

# 基于百科资源的名词性隐喻识别<sup>①</sup>

冯帅<sup>1</sup>, 苏畅<sup>1</sup>, 陈怡疆<sup>2</sup>

<sup>1</sup>(厦门大学信息科学与技术学院 智能科学与技术系, 厦门 361005)

<sup>2</sup>(厦门大学信息科学与技术学院 计算机科学系, 厦门 361005)

**摘要:** 针对现有隐喻计算知识库的广度和深度不足问题, 提出了引入了网络百科资源作为隐喻识别的世界知识库的方法. 利用信息检索技术, 从概念对应的百科页面中获取概念的背景世界知识, 计算两个概念的世界知识重合程度, 作为判断它们是否属于同一个概念域的依据, 进行隐喻的识别. 实验结果表明在使用百度百科作为世界知识库时, 隐喻/常规表达的识别正确率达到 81.06%, 算法的有效性得到证明.

**关键词:** 名词性隐喻; 隐喻识别; 百科资源; 词语相关度; 隐喻可接受度

## Approach to Recognizing Chinese Nominal Metaphor Based on Online-Encyclopedia

FENG Shuai<sup>1</sup>, SU Chang<sup>1</sup>, CHEN Yi-Jinag<sup>2</sup>

<sup>1</sup>(Cognitive Science Department of Xiamen University, Xiamen 361005, China)

<sup>2</sup>(Computer Science Department of Xiamen University, Xiamen 361005, China)

**Abstract:** Aim at the low coverage of current hand-coded knowledge base used in the metaphor recognition, a Chinese nominal metaphor recognition method based on exploiting the online encyclopedia resource is proposed. A noun is represented as a page in the encyclopedia. Capture the world knowledge from the page of encyclopedia, calculates the relatedness between the pages of the two nouns in the nominal reference by applying information retrieval methods. The relatedness determines the reference is metaphorical or not. The experimental results shows that the accuracy of metaphorical/normal references recognition can reach 81.06% by using BaiduBaik. It supports the validity, efficiency of this method.

**Key words:** nominal metaphor; metaphor recognition; encyclopedia resource; words relatedness; acceptability of metaphor

## 1 概述

隐喻的机器处理在自然语言处理中占有重要地位. 它在机器翻译, 文摘生成, 信息检索和问答系统中都获得重要的应用. 本文的研究对象名词性隐喻是隐喻的一种最基本类型, 名词在这类隐喻中充当主语、表语、宾语、同位语等成分<sup>[1]</sup>. 它是隐喻中最基础的一种类型, 常见于“X vx(is) Y”指称型模式, 包含了隐喻最基本的两个要素“本体”和“喻体”, 从中发掘“喻底”是人类对事物认知的研究中重要的一环. 例如, “这个女孩的脸蛋像苹果那样”, “女孩的脸蛋”是本体, “苹果”是喻体, 而“红润”是从隐喻中推理出来的喻底.

隐喻的机器识别是隐喻机器处理的基础. 本文利用百科资源与信息检索的技术, 从现有的百科资源中提取构成概念的世界知识, 判断两个概念是否在一个概念域中, 从而对名词性隐喻进行识别.

## 2 前人工作

目前最流行的隐喻识别模型是优先语义模型<sup>[2]</sup>. 隐喻理解模型 MET\*系统<sup>[3]</sup>对违反语义选择限制进行了进一步细分, 可以处理隐喻, 转喻, 字面义, 反常表达等. 然而, MET\*是基于手工构造的知识库, 在知识库的广度和深度上有所欠缺, 实际应用受到限制.

① 基金项目: 国家自然科学基金(61075058)

收稿时间: 2013-03-30; 收到修改稿时间: 2013-05-02

文献[4]利用 WordNet 中名词概念的上下位关系和词语的二元模型来判断隐喻. 对于“X is Y”类型的隐喻, 使用 WordNet 中名词的上下位关系来判断, 如果 Y 是 X 的上位, 则构成字面表达, 否则是隐喻表达. 但是对于约定俗成的隐喻(称之为“死隐喻”), 则将其作为反例(非隐喻).

