的应用。一些基本的自然语言处理任务，如词义消歧[22]、同义词检测[18]或自动拼写错误检测和纠正[5]依赖于单词语义相似度的评估。在知识管理领域可以找到其他直接的应用程序，例如同义词典生成[6]，信息抽取[1,37]，语义注释[34]和本体合并[12]和学习[32,35]，在这个过程中，应该从文本中发现或获得与已有概念相关的新概念。语义相似度也被应用于形式化概念分析[11]、结构化资源聚类[20]、问答[38]以及推荐系统[2]和多智能体系统[8]等领域。

在过去的几年里，已经有一些研究提出了语义相似度的度量方法。他们的目标是自动评估一个数值分数，该分数估计了一对概念之间的相似程度，作为在一个或几个知识来源中观察到的语义证据的函数。根据具体的知识资源开发和它们的使用方式，可以识别出不同的方法家族。简而言之，edgecounting度量方法基于将给定本体中包含的两个概念分开的分类链接的数量来进行相似性评估[16,17,27,40]。基于特征的方法根据共同特征和非共同特征的数量来估计相似度。通过特征，他们通常考虑在本体中建模的分类信息和从词典中检索到的概念描述(例如，注释)[25,29,39]。信息理论方法评估概念之间的相似性作为一个功能的IC，这两个概念在一个给定的本体中有共同的。IC通常从文本语料库中的概念分布计算[15,18,28]。最后，分布式方法只使用文本语料库作为源。他们推断语义是单词同现的功能。通过语料库，Web由于其大小、通用性和搜索引擎的可用性而被典型地利用，这些搜索引擎被查询以检索资源[4,7,31]。

本文提出了一种仅依靠本体论知识计算集成电路的新方法，旨在更好地捕捉概念的通用性和具体性。使用广泛使用的基准[19]，将这种方法与一组常用的相似度度量进行了比较。结果表明，该方法能够提供比相关工作更准确的人类判断相似度评价。

本文的其余部分组织如下。第二部分介绍了在文献中发现的主要基于IC的相似性度量。第三部分分析和讨论了相关工作中概念集成电路的计算方法，重点讨论了基于本体论的方法。第四部分讨论了在提出的IC计算模型中发现的一些问题，并提出了新的方法。第5节详细介绍了我们的方法的评估和比较，第6节讨论了结果。最后一部分包括结论和一些未来的研究方向。

2 基于信息内容的相似性度量

第一个将信息理论应用于语义相似度计算的工作是由Resnik[28]提出的，他指出概念相似度取决于它们之间共享信息的数量。通过利用背景本体，Resnik假定概念和bis之间的公共信息，这些信息由它们所属的分类法中包含的最小共同子集(Least common Subsumer, LCS)的IC(即包含两个概念的最特定的共同祖先)所代表。如果这两个概念没

有联系，LCS也不存在，那么它们就被认为是最大不同的。否则，它们的语义相似度计算为LCS提供的IC数量(1)：

$$Sim_{res}(a,b) = IC(LCS(a,b)) \quad (1)$$

Resnik度量的一个问题是，任何具有相同LCS的概念对都会产生完全相同的语义相似度。为了解决这个问题，Lin[18]、Jiang和Conrath[15]通过考虑每个评估概念的IC扩展了Resnik的工作。

林宣布两个概念之间的相似性应该被衡量为陈述它们的共性所需的信息量和充分描述它们所需的信息量之间的比率。作为这个定理的一个推论，他的测量方法一方面以与Resnik方法相同的方式考虑共性，另一方面，仅考虑每个概念的IC (2)：

$$sim_{lim}(a,b) = \frac{2\times sim_{res}(a,b)}{(IC(a)+IC(b))} \quad (2)$$

Jiang和Conrath[15]提出的措施是基于量化分类链接的长度，作为一个概念的IC和它的从属者之间的差异。当比较概念对时，他们通过从LCS的IC中减去每个概念的IC之和来计算它们的距离(3)：

$$dis_{j\&c}(a,b) = (IC(a)+IC(b)) - 2\times sim_{res}(a,b) \quad (3)$$

作者通常利用WordNet[10]作为本体。WordNet是一个领域独立的和通用目的的同义词典，它描述和组织了超过100,000个通用英语概念，这些概念以本体论的方式在语义上结构化。需要注意的是，语义相似度通常应用于未消除歧义的词对(w1和w2)，这些词对可能对应于WordNet中的几个概念(即多义词词)，作者计算相似度作为任何词义对组合获得的最大相似度(4)：

