

# 隐喻计算模型及其在隐喻分类上的应用<sup>\*</sup>

戴帅湘<sup>1</sup> 周昌乐<sup>1</sup> 黄孝喜<sup>2</sup> 杨芸<sup>1</sup> 王雪梅<sup>3</sup>

(厦门大学计算机与信息工程学院 厦门361005)<sup>1</sup>

(浙江大学计算机科学系 杭州310027)<sup>2</sup> (厦门大学人文学院 厦门361005)<sup>3</sup>

**摘要** 隐喻理解是语篇理解中的难点,也是认知科学研究中的热点。本文针对汉语中普遍存在隐喻的现象,从计算的角度,首先提出了一套汉语句子形式化方法,并在此基础上针对隐喻的特点构建出隐喻语义网络计算模型。为了阐述隐喻网络模型的应用,文章后半部分提出了一套基于隐喻网络模型的隐喻分类体系,并对其合理性以及实用性进行了分析,为后续的隐喻自动识别以及隐喻理解奠定了基础。

**关键词** 隐喻,隐喻计算,隐喻分类

## Computational Model of Metaphor and its Application in Metaphorical Classification

DAI Shuai-Xiang<sup>1</sup> ZHOU Chang-Le<sup>1</sup> HUANG Xiao-Xi<sup>2</sup> YANG Yun<sup>1</sup> WANG Xue-Mei<sup>3</sup>

(School of Computer Science and Information Engineer, Xiamen University, Xiamen 361005)<sup>1</sup>

(Department of Computer Science, Zhejiang University, Zhejiang 310027)<sup>2</sup>

(Department of Chinese Language and Literature, Xiamen University, Xiamen 361005)<sup>3</sup>

**Abstract** Metaphor is a difficult problem in natural language understanding, while it is also a hot topic in the cognitive scientific research. Aiming at the generally existent phenomenon of metaphor in Chinese, in term of computing [from the aspect of computing], this paper proposes a set of methods of formalizing Chinese sentence at first, and structures the computational model of metaphorical semantic network (MSN) according to the characteristics of metaphor. Then, we apply the MSN to metaphorical classification, present a systematic method of metaphorical classification, this lays a foundation for the later research in auto metaphorical classification and metaphorical understanding.

**Keywords** Metaphor, Metaphorical computation, Metaphorical classification, Metaphorical semantic network

## 1 引言

隐喻的一个重要语义特征是本体(Tenor)与喻体(Vehicle)之间具有明显的性质差异或者语境之间存在尖锐冲突,完全相同的事物之间显然不能构成隐喻,因而字面意义的差异性为隐喻成立的先决条件。与此同时,本喻体之间必须存在相似性,所谓“同从异出”,相似性是隐喻赖以成立的基本要素。隐喻意义的理解实际上就是将源领域(喻体)的经验知识映射到目标领域(本体),从而达到重新认识目标领域的特征的目的,而映射函数就是基于本体和喻体之间的某种相似性或相关性。此外,这里的相似性又有别于一般的词汇语义相似性,它不仅仅涉及词汇语义,更重要的是隐喻句中本体和喻体所指称的是事物或事件之间属性的相似。

目前,国外有关隐喻的计算研究自20世纪90年代以来已经取得了初步成就<sup>[1~2]</sup>。首先,自20世纪70年代以来,出现了一些隐喻语言理解的各种初步模型,如基于类比语义的方法,包括优选语义方法<sup>[3]</sup>、词汇语义方法<sup>[4]</sup>、类比推理方法<sup>[5,6]</sup>;基于连接主义的方法<sup>[7~9]</sup>;基于向量空间的方法<sup>[10]</sup>,通过潜在语义分析(LSA, Latent Semantic Analysis)来计算词语之间的相关程度。由于缺乏对隐喻本质的分析,这些方法从根本上来

说都是不足取的。另外在隐喻语言理解模型研究中,比较突出的是J. H. Martin提出的MIDAS系统(Metaphor Interpretation, Denotation, and Acquisition System),采用基于经典人工智能知识表示和符号处理技术,给出了一种英语隐喻语言理解方法,并具体应用于UNIX操作系统界面语言教学的对话系统中<sup>[1]</sup>。

其次,隐喻计算研究的另外一方面主要是关于隐喻语言的形式化表示和隐喻类比逻辑系统的构建。典型的有关于隐喻的语法、语义的形式化描述<sup>[11,12]</sup>,隐喻类比逻辑设计<sup>[13~15]</sup>。Steinhart比较系统地建立了一个隐喻类比逻辑系统<sup>[16]</sup>,他采用扩展的可能世界语义学来处理隐喻的“字面意义”和“表述意义”,进而提出了所谓隐喻的结构理论(STM, Structural Theory of Metaphor),针对结构性比较明显的隐喻,建立了一套相对完整的类比逻辑释义方法。所有这些方法都存在这样的缺点,即大都从认知角度去关注隐喻,在隐喻“表述意义”的描述和转释获取中,注重客观条件的类比匹配,而忽略了主体知识(主观性)的作用,因而没有能考虑引入认知状态<sup>[17]</sup>来更好地描述和推导隐喻的“表述意义”。

