

基于相关性约束的隐喻理解方法^{*}

苏畅¹, 王晓梅¹, 黄舒曼¹, 陈怡疆²

¹(厦门大学 智能科学与技术系, 福建 厦门 361005)

²(厦门大学 计算机科学系, 福建 厦门 361005)

通讯作者: 苏畅, E-mail: suchang@xmu.edu.cn, http://information.xmu.edu.cn/portal/node/100



摘要: 隐喻理解已成为语言学、认知学、计算机科学等研究的重要课题,也是自然语言处理中不可避免的任务。提出一种基于相关性约束的隐喻理解方法,利用隐含的相关角度计算目标域和源域的相关程度。首先,基于词、词的主题及语篇的主题扩展出多层次的语义表示;然后,利用上下文信息的相关关系,构建多层次的相关性模型,模型通过多种角度的相关关系将跨层次的语义信息关联起来;接着,采用 random walk 的方法,通过迭代计算获得隐含角度的相关关系;最后,选择与目标域具有最大相关度的属性作为隐喻理解的结果。将模型应用到隐喻理解任务中,实验结果表明,该方法能够有效地实现隐喻自动理解。

关键词: 相关性;隐喻理解;语篇;角度

中图法分类号: TP391

中文引用格式: 苏畅,王晓梅,黄舒曼,陈怡疆.基于相关性约束的隐喻理解方法.软件学报,2017,28(12):3167-3182. <http://www.jos.org.cn/1000-9825/5247.htm>

英文引用格式: Su C, Wang XM, Huang SM, Chen YJ. Metaphor comprehension method based on relevance constraints. Ruan Jian Xue Bao/Journal of Software, 2017,28(12):3167-3182 (in Chinese). <http://www.jos.org.cn/1000-9825/5247.htm>

Metaphor Comprehension Method Based on Relevance Constraints

SU Chang¹, WANG Xiao-Mei¹, HUANG Shu-Man¹, CHEN Yi-Jiang²

¹(Department of Cognitive Science, Xiamen University, Xiamen 361005, China)

²(Department of Computer Science, Xiamen University, Xiamen 361005, China)

Abstract: Metaphor comprehension has become an important issue of linguistics, cognitive science and computer science. It is also an unavoidable task of natural language processing. This paper presents a novel metaphor comprehension method to make full use of global information based on relevance constraints. The method uses implied perspective to calculate the relevance degree between the target and source domains. First, multi-level semantic representation is obtained based on the semantic representation of word, topic features of word and topic features of discourse. Next, the degree of relevance relations is calculated and the relevance model is generated. Additionally, relevance relations is used to connect cross-level nodes from different perspectives. Then, using random walk algorithm, the relevance relations are acquired from latent perspectives through iterative computations. Finally, the target attribute that has the maximum relevance degree with the target domain is selected as the comprehension result. Experimental results show that the presented method is effective in metaphor comprehension.

Key words: relevance; metaphor comprehension; discourse; perspective

隐喻作为自然语言中的一种常见现象,不仅是一种修辞手段,更是人类的一种认知方式。近年来,隐喻理解作为隐喻计算的核心子课题,是自然语言处理领域中不可避免的任务。Hobbs^[1]认为:隐喻理解是语篇理解的组

* 基金项目: 国家自然科学基金(61075058)

Foundation item: National Natural Science Foundation of China (61075058)

收稿时间: 2016-05-27; 修改时间: 2016-08-18, 2016-10-14; 采用时间: 2016-11-11; jos 在线出版时间: 2017-01-20

CNKI 网络优先出版: 2017-01-20 16:06:37, <http://www.cnki.net/kcms/detail/11.2560.TP.20170120.1606.012.html>

成部分,隐喻只有在上下文中才能被正确理解.语篇理解是与语言、认知、心理及语境密切相关的复杂问题,它依赖于许多过程的交互、推理和整合.首先,人们对语篇信息进行编码,建立语篇的心理表征;然后,通过推理连接语篇的不同部分,并激活背景知识来解释上下文信息.许多因素会影响到语篇理解,比如词汇、语义成分、上下文信息、感知信息、先验知识等.一些研究开始强调语言理解中涉身感知的重要性^[2,3].涉身观点认为,理解涉及知觉模拟、感觉运动模拟和情感模拟^[4,5].脑成像研究表明,人们在阅读时选择性地激活了神经网络中对应感知、动作和情感的部分.理解这些因素的影响和它们之间的交互作用,对于理解的理论发展和模型建立是非常重要的.从隐喻的功能性角度看,隐喻理解可视为交际行为^[6].换句话说,它满足关联理论^[7,8].虽然隐喻和人与人之间的沟通没有完全相同的结构,但是关联理论的一些观点也适用于隐喻理解.根据关联理论,人类的认知倾向于最大化的相关性,因此,感知、记忆检索、推理等过程可能启发式地选择潜在的相关信息,并以增强相关的方式处理这些信息^[9].隐喻理解可以看做在语篇信息的约束下,耗费尽可能少的努力,获得最大关联性的过程.

目前,研究人员已经提出了许多隐喻理解的模型,其方法可以分为两类:基于统计的方法和基于推理的方法.这些模型大部分孤立于语篇之外,没有充分利用上下文信息及语义的相关性.本文从语篇的层面出发,利用上下文信息构建多层次的语义网络模型,并在模型的基础上提出一种基于相关性约束的隐喻理解方法.

本文的主要贡献有以下两点.

- (1) 模型包含多层次的语义特征,包括属性、概念、感知信息以及全局的特征信息;
- (2) 方法有效地利用了从具体到抽象的全局信息,利用跨层次节点间的相关性获得隐喻的隐含表述义.

本文第 1 节简要介绍隐喻理解的相关工作.第 2 节介绍相关关系、角度以及相关性约束.第 3 节具体描述我们的方法,并运用到隐喻理解任务中.第 4 节分析实验的结果.第 5 节总结全文.

1 相关工作

语料库语言学的发展,带动基于统计的隐喻理解方法的通行.游维等人^[10]在对大规模语料库进行统计研究的基础上,建立了基于向量空间的模型,并构建了隐喻句生成系统.苏畅等人^[11]在隐喻互动理论的基础上提出了基于语料库和知识库的方法,创新性地使用合作机制来对隐喻进行理解.杨芸^[12]采用特征匹配的方法,提出了基于目标概念属性驱动的隐喻意义获取机制.贾玉祥等人^[13]从网页中获取隐喻实例并构建知识库,提出了基于知识库的隐喻自动理解和生成系统.Shutova^[14]提出了基于语料库的动词隐喻理解方法,基于优先规则区分字面释义和隐喻释义.Shutova 等人^[15]提出了基于向量空间模型的动词隐喻理解方法.通过因式分解模型来计算目标动词的潜在特征,获得释义的条件概率.Bollegala 和 Shutova^[16]从网页中提取大量的模板来表示语义关系,通过优先选择规则获得动词隐喻的理解结果.

基于推理的隐喻理解方法以规则推导为主.张威^[17]从隐喻的语义真值问题和逻辑全知问题入手,参考局部框架理论,引入池空间、理解算子和格式塔等概念,提出了汉语隐喻逻辑系统.苏畅^[18]从理解者的理解角度出发,着眼于被理解的句子及其表述义之间的关系,构建了隐喻认知逻辑.Veale 等人^[19]提出一个隐喻理解与生成的知识表示系统(talking points),通过插入、删除、替换等规则建立概念属性与概念之间的联系,以此完成隐喻的理解.Ovchinnikova 等人^[20]采用外展推理的方法将隐喻句映射到对应的概念隐喻,结合概念隐喻的目标域和源域信息,通过谓词逻辑和自然语言表达式解释隐喻.Su 等人^[21]描述了基于潜在语义相似的隐喻理解方法,如果目标域的属性特征通过 WordNet 的同义词扩展能够与源域的某一属性特征建立联系,就认为两者之间存在潜在的语义相似,而该源域属性就构成了隐喻的解释.

现有的方法大部分针对于词对形式的隐喻,然而由于隐喻的多义性和模糊性,隐喻的上下文信息对隐喻的准确理解有着不可忽视的作用.目前,部分模型引入了上下文信息,比如,杨芸^[12]和 Shutova^[14]利用了句子内部的局部信息帮助隐喻理解,一定程度上扩充了句子内部的上下文的线索,但未考虑语篇信息的约束.且隐喻理解的框架是从语境中选取最合理的理解角度,而目前基于角度的相关性推理研究是有限的.

