

# 基于知网概念多层结构和义原复修饰关系的 语义相似度计算方法

刘江鸣<sup>1</sup> 徐金安<sup>1†</sup> 张玉洁<sup>1</sup>

1. 北京交通大学计算机与信息技术学院, 北京 100044

†通信作者, E-mail:xja2010@gmail.com

## 摘要

为了使词语语义相似度计算更加贴近日常习惯, 本文提出了融合 HowNet 概念多层结构和义原复修饰关系的词语语义相似度计算方法。分析利用 HowNet 概念描述结构。概念多层结构化是指概念由义原以树状多层结构修饰, 义原复修饰关系是指义原由义原再修饰并细化与其上位义原的关系。因此本文使用上述 2 种结构化信息, 同时通过导入同义词词典来扩大义原的对义和反义关系的计算方法。实验结果表明, 与传统方法相比, 本文提出的语义相似度计算方法有效稳定, 效果良好。

## 1 介绍

在认知科学中, 语义计算是其必不可少的一部分。词语语义相似度计算是最根本和最直接的语义计算表现方式。由于词语语义相似度是主观性比较大的概念, 词语之间的关系十分复杂, 很难得到统一的标准。因此目前词语语义相似度主要通过日常生活中人们的认识作为标准。词语语义相似度在许多研究中广泛应用, Varelas (2005) 将语义相似度用于信息检索、杨超(2010)将语义相似度用于情感计算、M. Mohler(2009)将语义相似度用于简短问答系统, 此外在分词、句法分析、机器翻译等方面语义相似度计算也被广泛使用。由此看来词语语义相似度的研究越发得到重视。

国内外的语义相似度计算研究中, 利用的资源主要源于两个方面: 来自于网络或者语料库中

大规模的文本描述信息, D. Bollegala (2007)使用 web 搜索引擎获取网络中的文本信息用于语义相似度计算, Rohde(2009)基于语料库构造语义模型, S. A. Takale (2010)使用 web 中的大量的相关文档计算语义相似度; 来自于知识概念网络, Richardson(1994)、Varelas(2005) 和 Guti rrez,Y(2011)提出并使用 WordNet 计算英语语义相似度, Guan, Y (2002)、刘群(2002)和江敏(2008)提出并使用 HowNet 计算汉语语义相似度。语义相似度计算普遍所使用的计算方法主要有: M. Ruiz-Casado (2005)提出基于信息论的计算方法、Y. Li (2003)和 C. d'Amato (2009)提出基于逻辑描述的计算方法、另外还有基于语义距离的计算方法, 所使用的模型包括 DEERWESTER(1990)提出的 Latent Semantic Analysis(LSA)、Lund(1996)提出的 Hyperspace Analogue to Language(HAL), 所利用的资源包括 George(1995)介绍的 WordNet 和董振东(2008)提出最新的 HowNet。总结前人的工作, 本文将语义相似度计算主要的任务归结为: 构建完善且有效的知识库、提出基于知识库的语义计算方法和利用语料库或网络资源作为辅助用以解决知识库的局限性。

由于汉语中语法及其所描述的方式复杂多变且歧义严重。目前, 基于知识网络——HowNet 的语义相似度计算方法成为汉语语义相似度计算的主流方法。HowNet 是一个以汉语和英语的词语所代表的概念为描述对象, 以揭示概念与概念之间以及概念所具有的属性之间的关系为基本内容的知识库。HowNet 语义表现力十分丰富,

引起了广泛的关注。Huang (2008)提出构建中文词网方法并提出接口及应用, Xuexia(2011)提出基于 HowNet 并考虑义原子树结构的词语相似度计算模型等。

本文提出的语义相似度计算与传统方法相比,不同之处在于,提出方法在计算义原相似度时,利用了义原与义原之间的复修饰信息;在计算概念相似度时,采用自底向上,逐层计算义原在层次中的相似度,最终得到概念的相似度。因此,计算结果不仅在传统方法上利用义原之间的上下位关系和深度特征,而且考虑义原之间的细化描述,进而细化了义原间的相似度;再者不仅在传统方法上通过义原之间的相似度计算概念之间的相似度,而且考虑了义原描述概念时的层次特征,进而精确地计算概念之间的相似度。