目前国内从计算的角度对汉语隐喻进行研究尚属空白,鉴于此本文针对汉语隐喻的特征构建了一套适合于处理汉语

<sup>\*</sup> 本文得到国家自然科学基金资助(项目编号:60373080);福建省自然科学基金资助(项目编号:A0210005)。戴帅湘 硕士研究生,主要研究领域为人工智能算法、计算语言学;周昌乐 教授,博士生导师,主要从事人工智能领域的研究工作,研究方向为计算语言学、理论脑科学和认知逻辑学;黄孝喜 博士,主要研究领域为认知逻辑学、计算语言学。

隐喻的计算化模型。

## 2 隐喻计算模型

### 2.1 形式化方法

自然语言在处理过程中,首先需要对其语言进行形式化描述。自然语言的形式化包括两个方面的内容:结构形式化(句法表示)和内容的形式化(语义表示)<sup>[18]</sup>。在探讨隐喻的计算化方法前需要形式化地描述隐喻句,为了建模的方便,下文中对句子的处理将分别从语言学、对象描述语言(本体描述语言)以及图论的角度进行描述,将句子的成分分别描述为对象描述语言中的对象、方法以及属性,并对应为图论中的节点概念,不特别说明,文章中的说明性文字将从对象角度进行描述,而定义、定理以及证明等均采用图论中的符号,特殊地方将添加对应语言学上的解释。

一个有意义的句子是为了阐述某个中心对象的,而对对象的描述表现为对其属性和方法的说明,那么句子成分可以形式化为一个三层结构图。

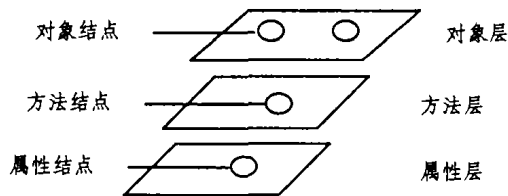


图1 句子分层结构图

图中节点在语言学、对象描述语言以及图论中的对应关系如下。

语言学	对象描述语言	图论
主语、宾语	对象	对象节点
定语	属性	属性节点
[状语]谓语[宾语 补语]	方法	方法节点

图1中对象层对应句子中描述的对象或者说概念的集合,即常识知识中的概念集,对象节点对应句子中描述的中心词(句子基本成分),包括主语和宾语。方法层是对象节点方法的集合,节点表现为句子中与主语搭配的谓语。属性层是对象属性节点的集合,表现为修饰对象的句子成分,是句子的附加成分,如定语,状语等。对于不带从句的短语、简单句子拆分成节点均是单个的对象、属性或者方法,如“踏入苍穹”可拆分为如下节点。

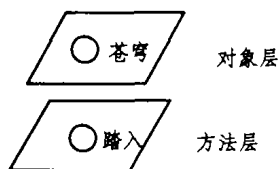
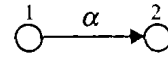


图2

由于汉语句子可分为单句和复句,而复句又是由单句组成的,那么复句节点也可以看作是简单节点的复合,关于节点的运算将在后文中具体介绍,本节的模型均是针对简单节点(单句和短语对应的节点)而言的。不考虑复句的情况,汉语中的短语(不含从句)、单句均可以将其拆分成上述图中的节点,并以一定的方式分布。节点间的连接决定于节点之间的语义关系,对应句子中成分之间的关系,语义关系用连接弧来表

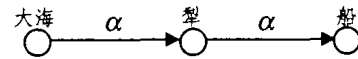
示。为了描述各个节点之间的语义关系,定义如下语义有向连接弧:

1. 调用弧,标记为  $\alpha$

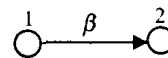


表示节点1被节点2直接使用或者说节点2直接使用节点1。对象描述语言中指节点1被指定为节点2的属性或者方法。对应语言学上表示句子中成分1被成分2直接调用,或者说成分2直接调用了成分1。

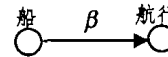
如:船犁大海



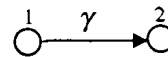
2. 常识弧,标记为  $\beta$



表示节点1可以衍生或者推导出节点2。对象描述语言中表示节点2是节点1的常识知识,如节点2是节点1固有的属性,方法等。语言学上表示成分1和成分2的搭配是合乎语法的正常搭配。如:

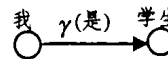


3. 同性弧,标记为  $\gamma$

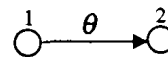


表示节点1和节点2等价或者同类属的关系。对象描述语言中表示节点1和节点2具有相似性质或者可比关系。语言学上表示成分1和成分2具有表达的意义是一样的或者是同类的可比概念。引入同性弧是为了表达语言学上具有 A is B 或者 A like B 这样结构的句子,其中 A、B 是同类可比对象,箭头方向指向 B。为了明确指出两个可比对象之间使用的谓词,也常把谓词写到弧上。

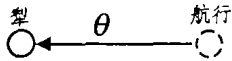
如:“我是学生”



4. 隐喻弧,标记为  $\theta$



表示节点1和节点2之间隐喻映射函数成立,箭头方向指向喻体概念。对象描述语言中表示节点1和节点2之间具有隐喻关系。语言学上表示成分1和成分2中分别为本体概念和喻体概念,一个句子是隐喻句当且仅当句子结构图中至少有一个这样的隐喻弧,这里的本体概念和喻体概念是指可以相互比较的一对同类词语,它们对理解句子的隐喻起到关键作用,但不一定要是名词,也可以是形容词或者动词或者词组等。如“船犁大海中”的本体概念是“航行”,喻体概念是“犁”,它们之间的隐喻弧如下:



在隐喻句中本体概念往往不出现,为了得到隐喻弧,需要在图中添加虚节点,如上图“航行”对应的是虚节点。一个隐喻弧总是对应一个本体概念和一个喻体概念,如果把其写成二元组的形式:(本体概念,喻体概念),则又可以称之为隐喻变量。

对于含有“像”、“如”、“是”等喻词作谓语,在修辞学上称为明喻的或者暗喻的句子,谓语在图中不以节点方式出现,而是把其标明在弧上。如:女人如花

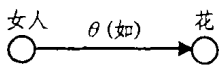
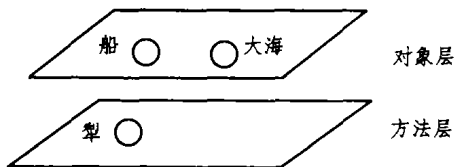


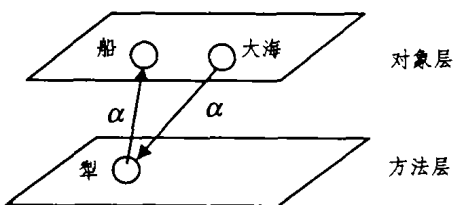
图3

通过把句子拆分为成分节点,并用节点间的弧来表示句子各成分之间的语义关系,可以构造一个句子的语义结构网络或者说语义结构图,为了讨论方便,下文图中和网络统称为网络。以“船犁大海”为例具体说明上述语义结构网络的构造过程如下:

第一步:拆分节点并分层排布



第二步:各个节点之间以相应的语义弧相连接



第三步:如果是隐喻句,为了体现隐喻中本体概念和喻体概念之间的映射关系,在本体概念未出现的情况下,需要在隐喻网络中加入虚节点,标注隐喻弧以及和虚节点相连的其他语义弧。

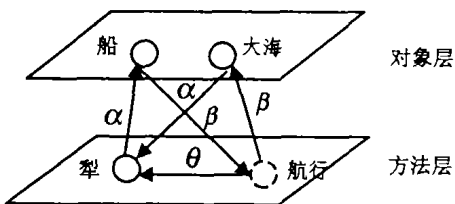


图4

“航行”在句子中未出现,为了清晰地表示隐喻句中本体概念和喻体概念之间的映射关系,引入“航行”这一方法节点作为虚节点。

## 2.2 隐喻网络模型

一个句子通过拆分为节点,并通过语义弧对各个节点进行连接,从而形成一个有向语义网络,这个网络称为结构语义网络(简称为语义网络, Semantic Network)。

定义2.1.1 语义网络(Semantic Network 简称为SN)

是一个三元组  $\langle V, \otimes, A \rangle$ , 记为  $SN = \langle V, \otimes, A \rangle$ , 其中:

1. 语义节点集合  $V, V \neq \emptyset, V = V(SN) = \{v_1^t, v_2^t, \dots, v_n^t\}$ ,  $v_i^t$  表示网络中的节点或者顶点,下标表示节点在句子中的先后次序,上标  $L = \{object, property, method\}$ , 表示节点所属的层是对象层、属性层还是方法层。

2. 语义弧集合  $A, A \neq \emptyset, A = A(SN) = \{a_1, a_2, \dots, a_n\}$ ,  $a_i$  为三元组  $(v_i^t, \otimes, v_j^t)$ , 或写为  $v_i^t \otimes v_j^t$ , 其中  $i \neq j$ 。