我们的方法基于 WordNet 语料库,充分结合统计和推理两种方案,突破词对和句子的限制,利用语篇级的上下文信息,通过多种角度的相关关系将语义信息联系起来,实现全局信息的跨层次关联,并构建了多层次的语义

网络模型,节点的多层次语义表示包括词的语义分布特征、词的主题特征以及语篇的主题特征,因此,我们的模型能够从上下文中捕捉到更多的语义信息,利用语篇信息的角度进行相关性推理。

2 相关性约束

首先,给出基本符号的说明,见表 1。

Table 1 Basic notations

表 1 基本符号说明

名称	描述
u	节点
U	节点集合
p, a, c, h, f P, A, C, H, F	不同层的节点,分别对应:感知层、属性层、概念/输入层、隐藏特征层和全局特征层
e	边
(u_i, u_j)	连接 u_i 和 u_j 的边
E	边的集合
θ	相关角度
Θ	相关角度的集合

定义 1(相关角度). 我们将相关角度 θ 定义为认识概念或情境的一种方式,例如性格、外观、颜色等。

定义 2(相关关系). 相关关系是一种语义关系,如果两个概念在某种角度上产生关联,就认为两者在该角度是相关的.相关关系的表达式为 $R(u_1, u_2, \theta)$,其中, u_1, u_2 表示节点, θ 表示 u_1 和 u_2 之间的相关角度. $R(u_1, u_2, \theta)$ 的值表示从角度 θ 看, u_1 和 u_2 的相关程度.任意两个节点具有相关关系,都可以表示成相关表达式。

相关是比相似更普遍的概念,语义相关度表明了两个概念在某个角度的相关程度^[22,23].当从同义关系的角度看,该相关关系等价于相似关系,即,相关性包含了相似性。

我们可以根据以下几种方式从语料库中获得相关关系的初始相关程度。

- (1) 从角度 θ 看, u_1 和 u_2 的共现概率;
- (2) 从角度 θ 看,如果 u_1 蕴含 u_2 ,那么 $R(u_1, u_2, \theta)=1$;
- (3) 从角度 θ 看,当 u_2 出现的情况下, u_1 出现的概率;
- (4) 从角度 θ 看, u_1 和 u_2 的上下文信息分布的相似性。

节点 u_1 和 u_2 间的相关关系可以看做约束满足问题 (X, D, C) 。

定义 3. 约束满足问题 (X, D, C) :

- (1) $X=\{\theta\}$ 是两个节点间的相关角度;
- (2) $D=\{d\}, d=\{\theta_i | i=1, \dots, n\}$ 是包含 X 的可能取值的有限集合;
- (3) $C=\{c\}$ 是约束条件.在这里, $c: R(u_1, u_2, \theta) > 0$ 。

定义 4. 相关约束:

- (1) $0 \leq R(u_1, u_2, \theta) \leq 1$;
- (2) 从角度 θ 看,如果 u_1 和 u_2 是相关的,那么 $R(u_1, u_2, \theta) > 0$;
- (3) 从角度 θ 看,如果 u_1 和 u_2 是不相关的,那么 $R(u_1, u_2, \theta) = 0$;
- (4) 从角度 θ 看,如果没有证据证明 u_1 和 u_2 是相关的,那么 $R(u_1, u_2, \theta) = 0$;
- (5) $\forall \theta \in \Theta, R(u_1, u_2, \theta) = 1$ 。

3 我们的方法

3.1 模型的结构

我们认为,理解模型需要从语言和篇章中提取多层次的语义^[24].我们在模型中关注节点间的相关关系,不同层次的节点通过不同角度的相关关系联系在一起。

相关性模型是一个基于语篇的多层语义网络(P, A, C, H, F, E),其中:节点 P 为属于感知层(perception level)的感知信息;节点 A 对应属性层(attribute level)的属性信息;节点 C 为输入层(concept/input level)中的概念信息;节点 H 对应隐藏特征层(feature(hidden))中的隐藏特征信息,其中,假定层数为 $x > 1$;节点 F 对应特征层(feature level)中的全局特征信息;边 $E \subseteq U \times U$ 对应节点间的相关关系.当且仅当 u_1 和 u_2 间有相关关系,模型中存在连接 u_1 和 u_2 的边 $(u_1, u_2) \in E$.基于初始的相关关系,模型通过计算,进一步获得稳定的相关关系.模型的基本结构如图 1 所示.

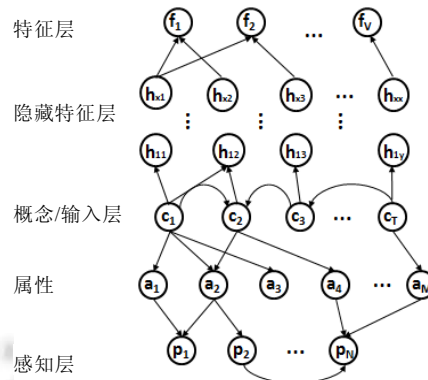


Fig.1 General structure of the model

图 1 模型的基本结构

下面,我们具体说明模型中的不同层次.

- 概念/输入层(concept/input level).

概念层中的节点对应输入的语篇信息.在这里,我们对输入的语篇信息进行简单的预处理,比如分词、去除停用词等.其中,停用词包括语气助词、介词、连词与副词等.最后,将预处理后所得到的词群转化为词向量.例如:对隐喻句“妈妈的爱如糖果一般,甜在我的心里”进行预处理.去除停用词“的”“如”“一般”与“在”,最终将得到词群“妈妈”“爱”“糖果”“甜”“我”与“心里”相关的词向量.

- 感知层(perception level).

感知层的节点对应具体的感知、情感信息.概念与味觉、听觉、触觉、嗅觉、视觉信息的相关关系建立涉身的语义关联.例如,抽象的情感词(属于属性层)是通过情感词的涉身含义理解的.我们考虑到概念间的相关性可能导致它们连接到相同或相似的涉身感知特性.又因为“通感”的存在,人们在描述和理解客观事物时,会把听觉、视觉、嗅觉、味觉、触觉等属于不同感官的感觉沟通关联起来.因此在我们的模型中,感知层也是重要的层次.

例如,在“妈妈的爱如糖果一般,甜在我的心里”句子中,“甜”对应具体的味觉信息,源域“糖果”与目标域“爱”借由“甜”这一感知信息产生了语义关联.

- 属性层(attribute level).

属性层的节点对应概念属性特征,比如描述概念的形容词、动词、副词等.

- 特征层(feature level).

除了最高层,其余的特征层都是隐藏层.隐藏层可以多于一层.特征层中的节点对应全局的特征,比如主题、文化、意图等.隐藏层支持进一步抽象全局特征.特征层存在向上一级抽象归纳的关系,其中,最底层对概念/输入层的节点进行抽象化,最高层是全局特征的总体抽象与概括.

如上所述,结合上下文信息可以捕捉到词汇丰富的语义信息.反过来,上下文信息的全局语义特征也会帮助决定词汇的语义范畴(例如词义消歧和实体消歧).输入的概念信息通过相关关系连接到较低层次的属性层和感知层节点,也连接到高层次的全局特征节点.模型的节点自底向上对应从具体到抽象的特征.

3.2 多层次语义表示

我们结合 3 种语义类型的向量表示(词、词的主题、语篇的主题)扩展出多层次语义向量表示,并计算节点间的初始相关程度.

3.2.1 词向量语义表示

我们根据每个词在语料中的分布信息获得词的语义表示.根据分布式假说^[25],具有相似分布属性的词通常有共享的语义含义.Mikolov 等人^[26]基于分布式假说提出了两种获得词向量语义表示的模型 Word2vec(<https://code.google.com/p/word2vec/>),分别是 skip-grams 和 continuous bag of words(CBOW).模型将每个词映射到几百维的向量中,而词的关系可以直接由词向量的计算获得.实验结果表明:两种模型的计算复杂度明显低于其他传统的神经网络模型,且在准确率上有较大的优势.本文采用 CBOW 模型获得词的向量表征,形式化为

$$V(w)=\{p_k|k:1,\dots,K\} \quad (1)$$

其中, K 是向量的维度, p_k 表示第 k 维的值.