文献[5]考察隐喻与语义距离及语义关系的关联, 利用《同义词词林》的语义距离和 HowNet 的语义关系来进行隐喻的判断. 只有当两种资源都对表达判断为隐喻时, 才认为是隐喻表达, 否则认为是常规表达.

基于 WordNet、HowNet 和《同义词词林》等分类树知识库的隐喻识别方法是现有研究的主流, 比较符合人类对事物的认识. 然而, 在分类树中找到的类别属性, 对于比喻思想来说可能不是最好的或者最明显的属性<sup>[6]</sup>. 例如, 对于以下隐喻:

集中营中这个医生就是个屠夫.

因为“医生”和“屠夫”都共有上位义“职业”, 这个共有属性的存在使得两个概念具有非常短的分类树路径, 无法正确识别出隐喻.

近年语义计算的研究开始转向使用网络百科资源. WikiRelate 算法<sup>[7]</sup>将一些在 WordNet 上具有代表性的方法都在 Wikipedia 上重新实现, 而 WPNRelate 算法<sup>[8]</sup>利用 Wikipedia 中出链入链和类别网信息来计算语义相似度, 实验结果表明优于大多数基于 WordNet 的算法. 文献[9]将 PageRank 和模式匹配等算法应用于百度百科中, 用以自动抽取同义词. 然而, 隐喻计算任务作为语义计算的一个应用, 目前仍没有将百科资源引入到隐喻计算的相关研究.

### 3 基于百科知识的名词性隐喻识别方法

#### 3.1 隐喻识别的本质

隐喻的一个本质属性是“同从异出”. 用与本体没有相关的事物作为喻体来描述本体, 在表层语义上构成了语义矛盾. 接收者接收到这个表达后, 根据所掌握的前置世界知识, 从构成表层语义矛盾的双方中发现相似性, 理解隐喻的深层意义<sup>[10]</sup>. 隐喻识别的本质, 就是判断表达中两个事物之间是否构成了语义矛盾.

#### 3.2 语义相关性在隐喻识别中的应用

如何判断语义矛盾, 是隐喻识别的主要问题. 文献[11]指出, 本体和喻体的语义上符合“同从异出”或“异中求同”的原则, 即本体和喻体来自不同的概念域,

同时本体和喻体在某一点上有相似性. 这指出了隐喻识别的主要手段是确定本体和喻体来自不同的概念域.

一个概念处于一个概念域中的前提是这个概念蕴涵了这个概念域中足够多的背景世界知识. 对于两个概念, 假如他们所蕴涵的共有世界知识越多, 他们越有可能处于同一个概念域; 而相反, 如果他们所蕴涵的共有世界知识越少, 他们来自不同的概念域的可能性越高. 基于分类树和共有知识的隐喻识别思想在文献[12]中被应用.

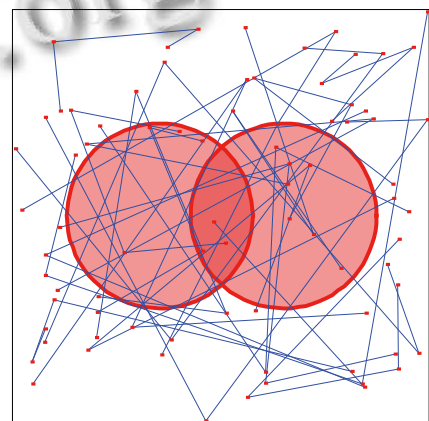


图 1 构成概念的世界知识和知识重合示意图

如图 1 所示, 红色大圆部分是两个不同的概念, 而红色小圆点为构成概念的世界知识, 蓝线是世界知识之间的语义联系(同义, 反义, 上下位, 整体部分等). 假如两个大圆的重合部分越小, 他们来自不同的概念域的可能性越高.

对概念间世界知识共有程度的量化, 与语义计算中的语义相关度的思想相类似, 因此, 我们将其称为隐喻识别中概念间的语义相关度.

#### 3.3 使用百科进行隐喻识别的原因

一切自然语言处理系统本质都是基于知识的系统, 如果机器要实现对一个表达的理解, 所需要的知识与人类完成同样任务所需要的知识是无差别的<sup>[13]</sup>.

