

学校编码: 10384
学 号: 31520171153195

分类号_____密级_____
UDC_____

厦門大學

硕 士 学 位 论 文

基于联想语义网的汉语隐喻理解研究

Chinese Metaphor Interpretation Research based on
Associative Semantic Web

指导教师姓名:

专 业 名 称: 计 算 机 技 术

论文提交日期: 2020 年 月

论文答辩时间: 2020 年 月

学位授予日期: 2020 年 月

答辩委员会主席: _____
评 阅 人: _____

2020 年 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

本人声明该学位论文不存在剽窃、抄袭等学术不端行为,并愿意承担因学术不端行为所带来的一切后果和法律责任。

声明人 (签名):

指导教师(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的涉密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。涉密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

隐喻广泛存在于人们生活中，是人们认识世界、了解世界的一种思维方式。80 年代认知科学的发展，加深了人们对隐喻认知特性的研究。大量与隐喻理解任务相关的模型也相继问世。但由于隐喻理解自身的难点，相较于隐喻识别任务，隐喻理解的相关研究较少。因此，对隐喻理解任务特点的分析以及相应解决措施的研究就显得尤为重要。

本文从认知隐喻学和语言学理论出发，通过研究汉语隐喻的特点以及理解过程，对汉语隐喻理解的难点进行具体分析。我们认为文化性是汉语隐喻中最为突出也是亟需解决的问题。本文以文化语义学理论为基础，将联想作为连接文化与隐喻的桥梁，通过构建联想语义网来解决汉语理解过程中的文化难题。联想语义网分为形容词联想语义网和动词联想语义网两个部分进行构建，使用同义词词林作为词汇扩展的核心部分。形容词网引入抽象度来描绘词语的联想过程，动词网则创新地引入动词图式来刻画动词的逻辑结构和语义信息。最后，通过 Neo4j 图形数据库进行可视化呈现。

从相似性出发，我们将汉语隐喻理解过程中的相似分为特征相似和关系相似，并依此对隐喻句进行划分。本文分别构造基于形容词网的相似度算法以及动词网的相似度算法来实现不同类型的汉语隐喻句的自动理解。对于自动理解过程中由于信息缺失问题导致难以获取理解结果的隐喻句，我们通过认知缺省推理出可信的理解结果。

综上所述，本文根据隐喻相似性的特点，对汉语隐喻句进行分类，同时利用构建的联想语义网对不同类型的汉语隐喻进行分类处理和自动理解。最后，针对缺失信息的汉语隐喻句，使用认知缺省逻辑推理出理解结果。实验结果证实了本文方法的有效性。

关键词：联想语义网；认知缺省推理；汉语隐喻理解

Abstract

Metaphor has been widely existed in people's life, and has become a way for people to think and understand the world. The development of cognitive science in the 1980s has deepened the research on the cognitive characteristics of metaphor. A large number of models related to the task of metaphor interpretation have also emerged. However, in recent years, due to the difficulties of metaphor interpretation, the relevant research on metaphor interpretation is less than that on metaphor detection. Therefore, it is very important to analyze the task characteristics of metaphor interpretation and study the corresponding solutions.

Based on the theories of cognitive metaphor and linguistics, this paper makes an in-depth study of the characteristics and understanding process of Chinese metaphor, and makes a concrete analysis according to the difficulties of Chinese metaphor interpretation. First of all, we think that culture is the most prominent and urgent problem in Chinese metaphor. Based on the theory of cultural semantics, this paper takes association as a bridge between culture and metaphor, and constructs association semantic network to solve the cultural problems in the process of Chinese metaphor interpretation. The associative semantic network is divided into two parts: adjective semantic web and verb semantic web. Synonym forest is used as the core part of vocabulary expansion. Abstract degree is introduced into adjective association semantic network to describe the association process of words, while verb association semantic network innovatively introduces verb schema to describe the logical structure and semantic information of verbs. Finally, the visual presentation is carried out through neo4j graphic database.

From the perspective of similarity, we divide the similarity in the process of Chinese metaphor interpretation into feature similarity and relation similarity, and then divide the metaphorical sentences. The similarity algorithm based on adjective net and verb net is used to realize the automatic understanding of different types of Chinese metaphorical sentences. In the process of automatic understanding, we use cognitive

default logic to infer credible understanding results for metaphorical sentences which are difficult to obtain understanding results due to the lack of information.

To sum up, this paper classifies Chinese metaphorical sentences according to the characteristics of metaphorical similarity. At the same time, the association semantic network is used to classify and automatically understand different types of Chinese metaphors. Finally, for the Chinese metaphorical sentences without information, we use cognitive default logic to infer the understanding results. The experimental results show that this method is effective.

Keywords: Association Semantic Web, Cognitive Default Reasoning, Chinese Metaphor Interpretation

目录

摘要.....	I
Abstract.....	III
第一章 绪论	1
1.1 问题的提出.....	1
1.2 隐喻基本理论	2
1.3 隐喻计算研究对象	3
1.3.1 隐喻识别	3
1.3.2 隐喻理解	4
1.4 本文的主要工作	4
1.5 本文的组织结构	5
第二章 隐喻理解相关研究	6
2.1 隐喻理解计算国外研究	6
2.1.1 基于推理的隐喻理解计算研究	6
2.1.2 基于统计的隐喻理解计算研究	7
2.1.3 基于语义的隐喻理解计算研究	7
2.2 隐喻理解计算中文研究进展	8
2.3 应用于隐喻理解计算的语义知识库研究	9
第三章 联想语义网的构建	12
3.1 联想语义网	12
3.1.1 联想与隐喻理解	12
3.1.2 语义网与隐喻理解	13
3.1.3 联想语义网	13
3.2 形容词联想语义网的构建	14
3.2.1 基本构建思路	14
3.2.2 词汇语料库的选择	14
3.2.3 形容词的扩展过程	15

3.2.4 形容词的扩展的剪枝	16
3.2.5 词语情感极性的引入	18
3.2.6 形容词的联想过程	19
3.3 动词联想语义网的构建	20
3.3.1 基本构建思路	20
3.3.2 动词扩展以及剪枝过程	21
3.3.3 论元获取	21
3.3.4 动词图式的引入	23
3.4 联想语义网的展示	26
3.5 联想语义网的评估	27
3.5.1 构建规模	27
3.5.2 评价方法	28
3.5.3 结果分析	28
第四章 基于联想语义网的汉语隐喻理解研究	30
4.1 汉语隐喻理解分析	30
4.1.1 汉语隐喻理解的过程	30
4.1.2 汉语隐喻理解的特点阐述	30
4.1.3 影响汉语隐喻理解计算的主要因素	31
4.2 汉语隐喻理解例句分析	32
4.2.1 汉语隐喻理解例句分类	32
4.2.2 汉语隐喻理解例句分析	33
4.3 基于联想语义网的汉语隐喻理解	34
4.3.1 特征相似性汉语隐喻算法流程	34
4.3.2 名词属性集合构建	35
4.3.3 基于形容词语义相似和情感的排序	36
4.3.4 关系相似性汉语隐喻算法流程	38
4.3.5 动作候选集合获取	39
4.3.6 基于动词语义相似的排序	40
4.4 认知缺省推理	42

4.4.1 认知缺省推理规则	43
4.4.2 基于认知缺省推理的汉语隐喻理解	45
4.5 实验与评估	50
4.5.1 测试集	50
4.5.2 评估方法	51
4.5.3 结果分析	51
第五章 总结与展望	54
5.1 本文的主要贡献	54
5.2 未来的研究方向	55
参考文献	58
附录 A 特征相似型汉语隐喻理解结果	62
附录 B 关系相似型汉语隐喻理解结果	65
附录 C 形容词联想网代表性部分	68
附录 D 动词联想网代表性部分	69
附录 E 词语 Cogbank 属性库与自建属性库对比	70
附录 F 部分词语获取的动作集合	71

CONTENTS

Abstract.....	III
Chapter1 Introduction.....	1
1.1 Background	1
1.2 Basic Theories of Metaphor	2
1.3 Research on Metaphor Computing	3
1.3.1 Metaphor Detection	3
1.3.2 Metaphor Interpretation	4
1.4 Main work.....	4
1.5 Chapters Outline.....	5
Chapter2 Research on Metaphor Interpretation.....	6
2.1 Foreign Studies on Metaphor Interpretation	6
2.1.1 Systems Based on Reasoning.....	6
2.1.2 Systems Based on Statistic.....	7
2.1.3 Systems Based on Semantics	7
2.2 Chinese Metaphor Computing Studies	8
2.3 Research on Semantic Knowledge Base.....	9
Chapter3 The Construction of Associative Semantic Web	12
3.1 Associative Semantic Web	12
3.1.1 Associative and Metaphor Interpretation	12
3.1.2 Semantic Web and Metaphor Interpretation	13
3.1.3 Associative Semantic Web	13
3.2 The Construction of Adjective Semantic Network.....	14
3.2.1 Construction Ideas	14
3.2.2 Choice of Corpus	14
3.2.3 Expansion Process of Adjectives	15
3.2.4 Pruning.....	16

3.2.5 Combining Affection	18
3.2.6 Association Process of Adjectives	19
3.3 The Construction of Verb Semantic Network	20
3.3.1 Construction Ideas	20
3.3.2 Choice of Corpus and Expansion.....	21
3.3.3 Argument Acquisition	21
3.3.4 Combining Verbal Schema	23
3.4 Visual Display of Associative Semantic Web	26
3.5 The Evaluation of Associative Semantic Web.....	27
3.5.1 Scale.....	27
3.5.2 Evaluation Method.....	28
3.5.3 Results Analysis	28
Chapter4 Chinese Metaphor Understanding based on Associative	
Semantic Web	30
4.1 Analysis of Chinese Metaphor Interpretation	30
4.1.1 Process of Chinese Metaphor Interpretation.....	30
4.1.2 Characteristics of Chinese Metaphor Interpretation	30
4.1.3 Main Factors impacts Metaphor Interpretation.....	31
4.2 Analysis of Chinese Metaphor examples.....	32
4.2.1 Classification of Chinese Metaphor examples.....	32
4.2.2 Analysis of Chinese Metaphor examples	33
4.3 Chinese Metaphor Understanding based on Associative Semantic Web .	34
4.3.1 Algorithm Flow About Feature Similiarity	34
4.3.2 Construction of Noun Attribute Set	35
4.3.3 Ordering by Adjective Association Web and Affection	36
4.3.4 Algorithm Flow About Relational Similiarity	38
4.3.5 Construction of Verb Candidate Set.....	39
4.3.6 Ordering by Verb Association Web	40
4.4 Cognitive Default Reasoning.....	42

4.4.1 Cognitive Default Reasoning Rules.....	42
4.4.2 Combining Cognitive Default Reasoning.....	45
4.5 Experiment and evaluation	50
4.5.1 Test Set.....	50
4.5.2 Evaluation Method.....	50
4.5.3 Results Analysis	51
Chapter5 Summary and Prospect.....	54
5.1 Main contributions of this paper	54
5.2 Future research direction	55
Reference.....	58
Appendix A	62
Appendix B	65
Appendix C	68
Appendix D	69
Appendix E	70
Appendix F	71

第一章 绪论

1.1 问题的提出

人工智能的初衷是通过模拟人类的认知以及思维过程,使计算机达到更高层次的水准。自然语言作为人类认知、逻辑思维活动的载体,不仅是人类区别于其他生物的本质体现,更是人类数千年智慧的呈现。在人工智能的研究领域中,自然语言学科具有极其特殊的重要地位^[98]。

隐喻,是语言中的一种特殊修辞手法。它集认知与文化为一体,贯穿历史的长河并深深根植于人类生活中。从 2700 多年前诗歌鼻祖《诗经·樛木》^[1]“南有樛木,葛藟累之”,到 1400 年前盛唐时期李白《望庐山瀑布》中的名句^[2]“飞流直下三千尺,疑是银河落九天”,从 90 年代现代文学《红玫瑰与白玫瑰》中,张爱玲所说的“娶了红玫瑰,久了,红的变成了墙上的一抹蚊子血,而白的还是窗前明月光”^[3],到现今日常生活中的“太阳花”、“电脑”、“桌肚”等,隐喻生动且形象地还原了人类的生活、行为,也逐渐成为人们了解自我、认识世界的一种思维方式。从认知角度对隐喻进行研究可以揭开人类思维过程、认知规律的神秘面纱,改善自然语言理解任务的表现。

上个世纪中后期,认知科学的发展更是推动学者们开启了隐喻认知和隐喻计算研究的新时代。Lakoff 和 Johnson 提出的概念隐喻理论^[4],Fauconnier 提出的概念整合原理以及概念整合网络^[5]以及 Gentner 等人提出的结构映射理论^[6]等,都使得人们对隐喻的认识更加深入,也为隐喻计算任务提供更为扎实的理论基础。隐喻计算任务主要关注两个方面,第一是如何通过计算机去识别、区分隐喻性的表达,即隐喻的识别;第二是使计算机可以理解隐喻性表达下的含义,即隐喻理解。

早期的隐喻计算任务多聚集于隐喻理解方面,多以规则和统计为主,比如 Fass^[7]基于规则的隐喻理解模型 Met5、Martin^[8]以规则推导为主的 MIDAS 隐喻计算系统、Mason^[9]基于统计的隐喻分析模型 CorMet 等等。随着认知科学理论的提出,学者开始关注基于语义的隐喻计算模型^{[10][11][12][13]},研究重心开始往隐喻识别任务偏移。而后深度学习的兴起,使得基于神经网络的隐喻识别模型^{[14][15][16][17]}开

始大量涌现,准确率也因此大幅度地提升。但对隐喻理解任务的研究却急剧下滑。溯其根本,主要有三点,第一,早期的隐喻理解模型以人工规则为主,耗时耗力;第二,神经网络模型效果虽好,但缺乏可解释性;第三,隐喻理解过程自身的难点,比如,文化和语境因素难以解决。因此,本文在认知隐喻学理论的基础上,针对汉语隐喻理解过程中的文化以及相似性难点,分别提出对应的解决方法,实现汉语隐喻的自动理解。针对由于缺省信息导致无法理解的隐喻句,我们引入认知缺省逻辑推理出可信的结果。

1.2 隐喻基本理论

传统隐喻学理论认为隐喻是一种纯粹的修辞手法^[18],在使用的过程中也只起到普通的修饰作用。古希腊哲学家亚里士多德认为隐喻是事物间可以借喻的相似之处。因此他的观点被称为相似论,相似论强调隐喻中两个不同概念之间特征的比较。

随着认知科学的发展,Richards 和 Black 率先对传统的隐喻理论提出质疑。他们认为,隐喻不仅仅是一种修辞手法或者语言现象,其本质上是人与人思想间的交流、互动。因此提出了互动论^[19],认为隐喻是将两个不同的“表象”带入一种互动的关联之中,这种关联体现为一个短语或词,而隐喻的意义正是这两个“表象”进行互动得出的结果。

80 年代,Lakoff 和 Johnson 提出了经典的概念隐喻理论^[4],认为人类通过自身的概念系统去思考、生活以及做出相应的行为,而这个概念系统是建立在隐喻之上的,并驳斥了传统的隐喻观念将隐喻仅仅看作是语言修饰作用。Lakoff 认为,“隐喻存在于日常语言中,属于日常语言的一部分,但它不仅仅是日常语言的替代品,而是构筑我们日常语言的基石,是不可或缺的主要元素。语言上所见的隐喻只是它在进行跨范围映射运动中的表象^[4]。”并且提出隐喻是借用一个范畴的经验来认识另一个截然不同的新范畴,是由“源域”到“目标域”的映射,且隐喻机制是“跨域映射”,隐喻性表达则是实现跨域映射的语言。

同期,Gentner^[6]的结构映射理论提出结构映射的过程是类比迁移,通过匹配不同对象来寻找结构上的相似,然后通过图式归纳提取源问题中的关系用于解决目标域问题。由此引出,隐喻映射的过程实质上是源域和目标域关系结构的匹配

过程。隐喻的理解就不仅限于特征的匹配，而是对特征之间关系的匹配，包括两个步骤：排列和映射。排列是通过匹配源域和目标域之间的特征，找出相关性最大的排列系统。然后将相关特征的关系结构进行类比，之后将源域的结构映射到目标域。

90 年代，Fauconnier 和 Turner 根据概念隐喻理论进行扩展，提出了概念整合模型^[6]，并认为其中的两域模型应是更大的概念投射的一个部分。模型是由个人知识背景构成框架，通过模拟人类获取概念的过程形成的心理空间网络。模型包含四个概念空间：输入空间 1，输入空间 2，类属空间以及合成空间。首先获取需要理解的表象，通过两个输入空间向类属空间映射，然后对初步获取的信息进行加工完善，最后在合成空间呈现整合共同的组织结构。该模型动态地对概念进行整合，为隐喻理解提供了有力的帮助。

除此之外，Grady 的基本隐喻理论^[20]、Ortony 的特征显著性不平衡理论^[21]等都为隐喻计算任务夯实了理论基础。

1.3 隐喻计算研究对象

隐喻计算研究对象主要分隐喻识别和隐喻理解。

1.3.1 隐喻识别

隐喻识别是隐喻计算任务^[22]的第一步，它的目的就是识别出给定文本中隐喻性的表达。当句子中包含明显的比喻指示词时，如 like、as...as、is 等等（在汉语中，是“像、似、是”等）隐喻识别较为容易。但隐喻往往很少包含隐喻指示词，因此识别隐喻的难度大大增加。