本文在第二节介绍 HowNet 最新的知识表示结构,并与英语中 WordNet 比较。介绍 HowNet 概念多层结构化信息和义原复修饰关系;第三节介绍如何利用概念多层结构化信息和义原复修饰关系计算词语语义相似度;第四节使用 SemEval 2012 语义相似度标准数据进行实验,并与传统的语义相似度计算方法进行比较,并分析实验结果;第五节给出本文的总结和未来工作。

## 2 HowNet 概念多层结构化信息和义原复修饰关系

HowNet 与 WordNet 不同, WordNet 是一个广泛覆盖英文单词的词汇网络,它不但类似于词典将英文单词罗列呈现,并且按照词的意义和词性等构成一个具有语义结构的词网络。WordNet 中词与词之间的关系主要有同义关系(Synonymy)、反义关系(Antonymy)、上下位关系(Hyponymy)、部分整体关系(Meronymy)、方式关系(Troponymy)和必要条件关系(Entailment),并将具有相同词性(名词,动词,形容词和副词)的词集,按照上述关系组成语义词网络,图 1 描述的是 WordNet 中形容词的部分网络结构。

HowNet 是一个广泛覆盖中文词语的知识库,它不同于 WordNet 按照关系连接词与词,而是用一种特定的方法描述词本身的意义及与其他词之间的联系。HowNet 中一个词蕴涵着多个概念,一个概念可以由一种或多种义原结合逻辑符号描述(图 2)。其中义原为知识体系中不可再分的最小单位,即描述概念的最小单位。义原分为实体类义原、事件类义原、属性类义原等。各类义原以树状的形式描述(图 3)

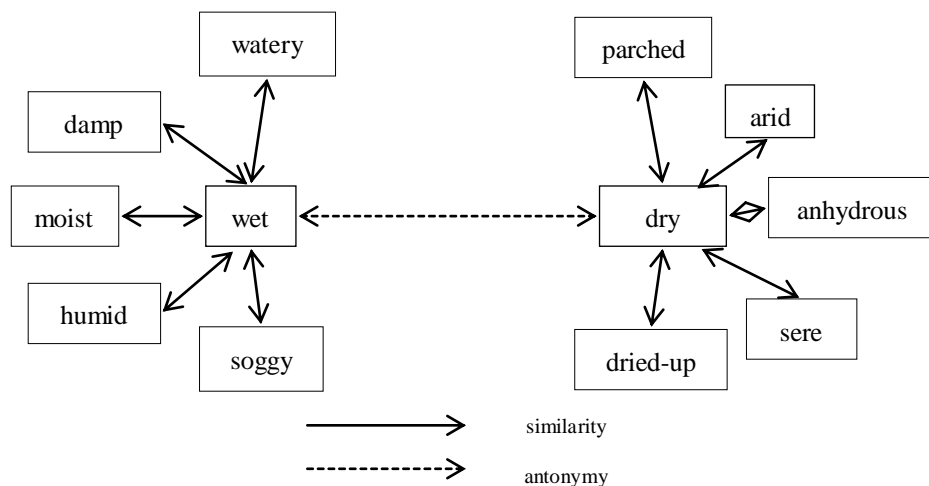


图 1 WordNet 中形容词的部分网络结构  
Fig.1 part of the network structure of the adjective in WordNet