$$sim(w_1,w_2) = \max_{(i,j)}(sim(s_{1i},s_{2j})) \quad (4)$$

3 IC计算

为了通过上述措施提供准确的结果，计算IC的方式是至关重要的。经典信息论方法[15,18,28]通过计算一个概念在语料库中出现概率的逆，得到该概念的IC (p(a))(5)。在这种情况下，不常用的术语被认为比常用术语更能提供信息。

$$IC_{(a)} = -log\ p(a) \quad (5)$$

需要注意的是，为了正确地表现，基于IC的度量需要IC值随着分类法中向下移动而单调增加(例如，8a,b|a是b =>IC(a)6IC(b)的上名)。在基于语料库的IC评估中，它是通过计算p(a)作为在给定语料库中遇到它或它的任何分类上的次词的概率来实现的。在实践中，任何名词在语料库中的每一个单独出现都被递归计算为它的每一个分类祖先的出现(6)[28]：

$$p(a) = \frac{\sum_{w\in W(a)} count(w)}{N} \quad (6)$$

其中w (a)为语料库中其意义被a所包含的术语集合，为分类法中包含的语料库术语总数。

由于文本语料库中包含单词和本体模型概念，为了准确计算概念出现概率，需要在语料库中识别词的语义。这就需要对语料库中出现的每个名词进行适当的消歧和注释。如果分类法或语料库发生了变化，就需要对受影响的概念进行递归的重新计算。因此，有必要对文本进行手工和耗时的分析，结果的概率将取决于输入语料库的大小和性质。此外，背景分类必须尽可能完整(即，它应该包括语料库中涉及的每个概念的大多数专门化)，以便在概念级别上提供可靠的结果。最后，语料库的内容要在本体范围上足够大，避免数据稀疏。大型和通用的语料库(如Brown corpus1)可能适合于WordNet[15]，但是涉及具体术语[24]的领域本体可能需要更具体的语料库。由于人工标记和语料库依赖性

和可用性而引起的可伸缩性问题阻碍了这些IC计算模型[33]的适用性。

Seco等人。[36]是第一个基于概念下位词的数量进行IC计算的方法。低(a)分类树的数量下概念a和max_nodes最大概念的分类，他们计算一个概念在以下方式（7）：

$$IC_{seco\_et\_al}(a) = \frac{\log(\frac{hypo(a)+1}{max\_nodes})}{\log(\frac{1}{max\_nodes})} = 1 - \frac{\log(hypo(a)+1)}{\log(max\_nodes)} \quad (7)$$

分母（对应于信息最丰富的概念：一片叶子）产生的IC值在0到1范围内的标准化。注意，分子还将概念本身视为下义词，以避免当a是叶时使用log（0）。这种方法只考虑分类法中给定概念的下同义词。因此，具有相同数量的下位词但不同程度的概念（即，一个相对于另一个出现在层次的上层）将是同样相似的。为了解决这个问题，周等人。[41]提出用分类法（8）中概念的相对深度来补充基于下位点的IC计算：

$$IC_{zhou\_et\_al}(a) = k(1 - \frac{\log(hypo(a)+1)}{log(max\_nodes)}) + (1-k)(\frac{\log(depth(a))}{log(max\_depth)}) \quad (8)$$

同样，WordNet是一个大型的、连贯的、详细的本体，具有相同的建模层次结构，它代表了一个很好的来源，其中内在IC计算的基础是[36]。事实上，正如它将在评估部分所显示的那样，在WordNet上的内在IC计算比基于语料库的方法会导致更好的相似性评估。

4. 一种基于本体的IC计算的新方法

尽管内在的IC计算已经导致了准确的评估，但我们认为仍有改进的空间。在本节中，我们将分析和讨论一些可能更好地捕获本体中建模的语义证据的策略。通过这一分析，我们提出了一种新的内在IC计算模型，与相关工作相比，它包含了一些改进。

第一个方面是，内在的IC计算模型依赖于一个概念的整个下对称树的大小。当建模大量具体概念时，知识专家根据它们的共同特征（以自下而上的方式）[14]，逐步将它们组织成集群（具有共同祖先的概念集）。随着每一个泛化，新的更抽象的内部概念作为公共包容者被引入到分类法中。因此，给定一组固定的具体概念来定义域的范围（即分类法的叶子），概念下义树的大小将取决于建模概念的细节程度、分支因素和内部结构的粒度。由于这种知识建模过程的结果，在某些情况下，一个包含者可能对应于一个特别的抽象(例如，。物理实体、抽象实体和事物都是WordNet)中根节点实体的下义词，很少出现在话语中。在WordNet中，大约21%的概念数量对应于内部分类节点[9]。这种层次的内部分类细节可能因一个本体而不同，甚至在同一本体中从一个分类树到另一个也不同。这影响了依赖于下马尾辫树的总大小的内在IC计算的一致性。另一方面，分类叶代表了一个领域中最特定的概念的语义。因此，属于一个域的叶子集可以准确地定义它的范围。我们认为，一个概念的下义词树的叶子足以描述和区分该概念与任何其他概念（具有不同的叶子集），而不管在分类法中包含的内部概念的数量。以此论证为前提，为了避免在评估概念的通用性时依赖于层次结构的内部细节，我们建议只考虑概念的下义树的叶子作为其IC的指示。在形式上，我们将一个概念的叶子定义为：