3. 语义算子集合  $\otimes, \otimes = \{\alpha, \beta, \gamma, \theta\}$ ,  $\alpha, \beta, \gamma, \theta$  分别对应上面四种语义弧中的语义运算。

语义网络具有如下性质:

性质1 一个句子具有唯一的语义网络相对应(唯一性)。

由语义网络的定义以及构造过程可知句意明确的句子只有一个语义网络,如果具有多个语义网络则说明句子一定有歧义,这里只讨论句意明确的句子,即假设所有句子都已经消歧。

性质2 由于句子的结构是可以嵌套的,即句子的各个成分仍然可以是句子,那么语义网络也具有可嵌套性(可嵌套性)。

如果语义网络表示的句子是隐喻句子,并且网络中添加了隐喻弧,这种语义网络称为隐喻结构语义网络,简称为隐喻网络(Metaphorical Semantic Network),即如果语义网络中存在隐喻弧  $v_i^t \theta v_j^t$ , 则称为隐喻网络,严格的数学定义如下:

定义2.1.2 隐喻网络(Metaphorical Semantic Network 简称为MSN) 是一个三元组  $\langle V, \otimes, A \rangle$ , 记为  $MSN = \langle V, \otimes, A \rangle$ , 其中:

1. 语义节点集合  $V, V \neq \emptyset, V = V(MSN) = \{v_1^t, v_2^t, \dots, v_n^t\}$ ,  $v_i^t$  表示网络中的节点或者顶点,下标表示节点在句子中的先后次序,上标  $L = \{object, project, method\}$ , 表示节点所属的层是对象层、属性层还是方法层。

2. 语义弧集合  $A, A \neq \emptyset, A = A(MSN) = \{a_1, a_2, \dots, a_n\}$ ,  $a_i$  为三元组  $(v_i^t, \otimes, v_j^t)$ , 或写为  $v_i^t \otimes v_j^t$ , 集合  $A$  中一定存在一条隐喻弧  $v_i^t \theta v_j^t$ , 其中  $i \neq j$ 。

3. 语义算子集合  $\otimes, \otimes = \{\alpha, \beta, \gamma, \theta, +\}$ ,  $\alpha, \beta, \gamma, \theta$  分别对应上面四种语义弧中的语义运算,在网络图中表现为语义弧。节点间的+运算表示节点的合并或称复合关系,在图中不表现为弧,只在节点复合时使用。

如果隐喻网络只由一对隐喻变量,以及它们之间的隐喻弧构成,则称为平凡隐喻网络,如图3所示。

隐喻网络除了具有语义网络的唯一性以及可嵌套性以外,还具有自己独特的性质:

性质1  $\forall v_i^t \alpha v_j^t, v_i^t \beta v_j^t, \exists v_k^t = v_i^t + v_j^t$ 。

即如果隐喻网络中节点  $v_i^t$  和  $v_j^t$  之间  $\alpha, \beta$  算子都成立,则  $v_i^t$  和  $v_j^t$  之间+算子成立,也就是说可以把此两个节点合并为一个节点而不影响隐喻网络的其他性质。

$\alpha$  算子表示的是两个句子成分间的搭配关系,而  $\beta$  算子是表示两个成分间的搭配是否正确的判断,如果两个成分之间  $\alpha, \beta$  都成立则说明它们是句子中搭配正常的部分,根据隐喻网络的定义,隐喻的产生是由于在相邻两个节点之间的搭配异常产生的,既然两个相邻的节点搭配正常,则可以断定此隐喻句子中的隐喻产生于别的弧线,则删除此两个成分对应的节点间的弧线,把此两个节点合并为一个更大的语义节点不会影响隐喻网络的其他性质。

性质2 经过节点合并后的隐喻网络是一个简单网络(没有环和多重弧)。

由语义网络的定义可知,语义网络是允许出现多重弧的,但对于隐喻网络而言,由于合并了具有多重弧的节点,因此隐喻网络节点之间不存在多重弧,是一个简单网络。

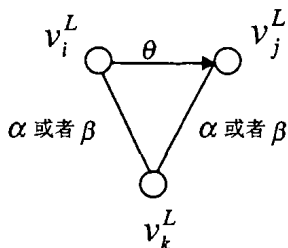
**性质 3** 隐喻弧只能出现在一个节点层上。

根据隐喻弧的定义,隐喻弧对应的隐喻矢量必须由两个同类对象构成,不同类的对象之间不具有语义可比性,从而不可能形成一对隐喻矢量,也就是说隐喻弧两端的节点必须属于同一个层。