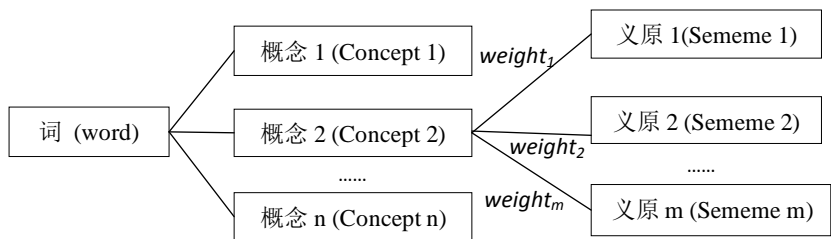


图 2 词的层次描述

Fig.2 the hierarchical description for a word

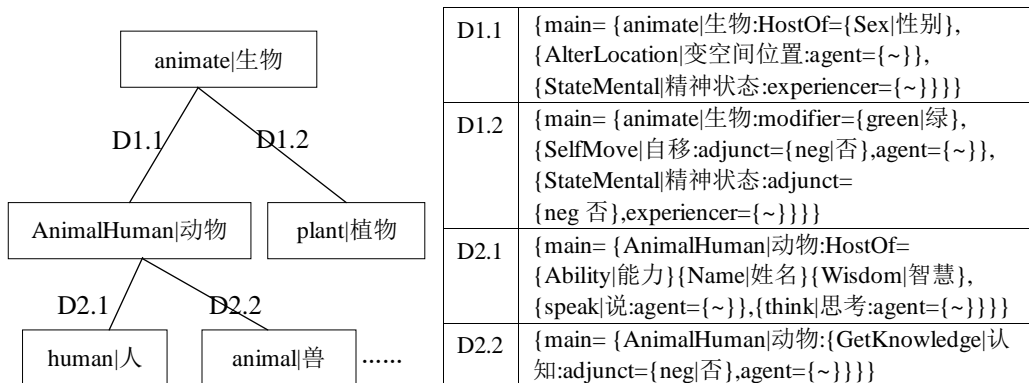


图 3 义原复修饰关系

Fig.3 the sememe multi-modify relationship

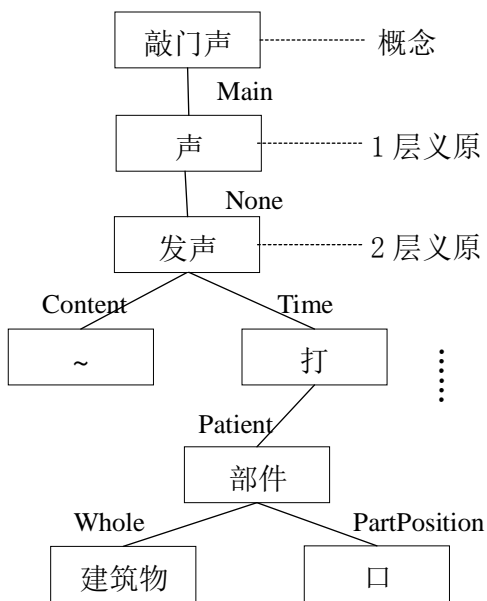


图 4 “敲门声”的概念描述

Fig.4 the concept description of “qiao men sheng”

本文所关心的是概念的描述形式，HowNet 对概念“敲门声”描述形式如图 4。刘群(2002)和江敏(2008)以描述方式为依据，将 HowNet 中的义原分为独立义原、关系义原和符号义原。本文认为无论是独立义原，关系义原或者符号义原，同种类义原之间或者不同种类义原之间，都是以树状结构层层修饰概念，本文定义为概念多层结构信息，这与(刘群,2002)和(江敏,2008)的假设不同。义原之间以上下位关系构成义原树，在事例关系义原和实体关系义原中，义原之间的存在细化修饰，图 3 给出了部分实体类义原的上下位关系和义原之间的细化修饰，本文定义为复修饰关系，这种复修饰关系是一种递归的修饰关系。

因此本文利用概念多层结构化信息和义原复修饰关系，提出计算词语语义相似度新方法。

### 3 词语语义相似度计算

HowNet 中词语由概念表示，词语之间的语

义相似度问题，相当于概念之间的相似度问题。概念由一个或多个义原描述。因此首先计算义原相似度，进而得出概念相似度，最终算出词语语义相似度，是目前公认的计算过程。本文在认同此计算过程的基础上，提出新的义原相似度计算方法和概念相似度计算方法。

### 3.1 义原（Sememe）相似度计算

目前，义原相似度计算采用义原树中义原间的上下位关系等结构信息，这种利用义原树结构的方法已被广泛使用。Dekang Lin(1998)利用 WordNet 中词的上下位关系计算词语相似度，刘群(2002)利用两义原在义原树中的距离计算义原相似度，江敏(2008)考虑了义原在义原树中的深度，葛斌(2010)考虑了义原的密集程度。

在 HowNet 中义原除了上文所提到的上下位关系、深度因素、密集程度等，还存在复修饰关系(图 3)，即义原如概念一般由多个义原修饰。因此本文在计算义原相似度时，不仅考虑义原树中的上下位等关系，还考虑义原的复修饰关系。这样能够使得词语语义相似度更加细化。根据以上考虑，本文提出义原之间相似度的计算公式 1。

$$Sim_s(S_1, S_2, depth) = \begin{cases} StructSim(S_1, S_2) & \text{if } depth = 0 \\ \beta_{depth} Sim_s(S_1, S_2, depth-1) + \beta_{modify} Modify(S_1, S_2) & \text{if } depth \neq 0 \end{cases} \quad (1)$$