从人工角度划分，隐喻识别可分为人工识别和计算机自动识别。在人工识别隐喻方面，Pragglejaz 团队提出的人工标注隐喻标准（MIP）^[23]已经被广泛应用^{[23][24]}。Stefanowitsch^[26]预选出可以作为源域的词类，再利用同义词典确定属于这些词类的词语，标注为隐喻词。Reining 等人^[27]则着眼于目标域，在属于目标域的词语周围设定窗口，并人工分析这些窗口中是否含有隐喻。人工标注隐喻是构建隐喻数据库的一种重要的方法。基于人工判别的方法费时费力，衡量的标准也难以统一。所以研究人员提出构建自动隐喻识别系统，使机器具备隐喻识别的

能力，来提升该任务的效率。目前大多学者通过结合神经网络模型来进行隐喻识别任务，效果也较之前有显著提升。

1.3.2 隐喻理解

隐喻理解所要解决的问题就是如何获取隐喻的真实含义。如果让计算机按照处理非隐喻性文本的方式来处理隐喻，则隐喻的矛盾语义和不确定性会对结果产生干扰。例如，隐喻句“护士是天使”，由于护士是一个职业，天使是不存在的一个虚拟事物，计算机无法将两者混为一谈。而人类往往会基于联想，集合它们关联的点：洁白、纯洁，将该句的意思理解为“护士是纯洁的”。隐喻理解就是将隐喻句改写成意思相同且符合正常语义的句子。目前隐喻理解任务大多使用基于语义的理解模型^{[12][13][36][39]}对隐喻进行解读，但也存在许多的难点等待攻克，比如，语境、文化背景等。

1.4 本文的主要工作

近几年对隐喻识别研究的研究逐渐引起注意，但隐喻理解任务受到的关注度较少。深度学习的兴起使得基于神经网络的隐喻识别模型的性能大大提升，但神经网络存在的可解释性问题以及无法实现文本到文本的模式，导致无法很好的运用到隐喻理解当中。而基于语义的隐喻理解工作在解决理解问题时仍有欠缺。本文主要研究汉语隐喻的理解过程，有针对性地解决汉语隐喻理解中的难点，最后得出汉语隐喻的理解结果。

本文的主要工作包括以下几方面：

1. 利用同义词词林中的相似关系，构建包含文化因素的联想语义知识库，并在此基础上构建形容词联想语义网和动词联想语义网。形容词网中引入抽象度算法形成形容词的联想过程，并赋予每个形容词情感极性。动词网引用 Lakoff^[73]图式理论来划分基本动词图式，手动对 3200 个动词进行分类，并归纳总结对应的图式逻辑结构。

2. 深入研究汉语隐喻理解过程中的相似性。根据 Gentner 结构映射理论^[6]，对汉语隐喻句进行分类，并针对特征相似和关系相似设计不同的算法，来实现汉语隐喻的自动理解。

3. 将认知缺省逻辑应用于缺省信息的汉语隐喻理解过程中。

本文的创新之处主要体现在：

第一，引入动词图式理论作为动词网的核心。并根据动词图式的类别以及定义，规定相应的动词逻辑结构来丰富动词的语义信息。第二，针对理解过程中由于信息缺失未能获取理解结果的汉语隐喻句，使用认知缺省逻辑推理出合理的理解结果。

1.5 本文的组织结构

本文主要研究的是汉语隐喻理解，详细阐述了与隐喻相关的基本认知理论，并介绍了隐喻理解任务在国内外的研究进展。本文在隐喻认知理论和前人工作的基础上，提出构建了联想语义知识库来应对汉语隐喻中文化难点，并以此为基础，设计了以词语抽象度为核心的形容词网以及以动词图式为核心的动词网。在隐喻理解方面，本文依据结构映射理论，对汉语隐喻理解过程中的相似性进行区分，并通过形容词网和动词网分别对特征相似型和关系相似型汉语隐喻句进行自动理解。最后，针对缺失信息的隐喻句，结合认知缺省逻辑推理得出汉语隐喻的理解结果。

论文组织结构和内容如下：

第一章论述了隐喻的研究背景以及重要性、隐喻相关的认知理论以及隐喻计算任务的研究对象。

第二章介绍了隐喻理解任务的研究进展，详细地对国内外隐喻理解的研究现状进行分析，并介绍了隐喻理解计算过程中可利用的知识库资源。

第三章论述了联想、语义网与隐喻的关联，以及构建联想语义网想法的由来，详细介绍了联想语义网的内容以及构建过程，并对该网进行可视化展示以及分析。

第四章分析了汉语隐喻的理解过程以及痛点，论述了汉语隐喻例句的相似性特点并进行分类。通过自主构建的联想语义网进行汉语隐喻自动理解工作，并引入认知缺省逻辑来处理缺失信息的汉语隐喻句。最后对实验得出的结果进行评价和分析。

第五章对本文工作进行总结，点明本文的贡献以及不足之处，并对未来工作进行展望。

第二章 隐喻理解相关研究

2.1 隐喻理解计算国外研究

2.1.1 基于推理的隐喻理解计算研究

Narayanan^[28]提出的 KARMA 系统将推理重心放在源域上，通过人工构建的知识库获取相应的知识并投射到目标域上。该系统擅长处理动词性隐喻。比如隐喻句“government is stumbling in its efforts (政府摇摇晃晃)”中，动词 stumble 作为隐喻的源域，表示绊倒，而“绊倒”在概念隐喻中就意味着失败。经过一系列的逻辑推理，系统得出“政府失败”这个理解结果。

Steninhart^[29]创新地将类比理论引入隐喻理解中，提出了隐喻的结构理论，并强调源域和目标域之间结构相似的重要性。依据该理论 Steninhart 构建了 NETMET^[28]系统。该系统通过人工构建的知识库获取隐喻句相关的命题，组成可能世界。然后通过类比映射构建源域和目标域之间的关联，最后通过推理得出隐喻理解的结果。但由于该系统采用的类比推理过于关注结构上的相似性，所以仅仅适用于结构性较明显的隐喻，对于一般隐喻的处理较为欠缺。

Ovchinnikova 等人^[30]构建了一个基于溯因推理^[31]多语言隐喻理解的系统。该系统首先对包含隐喻文本的片段进行语法分析，并生成逻辑表达形式。然后结合知识库输入到推理系统中获取可能性最高的推理结果，将其映射为概念隐喻。最后将得到的概念隐喻翻译为逻辑谓词，并转换为自然语言表达形式。系统在英语、俄语两种语言的测试集上进行了测试，取得了较好的效果。

基于这项工作，Gordon^[32]提出了一个高精度、多语言的溯因推理映射系统，目标是将诸如“战胜贫穷”等概念隐喻映射到<源域, 目标域>词对(比如<战争, 贫穷>的形式)。该系统能够识别 67 种源概念和 7 种目标概念，支持英语、西班牙语、俄语和波斯语，且在识别西班牙语目标域的效果最好，准确率达 0.99。该篇关注句中源域和目标域的跨度(即一个句子最小的摘录，可触发源域或目标域概念)和新知识库的创建，它使用解析器过滤只留下与跨度直接相关的文字内容，并生成逻辑形式^[30]。实验结果表明过滤模型显著提高了源域识别的精度。该推

理机制支持用已知知识来学习新概念的观点,这与认知隐喻理论相一致。该方法体现了基于逻辑的系统在模拟人类理解隐喻过程中的潜力:可以基于给定的知识和推理规则,推理并理解新知识。

2.1.2 基于统计的隐喻理解计算研究

Ortony^[21]提出隐喻中源域的突显特征和目标域的非突显特征是相关联的,而隐喻理解过程则是比较源域和目标域之间突显不平衡特征的过程。依据该理论,Terai^[33]构建了一个基于语料库统计分析的理解模型,主要关注“A is B”形式的隐喻。语料库数据从1993-2002年日本报纸《每日新闻》中获取,并通过语法分析工具提取出与形容词-名词修饰相关的数据。系统首先通过对数据的统计分析,计算概念与属性值之间的概率关系。然后将源域和目标域作为输入节点,属性值作为输出节点,源域和目标域属性值的概率作为输入权重来模拟隐喻理解动态交互的过程。最后通过心理实验,验证了该模型的有效性。

在该任务的基础上,Terai^[34]更深入地去探索隐喻理解过程,并假设隐喻理解分为两个步骤。首先是分类过程,根据Kayser^[35]将目标词分配给特定的类别;然后就是动态交互的过程,分配给特定类别后的目标会受到特征之间动态交互的影响。依据此假设,提出了两者结合的隐喻理解模型,并通过心理实验再次验证了该模型的有效性。

2.1.3 基于语义的隐喻理解计算研究

Shutova 提出一个无监督的隐喻理解系统^[36],可以将隐喻性文本改成意义相同的非隐喻性文本。系统通过训练语义相似向量和 UKWac 语料库^[37]得到的两个理解模型。然后根据隐喻词所在的上下文,计算出与该词原意契合度较高的候选释义,然后利用选择优先模型(Selectional preference)^[38]评价各个释义的字面含义。该系统主要关注主语-动词和动词-宾语类型的隐喻,对其他形式的隐喻未做测试,系统准确率达到0.52。

Bollegala 等人^[39]使用互联网资源构建的无监督隐喻理解模型与 Shutova^[36]有所相似。针对输入的单词对(由隐喻性动词和它所修饰的名词构成),该模型从网络上查找包含该词对的语句片段。然后从片段中提取词汇模式,以表示单词对

之间的依存语义关系,提取出来的模式按其代表性评分选出最优的模式。用选择偏好模型评价释义的字面含义,最后通过词汇替代方式识别正确的候选释义。该系统使用网络资源,能够发掘比人工编辑的资源库更丰富的改述方法,提供了更多可行的释义选择。Bollegala 等人^[39]和 Shutova^[36]的共同优势在于系统是完全无监督的,方便移植到其他 NLP 工作中进行隐喻处理。

Su 等人^[12]认为隐喻理解的结果应是源域与目标域的合作,而隐喻理解的重点就是提取它们的相似性。他们将隐喻理解分为两个步骤:属性提取和属性转移。系统从属性数据库₁和 Sardonicus₂中获取源域属性。然后,系统计算从目标域和源域各个属性之间的关联性。基于词向量模型使用余弦相似度来进行相关性的度量。系统选取最相关的属性作为隐喻的最佳释义。该系统分别在中文和英文隐喻集测试,精度分别为 0.87 和 0.85。

Su 等人^[13]还构建了一个基于潜在语义相似的中文隐喻理解系统,用以解释名词性隐喻和动词性隐喻。该篇将“潜在语义相似性”定义为“源属性与目标属性存在的相似之处”。他们认为,如果两个词语存在潜在语义相似的关系,那么它们在 WordNet 中存在语义关联。系统抽取隐喻文本中与源域和目标域相关的属性知识,根据 WordNet^[40]对二者所在的同义词集进行扩展,得到了它们的潜在语义相似关系,从而得出隐喻理解的结果。系统处理的是单句隐喻,考虑到在上下文缺失的情况下,一个隐喻可以由不止一种方式来解释。因此,系统同时结合谷歌距离^[41]对所有可能的释义进行排序。系统在名词和动词隐喻处理上分别达到了 0.85 和 0.86 的精度。

2.2 隐喻理解计算中文研究进展

张威等人^[42]从逻辑角度出发,赋予隐喻逻辑新的定义,并制定相应规则用于汉语隐喻理解计算任务。该篇主要关注名词性和动词性汉语隐喻,通过语料库和 HowNet^[43]获取与源域、目标域相关的词语(包括实体和关系)组成相应的池空间。然后比较两个池空间中的词语,若存在相交的词语,则认为该隐喻可以被理解。最后语义相关度最高的作为理解结果。实验结果表明该隐喻逻辑有较强的解释能

¹ 由厦门大学自然语言处理实验室开发

² <http://afflatus.ucd.ie/sardonicus/tree.jsp>

力，有利于汉语隐喻理解任务。

苏畅^[44]认为上下文信息对隐喻理解有着重要影响，因此利用上下文信息构建多层次的语义网络模型，并在此基础上，通过相关性约束的方法对名词、动词、形容词隐喻进行理解。多层语义网络模型分为五个部分，输入层（概念信息）、感知层、属性层、隐藏特征层以及特征层。输入层获取语篇信息，并进行预处理获取词群以及它们的词向量；感知层获取感官和情感信息；属性层则获取概念的属性特征；隐藏特征层获取全局特征，比如文化、主体等信息；特征层则为隐藏特征层的总体概括。实验证明该模型能有效地实现隐喻的自动理解。

贾玉祥^[45]认为明喻相较于隐喻更容易识别，且自身是带有标记的，可以为隐喻理解提供很好的信息。该篇首先从网页中获得大量明喻实例，并通过自定的语法规则自动构建汉语明喻知识库。然后利用该知识库抽取出隐喻源域和目标域相似的特征信息，应用于汉语隐喻理解和生成任务。实验结果表明，该模型在理解和生成任务都有较高的准确率，且可扩展性较好。

Mao^[46]提出了一种无监督学习模型在词语级别去识别和解释隐喻。首先，输入一个句子，并标注目标单词（文本中隐喻性有待确定的词以及相关语境词）。其次，使用 word2vec^[47]训练词汇，获取词汇的基本向量表示。同时通过 WordNet 获取该词汇的同义词以及上位词集合作为候选释义的集合。然后，确定候选释义集合中最符合的词语，即原文中最有可能解释原句的词语。最后，计算该词与目标词的余弦相似度，低于阈值则为隐喻意义，并将得到的候选释义替代源隐喻词，即得到隐喻理解结果。实验表明该模型对隐喻词的理解结果也有助于谷歌翻译和 Bing 翻译系统（英译中），准确率提高了 9% 以上。

2.3 应用于隐喻理解计算的语义知识库研究

WordNet^[40]在隐喻识别和隐喻理解中被广泛地用于捕捉词汇语义。它是由普林斯顿大学开发的一个大型的英语词汇库，从认知的角度将名词、动词、形容词和副词划分到表达不同概念的同义词集，同义词概念之间由语义和词法关系相互联接。它的体系简洁，层级关系明显，WordNet 中的单词距离也可以用来衡量单词的语义关联性，这种测量方法在隐喻识别任务中被广泛应用^[49]。

VerbNet^[49]是目前最大的英文在线动词词汇库。数据库中的每个动词类都用

题元角色描述、描述参数的限制条件以及由语法和语义信息构成的框架等来描述。它也含有向其他词汇资源的映射,比如 WordNet^[40]、Xtag^[50]等。VerbNet 和 WordNet 一样拥有明显的层级关系,并关注于动词方面,因此这个数据库在处理动词隐喻计算任务时特别有用^[53]。

FrameNet^[51]是基于框架语义学构建的英语词汇知识库,它认为绝大多数单词的意义可以通过语义框架呈现。知识库对拥有相同认知结构、相同类型的语义角色的词语进行归类,并用一个框架对这类词进行概述。同时它还会构建不同框架之间动态、静态的关联。FrameNet 语义关系丰富,可以为隐喻理解任务提供大量的语义信息。

HowNet^[43]主要关注于词语的概念,是一个以义原为基本单位构建的中英双语知识库。它将概念与概念之间的关系以及概念及其属性之间的关系形成了一个网状的常识系统。与 WordNet 不同,HowNet 突破了词汇屏障,将词语进一步细分,使用最小意义单位的义原来进行标注,并对获取的义原进行规整和确认。让我们可以更深入的去了解词汇背后的语义信息。

哈工大同义词词林^[52]是根据《同义词词林》^[54]扩展而来的汉语知识库。它按树状结构对词条进行收录,词条不但包含词语的相似词,也包含与其相关的词汇。词林将词语分为五层结构,级别越高,对词语的刻画越细,第五层往往只有一个词语,通常被称为原子词群。它在信息检索、文本分类等研究领域有一定的应用。

Cogbank^[80]是一个自主构建的汉语认知属性知识库。它通过指定隐喻句的句法结构“目标域+像+源域+一样+属性”,从网页中爬取了近百万条“源域词语-属性”对。经过人工校对和信息整理,获得二十多万条词语认知属性对,共包含近万个词语以及认知属性。Cogbank 认知属性库为认知语言学与隐喻的研究提供了便利。

第三章 联想语义网的构建

人类交流、理解语言的能力不是与生俱来的。婴儿自出生伊始并不能与他人进行正常的沟通交流,但随着对外界事物的认知学习,逐渐建立自己的语言系统,并拥有自己的心理词典之后,便可以无障碍地与他人进行认知交流。隐喻作为语言更为复杂的一个分支,更需要人类拥有相似的文化、知识信息,才能对其进行解读。所以,要想使计算机在一定程度上理解汉语隐喻,构建属于计算机的汉语心理词典(也就是语料库)就显得尤为重要。我们选用同义词词林作为原始语料,结合抽象度以及动词图式,对信息进行筛选和扩充,以构建联想语义知识库,并最终形成以词语为核心的知识网络。

3.1 联想语义网

3.1.1 联想与隐喻理解

联想具有语义连接的功能^[57]。生物学学家研究证明,语言信息加工的过程中,神经元起了关键作用。且生物学家们认为神经元之间相互连接,并进行激活过程的中介就是联想。它将相关的语义信息进行组合,形成神经团簇,使人们可以处理相应的外部信息^[58]。心理学家认为联想是由某个人、某件事或某个概念而引起的其他相关人或物或其他概念的一种思维活动,人处理信息的过程就是通过联想推理语义进行选择连接的过程。心理学^[59]还提出联想的两条基本定律:一是接近律,因概念在时空中同时发生或比较接近而被联想;二是相似律,因概念相似而被联想。语言学家^[55]认为语义也是具有联想性的,通过心理联想,外部输入的语言信息才能通过语义进行关联,并将联想分为三类,语音、语义和语境联想。

