

# 汉语名词的隐喻知识表示及获取研究\*

汪梦翔<sup>1, 2</sup>, 饶琪<sup>2</sup>, 顾澄<sup>3</sup>, 王厚峰<sup>2</sup>

(1. 北京联合大学 师范学院, 北京, 100011

2. 北京大学 计算语言学研究所, 北京, 100871

3. 清华大学 五道口金融学院, 北京, 100083)

**摘要:** 隐喻知识的表示和获取是进行隐喻计算的基础。本文把隐喻知识看做是本体和喻体的特征和属性之间的关联。我们主要通过惯用语导入和句法模式识别两种机制, 来获取名词的隐喻知识。惯用语的隐喻比较固定和单一, 我们从专门的词典释义中就能获取准确的隐喻特征和属性。而一般名词的隐喻知识比较复杂, 我们主要依托语料库和搜索引擎, 通过关键词和句法匹配来获取同一名词不同的隐喻特征及对应属性。该项工作的结果, 对于隐喻句隐喻意义的获取和名词语义属性特征的描述体系具有一定的价值。

**关键词:** 中文信息处理; 隐喻计算; 模式识别; 惯用语

中图分类号: TP391

文献标识码: A

## Study of Metaphorical Knowledge Expression and Acquisition for Chinese Nouns

Wang Mengxiang<sup>1, 2</sup>, Rao Qi<sup>2</sup>, Gu Cheng<sup>3</sup>, Wang Houfeng<sup>2</sup>

(1. Teachers' College, Beijing Union University, Beijing 100011, China;

2. Institute of Computational Linguistics, Peking University, Beijing 100871, China;

3. PBC School of Finance, Tsinghua University, Beijing 100083, China)

**Abstract:** The Acquisition and Expression of metaphorical knowledge is the basis of metaphor computation. In this paper, metaphor knowledge is regarded as the relationship between Original and target domain, which conclude the features and attributes. In order to get Chinese concrete nouns' metaphor features, we use 2 methods: one is idiomatic expression importing and the other is syntax pattern matching. For the metaphor knowledge of idiom, we mainly derive the accurate metaphorical features and attributes from idiomatic dictionary definitions. As for general concrete nouns, metaphor knowledge is more complex. We mainly rely on corpus and search engine to acquire different metaphorical features and corresponding attributes of the same noun by keyword and syntactic matching. This work may make sense in the construction of Chinese nouns' semantic feature system and the understanding of Chinese metaphorical sentences.

**Keywords:** Chinese information processing; computational metaphor; pattern matching; idiomatic.

### 1. 引言

(1) 他是一只白眼狼。

(2) 他是一头猪。

这里两个句子都是“X是Y”型隐喻句。“白眼狼”这个惯用语多半比喻忘恩负义的人, 我们一般理解为“他像白眼狼一样忘恩负义”, 但是后句我们可以理解的意义很多, 比如: “他像猪一样脏”、“他像猪一样蠢”、“他像猪一样肥”、“他像猪一样丑”等。

文献[1]中指出隐喻其实就是一种两个概念的特征继承, 传统的隐喻知识只是描述X和Y是否存在语义的映射关联, 但是这种知识表示不能反映X和Y之间建立关联的特征项, 而且也不一定Y上所有的特征都会被X所具备, 因为X对Y特征的继承是可选择性的, 于是就会出现像(2)句的多种理解。基于此, 本文想提出一种新的隐喻知识表示方式, 以便能够更深入的反映为什么X与Y能存在某种隐喻关联, 以及他们在语义特征中的相似性问题。

另外, (1)句中的惯用语的隐喻特点也引起了我们的思考。惯用语因由社会约定俗成而产生, 所以具

---

\* 基金项目: 国家自然科学基金(61602040); 国家863项目(2015AA015402); 北京市教委科技计划项目(KM201711417006)

备语义的固定性和深层的隐喻性<sup>[2]</sup>。可见惯用语一般都伴随隐喻，而且所形成的隐喻含义比较固定，一般很难随着语境发生改变。比如“中山狼”、“二把刀”等名词性结构的惯用语比喻的对象都不超过2种，这比起一般名词“天气”、“时间”、“狗”等的隐喻义要单一得多。比如文献[3]中就指出“时间”所能映射对应的隐喻概念多达十余种，有“动体、人、有价物、工具、容器、植物”等。而且“时间”放入不同的语境中，就会有着不同的隐喻对象。而一般像“白眼狼”、“铁公鸡”这种所映射的隐喻概念就比较单一和固定又不受语境影响，比如“铁公鸡要请客”、“他平时就是个铁公鸡”、“铁公鸡走了”中的“铁公鸡”一般人理解都是隐喻“吝啬的人”。鉴于此，本文在提取隐喻知识时，注重了静态的惯用语的释义数据导入，并且和动态的日常性语料结合起来提取名词的隐喻知识，从而为隐喻知识计算提供更深层的语义信息和新的知识来源。