3.2.2 词的主题特征

主题模型认为篇章中的每个词都是通过“以一定概率选择某个主题,并从这个主题中以一定概率选择某个词语”的过程得到.我们可以将每一篇文档视为不同主题构成的概率分布,而每一个主题又可以表示成不同单词构成的概率分布.反过来,可以获得每个词属于不同主题的概率分布.我们将词的主题分布概率作为词的主题特征.在本文中,使用 LDA 主题模型(<http://gibbslda.sourceforge.net>)来获得每个词的主题概率,通过简单标准化获得词的主题分布概率,形式化如下:

$$T(w)=\{p(t_k|w)|k:1,\dots,R\} \quad (2)$$

其中, t_k 表示第 k 个主题, R 是主题总数, $p(t_k|w)$ 表示词 w 属于第 k 个主题的概率.

3.2.3 语篇的主题特征

语篇的主题特征是一种抽象的语义特征,在一定程度上帮助决定词汇的语义范畴.例如,Lau 等人^[27]利用句内的主题特征进行词义消歧.对于包含隐喻的语篇,我们首先提取语篇的词语,表示为 $w_1, w_2, w_3, \dots, w_N$ (去除停用词).基于词的主题分布概率,我们获得语篇的主题特征,形式化如下:

$$D=\{d_k|k:1,\dots,R\} \quad (3)$$

$$d_k = \frac{1}{N} \sum_{i=1}^N p(t_k | w_i) \quad (4)$$

其中, N 是提取的词的总数.

3.2.4 多层次语义特征的整合

考虑到全局的相关性,我们将语篇的主题特征作为权值更新词的主题特征,形式化如下:

$$T'(w)=\{d_k \times p(t_k|w)|k:1,\dots,R\} \quad (5)$$

基于已有的 3 种类型的语义特征表示,给定词 w ,我们采用语义连接的方式获得 w 的多层次语义表征,形式化如下:

$$V^E(w)=(V^K(w), T^R(w), T'^R(w)) \quad (6)$$

其中, $E=(K+R+R)$.

3.2.5 初始相关程度计算

给定词 w_i 和 w_j ,通过正则化 $V^E(w_i)$ 和 $V^E(w_j)$ 的余弦距离来获得相关关系的初始相关程度,公式如下:

$$R(w_i, w_j, \theta) = \frac{\cos(V^E(w_i), V^E(w_j)) + 1}{2} \quad (7)$$

3.3 模型的生成

- 假设

给定一个隐喻,该隐喻能够被认知主体理解的充分必要条件是:隐喻中的目标域和源域在当前的语境下具有相关性,且这种语义相关能够被认知主体所捕获.

- 特征层

如第 3.2 节所述,我们扩展 3 种基本语义表示形成多层次的语义表示,这种语义表示方法将特征层的信息与其他层的节点潜在地联系在一起。

- 概念/输入层

给定包含隐喻的语篇,我们对语篇进行分词并去除停用词,获得的词 c_1, c_2, \dots, c_N 构成输入层。

- 属性层和感知层

我们从知识库中获得目标域和源域的属性,基于情感的角度生成属性层.保留与源域的属性具有相关关系的目标域属性,构成属性层和感知层.本文选择斯坦福大学 NLP 实验室开发的情感分析工具 Stanford CoreNLP (<http://nlp.stanford.edu/software/corenlp.shtml#sentiment>)作为情感评估的来源.CoreNLP 将情感值量化为 5 类:0(非常消极)、1(消极)、2(无明显的情感倾向)、3(积极)、4(非常积极)。

我们假设 w_1 是 w_2 的同义词,那么:

- (1) 从情感的角度看,如果 w_1 和 w_2 都是积极的,那么 $R(w_1, w_2, \theta)=1$;
- (2) 从情感的角度看,如果 w_1 和 w_2 都是消极的,那么 $R(w_1, w_2, \theta)=1$;
- (3) 从情感的角度看,如果 w_1 或者 w_2 没有明显的情感倾向,那么 $R(w_1, w_2, \theta)=1$ 。

基于上述的定义和假设,我们给出属性层的生成算法,即,通过计算目标域属性和源域属性之间的相关关系生成属性层.算法基于 WordNet(a lexical database for English. <http://wordnet.princeton.edu/>)的同义词扩展,扩展中涉及属于感知层的潜在的具体感知信息和情感信息。

根据文献[21]的理论,我们将属性层的生成算法限制在 6 次扩展以内.如果目标域的属性经过 6 次扩展仍未与源域的属性产生相关关系,那么我们认为该目标域的属性与源域不存在相关性,从属性层中移除.属性层的生成算法见算法 1。

算法 1. 属性层的生成算法.

Require:

- 目标域的属性集合: T ;
- 源域的属性集合: S ;
- 扩展中的集合: $P=\emptyset, Q=\emptyset, Q'=\emptyset$;
- 生成的属性层集合: A ;

Ensure:

- 1: for $i=1$ to $|T|$ do
- 2: $P=P \cup t_i$;
- 3: for $j=1$ to 6 do
- 4: for $l=1$ to $|P|$ do
- 5: 获得 p_l 的同义词集 Q'
- 6: for $k=1$ to $|Q'|$ do
- 7: if 从情感的角度看, $R(p_l, q_k, \theta)=1$ then
- 8: if $q_k \in S$, then 将 t_i 加入属性层 A ; goto (17);
- 9: else
- 10: 从 Q' 中删除 q_k ;
- 11: end if
- 12: end for
- 13: $Q=Q \cup Q'$;
- 14: end for
- 15: $P=Q$;

16: end for
 17: $P=\emptyset, Q=\emptyset, Q'=\emptyset;$
 18: end for

3.4 基于相关性约束的隐喻理解

文献[28]将隐喻分为3种类型:I类、II类和III类隐喻.在I类隐喻(又称名词隐喻)中,一个名词通过动词“be”与另一个名词产生关联,例如“*John is a lion.*”在II类隐喻(又称动词隐喻)中,动词作用在名词上构成隐喻,如“*How can I kill a process?*”III类隐喻(形容词隐喻)中,由名词和修饰它的形容词构成隐喻,如“*The book is dead.*”我们基于相关性提出名词隐喻(I类)理解方法.文化的差异性也会影响到隐喻的理解,在本文中,我们认为文化信息隐含在语篇中.

隐喻理解可以看做是寻找隐喻要突显的目标域属性, $a=a_i$,可以形式化为:寻找在当前语境下,与目标域具有最大相关程度的目标域属性,该角度 θ 即为隐喻的理解角度:

$$a = \operatorname{argmax}_{a_i} \{R(\text{Target}, a_i, \theta)\} \quad (8)$$

我们基于全局的相关性,通过迭代更新相关关系的相关程度,计算方法如下.

如果 $R(u_1, u_2, \theta_i) \wedge R(u_2, u_3, \theta_j)$,那么从 $\theta_i \circ \theta_j$ 角度上看, u_1 和 u_3 具有相关关系,表示为 $R(u_1, u_3, \theta_i \circ \theta_j)$.其中, $R(u_1, u_3, \theta_i \circ \theta_j)$ 的值等于 $R(u_1, u_2, \theta_i) \times R(u_2, u_3, \theta_j)$.

我们采用随机游走(random walk)的方法获得语篇和潜在信息之间稳定的相关程度.从目标域出发,我们可以获得目标域与其他节点之间的相关关系和相关程度.我们选择与目标域具有最大相关程度的目标域属性作为隐喻的理解结果.

模型的节点 $\text{Node}=\{\text{node}_k|k<N+G\}$ 包含输入层节点 $\{c_1, c_2, \dots, c_N\}$ 以及属性层节点 $\{a_1, a_2, \dots, a_G\}$. M 为 $(N+G) \times (N+G)$ 的相关程度矩阵,其中,

$$M_{ij} = \frac{R(\text{node}_i, \text{node}_j, \theta)}{\sum_{k, k \neq j}^{N+G} R(\text{node}_j, \text{node}_k, \theta)}, i \neq j \quad (9)$$

$$M_{ii} = 0, i < N+G \quad (10)$$

那么,目标域的相关程度向量 $R=\{R(\text{Target}, \text{node}_k, \theta)|k<N+G\}$ 的计算如下:

$$R=cMR+(1-c)s \quad (11)$$

其中, s 是 $(N+G) \times 1$ 的初始向量,对应目标域和其他节点间初始相关程度; c 是阻尼因子,值的范围在 0~1 之间.在本文中,我们将阻尼因子设为 0.8.在初始的 R 中, $R_{\text{target}}=1, R_j=0(j \neq \text{target})$,表示从目标域节点开始随机游走.迭代计算,直到相关程度向量 R 达到稳定状态.在稳定的情况下, R 的各个维度的数值代表着其对应节点与目标域之间的相关程度.我们选择具有最高相关程度的属性节点作为理解结果.