在隐喻理解中联想也起着重要的连接作用。当人们理解隐喻句“律师是鲨鱼”时,律师和鲨鱼两个词语的信息必先经过大脑的加工,通过自身心理词典中信息的匹配,鲨鱼是凶恶的,进一步联想到律师的表现,得出律师是凶恶的结果。从外部获取信息到内部信息匹配加工的过程通过联想进行,因此联想也是隐喻理解过程的主要机制。心理学的相似律也被广泛应用在隐喻理解模型中,通过找寻概念的相似之处得出理解的结果。

3.1.2 语义网与隐喻理解

互联网的诞生使得科技和生活快速发展。它可以提供我们海量信息，为获取知识提供了便利。但是网页并不能像人一样可以轻松地理解数据表达的含义，因此大量繁琐的筛选信息过程都需要人类的参与。如今面对网页成几何倍数地增长，如何快速准确地获取有价值信息成为人们思考的重中之重。假如计算机能读懂网页中的知识，信息获取将会更加智能和便捷，世界也将会发生更为深刻的变革，于是一个新的概念“语义网”诞生了。万维网的发明人 Tim Berners-Lee 于 1999 年在他的书《Weaving the Web》中提到，“如果说 HTML(超文本语言)和 web(网页)将整个在线文档变成一本巨大的书，那么 RDF(资源描述框架)，inference languages(推理语言)以及 schema(数据库)，就可以容纳世界上所有的数据信息，且变成一个巨大的数据库供人们使用^[60]。”因此，语义网是一个以数据为核心的，可以使计算机理解并处理的网络。与语义网相同，隐喻理解的目的也是让计算机可以理解其字面含义。隐喻理解需要海量的知识作为基础，而语义网中大量的数据正好可以提供其所需要的语义信息。

3.1.3 联想语义网

联想是隐喻理解过程中的重要机制^[57]，语义网又是隐喻理解的有力工具，因此我们认为可以通过将联想与语义网进行结合，来对隐喻进行更高效的解读。通过 3.1.1 节中，心理学家以及语言学家^{[55][59]}对联想过程的解读，我们发现在该过程中，词语往往更容易联想到与之语义相近的词。本论文通过同义词词林对词语的相似关系进行抽取，并以相似关系作为词语扩展过程的基础。

Su 等人^[61]认为人类在认知过程中，通过对事物的涉身描述来了解该事物，且这种涉身描述一般都是形容词。靳光瑾^[62]中指出动词是语言中最重要的存在，它处于句子结构和语义的中心。因此，将联想语义网分成形容词和动词两个部分进行构建，通过研究形容词来体现对事物的认知过程，研究动词来更深入地分析语义信息。

3.2 形容词联想语义网的构建

3.2.1 基本构建思路

Su 等人^[61]以形容词为中心，以人类五感为基础，通过抽取 WordNet^[40]中的同义词和上位词集合，构建了基于形容词的涉身知识网络，并将该形容词网应用于隐喻理解任务中。该篇创新地提出形容词是人们认知事物的基础，并证明了人类认知的过程是由具体到抽象的，但也存在一定的问题，比如，该篇形容词网是通过 WordNet 中同义词构成的，在英语隐喻的理解应用是可行的，若要在汉语中使用，需要进行翻译得出理解结果，而翻译过程中会丢失相应的文化特征。针对该问题，本文重新构建了一个可应用于汉语隐喻理解的形容词网，并可以同时囊括文化语义信息。

首先，我们选用现有语料库，筛选出常用形容词组成语料库的种子集。其次，我们认为情感在词汇联想过程中起着一定引导的作用，所以本文对词汇本体以及其扩展词都进行了情感标注。最后，我们以“具体-抽象”为主线构建形容词网的基本结构。

表 3.1：构建形容词联想语义网的过程

算法：形容词联想语义网的构建过程
<ol style="list-style-type: none"> 1. 通过现有词典和文本中获取常用形容词。 2. 使用哈工大版同义词词林对形容词进行扩展。 3. 对原有的形容词和扩展词进行情感标注。 4. 根据具体到抽象这条主线，使用抽象度计算的方法引导形容词网的走势。

3.2.2 词汇语料库的选择

形容词由北大汉语语法信息词典^[65]中的 1550 个常用形容词以及由文学文本中获取的 1000 多个形容词组合而成，筛除重复词汇之后一共得到 1850 个常用形容词。

构建一个既精准又包含大数据量的语料库往往需要较大的人力和时间，

WordNet、HowNet 等大型语料库都耗费几年甚至十几年的时间。短时间内构建一个高质量、大数据的语料库是不现实的。因此本文选取现有的资源，并对其进行加工利用。通过对比现有语料库，本文选取了哈工大版同义词词林。同义词词林中有着丰富的汉语语义信息，可以服务于词语的扩展过程。哈工大同义词词林^[52]更是在同义词词林的基础上进行扩展。首先，它将词汇的分为大、中、小三类，大类用大写字母表示，共 12 类；中类用小写字母表示，是对大类词语的细分，共 97 类；小类用数字表示，更细致的划分了中类的词语，共 1400 类。然后将小类段落作为下一层的分类，段落中的行作为最后一层的分类。综上，同义词词林一共被划分成五层结构，大、中、小分别对应前三层，层级越高，对词义的刻画越细致。最后，第五层的每个词条根据“相同”、“同类”、“独立”按“=”、“#”、“@”标记。

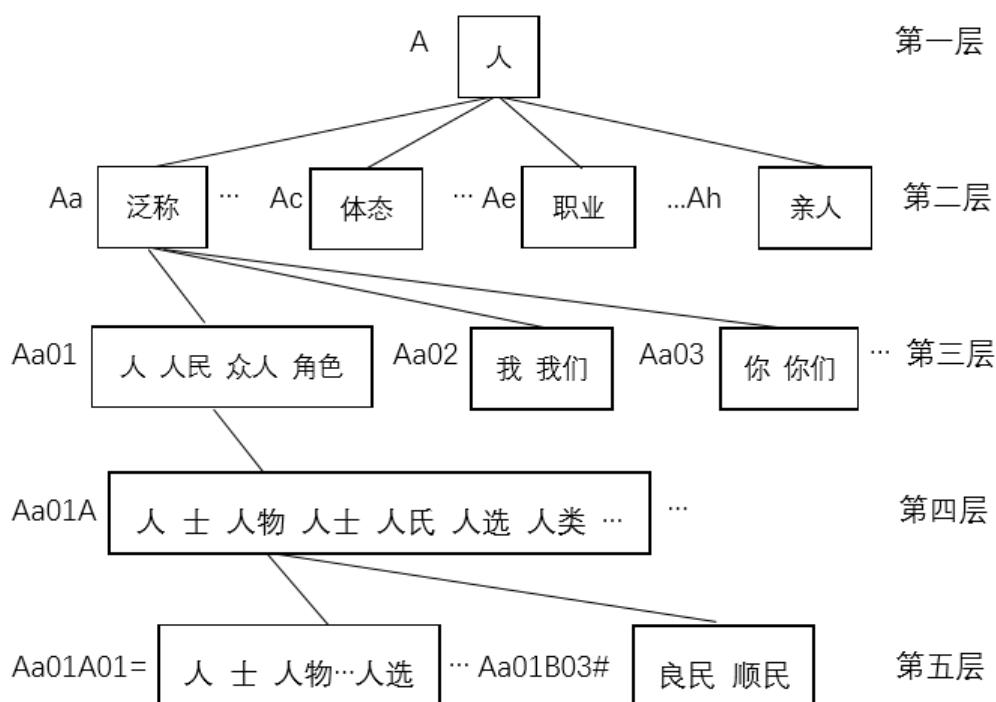


图 3.1: 哈工大同义词词林结构图

3.2.3 形容词的扩展过程

本文利用哈工大同义词词林中的第五层原子词群抽取出词语的相似关系来模拟词汇的扩展过程。首先，我们从同义词词林中获取词条编码为“=”的目标

词条，作为目标词的相似信息。其次是词语的扩展。第一步获取到的词条中除目标词以外的其他词语作为词语扩展过程中的第一层，且各个词条是相互独立的。然后遍历第一层的词语，若可以在词林中找到相似关系，则作为可扩展的下一层词语。根据“六度空间理论”，任意两个人之间的间隔不会超过五个人，我们认为词语也一样，因此通过该步骤一直扩展至第六层。下图为“矮”的部分扩展过程。

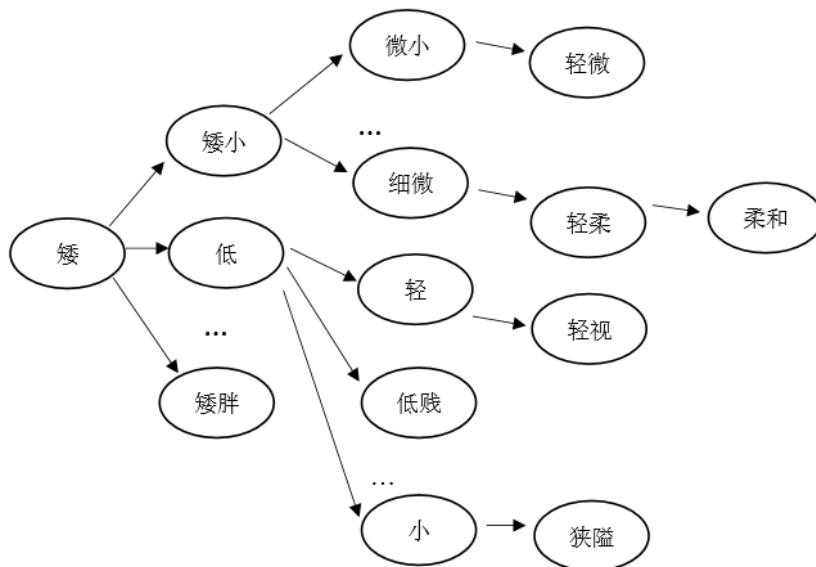


图 3.2: 形容词“矮”的部分扩展过程

3.2.4 形容词的扩展的剪枝

由形容词的扩展过程，我们可以观察到，对一个词扩展的层级越高，获取的信息就越多（当扩展至第三层时就能达到上千的词汇，至第六层词汇的个数可达上万），同时也伴随着一些噪音的出现。比如，形容词“矮”扩展的过程中，会出现“矮-低-轻-轻柔-柔和”的过程，而矮与柔和的关联性并不大，且影响后期理解模型的速度。因此在扩展至第三层的时候，就要终止该过程。所以我们在词语扩展过程中需要对获取到的信息进行筛选，以提升效率。但由于获取的数据量较大，人工的方法显然是不可行的，我们需要自动化方法将不相关的词汇进行删减。

本文将该问题看作是一个二分类问题，来判断词语是否相关。使用两种相似度算法（词向量相似度以及同义词词林相似度算法）得出词语相关的概率，将两

个概率作为相似特征，得出相关的综合概率。词向量选用的是^[66]训练好的中文词向量（包含了百度百科、人民日报、维基百科、文学作品等数十领域的语料）。而同义词词林中本身就包含词语的编码，因此可以计算编码相关度来得出扩展的词汇与目标词汇的相关值。

基于逻辑斯蒂的二分类具有简单易用、占用内存小，且可解释性较好的特点。本文训练逻辑斯蒂模型，将词向量相似得分、同义词词林相似得分作为输入的两个特征，来预测词语的相关的概率值。

设 P 为词语相关性的概率值，则条件概率分布如下：

$$P(Y = 0|R) = 1 / (1 + e^{(b + \sum_{i=1}^n (w_i R_i))}) \quad (3-1)$$

$$P(Y = 1|R) = 1 / (1 + e^{-(b + \sum_{i=1}^n (w_i R_i))}) \quad (3-2)$$

我们从获取的词中随机取出 2500 个词按 1: 9 划分测试集和训练集，并对抽取的词语进行人工标注。下面简要叙述标注过程，若抽取的词汇与目标词不相关，则赋值 0，其他存在相关性（相似关系或相关关系）则赋值 1。比如，“矮-矮小-小-暂时”的扩展过程中，前三个是相关的，标为 1，后面“矮”与“暂时”没有关联，则标为 0。输入两个相关变量，使用逻辑斯蒂回归算法得出相关概率，通过人工设定阈值，使得对数损失函数最小化。最后得出阈值在 0.33 以上为存在相关的词汇。

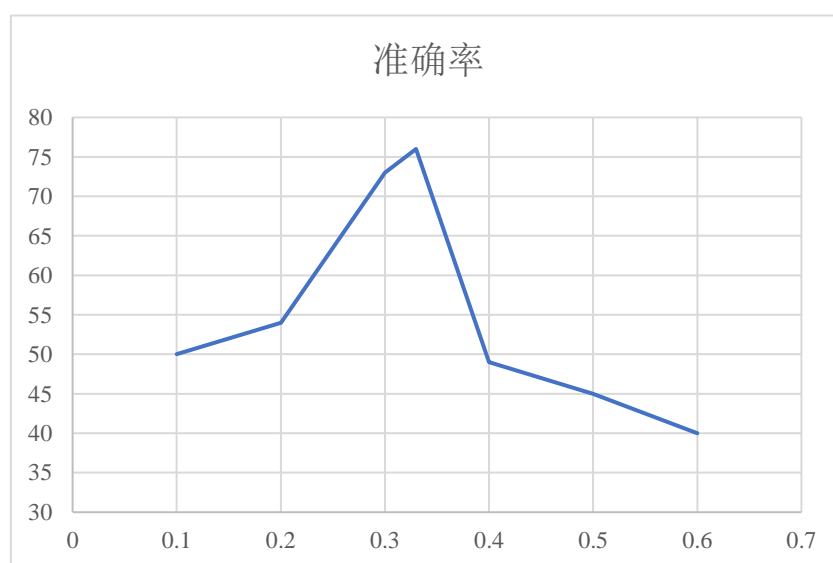


图 3.3: 形容词相关性阈值

3.2.5 词语情感极性的引入

隐喻能够传递较为强烈的情感。为了使形容词网可以更好地应用于隐喻计算任务，我们在它的构建过程中引入了情感元素。人工情感标注带有一定的个人主观倾向，且在词汇量较大的前提下，一一标注耗时过久。考虑到本文形容词以及扩展词数量较大，不适合使用人工标注。所以本文采用情感词典先对存在于词典中的形容词进行情感极性的标记。情感字典选用的是大连理工情感词汇本体库^[67]（后面简称情感本体库），包含一个词的词性种类、词义数、情感分类、极性等。情感分类有 21 种，包含快乐（PA）、安心（PE）、尊敬（PD）等，情感极性分为四种，0 代表中性，1 代表积极，2 代表消极，3 代表两者都有。本文只考虑积极和消极的情感，因此获得了 1019 个已标记情感信息的形容词。但仍由 800 多个词汇的情感信息未获取。

因此，本文使用 SnowNLP 工具^[68]来获取形容词的情感。SnowNLP 是一个中文自然语言处理的 python 库，在处理中文情感分析任务中有一定的优势。它自带了中文正负情感的训练集，通过输入目标词语或句子，使用贝叶斯模型计算该词或句子正向的概率。

本文输入目标词汇 w ，计算公式如下：

$$P(c_1|w) = \frac{P(w|c_1) * P(c_1)}{(P(w|c_1) * P(c_1) + P(w|c_2) * P(c_2))} \quad (3-3)$$

c_1 表示积极， c_2 表示消极。 $P(w|c_1)$ 表示在积极情况下，词语 w 的概率。 $P(c_1)$ 表示积极的概率。通过计算条件概率，得出情感正向的评分。

获取情感词典已标记情感词语的评分，并通过原标记信息进行积极、消极类别的划分。当阈值为 0.5 的时，准确率最高。最后，计算出剩下的 800 个词的情感得分，大于等于 0.5 时为积极。

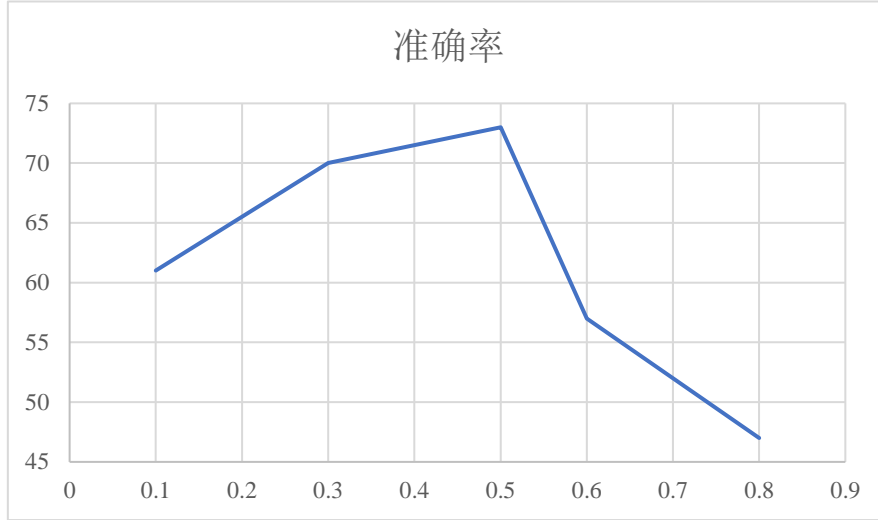


图 3.4: 情感极性阈值

3.2.6 形容词的联想过程

Su 等人^[61]提出并证明对事物认知的过程，往往是由具体到抽象的。马清华^[55]也曾提出词汇的联想过程是由具体往抽象发展的。所以，本文我们认为形容词的联想过程不应该仅仅是相似关系的扩展，更应遵循“具体-抽象”这条主线。因此，我们使用 Wang^[64]的方法来计算每个词的抽象度，使得形容词的扩展过程满足由具体到抽象发展的过程，以此还原形容词的联想过程。

