

# 基于语义关系图的词汇语义相关度计算研究\*

李佳媛, 张仰森

(北京信息科技大学 智能信息处理研究所, 北京 100192)

**摘要:** 词汇的语义计算是中文信息处理领域的重要问题之一, 目前的研究主要针对词汇语义的相似度计算方面, 对词汇语义的相关度计算方法研究不够。为此, 本文以《知网》中概念之间以及概念的属性之间的语义关系和大规模语料中统计出的词语依存关系为基础, 构建了一张语义关系图。然后, 将此语义关系图与图理论的相关知识相结合, 提出了一个基于语义关系图的词汇语义相关度计算的方法。实验结果表明, 该方法在语义相关度计算方面效果较好, 得到的语义相关度结果较为合理。

**关键词:** 语义相关度; 语义关系图; 知网

## Study of word semantic relevancy computation based on lexical semantic relationship graph

LI Jiayuan, ZHANG Yangsen

(Institute of Intelligence Information Processing,  
Beijing Information Science and Technology University, Beijing 100192, China)

**Abstract:** Lexical semantic computation is one of the important issues of the Chinese Information Processing field. The present study is focused on semantic similarity computation of word, semantic relevancy computation is not paid enough attention. For this reason, we built up a lexical semantic relationship graph based on words dependency relations and the semantic relations of inter-conceptual and inter-attribute of concepts, which come from HowNet. Then, we applied related knowledge of graph theory to the semantic graph, and proposed a word semantic relevancy computation method based on the semantic relationship graph. Experimental results show that this method has a better performance in word semantic relevancy computation, and semantic relevancy computation results are more reasonable.

**Keywords:** Word semantic relevancy; lexical semantic relationship graph; HowNet.

### 1 引言

在任意一个自然语言文本中, 不同词语之间的相关性是不同的, 从语义的角度将它称为语义相关度<sup>[1]</sup>。词汇的语义相关度研究是自然语言处理的基础, 在信息检索、文本分类、文本聚类 and 句法消歧等应用领域的研究中, 语义相关度都有着非常重要的意义。

语义的相关度和语义的相似度是两个不同的概念, 但二者又有着紧密的联系。词语相关度反映了两个词语相互关联的程度, 即词语之间的组合特点, 它可以用这两个词语在同一个语境中共现的可能性来衡量。词语相似度是指两个词语的相似程度, 通常指两个词语的语义本身所具有某些共同特性, 相似度反映的是词语之间的聚合特点。就它们所表示的范畴来说, 语义相关度是更一般的概念, 而语义相似度是语义相关度的一种特例, 也就是, 语义相关度包含了语义相似度的概念。Resnik<sup>[2]</sup>用轿车、汽油和自行车的例子解释了两者的区别: “轿车依赖于汽油作为燃料, 显然它们之间的相关性比轿车与自行车更为紧密, 但人们却普遍认为轿车与自行车之间的相似性大于轿车与汽油。”这个例子表明, 相关性不能等同于相似性。即使轿车与汽油是紧密相关的,

---

\* **基金项目:** 国家自然科学基金项目 (61070119); 北京市属高等学校创新团队建设与教师职业发展计划项目 (IDHT20130519); 北京市较为专项基金 (PXM2012-014224-000020)

**作者简介:** 李佳媛 (1988—), 女, 研究生, 主要研究方向为中文信息处理; 张仰森 (1962—), 男, 教授, 主要研究方向为中文信息处理、人工智能。

但由于这两者之间没有共同的特性，人们不会认为它们是相似的。而轿车和自行车都是交通工具，都有轮子并且可以载人，因此它们是相似并且相关的。在评价两者的时候，可以把词语相似度作为词语相关度评价的一个方面，也就是说如果词语间的语义越相似，那么它们的相关度在一定程度上也越大，相似度的大小在一定程度上影响相关度的度量。