使用随机游走的方法具有以下两种优势.

- (1) 实现了同层次和跨层次之间的随机相关计算;
- (2) 因为任意两个节点都可能在某一些角度上形成相关,我们在计算的过程中不需要获得角度的具体信息,而随机游走实现了角度的随机性.

隐喻理解算法见算法 2.

算法 2. 隐喻理解算法.

Require:

包含隐喻的语篇,已标注的目标域和源域.

Ensure:

- 1: 从语篇中提取词 c_1, c_2, \dots, c_N (包括目标域和源域)构成概念/输入层.
- 2: 从知识库中获取目标域和源域的属性.
- 3: 生成模型的属性层.

- 4: 获得节点间的初始相关程度.
- 5: 获得相关程度矩阵 M 和向量 s .
- 6: 从目标域开始进行随机游走, $R_{target}=1, R_j=0(j \neq target)$.
- 7: $d_{R'R} = \text{MAX}$;
- 8: while $d_{R'R} > \varepsilon$ do
- 9: $R' = cMR + (1-c)s$;
- 10: 计算 R 和 R' 间的欧式距离(Euclidean distance) $d_{R,R}$;
- 11: $R = R'$;
- 12: end while
- 13: R 中具有最高相关程度的属性 $a = a_i$, 形成隐喻的理解结果“Target is a_i ”.

4 实验

本文使用《读者》(中文杂志.URL:www.duzhe.com)作为语料库,收录百余篇文章,共 23.6MB.使用厦门大学 NLP 实验室的分词工具 Segtag 对语料进行预处理:分词、命名实体识别、去除停用词,搭建中英文对照属性库.我们通过 word2vec 将词 w 映射到 100 维的向量空间(即 $K=100$),设置 LDA 的主题数量为 100(即 $R=100$).在这部分,我们结合例子说明实验的具体过程.

考虑下面的例子:

航空工业是国防科技工业大家庭的重要一员,它的发展水平体现了一个国家的综合国力.航空发动机被视为心脏,其性能在很大程度上决定了飞机的飞行速度、飞行高度、机动性、航程、有效载荷、安全等性能,直接影响军用飞机的作战能力和出勤率.某种意义上说,一个国家没有独立自主发展的航空发动机事业,就没有独立自主发展的航空工业,也就不能有独立的空防力量.

(The aviation industry is an important member of the big family of defense science industries, the development level of which reflects a country's comprehensive national strength. The aircraft's engine is regarded as the heart; to a great extent, its performance determines the aircraft's performance, such as flight speed, flight altitude, maneuverability, range, payload and security. The engine directly affects the operational effectiveness and attendance rate of military aircraft. In a sense, without independent development of the aircraft engine, a country cannot have independent development of the aviation industry and independent air defense forces.)

其中,标注的目标域是“发动机(engine)”、源域是“心脏(heart)”.带下划线的词为构建概念/输入层的上下文词.基于词的主题特征,我们根据第 3.2 节获得了该语篇的主题特征向量.

我们从在线知识库 Lex-Ecologist(the Lex-Ecologist is a context attribute query system produced by Creative Language Systems Group. <http://afflatus.ucd.ie/lexeco/index.jsp>)和 English CogBank(the database is extracted from Google. <http://www.cognitivebase.com>)^[29]中获得目标域和源域的属性(部分属性信息如图 2 所示).

结合 WordNet 和 Stanford CoreNLP,根据属性层生成算法,我们获得与源域属性相关的目标域属性(图 2 中带下划线的属性),这些属性构成模型的属性层.

接着,我们计算节点之间的初始相关程度.基于这些相关关系,我们通过随机游走更新节点间的相关程度,迭代获得模型的稳定状态.最后,将英文属性与之前所建立的中文属性库进行匹配.稳定状态下,目标域与属性之间的相关程度排序结果见表 2.

Attributes of Target(engine) frightful major electric efficient complex smooth damaged heavy dangerous universal extensive unique powerful fine new analytical reliable larger different international straight modern basic efficient important traditional compact advanced continental mechanical rudimentary complex compact ...
Attributes of Source(heart) vital broken near essential aggressive passionate sacred industrial human immaculate important old good lost commercial heavy novel lonely financial pure kind historical cultural new great cold spiritual whole warm left big political dead foolish dark open severe secret traditional weak national loving dear real economic green sad red ...

Fig.2 Part of the attributes of target and source domains

图 2 目标域和源域的部分属性

Table 2 Rank of relevance degrees between the attributes and target domain

表 2 目标域和源域的部分属性属性与目标域间的相关程度排序

属性	相关程度
重要(important)	0.0214371
高效(efficient)	0.0146519
破坏性(disruptive)	0.0146124
常规(conventional)	0.0145793
劣等(inferior)	0.0133275
机械(mechanical)	0.0129293
紧凑(compact)	0.01225507
坚固(rugged)	0.0122875
备用(spare)	0.0121588
强大(powerful)	0.0118181
关键(critical)	0.0115586
可靠(reliable)	0.011182
基本(fundamental)	0.0105748
危险(dangerous)	0.0104118
...	...

我们选择与目标域之间具有最高相关程度的属性节点作为理解结果——“发动机是重要的(engine is important)”。

4.1 结果与分析

4.1.1 测试数据

实验使用的测试集是从网络、博客、课本中抽取的 256 个语篇,其中包含 300 句名词隐喻,人工标注了隐喻的目标域和源域。数据涵盖了新闻、评论、小说、散文等多种类型。部分测试集及隐喻理解结果在附录 A 中给出。

4.1.2 评价方法

采用对测试结果进行人工评测的方法评价实验结果。我们邀请了 5 位评价者,这些评价者都是以汉语为母语,并具有一定的语言学基础。

首先,要求评价者对语料进行隐喻用法和字面义用法标注,通过检验标注结果的一致性来测试人工评测的可靠性,并用 κ 参数^[30]衡量。5 位评价者的一致性达到 0.66(κ),此一致性程度可以认为评价者的评价是可靠的。评价不一致的主要原因是:有些常规隐喻已经固化成常用用法、词组等形式,因此有些评价者标注为字面义表达。

在评价可靠的基础上,我们给出测试数据及结果,要求评价者按照以下标准对解释结果进行评价,每个评价者单独给出隐喻理解结果的可接受程度评分。

- (1) 评价者结合语篇信息,按照自己对该隐喻的理解,根据直觉对结果进行评分;
- (2) 可接受度分值为 1~5,其中,1 表示“该理解结果非常难以接受”;2 表示“该理解结果难以接受”;3 表示“对该理解结果不置可否”;4 表示“能够接受该理解结果”;5 表示“非常能够接受该理解结果”。

我们将可接受度评分分为 5 个级别,而不是传统的“是”与“否”的二值评价。相比于二值评价,我们的评价方法的评价粒度更细,能够减少由于评价者对隐喻理解结果认可程度的不一致造成的误差。而且我们可以将评价

结果转换为二值评价结果.在本文中,我们认为平均可接受程度大于等于 3 的隐喻理解结果是正确的;否则,理解结果就是不正确的.

4.1.3 结果分析

为了更好地说明方法的性能,我们将该方法与下列 3 种方法作比较.

- (1) 随机:随机选择目标域的属性作为理解结果;
- (2) 无属性层计算:在模型生成部分,保留目标域的所有属性构建属性层;
- (3) 不考虑上下文信息:在模型生成部分,仅保留目标域和源域构建概念/输入层.

评价的指标是准确率,即,理解结果的可接受程度不小于 3 的隐喻句数目与测试句的总数目的比值.

实验结果见表 3.

Table 3 Results of four methods

表 3 4 种方法的结果

	随机方法	无属性计算方法	不考虑上下文方法	我们的方法
准确率	0.18	0.44	0.52	0.89

实验结果表明:我们的方法在一定程度上具有较高的准确率,对比其他 3 种方法有较大的提高.其中,与 WA 方法比较,说明我们的方法能够有效地获取目标域属性和源域属性的相关性,从而帮助理解隐喻.与 WI 方法的比较说明,我们的方法有效地利用语篇的上下文信息帮助理解隐喻.尽管有些隐喻的理解结果仍然是隐喻的,但是理解结果的具体程度高于隐喻句本身,而越具体的内容越容易被理解.对于某些隐喻句,因为目标域的某些属性未被提取,导致我们的方法不能获得合适的理解结果.比如在如下例子中:

“有人说爱是一朵艳丽的花朵,开放的时候,开始含苞迷人,绽放绚丽;当成为果实的时候,就再也没有含苞的迷人,绚丽的心动了.那么爱情的果实其实就是婚姻,是不是婚姻就不再有迷人和绚丽了呢?当然不是,关键是你如何看待这个爱情的果实.花的果实也有美丽的,如果像罂粟一样,有着绚丽迷人的开始,果实的色彩却会让人不寒而栗.如果**爱情是花朵,婚姻是果实**,这应当是一种美好的寓意.花是美丽的,果实是希望的结果,是一种美丽的结局.按照这种比喻,婚姻应当是美丽的延续,是使花朵再次成为美丽的基石,这样的结果是为了比爱情更多的美丽而诞生的准备,这就要看你是否懂得去经营这朵美丽了,是就此成为解饥吞咽,还是美好的温床.”