设抽象词集为 ab_i ，具体词集为 con_i ，则目标词 m 的抽象值 P 为：

$$P = norm \left(\sum_{i=0}^n sim(m, ab_i) - \sum_{i=0}^n sim(m, con_i) \right) \quad (3-4)$$

抽象词集和具体词集是 Wang^[64]通过人工标注的各 20 个抽象概念和具体概念的集合。 sim 是通过计算目标词汇与抽象词集和具体词集的词向量相似度的值，来判断该词是更具体还是更抽象。如果词语更抽象，则词语与抽象词集的平均相似值就越大，与具体词集的平均相似值越小。 $Norm$ 表示是对得到的数据进行归一化处理。最终计算的得分越靠近-1 则表示更具体，越接近 1 则表示更抽象。

在得到扩展词汇的抽象度之后，我们开始对扩展过程加工。第一步，比较目标词和扩展词的抽象度，将小于本身的作为形容词的前面一层，大的作为后一层。然后，将形容词作为节点，每个词条选出一个代表性的词按抽象度作为前驱节点或后续节点，词条中其他的词语作为该节点的分支，形成形容词的联想过程。

3.3 动词联想语义网的构建

3.3.1 基本构建思路

动词网构建的目的也是为了服务隐喻计算任务的。目前，较少有人涉及中文动词性隐喻理解计算任务。主要原因是缺少对应的动词知识库使得该任务难以下手。因此建立一个汉语动词知识库是极为重要的一项任务。

VerbNet^[49]是由科罗拉多大学 Martha Palmer 等学者手工构建而成的至今为止最大的在线英语动词词典。它拥有丰富语义信息，包含了谓词的论元结构、选择限制、题元角色等，并且可以非常便捷地链接到其他的词汇资源中，比如 WordNet^[48]、XTAG^[50]、FrameNet^[51]、PropBank^[69]等。

参考 VerbNet，本文构建的动词知识库需要包括论元结构、选择限制这些丰富的语义信息。在此基础上，我们还会增加动词图式作为我们的创新点。由于纯手工构建耗时较多，我们采用了半自动化的方式构建动词知识库，来减轻一定的工作量。同样我们通过同义词词林获取动词的相似关系。其次，我们通过读者等语料资源获取动词的论元信息，并按同义词词林给出的词语类别对论元进行归类。然后，按照百度词典获取的动词释义，对 3200 个动词进行人工动词图式划分，根据论元信息形成其逻辑结构。最后生成以动词为核心的知识网络。

表 3.2: 构建动词网的主要过程

动词网构建方式
1. 自动获取常用动词种子集作为网络第一层。
2. 使用同义词词林对动词进行同义扩展。
3. 使用读者等语料资源抽取动词相关论元信息，按按同义词词林词语第一层的类别对论元进行归类。
4. 使用百度词典抽取动词释义，并划分动词图式，通过获取的论元信息组成动词逻辑结构。
5. 生成以动词为核心的知识网络。

3.3.2 动词扩展以及剪枝过程

动词由北大汉语语法信息词典^[66]中的常用动词以及由文学文本中抽取的动词组合而成，筛除重复词汇之后一共得到 3300 个常用动词。

同样使用哈工大版的同义词词林，获取目标动词的同义词。该过程与获取形容词相似关系一致。首先通过编码获取目标动词在同义词词林中相似的词条，作为动词扩展的第一层，且各个词条相互独立。然后遍历第一层词条中的词语，获取不同于第一层扩展词汇的同义信息，扩展为第二层，然后依次扩展至第六层。

扩展过程中同样存在的不相关的信息需要剔除。随机抽取 2500 个动词与该动词的目标词对其进行标注，按 1: 9 划分测试集和训练集。标注方法与形容词一致。我们仍使用同义词词林相似度和词向量相似度计算方法作为判断的标准，并通过逻辑斯蒂回归算法得出词语相关的概率，人工调整阈值使得对数损失函数最小化。最后得出阈值在 0.35 以上为相关词汇。

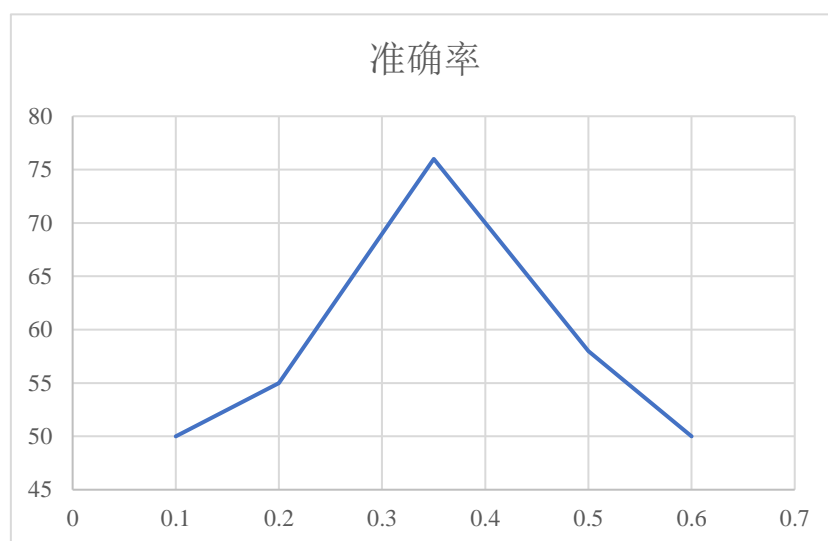


图 3.5: 动词相关性阈值

3.3.3 论元获取

动词的联想过程往往和与它衔接的名词有很大的关联。比如，当我们看到动词“吃”时，我们往往会联想到食物；当我们看到动词“跳水”时，往往会联想到“运动员”、“跳台”等。因此，除了获取相似关系的动词外，我们希望通过获

取动词的论元信息，作为动词网的辅助信息。

本文使用读者、青年文摘等文学作品、经济期刊作为语料库（共 1.3 亿字），获取与动词相关的例句，并使用 stanford 分析器^[70]抽取对应的论元信息。由于 stanford 分析器对于中文信息处理准确率不高，因此通过设置预实验来对研究动词与论元在文中共同出现的频率 P 对抽取论元准确率的影响。频率 P 取值范围为 {3, 5, 7}。选择 20 组动词种子集，抽取在语料库中包含该动词的语句，利用 stanford 分析器抽取对应的主语、宾语以及各自出现的频率（若满足主谓宾结构，则另外再输出（主语，宾语）二元组以及对应频率信息）。通过人工评判论元是否与动词相关，将无关的论元信息视为噪声，计算在不同频率阈值下 20 组动词的论元的平均准确率，部分结果如表 3.3，平均准确率如图 3.6。

表 3.3：部分动词-论元的频率

动词	频率 3	频率 5	频率 7
教育	0.21	0.66	0.66
吸收	0.34	0.48	0.50
认识	0.62	0.75	0.83
明白	0.62	0.82	0.92
保护	0.55	0.69	0.83

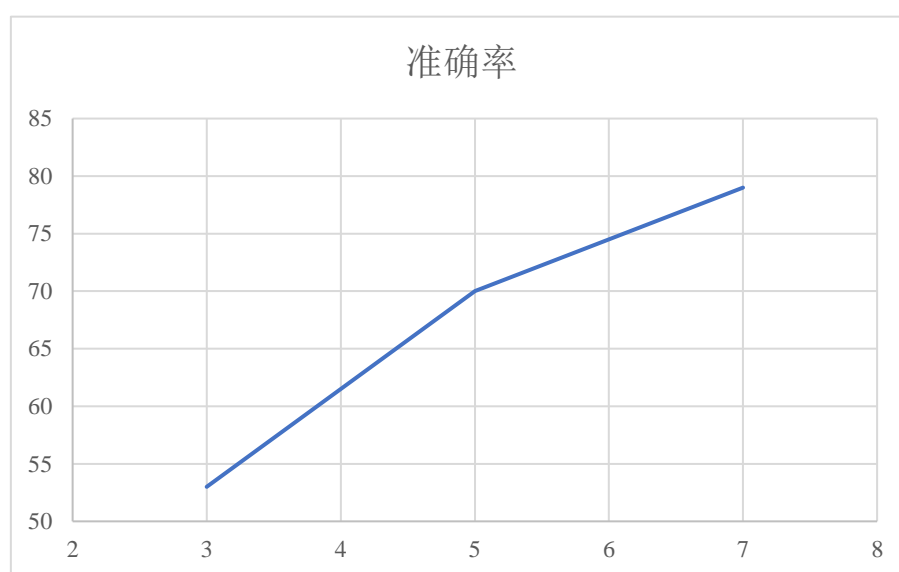


图 3.6：频率对论元准确率的影响

由图可知，随着频率的增加，论元的准确率也随之增加。因此，本文在获取到论元信息后，将论元按频率排序，获取前三个频率最高的词汇，然后通过人工筛选组成动词的论元信息。

同义词词林中对词语进行类别的归纳，并划分为大、中、小三类，依此为前者的扩充。因此我们使用同义词词林中的大类，作为论元的划分类别的依据。论元分别为人、物、时空、抽象，共四大类。然后将获取到的主语和宾语分别映射到同义词词林的词中找到所属的大类类别，并按大写字母（A, B, C）进行标注。论元标注集部分内容如表 3.4 所示。

表 3.4：论元标注集部分内容

A 人	B 物	C 时空	D 抽象
人民	生物	时代	这里
我	物体	时候	案件
儿童	产品	平时	过程
居民	物资	期限	命运
职工	泉水	初期	经验
教师	树木	春	人生

3.3.4 动词图式的引入

动词图式，顾名思义就是代表动词的经验图式。图式这个概念最初是由康德提出的，他认为图式是一种先验的范畴。在认知发展理论^[71]中，皮亚杰将图式定义为人们的行为模式或者心理结构，且该结构是有组织、可重复的。简单来说，图式就是一种认知结构的单元。一个人从出生到长大所经历的全部图式组成了这个人的认知结构。且该结构不是一成不变的，而是动态的、时刻更新的。在该理论中，皮亚杰认为行为主义 $S \rightarrow R$ 公式是有缺陷的，并提出 $S \rightarrow (AT) \rightarrow R$ 的公式，细化认知的过程。该公式意为在一定的刺激（S）下，个体能够感受并进行同化过程（A），然后将经验信息存储进认知结构（T）中，才能作出相应的反应（R）。

他对图式、同化、顺应、平衡四个概念进行重新定义，并用来形容人类认知结构的的活动过程。Johnson 认为经验图式是在所有概念之前，独立于概念之外，并且会形成一种经验结构并一直存在，新的概念也只是在这种经验结构上加强。这就是 Johnson 对意象图式的基本见解，他在《心灵中的身体：意义、想象和理性的身体基础》中还给出了一些意象图式的类别，包含容器图式、部分-整体图式、起点-路径-目标图式、空间图式等^[72]。Lakoff 在《女人、火与危险事物》^[73]中认为容器图式、部分-整体图式、起点-路径-目标图式是三个最为基本的图式，并能激活其他附属图式，比如过程图式、前后图式、力图式等等。

我们认为图式存储着个体对事物认知的过程，所以人们可以轻而易举的通过自己的经验图式获取对事物的认知，但是对于计算机却存在着困难。因此若能够提取图式信息，则对机器自然语言理解意义重大。本文通过构建动词图式来存储动词的相关信息，使得动词联想过程更为饱满，能够包含更多的语义信息。

首先通过 Lakoff 对图式的基本分类，再结合动词的特点，选定动词图式划分为容器图式、链接图式、起点-路径-目标图式以及整体-部分图式^[73]。

定义 1：容器图式由内部、界限、外部构成。

它的基本逻辑就是，事物要么在容器内，要么在容器外，在容器内可以继续停留在容器内或者离开容器，在容器外也可以继续在容器外或者进入容器内。

因此本文按活动状态细分，共四种情况：容器内事物在容器内活动；容器内事物分离至容器外；容器外事物在容器外活动；容器外事物进入容器内。

动词逻辑结构：动词 V，情况 C（X：活动者，A：容器）

表 3.5：容器图式例句

例句	动词逻辑结构
帝王拥有权力	拥有，在内（权力，帝王）
土壤长出新芽	长出，在内转在外（新芽，土壤）
烟雾笼罩村庄	笼罩，在外（烟雾，村庄）
我吸收知识	吸收，在外转在内（知识，我）

定义 2：起点-路径-目标图式包括起点、终点以及路径（连接起点和终点的

过程) 和方向。

基本逻辑是沿着一条路径从起点到终点, 那么必定会经过所有的中间点。

动词逻辑结构: 动词 V (A: 源头, B: 目的地, C: 施事, X: 方向)

表 3.6: 起点-路径-目标图式

例句	动词逻辑结构
我从武汉出发去上海	出发 (武汉, 上海, 我, -)
运动员跳水	跳水 (跳台, 水池, 运动员, 向下)

定义 3: 整体-部分图式包括整体和部分。

基本逻辑是部分属于整体, 整体存在部分必定存在, 部分分离整体则必然消失。按照整体是否存在分为两种情况: 整体存在; 整体不存在。

动词逻辑结构: 动词 V (A: 整体状态, (Ai: 第 i 个部分))

表 3.7: 整体-部分图式例句

例句	动词逻辑结构
我和他组成家庭	组成 (整体, (我, 他))
我和他离婚	离婚 (非整体, (我, 他))

定义 4: 链接图式是由两个实体以及它们之间的连接组成的。

它的基本逻辑是, 如果实体 a 与实体 b 相联系, 那么实体 a 受到 b 的限制并且依赖于 b。

动词逻辑结构: 动词 V (A: 事物 1, B: 事物 2)

表 3.8: 链接图式例句

例句	动词逻辑结构
我爱她	爱 (我, 她)
我参加会议	参加 (我, 会议)

本文按照划分的四种动词图式,对现有的 3200 个动词进行人工图式归类。得出容器图式的动词有 668 个,链接图式的动词有 1205 个,起点-路径-目标图式的动词有 276 个,整体-部分图式的动词有 69 个,不符合以上动词图式定义的 982 个。并按照获取到的论元信息,补充动词图式对应的逻辑结构,共计获得 2218 个动词的动词逻辑结构。

3.4 联想语义网的展示

本文选用 Neo4J 图形数据库来可视化的形式展现联想语义网中形容词和动词的扩展过程。Neo4J^[74]是一个高性能的图形数据库，可以将结构化的数据存储在网络中，并且在关系数据的处理上远远高于其他数据库的性能。

图 3.7 展示了形容词“矮”的部分联想过程。箭头方向代表抽象度的走向，指向的词语表示更抽象。箭头中间代表词语的关系，sim 表示相似关系，节点词中包含该词的 id、情感极性以及抽象度值。其中“矮-低-卑微”的扩展过程也符合我们对形容词“矮”的认知。

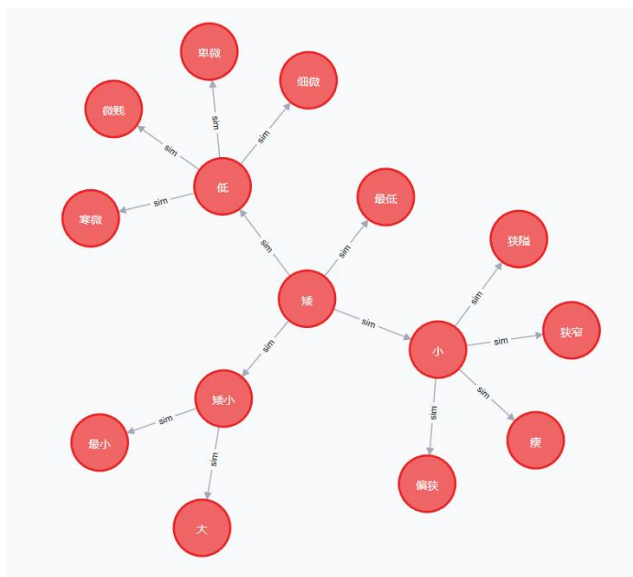


图 3.7: 形容词“矮”的部分联想过程

图 3.8 为动词“爱”的部分联想过程。箭头方向代表下一层级，箭头中间代表词语的关系，sim 为相似关系，动词的图式信息以及逻辑结构存储在节点中。图 3.9 为动词“爱”联想过程中部分的动词图式信息。我们发现拥有相似关系的节点，也拥有相似动词图式和对应的逻辑结构。