目前，对于语义相关度的评价方法主要分为两种：基于统计的方法和基于语义词典的方法。基于统计的方法是建立在“两个词语经常同时出现，则这两个词语往往相关”这一假设的基础上，这种方法主要利用文档集中词语间共现性的统计数据来确定词语间的相关度，但是这种方法只是利用文档中包含的内容信息，而忽略了词语之间的具体关系以及关联的语义依据<sup>[3]</sup>，并且当统计样本不足时，其计算结果就会出现较大误差。而基于语义词典的方法是依据已有的词典，如 WordNet、HowNet 等知识体系来计算词语在语义上的相关程度。许云<sup>[4]</sup>等人基于语义相似度以及知网的语义信息，提出了语义相关度的计算方法，即利用知网义原纵向与横向关系以及词语的实例信息来计算词语之间的语义相关度。

本文充分利用知网中丰富的语义关系，同时借鉴文献[5]中的方法，提取出知网中的语义关系，建立了一个语义关系图，在此基础上，我们通过对大规模语料进行依存语法分析，统计出其中存在的依存语义关系，将统计出的依存语义关系加入图中，对语义关系图做了进一步的扩展。该语义关系图具有图结构的基本特征，图的一般理论都适用于该语义关系图。在语义关系图的基础上，本文将图理论知识与语义关系图中的信息相结合，提出了一种基于语义关系图的词汇语义相关度计算方法，并通过实验验证了该方法是有效可行的。在下面的几节中，第 2 节介绍了语义关系图的构建及其扩展，以及基于语义关系图的词语相关度计算方法，第 3 节介绍了实验结果，第 4 节做了简要的总结和工作展望。

## 2 词汇语义相关度计算

### 2.1. 语义关系图的构建

在自然语言中，一个词语往往具有多个含义，在具体的语言环境中它们对句意的表达作用也往往是多种多样的，词与词之间的关系更是错综复杂，比如同义、反义、施事、受事、句法关系等等。为了表示这些复杂的关系，本文采用了在表现复杂关系方面具有天然优势的图结构。在构建语义关系图时，首先分析研究了《知网》对词语语义的表示方式，根据《知网》中词语的语义信息以及知网对词语语义的表示特点，借鉴了文献[5]中的方法，提取出知网中的语义关系，构建出了基于知网的语义关系图；然后，通过分析大规模的语料，统计出其中的依存词语搭配，并将这些词语搭配及其依存关系扩展到已构建好的语义关系图中，达到进一步丰富语义关系图的目的，并完成语义关系图的构建。

本文所构建的语义关系图由表示语义的节点和表示语义节点间关系的边组成，整个语义关系图以边为单位采用三元组的方式存储，也就是每一个语义关系三元组（结点 1，结点 2，语义关系）作为一个存储记录。采用这样的存储方式便于对已构建好的语义关系图进行扩展，在扩展时只要将新的语义关系三元组加入图中即可达到扩展语义关系图的目的。

#### 2.1.1. 《知网》中的语义关系

《知网》是一个以英汉双语所代表的概念以及概念的特征为基础的，以揭示概念与概念所具有的特性之间的关系为基础内容的常识知识库<sup>[6]</sup>。通过对《知网》词典中概念表示方法的研究，我们发现《知网》按照 KDML 的规范描述概念时，对概念采用嵌套式的结构表示，也就是，一个复杂的概念用较简单的概念进行解释，较简单的概念再用更简单的概念解释，直到用义原表示出来<sup>[5]</sup>。这种结构是一种隐含的树结构，称之为概念树<sup>[5]</sup>。如下面的例子：

W\_C=拳台

DEF={facilities| 设施 :domain={boxing| 拳击 },{compet| 比赛 :location={~}},{exercise| 锻炼:location={~}}}

在这个概念描述中，KDML 表示了这样的含义：拳台是一个设施，这个设施所属的领域是拳击领域，这个设施是比赛的地方，这个设施也是锻炼的地方。也就是说，拳台是一个用来进行拳击比赛和拳击锻炼的设施。