对于并列的两个隐喻“爱情是花朵,婚姻是果实”,通过我们的方法所得到的理解结果是:爱情是美丽的,婚姻是亲密的.前者的理解结果符合语义表达.后者的隐喻表达中,目标域“婚姻”的属性之一“结果”未被提取,且“亲密”和水果的某个属性能在 6 次扩展产生联系,在随机游走的过程中具有最高相关程度,导致理解的结果产生了偏差.

因此,如何改进属性的抽取机制也是未来不可或缺的工作.

5 总 结

本文利用全局信息的相关性构建多层次的相关性模型,并运用到隐喻理解任务中.相关性模型包括多层次语义特征,不仅包括概念、概念的属性特征以及感知特征,还包含尽可能多的全局特征信息,如上下文信息、背景知识、主题、文化、情感等.模型通过多种角度的相关关系将语义信息联系起来,实现了全局信息间的跨层次关联.我们在模型的基础上提出一种基于相关性约束的隐喻理解方法,充分利用全局信息实现隐喻理解.实验结果证明,我们的方法能够有效地理解隐喻的隐含表述义.

References:

- [1] Hobbs JR. Metaphor and Abduction. Communication from an Artificial Intelligence Perspective. Berlin, Heidelberg: Springer-Verlag, 1992. 35–58. [doi: 10.1007/978-3-642-58146-5_3]
- [2] Horchak OV, Giger JC, Cabral M, Pochwatko G. From demonstration to theory in embodied language comprehension: A review. Cognitive Systems Research, 2014,29:66–85. [doi: 10.1016/j.cogsys.2013.09.002]

- [3] Zhou CL. Metaphor as a mean of cognition and its embodiment analysis. *Mind and Computation*, 2008,2(3):272–278 (in Chinese with English abstract).
- [4] Glenberg AM, Witt JK, Metcalfe K. From the revolution to embodiment 25 years of cognitive psychology. *Perspectives on Psychological Science*, 2013,8(5):573–585. [doi: 10.1177/1745691613498098]
- [5] Zwaan RA. Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences*, 2014,18(5): 229–234. [doi: 10.1016/j.tics.2014.02.008]
- [6] Blakemore D. Discourse and relevance theory. In: *The Handbook of Discourse Analysis*, 2003,18:100. [doi: 10.1002/9780470753460.ch6]
- [7] Sperber D, Wilson D. *Relevance: Communication and Cognition*. 2nd ed., Oxford: Blackwell Publishers, 1995. 118–123.
- [8] Wilson D, Sperber D. Relevance theory. In: *Handbook of Pragmatics*. Oxford: Black Publishers, 2004. 249–287.
- [9] Wilson D, Carston R. Metaphor and the ‘emergent property’ problem: A relevance-theoretic treatment. *The Baltic Int’l Year Book of Cognition, Logic and Communication*, 2008,3(2007):1–40. [doi: 10.4148/biyelc.v3i0.23]
- [10] You W, Zhou CL. Chinese metaphor generation model and its implementation based on statistics. *Mind and Computation*, 2007,(1): 133–141 (in Chinese with English abstract).
- [11] Su C, Zhou CL. Chinese nominal metaphor comprehension method based on cooperation mechanism. *Application Research of Computers*, 2007,24(9):67–70 (in Chinese with English abstract). [doi: 10.3969/j.issn.1001-3695.2007.09.021]
- [12] Yang Y. Computational model of Chinese metaphor recognition and interpretation [Ph.D. Thesis]. Xiamen: Xiamen University, 2008 (in Chinese with English abstract). [doi: 10.7666/d.y1445419]
- [13] Jia YX, Yu SW. Metaphor comprehension and production based on examples. *Computer Science*, 2009,36(3):138–141 (in Chinese with English abstract). [doi: 10.3969/j.issn.1002-137X.2009.03.037]
- [14] Shutova E. Automatic metaphor interpretation as a paraphrasing task. In: *Proc. of the Human Language Technologies: The 2010 Annual Conf. of the North American Chapter of the Association for Computational Linguistics*, 2010. 1029–1037.
- [15] Shutova E, Van de Cruys T, Korhonen A. Unsupervised Metaphor Paraphrasing using a Vector Space Model. *COLING (Posters)*, 2012. 1121–1130.
- [16] Bollegala D, Shutova E. Metaphor interpretation using paraphrases extracted from the Web. *PloS One*, 2013,8(9):e74304. [doi: 10.1371/journal.pone.0074304]
- [17] Zhang W, Zhou CL. The logic description of Chinese metaphor comprehension. *Journal of Chinese Information Processing*, 2004, 18(15):23–28 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-0077.2004.05.004]
- [18] Su C. The research of Chinese nominal metaphor computation [Ph.D. Thesis]. Xiamen: Xiamen University, 2008 (in Chinese with English abstract). [doi: 10.7666/d.y1445417]
- [19] Veale T, Hao Y. A fluid knowledge representation for understanding and generating creative metaphors. In: *Proc. of the 22nd Int’l Conf. on Computational Linguistics, Vol.1*. 2008. 945–952. [doi: 10.3115/1599081.1599200]
- [20] Ovchinnikova E, Israel R, Wertheim S, Zaytsev V, Montazeri N, Hobbs J. Abductive inference for interpretation of metaphors. In: *Proc. of the 2nd Workshop on Metaphor in NLP*. 2014. 33–41. [doi: 10.3115/v1/w14-2305]
- [21] Su C, Tian J, Chen Y. Latent semantic similarity based interpretation of Chinese metaphors. *Engineering Applications of Artificial Intelligence*, 2016,48:188–203. [doi: 10.1016/j.engappai.2015.10.014]
- [22] Ballatore A, Bertolotto M, Wilson DC. An evaluative baseline for geosemantic relatedness and similarity. *GeoInformatica*, 2014, 18(4):747–767. [doi: 10.1007/s10707-013-0197-8]
- [23] Strube M, Ponzetto SP. WikiRelate computing semantic relatedness using wikipedia. In: *Proc. of the National Conf. on Artificial Intelligence*. Menlo Park: AAAI Press, 2006. 1419–1424.
- [24] Frank SL, Koppen M, Noordman LGM, Vonk W. Modeling multiple levels of text representation. In: *Proc. of the Higher Level Language Processes in the Brain: Inference and Comprehension Processes*. 2007. 133–157.
- [25] Sahlgren M. The distributional hypothesis. *Italian Journal of Linguistics*, 2008,20(1):33–54.
- [26] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv: 1301.3781*, 2013.
- [27] Lau JH, Collier N, Baldwin T. On-Line trend analysis with topic models: Twitter trends detection topic model online. In: *Proc. of the COLING*. 2012. 1519–1534. [doi: 10.3115/v1/d14-1212]
- [28] Krishnakumar S, Zhu X. Hunting elusive metaphors using lexical resources. In: *Proc. of the Workshop on Computational Approaches to Figurative Language*. 2007. 13–20. [doi: 10.3115/1611528.1611531]

[29] Li B, Chen J, Zhang Y. Web Based Collection And Comparison Of Cognitive Properties in English and Chinese. In: Proc. of the NAACL-HLT 2012 Joint Workshop on Automatic Knowledge Base Construction and Web-Scale Knowledge Extraction (AKBC-WEKEX). Montreal, 2012.

[30] Siegel S. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill, 1956. 174-193.

附中文参考文献:

[3] 周昌乐.作为认知手段的隐喻及其涉身性分析.心智与计算,2008,2(3):272-278.

[10] 游维,周昌乐.基于统计的汉语隐喻生成模型及其系统实现.心智与计算,2007,(1):133-141.