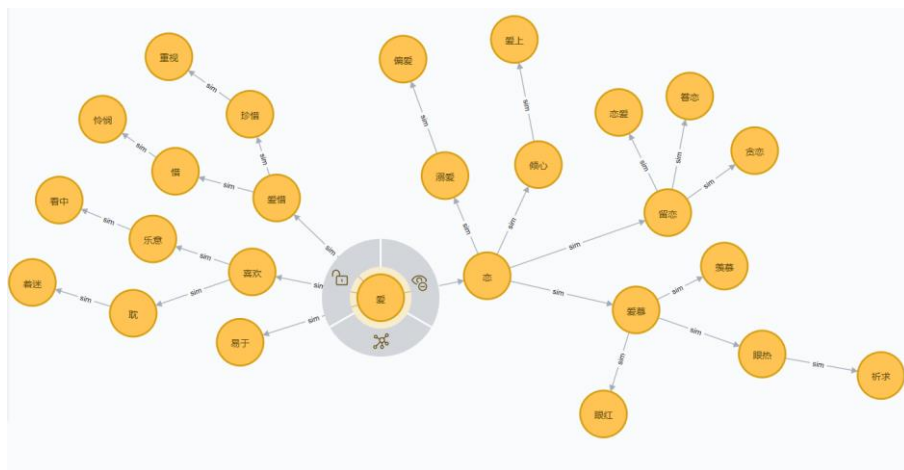


图 3.8: 动词“爱”的部分联想过程

表 3.9: 动词“爱”的部分扩展词的动词信息

动词	扩展动词	动词（扩展动词）图式	动词（扩展动词）逻辑结构
爱	爱	链接图式	(A, B)、(A, A) …
	喜欢	链接图式	(A, B)、(A, A) …
	珍惜	链接图式	(A, B)、(A, C) …
	重视	链接图式	(A, B)、(A, A) …

3.5 联想语义网的评估

3.5.1 构建规模

本文构建的联想语义网分为两个部分，形容词联想语义网和动词联想语义网。形容词和动词种子集由两部分组成，分别为北大汉语语法信息词典中常用词汇^[65]和文学作品中的词汇，共 1850 个形容词和 3200 个动词。形容词网是依据抽象度进行延伸，因此我们将 360 个更为具体的形容词作为形容词网构建的基础，并通过联想过程扩展得到近百万个词汇。考虑到动词的联想，不仅包括动词的相似关系，大多仍与论元相关。因此构建的动词网方式有异于形容词。动词网由 3200 个动词构成，其中包含动词图式信息的动词有 2218 个，且有对应的逻辑结构信息。

3.5.2 评价方法

本文采用人工评价的方法对我们自主构建的联想语义网进行评估。随机抽取形容词网和动词网中各 50 个词语的扩展过程。邀请 5 位母语为汉语，且具有 NLP 研究背景的相关人员对抽取的信息进行评估。

评价过程中，我们要求评价者依照以下原则进行评估：

1. 根据自己的理解，该目标词呈现的扩展过程是否满足本人对该词的联想过程，且该联想过程是否满足“具体-抽象”的发展过程。

3.5.3 结果分析

综合 5 位评估者的评估结果，100 个扩展词中 90 个是满足的，还有 10 个未满足联想过程。5 位评估者的标注一致性 K 为 0.74，符合评价一致性^[76]的标准。因此，认可 5 位评估者对语义网的评估。

通过评估，我们认为词汇在联想语义网中扩展的过程是在认可范围内的，且经过三次扩展就可以获取到丰富的语义信息。由于篇幅限制，附录 C 列出了形容词网中词语联想比较有代表性的部分（加粗表示包含文化信息部分）。附录 D 列出了动词网中代表性的动词、相关图式以及逻辑结构信息（加粗表示包含文化信息部分）。从附录中，我们也可以看出联想语义网包含一定的文化信息，比如，红的联想过程。在中国文化中，红色既代表着喜庆，体现古代婚嫁喜事的风采，又代表着革命战争年代的一种热血的精神。因此就有了“红-吉利-吉庆-喜庆”、“红-革命”的联想过程。

但是联想语义网中仍存在一些不足之处，下面分四点对其进行叙述：

- （1）语料库中词语信息缺失。本文选用同义词词林作为我们构建联想语义网的语料库。但由于语料库中的知识不可能涵盖所有的内容，因此存在信息缺失的情况，从而导致联想语义网中联想过程的中断。比如，对于词语“薄情”，从语料库中抽取的信息只有“喜新厌旧”，而喜新厌旧并没有下级的扩展词，因此仅有“薄情-喜新厌旧”的联想过程。

- （2）扩展过程中仍存在不相关的词。虽然已经通过相关性筛选掉大部分相关性低的词语，但仍存在漏网之鱼。比如，动词“哀”的联想过程“哀-哀悼-哀求-请求-告”。从“请求”开始，与目标动词“哀”的关联性开始变弱，而后动词

“告”与动词“哀”并没有关联，但未通过相似度算法筛检出。因此，未来还需对联想语义网进一步地完善和删改。

（3）本文构建联想语义网的种子集为常用词汇。因此，也会缺失一些生僻信息，从而导致使用自主构建的语料库仍会受到限制。

（4）动词图式的划分。通过观察 3200 个动词，我们发现仍有 900 多个词汇并没有包含在我们定义的动词图式中。在已有的动词图式中，我们发现存在词语由于拥有不同含义，因而拥有多种图式的情况。比如动词“留”，在它表示“停止在某一个地位不动”的含义时，为起点路径图式；表示“保留”或“接受”的含义时，为容器图式。但本文对每个动词仅赋予单个图式，从而导致部分信息的缺失。这将是未来工作两个需要注意的地方，如何对图式进行更细致的划分以及如何定义单个动词在不同语义下的动词图式信息。

第四章 基于联想语义网的汉语隐喻理解研究

4.1 汉语隐喻理解分析

4.1.1 汉语隐喻理解的过程

在日常生活的交流中，人们常常会使用自身经验来理解隐喻。将主观的经验主义和感觉运动经验相互关联，使人们可以在无意识地情况下，就能使用隐喻思维来自动理解隐喻或生成隐喻^[4]。胡壮麟^[86]提出隐喻理解是将陌生的事物和我们所熟知的事物进行对比，以此来加深对陌生事物的认知，是未知到新知的过程。因而，我们认为汉语隐喻的理解过程，分为如下三个步骤。

1. 经验图式获取。经验图式是一个人在日常生活中累积的经验信息，且该信息包含该个体所处环境的文化信息，可以用作对事物的判断。在汉语隐喻的理解过程中，我们只有具备了与汉语相关的经验图式，才能对其进行正确的解读。因此，在汉语框架下获取与隐喻相关的经验图式，是汉语隐喻理解过程中的第一步。

2. 联想过程。该过程就是对源域信息和目标域信息进行联想。史鑫^[87]将联想分为相似性联想和创造相似性联想，前者通过获取的经验图式，找寻两者的相似性，是一个激活以及调用的过程。创造相似性联想则是通过一定的主观想法来创造两者的联想过程。无论是相似性联想，还是创造相似性联想过程，最终获取的都是两者的相似信息。

3. 得到汉语理解结果。通过联想过程获得源域和目标域的相似信息后，我们需要对该信息进行筛选，得出最符合的理解结果。

因此，汉语隐喻的理解过程本质上就是受喻者通过自己拥有的经验图式进行联想，并获取相似结果的过程，且该过程必然受到汉语文化的影响。

4.1.2 汉语隐喻理解的特点阐述

首先，汉语隐喻理解最明显的特点就是文化性。汉语隐喻诞生于中华民族这个文化群体，受其影响以及制约，且普遍存在于该文化群体的所有个体中。人们创造隐喻来表达共同认识到的观念，这个过程是在拥有共同意识形态的社会群体

间发生的，也就是说，汉语隐喻理解是带有独特的汉语文化烙印，需要我们身处于相同汉语文化框架中，拥有共享的概念结构及世界观才能对其进行正确的解读。

其次，歧义性也是汉语隐喻理解的特征之一。语言的意义本身就具有不确定性。每个人都是单一的个体，由于其成长经历、生活经验不同，导致对事物的理解、认知也存在着偏差。隐喻作为语言的一部分，自然也存在着不确定性。因此在失去语境的情况下，对语言的理解结果是千差万别的。比如，“小明是一只猪。”通过联想猪的特征为懒惰，对该句的解读则是小明是懒惰的；联想到猪能吃的特点，对该句的理解则是小明是特别爱吃的；联想到猪是可爱的象征，则对该句的理解为小明是可爱的。由于各种不确定的因素，使得施喻者表达的意义和受喻者理解的意义产生了不同的情形，导致了汉语隐喻理解的歧义性。

最后，汉语隐喻理解具有互动性。最初，人们认为理解隐喻的意义就是比较源域和目标域的相似性，从而匹配得出一致的特征作为理解结果，即特征匹配理论。Black 曾提出从新的角度去看待源域和目标域，不去比较两者的相似点，而是在两者之间创造相似性，使之进行互动。根据他的观点，受喻者在考虑到目标域的情况下选取源域部分特征，并构建平行且适用于目标域的隐含综合体，然后该综合体又反过来作用于源域并引起其内部的变化^[19]，得出最后的理解结果。汉语隐喻理解作为隐喻理解的一部分，同样也具有互动性的特征。

4.1.3 影响汉语隐喻理解计算的主要因素

根据汉语隐喻的理解过程和特点，我们可以得出以下三点影响汉语隐喻理解计算的因素：

1. 相似性因素

认知心理学中指出相似性为两个或多个心理表征之间存在的心理接近性或者邻近度^[78]。根据 black 的互动论，隐喻的相似性既包含物理相似性，又包含心理相似性。其中，物理相似性指的就是源域和目标域存在的共同特征，如颜色、形状、功能等方面的相似。心理相似性是指两者不存在实际的相似特征，而是人为地从主观上去创造了两个域的共同特性，使得心理感受上是相似的^[19]。结构映射理论中也提到了隐喻相似性，并将相似性分为特征相似和关系相似。前者指的是属性的相似，后者指的是关系结构的相似。人们通过事物间的相似性创造隐喻，

因而无论是特征相似还是关系相似都是汉语隐喻理解的需要考虑的因素。

2. 文化因素

由于不同的民族身处的环境不同,从而导致其意识形态、文化存在差异,对隐喻的理解大相径庭。比如,汉语中的“狗”常常用作贬义,代表低贱的意思;而英语中“狗”是褒义的,代表忠诚。因而当我们提到汉语隐喻“狗仗人势”时,如果本身不具备相关的汉语文化知识,无法正确理解该隐喻。要想对其进行解读,文化的重要性不言而喻。因此对汉语隐喻的理解是无法脱离汉语文化背景的。

3. 语境因素

Kittay^[79]曾指出在人们日常交流中,对语言的理解总可以在更大的语境中得以修正。显然,对汉语隐喻的理解,很大程度上受到了上下文的影响和钳制。不同的受喻者对同一个隐喻喻体有着不同的理解,在较少的语境下,人们只能依靠自己的判断对隐喻进行解读,然而这个解读往往伴随着歧义性。只有在具体语境下,我们才能判断施喻者真正想要表达的含义。侯友兰^[94]曾对语境在修辞中起到的作用进行总结,语境可以消除歧义、衍生新义、使得词语产生情景意义等等。因此语境提供的信息越多,受喻者对汉语隐喻的理解越透彻。

4. 2 汉语隐喻理解例句分析

4. 2. 1 汉语隐喻理解例句分类

相似性是构成隐喻的基础,同样的,在汉语隐喻理解的过程中相似性因素是贯穿始终的。结构映射理论^[6]中提到隐喻的相似性包括特征相似性和关系相似性。特征相似性是指源域和目标域存在相似的属性(比如,颜色、形状、功能等方面的相似);关系相似性是指源域和目标域存在关系结构上的相似。在汉语隐喻中同样存在特征相似型和关系相似型隐喻。但目前对汉语隐喻的研究,尤其是对其名词隐喻的研究,都只关注于源域和目标域特征上的相似,而忽略了两者关系结构上的相似。而在理解中,如果使用特征相似去解读关系相似的隐喻句是无法得出较好的理解结果的,反之亦然。因此,我们认为对汉语隐喻的理解也是要按实际情况分类型进行解读,不能一概而论。本文通过对汉语隐喻例句进行分析,并依据认知语言学理论,将隐喻例句分为特征相似型隐喻与关系相似型隐喻两种。

4.2.2 汉语隐喻理解例句分析

1. 特征相似型隐喻

表 4.1: 特征相似型隐喻

隐喻句	源域	目标域	隐喻理解结果
时间是金钱	金钱	时间	时间是重要的
律师是鲨鱼	鲨鱼	律师	律师是凶恶的

特征相似型是人们可以客观地匹配出源域和目标域相似的属性。就像例句“时间是金钱”，金钱最显著的特征就是宝贵，通过宝贵联想到它的重要性，时间也有重要的特性，因此得出理解结果，时间是重要的。“律师是鲨鱼”，鲨鱼的特征是凶恶的，由此联想到律师也可以是凶恶的，从而得出理解结果，律师是凶恶的。

2. 关系相似型隐喻

表 4.2: 关系相似型名词例句

隐喻句	源域	目标域	理想隐喻理解结果
老师是园丁	园丁	老师	老师培养学生就像园丁培育花朵
学生是花朵	花朵	学生	老师培养学生就像园丁培育花朵

表 4.3: 关系相似型动词例句

隐喻句	源域动作	目标域场景	理想隐喻理解结果
金融危机肆虐全球市场	肆虐	金融危机，市场	金融危机影响市场就像狂风肆虐树林
吃透汉字	吃	汉字	人理解汉字就像人吃东西

关系相似型隐喻针对的是源域和目标域相似的关系结构。在关系相似型的名

词隐喻例句中,如果仅仅通过特征相似得出的理解结果是,老师是辛勤的以及学生是娇嫩的。虽然关系相似型的名词隐喻可以使用特征型隐喻理解方式,但并没有能完全的传达出该隐喻真正的含义。如果通过关系相似,将老师和学生与园丁和花朵进行对比,得出老师教育学生就像园丁浇灌花朵。该理解结果较于特征型理解结果更为贴切。在动词隐喻句中,往往都是以关系结构为主的。我们通过找寻源域和目标域相似的关系以及场景对其进行解读。就如表 4.3 中的例句“金融危机肆虐全球市场”和“吃透汉字”,通过句中的目标域场景联想到相关的动作,金融危机影响市场以及理解汉字,然后通过源域动作联想源域的场景,狂风肆虐树林和吃东西。将得出的源域信息和目标域信息进行比对,得出最后的理解结果。

4.3 基于联想语义网的汉语隐喻理解

本文主要研究汉语中的名词隐喻和动词隐喻。我们认为汉语名词隐喻中既包含特征相似型隐喻,又包含关系相似型隐喻,动词隐喻中则往往都是关系相似型的。而特征型和关系型隐喻的理解过程是不相同的。因此,对汉语隐喻的理解过程需要分类型进行处理。首先,我们对汉语隐喻句中的源域、目标域进行人工标注。然后,根据不同的类型,从自主构建的联想语义网获取需要的信息,作为候选理解结果。最后,通过源域信息和候选信息的在联想语义网中的相似距离结合情感(图式信息)进行排序,选择最优信息作为汉语隐喻理解的结果。

4.3.1 特征相似性汉语隐喻算法流程

特征相似型汉语隐喻理解算法流程包含以下四个步骤。

Step1: 数据预处理

对于每个汉语隐喻句,人工标注出源域 S 和目标域 T。

Step2: 源域和目标域属性获取

从 Cogbank 属性知识库和自主构建的属性库中分别抽取出源域的属性 C_i 和目标域的属性 D_i ,并形成相应的属性集合。若在两个属性库中未找到源域和目标域属性,则直接输出并标记“*”,表示信息缺失。

Step3: 通过联想语义网寻找其关联

获取到源域的属性 C 集合，并通过自主构建的形容词网进行迭代拓展，直到有符合目标域的属性 D 出现，则进入理解的候选集合 P。

Step4: 排序和选择

对于获取的语解释义的集合，综合考虑源域属性 S 与候选释义 P 在形容词网中的语义距离和情感因素进行排序，选择最优的结果作为汉语隐喻理解的结果。

图 4.1 为特征相似型汉语隐喻理解过程的整体流程图。

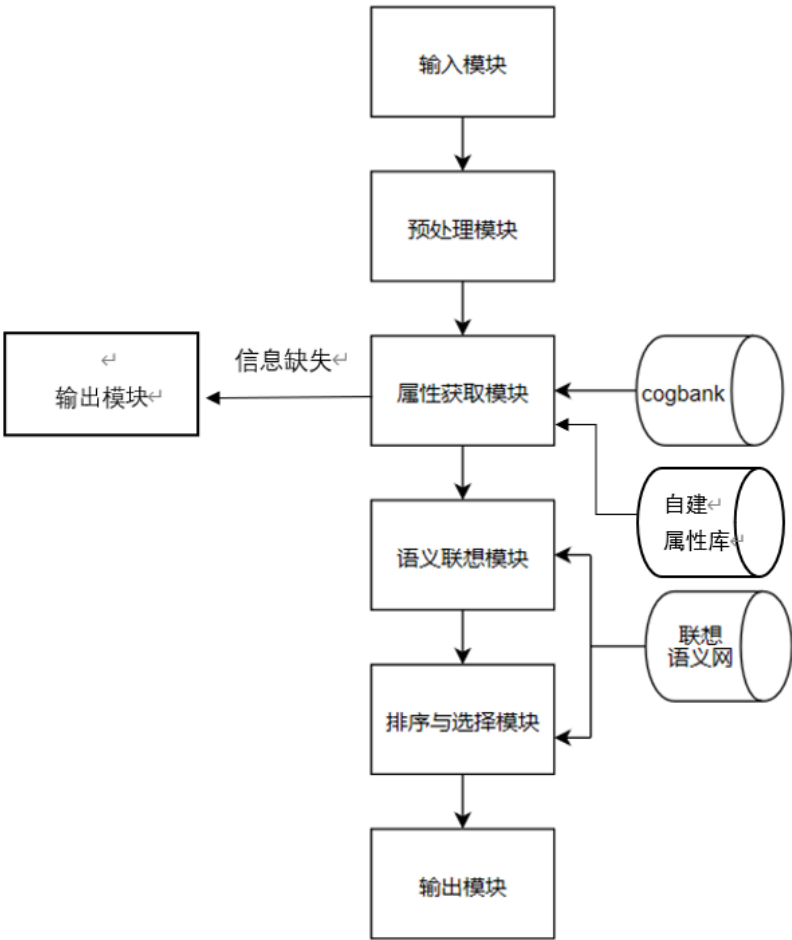


图 4.1 特征型汉语隐喻理解流程图

4.3.2 名词属性集合构建

Cogbank 属性知识库通过网页抽取特定格式的明喻，收集与名词相关的属性，并包含近万个词语。但在汉语隐喻的理解过程中，我们发现存在部分词在

Cogbank 属性库中缺失属性或属性较少的情况，从而导致理解结果不尽如人意。因此，本文通过自动抽取名词的属性，来增加相关的属性信息，以提升汉语隐喻的理解结果。我们采用读者、青年文摘等语料库来进行名词的属性抽取工作。首先，从语料库中获取包含目标词的句子，然后对句子进行分词，并标注词性。由于使用 stanford 分析器并没有达到理想的抽取效果，因此我们通过对句子进行分词并标注词性，然后设定语法规则来抽取名词属性知识。jieba 分词工具^[90]是一款分词能力非常出色的 python 库。我们通过 jieba 分词工具对本文进行处理，并设定了三条规则来抽取名词属性，并对提取的属性信息进行筛选。如果在语料库中无法找到名词的属性，则直接输出，等待后续理解。表 4.4 为名词与属性的语法抽取规则。附录 E 展示了部分在 Cogbank 属性库中的属性信息与自动抽取的属性信息的对比。