## 2. 相关研究

在中文信息处理领域，目前对于汉语隐喻方面的研究，主要集中于隐喻句的机器识别和分类上<sup>[4]、[5]</sup>，对于隐喻知识的表示和获取方法的专门研究相对较少，不过隐喻识别和隐喻知识获取紧密相关。

总体来看国内对隐喻知识的提取有两种思路，第一种思路是基于规则的，这规则依托的是形式上的判断，一般是根据特定的特征词或语法结构，比如“像、是”等关键词或“如……般”等常用隐喻结构来识别和提取隐喻知识。李斌<sup>[6]</sup>、杨芸<sup>[7-9]</sup>就是用的这种思路。这种思路最为基础，但也存在一些问题：第一就是现实语料中符合这些框架的不是很多；第二，即使带有特征词，判断起来还是会有难度。比如“张三是地主”、“张三是神”同样的形式，但不是都含有隐喻色彩。贾玉祥<sup>[10-11]</sup>，针对这两个问题进行了改进，一方面从web海量数据中根据规则提取高质量的隐喻知识，另一方面利用词语的上下文关系特征来改进了“A是B”型的隐喻判断。比如面对“X是Y”型的句子，如果X是Y的下位，则“X是Y”的搭配为正常搭配，否则为隐喻。比如“香蕉是水果”是正常搭配，“世界是个舞台”就是隐喻搭配，需要抽取出来。但是有时候X、Y处于同一个语义类下也有可能是隐喻关系。比如“集中营中的医生就是屠夫”，这里的“医生”和“屠夫”都属于“职业”这个类别的下属分类，但是却带有隐喻性。

第二种思路是基于机器学习的方法，以统计为手段，为特定名词选择一定的语料，标记隐喻和非隐喻用法，用标记语料训练分类器，从新的名词性短语中识别出隐喻，进而就可以提取隐喻关联。王治敏<sup>[12-13]</sup>就是采用这种方法进行识别和提取。这种思路虽然识别准确，提取的质量也高，但是首先要为选取的每个名词人工标注一定数量的语料，才能进行相应的机器学习，因此非常耗费人力。

不过，上述思路都只是找到简单的词语隐喻映射关联，无法找到它们之所以建立关联所其核心作用的特征，而想要进一步的挖掘带有特征和属性的隐喻知识还需要对本体和喻体以外的信息进行梳理。另外，虽然很多人通过语料或者借助语义知识库来提取隐喻，但是都没有借助惯用语这种隐喻固定而且特征明显的语义资源，所以我们可以把它纳入到我们的抽取范围当中。本文就是针对惯用语和一般性的名词进行了相关的隐喻知识抽取工作，不过两者特点不一样，所以我们进行了分别对待。

## 3. 名词隐喻知识的知识表示

以往隐喻知识的表示基本都是依托概念隐喻理论，只涉及源域和目标域两个概念<sup>[14]</sup>。但是其实分属于两个不同概念域的X和Y，要建立某种语义上的映射关联而达到一种语义隐喻的效果，一般需要它们语义蕴含的某种特征有着某种相似或交集<sup>[15]</sup>。为此我们在描述词语的隐喻知识时，除了要明确哪些词语能够彼此建立这种语义映射关联，更应该提取其中蕴含的共同特征。而这些特征又是附着在某些属性上的，比如“猪”有“丑”的特征，也有“肥”的特征，这些特征其实都可以理解为“猪”这个概念中的“外形”属性的不同取值。所以我们说“他像猪一样丑”，其实是概念“他”和概念“猪”在“外形”属性上有着共同的特征值“丑”。具体如下图所示：