将“拳台”这个概念用概念树重新表示如图 1：

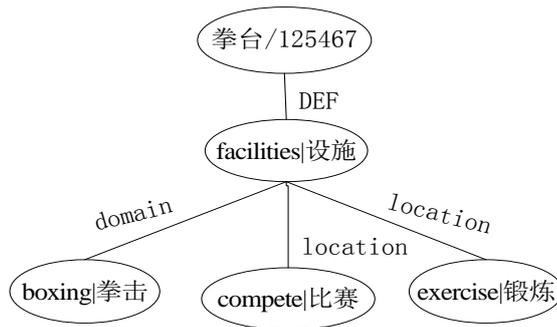


图1 概念“拳台”的概念树表示

### 2.1.2. 《知网》语义关系提取

通过上面的例子，可以发现：在概念树上，每一个父节点与其子节点之间必定有一个表示语义关系的关键词。因此，在遍历概念树提取语义关系的时候，就可以把关键词作为提取语义关系的标志，即当遇到语义关系词时，则考察该关系词所连接的两个节点是否可作为语义节点与该关系词构成一条语义关系记录加入语义关系图中。下面具体说明本文在基于文献[5]提出的方法的基础上提取语义关系的步骤：

以概念 $\{s1:r1=\{s2:r2=\{s3\}\}\}$ 为例，按照 KDML 规范，每一对括号所包括的部分都是一个概念，则在该例中三对括号所包括的内容“ $\{s1:r1=\{s2:r2=\{s3\}\}\}$ ”、“ $\{s2:r2=\{s3\}\}$ ”、“ $\{s3\}$ ”则是三个不同的概念。其中  $s1, s2, s3$  是义原； $r1, r2$  是关系词， $r1$  是表示  $s1$  和 $\{s2:r2=\{s3\}\}$ 之间关系的关键词， $r2$  是表示  $s2$  和 $\{s3\}$ 之间关系的关键词。在提取语义关系时：

- (1) 如果关系词后面所连接的概念只是一个义原，则直接提取语义关系。例如： $\{s2:r2=\{s3\}\}$ ，关系词  $r2$  后的概念“ $\{s3\}$ ”只有义原“ $s3$ ”表示，那么可以直接提取语义关系 $\{s2,s3,r2\}$ ；
- (2) 如果关系词后面所连接的概念不只是一个义原，而是多个义原，这时需要考察关系词后面所连接的概念是否可以用某个义项表示，若可以用特定义项表示，则可提取语义关系，否则，不提取该关系词的语义关系。例如： $\{s1:r1=\{s2:r2=\{s3\}\}\}$ ，若 $\{s2:r2=\{s3\}\}$ 可用义项 B 表示，那么可提取的语义关系 $\{s1,B,r1\}$ 和 $\{s2,s3,r2\}$ ；若 $\{s2:r2=\{s3\}\}$ 找不到义项表示时，则只能提取语义关系 $\{s2,s3,r2\}$ ；
- (3) 如果关系词所在的整个概念可用某个义项表示时，则可将关系词前面的义原替换为该义项并提取语义关系。例如：若 $\{s1:r1=\{s2:r2=\{s3\}\}\}$ 可用义项 A 表示， $\{s2:r2=\{s3\}\}$ 可用义项 B 表示，则可提取语义关系 $\{A,B,r1\}$ ， $\{s1,B,r1\}$ ， $\{B,s3,r2\}$ ， $\{s2,s3,r2\}$ ；
- (4) “DEF”关系，对于每个义项都需提取该义项与其第一基本义原的 DEF 语义关系；
- (5) 反义、对义、同义关系的提取，这三种关系可直接从《知网》的 Antonym set、Converse Set、Synset Set、Taxonomy Antonym、Taxonomy Converse 文件中直接提取。

通过研究《知网》收录的词语及其语义描述，我们发现其中有些词语的几个义项的中文词、词性以及概念描述等完全相同，只有对应的英文词不同而表示为不同的义项。由于本文所做工作的主要目的是为了计算词语的语义相关度，与该词语的词性及其对应的英文词无关，因此在提取语义关系之前，我们先将中文词相同且概念描述也相同的但编号不同的概念进行合并，并重新为其编号，然后再提取语义关系构建语义关系图。在本文中，每一条语义关系记录采用语义三元组的形式存储，存储格式如表 1 所示，采用这样的存储方式便于对已构建好的语义关系图进行扩展，在扩展时只要将新的语义关系三元组加入图中即可达到扩展语义关系图的目的。

表1 语义关系的存储格式

关系左项	关系右项	关系词
拳台	设施	DEF
.....	.....	.....