其中  $StructSim(S_1, S_2)$  表示义原间的上下位、深度等结构相似度， $Modify(S_1, S_2)$  表示义原间的复修饰关系相似度， $depth$  表示细化的深度， $\beta_{depth} \geq 0$ ，

$\beta_{modify} \geq 0$ ， $\beta_{depth} + \beta_{modify} = 1$ 。当  $S_1, S_2$  均不存在义原复修饰时，认为  $Modify(S_1, S_2) = 1$ 。

上下位结构相似度  $StructSim(S_1, S_2)$  的计算，参考(江敏,2008)的计算方法如式 2。

$$StructSim(S_1, S_2) = \frac{\alpha \cdot (depth(S_1) + depth(S_2))}{\alpha \cdot (depth(S_1) + depth(S_2)) + dist(S_1, S_2) + |depth(S_1) - depth(S_2)|} \quad (2)$$

其中  $depth(S_1)$  表示  $S_1$  在义原树中的深度， $dist(S_1, S_2)$  表示义原  $S_1$  与  $S_2$  的距离。

复修饰关系相似度  $Modify(S_1, S_2)$  中义原  $S_1, S_2$  由其他义原修饰并且修饰形式与概念相同，因此

本文定义  $Modify(S_1, S_2)$  的计算方法与概念之间相似度计算方法相同，概念之间相似度计算方法将在 3.2 节详细介绍。

此外在 HowNet 中义原之间除上下位关系外，对义和反义关系十分重要的。因此在计算两义原  $S_1, S_2$  相似度时做如下操作：

- 当  $S_1, S_2$  为对义或反义关系时，或者当  $S_1, S_2$  在义原树路径上存在对义或反义关系时，将其相似标记为“-”
- 在 HowNet 的概念描述中，存在使用具体词描述概念。因此规定具体词与具体词的相似度计算中，两具体词相等时相似度为 1，否则为 0；具体词与义原相似度为一个较小常数  $\gamma$

通过观察 HowNet 中对义和反义表，发现在对义和反义表中，义原“美”和义原“难看”无法找到其对应关系，并且其同属于属性值义原类别，在其义原路径上也不存在反义对义关系。因此通过导入同义词词典，将属性值义原还原为基本同义词（反义对义表中存在的词），即“美”和“难看”还原成“美”与“丑”，之后再判断反义对义关系。

### 3.2 概念（Concept）相似度计算

上节已经可以得到两个义原的相似度，本节介绍在义原相似度的基础上计算概念之间的相似度。

如图 5 中所描述的，概念由义原以树状多层结构描述，这不是简单的层次结构，而是对概念的细化定义。因此本文定义除概念根节点以外的第一层义原作为一个“main”关系描述概念，第二层义原由 HowNet 定义的关系描述第一层义原，以此类推。这样就形成了一种递归描述结构。较传统方法，不同的地方在于本文按照自底向上，逐层计算义原在层次中的相似度，最终得到概念的相似度。因此本文所提出计算方法（不仅是通过义原之间的相似度计算概念之间的相似度，还考虑了义原描述概念时的层次特征，进而更加准确、更加细化地计算概念之间的相似度。

总结 HowNet 概念描述及结构树中的节点描述分为三种形式：

- 1) {义原}
- 2) {义原: {义原}, {义原}, ...}
- 3) {义原: 关系 1={义原|符号}{义原|符号},  
关系 2={义原|符号}{义原|符号}, ...}

其中符号有~, \$, 表示动作的任何具体内容, 本文认为其与任何义原相似度为 1。在概念的描述上, 为了计算方便本文将无具体关系的义原描述认为是一种“none”关系描述, 即将概念描述形式的 1、2 与 3 合并为

{义原: (none={义原|符号}{义原|符号}), 关系 1={义原|符号}{义原|符号}, 关系 2={义原|符号}{义原|符号}, ...}

其概念描述中(\*)中的内容可以省略。因此图 4 可简单描述为: 其中边信息为描述关系, 根节点为概念, 非根节点为义原或者符号。

概念相似度具体计算方法步骤如下:

- 步骤1: 对于某个节点义原, 将描述关系相同的直接子节点义原归为同一个义原集合, 借鉴(刘群,2002)中义原集合间相似度的方法, 当描述关系无存在对应时默认相似度为一个较小常数  $\delta$ 。
- 步骤2: 将所有关系的相似度计算加权平均值, HowNet 中将符号义原和关系义原合并并且细化为多种关系。因此在概念多层树结构中, 节点义原相似度计算可以简化计算公式。

$$Sim_{s,node}(S_1, S_2) = \frac{1}{2} \left( \frac{1}{N} \sum_{i=1}^N Sim_{rela,i}(S_1, S_2) + Sim_s(S_1, S_2, depth) \right) \quad (3)$$

其中  $N$  表示有  $N$  种不同关系描述,  $depth$  为预先设定的细化程度参数。 $Sim_{s,node}(S_1, S_2)$  表示义原树中节点  $S_1$  与义原节点  $S_2$  之间第  $i$  种描述关系的相似度。由于义原树中的义原节点 (非根节点) 同概念节点 (根节点) 一样, 由下层义原以某种关系描述, 那么概念相似度  $Sim_c(*) = Sim_{s,node}(*)$ , 在 2.1 中提到的义原复修饰关系中, 义原具有概念一般的描述形式。因此  $Modify(*) = Sim_c(*)$ 。

- 步骤3: 依据步骤 1-2, 在得到节点义原相似程度的基础上, 自底向上地计算父节点义

原的层次相似度, 从而得到根节点概念的相似度。

在 HowNet 中义原分为实体类义原、事件类义原、属性类义原、属性值类义原、次要特征义原和专有名词义原。本文认为在相同类别义原中, 平等地计算义原间的相似度存在一个问题如表 1。

概念	描述
硬度	{硬度:host={物质}}
名声	{名声:host={群体}{人}}

表 1 概念“硬度”和“名声”的描述

Table 1 the description of “Hardness” and “Reputation”

在传统方法中硬度与名声的相似度为 0.565, 基本上达到的粉红色和深红色的相似度 0.580, 这不符合日常生活习惯, 因此通过观察它们的概念描述信息, 发现由于属性类义原, 在概念描述中, 为第一层节点义原, 所占的比重过大。属性类义原往往之后存在属性值类等义原补充说明, 因此本文对于属性类、次要特征类等义原, 加入  $\varepsilon$  惩罚因子。

本文借鉴前人提出的语义相似度计算公式, 引入极性符号, 通过在得到两个概念之间相似度的基础上, 计算出词语语义相似度。

$$Sim_w(W_1, W_2) = \pm \max_{i=1..n, j=1..m} |Sim_c(C_{1i}, C_{2j})| \quad (4)$$

其中  $C_{1i}$  为词语  $W_1$  的第  $i$  个概念,  $C_{2j}$  为词语  $W_2$  的第  $j$  个概念,  $\pm$  符号取决于词语之间最大概念相似度的符号,  $Sim_c(*)$  表示计算两个概念的相似度。

## 4 实验

在英语语义相似度计算中, 使用认知实验得到的数据作为语义相似度的标准评测数据, 但是在汉语中没有标准评测数据集, 因此依靠使用人工标注数据进行评测, 并规定语义相似度范围为 0 到 1 之间, 0 表示相似度最小, 1 表示相似度最大。实验中传统方法参数参考江敏(2008), 本文方法参数默认设定如表 2 所示, 首先规定结构相似度和复修饰关系相似度的权重值相等, 之后通过实验观察参数影响。

4.1 SemEval 2012

SemEval 2012 在 Evaluating Chinese Word Similarity 任务中, 提供 300 对中性词汇语义相似度的人工标注结果。其结果由 27 名人员独立标注, 并取其平均值作为人工标注结果。本实验剔除 HowNet 中未登入词后选取 130 对词汇, 其中包括 40 个词对相似度为 1-0.6、50 个词对相似度为 0.6-0.3 和 40 个词对相似度小于 0.3, 并按照人工标注的结果排序。实验分别以前 20、40、60、80、100 和 130 对词为计算对象。实验结果如图 5 所示。

图 5 中纵轴表示结果与评测数据的相似度差值, 横轴为实验的词对数量。图 5 表明相似度较大的词对 (前 40) 中, 相似度差值在 0.095 左右。最高的平均差异值在 0.1 左右, 这是正常的, 针对相似度比较模棱两可的词对, 由于词义的理解是有争议的, 因此与标准数据差异会增加。实验表明本文提出的语义相似度计算是有效的, 并且最高差值与最低差值的范围不超过 0.01, 说明此语义相似度计算方法是稳定的。与传统方法相比, 无论是在稳定性还是准确性上都有明显提高。