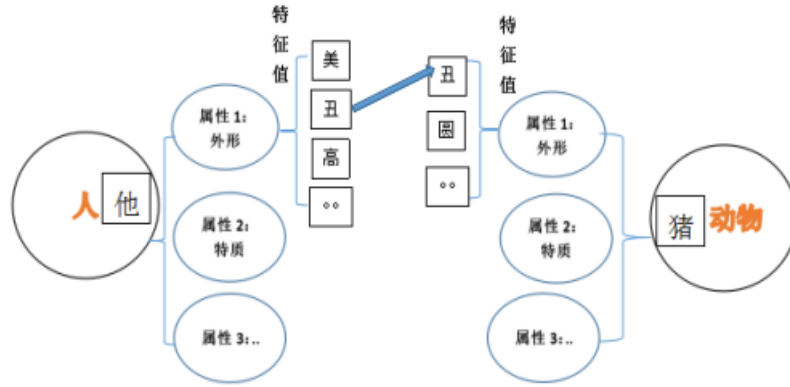


图1 “他像猪一样丑”的隐喻映射分析

基于上述思路，本文隐喻知识的描述不是表面意义的语义关联映射（如“他像猪一样”，不是简单的描述为“他”可以映射“动物”），而是要展示概念内部更深层次的语义关联，包括概念隐喻中具有相同属性以及相关特征值的关系。本文所指名词所蕴含的隐喻知识，主要是指要获取它的隐喻特征以及属性。比如“他像一只狼一样凶狠”，这里虽然是“他”和“狼”构成隐喻映射，但是这种隐喻知识包含的是相关特征以及内部属性的关联性，这里“凶狠”就是把“他”和“狼”建立起映射的特征之一，其属性值可以描述为“特质”。再比如“他像狼一样反映迅速”中的“迅速”作为“狼”的又一个隐喻特征，只不过属性由“特质”变成了“速度”。一般情况下，不同的映射关联特征对应不同的映射属性。例如，对于一个词语“花朵”它的隐喻特征可以有“鲜艳”、“美丽”、“芬芳”，而对应的隐喻属性分别有“色彩、形状、气味”。

我们可以把“ $X \rightarrow Y$ ”的隐喻知识标记为： $X \rightarrow Y = X_{(属性+特征)} \cap Y_{(属性+特征)}$

在具体的描述中，我们可以将“ $X \rightarrow Y$ ”隐喻知识采取二维表格的形式：

表1：概念隐喻的知识表现形式

X概念	属性	特征	Y语义映射概念
猪	外形	丑、胖、重、	人
	特质	愚蠢、可爱、能吃、懒、没用	
	速度	慢	
	...		
狼	外形	帅、威武、霸气	王
	速度	快、迅速	
	特质	凶狠	
	...		
花朵	外形	美	人
	色彩	鲜艳、灿烂	
	气味	芬芳	
	...		
白眼狼	特质	忘恩负义	人
铁公鸡	特质	小气	人

在表中，我们可以看到对于一个概念的隐喻知识，都是通过属性以及属性特征的取值来展现的。其中隐喻特征可定义为人们长期以来对某名词的认知特征中常用于隐喻修辞中描述本体特性和部分认知特征，这种特征通常用谓词来表示。如花朵的隐喻特征包括“美丽”、“娇弱”等。在这里，一般性名词的特征众多，而惯用语中的名词隐喻特征比较单一，这主要是由于惯用语的语义较为固定，不受语境制约引起的，所以惯用语的隐喻知识提取比较简单。这个表中的特征值，一般取的是X和Y所具备特征的交集，而不是全部。

隐喻知识中的属性本文定义为具体事物的宏观属性，即该事物所属群体的属性，而不是某一个体属性，由描述属性的名词构成。比如“小李就是一朵花”这里“小李”的属性应该是一般人都具有的属性，所以

其隐喻的语义映射应该是“花”和“人”之间关联。属性一般由名词性成分表示，多半都是特征所描述的对象。比如“美”主要是描述“外形”的。另外由于惯用语比喻的对象都比较固定，语义上可选的特征也比较固定，基本上都对应的是这个名词性词组的特点，所以在描述属性时，我们为了简单起见，都标为“特质”。比如“铁公鸡”的特征一般固定为“吝啬”，其属性归属按照非自然属性的归类，原则上是要归入到“心理态度”，但是这里我们都统一写为“特质”，“特质”接近于心理认知的特性

一般情况下，只有特征值有相似性，那么才会发生隐喻关联。比如“人”和“花”都有相似特征“美”，那么它们两者可以发生隐喻关联。另外，两个发生隐喻关联的概念还不能是从属关系，这在文献[11]中就已经论证。比如“白眼狼”从属于“狼”，即使和“狼”有共同的特征“凶狠”也不能发生隐喻关联，比如“这头白眼狼像一头狼”就说不通。

#### 4. 名词隐喻知识的获取

本文将名词的隐喻知识获取分为两种途径，一种是面向惯用语中的名词性词组，一种是面向动态文本中的普通名词。同时，隐喻知识的提取包括隐喻特征的获取和隐喻属性的提取两个方面。

##### 4.1 基于惯用语导入的隐喻知识获取

目前惯用语词典一般都会将比喻意义列出，所以从词典的释义中，我们就可以进行提取，并且建立语义映射关联。

因为惯用语长短不一，有时甚至是一句话，如“吃不了兜着走”，所以我们对惯用语进行了筛选，主要从《现代汉语惯用语规范词典》<sup>[16]</sup>和《新华惯用语词典》<sup>[17]</sup>，提取了1650个名词性惯用语词组，提取的多以三字词居多，也有四字词和二字词。通过分析释义，我们发现有些惯用语所隐含的隐喻对象并不只一个，比如“馋嘴猫”解释为“贪吃嘴馋的人或好色的男人”，那么这里虽然都是指人，但是隐喻的特征却不一样，一个是“贪吃嘴馋”、一个是“好色”。这些惯用语中表示一个隐喻特征的有1240个，表示两个隐喻特征的有360个，两个以上的50个。具体统计如下：