表 4.4：名词-属性语法抽取规则

规则	例子
/a **/n (/nr)	/a 阳光/nr
/a 的/uj **/n (/nr)	/a 的/uj 阳光/nr
像/v **/n (/nr) 一样/r	像/v 阳光/nr 一样/r

4.3.3 基于形容词语义相似和情感的排序

在对源域和目标域扩充属性集之后，我们进行下一步骤，通过联想语义网找寻两者之间的关联并自动筛选出最符合的理解结果。

首先，我们获取源域的属性集合，并对属性集合中的词语在形容词网中进行扩展，直到出现符合目标域属性集的词语出现。将符合的属性添加至候选属性集合。若扩展六次尚未成功，则表示未找到关联属性，当作属性缺失问题处理。然后，我们需要在候选属性集中选择一个最符合的属性作为理解结果。因此，需要对候选的属性集进行量化排序。本文排序的标准包括两方面，第一个，源域属性与候选属性在形容词网中的平均语义相似度，第二个是它们的情感信息。

Palmer^[96]提出了一种基于语义距离的相似度算法，并定义如下公式。

$$sim(s_1, s_2) = \frac{2H}{2H+N_1+N_2} \quad (4-1)$$

H 表示深度，即词语义项 s_1 和 s_2 最近公共父节点到根节点的距离， N_1 与 N_2 表示词语义项 s_1 和 s_2 与最近公共父节点的路径距离。陈宏朝^[91]认为该方法未能对路径和深度进行动态调节，因此提出了一种动态的相似度计算方法。

$$sim(s_i, t_i) = \frac{Depth(LCP(s_i, t_i)) + a}{Depth(LCP(s_i, t_i)) + a + Path(s_i, t_i) + b} \quad (4-2)$$

其中，Depth 表示为 s_i 和 t_i 距离最近的公共父节点到根节点的深度距离，Path 表示词语义项 s_i 和 t_i 到最近公共父节点的路径距离。a, b 为调节参数。相似度值越接近 1 则表示越相似。

本文基于陈宏朝^[91]提出的语义相似度计算的方法，增加了情感因素的考量。因此，词语语义相似度 $G(s_i, t_i)$ 计算公式如下：

$$G(s_i, t_i) = n * sim(s_i, t_i) + m * sem(s_i, t_i) \quad (4-3)$$

其中， $sim(s_i, t_i)$ 表示源域属性与候选属性在形容词网中的语义距离， $sem(s_i, t_i)$ 表示情感的相似值，n、m 表示距离相似和情感因素的权重，并满足 $n+m=1$ 。

以下是对 n 和 m 的取值的实验过程的简要描述。首先，我们选择测试集外 10 个隐喻句以及可接受的理解结果。由于形容词网本身就是根据相似关系进行扩展，所以将权重 n 的值设为大于 0.5。通过人工判定不同权重下得到的理解结果是否符合认知，选择准确率最高的参数作为 n 和 m 的权重。经实验，取 $n=0.9$, $m=0.1$ 。因此，本文基于形容词语义距离的相似度计算公式如下：

$$G(s_i, t_i) = 0.9 * sim(s_i, t_i) + 0.1 * sem(s_i, t_i) \quad (4-4)$$

$G(s_i, t_i)$ 的取值范围为 $[0, 1]$ ，越接近 1 表示越相似。最后，按由高到低对候选词的形容词语义相似度进行排序。表 4.5 列出了特征型汉语隐喻理解的部分结果。对于表中汉语隐喻句“关公是红脸的”，虽然源域“红脸”的属性只有红，但仍能通过本文构建的形容词网进行扩展，并得出勇敢、无畏、正直三个与之相关的文化属性。最后，得到理解结果为“关公是正直的”。因此再次验证本文构建的形容词网是包含文化的，并可以处理汉语隐喻中的文化问题。

表 4.5：特征型汉语隐喻的理解结果

汉语隐喻句	源域属性	候选属性集	理解结果	基于形容词网的语义相似度
爱情是甘露	澄清、甜、芬芳、清纯、纯	甜、鲜明、寂寞、痛、永恒、鲜美、美、完美、热烈、深刻...	爱情是甜的	0.68
			爱情是鲜明的	0.61
			爱情是寂寞的	0.46
			爱情是痛的	0.52
母爱如水	流动、平凡、清澈、温暖、伟大...	细腻、平等、多、伟大、无边	母爱是细腻的	0.41
			母爱是平等的	0.09
			母爱是多的	0.38
			母爱是伟大的	0.35
红脸的关公	红	勇敢、无畏、正直	关公是勇敢的	0.51
			关公是无畏的	0.51
			关公是正直的	0.56

4.3.4 关系相似性汉语隐喻算法流程

关系相似型汉语隐喻理解算法流程包含以下四个步骤。

Step1: 数据预处理

对于汉语名词隐喻句,人工标注出源域 S 和目标域 T;对于汉语动词隐喻句,人工标注出源域动作 V 和目标域场景 C。

Step2: 源域和目标域的动作获取

如果处理的是名词性隐喻,则需要从语料库中抽取出源域和目标域的可能的动作集合。动词性隐喻,则根据目标域场景抽取可能的动作集合。

Step3: 动作排序与选择

从动词网中获取到与源域和目标域的相关的动作集合,通过基于动词网相似度算法来对候选动作进行排序,选取相似度最高的作为目标域动作。若未能获取到相关的动作集合,则输出源域动作和目标域的可能动作集合,然后结束。并标

记“*”，表示信息缺失。

Step4: 关系型隐喻理解

从动词网中，获取源域的动词图式作为源域场景，得出最后的理解结果。

图 4.2 为关系相似型汉语隐喻理解过程的整体流程图。

```
graph TD; Input[输入模块] --> Pre[预处理模块]; Pre --> ActionAcq[动作获取模块]; Corpus[(语料库)] --> ActionAcq; ActionAcq --> ActionSel[动作选择模块]; ActionSel -- "信息缺失" --> Output1[输出模块]; ActionSel --> RelationAna[关系类比模块]; Corpus2[(联想语义网)] --> ActionSel; Corpus2 --> RelationAna; RelationAna --> Output2[输出模块];
```

图 4.2: 关系型汉语隐喻理解流程图

4.3.5 动作候选集合获取

与形容词网抽取名词的属性信息相同，本文仍采用读者等语料库以及 jieba 分词工具^[90]抽取动作信息。我们抽取与场景信息相关的动作（名词关系型隐喻中，则抽取源域和目标域的动作；动词关系型隐喻中，抽取目标域场景的动作），并将获得的动作进行比对，筛选相同或类似的动作信息作为动作候选集合。若目标域场景只有一个，则只抽取该词语的动作信息。通过对语料库中的句子进行分析，我们发现动词与名词的搭配比较灵活。因此，我们抽取与名词距离范围为 5 的动

39

作信息，并提取至少出现两次以上的动作，通过筛选进入候选集合。附录 F 展示了本文抽取的部分动作信息。

4.3.6 基于动词语义相似的排序

在获取动作候选集合之后，我们通过两个步骤来对动作集合进行排序。首先，通过动词网对源域进行六次扩展，并匹配符合目标域的动作集合的词语，添加至候选动作集合。名词隐喻则直接使用候选集进行扩展。其次，本文结合动词图式对陈宏朝^[91]基于语义距离的相似度算法进行修改，公式如下：

$$G(s_i, t_i) = n * sim(s_i, t_i) + m * V(s_i, t_i) \quad (4-5)$$

其中， $sim(s_i, t_i)$ 表示源域动作与侯选动作在动词网中的语义距离， $V(s_i, t_i)$ 表示两者图式信息的相似值， n 、 m 表示语义相似和图式因素的权重，并满足 $n+m=1$ 。

由于一个动词可能存在不同的动词图式，本文将动词图式的相似值分为 0.5 和 1，表示两者图式不同和相同的情况。以下是对 n 和 m 的取值的实验过程的简要描述。同样选择测试集外 10 个隐喻句以及可接受的理解结果。通过人工判定不同权重下得到的理解结果是否符合认知，选择准确率最高的参数作为 n 和 m 的权重。经实验，取 $n=0.8$, $m=0.2$ 。因此，本文基于动词语义距离的相似度计算公式如下：

$$G(s_i, t_i) = 0.8 * sim(s_i, t_i) + 0.2 * V(s_i, t_i) \quad (4-6)$$

$G(s_i, t_i)$ 的取值范围为 $[0,1]$ ，越接近 1 表示动作相似度越高。最后，从动词网中获取源域动作的论元信息组成最后的理解结果。表 4.6、4.7 分别是关系型汉语隐喻句的理解结果。由表 4.7 可以看出，“船犁大海”以及“吃透精神”的源域动作“犁”和“吃”都带有独特的汉语文化气息。且通过动词网得出的理解结果“船驶入大海就像犁犁田”以及“学习精神就像吃东西”，两者都符合我们对这两个动作的认知。因此，我们认为本文构建的动词网同样可以处理汉语隐喻中的文化问题。

表 4.6: 关系型汉语隐喻的理解结果(名词)

汉语隐喻句	源域动作集合	目标域动作集合	源域动作相似得分	目标域动作相似得分	理解结果
医生是屠夫	杀、宰杀、进入	治疗、发现、诊治	杀 0.39	治疗 0.37	医生治疗病人就像屠夫杀猪
			宰杀 0.30	发现 0.1	
			进入 0.30	诊治 0.36	
思想是潮水	涨起、涌、淹没、汹涌、退去	保持、动摇、滋长、放松...	涌 0.352	准备 0.106	思想滋长就像潮水涨落
			涨落 0.602	滋长 0.241	
			-	解放 0.098	
爱情是灯	关掉、熄灭、点亮、点燃	开始、发生、幻灭、破裂...	点亮 0.1	幻灭 0.171	爱情幻灭就像灯熄灭
			熄灭 1	开始 0.168	
			关掉 0.1	发生 0.170	

表 4.7: 关系型汉语隐喻的理解结果(动词)

汉语隐喻句	候选理解结果	动作相似度	理解结果
船犁大海	船驶入大海就像犁犁田	0.282	船驶入大海就像犁犁田
	船行驶大海就像犁犁田	0.227	
	船沉没大海就像犁犁田	0.01	
公司萌芽	公司开拓就像种子萌芽	0.379	公司开拓就像种子萌芽
	公司注册就像种子萌芽	0.266	
	公司倒闭就像种子萌芽	0.271	
	公司破产就像种子萌芽	0.273	
吃透精神	学习精神就像吃东西	0.356	学习精神就像吃东西
	发扬精神就像吃东西	0.242	
	开拓精神就像吃东西	0.353	
	超越精神就像吃东西	0.225	

4.4 认知缺省推理

通过研究基于联想语义网的汉语隐喻理解的过程,我们发现由于知识库中的信息总是不完善的,因此存在 Cogbank 属性库和自主构建属性库中缺失目标属性以及同义词词林中缺失信息导致构建的联想语义网缺失信息的情况。而这些情况都会使汉语理解过程中断。但在现实生活中知识缺失的情况常常存在,人们仍可以对隐喻进行解读。所以我们在知识缺失的情况下,同样能够对汉语隐喻进行理解,因此引入认知缺省推理来解决该问题。

认知逻辑的研究始于 20 世纪 40 年代,80 年代开始广泛运用于计算机科学、人工智能、哲学、语言学和经济学等学科中。它拥有良好的公理系统,并为知识和信念的推理与分析过程提供语义信息。Hilpinen^[92]认为在认知逻辑中,“a 知道 p”表示为:(1) a 可以接受 p;(2) 对于 p, a 有充分的依据;(3) p 为真。Hintikka^[84]通过结合二元认知算子,来表达命题的认知形式,奠定了认知逻辑的基础。用 $K_i a$ 表示主体 i 在可能世界中,知道情形 a 的存在。 $B_i a$ 表示主体 i 在可能世界中,相信情形 a 的存在。而后人们开始对认知算子进行语义解读,给出了常见的 K 公理^[93],并建立认知逻辑对应的公理系统。20 世纪中后期,人们开始将认知逻辑与其它非经典逻辑结合来处理问题,也逐渐成为认知逻辑研究的新热潮,并使得认知逻辑可以应用于实际生活中。

缺省逻辑最早是 Raymond Reiter(加拿大计算机科学家和逻辑学家)在 1980 年提出的,用于在缺失主体信息情况下的推理。它可以表示为缺省的,某个事物为真。意思就是某一事实 S 在缺省状态下成立,如果没有证据表明 S 不成立,那么 S 总是成立的。比如,在一般情况下鸟都是会飞的,如果 x 是鸟,且 x 会飞与现有的知识不冲突,那么就会推出 x 会飞的结论。但如果后续系统知道 x 不会飞时,那么该结论则不成立。

缺省推理是在缺少证据和没有违反限制信息的条件下进行推理。因此在推理过程中缺乏主体对知识信念的刻画能力,从而导致新增信息后推理很可能被撤销。而认知逻辑可以描述主体知识与信念的认知状态的更新过程,于是人们尝试引入认知逻辑与 Reiter^[95]缺省理论相结合来弥补缺省推理的不足之处。认知缺省逻辑的主要思想就是在缺省推理规则中增加对知识和信念的刻画。Meyer 等人^{[99][100]}提出了在认知逻辑框架下对缺省信息进行处理,使用 $P_i \phi$ 来表示“在认知

框架 i 下推测缺省信息 Φ ”，突出了知识的重要性。李永礼等人^[101]、吴梅红^[85]和张敏^[82]则在 Reiter 缺省理论的基础上，结合模态算子对缺省规则进行扩展，来更好地刻画主体知识。

4.4.1 认知缺省推理规则

文献^{[81][85][95][101]}给出认知逻辑的公理，以及缺省理论和认知缺省理论的相关定义及规则。

1. 语法

对于认知逻辑和缺省逻辑，引入如下符号集：

原始命题集 $I = \{a, b, c, \dots\}$

模态算子：K（知道），M（可能），B（相信）

逻辑联结词： \Rightarrow （蕴含）， \neg （非）， \wedge （交）

2. 认知逻辑公理及规则：

$$(1) (Ka \wedge K(a \Rightarrow b)) \Rightarrow Kb$$

$$(2) (Ba \wedge B(a \Rightarrow b)) \Rightarrow Bb$$

$$(3) Ka \Rightarrow a$$

$$(4) Ka \Rightarrow Ba$$

规则：

$$(1) a \text{ 与 } a \Rightarrow b \text{ 可推出 } b$$

$$(2) a \text{ 可推出 } Ka$$

3. 缺省理论定义

定义 1. 缺省模型为二元组 $\langle D, W \rangle$ ，其中 D 为缺省规则集； W 为公式集，缺省推理规则的一般形式^[95]为： $(a: M\beta_1, \dots, M\beta_m) / \gamma$ 。

a 为规则的先决条件， β 为规则的缺省条件， γ 为规则的结论， M 为模态算子，意为“可能”。该规则的意思是在先决条件满足的情况下，且现有的知识与系统不相悖，则可以得出结论成立。

4. 认知缺省模型以及推理规则

定义 2. 缺省模型为二元组 $\langle D, W \rangle$, 其中 D 为缺省规则集; W 为公式集, 认知缺省推理规则的一般形式^[95]为: $(Ka: M\beta / K\gamma)$ 。