[11] 苏畅,周昌乐.基于合作机制的汉语名词性隐喻理解方法.计算机应用研究,2007,24(9):67-70. [doi: 10.3969/j.issn.1001-3695.2007.09.021]

[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. [doi: 10.7666/d.y1445419]

[13] 贾玉祥,俞士汶.基于实例的隐喻理解和生成.计算机科学,2009,36(3):138-141. [doi: 10.3969/j.issn.1002-137X.2009.03.037]

[17] 张威,周昌乐.汉语隐喻理解的逻辑描述初探.中文信息学报,2004,18(15):23-28. [doi: 10.3969/j.issn.1003-0077.2004.05.004]

[18] 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. [doi: 10.7666/d.y1445417]

附录 A:基于潜在角度的隐喻理解结果及其可接受度

Context							
航空工业是国防科技工业大家庭的重要一员,它的发展水平体现了一个国家的综合国力.航空发动机被视为飞机的“心脏”,其性能在很大程度上决定了飞机的飞行速度、飞行高度、机动性、航程、有效载荷、安全等性能,直接影响军用飞机的作战能力和出勤率.某种意义上说,一个国家,没有独立自主发展的航空发动机事业,就没有独立自主发展的航空工业,也就不能有独立的空防力量.因此,党和国家非常重视发展航空工业,非常重视发展具有完全自主知识产权的航空发动机事业.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
发动机是心脏	发动机是重要的	4.0	5.0	4.0	5.0	5.0	4.6
Context							
这个倔驴不好对付.克鲁格曼生性桀骜不驯,言辞尖锐,被业界称为倔驴.在布什上台后不久,他毅然拿起了笔,在《纽约时报》专栏中以犀利的文笔和缜密的思维,深入剖析布什政府的内外政策.从布什政府的减税政策到伊拉克战争,他经常在大众媒体人云亦云之际一针见血地戳穿布什政策背后的骗人把戏,与华盛顿的“御用”经济学家们形成鲜明对比.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
他是驴	他是倔的	5.0	5.0	5.0	3.0	5.0	4.6
Context							
夜色像一朵野花那样柔和地合拢来了.我坐在船头,静静地等待着酉水的月亮.酉水,在缓缓地流淌,喃喃自语着,仿佛一对久违的恋人,在花前月下拥抱着在一起的窃窃私语声.波浪,在轻轻地荡漾,好像一位慈母哄着怀中的婴儿,轻柔地亲吻着泊在岸边的渔船,两岸的柳树和水草.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
夜是野花	夜是宁静的	5.0	3.0	5.0	3.0	4.0	4.0
Context							
他是狼,他们是狼群.千千万万,以守护自己国家之名,甘愿拿生命做赌注,以死赴战.“杀——!”冷锋撕破喉咙发出的叫喊,用刺刀以最原始的方式冲向敌人的时候,我仿佛看到他的,不是孤身一人.在他的身后跟着的,是一群为建立我们国家之最初,浴血抗战的军人.他的喊声,是最为原始的冲锋号,坚定又嘹亮,让保卫国家的狼群们,以护国家之名,勇敢地冲向侵犯领土的敌人.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
他是狼	他是强大的	4.0	3.0	4.0	4.0	4.0	3.8
Context							
满院的丁香树.每年4月初是法源寺丁香花开的时节,寺中紫色、白色丁香花300余株争相盛开,在北京法源寺赏丁香是一场盛事.可惜现在我来晚了,在四月初来法源寺赏丁香是最好的时间,整座寺院丁香如雪,花瓣似玉,娇怜欲滴,幽香馨鼻,让游人痴醉于其中.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
丁香如雪	丁香是白色的	5.0	5.0	5.0	4.0	4.0	4.6
花瓣似玉	花瓣是娇嫩的	5.0	3.0	5.0	3.0	4.0	4.0

Context							
有人说爱是一朵艳丽的花朵,开放的时候,开始含苞迷人,绽放绚丽,当成为果实的时候,就再也没有含苞的迷人,绚丽的心动了.那么爱情的果实其实就是婚姻,是不是婚姻就不再有迷人和绚丽了呢?当然不是,关键是你如何看待这个爱情的果实.花的果实也有美丽的,如果像罂粟一样,有着绚丽迷人的开始,果实的色彩却会让人不寒而栗.如果 爱情是花朵,婚姻是果实 ,这应当是一种美好的寓意.花是美丽的,果实是希望的结果,是一种美丽的结局,按照这种比喻,婚姻应当是美丽的延续,是使花朵再次成为美丽的基石,这样的结果是为了比爱情更多的美丽而诞生的准备,这就要看你是否懂得去经营这朵美丽了,是就此成为解饥吞咽,还是美好的温床.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
爱情是花朵	爱情是美丽的	5.0	5.0	5.0	4.0	5.0	4.8
婚姻是果实	婚姻是亲密的	3.0	1.0	2.0	1.0	2.0	1.4

Context							
如果你见过一次木棉树,你会不容易忘记它的样子.木棉树的主干非常挺直,有青松的气概,它的 树干上那些鳞片,像武士那有大钉的铁甲 ,显得凛然不可侵犯的样子.但是,它并不是满身刺,整棵树也不是很张扬,它的 旁枝在主干上很齐整地横生出去,也是直直的,像健美运动员伸展的手臂 .横枝不多,秋冬天气落尽叶子的木棉树,显得很坚韧.木棉花是春天的象征,花一开,春天就到了.每年三月,木棉树那没有叶子的枝条上开出了朵朵 大红花,像一支支燃烧的火炬 ,格外引人注目.我喜欢木棉树,喜欢它的挺拔,坚韧.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
刺是大钉	刺是尖锐的	5.0	5.0	5.0	4.0	5.0	4.8
枝干是手臂	枝干是直的	5.0	5.0	5.0	4.0	5.0	4.8
木棉是火炬	木棉是鲜艳的	5.0	3.0	5.0	4.0	4.0	4.2

Context							
你们瞧瞧树上的木棉花,金黄金黄的 木棉花像是树顶上的金灿灿的太阳 ;火红火红的 木棉花像是一团烈火 一团团火球在顶上燃烧着,它红得那么庄严、稳重,它红得那么鲜艳夺目,难怪人们都叫它“英雄花”.清晨回到学校,在远方遥望着那美丽的木棉花,它开得这样鲜艳,火热,阳光一照,仿佛一个个火炬在眼前灼灼闪耀,微风一吹,又好像 千万面旗帜在风中飞扬 .							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
木棉是太阳	木棉是金黄的	5.0	5.0	5.0	1.0	3.0	3.6
木棉是烈火	木棉是红的	5.0	5.0	5.0	5.0	4.0	4.8
木棉是旗帜	木棉是红的	5.0	5.0	4.0	5.0	4.0	4.6

Context							
夏天的一个清晨,太阳冉冉升起,太阳光洒在水面上,波光粼粼,金光闪闪.湖面有一朵朵荷花是小鱼儿休息的屋子吧?荷花上好像有一群如花似玉的姑娘在跳舞吧? 湖水像一面镜子 映出了蓝天,白云还有变换的山峦.哟,清澈的湖水里有许多小鱼在游,一条银白色的小鱼跃出水面,又在浪花中消失,好像在给游人表演呢!湖的中央有一个亭子,小孩正在开开心心地玩耍,大人正在谈笑风生,老年人在专心致致地下棋.亭子不远处有一座假山,假山被无名的花儿点缀的更加迷人,迷人的四周群山环绕,树木茂盛.那边的山石像一只展翅欲飞的雄鹰,这边的山石像一只正要跳起的青蛙,半山腰的石兔,石龟好像正在赛跑呢!							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
湖水是镜子	湖水是清澈的	4.0	5.0	5.0	5.0	5.0	4.8

Context							
创业者的万丈豪情大多来自对汽车后市场的俯视,有高度是好事,但看到的只是风景.消费者的痛点;汽车后市场的无序竞争,没有一个哪怕是在一个细分市场能够呼风唤雨的重量级企业,更别说服务品牌;汽车后市场无论 B2B,C2B,O2O,还是 IT 技术(包括硬件和软件),思维理念等等,都需要变革,因此革命与颠覆的言论有了市场,创业的切入口到处都是.再加上巨大的,就在眼前的刚性需求,几乎 20 年来各种产业调研报告显示的高速增长数据,预期更是美好,让 汽车后市场对创业者的吸引力无比巨大.笔者分明从许多想要在汽车后市场创业的朋友的眼眸中看到了一望无际的大草原 ,水草丰美,遍地牛羊.不少行内的朋友,一聊模式没有不“高大上”的,一聊运营行业内的全是土鳖,一聊管理行内的都是笨猪.不少行内兄弟交流也是一样论调,汽车后市场奋斗的兄弟的确没有做好,辩解无用.无论是朋友,还是兄弟都是想在汽车后市场有一番作为,笔者在此将分享一下汽车后市场风景之后的生态,仅代表个人观点,欢迎斧正!							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
市场是草原	市场是巨大的	4.0	5.0	5.0	4.0	5.0	4.6