**表 2: 惯用语中的名词性词组的隐喻特征分布**

惯用语中的名词性词组	一个隐喻特征	两个隐喻特征	两个以上隐喻特征
三字词	1095	309	32
四字词	105	40	16
二字词	25	10	1
其他字	15	1	1

可见，大部分汉语惯用语只有一个隐喻特征，如“白眼狼”的隐喻特征只有“忘恩负义”，而作为狼的“凶残”、“敏捷”等特征不会被使用于隐喻中，如“他跑的飞快，像一只白眼狼”这样的句子是不会出现的。也就是说，惯用语中的隐喻知识它们只出现在特性隐喻中，通过喻体事物（惯用语）的最显著的特征体现出来，而且具有一定的排他性，因此对于惯用语中隐喻特征的提取，我们只需要根据惯用语的释义就可以确定具体名词的隐喻特征了。对于惯用语的释义格式，我们做了相关考察，名词性惯用语多以偏正和联合结构居多，一般在释义中多以“喻、比喻、形容、指、意谓”引出映射性概念 Y，而且 Y 多存在于“的”字结构，而隐喻的特征多为 Y 的修饰性成分，我们找了五种名词性结构进行对比，如下所示：

**表 3: 五种名词性结构惯用语释义列举**

惯用语中的名词性词组	语法结构	释义	例句
犬子 NN	偏正	喻无能的儿子	这些英雄们的后来各个了得,可谓虎父无~
鬼画符 NVN	偏正	形容潦草拙劣的字画	当中挂了一幅~的画
花花太岁 AN	偏正	指好色、凶恶的富家子弟或好色之人	谁敢与他口角,他就是个~
黑漆皮灯笼 ANAN	偏正	比喻不明就里或无知、糊涂的状态	大家~,哪里知道内中之事。
老婆孩子热炕头 NNAN	联合	意谓一家人其乐融融的幸福生活	他现在生活状态是~,甭提多幸福了。

我们可以看出,从释义中只要惯用语的结构是名词性的,那么其解释多半是“A的B”模式,这样我们基本上抽取“的”字结构并进行适当过滤就可以得到我们想要的知识了。识别算法如下:

Step 1 预处理:将惯用语词典进行词性标注等预处理。

Step 2 确定隐喻文段:利用关键词“喻、比喻、形容、指、意谓”等进行判断,提取这些关键词后面的文档。

Step 3 提取隐喻特征:利用关键词(“或”)和特定语法结构(主要是“的”字结构)从惯用语词典中获取惯用语释义作为其隐喻特征。一般情况下“或”确定特征数量(有“或”字的多为两个隐喻特征)，“的”确定隐喻特征值以及所映射对象(一般“的”前的修饰性成分为隐喻特征值，“的”后为隐喻对象)

至于惯用语的隐喻属性提取,由于大部分都是一个隐喻特征,较为单一,为了处理的方便我们统一标为“特质”。另外,其实名词的属性定义在具有特定含义的具体名词惯用语身上往往并不成立。因为属性定义为该事物所属群体的属性,然而惯用语由于对其内涵进行了限定,失去了大部分其所属群体的属性。例如“白眼狼”作为惯用语,失去了其所属群体狼的部分属性(如动作、外貌),只保持了狼的特征属性。所以惯用语“白眼狼”的属性唯一且固定为“特质”。

按照上述方法,我们把上面五个惯用语释义信息的隐喻知识进行相关的提取和分类,结果如下:

表 4: 五种名词性结构的惯用语隐喻知识提取结果

X 概念	属性	特征	Y 语义映射概念
犬子	特质	无能	儿子
鬼画符	特质	潦草拙劣	字画
花花太岁	特质	好色、凶恶	富家子弟;人
黑漆皮灯笼	特质	不明就里;无知、糊涂	状态
老婆孩子热炕头	特质	一家人其乐融融	生活

## 4.2 基于句法模式识别的隐喻知识获取

其实从大规模的语料中看,惯用语的使用范围有限,对于一般名词我们是采用句法模式的提取方法作为补充,靠一些常见的表示比喻的动词“像”、“如”等,或者依托一些语法结构特征来提取。而且一般性名词不同于惯用语,它一般有多种隐喻意义,所以必须要对隐喻特征和隐喻属性进行分别讨论。

### 4.2.1 隐喻特征提取

基于句法模式识别的隐喻特征提取主要是依靠直接出现在隐喻句中的特征词来获取喻体的隐喻特征,在提取方法上参考了李斌<sup>[6]</sup>的思路。我们通过考察确定了五种隐喻句的语法模式,如下所示:

表 5: 明确隐喻句的语法模式

序号	基本模式描述	例句
1	A vx B uyy C A 名词或代词, B 为具体名词, C 为形容词	她像花朵一样美丽。
2	A vx C ude1 B A 名词或代词, B 为具体名词, C 为形容词	她如同美丽的花朵。
3	A C ude3 vx B ( uyy ) A 名词或代词, B 为具体名词, C 为形容词	她美丽得像花朵(一样)。
4	vx B uyy C ude1 A A 名词或代词, B 为具体名词, C 为形容词	像花朵一样美丽的小女孩。
5	特殊模式	骨瘦如柴、骄阳似火

注: vx=“像、如同、似”等比喻动词, u=助词, yy=一样, de1=的, de3=得

基于句法模式识别的汉语隐喻特征获取算法如下:

Step 1 预处理: 获取隐喻句语料库, 将语料库中语料分割为以句子为单位。然后依次对各句进行切词和词性标注。

Step 2 句法模式识别: 根据各句中的词语位置和词性标注, 进行模式识别, 剔除无法匹配的句子。

Step 3 提取隐喻特征: 根据不同隐喻模式, 提取相应的喻源词语(具体名词)和其在句中的隐喻特征。

## 4.2.2 隐喻属性提取

在汉语一般名词的隐喻特征提取时, 往往会提取出多个特征, 为了隐喻意义获取的准确度, 我们还需要为每个隐喻特征加上对应属性值, 加以区分。

在获取属性值时, 对无法归纳至群体的名词, 可采用单个词语属性挖掘的方法获取隐喻属性; 对于可归纳至某一群体的名词, 通过群体属性挖掘方法, 可提高获取隐喻属性的准确度。之后, 可以通过搜索引擎海量的语料库, 提高不同候选属性和隐喻特征相关性计算的准确度。

### 1) 单个词语属性挖掘方法

属性值的获取是通过基于语料库的、模板驱动的知识挖掘方法。基本思想是, 具体名词的属性是指其宏观属性, 由描述属性的名词构成。如“小姑娘”的属性包括“身高”、“年龄”和“外貌”等等, 可以根据人们的用语习惯构建模板, 直接从语料库中提取这些属性信息。模板主要有以下两种:

表 6 属性值提取模板

序号	模板	例句
1	目标具体名词/n 的/ude1 属性/n	火车的速度是每小时 120 千米。
2	目标具体名词/n 属性/n	菊花颜色丰富, 有红、黄、白等。

在大规模语料库中, 通过模板进行挖掘后, 可设定阈值, 过滤出现频率低的词汇。之后可根据特征值, 再在语料库中进行文本挖掘, 统计特征值和属性候选集中各候选项出现在同一句子中的次数。出现频率最高的候选项作为具体名词在该特征下的属性被记录。

单个词语汉语隐喻属性获取算法如下:

Step 1 预处理: 获取隐喻句语料库, 将语料库中语料分割为以句子为单位。然后依次对各句进行切词和词性标注。

Step 2 模式识别: 在具体名词出现的句子中, 根据具体名词在句子中出现位置和词性标注, 进行模式识别, 选取符合模板的候选属性词。

Step 3 过滤: 选取阈值, 将出现频率低于阈值的候选词语删除, 只保留出现频率高于阈值的词语, 组成候选属性集。

Step 4: 确定属性值: 再次对语料库中的句子进行文本挖掘, 统计候选属性集中各项和特征值出现在同一语句中的频率。出现频率最高的候选项被选取为属性值。

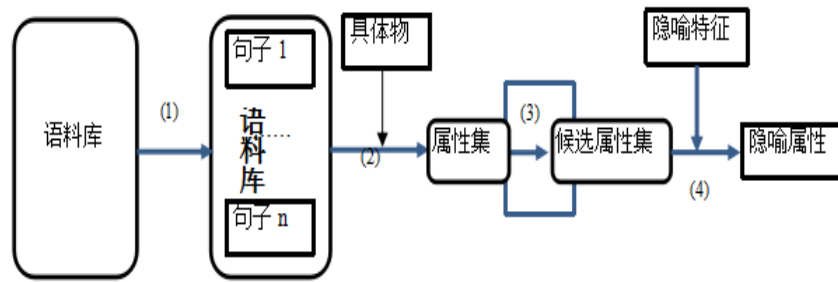


图 2 基于单个词语的汉语隐喻属性获取算法

## 2) 群体属性挖掘方法

在对单个具体名词的属性挖掘过程中, 我们发现属于同一群体的名词在隐喻修辞中使用到的属性非常相似。

群体的属性挖掘, 就是指从群体的个体实例的语言表述中, 通过大规模语料的自动提取, 获得其共同属性。相比于单个词语的属性获取, 群体的属性挖掘可以过滤单个词语属性中并不常使用于隐喻中的“噪音”, 提高属性提取的准确性。

基于群体属性的汉语隐喻属性获取算法如下:

Step 1 归类: 将名词划分至所属群体。

Step 2 单个词语属性模式识别: 在具体名词出现的句子中, 根据具体名词在句子中出现位置和词性标注, 进行模式识别, 选取符合模板的候选属性词。

Step 3 群体属性获取: 根据群体中每个实例的候选属性集, 设定阈值, 选取出现频率超过阈值的属性作为群体共同属性, 产生群体属性集。

Step 4 确定属性值: 再次对语料库中的句子进行文本挖掘, 统计群体属性集中各项和具体名词特征值出现在同一语句中的频率。出现频率最高的候选项被选取为属性值。

## 3) 基于搜索引擎的属性挖掘方法

我们通过实验发现, 实际群体属性集的规模不大, 于是我们想着借助搜索引擎, 通过具体模式搜索在海量数据的情境下帮助在群体属性中获取某一隐喻特征的隐喻属性。贾玉祥<sup>[10]</sup>和李斌<sup>[19]</sup>都尝试过通过搜索引擎的思路, 不过针对的对象有所差别。

基于搜索引擎的属性挖掘算法如下:

Step 1 -Step 3 同方法 2

Step 4 确定属性值: 将具体名词、特征属性和群体属性集中的每个选项都输入到搜索引擎中进行搜索, 过滤掉不符合的搜索条目后, 记录搜索条目数量, 选取搜索条目最多的候选项选取为属性值。

通过搜索引擎, 虽然只修改方法 2 的第四步, 但是效果改进还是较为巨大。专业的搜索引擎比我们的语料库相比, 存储了海量的文本数据, 并且具有更加优化的关键词搜索技术, 因而在获取相关性上效果更好。

## 5. 实验与分析

本次实验中的中文分词和词性标注相关工作, 选用了 NLPPIR 汉语分词系统。《新华惯用语词典》为惯用语主要释义来源, 自建语料库选取了人民日报的部分内容和文学作品集。(参考附录)

### 5.1 基于惯用语的隐喻知识提取

我们提取了 1650 个惯用语，在基于惯用语名词的隐喻特征直接导入时，直接将现有的惯用语的释义作为隐喻特征导入，隐喻属性标志统一为“特质”。基于惯用语导入的隐喻特性的部分生成结果表 7 所示：

表 7： 惯用语隐喻知识抽取

惯用语词组	释义	特征	属性	映射概念
安乐窝	比喻舒适温暖的生存环境	舒适、温暖	特质	生存环境
八败命	指一个人因八字不好，对其家人、亲	严重不幸	特质	命相
八辈子	比喻时间久远	久远	特质	时间
巴着眼	睁大着眼，形容期待的眼神	期待	特质	眼神
白鼻头	旧剧里奸险的人，现用以比喻叛徒、	奸险	特质	叛徒；特务
白眼狼	比喻忘恩负义的人	忘恩负义	特质	人
白骨精	比喻阴险毒辣的坏人	阴险毒辣	特质	坏人
百世师	指德高望众、学问渊博，可做百世师	德高望众、学问渊博	特质	人
挡箭牌	比喻掩饰的借口；也比喻庇护的场所	掩饰；庇护	特质	借口；场所
钉子户	长期违反规定办事难以处理的单位或	难以处理；不肯搬迁	特质	单位；个人；
顶梁柱	比喻起主要作用的骨干人物	起主要作用	特质	人
定心丸	指使人安定放心的事物。	安定、放心	特质	事物
拦路虎	过去喻指拦路打劫的土匪；现在喻指	打劫；阻碍前进	特质	土匪；事物；
老狐狸	比喻非常奸猾的人	奸猾	特质	人
里程碑	比喻历史中发生的具有标志性的大事	作为标志	特质	大事
马后炮	指事后喜欢采取行动的人，往往迟而	事后、无用	特质	人
马大哈	指做事粗心大意的人	粗心大意	特质	人
脓包货	比喻没用的人	没用	特质	人
破罐子	比喻失过身的妇女	失过身	特质	妇女
墙头草	比喻立场不稳、缺乏主见、左右摇摆	立场不稳、缺乏主见、左右摇摆	特质	人
书呆子	指只懂读书，不通世故的读书人	不通世故	特质	读书人
同路人	指志同道合的人，或与与时俱进的事物	志同道合；与时俱进	特质	人；事物
五分钟热度	形容不能持久的热情	不能持久	特质	热情
下三烂	比喻不光彩的事或者不争气的人	不光彩；不争气	特质	事；人
直肠子	比喻性格豪爽、直爽的人	豪爽、直爽	特质	人
孺子牛	比喻甘于奉献的人	甘于奉献	特质	人

我们对能正确从名词性惯用语的释义中成功提取隐喻知识的情况作了统计，其中准确率 92.6、召回率 89.4、F 值为 90.97。这个数值比一般目前单纯的在海量文本中提取隐喻知识要高得多，这主要是有惯用语自身特点决定的。