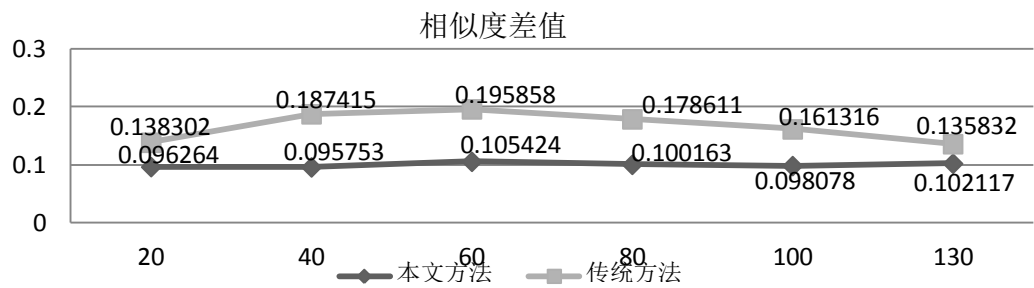


图 5 实验结果与 SemEval2012 相似度差值曲线  
Fig.5 difference between the experiment result and standard data in SemEval2012

	$\alpha$	$\gamma$	$\delta$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_{depth}$	$\beta_{modify}$	$\varepsilon$	Depth
传统	0.5	0.01	0.01	0.7	0.17	0.13	—	—	—	—
本文	1.6	0.01	0.01	—	—	—	0.5	0.5	0.25	1

表 2 实验参数设定  
Table 2 experiment parameters

$\beta_{depth}$	$\beta_{modify}$	词对数量					
		20	40	60	80	100	130
0.1	0.9	0.119334	0.119463	0.130196	0.124995	0.114531	0.111135
0.2	0.8	0.113566	0.112729	0.122762	0.117729	0.109338	0.108035
0.3	0.7	0.107799	0.106976	0.116765	0.111440	0.105020	0.105619
0.4	0.6	0.102031	0.101365	0.111094	0.105601	0.101183	0.103578
0.5	0.5	0.096264	0.095753	0.105420	0.100163	0.098078	0.102117
0.6	0.4	0.090496	0.091541	0.100686	0.096294	0.096423	<b>0.101841</b>
0.7	0.3	<b>0.090174</b>	0.090175	0.097845	0.093847	<b>0.095931</b>	0.102481
0.8	0.2	0.092316	<b>0.090040</b>	0.096216	0.092310	0.096331	0.103818
0.9	0.1	0.097070	0.091212	<b>0.096023</b>	<b>0.091995</b>	0.097971	0.106124

表 3 参数对相似度差值的影响  
Table 3 The influence of parameter on similarity

本文方法不改变其它参数的情况下,调整 $\beta_{depth}$ 和 $\beta_{modify}$ 的参数,得到实验结果如表3。通过实验发现在词对的相似度比较小时,细化的计算结果能够很好地降低与标准数据的差值。

## 4.2 中性名词相似度比较

传统方法和本文方法的实验结果相比(表4),男人、父亲和男亲属,传统方法中其相似度都为1。本文通过细化分析男亲属与父亲的相似度较男人与父亲高0.08,这更符合日常生活。珍宝和宝石的相似度却降低了。在HowNet概念层次定义中。

概念“宝石”描述式为:

{材料:MaterialOf={珍宝}}

概念“珍宝”描述式为:

{珍宝}

本文所使用的方法在义原复修饰关系中

义原“材料”描述式为:

{人工物:{创造:PatientProduct={人工物},material={~}}}

义原“珍宝”描述式为:

{钱财:modifier={贵}{珍},{留存:patient={~}}}

可以看出义原的描述式的相似程度较低,因此导致概念相似度降低。然而对于“跑”的描述式为{跑}、“跳”的描述式为{跳},其中{跑}和{跳}的描述式相同,因此相似度相对提高。

此外概念“粉红色”和概念“深红色”也大大提升了相似度。因此加入多层义原细化了相似度的计算。通过加入惩罚因子,用以区分不同类别的义原对概念的修饰权重。概念“名声”和概念“硬度”,以及“风度”和“面积”的相似度在不影响实体义原修饰的情况下降低了。

在表5中,HowNet加大了对义和反义义原的数量,因此使得概念“三伏”和概念“冬眠”修饰中产生了反义的关系,相似度的绝对值明显提高,这说明多层义原的修饰是有效的,并通过导入同义词典,还原“父亲”和“母亲”的基本义原后得到对义关系。