对表层语义判断是否存在语义矛盾, 主要是根据表达的语义违背了我们过去所掌握的世界知识, 两个概念来自不同的概念域. 隐喻的机器处理在知识库的使用上, 经历了从手工构造概念映射库(Master Metaphor List, ATT-meta 等), 到手工构造的语义网知识库(WordNet, HowNet, 同义词词典, CCD 等)的过程.

手工知识库可以看作是世界知识的一个结构化和组织化的反映,然而,手工构造使得其具有代价高,规模小的缺点,不能用于实际应用.而作为知识最自然和最能被人们所接受的表现,百科知识在隐喻的计算中却从来没有获得应用.

与其他语义计算任务一样,通过对百科进行一定的处理,可以实现过去手工知识库所具有的作用.过去基于知识库的隐喻识别研究,判断两个概念是否在同一个概念域中的方法都是发展自语义计算中基于手工知识库的语义相关/相似计算方法,而这些相关/相似计算方法都已经在百科知识库中重新实现.借助目前基于百科知识语义计算的研究,将其引入到隐喻的识别任务中,以弥补过去基于手工知识库在广度和深度上的不足.

百科中对于每个概念的描述相对于手工知识库更加丰富,同时也包含有更多在描述这个概念时候人类认知中对这个概念的主观思想.隐喻的识别是一种认知的行为,更应该注重从认知的角度出发而选择资源.“医生是屠夫”中,“屠夫”在百度百科的释义中就包含了“血腥”、“屠杀”等字眼.这些丰富的释义可以使得对本体和喻体的认知更充分.而如果只依赖手工知识库中他们所共有上位义“职业”,会很容易的被判断为常规表达.

百科中的每个页面都表示单独的一个概念,在大广度上提供了高可读性、目录化组织和内容丰富的释义,因此,使用百科作为概念的背景世界知识是合理的.除此以外,百科也在深度和广度上不断稳定增强.

### 3.4 基于百科的名词性隐喻识别语义相关研究

综上所述,使用百科作为隐喻识别的世界知识库的基本过程是:从百科中挖掘出概念的背景世界知识,量化概念间世界知识的共有程度,以此作为判断两个概念在同一概念域的可能性的标准,从而确定是否构成隐喻.

百科中的页面包含了三种资源,分别是类别资源、释义文本资源和释义链接资源.针对三种不同的资源,拟定方法来计算表达中被指称者与指称者间的三种相关度,分别为类别相关度、文本相关度和链接相关度.类别相关度根据路径算法计算,文本相关度和链接相关度根据共有知识算法计算.图 2 为从抽取文本释义和链接释义计算知识共有程度的示意图.共有的世界知识越少,则相关度越低,越有可能处于不

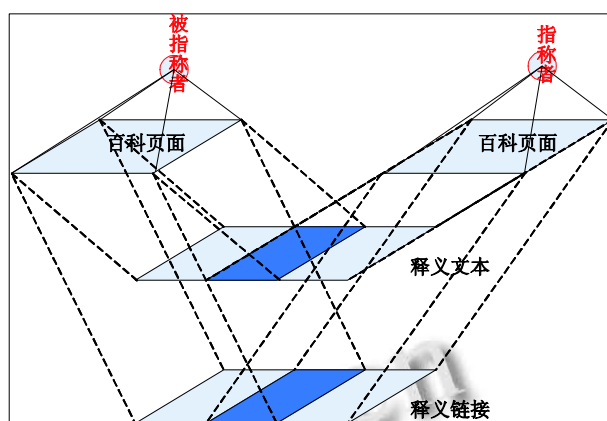


图 2 释义文本和释义链接抽取重合示意图

同的概念域中,构成隐喻表达.参考机器学习中 bagging 方法,利用投票机制进行判断:

$$\text{表述是} \begin{cases} \text{隐喻表达, } Relateness_{category} = 0 \text{ 且} \\ \quad Relateness_{text} < a \text{ 且} \\ \quad Relateness_{link} < b \\ \text{常规表达, 否则} \end{cases} \quad (1)$$

$\alpha$  和  $\beta$  分别是文本相关度  $Relateness_{text}$  和链接相关度  $Relateness_{link}$  的阈值.只有当类别不相关,文本资源和链接资源的共有世界知识达不到要求时才被判断为隐喻.