**性质 4**  $\forall MSN = \langle V, \otimes, A \rangle, \exists v_i^L \otimes \dots v_i^L \theta v_j^L \dots \otimes v_k^L$ , 其中为非平凡网络。

上述性质说明除平凡隐喻网络外,隐喻网络的各个节点之间一定存在一个含隐喻弧的封闭语义回路,简称为隐喻回路。

由隐喻网络的定义可知隐喻网络中一定存在一个隐喻矢量,设其为 $(v_i^L, v_j^L)$ ,于是对非平凡隐喻网络而言,一定存在除 $v_i^L, v_j^L$ 外的第三个节点,设为 $v_k^L$ ,另外由性质3可知 $v_i^L$ 和 $v_j^L$ 属于同一层,那么 $v_k^L$ 与 $v_i^L, v_j^L$ 肯定不在同一层,否则网络对应的就不是有效的句子,于是对应句子间的语义关系 $v_i^L, v_j^L$ 之间, $v_j^L, v_k^L$ 之间一定 $\alpha$ 或者 $\beta$ 算子成立,从而使得 $v_i^L, v_j^L, v_k^L$ 之间具有如下回路:



### 2.3 隐喻网络的矩阵表示

隐喻网络也可以用矩阵来表示,这样可以用矩阵运算来研究有限隐喻网络的结构特征以及性质。

**定义 2.2.1** 设隐喻网络  $MSN = \langle V, \otimes, A \rangle$ , 则  $MSN$  的邻接矩阵  $T = (t_{ij})_{n \times n}$ , 其中  $t_{ij}$  是  $v_i^L$  和  $v_j^L$  之间的语义算子。如图4, 设其节点按顺序依次为{“船”, “犁”, “大海”, “航行”}, 则其邻接矩阵  $T_{4 \times 4}$ ,

$$T_{4 \times 4} = \begin{pmatrix} 0 & \alpha & 0 & \beta \\ 0 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & \theta & \beta & 0 \end{pmatrix}$$

## 3 隐喻分类体系

从理解方式上讲,隐喻有许多表现形式,如明喻、暗喻、借喻、事喻、物喻等等。另外,除了隐喻句以外,汉语语句还有可分为直陈句、象征句等等,这些都是从修辞学角度对句子进行分类的,对于计算化研究来说意义不大。此外,西方从认知的角度对隐喻进行了深入的研究,把隐喻主要归为传导隐喻、概念隐喻、基本隐喻、诗性隐喻以及根隐喻等<sup>[10]</sup>。西方主要把隐喻看作是一种思维方式,因此这些研究偏重于从心理学出发,构建隐喻思维的类比计算模型,而忽略了语言理解中隐喻句的处理。在我国,有关隐喻的计算化研究尚属空白,这对于深化中文信息处理技术的发展极为不利。从这个意义上讲,作为汉语隐喻理解的准备,首先开展汉语隐喻的分类和识别研究工作,不但十分必要,而且有着重要的学术意义和理论建设性意义,可以填补我国在该领域方面的空白,并有望做出开创性的成绩。

### 3.1 隐喻的分类方法

根据隐喻网络的定义以及性质,可以从隐喻结构及其语义关系对隐喻进行分类,为了不至于引起歧义,同时不失一般性,分类前给短语、单句以及复句以明确的结构限定:

语言学上的短语,用对象描述语言表示为:

对象+属性、属性+对象

的形式;

语言学上的单句,用对象描述语言表示为:

对象+方法、对象+方法+对象

的形式。对于根据上下文显然可知而省略了某些句子成分的省略句。如:“这家伙溜得狗一样快”这句话在文章或对话中往往省略主语,变成“溜得狗一样快”。类似这种形式,将在省略句的前面加上任一指代人的代词补全主语,以对应对象描述语言中的“对象+方法”的形式。另外诸如“猫着腰”的动词词组亦当作省略主语的句子处理。

语言学上的复句,是由多个单句组成,可以看作单句的嵌套行为。

下面首先从语言学以及对象描述的角度分别对短语、单句以及复句中的隐喻进行分析,获取隐喻网络的简单类别以及复合类别,然后从图论角度进行描述,以构建隐喻形式化分类体系。

1. 短语隐喻 短语隐喻的理解主要在于短语中心词与属性之间的关系,主要表现为两种形式:一是同名属性值的移用,即中心对象属性值的改变或者增加,如“东方威尼斯”中威尼斯的位置属性的取值改为“东方的”;另外一种是属于的移用,即中心对象的属性的增加,如“民族的肤色”中给“民族”这一个对象增加了一个“肤色”的属性。这两种形式对应的隐喻网络具有显著的区别,即隐喻弧所在的层不一样,一个在对象层,一个在属性层。