Context							
婚姻就是一场战争 ,能将自己的意志强加在别人头上,这就是权力.依照这个定义,权力几乎无所不在,它绝不仅仅存在于国与国之间、公司与公司之间,而是存在于一切人际关系中,包括婚姻家庭中,“婚姻就是一场男人和女人争夺话语权、经济权、掌控权的战争.”这是来自热播剧《婚姻保卫战》的台词.剧中许小宁、郭洋、老常这 3 个家庭的老公最终沦落为“家庭煮夫”,看着强悍的老婆冲锋陷阵,只能可怜巴巴地反抗、挣扎.以往男强女弱的婚姻关系在这里彻底颠覆.男女势均力敌,或者女性占有的社会资源越来越多,在家庭中的主导地位越来越强势,婚姻生活中就会出现分工交叉甚至角色互换的情况,传统婚姻中两性角色的界限越来越模糊.婚姻家庭是社会存在的基本形式,一个家庭就是一个国家中的小个体,其中的权力分配同国家管理一样,也存在各有分工各负其责,只是一个权力大小的问题.一个完整和谐的家庭,离不开权力的平衡问题,夫妻双方,会利用自己的智慧维持婚姻这块跷跷板的平衡.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
婚姻是战争	婚姻是平等的	4.0	1.0	2.0	1.0	3.0	1.8

Context						
清晨,田野里,麦叶身上的露水,像万颗珍珠在阳光下闪耀着晶莹灿烂的光辉.在那太阳下面展开着的草的叶片之上,在那生气勃勃的麦的新芽之上,露珠好像串在线上的玻璃小珠一样颤抖着.叶子轻轻地摇动了几下,几颗小露珠就调皮地躲进了草丛,再也寻找不到它们了.						

Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
露水是珍珠	露水是闪耀的	5.0	5.0	5.0	3.0	5.0	4.6
露珠是玻璃	露水是纯净的	3.0	5.0	5.0	2.0	4.0	3.8

Context						
诺拉-盖特利的养父母另有四个孩子,其中两个也是领养的.在美国养父母家中生活一年后,诺拉-盖特利与养父母关系生变.她被指殴打家里另一个孩子,于是短暂离家.诺拉 14 岁时,养父母没有说明任何原因,就把她带到田纳西州,交给汤姆和黛博拉-史密兹夫妇.诺拉-盖特利说,当她被养父母转让给史密兹夫妇后,她的生活成了一场噩梦.按照美国的领养法律,养父母如果将领养儿童转给其他的领养人需要签署法律文件,而跨州转换被领养人是非法的.但事实上,养父母私下将领养儿童转给其他领养人却是屡见不鲜.史密兹夫妇当时已有 9 个领养的孩子,后来增加到 17 个,而拉诺拉-盖特利的养母黛博拉则是利用严厉的体罚和精神压力来管理家中的十几个孩子.拉诺拉-盖特利说养母曾令她到院子挖坑,要她自掘坟墓.养母明确地告诉她,我不会在乎你的死活,没有人会知道你生活在这里.诺拉-盖特利除了必须照顾较小的孩子外,还曾被彻夜用牙刷刷地板.黛博拉还不时把她的右脚支架拿走,使她无法走动.						

Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
生活是噩梦	生活是可怕的	5.0	3.0	4.0	2.0	4.0	3.6

Context						
生命又如一棵弱不禁风的小草,那么脆弱,稍不注意,它便会从你手中悄悄溜走.有了安全,我们才能快乐地成长;有了安全,父母才能放心地工作;有了安全,我们国家才会和谐安康、繁荣昌盛……安全,像一条长长的纽带,维系着我们的生死存亡、喜怒哀乐.						

Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
生命是小草	生命是脆弱的	5.0	5.0	5.0	5.0	5.0	5.0

Context						
历史已有记载,他是敦煌石窟的罪人.我见过他的照片,穿着土布棉衣,目光呆滞,畏畏缩缩,是那个时代到处可以遇见的一个中国平民.他原是湖北麻城的农民,逃荒到甘肃,做了道士.几经转折,不幸由他当了莫高窟的家,把持着中国古代最灿烂的文化.他从外国冒险家手里接过极少的钱财,让他们把难以计数的敦煌文物一箱箱运走.今天,敦煌研究院的专家们只得一次次屈辱地从外国博物馆买取敦煌文献的微缩胶卷,叹息一声,走到放大机前,完全可以把愤怒的洪水向他倾泻.但是,他太卑微,太渺小,太愚昧,最大的倾泄也只是对牛弹琴,换得一个漠然的表情.让他这具无知的躯体全然肩负起这笔文化重债,连我们也会觉得无聊.这是一个巨大的民族悲剧.王道士只是这出悲剧中错步上前的小丑.一位年轻诗人写道,那天傍晚,当冒险家斯坦因装满箱子的						

Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
愤怒是洪水	愤怒是强烈的	5.0	5.0	5.0	5.0	4.0	4.8
他是小丑	他是令人愤怒的	5.0	1.0	4.0	2.0	3.0	2.8

Context						
石油成长记:你所不了解的“黑色黄金”.关于石油的成因,目前主要有两种假说:一种是无机成因论,认为石油是在基性岩浆中形成的.上世纪60年代,曾有前苏联科学家研究认为,生出石油所需要的压力与地幔层的压力相似,继而认为石油是在地幔里不断生成的,并通过断层或者“转移通道”,被挤压到地壳的浅层;另一种是更为主流的有机成因论,即认为石油像煤和天然气一样,是古代有机物通过漫长的压缩和加热过程后逐渐形成的.史前的海洋动物和藻类死后,尸体埋藏在海底,经过漫长的岁月的作用,最终形成石油.它与我们更为熟悉的煤一样,是不能再生却又不可或缺的宝贵财富.						

Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
石油是黄金	石油是昂贵的	5.0	5.0	5.0	5.0	4.0	4.75

Context						
她有美丽的面容,是在座不能相比的,她有一头乌黑的长发,迷人的双眼,她的眼睛像星星一样闪动.她笑的时候最美丽,那种美丽好似天上的仙女,她的身材很好,一颦一笑,婀娜多姿.她的双手是非常纤细,非常柔软.好久没有见到她了,我很想念她,我也深深的爱着她,直到永远,永不放手.						

Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
她是仙女	她是美丽的	5.0	5.0	5.0	5.0	5.0	5.0
眼睛是星星	眼睛是明亮的	5.0	5.0	5.0	5.0	4.0	4.8

Context							
我期待未来.窗外的风追逐阳光,撞击窗户发出温和的轻响.窗台上的风信子不知在期待什么,只是,迟迟,花未开.雏鹰想长成雄鹰,展翅高飞;树苗,渴望长成苍天大树.....而孩子们,则期待着长大,去探索外面的世界.每个人都有期待的时候,而每个人期待的也是不同的.小时候,我期待着长大;长大了,我期待着在我生命的舞台上发光发亮.期待是不会停止的,它会陪伴着你到生命结束的那一刻.....期待是一种补偿,像一双手轻轻的缝合了遗憾,期待是一种奋斗,它将用不眷顾坐享其成的人, 期待是一种欣慰,像一阵春风吹开了眉心的忧愁 ,期待是一种幸福它将引领你走出痛苦.....							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
期待是春风	期待是温和的	4.0	3.0	5.0	3.0	3.0	3.6

Context							
“邀来满天忘言的繁星”,台北进入了妩媚的时刻.一周期间每天一色的 101 大楼今夜蓝光黯然,广场上已经年味十足了,各种羊年的装饰和灯光都在喧嚣着:要过年了.选中了一个角度,无遮无拦, 仔细品味这个曾经拥有诸多元素排世界第一的高楼.抬头仰望,它光华青幽,暗蓝的夜空下,无言的巨人矗立穹顶之下 ,它的沉默仿佛是一股力量,给人一种没有张扬的内敛和张力.台北的第一天,就这样过去了.据说,台北明天有雨.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
楼是巨人	楼是高的	5.0	5.0	5.0	5.0	5.0	5.0