提取出《知网》中存在的语义关系后，这些关系互相关联后则形成了一张网，称之为语义关系图。该语义关系图符合图的一般特点，具有图结构的一般性质。由于该图的是以语义三元组为单位进行存储，因此该图具有良好的可扩展性，并且可以融合其他语义资源的语义关系，进一步完善词语间的语义关联信息，得到两个词语间更为客观的相关度。

### 2.1.3. 语义关系图的扩展

在自然语言领域中，词语以及概念由于所处的语言环境不同，它们之间所表现出来的关系也是非常错综复杂的，虽然《知网》着力反映了概念之间和概念的属性之间的各种关系，但是由于不可能穷尽概念之间或概念的属性之间的所有关系，加上人力物力的限制，《知网》中列举出的关系只是最基本的，还有一些在语言使用过程中所用到的语义关系，在知网中没有体现出来，或者某些词语间的语义关联方式与《知网》中的关联方式不同，也就是在语义关系图中两个语义节点通过不同的路径相互连通。为了使语义关系图中的语义关联信息更全面，则需要对基于《知网》构建好的语义关系图做进一步的扩展，来丰富其中的语义关联信息。

在以往分析语料时，我们发现如果两个词语经常同时出现在一句话中，则这两个词语之间往往具有一定的关系。虽然在具体的语句中并没有说明一句话里的词语间具有何种关系，但是只有当词语间存在内在的语义关联，才可以组合形成一句话并表达一个完整的句意。另外，依存语法认为句子的成分之间存在依存关系，这种依存关系可以反映出句子中各成分之间的语义修饰关系<sup>[7]</sup>。基于以上结论，本文采用哈尔滨工业大学所作的依存句法分析器，对主要来自于人民日报的语料进行句法分析，从中提取出具有依存关系的词语搭配对，每一个词语搭配对与它们的依存关系则构成了一个语义关系三元组，将这些语义关系三元组加入到基于《知网》构建的语义关系图中，就实现了语义关系图的扩展。具体的扩展步骤如下：

(1) 依次对人民日报语料中的每一句话进行依存句法分析，得到每一句话的依存句法树；

(2) 根据每一颗依存句法树中的词语依存信息，从中提取出实词依存搭配对，并统计其出现的频次；

(3) 为每依存搭配对中两个词语间的语义关系赋予一个关系词——“依存”，这样每个搭配对及其关系词则构成一个语义关系三元组；

(4) 依次将频次大于一定阈值的语义关系三元组加入到基于《知网》构建的语义关系图中。

经过以上处理，则实现了对语义关系图的扩展，丰富了语义关系图中词语与概念的语义关联关系，并得到了一张相对完善的语义关系图。构建好语义关系图，我们就可以把图论的知识与图中词语以及概念之间错综复杂的关系相结合设计词语相关度计算的算法，并最终实现词汇语义相关度的计算。

## 2.2. 词语相关度计算算法设计

### 2.2.1. 基本定义

为了更好地阐述算法和便于理解算法，下面先给出算法中将要涉及到的一些基本概念和定义：

定义 2.1 (语义相关度)：给定两个实体  $E_1$  和实体  $E_2$ ，它们之间的语义相关度通过函数  $Rel(E_1, E_2): S \times S \rightarrow [0, 1]$  表示，它体现了集合  $S$  中两个实体  $E_1$  和  $E_2$  之间的语义相关程度<sup>[8]</sup>。

结合图论中两点连通的概念以及文献<sup>[9]</sup>，本文给出语义连通的定义如下：

定义 2.2 (语义连通)：在语义关系图中，有结点  $E_1$ 、 $E_2$ 、 $E_3$ .....、 $E_n$ ，如果从结点  $E_1$  到结点  $E_2$  有路径，则称  $E_1$  和  $E_2$  是语义连通的。

定义 2.3 (连通路程长)：在语义连通的结点  $E_1$  和结点  $E_2$  之间，连通路程长用  $L$  表示，并且是一个实数值型函数， $L(E_1, E_2) \in (0, \infty)$ ，表示  $S$  中的结点  $E_1$  和  $E_2$  之间的语义距离的度量。在本文中，连通路程长  $L$  等于结点  $E_1$  和  $E_2$  的连通路程上语义关系边的数量。