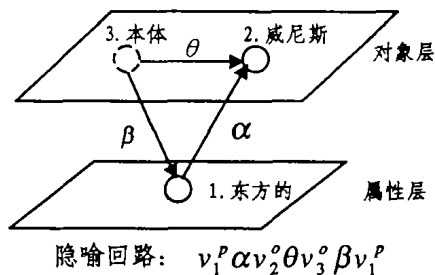


图5 东方的威尼斯

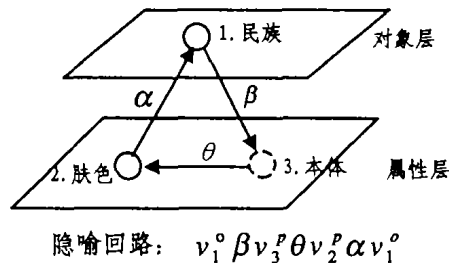


图6 民族的肤色

2. 单句隐喻 单句隐喻主要有类映射和方法映射两种。类映射是指隐喻中本体对象和喻体对象的直接作用,而方法映射是指中心对象与其方法之间的作用,这两种不同的作用导致隐喻意义产生的机制也不同。

2.1 类映射的隐喻一般具有 A-is-B or A-like-B 的形

式,分别如图7和8,其中图8是平凡隐喻网络,只含有一条隐喻弧,不具有隐喻回路。

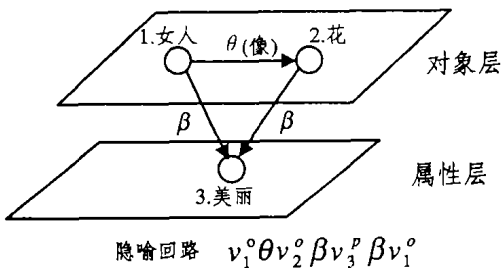


图7 女人如花一样美丽

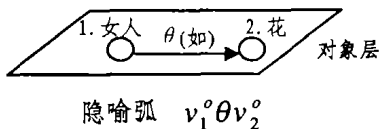


图8 女人如花

2.2 方法映射产生的隐喻表现为中心词与方法的搭配异常,方法又因是否带宾语而产生显著不同的两种隐喻结构,分别如图9和10。图9是主谓,谓宾同时搭配异常,图10只是主谓搭配异常。

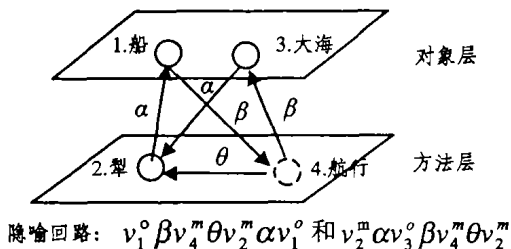


图9 船犁大海

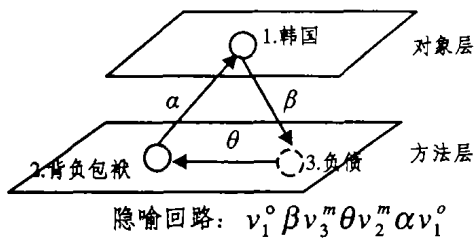


图10 韩国背负几十亿美元的包袱

图9中有两条隐喻回路  $v_1^o \beta v_4^m \theta v_2^m \alpha v_1^o$  和  $v_2^m \alpha v_3^o \beta v_4^m \theta v_2^m$ 。注意到图10中实际上应该还有一个对应“包袱”的节点,但是由于“背负”和“包袱”之间  $\alpha, \beta$  算子都成立,根据隐喻网络的节点合并原则进行了合并,从而图10中只有一条隐喻回路  $v_1^o \beta v_3^m \theta v_2^m \alpha v_1^o$ 。

3. 复合隐喻 复合隐喻主要是指由短语隐喻以及单句隐喻嵌套复合而成的隐喻,如图11。但是复合隐喻还包含一种隐喻不属于隐喻嵌套,如图12,它们的节点是复合节点(如果句子的某个成分又是句子,则称此成分对应的节点为复合节点),但映射关系却是类似类映射的简单关系,这里称之为关系映射。

3.1 复合映射:

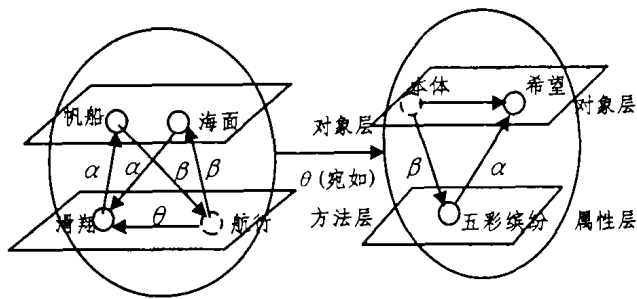


图11 一叶叶帆船在海面滑翔,宛如撑着五彩缤纷的希望

3.2 关系映射:

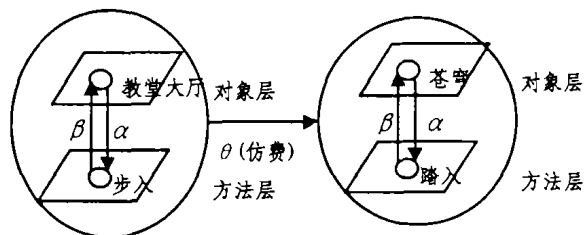
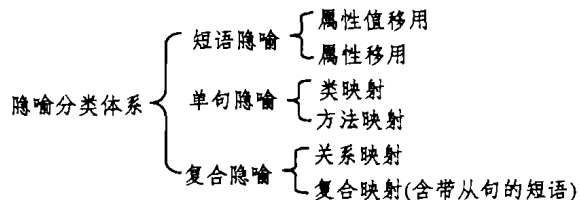


图12 步入教堂大厅,仿佛一步踏入幽深飘渺的苍穹

关系映射和类映射的概念相差无几,只是隐喻网络节点是复合节点(多个节点通过“+”运算复合而成的节点)而已,即节点对应的句子成分是由一个和多个句子构成的。关系映射隐喻可以解决“亡隐喻”的归属问题,即认为某些成语典故是关系之间的映射,即本体概念和喻体概念均可以是句子或者篇章。

3.2 分类体系及其合理性分析

综上所述,隐喻句的分类体系可表示为如下分类图:



上述分类体系是构建在隐喻语义网络模型的基础上的,从隐喻网络模型的构造以及节点间的作用关系可知节点类型以及节点间的关系均是有限的,结合隐喻本身的认知特征对隐喻网络进行分类从而构建隐喻分类体系,这种分类方法不但具有较为严格的数学依据,同时易于程序实现,为进一步研究隐喻的机器自动识别,甚至隐喻的机器理解奠定了形式化基础。如果能够找到一种可行的方法从语义网络构造出隐喻网络的话,那么就可以在普通语言中提取隐喻含义,从而在一定层次上研究隐喻的机器识别以及隐喻的机器理解,从而推动自然语言理解的研究进程,而隐喻分类体系的建立正是要简化从语义网络到隐喻网络的构造方法,其分类的前提和依据也正是基于此原则的。

结束语 隐喻网络模型是针对汉语隐喻计算化所提出的一个形式化语义网络,与语义 Web 纯粹的知识表述结构不同,本文中的语义网络是从句子的结构以及理解方式出发构建的,它的建立对计算化研究隐喻理解以及汉语的篇章理解均有好处,同时也能够方便地引入语义 Web 中的世界知识,以构建更好的机器语言理解模型。对于隐喻的自动分类识别与释义推导正在研究当中,并已经取得了初步成绩,我们将另文发表。

用卡使用或者电信服务;预测市场动向,在市场分析中分析客户的流失等异常行为;或者在医疗分析中发现对多种治疗方

式的不寻常的反应等等。通过对这些数据进行研究,发现正常的行为和模式,实现异常数据挖掘功能。

表1 基于属性的异常点检测算法与文[7,8]中基于距离的异常点检测算法比较

实验	算法	参数设定	检测结果	消耗时间
实验一	基于距离	distance=0.25,p=30	检测出异常数据54个	42分钟
		distance=0.38,p=18	检测出异常数据2个	26分钟
		distance=0.35,p=20	检测出异常数据37个	38分钟
	基于属性	p=10	检测出异常数据8个,其中一个检测错误	2分钟
实验二	基于距离	distance=0.25,p=20	检测出异常数据6个	20分钟
		distance=0.25,p=30	检测出异常数据7个	38分钟
		distance=0.30,p=20	检测出异常数据3个	14分钟
	基于属性	p=10	检测出异常数据6个	1分钟

本文提出的基于属性的异常检测算法实现了在一个未知的数据集上进行异常点数据检测的功能。它通过分析数据对象的各个属性,对数据进行异常检测,然后利用异常标记数组对数据集进行数据分离,并进行输出。实验表明,基于属性的异常检测算法比现有的一些算法在执行时间、检测精度等方面具有明显的优势,目前我们在算法的准确率、易用性等方面正在进行进一步的研究。