### 3.5 基于百科类别的相关度计算

在实际处理中,会遇到很多带有上下位关系的表达,其中被指称物是下位,指称物是上位,如:“病菌是生物”,“金钱是财富”等.两个概念必然处于同一个概念域中.在识别的过程中,这种带有上下位关系的表达都可以优先被确定为常规表达.

现有的网络百科,继承了传统百科资源中的分类体系.而且在这个基础上,有了进一步改进.网络百科中的概念都有它所述的分类,同时每个概念不仅只属于一个类别,可以分属于多个分类树.百科中的父类和子类的关系等同于过去语义网中的上下位关系.因此可以在处理的时候利用文献[4]的识别方法排除带有父类-子类关系的常规表达.当指称者是被指称者的父类,或者祖先类时,类别相关度为 1,否则类别相关度为 0.

$$Relateness_{category}(A, B) = \begin{cases} 1, \text{ 指称者是被指称者的父类,} \\ \quad \text{或祖先类} \\ 0, \text{ 否则} \end{cases} \quad (2)$$

### 3.6 基于百科文本的相关度计算

百科中的每个文章都只阐述一个概念, 在文本处理中的 TF-IDF(term frequency-inverse document frequency) 值是衡量文章中不同词语和文章的相关程度. 在一个概念域中, 有一些词语对定义这个概念域有重要影响. 如果两个概念共有某些对他们的影响力很大的词语, 就有更大的可能性处于同一个概念域中. 因此, 利用 TF-IDF 获得概念释义文本中各个词语的重要程度, 再根据两个概念的 TF-IDF 向量来确定重要词语的重合程度, 可以作为判断是否处于同一个概念域中的依据.

假设百科中词语的总数为  $N$ , 则可以表示成一个  $N$  维的空间, 将概念  $X$  对应的释义文本表示为这个  $N$  维空间中的一个向量  $T = \langle t_1, t_2, \dots, t_N \rangle$ . 其中元素  $t_i$  表示的百科中的第  $i$  个词语  $w_i$  对概念  $X$  作出定义的贡献值:

$$t_i = tf_i \times idf_i = \frac{k_i}{\sum_{i=1}^N k_i} \times \log\left(\frac{N}{m_i}\right) (1 \leq i \leq N) \quad (3)$$

$k_i$  为  $w_i$  在该释义文本中出现的次数,  $\sum_{i=1}^N k_i$  表示的是这个释义文本中的总词数.  $m_i$  为百科中含有  $w_i$  的文档数.  $tf_i$  (term frequency, 词频) 是  $w_i$  在该释义文本出现的频率;  $idf_i$  (inverse document frequency, 逆向文件频率) 是  $w_i$  在百科总文档中的普遍重要性. 某一页面内的高频率词语, 以及该词语在整个百科文档集合中的低文档频率, 可以产生出高 TF-IDF 值, 表明该词语对于这个页面具有高重要性. 因此, 有如下的  $Relateness_{text}$  算法:

1. 当得到被指称者 A 和指称者 B 所命中的百科页面, 利用分词工具分词和剔除停用词后, 求出他们各自的  $N$  维 TF-IDF 向量  $\bar{A}$ ,  $\bar{B}$ .

2. 两个概念的释义文本相关度可以转化为所命中页面之间的相似度确定, 页面的相似度通过成熟的夹角余弦方法来获得:

$$Relateness_{text}(A, B) = \frac{\sum_{i=1}^N a_i \times b_i}{\sqrt{(\sum_{i=1}^N a_i^2) \times (\sum_{i=1}^N b_i^2)}} \quad (4)$$

### 3.7 基于百科链接的相关度计算

百科中含有大量的链接信息, 每一个链接都质心百科内的页面中, 这类类似于科研论文中的引用与参考文献. 一个概念对应页面中的链接(简称释义链

接)的集合, 可以看作构成和支撑这个概念的前置世界知识. 和科研论文一样, 一个概念有更高的概率引用本概念域中的概念. 因此, 挖掘两个概念释义链接集合的共有部分, 可以看作量化支撑两个概念的世界知识的共有部分, 从而判断概念是否处于同一个概念域中.