## 4.3 褒贬词相似度比较

虽然HowNet加大了对义和反义的规模,但是对义和反义仍然没有细化。因此在HowNet中加入同义词典还原放宽了反义和对义的限制,例如义原“残疾”与“健康”是严格的反义关系,然而修饰概念“舒服”的义原集中,存在义原“健壮”,那么通过同义词的还原,将“健壮”还原成“健康”从而找到对应的反义关系,使得相似度提高。

## 5 结论与未来工作

HowNet加入了大量的词汇信息和语义知识,内容的描述十分详尽。通过分析其描述规则和语义特征,能够获得大量的词汇知识。此外对词语的读音、特殊例句、中英对应关系等都做出了大量的描述。本文总结国内外语义计算的方法,在前人研究的基础上,提出一种新的词语语义相似度计算方法,并对计算中的细节加以改进:

- 1) 概念是由义原描述,是一种树状层次结构的描述关系,并且越往深层的描述是一种细化概念的过程。因此本文在树状层次结构中自底向上地计算相似度,从而得到根节点(概念)的相似度。
- 2) 义原在义原树上的信息,除了义原距离、义原深度、义原密集等,还存在类似概念的义原复修饰关系(与其直接父节点的修饰关系)。利用此关系并通过词语语义相似度实验分析,表明能够细化概念的分析,提高分析准确度。
- 3) 义原存在多种类别,认为不同类别的义原对概念的修饰存在不同的权重。
- 4) HowNet中对义和反义关系的描述粗且少,因此加入同义词典,将对义和反义关系的限制放宽,在计算词语语义相似度中得到好的效果。

本文使用SemEval 2012提供的语义相似度人工评测数据实验,说明语义相似度计算的有效性。并利用改进的方法与江敏比较,证明该方法能有效影响并细化词语语义相似度计算。

在计算概念相似度时,可以利用的结构化信息十分丰富,不仅仅局限于同层次义原间的计算,还存在子树的约束。某些概念树状描述为其他概

念树状描述的子树,使用概念描述树状结构信息准确地对齐两个概念间的描述的子树,本文认为能够进一步提高词语语义相似度计算方法。未来工作包括:考虑子树结构的对齐信息,将其有效地运用于机器翻译和情感计算中。

词语 1	词语 2	传统	本文	
			未加 $\varepsilon$ 因子	加入 $\varepsilon$ 因子
男人	父亲	<b>1.000 000</b>	<b>0.915 942</b>	<b>0.915 942</b>
男人	男亲属	<b>1.000 000</b>	<b>0.915 942</b>	<b>0.915 942</b>
男亲属	父亲	<b>1.000 000</b>	<b>0.994 369</b>	<b>0.994 369</b>
发明	创造	0.850 000	0.864 609	0.864 609
香蕉	苹果	1.000 000	1.000 000	1.000 000
医生	医治	0.004 422	0.075 000	0.075 000
珍宝	宝石	0.500 000	0.401 928	0.401 928
中国	联合国	0.257 500	0.353 967	0.353 967
深红色	粉红色	<b>0.580 867</b>	<b>0.966 216</b>	<b>0.966 216</b>
跑	跳	0.818 182	0.967 532	0.967 532
跳槽	拔脚	0.158 175	0.342 398	0.342 398
风度	面积	<b>0.427 833</b>	<b>0.736 325</b>	<b>0.416 472</b>
名声	硬度	<b>0.565 357</b>	<b>0.632 849</b>	<b>0.312 996</b>

表 4 中性词相似度结果

Table 4 the results of similarity between neutral words

词语 1	词语 2	传统	本文
美丽	贼眉鼠眼	-1.000 000	-1.000 000
美丽	优雅	<b>0.750 000</b>	<b>0.852 941</b>
高尚	卑鄙	-0.912 500	-0.872 881
舒服	残废	<b>-0.222 203</b>	<b>-0.614 747</b>
三伏	冬眠	<b>0.005 838</b>	<b>-0.655 555</b>
父亲	母亲	<b>0.888 888</b>	<b>-0.994 369</b>