定义1：假设C是本体论的概念集，我们将概念a的叶集定义为：

$$leaves(a) = \{l \in c | l \in hyponyms(a) \wedge l \text{ is a leaf}\}$$

l为叶，如果$hyponyms(l \in \emptyset)$.

关于叶计数，一个概念的下义树的一些内部节点存在多重分类遗传，这可能导致从该概念到一个叶存在多条路径。为了避免冗余，在创建leaves(a)时考虑一次。

根据与相关著作类似的原则，我们认为

下义树中有许多叶子的概念是一般的(即它们的IC较低)，因为它们包含了许多显著术语的含义。另一方面，叶子将呈现同样的最大IC，因为它们是完全区别（即，不进一步专门）的任何其他概念。

以IC出现的概率为理论公式。（5）），并为适应这一前提，我们提出以下表达式：

定义2，一个概念a的IC被定义为

$$IC(a) = -\log(a) \cong -\log\left(\frac{|leaves(a)|+1}{max\_leaves+1}\right) \quad (9)$$

其中，max_leaves表示与层次结构的根节点对应的叶数，它作为一个规范化因子。我们在分子中添加1，以避免在计算叶时的log（0）值。

从另一个的角度来看，一个分类单元中一个概念的深度实际上对应于其分类分类单元的数量（当不考虑多重遗传时）。所以，桑切斯等人的体积越大。/基于知识的系统24（2011）297-303 299 数量包含的概念超过一个给定的概念，其具体程度越高，因为它是许多专门化的结果。在形式上，我们用以下方式来定义概念包容：

定义3：一个概念的IC被定义为：

$$IC(a) = -\log\frac{\frac{|leaves(a)|}{|subsumers(a)|}+1}{max\_leaves(a)+1} \quad (10)$$

表达式（10）是根据最一般的概念（根节点）的维数进行标准化的，对于|leaves(root_node)|=max_leaves和|subsumbers(root_node)|=1来说。

5. 评价

为了评估所提出的IC计算方法的行为，并将其与相关工作（基于内在和基于语料库的工作）进行比较，我们将不同的IC计算范式应用于第2节中引入的经典基于IC的度量，在相似性评估中评估它们的准确性。然而，正如博莱加拉等人所述。[4]，由于语义相似函数的概念是主观的，因此一个客观的评价语义相似函数的准确性是困难的。为了进行公平的比较，我们采用了Miller和查尔斯[19]基准，它由30对英语名词对（取自[30]）组成，它们的相似性由38名学生从0（语义无关）到4（高度同义）进行了评估。相似性评估的准确性通过计算人类判断和计算机测量提供的结果之间的相关性来量化相似性评估的准确性。这使得对不同的相似性度量和集成电路计算范式的客观评估成为可能。

6. 结语

概念IC的概念在过去已经在一些评估语义相似性[11,15,17,28]的应用中被成功地利用。在这些情况下，从知识资源中进行准确的IC估计对于提供可靠的相似性评估至关重要。本文提出了一种基于本体中建模的知识计算集成电路的新方法。它的设计考虑并解决了相关工作中确定的一些问题，旨在更好地捕获关于作为其分类特征的功能在本体中隐含建模的概念的具体性/通用性的证据。我们的方法避免了依赖于调优参数的一个事实，即由于缺乏公司相关性，配置了一种通用的、可扩展的和易于适用的方法，以一种内在的方式计算IC。

该评估基于一个广泛使用的基准，并应用于经典的基于IC的相似性度量，维持了该方法的直觉，其与人类判断的相关性在测试中最高，非常接近人类之间的评级协议。一方面，它能够改进其他固有的集成电路计算方法。另一方面，当应用于基于IC的度量时，它配置了一种纯的基于本体的方法，而不依赖于外部资源、手动标记或加权参数。

作为未来的工作，我们计划研究该方法与特定领域的其他本体(如MeSH或SNOMED)的行为。我们还考虑了组合由几个重叠的本体所提供的知识的可能性。这将捕获可能有助于改进评估的其他语义证据。

参考文献：

（1） J. Atkinson, A. Ferreira, E. Aravena, Discovering implicit intention-level knowledge from natural-language texts, Knowl.-Based Syst. 22 (7) (2009) 502–508.