从概念  $X$  所命中页面中获取到链接后, 链接较少, 为了避免数据稀疏, 通过 BFS(breadth-first search, 广度优先搜索) 算法来生成以  $X$  所命中的(第 1 层)的一棵树, 其中离根越近的节点与  $X$  的相关性越大. BFS 算法以一个根节点开始, 访问节点的所有邻居(出链), 然后对于每一个邻居, 以其为根节点进行 BFS. 根据节点所在的层次高度赋予权值, 表示节点与  $X$  的相关程度. 出于时间复杂度考虑, 仅考虑树高 3 层的情况.

假设百科共有  $M$  个页面.  $X$  可以表示为一个  $M$  维的向量  $L = \langle l_1, l_2, \dots, l_M \rangle$ ,  $l_i$  为页面  $p_i$  对  $X$  的贡献值:

$$l_i = \frac{c_{ij}}{\lambda_j} \times 2^{-j} \times \lg\left(\frac{M}{m_i}\right) \quad (5)$$

当  $p_i$  表示的节点在树上不同的层次出现时, 只保留离根最近的节点, 其层数记为  $j$ ,  $c_{ij}$  为  $p_i$  在第  $j$  层出现的次数.  $m_i$  为百科中包含  $p_i$  链接的页面的数目(入链),  $\lambda_j$  为第  $j$  层节点总数.

$$\lambda_j = \sum_{i=1}^M c_{ij} \quad (6)$$

$c_{ij} / \lambda_j \times 2^{-j}$  表示的是  $p_i$  这个节点在这棵生成树上出现的频次;  $\lg(M / m_i)$  表示的是  $p_i$  对于整个百科的普遍性,  $m_i$  越大, 则普遍性越高, 对于  $X$  的重要性越低.  $p_i$  在这棵树上出现的越多, 离  $X$  越近, 在整个百科中出现的次数越少, 对  $X$  的贡献值越高.

因此有如下的  $Relateness_{link}$  算法:

1. 当得到被指称者 A 和指称者 B 所命中的百科文档后, 利用 BFS 算法获取各自的链接树;

2. 根据链接树, 分别获得 A, B 的释义链接向量  $\bar{A}$ ,  $\bar{B}$ ;

3. 通过计算向量的夹角余弦来计算这两个名词的释义链接相似性.

$$Relateness_{link}(A, B) = \frac{\sum_{i=1}^M a_i \times b_i}{\sqrt{(\sum_{i=1}^M a_i^2) \times (\sum_{i=1}^M b_i^2)}} \quad (7)$$

## 4 实验数据

### 4.1 百科知识库

本实验主要使用百度百科作为知识库. 它是目前最大的中文网络知识库, 共收录词条超过 5600 万条, 而且在医学等方面邀请专家审核, 从而确保内容具有良好的质量.

### 4.2 测试数据

测试使用文献[14]中的名词性隐喻表达作为隐喻样本库. 库中的条目被整理成“A 是 B”的形式, A 与 B 都是实名词而不是指代(含有指代的表达应该进行指代消解). 每个条目含有“中文隐喻句可接受度”(简称隐喻可接受度), 是一个由 15 个人所作出评价的平均值, 取值范围为[1, 5], 其中 1 表示这个句子“很难被接受为隐喻”, 5 表示这个句子“很容易被接受为隐喻”. 只选取隐喻可接受度在 3.4 及以上的条目作为测试数据, 确保数据能被大部分人接受. 测试中并没有“死隐喻”<sup>[4]</sup>的区分, 实际上, 在对“死隐喻”没有做出严格定义且能够被大部分人所接受的前提下, 从测试数据中将被认为是“死隐喻”的隐喻样本排除在外会影响到实验的客观性. 以下是部分的测试样本:

顾客是上帝.  
兄弟是手足.  
医生是屠夫.

测试使用了 Words-240<sup>[15]</sup>作为常规样本库. 每个样本都包含一对词语和一个评测值, 这个值是由 20 个人做出评判后的平均值. 评测值的范围是[0, 10], 0 表示的是这对词语“毫不相关”, 10 表示的是这对词语是“同义词”. 同样, 只选用评测值在 5.33 以上的条目作为测试数据, 保证数据能够被大部分人接受.