表 5 褒贬词和对义词相似度结果

Table 5 the results of similarity between emotion words and opposite word

致谢

SemEval 2012 在 Evaluating Chinese Word Similarity 任务中的人工标注数据,由乐山师范学院金澎博士提供

参考文献

Varelas, G., Voutsakis, E., Raftopoulou, P., Petrakis, E., Milios, E.: Semantic Similarity Methods in WordNet

and their Application to Information Retrieval on the Web. In: 7th ACM Intern. Workshop on Web Information and Data Management (WIDM 2005), Bremen, Germany (2005) 10–16

杨超,冯时,王大玲等. 基于情感词典扩展技术的网络舆情倾向性分析[J]. 小型微型计算机系统. 2010, (4):691-695

M. Mohler and R. Mihalcea (2009), “Text-to-text Semantic Similarity for Automatic Short Answer Grading,” in Proc. 12th Conf. European Chap. ACL, Athens, Greece, 567–575.

D. Bollegala, Matsuo Y., and M. Ishizuka. 2007. Measuring semantic similarity between words using web search engines. In Proceedings of WWW’2007

Rohde, D. L. T., Gonnerman, L. M., & Plaut, D. C. (2008). An improved method for deriving word meaning from lexical co-occurrence. Unpublished manuscript.

S. A. Takale and S. S. Nandgaonkar. Measuring semantic similarity between words using web documents. International Journal of Advanced Computer Science and Applications (IJACSA), 1(4), 2010

Richardson, R., Smeaton, A. F., & Murphy, J. (1994). Using WordNet as a knowledge base for measuring semantic similarity between words. Working paper CA-1294, Dublin City University, School of Computer Applications, Dublin, Ireland.

Gutiérrez, Y., S. Vázquez and A. Montoyo. Sentiment Classification Using Semantic Features Extracted from WordNetbased Resources. In 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011). ACL. 2011: 139-145. Portland, Oregon.

Guan, Y., et al.: Quantifying Semantic Similarity of Chinese Words from Hownet. In: IEEE Proceedings of ICMLC02, Beijing, vol. 1, pp. 234–239. IEEE Computer Society Press, Los Alamitos (2002)

董振东, 董强. 知网(HowNet)[DB/OL]. (2008-09-15) 刘群, 李素建. 基于《知网》的词汇语义相似度的计算[C]//第三届汉语词汇语义学研讨会. 台北, 2002.



- 江敏, 肖诗斌, 王弘蔚等. 一种改进的基于知网的词语语义相似度计算[J]. 中文信息学报, 2008, 22(5): 84-89.
- M Ruiz-Casado, E. Alfonseca, and P. Castells. 2005. Using context-window overlapping in Synonym Discovery and Ontology Extension. In Proceedings of RANLP-2005, Borovets, Bulgaria,.
- Y. Li, Z. A. Bandar, and D. McLean. An approach for measuring semantic similarity between words using multiple information sources. IEEE Transactions on Knowledge and Data Engineering, 15(4):871–882, 2003.
- C. d’Amato, N. Fanizzi, and F. Esposito. A semantic similarity measure for expressive description logics. In A. Pettorossi, editor, Proceedings of Convegno Italiano di Logica Computazionale (CILC05), Rome, Italy, 2005.
- DEERWESTER, S., DUMAIS, S., LANDAUER, T., FURNAS, G., AND HARSHMAN, R. 1990. Indexing by latent semantic analysis. J. Amer. Soc. Info. Sci. 41, 391–407.
- Lund, K., & Burgess, C. (1996) Producing high-dimensional semantic spaces from lexical co-occurrence. Behavior Research Methods, Instruments, & Computers, 28, 203-208.
- Huang, C.-R., Hsieh, S.-K., Hong, J.-F., Chen, Y.-Z., Su, I.-L., Chen, Y.-X. and Huang, S.-W.. Chinese Wordnet: Design, Implementation, and Application of an Infrastructure for Cross-lingual Knowledge Processing. Proceedings of the 9th Chinese Lexical Semantics Workshop. 2008.
- Xuexia Gao. Tao Kuang. Research of Word similarity Model Based on HowNet. IEEE, 2011, 1: 287-290.
- MILLER, G. A. 1995. WordNet: A lexical database for English. Communications of the ACM 38, 11, 39–41.
- Dekang Lin. An Information-Theoretic Definition of similarity semantic distance in wordNet[C]//Proceedings of the Fifteenth International Conference on Machine Learning. 1998.
- 葛斌, 李芳芳等. 基于知网的词汇语义相似度计算方法研究[J]. 计算机应用研究, 2010, 27(9): 3329-3337