## 5.2 基于句法模式识别的一般性名词的隐喻知识提取

### 5.2.1 隐喻特征提取

我们从语料中提取了 3111 个符合隐喻模式的句子，根据句法模式识别方法把这些符合条件的句子中的隐喻特征提取了出来，部分生成结果如下所示：

表 8： 基于模式识别的一般性名词的隐喻特征提取

	具体名词	隐喻特性	映射概念	例句
1	云彩	洁白	羽毛	云彩像一朵朵洁白的羽毛，轻轻的飘浮在空中。
2	太阳	热	火炉	太阳像火炉似的热的人睁不开眼。
3	苹果	大、红	灯笼	树上的苹果像灯笼似的又大又红。
4	晚霞	深红	血	晚霞如同一滩血一样深红
5	嘴巴	大	香肠	他的嘴巴大的像两根香肠

在实验中，我们能够成功提取到隐喻特征并基本正确的句子有 3028 个，其中准确率为 97.3%，可以说基于模式识别的隐喻特征提取具有较高的准确度。

### 5.2.2 隐喻属性集的获取

#### (1) 单个词语的隐喻属性候选集生成



我们根据方法（1）从 3111 个模式隐喻句中提取了 2480 个单个词语的属性，通过人工校对，全部正确的只有 1382 个，准确率只有 55.7%，可以说准确率不是很高。但是前面三个属性的正确率却达到了 82.4%。部分生成结果举例如下：

表 9 部分名词属性候选集

	具体名词	属性候选集
1	汽车	速度 体积 排量 价格 品牌
2	母亲	形象 亲属 年纪 生活 白发

从表 9 中可以看出，单个词语的属性集中的候选项，部分可以作为准确描述具体名词隐喻特征的隐喻属性。例如具体名词“汽车”在隐喻特征为“飞驰”的时候，“速度”是其隐喻属性；具体名词“母亲”在隐喻特征为“慈祥”的时候，“形象”是其隐喻属性。

但是也应看到，候选属性集内有一些不会成为真正的隐喻属性。例如“白发”虽然常与“母亲”搭配出现，但是不会成为其隐喻属性。

### （2）群体的隐喻属性候选集生成

上面数据显示，对单个词语生成隐喻属性候选集时，容易产生跟词汇内在含义相关的“噪音”，如在文学作品中，“母亲”常与名词“白发”搭配出现，然而“白发”并不是“母亲”的宏观属性，所以也不会成为其隐喻属性。但是若将具体名词先划分为一些群体，在通过选取个体实例中共同出现的属性作为群体的属性集，就可以过滤这类情况，提高准确度。

本次实验将一般具体名词划分为人、动物、景物、物品（歌曲等无形物也算作物体）四个群体，每个群体包含的部分实例如下表所示：

表 10 群体与部分实例

1	人	小姑娘 仙女 母亲 园丁 士兵。。
2	动物	小鸟 小绵羊 蛇 百灵 狗。。。
3	景物	山水 大树 向日葵 海洋 云彩。。。
4	物品	喇叭 灯笼 桥 伞 凉棚。。。

每个群体根据实例属性集产生的群体属性集如下表所示：

表 11 群体属性集部分属性

	群体	属性
1	人	外貌 性格 地位 手 表情
2	动物	特性 外形 动作 声音 速度
3	景物	形状 颜色 分布 形成 高度
4	物品	颜色 外形 声音 温度 味道 材质

具体名词按照群体属性集获取的隐喻属性准确度有了较大的提高，人工核查发现，隐喻属性包含在群体属性集前 3 名的比例超过 50%。也就是说属性集的前 3 名基本上涵盖了具体名词所有一半属性。

### （3）基于搜索引擎的隐喻属性候选集生成

隐喻属性是从隐喻属性集中选取和隐喻特征共同出现的频率最高的候选项产生的。对于无法归类至群体的候选项，选取其单个词语候选项，否则选取其群体候选项，或基于搜索引擎进行属性挖掘。我们根据方法三，使用必应搜索引擎，结果如下表所示：

表 12 搜索引擎提取属性举例

	群体	实例	准确度
1	人	小姑娘 仙女 母亲 园丁 士兵	78.9%
2	动物	小鸟 小绵羊 蛇 百灵 狗	80.2%
3	景物	山水 大树 向日葵 海洋 云彩	68.1%
4	物品	喇叭 灯笼 桥 伞	64.3%
5	未分类	眼睛 手 歌曲	61.3%
6	总体		70.6%

根据结果可以看出，对群体分类的方式加上搜索引擎的搜索方式有效提高了隐喻属性获取的准确度，总准确度达到了 70.6%左右。这一结果表明，虽然三种方法结合可以有效提高属性值获取的准确度，但是要想具体名词隐喻知识中的属性全部找到还是有一定难度。不过总体上，惯用语的隐喻知识难度还是比一般名词要简单，而且准确度要远高于一般名词的隐喻知识提取。