## 5 实验及结果分析

### 5.1 实验步骤

实验拟定了如下步骤:

#### ① 数据预处理

接受了一个指称表达, 系统分别获取被指称者 A 与指称者 B 所命中的页面. 利用系统提供的 API 或者分析器来抽取纯文本和链接信息.

#### ② 计算基于类别的概念相关度

利用 3.5 节的类比相关度算法判断指称者是否是被指称者的父类或祖先类, 是的则是常规表述, 算法结束, 否则进入③.

#### ③ 计算基于文本的概念相关度

①中获取指称者和被指称者的文本后, 利用 3.6 节基于文本的算法计算指称者与被指称者的文本相关度  $Relateness_{text}(A, B)$ .

#### ④ 计算基于链接的概念相关度

①中抽取到的指称者和被指称者的链接后, 利用 3.7 节的基于链接的算法计算指称者和被指称者的链接相关度  $Relateness_{link}(A, B)$ .

#### ⑤ 隐喻的判断

获取  $Relateness_{text}(A, B)$  与  $Relateness_{link}(A, B)$  后, 利用 3.4 节进行名词性隐喻的判断. 只有当  $Relateness_{text}(A, B)$  小于  $\alpha$  且  $Relateness_{link}(A, B)$  小于  $\beta$  时, 表达被判定为名词性隐喻, 否则为常规表达. 算法结束.

### 5.2 结果分析

4.2 节中的隐喻测试集中, 扣除了在百度百科中未登录名词后, 共计有 153 例; 在常规表达测试集中随机选取 153 条指称者和被指称者在百科中均有登陆的样本, 构成用于识别的 306 条样本. 本文使用召回率和准确率作为实验的有效性指标:

$$\begin{aligned} \text{准确率} &= \frac{\text{系统正确识别的隐喻(或非隐喻)}}{\text{系统识别的隐喻(或非隐喻)}} \times 100\% \\ \text{召回率} &= \frac{\text{系统正确识别的隐喻(或非隐喻)}}{\text{实际的隐喻(或非隐喻)}} \times 100\% \\ F &= \frac{2 \times \text{准确率} \times \text{召回率}}{\text{准确率} + \text{召回率}} \times 100\% \end{aligned}$$

表 1 为在使用百度百科作为知识库  $\alpha=0.4$ ,  $\beta=0.3$  的情况下, 名词性隐喻识别的结果. 结果显示本算法能够有效识别隐喻与非隐喻, 语义关联可以作为隐喻判定的一个标准.

表 1 使用百度百科作为知识库,  $\alpha=0.4$ ,  $\beta=0.3$  的情况下, 名词性隐喻识别的结果

	正确率	准确率	召回率	F
隐喻表				
达	248/306	126/157=80.25%	126/153=82.35%	81.29%
常规表	=81.06%			
达		122/149=81.88%	122/153=79.74%	80.80%

表 2 和表 3 是部分隐喻/常规表达的识别结果, 两个表中均未列出  $Relateness_{category}$  的结果, 因为  $Relateness_{category}$  的结果均为 0, 进入使用文本相关度和链接相关度进行识别的过程. 实验主要分析识别错误的样本. 对于被误识别为常规的隐喻表达, 是一些如“质量就是生命”谚语型的隐喻表达, 新颖性不强, 反映了有一部



分的隐喻表达由于长期使用,已经转化为常规表达,这种隐喻亦称为“死隐喻”<sup>[4]</sup>;而“汉奸”和“走狗”本身在中文中经常同时出现,所以关联性很强,而被误判为常规表达。对于被误识别为隐喻的常规表达,大部分是因为相关两个概念的世界知识有所欠缺,而导致误判,这也体现了,从更大范围的知识库(如互联网)中获取更多的背景知识,对常规表达的识别非常重要。