Context							
黄昏时候, 洪水如暴虐的猛兽 ,终于撕开了江堤,一个小小的村庄成了一片汪洋泽国.受灾的人们三三两两聚在堤上,凝望着水中家园.忽然,一个黑点正顺着波浪飘过来,定睛一看,是个蚁球.“蚁球?”人们十分不解.“蚁球这东西很灵性.”一位老者解释说,“以前发大水,我也见过一个,洪水来时,一窝蚂蚁迅速抱团,随波漂流.在只要一靠岸,或者碰上一个漂流物,蚂蚁就得救了.说话间蚁球已飘了过来,越来越近,看清了;一个小足球大的蚁球!黑乎乎的蚂蚁密匝匝的抱在一起,风起波涌,蚁球漂流,不断有小团蚂蚁被浪打开,像铁器上的油漆片剥离开去.人们看得惊心动魄.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
洪水是猛兽	洪水是猛烈的	5.0	5.0	5.0	4.0	5.0	4.8

Context							
我期待已久的春节终于到来了,一吃过晚饭,耳边就传来了阵阵响声.我到窗边一看,只见随着雷鸣般的声音,一道亮光飞向天空, 烟火瞬间绽放成诗 ,一个又一个的烟花飞向天空,展示出不同的魅力.有的像火球、有的像金蛇,还有的像一朵朵盛开的金菊花,美妙极了!很快,这些美丽的烟花慢慢地变成了一阵流星雨洒落人间.这梦幻般的夜空再配上五颜六色的烟花,真让人过目难忘啊!							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
烟火是诗	烟火是唯美的	5.0	5.0	5.0	5.0	4.0	4.8

Context							
中秋节晚上,我观察了月亮.月亮圆溜溜的,因为八月十五月儿圆,月亮发出淡淡的光明,给人十分安静.月亮帮我们照亮了道路.月亮外表有点淡淡的黄色,给人感觉非常凉.我发现不管我走到哪儿,月亮总紧跟着在我身旁.月亮上有一点一点的黑点,是不是嫦娥在月亮上和我们一起庆祝中秋呢? 月亮圆圆的,好像黄色的月饼挂在天上.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
月亮是月饼	月亮是圆的	5.0	5.0	5.0	4.0	4.0	4.6

Context							
寂寞有时是一种异常美的境界.若从最远处窥望湛蓝的深湖,似乎会让人感到是很难进入的.一切景物,在翩翩舞动天鹅的翅膀下浮游,残雪斑斑,落在天鹅们的身上,也落在 长诗短歌般的山水之中.天鹅瞄着湖泽,优雅雅地舒展公主似的形影,感到不寻常的惬意.偶间,有庄重的王子腹收羽毛 ,将背骨挺得笔直,向公主显现英俊洒脱.它们同临一泓湖水,有时心怀幽情,恪守行规,有时也会意会神,雌雄彼此调护.甜柔、富有人情味,一种神秘情绪的陶醉,让人悟得高洁和温和、妍丽和尊严以及雍容和自在的妙处.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
天鹅是公主	天鹅是优雅的	5.0	5.0	5.0	5.0	5.0	5.0
天鹅是王子	天鹅是优雅的	5.0	5.0	5.0	5.0	5.0	5.0
山水是长诗短歌	山水是壮美的	5.0	5.0	5.0	4.0	5.0	4.8

Context							
夏天,夜色已笼罩着大地,一切都安静了,透着种种神秘的气息.云团缓缓地移动着, 被吞没多时的满月一下子就跳了出来,像一个刚出炼炉的银盘 ,辉煌灿烂,银光闪耀,把整个大地都照得亮堂堂的.荷叶上的青蛙,草丛里的蚂蚱和树枝上的小鸟,都被这突然降临的光明惊醒,欢呼、跳跃、高声鸣唱起来.几朵银灰色的、薄薄的云绕在它的身旁.							
Metaphors	Results	Acceptability					
		1	2	3	4	5	Average
月亮是银盘	月亮是银色的	5.0	5.0	5.0	4.0	4.0	4.6

Context

我听得入迷,一转头我发现**月亮已经升得很高了,它就像一块美丽的钻石挂在天空**,再看深蓝色的天空,已经成了玉皇大帝的天池,真是美呀!皎洁的月光照着我们,我和爸爸仿佛穿着洁白的圣衣。月亮这么美,怪不得嫦娥要奔月呢!

Metaphors	Results	Acceptability					Average
		1	2	3	4	5	
月亮是钻石	月亮是明亮的	5.0	5.0	5.0	5.0	4.0	4.8

Context

知识是珍宝,而实践才是获取它的钥匙。获得知识固然重要,但最重要的还在于运用。有了上面学到的一系列超级阅读技巧,你现在就能进行本章所要求的阅读。

Metaphors	Results	Acceptability					Average
		1	2	3	4	5	
知识是珍宝	知识是宝贵的	5.0	5.0	5.0	5.0	5.0	5.0
实践是钥匙	实践是重要的	5.0	5.0	5.0	4.0	5.0	4.8

Context

我又一次走向云梯,我一把抓住了第一梯...到最后一梯时,我又开始紧张,但想到妈妈的鼓励时,我就使出了吃奶的力气,稳稳地抓住了最后一梯。当我下来时,妈妈一下子抱住了我,说:“太棒了!儿子!你成功了!”我的脸上露出了高兴的笑容。妈妈一边欢呼着,一边为我鼓掌。**妈妈的爱如糖果一般,甜在我的心里。**

Metaphors	Results	Acceptability					Average
		1	2	3	4	5	
爱是糖果	爱是甜的	5.0	1.0	5.0	4.0	5.0	3.8

Context

那里有许多令人难忘的美景,但我似乎只对那长廊前的湖水情有独钟。秋天,蓝蓝的天空映照在湖面上,**湖水闪着晶莹的光芒,是美神掉落的泪珠么?**湖边的树上果实累累,石榴裂开了嘴,大大的苹果把树压弯了腰,一串串葡萄,似美女头上的挂饰,随风飘舞,多么美丽的图画啊!

Metaphors	Results	Acceptability					Average
		1	2	3	4	5	
湖水是泪珠	湖水是透明的	5.0	5.0	3.0	1.0	4.0	3.6

Context

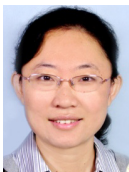
所有投资人都有一个可怕的敌人,而且太容易低估这个敌人,这个敌人就是通货膨胀。对散户来说,这个敌人特别危险。通货膨胀才是个人投资者的大问题,日常或波动性的股价改变虽然惊心动魄,最让投资人担心,却不是个人投资者最大的忧患。通货膨胀的侵蚀力量才真的令人害怕,以5%的通货膨胀率来说,你的钞票的购买力在不到15年内,就会少掉一半,在随后的5年内,又再少掉一半。通货膨胀如果是7%,只要经过21年,也就是从61岁“提早”退休,到82岁为止,你的钞票购买力就会降到只有目前的1/4。然而82岁却是日渐常见的平均寿命。这点显然是严重的问题,特别是个人已经退休,没有办法再增加资本,抵消通货膨胀对购买力可怕的侵蚀。

Metaphors	Results	Acceptability					Average
		1	2	3	4	5	
通货膨胀是敌人	通货膨胀是危险的	5.0	5.0	5.0	4.0	5.0	4.8

Context

在西班牙马洛卡,**蜿蜒曲折的高速公路长蛇一般扭动出不可思议的弧线盘绕在狭窄的山崖之间**,形成了史诗一般气势磅礴的360度弯道,这就是被称为Calobra之节的Calobra大道。建于岩石地貌上盘折的沥青混沙石路面且急且陡。由于全路的建设目标是为了尽力避免使用隧道,全路段平均下来也有7%的坡度,如果这也吓不倒身为红牛作死队的你,我们的建议是:骑自行车。

Metaphors	Results	Acceptability					Average
		1	2	3	4	5	
高速公路是蛇	高速公路是蜿蜒的	5.0	5.0	5.0	5.0	4.0	4.8



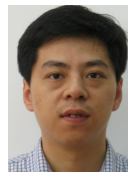
苏畅(1974—),女,福建厦门人,博士,副教授,主要研究领域为自然语言理解,隐喻计算,情感计算。



黄舒曼(1991—),女,硕士,主要研究领域为隐喻计算,自然语言处理。



王晓梅(1994—),女,硕士,主要研究领域为隐喻计算,自然语言处理。



陈怡疆(1972—),男,博士,副教授,主要研究领域为自然语言处理。