（2） Y. Blanco-Fernández, J.J. Pazos-Arias, A. Gil-Solla, M. Ramos-Cabrer, M. López- Nores, J. García-Duque, A. Fernández-Vilas, R.P. Díaz-Redondo, J. Bermejo- Muñoz, A flexible semantic inference methodology to reason about user preferences in knowledge-based recommender systems, Knowl.-Based Syst. 21 (4) (2008) 305–320.

（3） A. Blank, Words and Concepts in Time: Towards Diachronic Cognitive Onomasiology, in: R. Eckardt, K. von Heusinger, C. Schwarze (Eds.), Words and Concepts in Time: towards Diachronic Cognitive Onomasiology, Mouton de Gruyter, Berlin, Germany, 2003, pp. 37–66.

（4） D. Bollegala, Y. Matsuo, M. Ishizuka, Measuring semantic similarity between words using web search engines, in: Proc. of 16th international conference on

World Wide Web, WWW 2007, ACM Press, Banff, Alberta, Canada, 2007, pp. 757–766.

（5） A. Budanitsky, G. Hirst, Semantic distance in WordNet: an experimental, application-oriented evaluation of five measures, in: Proc. of Workshop on WordNet and Other Lexical Resources, Second meeting of the North American Chapter of the Association for Computational Linguistics. Pittsburgh, USA, 2001, pp. 10–15.

（6） J.R. Curran, Ensemble methods for automatic thesaurus extraction, in: Proc. of Empirical Methods in Natural Language Processing, EMNLP 2002. Association for Computational Linguistics, Philadelphia, PA, USA, 2002, pp. 222–229.

（7） H.-H. Chen, M.-S. Lin, Y.-C. Wei, Novel association measures using web search with double checking, in: Proc. of 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, COLING-ACL 2006. ACL, Sydney, Australia, 2006, pp. 1009–1016.

（8） J. Debenham, C. Sierra, Merging intelligent agency and the Semantic Web, Knowl.-Based Syst. 21 (3) (2008) 184–191.

（9） A. Devitt, C. Vogel, The topology of WordNet: some metrics, in: Proc. of 2nd Global Wordnet Conference, GWC 2004, Masaryk University, Brno, Czech Republic, 2004, pp. 106–111.

（10） C. Fellbaum, WordNet: An Electronic Lexical Database, MIT Press, Cambridge, Massachusetts, 1998.

（11） A. Formica, Concept similarity in formal concept analysis: an information content approach, Knowl.-Based Syst. 21 (1) (2008) 80–87.

（12） M. Gaeta, F. Orciuoli, P. Ritrovato, Advanced ontology management system for personalised e-Learning, Knowl.-Based Syst. 22 (4) (2009) 292–301.

（13） R.L. Goldstone, Similarity, interactive activation, and mapping, J. Exp. Psychol. Learn. Mem. Cogn. 20 (1) (1994) 3–28.

（14） A. Gómez-Pérez, M. Fernández-López, O. Corcho, Ontological Engineering, second ed., Springer-Verlag, 2004.

（15） J.J. Jiang, D.W. Conrath, Semantic similarity based on corpus statistics and lexical taxonomy, in: Proc. of International Conference on Research in Computational Linguistics, ROCLING X. Taipei, Taiwan, 1997, pp. 19–33.

（16） C. Leacock, M. Chodorow, Combining Local Context and WordNet Similarity for Word Sense Identification, WordNet: An Electronic Lexical Database, MIT Press, 1998. pp. 265–283.

（17） Y. Li, Z. Bandar, D. McLean, An approach for measuring semantic similarity between words using multiple information sources, IEEE Trans. Knowl. Data. Eng. 15 (4) (2003) 871–882.

（18） D. Lin, An information-theoretic definition of similarity, in: Proc. of Fifteenth International Conference on Machine Learning, ICML 1998, Morgan Kaufmann, Madison, Wisconsin, USA, 1998, pp. 296–304.

（19） G.A. Miller, W.G. Charles, Contextual correlates of semantic similarity, Lang. Cogn. Process. 6 (1) (1991) 1–28.

（20） R. Nayak, W. Iryadi, XML schema clustering with semantic and hierarchical similarity measures, Knowl.-Based Syst. 20 (4) (2007) 336–349.