假设 2.1：在语义关系图中，如果结点  $E_1$  和  $E_2$  之间有至少一条语义连通路程，则认为结点  $E_1$  和  $E_2$  是语义相关的。

以上定义和假设是下文进行语义相关度计算的基础。

### 2.2.2. 设计原理

本文认为：如果两个实体  $E_1$  和  $E_2$  相关，那么在语义关系图上，结点  $E_1$  和  $E_2$  之间应该存在至少一条连通路程，或者  $E_1$  与以  $E_2$  为中心的一定连通路程长度范围内的结点集合中的某个元素的相似度很大（或者  $E_2$  与以  $E_1$  为中心的相关结点的相似度很大）。本文的研究就是在此基础之上的，为了对两个实体之间的语义相关度进行量化，以下详细介绍语义相关度计算算法设计的基本原理：

(1) 在语义关系图中，两个语义连通的实体结点之间的连通路程越多，则认为两个实体越相关；反之，则认为两个实体之间的相关度越低。

在两个存在语义连通路径的实体之间，假设它们的连通路径长都相等，那么，连通路径的数量和两个实体之间的相关度的大小成正比，即认为语义相关度的值随着连通路径数量的增加而增大，随着连通路径数量的减少而减小。例如图 2，图 (a) 中结点 A 和 B 的连通路径有 2 条，且长度都为 2；图 (b) 中，结点 A 和 B 的连通路径有 3 条，且长度也同样都为 2，在这样情形下，我们自然会认为图 (b) 中 A 和 B 的相关度要大于图 (a) 中 A 和 B 的相关度，因为在相同关联路径长的前提下，图 (a) 中 A 和 B 比图 (b) 中 A 与 B 之间存在有更多的语义关联。

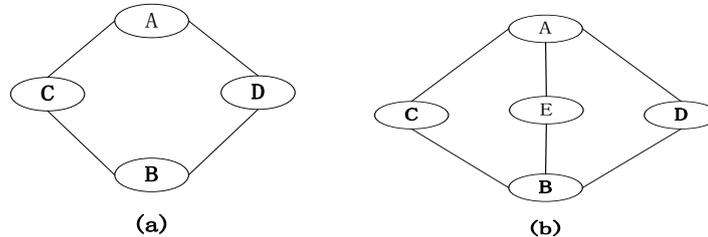


图 2 语义连通路径的数量与语义相关度的关系

(2) 在两个存在语义连通路径的实体之间，假设它们的连通路径数量都相等，那么，连通路径长度和相关度的大小成反比，即认为语义相关度的大小随着连通路径长度的增大而减小，随着连通路径长度的减小而增大。例如图 3，图 (a) 中结点 A 和 C 的连通路径有 1 条，且长度为 2；图(b)中，结点 A 和 C 的连通路径也有 1 条，但长度为 1，在这样的情形下，我们认为图 (b) 中 A 和 C 的相关度要大于图 (a) 中 A 和 C 的相关度。因为在相同数量的关联路径的前提下，图 (b) 中 A 和 C 是直接相关的，而图 (a) 中 A 和 C 则是依赖于其他结点才能相互关联。

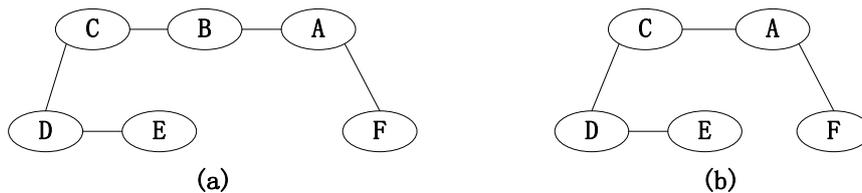


图 3 语义连通路径长度与相关度的关系

(3) 在语义关系图中，若两个实体结点之间没有连通路径，则有两种可能：

① 在构建语义关系图时，语义资源有限导致语义关系图并不能够穷举出所有的语义关系，致使某些语义关联缺失，从而使得某些有关联的词语失去语义关联，表现在语义关系图上即为两词语的结点之间没有语义连通路径；