### 参考文献

- 李炎,李皓,等. 异常检测算法分析. 计算机工程,2002,028(006): 5~6,32
- 李之棠,刘颖. 入侵检测中的模糊数据挖掘技术. 计算机工程与科学,2002,024(002): 18~21
- Han Jiawei, Kamber M. Data Mining Concept and Technique. 北京:高等教育出版社,2001
- 魏葵,宫学庆,等. 高维空间中的离群点发现. 软件学报,2002,013(002): 280~290
- Witten Ian H, Frank Eibe. Data mining: practical machine learning tools and techniques with Java implementations. Morgan Kaufmann, 1999
- Collections of datasets. <http://www.cs.waikato.ac.nz/ml/we-ka/>
- Knorr E,Ng R. A unified notion of outliers: Properties and computation. In: Proc. 1997 Int. Conf. Knowledge Discovery and Data Mining (KDD'97), Newport Beach, CA, Aug. 1997. 219~222
- Knorr E,Ng R. Algorithms for mining distance-based outliers in large datasets. In: Proc. 1998 Int. Conf. Very Large Data Bases (VLDB'98), New York, Aug. 1998. 392~403
- Petrovskiy M I. Outlier Detection Algorithms in Data Mining Systems. Programming and Computing Software New York, 2003, 029(004): 228~237
- Kollios G, Gunopulos D, Koudas N, Berchtold S. Efficient Biased Sampling for Approximate Clustering and Outlier Detection in Large Data Sets. IEEE Transactions on Knowledge and Data Engineering, 2003, 015(5): 1170~1187
- Arning A, Agrawal R, Raghavan P. A linear method for deviation detection in large database. In: Proc. 1996. Int. Conf. Data Mining and Knowledge Discovery (KDD'96), Portland, OR, Aug. 1996. 164~169
- 李翠平,李盛恩,王珊,杜小勇. 一种基于约束的多维数据异常点挖掘方法. 软件学报, 2003,014(009): 1571~1577
- 黄守坤. 异常数据挖掘及在经济欺诈发现中的应用. 统计与决策, 2003(004): 32~33
- 黄莹. 基于数据挖掘的异常检测模型. 电子工程师, 2003, 029(006): 11~13
- 孔学峰. 数据挖掘及其在信用卡风险控制中的应用. 中国金融电脑, 2003(010): 21~22, 33
- 宋世杰,胡华平,胡笑蕾,金士尧. 数据挖掘技术在网络型异常入侵检测系统中的应用. 计算机应用, 2003, 023(012): 20~23

(上接第163页)

### 参考文献

- Martin J H. A Computational Model of Metaphor Interpretation. Boston, Academic Press, 1990
- Nehaniv C L. Computation for Metaphors, Analogy, and Agent. Springer, 1999
- Wilks Y. Making preferences more active. Artificial Intelligence, 1978, 11
- Fass D. Collative Semantics: A Semantics for Natural Language. [PhD thesis]. New Mexico State University, CRL Report No. MCCS-88-118, 1988
- Carbonell J G, Minton S. Metaphor and Commonsense reasoning. In Hobbs & Moore, Formal Theories of the Commonsense World, Norwood, NJ: Ablex, 1985. 405~426
- Gentner D. Structure-mapping: A theoretical framework for analogy. Cognitive Science, 1983, 7: 155~170
- Holyoak K, Thagard P. Analogical mapping by constraint satisfaction. Cognitive Science, 1989, 13: 295~355
- Indurkha B. Metaphor and Cognition: Studies in Cognitive Systems. Kluwer Academic Publishers, Dordrecht: The Netherlands, 1992
- Veale, Tony. Metaphor, Memory and Meaning: Symbolic and Connectionist Issues in Metaphor Interpretation. [PhD. Dissertation]. 1995
- Kintsch W, Bowles A. Metaphor comprehension: What makes a metaphor difficult to understand? Metaphor and Symbol, 2002, 17: 249~262
- Mori T, Nakagawa H. A formalization of metaphor understanding in situation semantics. In: Barwise, J. et al. eds. Situation Theory and its Applications, vol. 2, CSLI Lecture Notes 26, Stanford, CA: CSLI Publications. 1991
- van Dijk T A. Formal semantics of metaphorical discourse. Poetics 1975, 4: 173~198
- Hall R P. Computational approaches to analogical reasoning: A comparative analysis. Artificial Intelligence, 1989, 39: 39~120
- Indurkha B. Approximate semantic transference: A computational theory of metaphors and analogies. Cognitive Science, 1987, 11: 445~480
- Isabel D'Hanis. A logical approach to the analysis of metaphors. In: Magnani, L, ed. Logical and computational aspects of Model-based Reasoning, Kluwer Academic, Dordrecht, 2002. 21~37
- Steinhart E C. The Logic of Metaphor: Analogous Parts of Possible Worlds, Kluwer Academic Publishers, 2001
- 周昌乐. 认知逻辑导论. 清华大学出版社, 2001
- 周昌乐. 心脑计算. 北京: 清华大学出版社, 2002
- 胡壮麟. 认知隐喻学. 北京: 北京大学出版社, 2004