K (知道)、 M (可能) 为模态算子。该规则的意思是在缺乏相关信息的情况下, 如果缺省规则 β 与现有主体知识 a 不相悖, 则认为该结果 γ 是成立的。

5. 应用于汉语隐喻理解的认知缺省模型及推理规则

前人^{[101][85]}对认知缺省逻辑的研究工作仅限于理论方面, 并未使用到实际应用中。我们认为在缺失信息的汉语隐喻理解过程中使用认知缺省逻辑可以推理出更合理的结果。因此本文基于李永礼^[101]的认知缺省推理规则进行扩展, 提出可用于汉语隐喻理解过程的认知缺省模型, 并对理解过程中缺省信息的语义相关性进行定义。

定义 3. 认知缺省模型为一个三元组 $M = (W, D, R)$ 。

其中, W 表示主体信念集, D 表示缺省规则的集合, R 是为 w 上的一个二元关系, 用于认知过程中对缺省规则的判断。模型的含义为: 若 aR_d 为真, 则表示在主体信念集中, 缺省规则 d 不违反主体知识 a_i 语义限制信息, 则该结果是可达的。

定义 4. 认知缺省推理规则形式为: $(Ka_i: Md_1, \dots, Md_m) / B\gamma$ 。

K (知道)、 M (可能)、 B (相信) 为模态算子。解释为: 在缺乏相关信息的情况下, 如果主体拥有知识 a_i , 且缺省规则 d_m 与现有主体知识不相悖 (即主体知识与结果 γ 有语义相关性), 则相信该结果 γ 是成立的。

定义 5. 结合认知公理, 主体知识 a 与缺省信息 b 的语义相关性分为如下两种:

$$K_wRel(a, b) = (Ka \wedge K(a \Rightarrow b)) \Rightarrow Kb$$

表示主体信念 w 知道知识 a 和词语 b 有语义相关性;

$$M_wRel(a, b) = (Ba \wedge B(a \Rightarrow b)) \Rightarrow Bb$$

表示主体信念 W 认为知识 a 和词语 b 可能有语义相关性；
其中， $K=\neg (\neg K)$ ， $M=\neg (\neg M)$ 。

4.4.2 基于认知缺省推理的汉语隐喻理解

影响汉语隐喻理解的因素之一就是语境，语境所提供的信息越少，我们对隐喻的理解就越困难。本文通过引入认知缺省推理来对存在缺省信息的汉语隐喻句进行二次理解。希望在知识缺失的情况下，同样理解出汉语隐喻的结果。

图 4.3 为认知缺省推理的过程图。

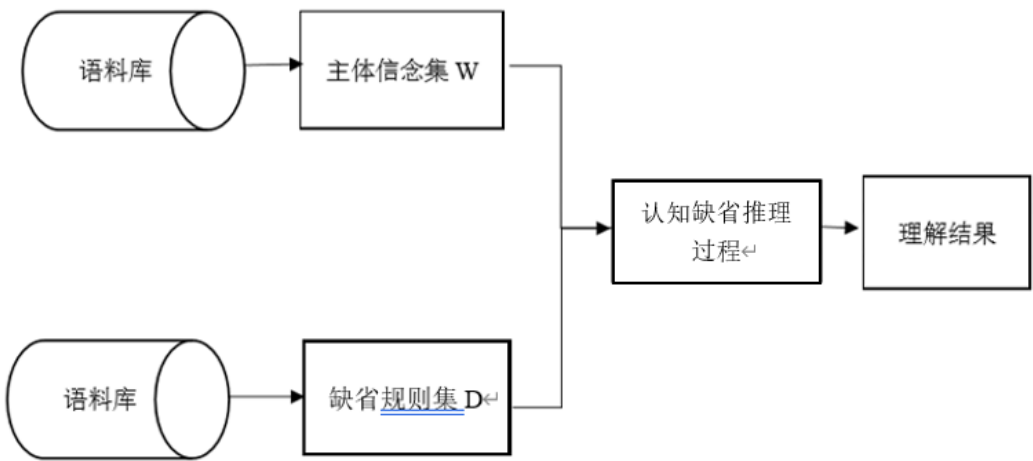


图 4.3： 认知缺省推理过程

首先，我们对汉语隐喻句以及理解过程中需要用到的公式进行定义。

定义 6. 将汉语隐喻认知模型定为四元组 $\langle S, T, S_I, T_I \rangle$,

$S_I = \{s_i \mid s_i \text{ 为汉语隐喻句中源域的认知信息集合}\}$,

$T_I = \{t_i \mid t_i \text{ 为汉语隐喻句中目标域的认知信息集合}\}$,

其中，若为特征型汉语隐喻句，则 S 为源域(source), T 为目标域(target),
 S_I 和 T_I 集合中包含属性；若为关系型隐喻句，则 S 为源域动作， T 为目标域场景
 S_I 集合中包含动词相应的主语和谓语信息， T_I 集合中包含目标域可能动作集合。
 S_I 和 T_I 的信息从 Cogbank 属性库、自主构建属性库以及联想语义网中获取，若信息缺失，则为空集。

定义 7. 词语相似：将 $\text{sim}(a, b)$ 定义为两个词语相似的公式：

$\text{sim}(a, b) = 1$ 时，两个词语相同；

$\text{sim}(a, b) = 0$ 时，两个词语不相似；

$0 < \text{sim}(a, b) < 1$ ，两个词语存在相似的可能

根据联想语义网理解的结果，我们对由于信息缺失而导致无法得出理解结果的两种情况，使用认知缺省推理得出理解结果。

1. 缺省情况为 Cogbank 属性库与自主构建的属性库中均不存在源域或目标域的属性，推理过程如下：

(1) 获取汉语隐喻认知模型 $\langle S, T, S_i, T_i \rangle$ 作为主体信念集

S 为源域， T 为目标域；

$S_i = \{s_i | s_1, s_2, s_3, \dots, s_i\}$, $T_i = \{t_i | t_1, t_2, t_3, \dots, t_i\}$ 分别表示源域、目标域可获取的认知属性集合，若无则记为空集。

主体信念集合由汉语隐喻认知模型中的源域、目标域以及对应的属性组成，表示为 $W = \{a_i | a_1, a_2, \dots, a_i\}$ 。

(2) 构建候选属性集

若缺省域为源域（目标域），则从同义词词林获取该域同义词集合 C ，然后获取同义词的相关属性集合作为候选属性集 $P = \{p_1, p_2, p_3 \dots p_m\}$ ；

(3) 通过认知缺省模型 $\langle W, D, R \rangle$ ，开始进行认知缺省推理。

W 为主体信念集； D 为缺省规则集； R 是为 w 上的一个二元关系，用于认知过程中对缺省规则的判断。这里将被认可的属性组成理解集合 $F = \{f_1, f_2, \dots, f_m\}$ 。
 K, B 为知道和相信算子。判断过程如下：

若 $W \cap P \neq \emptyset$ ，则 $K_w \text{Rel}(s_i, p_m)$ 为真， $p_m \subseteq KF$

表示候选属性集 $P = \{p_1, p_2, p_3 \dots p_m\}$ 可以在主体信念集 W 中找到相关知识，则认为主体信念集知道源域知识 s_i 与候选属性信息 p_m 是语义相关的，则认为该结果是符合的，进入集合 F ；

若 $W \cap P = \emptyset$ ，且 $(Ka_i: M d_m) \Rightarrow Bp_m$ ，则 $M_w \text{Rel}(s_i, p_m)$ 为真， $p_m \subseteq BF$

表示候选属性集 $P=\{p_1, p_2, p_3 \cdots p_m\}$ 在主体信念集 W 中无法找到相关知识, 通过认知缺省推理规则判断候选属性 p_m 与主体信念集中的知识 a_i 是否违背语义限制 d_m , 即是否与源域及其情感信息相违背。若不违背, 则认为源域知识 s_i 与候选属性 p_m 可能存在相关关系, 进入集合 F , 并相信该结果; 否, 则不进入该集合。

(4) 选取最终释义 Z

使 $\sum_{i=0}^n \text{sim}(f_m, s_i)$ 的值最大, 即 $Z^* = \text{argmax}_{zi} \sum_{i=0}^n \text{sim}(f_m, s_i)$, 得出最后的理解结果。

例 1. 还有几家的红白大礼, 至少要二三千两银子用。 — 《红楼梦》

认知缺省推理过程如下:

(1) 首先, 获取隐喻四元组 $\langle S, T, S_1, T_1 \rangle$ 为主体信念集:

$S=\{“婚礼”, “丧礼”\}$, $T=\{“红礼”, “白礼”\}$;

$S_1=\{“庄重”、“美丽”、“热闹”、“喜庆”\cdots, “悲伤”、“隆重”、“简朴”\cdots\}$,

$T_1=\emptyset$

(2) 构建候选属性集

该句源域存在信息缺省情况, 则从同义词词林获取该域同义词的属性集合作为候选属性集 $P=\{p_1, p_2, p_3 \cdots p_m\}$;

$C_{\text{红礼}}=\{“庆典”、“盛典”、“婚礼”、“厂庆”\cdots\}$

$C_{\text{白礼}}=\{“奠基礼”、“葬礼”、“丧礼”\cdots\}$

$P_{\text{红礼}}=\{“好玩”、“喜庆”、“美丽”、“热闹”\cdots\}$

$P_{\text{白礼}}=\{“悲伤”、“隆重”、“简朴”\cdots\}$

(3) 认知缺省推理

$W_{\text{红礼}}\{“婚礼”、“红礼”、“庄重”、“美丽”、“喜庆”\cdots\} \cap P_{\text{红礼}}\{“好玩”、“喜庆”、“美丽”、“隆重”\cdots\}=\{“美丽”、“喜庆”、“隆重”\cdots\}$,

则 $K_w \text{Rel}(\{“庄重”、“美丽”、“喜庆”\cdots\}, \{“美丽”、“喜庆”\})$

$=(K\{“庄重”、“美丽”、“喜庆”\cdots\} \wedge K(\{“庄重”、“美丽”、“喜庆”\cdots\})$

$$\Rightarrow \{ \text{“美丽”、“喜庆”} \}) \Rightarrow K \{ \text{“美丽”、“喜庆”} \}$$

$$\Rightarrow \{ \text{“美丽、喜庆...”} \} \subseteq KF_{\text{红礼}}$$

$W_{\text{白礼}} \{ \text{“丧礼”、“白礼”、“悲伤”、“隆重”、“简朴”...} \} \cap P_{\text{白礼}} \{ \text{“悲伤”、“隆重”...} \} = \{ \text{“悲伤、隆重...”} \},$

则 $K_{\text{W}} \text{Rel}(\{ \text{“悲伤”、“隆重”、“简朴”...} \}, \{ \text{“悲伤”、“隆重”...} \})$

$= (K \{ \text{“悲伤”、“隆重”、“简朴”...} \} \wedge K(\{ \text{“悲伤”、“隆重”、“简朴”...} \}) \Rightarrow \{ \text{“悲伤”、“隆重”...} \}) \Rightarrow K \{ \text{“美丽”、“喜庆”} \}$

$$\Rightarrow \{ \text{“悲伤、隆重...”} \} \subseteq KF_{\text{白礼}}$$

(4) 选取最终结果。

$$Z_{\text{红礼}} = \text{argmax}_{Zi} \sum_{i=0}^n \text{sim}(\text{“喜庆”}, \text{“庄重、美丽、喜庆...”}) = \text{喜庆}$$

$$Z_{\text{白礼}} = \text{argmax}_{Zi} \sum_{i=0}^n \text{sim}(\text{“悲伤”}, \text{“悲伤、隆重、简朴...”}) = \text{悲伤}$$

得出理解结果：红礼是喜庆的，白礼是悲伤的。

2. 若缺省情况为无法在联想语义网以及同义词词林中，找到属性或动作的同义关系，则推理过程如下：

(1) 获取汉语隐喻认知模型 $\langle S, T, S_i, T_i \rangle$ 作为主体信念集

考虑源域同义关系缺失情形如下（目标域相反即可）：

特征型汉语隐喻认知模型获取步骤同缺省情况 1 的第一步；

关系型汉语隐喻认知模型中 S 为源域动作， T 为目标域场景信息，

$S_i = \{ (c_1, e_1), \dots, (c_i, e_i) \}$ 表示源域动作相关的场景，

$T_i = \{ t_1, t_2, t_3, \dots, t_n \}$ 表示目标域可能动作的集合，

主体信念集合由源域、目标域以及目标域相关知识（属性或场景信息）组成，表示为 $W = \{ a_i | a_1, a_2, \dots, a_i \}$ 。

(2) 构建候选属性集

将目标域属性赋予候选动作集合 P 。

$$P = \{p_1, p_2, p_3 \cdots p_m\} = T_i \{t_1, t_2, t_3, \cdots t_i\}$$

(3) 通过认知缺省模型 $\langle W, D \rangle$, 开始进行认知缺省推理。

W 为主体信念集; D 为缺省规则集; R 是为 w 上的一个二元关系, 用于认知过程中对缺省规则的判断。这里将被认可的动作组成理解集合 $F = \{f_1, f_2, \cdots f_m\}$ 。 B 为相信算子。

若 $(Ka_i: M d_m) \Rightarrow Bp_m$, 则 $M_w \text{Rel}(s_i, p_m)$ 为真, $p_m \subseteq BF$

表示为在主体信念集 W 中, 若知识 a_i 与候选动作 p_m 未违反语义限制 d_m (即场景信息), 则认为源域知识 s_i 与候选动作 p_m 是可能语义相关的, 因此将该动作纳入集合 F 中, 并相信该结果。

(4) 选取最终释义

使 $\sum_{i=0}^n \text{sim}(f_m, s_i)$ 的值最大, 即 $Z^* = \text{argmax}_{Zi} \sum_{i=0}^n \text{sim}(f_m, s_i)$, 得出最后的理解结果。

例 2. 生活是个刽子手。

认知缺省推理过程如下:

(1) 汉语隐喻认知模型 $\langle S, T, S_i, T_i \rangle$

$S =$ “刽子手”, $T =$ “生活”

$T_i = \{$ “真实”、“简单”、“质朴”、“残忍” $\cdots\}$, $S_i =$ “绝情”

主体信念集 $W = \{$ “刽子手”、“生活”、“绝情” $\}$ 。

(2) 构建缺省信息集

由于源域的属性“绝情”在联想语义网以及同义词词林中都无法搜寻到同义关系, 则将目标域属性赋予候选属性集合 P 。

$P = \{$ “真实”、“简单”、“质朴”、“残忍” $\cdots\}$

(3) 认知缺省推理

获取到主体信念集以及候选属性集后, 通过判断得出集合 F 。

$(K\{\text{“刽子手”、“生活”、“绝情”}\}: M\{\text{“人”、“状态”、“消极”}\}) \Rightarrow B\{\text{“残忍”、“残酷”}\}$

$\Rightarrow M_w \text{Rel}(\text{“绝情”}, \text{“残忍”})$ 为真

$\Rightarrow M_w \text{Rel}(\text{“绝情”}, \text{“残酷”})$ 为真

$\Rightarrow \{\text{“残忍”}, \text{“残酷”}\} \subseteq BF$

(4) 得出理解结果

$$Z_{\text{生活}} = \operatorname{argmax}_{zi} \sum_{i=0}^n \operatorname{sim}(\text{“绝情”}, \text{“残忍”}) = \text{残忍}$$

得出理解结果，生活是残忍的。

4.5 实验与评估

4.5.1 测试集

本文主要处理汉语隐喻中的名词隐喻和动词隐喻，并根据隐喻句的特点划分为特征相似型和关系相似型汉语隐喻句。特征型汉语隐喻共 150 句，由名词隐喻组成，关系型汉语隐喻共 150 句，由名词隐喻和动词隐喻组成。通过基于联想语义网的方法对汉语隐喻进行处理，未得到理解结果的句子自动划分到认知缺省逻辑模块。表 4.8 为本文测试集的分布情况。附录 A 和附录 B 列出了部分测试集以及对应汉语隐喻理解的结果（“*”表示缺省信息的汉语隐喻句，斜线表示与文化相关的汉语隐喻句）。

表 4.8：测试集分布

汉语隐喻类型	基于联想语义网的理解	认知缺省逻辑
特征型汉语隐喻	139	11
关系型汉语隐喻	132	18

4.5.2 评估方法

本文采用人工评价的方法对理解结果的认同度进行评估,并邀请 5 位母语为汉语,且从事自然语言处理研究的学生对汉语隐喻的理解结果进行评估。评估之前,我们对 5 位评估者进行评价一致性实验^[76]。给出相同的隐喻测试问卷,要求他们对其中的句子进行判断,是否为隐喻表达。根据测试的结果,5 位评估者的标注一致性 K 为 0.68,符合评价一致性的标准。因此,认可 5 位评估者对测试集的评估。

评价过程中,我们要求评估者依照以下原则对理解结果进行评估:

1. 根据自己的理解,对汉语隐喻理解的结果进行评估。
2. 根据认同度由低到高进行标注为 1-5。其中,1 代表“不能认同”;2 代表“较难认同”;3 代表“中立”;4 代表“可以认同”;5 代表“非常认同”。

考虑存在理解的差异性,本文通过设置认同度评分来对汉语隐喻理解的结果进行评估,将评估结果细分为 5 个程度,降低个体认知对理解结果带来的影响。

4.5.3 结果分析

通过对 5 位评估者的评估结果进行平均求和,若汉语隐喻理解结果的认同度大于 3,则认为该结果是被认可的,反之则不被认可。由此得出,150 句特征型汉语隐喻句中 140 句是被认可的,10 句不被认可(其中属性缺省 11 句,11 句都被认可);150 句关系型汉语隐喻句中 127 句可以被认可,23 句不被认可(其中信息缺省 18 句,15 句被认可)。表 4.9 为本文方法的准确率。