（21） B. Partee, A. ter Meulen, R. Wall, Mathematical Methods in Linguistics, Kluwer Academic Publisher, 1990.

（22） S. Patwardhan, S. Banerjee, T. Pedersen, Using measures of semantic relatedness for word sense disambiguation, in: Proc. of 4th International Conference on Computational Linguistics and Intelligent Text Processing and Computational Linguistics, CICLing 2003, Springer Berlin/Heidelberg, Mexico City, Mexico, 2003, pp. 241–257.

（23） S. Patwardhan, T. Pedersen, Using WordNet-based context vectors to estimate the semantic relatedness of concepts, in: Proc. of EACL 2006 Workshop on

Making Sense of Sense: Bringing Computational Linguistics and Psycholinguistics Together, Trento, Italy, 2006, pp. 1–8.

（24） T. Pedersen, S. Pakhomov, S. Patwardhan, C. Chute, Measures of semantic similarity and relatedness in the biomedical domain, J. Biomed. Inform. 40 (3) (2007) 288–299.

（25） E.G.M. Petrakis, G. Varelas, A. Hliaoutakis, P. Raftopoulou, X-similarity: computing semantic similarity between concepts from different ontologies, J. Digit. Inf. Manage. 4 (2006) 233–237.

（26） G. Pirró, A semantic similarity metric combining features and intrinsic information content, Data Knowl. Eng. 68 (11) (2009) 1289–1308.

（27） R. Rada, H. Mili, E. Bichnell, M. Blettner, Development and application of a metric on semantic nets, IEEE Trans. Syst. Man Cybern. 9 (1) (1989) 17–30.

（28） P. Resnik, Using information content to evaluate semantic similarity in a taxonomy, in: Proc. of 14th International Joint Conference on Artificial Intelligence, IJCAI 1995, Morgan Kaufmann Publishers Inc., Montreal, Quebec, Canada, 1995, pp. 448–453.

（29） M.A. Rodríguez, M.J. Egenhofer, Determining semantic similarity among entity classes from different ontologies, IEEE Trans. Knowl. Data. Eng. 15 (2) (2003) 442–456.

（30） H. Rubenstein, J. Goodenough, Contextual correlates of synonymy, Commun. ACM 8 (10) (1965) 627–633.

（31） M. Sahami, T.D. Heilman, A web-based kernel function for measuring the similarity of short text snippets, in: Proc. of 15th International World Wide Web Conference, WWW 2006, ACM Press, Edinburgh, Scotland, 2006, pp. 377– 386.

（32） D. Sánchez, A methodology to learn ontological attributes from the Web, Data Knowl. Eng. 69 (6) (2010) 573–597.

（33） D. Sánchez, M. Batet, A. Valls, K. Gibert, Ontology-driven web-based semantic similarity, J. Intell. Inf. Syst. (2009) doi:10.1007/s10844-009-0103-x.

（34） D. Sánchez, D. Isern, M. Millán, Content annotation for the Semantic Web: an automatic web-based approach, Knowl. Inf. Syst. (2010) doi:10.1007/s10115- 010-0302-3.

（35） D. Sánchez, A. Moreno, Learning non-taxonomic relationships from web documents for domain ontology construction, Data Knowl. Eng. 63 (3) (2008) 600–623.

（36） N. Seco, T. Veale, J. Hayes, An intrinsic information content metric for semantic similarity in WordNet, in: Proc. of 16th European Conference on Artificial Intelligence, ECAI 2004, including Prestigious Applicants of Intelligent Systems, PAIS 2004, IOS Press, Valencia, Spain, 2004, pp. 1089–1090.

（37） M. Stevenson, M.A. Greenwood, A semantic approach to IE pattern induction, in: Proc. of 43rd Annual Meeting on Association for Computational Linguistics, COLING-ACL 2005, Association for Computational Linguistics, Ann Arbor, Michigan, USA, 2005, pp. 379–386.

（38） A.G. Tapeh, M. Rahgozar, A knowledge-based question answering system for B2C eCommerce, Knowl.-Based Syst. 21 (8) (2008) 946–950.

（39） A. Tversky, Features of similarity, Psycol. Rev. 84 (4) (1977) 327–352.

（40） Z. Wu, M. Palmer, Verb semantics and lexical selection, in: Proc. of 32nd annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Las Cruces, New Mexico, 1994, pp. 133–138.

（41） Z. Zhou, Y. Wang, J. Gu, A new model of information content for semantic similarity in WordNet, in: Proc. of Second International Conference on Future Generation Communication and Networking Symposia, FGCNS 2008, IEEE Computer Society, Sanya, Hainan Island, China, 2008, pp. 85–89.