表 2 部分隐喻表达的识别结果

A	B	$Relateness_{text}$	$Relateness_{link}$	识别结果
顾客	上帝	0.015890	0.085614	隐喻
兄弟	手足	0.131824	0.190000	隐喻
医生	屠夫	0.021313	0.032751	隐喻
老师	蜡烛	0.028928	0.243685	隐喻
江山	图画	0.129772	0.296052	隐喻
商人	狐狸	0.024314	0.208247	隐喻
灾难	噩梦	0.069515	0.000000	隐喻
交通	血脉	0.023884	0.216912	隐喻
质量	生命	0.154064	0.456341	常规
汉奸	走狗	0.632706	0.174449	常规

表 3 部分常规表达的识别结果

A	B	$Relateness_{text}$	$Relateness_{link}$	识别结果
李白	诗	0.206134	0.654080	常规
发动机	汽车	0.336212	0.681899	常规
演唱会	歌手	0.051763	0.441589	常规
环境	生态	0.261340	0.437551	常规
义务	权利	0.486027	0.636053	常规
奥运会	北京	0.061678	0.625668	常规
行星	地球	0.293114	0.784055	常规
图片	照片	0.180090	0.338446	常规
计算机	实验室	0.080480	0.083449	隐喻
性别	平等	0.085257	0.193456	隐喻

### 5.3 与过去工作的比较

与过去基于手工知识库的方法的实现思路相同,本文也是基于指称物和被指称物之间是否处于同一个概念域作为衡量表达是否隐喻的根据。而在实现的方法上本文使用的是解释概念的百科资源中,所共有的信息作为判断两个概念是否处于同一个概念域的标准,从而识别隐喻;过去基于手工知识库的识别方法是利用被指称物和指称物在知识库中的路径长度作为判断是否处于同一概念域的标准。对于同一份隐喻样本集,基于手工知识库的方法获得的准确率为 73.28%<sup>[5]</sup>,而本方法的准确率为 80.25%。

然而,隐喻的识别是一个集成性的任务,单纯依

赖一种方法都不能达到理想的效果。两种知识库本身固有的特点所决定他们适合不同的识别任务。因此,将两种方法集成起来是解决问题的途径之一。

## 6 结语

本文利用现有的网络百科资源,提取名词性表达中的指称者与被指称者对应的解释页面,分解为类别、文本与链接三部分信息,通过信息检索的方法来确定概念背景世界知识间的相关度,判断是否处于同一个概念域中,从而确定名词性表达是否隐喻。从实验结果分析,本算法能有效对名词性隐喻进行识别。

本算法仍存有如下不足和改进的空间:

1. 没有对释义文本中的语义进行进一步深入的挖掘,下一步的研究将会挖掘释义文本中更多的语义信息用于名词性隐喻识别的研究。

2. 对于网页文档而言,权重的计算方法应该体现出文档的结构特征,特征词在不同部分中对文档内容的反映程度也不尽相同。

3. 基于链接的算法由于需要访问释义页面的出链,所以计算量比较大。如果将释义页面的出链预先保存起来,可以减少计算量。

将网络百科和信息检索技术用于隐喻的机器处理,本文只是个初始的探索。上述的实验结果表明,基于信息检索的名词性隐喻识别计算,具有优点的同时,也具有一定的不足。全面的隐喻识别计算仍需要多种方法的判断。探索百科资源,挖掘百科资源中的语义信息是后续研究中值得关注的课题。

## 参考文献

- 束定芳.隐喻学研究(第一版).上海:上海外语教育出版社,2000:60-66.
- Wilks Y. Making preferences more active. *Artificial Intelligence*, 1978, 11(3): 197-223.
- Fass D. Met\*:a method for discriminating metonymy and metaphor by computer. *Comput Linguist*, 1991, 17(1): 49-90.
- Kriahnakumaran S, Zhu X. Hunting elusive metaphors using lexical resources. *Proc. of the Workshop on Computational Approaches to Figurative Language*. Stroudsburg:Association for Computational Linguistics, 2007. 13-20.
- 贾玉祥,俞士汶.基于词典的名词性隐喻识别.中文信息学报, (下转第 59 页)