## 6. 结论

隐喻从某种程度上讲虽然是一种语义映射关联，但是其中起关联作用的则是语义特征的相似性，这些特征其实就是隐喻知识。本文主要是将汉语具体名词间所映射的隐喻知识以隐喻特征和相关属性的形式加以表现，通过惯用语导入和特定句法模式导入两种方法引入隐喻知识，并从中提取隐喻特征和对应属性。针对惯用语，本文主要是基于语义词典的释义来提取特征，同时由于惯用语的隐喻特征具有唯一性，其对应的属性也相对唯一，所以可以说惯用语的隐喻知识是静态的。而一般名词由于语境不一样，他们放在不同的句子中可能隐喻的对象往往不同，因此这类一般名词的隐喻知识是动态的。针对一般句子中名词的动态隐喻知识，我们是通过一些规则性的隐喻构式来提取具体名词的隐喻特征，在获取属性值时，我们根据具体情况采用单个词语属性挖掘的方法、群体属性挖掘方法和搜索引擎挖掘方法来获取相关属性值。实验表明，在特征识别上取得较好效果，但在具体对应的属性值挖掘上总的准确率不高，说明获得具体名词对应的隐喻属性难度要远大于隐喻特征的获取。这就是为什么我们容易说“他像狼一样狠毒”，却很难列举出“狼的特质除了狠毒，还包括什么”。

## 参 考 文 献

- [1]. 郭振伟.我国隐喻语料库研究现状与展望[J].浙江师范大学学报（社会科学版），2013，3(38): 38~43
- [2]. 李行健.惯用语的研究和规范问题[J], 语言文字应用, 2002（1）：55-60
- [3]. 裴霜霜, 周榕. 借助Sketch Engine 和WordNet 进行隐喻的概念模型和语义映射分析----以目标域TIME 为例[J].外语教学, 2012,33(2):13-17
- [4]. 田嘉, 苏畅, 陈怡疆. 隐喻计算研究进展[J].软件学报, 2015, 26(1): 40~51
- [5]. 王治敏. 隐喻的计算研究与进展[J]. 中文信息学报, 2006,20(4): 16~24
- [6]. 李斌, 于丽丽, 石民, 曲维光. “像”的明喻计算[J]. 中文信息学报, 2008,22(6): 27~32
- [7]. 杨芸. 汉语隐喻识别与解释计算模型研究[D].博士学位论文.厦门.厦门大学.2008
- [8]. 杨芸, 周昌乐, 王雪梅, 戴帅湘. 基于机器理解的汉语隐喻分类研究初步[J].中文信息学报, 2004,18(4): 31~36
- [9]. 杨芸, 周昌乐, 李剑锋. 基于隐喻角色依存模式的汉语隐喻计算分类体系[J].语言文字应用, 2008, (3) :125~133

- [10]. 贾玉祥,俞士汶. 基于实例的隐喻理解与生成, 计算机科学, 2009,36(3):138-141
- [11]. 贾玉祥,俞士汶. 基于词典的名词性隐喻识别, 中文信息学报, 2011,25(2):99-104
- [12]. 王治敏. 名词隐喻相似度及推理识别研究[J]. 中文信息学报, 2008,22(3): 37~43
- [13]. 王治敏. 汉语名词隐喻的语义映射分析[J]. 语言教学与研究.2009.3: 89~96
- [14]. 王金锦, 杨芸, 周昌乐. 隐喻字面语义表示与生成[J]. 中文信息学报, 2009,23(3): 95~102
- [15]. 张威, 周昌乐. 汉语隐喻理解的逻辑描述初探[J]. 中文信息学报, 2004,18(5): 23~28
- [16]. 李行健. 现代汉语惯用语规范词典[M], 长春: 长春出版社, 2002
- [17]. 温端政. 新华惯用语词典[M], 北京: 商务印书馆, 2007
- [18]. 戴帅湘, 周昌乐, 黄孝喜, 杨芸, 王雪梅. 隐喻计算模型及其在隐喻分类上的应用[J]. 计算机科学, 2005, 32(5): 159~166
- [19]. 李斌, 陈家骏, 陈小荷. 基于互联网的汉语认知属性获取及分析[J]. 语言文字应用, 2012,(3): 134~143
- [20]. 李斌, 宋丽, 曲维光 等. 基于认知属性库的原型范畴研究[J]. 中文信息学报, 2016,30(6): 90~99
- [21]. 苏畅. 汉语名词性隐喻的计算方法研究[D]. 博士学位论文. 厦门. 厦门大学. 2008

### 附录：主要采样资源

《新华惯用语词典》	《现代汉语惯用语规范词典》
《中国惯用语快读》	
《老舍文集》	《王小波作品全集》
《朱自清文集》	《沈从文文集》
《鲁迅文集》	《冰心文集》
《围城》	《文化苦旅》
《张小娴文集》	《世界动物故事 100 篇》
《荒废集》	《人民日报》1998 年 1 月语料库