② 两实体语义不相关。

对于第三种情况，本文采用相似词替换的方法计算相关度。首先，以其中一个实体结点 A 为中心，找出其固定连通路径长度内的所有相关结点，从这些结点中选出与另一个实体结点 B 语义最相似的结点 C，当 B 与 C 的相似度  $\text{Sim}(B,C)$  大于一定阈值时，则计算出 A 与 C 的相关度  $\text{Rel}(A,C)$ ；然后以 B 为中心，采用同样的方法，计算出  $\text{Sim}(A,C')$  与  $\text{Rel}(B,C')$ ；最后，从  $\text{Sim}(B,C) \times \text{Rel}(A,C)$  与  $\text{Sim}(A,C') \times \text{Rel}(B,C')$  中选择较大者作为 A 与 B 的相关度；若  $\text{Sim}(B,C)$  与  $\text{Sim}(A,C')$  都小于设定的阈值，即没有与 A 或 B 非常相似的词时，则认为 A 与 B 的相关度为 0，即 A 与 B 语义不相关。本文在计算两个词语的相似度时采用文献[10]的方法。

### 2.2.3. 词语相关度计算的算法

在语义关系图的基础上，结合图论的相关知识及前面论述的定义及原理，可以设计词语相关度计算的算法，实现词语相关度的计算。本文主要考察的是两个词语在语义关系图中的关联路径的数量和每条路径的长度这两个因素，即在给定两个词语后，通过遍历语义关系图，得到这两个词语的语义连通路径数 N 和每条连通路径的长度  $L_i$  后，则可通过 N 和  $L_i$  计算出两个词语的相关度。

为了使算法比较好实现和强调路径长度的影响，在计算中不考虑连通路径中长度超过 6（包含 6）的路径，并且为连通路径长为 1~5 的路径分别赋予系数  $\beta_k$  ( $1 \leq k \leq 5$ )。由此，可得到每条路径的带权长度为  $\beta_k L_i$ ，其中， $k \in [1,5]$ ， $i \in [1,N]$ 。因此，

$$\text{总路径长 } L: L = \sum_{i=1}^N \beta_k L_i \quad 1 \leq k \leq 5; \quad (1)$$

$$\text{平均路径长 } \bar{L}: \bar{L} = \frac{1}{N} \sum_{i=1}^N \beta_k L_i \quad 1 \leq k \leq 5; \quad (2)$$

因此,当两个词语在语义关系图中存在长度小于 6 的语义连通路程时,它们的相关度为:

$$\text{Rel} = \frac{\log_2(N+1)}{\log_2(N+1) + 2 \frac{1}{N} \sum_{i=1}^N \beta_k L_i} \quad 1 \leq k \leq 5 \quad (3)$$

基于语义关系图的词语相关度计算的算法 1 描述如下:

#### 算法 1 基于语义关系图的词语相关度计算算法

- 
- Step1. 计算两个实体结点 A 和 B 在语义关系图中的连通路程数 N, 和每条连通路程的长度  $L_i$ , 若  $N > 0$ , 则转到 Step2; 否则, 转到 Step3;
- Step2. 利用公式 (3) 计算 A 与 B 的相关度  $\text{Rel}(A,B)$ , 转 Step8;
- Step3. 查找结点 A 在连通路程长为 5 以内的连通路程上的所有结点, 这些结点组合为集合 U;
- Step4. 计算结点 B 与集合 U 中每个元素结点的相似度, 并标记相似度最大的结点 C 及其相似度的值 Sim, 若  $\text{Sim} > 0.85$ , 则转到 step5, 否则, 转到 Step6;
- Step5. 重复步骤 1~2, 利用公式 (3) 计算 A 与 C 的相关度  $\text{Rel}(A,C)$ , 并计算  $\text{Sim} \times \text{Rel}(A,C)$ , 记为  $\text{Rel}_i(A,B)$ , ( $i=1,2$ );
- Step6. 结点 A 和 B 角色互换, 采用同样的方法, 重复步骤 3~5;
- Step7. 计算 A 与 B 的相关度  $\text{Rel}(A,B) = \text{Max Rel}_i(A,B)$ , ( $i=0,1,2$ ), 转 Step8;
- Step8. 结束。
- 