#### 4 结束语

本文结合 JBPM 工作流相关技术, 在 J2EE 开发环境下开发出具有可视化设计功能的流程设计器, 有利于开发人员高效完成具体业务办理流程的设计和工作流管理系统的发布, 提高了开发效率, 此流程设计器目前已成功应用于广州市药监局二期电子政务管理平台下的稽查、审评认证和行政审批三个子系统。接下来的主要工作: ①进一步完善流程设计器对具体业务办理流程可视化界面的设计; ②对流程设计器中的相关组件进行二次开发, 进一步提高具体业务工作流系统的运行效率。

#### 参考文献

- 1 唐协平,张鹏翥.电子政务需求研究综述.计算机应用研究, 2011, 25(2): 99-104.
- 2 Cumberlidge M. Business Process Management with Jboss JBPM. UK: Packt Publishing Ltd, 2007.
- 3 刘军,王广杰.基于 JBPM 和轻量级 J2EE 的协同办公系统的研究.福建电脑,2007,12:99-100.
- 4 Johnson R, Hoeller J, Arendsen A. Spring Java/J2EE application Framework. 2004.
- 5 强锋科技,那静.Eclipse SWT/JFace 核心应用(第 1 版).北京:清华大学出版社,2007:446-463.
- 6 陈冈.Eclipse RCP 应用系统开发方法与实战(第 1 版).北京:电子工业出版社,2007:20-26.
- 7 Strube M, Ponzetto SP. WikiRelate computing semantic relatedness using wikipedia. Proc. of the 21st National Conference on Artificial Intelligence-Volume 2. Palo Alto: AAAI Press, 2006: 1419-1424.
- 8 盛志超,陶晓鹏.基于维基百科的语义相似度计算方法.计算机工程,2011,37(7):193-195.
- 9 陆勇,章成志,侯汉清.基于百科资源的多策略中文同义词自动抽取研究.中国图书馆学报,2010,36(185):56-62.
- 10 苏畅.汉语名词性隐喻的计算方法研究[博士学位论文].厦门:厦门大学,2008.
- 11 王治敏.名词隐喻相似性及推理识别研究.中文信息学报, 2008,22(3):37-43.
- 12 杨芸.汉语隐喻识别与解释计算模型研究[博士学位论文].厦门:厦门大学,2008.
- 13 孙茂松,黄昌宁,方捷.汉语搭配定量分析初探.中国语文, 1997,1:29-38.
- 14 李玉莲.本体和喻体在中文隐喻句理解中的应用[硕士学位论文].重庆:西南大学,2007.
- 15 汪祥,贾焰,周斌,丁兆云,梁政.基于中文维基百科链接结构与分类体系的语义相关度计算.小型微型计算机系统, 2011,32(11):2237-2242.

(上接第 13 页)

2011, 25(2): 99-104.

6 Veale T, Hao Y. Comprehending and generating apt metaphors: a web-driven, case-based approach to figurative language. Proc. of the 22nd National Conference on Artificial intelligence-Volume2. Palo Alto: AAAI Press, 2007: 1471-1476.

7 Strube M, Ponzetto SP. WikiRelate computing semantic relatedness using wikipedia. Proc. of the 21st National Conference on Artificial Intelligence-Volume 2. Palo Alto: AAAI Press, 2006: 1419-1424.

8 盛志超,陶晓鹏.基于维基百科的语义相似度计算方法.计算机工程,2011,37(7):193-195.

9 陆勇,章成志,侯汉清.基于百科资源的多策略中文同义词自动抽取研究.中国图书馆学报,2010,36(185):56-62.

10 苏畅.汉语名词性隐喻的计算方法研究[博士学位论文].厦门:厦门大学,2008.

11 王治敏.名词隐喻相似性及推理识别研究.中文信息学报, 2008,22(3):37-43.

12 杨芸.汉语隐喻识别与解释计算模型研究[博士学位论文].厦门:厦门大学,2008.

13 孙茂松,黄昌宁,方捷.汉语搭配定量分析初探.中国语文, 1997,1:29-38.

14 李玉莲.本体和喻体在中文隐喻句理解中的应用[硕士学位论文].重庆:西南大学,2007.

15 汪祥,贾焰,周斌,丁兆云,梁政.基于中文维基百科链接结构与分类体系的语义相关度计算.小型微型计算机系统, 2011,32(11):2237-2242.