### 3 实验及结果分析

按照 2.1 节所述的语义关系提取规则及语义关系图的扩展方法, 经过对 HowNet2012 以及人民日报 2000 年语料中数据的处理, 本文所构造的语义关系图共包括 836147 个语义关系。其中基于 HowNet2012 提取的语义关系共有 524921 个, 语义关系种类共有 98 种。

在已经构建好的语义关系图的基础上, 根据 2.2 节所述的词语相关度的计算方法, 我们实现了一个基于语义关系图的词语语义相关度计算程序, 通过该程序可以计算两个词语的语义相关度。实验中所用到的参数设置如下:  $\beta_1=1.0$ ,  $\beta_2=1.2$ ,  $\beta_3=1.4$ ,  $\beta_4=1.7$ ,  $\beta_5=2.0$ 。实验结果如表 2 所示:

表 2 语义相关度计算的实验结果

词语 1	词语 2	相关度
足球比赛	比分	0.450191
足球比赛	直播	0.421904
足球比赛	场地	0.301722
足球比赛	规则	0.396271
足球比赛	法庭	0.100823
滑冰	足球比赛	0.120737
滑冰	流畅	0.396185
滑冰	速度	0.420762
滑冰	摔倒	0.376215
滑冰	法庭	0.148226
足球比赛	流畅	0.012571

由表 2 中数据可以看出, 绝大部分结果是比较符合习惯上对相关度的主观判断, 且实验结果比较平稳, 不会出现极端值的问题。但从实验结果也可以看出, 部分结果不够理想, 例如: “滑

冰”和“法庭”的相关度比“足球比赛”和“法庭”的相关度高，导致这种情况的原因主要有两点：

- (1) 知网对有些词语的定义不够合理，导致词语间的语义关系产生误差；
- (2) 在通过依存句法分析器分析人民日报语料统计语义关系时，分析出错误的语义依存搭配关系，导致将错误的语义关系扩展到了语义关系图中，致使计算出不合理的相关度结果。

#### 4 结论及工作展望

本文以《知网》中概念之间以及概念的属性之间的语义关系和大规模语料中统计出的词语依存关系为基础，构建了一张语义关系图，然后，将此语义关系图与图理论的相关知识相结合，提出了一个基于语义关系图的词汇语义相关度计算的算法。实验表明，采用本文中所述的方法得到的相关度计算结果较为合理。

在接下来的工作中，我们计划进一步丰富语义关系图中的语义关联信息，比如根据语义关系为语义关系边赋予权重，然后应用图理论知识优化相关度计算的算法，期望得到更加真实有效的词汇语义相关度。

#### 参考文献

- [1] 张运良, 张全. 基于 HNC 理论的语义相关度计算方法[J]. 计算机工程与应用, 2004, 34: 1-4.
- [2] Resnik P. Using information content to evaluate semantic similarity in a taxonomy[C]// Proceedings of the 14th International Joint Conference on Artificial Intelligence. volume 1, Montreal, Canada, August 1995:448-453
- [3] 田萱, 杜小勇, 李海华. 语义查询扩展中词语-概念相关度的计算[J]. 软件学报, 2008, 19(8): 2043-2053.
- [4] 许云, 樊孝忠, 张锋. 基于知网的语义相关度计算[J]. 北京理工大学学报, 2005, 20(5): 411-414.
- [5] 王宏显, 周强, 邬晓钧. 《知网》语义关系图的自动构建[J]. 中文信息学报, 2008, 22(5): 90-96
- [6] 董振东, 董强. 知网[EBPOL]. [http://www.keenage.com/html/c\\_index.html](http://www.keenage.com/html/c_index.html)
- [7] 李彬, 刘挺, 秦兵, 李生. 基于语义依存的汉语句子相似度计算[J]. 计算机应用研究, 2003, (12): 15-17
- [8] 肖江涛. 基于本体的语义相关度算法研究[D]. 长沙: 国防科学技术大学研究生院, 2010, 1-81
- [9] 钟茂生, 刘慧, 刘磊. 词汇间语义相关关系量化计算方法[J]. 中文信息学报, 2009, 23 (2): 115-122.
- [10] 刘群, 李素建. 基于《知网》的词汇语义相似度计算[J]. 计算语言学及中文信息处理, 2002, 7: 59-76.