表 4.9: 本文方法的准确率

	联想语义网 (Cogbank)	联想语义网 (结合自建知识库)	联想语义网 (结合认知缺省推理)
特征型隐喻	74.6%	86%	93%
关系型隐喻	—	75%	85%

为了测试联想语义网对于汉语特征型隐喻理解和关系型隐喻理解的性能,我

们采用以下基准值作为比较（由于未有关于名词关系型隐喻的理解方法，基准只比较动词关系型隐喻的结果）：（1）对于特征型汉语隐喻，未将情感作为约束条件；（2）对于特征型和关系型汉语隐喻，使用 WordNet 作为语料库来进行理解计算的。通过人工对基准的理解结果进行评估，表 4.10 对比了本文方法与基准值准确度。

表 4.10： 准确度对比

	汉语隐喻	基准 1	基准 2	联想语义网
准确度	特征型隐喻	62%	79%	86%
	关系型隐喻	—	85%	87.5%

通过从表 4.9 的结果可以看出，在增加自建知识库信息之后，特征型隐喻的准确率有了一定的提升。且大部分汉语隐喻的理解结果都是在可认同程度内的。结合联想语义网与认知缺省推理的方法，在汉语隐喻的理解结果上也有较好的准确率。在表 4.10 中，基于联想语义网的汉语隐喻理解系统的性能也超过了两个基准值，尤其是采用汉语语料库在汉语隐喻的文化理解方面有一定程度的提升。但是仍存在以下几点不足之处。

1. 语料库的影响。本文的基础语料信息由同义词词林、Cogbank 属性库组成。虽然它们涵盖了大部分的知识信息，但仍存在知识不全的情况导致理解结果的偏差。比如，汉语隐喻句“男儿是泥做的骨肉”，结合 Cogbank 属性库和自建属性库，获取“泥土”的属性为：{“质朴”、“平凡”、“朴实”…}，从而得出理解结果，男人是厚实的。但在红楼梦中，该句意为男人是混浊的。由此可以看出，汉语的理解过程会受到语料库的约束。

2. 语境问题。由于隐喻理解结果的多样性导致不同语境下的理解结果是多样化的。比如，对于特征型隐喻“生活是杯茶”，我们通过源域与目标域的联想，得出理解结果为：生活是简单的。该句的认可度均分仅为 2.6，由于评估者理解的角度不同，它的理解结果还可以为“生活是苦涩的”或“生活是平淡的”等等。因此，要想提高汉语隐喻理解的准确性，仍需要大量语境信息的加持。

3. 关系型汉语隐喻的准确率仍有提升的空间。本文融合了动词语义距离与图式信息作为动词相似度计算的方法。虽能有效地辅助动词的筛选，但存在由于缺

失图式信息导致理解结果出现偏差的情况。比如，隐喻句“绑架银行”，动词网得出的理解结果为“成立银行就像绑架人”。由于缺失绑架的动词图式信息，使得相似度计算中缺失图式部分的相似值，从而导致理解出错。其次，本文对其动词图式的划分较为单一。单个动词可以拥有多个图式信息，本文仅选取最具代表性的图式信息，从而减少了其它相似的信息。

第五章 总结与展望

5.1 本文的主要贡献

本文的主要工作是对汉语隐喻进行研究。结合认知学理论以及隐喻学理论^{[77][79][86]}，本文概述了汉语隐喻的理解过程以及特点，并依据特点对症下药，实现汉语隐喻的自动理解和推理过程。根据认知隐喻学^{[56][57]}和文化语义学^[55]，我们认为联想作为隐喻和文化的桥梁，是解决汉语隐喻理解任务中文化难点的强大工具。因此，本文将联想作为汉语隐喻理解的核心机制，并使用同义词词林来构建联想语义网，为本文的汉语隐喻理解任务打下基础。其次，我们根据认知学理论对隐喻相似性的研究，对汉语隐喻句进行划分，并基于联想语义网相似度算法实现隐喻的自动理解过程。最后，对由于缺失信息导致未能进行理解的隐喻句，本文使用认知缺省逻辑推理出符合认知的理解结果。本文主要有如下几点贡献：

1. 通过构建联想语义网，来为汉语隐喻理解任务提供文化和语义信息。联想语义网是以同义词词林作为核心语料，分为形容词联想语义网和动词联想语义网两部分进行构建。形容词网以相似关系作为词汇的扩展过程，引入词语抽象度约束词汇的联想过程。动词网以动词图式为核心，根据图式理论对动词的逻辑结构进行刻画，并自动获取论元对其进行填充。实验表明联想语义网中词汇的联想过程是符合人类认知过程，且包含一定文化信息的。

2. 依据 Gentner 结构映射理论^[6]，深入探讨汉语隐喻理解中的相似特性，并依据该特性将汉语隐喻句分为：特征相似型汉语隐喻和关系相似型汉语隐喻。本文认为不同类别的隐喻句的理解过程也是不尽相同的。因此提出基于形容词网和动词网的隐喻理解算法，来实现汉语隐喻的自动理解过程。实验表明，基于联想语义网的隐喻理解结果是有效的。

3. 由于语料库不总是全面的，因此存在信息缺失导致无法得出理解结果的情况。但人们往往在缺失信息的情况下，仍能对隐喻做出理解。因此我们同样希望计算机可以和人类一样做出相应的判断。本文引入认知缺省逻辑，并根据缺省情形制定相应推理模式，以解决信息缺失的问题，实现汉语隐喻的推理理解。实验表明，结合认知缺省推理的汉语隐喻理解模型在理解任务中有较好的准确率。

5.2 未来的研究方向

本文提出的联想语义网符合词语的联想过程,在汉语隐喻理解任务中也取得了较好的效果,但仍存在一些需要改进的地方。因此,本文提出以下几点不足之处:

1. 语料知识扩充的问题。首先,本文并不是通过自主构建的语料库进行信息提取,而是使用现有的语料库来获取知识。通过对本文实验结果的分析,我们发现汉语隐喻的理解较为依赖语料库。若语料库中信息缺失,理解结果则会大打折扣。而现有的语料库都不能将所需要的知识都包含在内。因此仍需要通过自主构建相关信息对语料库进行扩充。本文虽通过其他语料进行知识抽取,来服务理解工作,但抽取信息的质和量还有待提升,这是未来工作需要注意的地方。其次,知识每天都在更新,我们无法通过旧的语料库获取到新的知识。因此,联想语义网也不例外,若没有新知识的扩充,我们的模型很难在精度上保持长久地领先。所以在未来的工作中,我们仍需持续地对联想语义网的内容进行动态更新。

2. 基于联想语义网的相似度算法仍有改进的空间。在特征型隐喻理解中,基于形容词网的相似度算法虽然动态结合了形容词网的语义距离和词汇情感,可以有效地约束理解的过程。但对于其中语义距离的参数并未进行调整,动词网中也是如此。因此对语义距离的参数的动态调整是我们未来还需改进的。其次,动词网相似度算法中动词图式相似部分。由于图式本身的复杂性和多样性,我们仅仅划分了两个大类,并没有更细致地划分。因此,在未来工作中,我们将更深入的扩展图式相似部分来完善基于联想语义网的相似度算法。

3. 汉语理解任务中的语境问题。由于隐喻本身存在歧义性,在不同语境下的理解结果是有偏差的。比如,“她是冷的”,通过人外在属性可以理解为“她的身体是冷的”;通过人的性格可以理解为“她是冷漠的”。理解的角度不同导致理解结果不同。因此对隐喻的正确理解是无法脱离语境存在的。在未来工作中,如何利用语篇信息来突显模型的动态性,是我们的改进方向之一。

4. 汉语隐喻理解任务中的文化问题。通过本文实验,我们发现基于联想语义网的汉语隐喻理解是有成效的。但在遇到一些生僻的词语时,由于缺乏与词语相关的语义信息,导致隐喻理解任务无法顺利进行。因此,我们认为文化知识库的构建是十分有必要的。在未来的工作中,我们也会致力于汉语文化知识库的构建

工作，并对汉语隐喻与文化进行更深入的研究。

参考文献

- [1] 李白. 望庐山瀑布[DB/OL]. <https://so.gushiwen.org/gushi/shijing.aspx>
- [2] 佚名. 诗经[DB/OL]. https://www.gushiwen.org/GuShiWen_8ef0998901.aspx
- [3] 张爱玲. 红玫瑰与白玫瑰[M]. 2013.
- [4] Lakoff G, Johnson M. Metaphors we live by[M]. University of Chicago press, 2008.
- [5] Fauconnier G, Turner M. Conceptual integration networks[J]. Cognitive science, 1998, 22(2): 133-187.
- [6] Gentner D. Structure-mapping: A theoretical framework for analogy[J]. Cognitive science, 1983, 7(2): 155-170.
- [7] Fass D. met*: A method for discriminating metonymy and metaphor by computer[J]. Computational Linguistics, 1991, 17(1): 49-90.
- [8] Martin J H. A computational model of metaphor interpretation[M]. Academic Press Professional, Inc., 1990.
- [9] Mason Z J. CorMet: a computational, corpus-based conventional metaphor extraction system[J]. Computational linguistics, 2004, 30(1): 23-44.
- [10] Shutova E, Van de Cruys T, Korhonen A. Unsupervised metaphor paraphrasing using a vector space model[C]. Proceedings of COLING 2012: Posters. 2012: 1121-1130.
- [11] Bollegala D, Shutova E. Metaphor interpretation using paraphrases extracted from the web[J]. PloS one, 2013, 8(9).
- [12] Su C, Huang S, Chen Y. Automatic detection and interpretation of nominal metaphor based on the theory of meaning[J]. Neurocomputing, 2017, 219: 300-311.
- [13] Su C, Tian J, Chen Y. Latent semantic similarity based interpretation of Chinese metaphors[J]. Engineering Applications of Artificial Intelligence, 2016, 48: 188-203.
- [14] Do Dinh E L, Gurevych I. Token-level metaphor detection using neural networks[C]. Proceedings of the Fourth Workshop on Metaphor in NLP. 2016: 28-33.
- [15] Sun S, Xie Z. Bilstm-based models for metaphor detection[C]. National CCF Conference on Natural Language Processing and Chinese Computing. Springer, Cham, 2017: 431-442.
- [16] Bizzoni Y, Ghanimifard M. Bigrams and BiLSTMs Two neural networks for sequential metaphor detection[C]. Proceedings of the Workshop on Figurative Language Processing. 2018: 91-101.
- [17] Swarnkar K, Singh A K. Di-LSTM contrast: A deep neural network for metaphor detection[C]. Proceedings of the Workshop on Figurative Language Processing. 2018: 115-120.
- [18] Halliwell S. The Poetics of Aristotle: translation and commentary[M]. UNC Press Books, 1987.
- [19] Richards I. The philosophy of rhetoric[M]. New York Oxford University Press, 1937.
- [20] Grady J. Foundations of meaning: Primary metaphors and primary scenes[J]. 1997.
- [21] Ortony A. Metaphor: A multidimensional problem[J]. Metaphor and thought, 1979: 1-16.
- [22] Cirstea B I, Chiru C G. Metaphor Detection[C]. 2013 19th International Conference on Control Systems and Computer Science. IEEE, 2013: 210-217.
- [23] Semino E. MIP: a method for identifying metaphorically used words in discourse[J]. Metaphor and symbol, 2007, 22(1): 1-39.

- [24] Amin T G, Jeppsson F, Haglund J, et al. Arrow of time: Metaphorical construals of entropy and the second law of thermodynamics[J]. *Science Education*, 2012, 96(5): 818-848.
- [25] Klebanov B B, Flor M. Word association profiles and their use for automated scoring of essays[C]. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2013: 1148-1158.
- [26] Stefanowitsch A. Words and their metaphors: A corpus-based approach[J]. *Trends in Linguistics Studies and Monographs*, 2006, 171: 63.
- [27] Reining A, Lönneker B. Corpus-driven metaphor harvesting[C]. *Proceedings of the Workshop on Computational Approaches to Figurative Language*. 2007: 5-12.
- [28] Narayanan S S. KARMA: Knowledge-based active representations for metaphor and aspect[J]. 1999.
- [29] Steinhart E. NETMET: A program for Generating and Interpreting Metaphors[J]. *Computers and the Humanities*, 1994, 28(6): 383-392.
- [30] Ovchinnikova E, Israel R, Wertheim S, et al. Abductive inference for interpretation of metaphors[C]. *Proceedings of the Second Workshop on Metaphor in NLP*. 2014: 33-41.
- [31] Ortony A, Stock O, and Slack J. Communication from an artificial intelligence perspective: theoretical and applied issues[M]. Springer Science & Business Media, 1992.
- [32] Gordon J, Hobbs J R, May J, et al. High-precision abductive mapping of multilingual metaphors[C]. *Proceedings of the Third Workshop on Metaphor in NLP*. 2015: 50-55.
- [33] Terai A, Nakagawa M. A neural network model of metaphor understanding with dynamic interaction based on a statistical language analysis: targeting a human-like model[J]. *International Journal of neural systems*, 2007, 17(04): 265-274.
- [34] 寺井あすか, 中川正宣. 確率的概念構造に基づく比喩理解の計算モデル—日本語コーパスの統計解析を用いて—[J]. *認知科学*, 2010, 17(1): 129-142.
- [35] Glucksberg S, Keysar B. Understanding metaphorical comparisons: Beyond similarity[J]. *Psychological review*, 1990, 97(1): 3.
- [36] Shutova E, Van de Cruys T, Korhonen A. Unsupervised metaphor paraphrasing using a vector space model[C]. *Proceedings of COLING 2012: Posters*. 2012: 1121-1130.
- [37] Baroni M, Bernardini S, Ferraresi A, et al. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora[J]. *Language resources and evaluation*, 2009, 43(3): 209-226.
- [38] Shutova E. Automatic metaphor interpretation as a paraphrasing task[C]. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010: 1029-1037.
- [39] Bollegala D, Shutova E. Metaphor interpretation using paraphrases extracted from the web[J]. *PloS one*, 2013, 8(9).
- [40] Fellbaum C. WordNet[J]. *The encyclopedia of applied linguistics*, 2012.
- [41] Cilibrasi R L, Vitanyi P M B. The google similarity distance[J]. *IEEE Transactions on knowledge and data engineering*, 2007, 19(3): 370-383.
- [42] 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* (05): 24-29.
- [43] Dong Z, Dong Q, Hao C. HowNet and the Computation of Meaning[J]. 2006.
- [44] 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[J]. *软件学报*, 028(012): 3167-3182.

- [45] 贾玉祥, 俞士汶. 基于实例的隐喻理解与生成[J]. 计算机科学, 036(003):138-141.
- [46] Mao R, Lin C, Guerin F. Word embedding and wordnet based metaphor identification and interpretation[C]. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018: 1222-1231.
- [47] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [48] Fellbaum C. WordNet[J]. The encyclopedia of applied linguistics, 2012.
- [49] Schuler K K. VerbNet: A broad-coverage, comprehensive verb lexicon[J]. 2005.
- [50] Doran C, Egedi D, Hockey B A, et al. XTAG system-a wide coverage grammar for English[J]. arXiv preprint cmp-lg/9410010, 1994.
- [51] Fillmore C J, Johnson C R, Petruck M R L. Background to framenet[J]. International journal of lexicography, 2003, 16(3): 235-250.
- [52] 哈工大社会计算与信息检索研究中心同义词词林扩展版[ER/OL]. (2015-09-13), <http://www.datatag.com/data/42306/>.
- [53] Schule K K. VetbNet[EB/OL]. <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>
- [54] 梅家驹. 同义词词林[M], 上海辞书出版社, 1983
- [55] 马清华. 文化语义学[M]. 2000.
- [56] 武锐. 俄汉颜色词文化伴随意义对比初探[J]. 俄语学习, 1996(2).
- [57] 王小璐. 汉语隐喻认知的神经机制研究[D]. 浙江大学, 2007.
- [58] Hawkins J, Blakeslee S. On intelligence: How a new understanding of the brain will lead to the creation of truly intelligent machines[M]. Macmillan, 2007.
- [59] 刘益民, 张旭东, 程甫. 心理学概论[J]. 2006.
- [60] Berners-Lee T, Hall W, Hendler J A, et al. A framework for web science[J]. Foundations and Trends® in Web Science, 2006, 1(1): 1-130.
- [61] Su C, Tian J, Chen Y. An Adjective-Based Embodied Knowledge Net[M]//Natural Language Processing and Chinese Computing. Springer, Berlin, Heidelberg, 2014: 46-53.
- [62] 靳光瑾. 现代汉语动词语义计算理论[M]. 2001.
- [63] 邵敬敏. 动量词的语义分析及其与动词的选择关系[J]. 中国语文, 1996(2):100-109.
- [64] Wang X, Su C, Chen Y. A method of abstractness ratings for Chinese concepts[C]//UK Workshop on Computational Intelligence. Springer, Cham, 2018: 217-226.
- [65] 俞士汶, 朱学锋. 现代汉语语法信息词典规格说明书[J]. 中文信息学报, 1996, 10(2):1-22.
- [66] Li S, Zhao Z, Hu R, et al. Analogical reasoning on chinese morphological and semantic relations[J]. arXiv preprint arXiv:1805.06504, 2018.
- [67] 徐琳宏, 林鸿飞, 潘宇, 等. 情感词汇本体的构造[J]. 情报学报, 2008, 27(2): 180-185.
- [68] SnowNLP[DB/OL]. <https://pypi.org/project/snownlp/0.11.1/>
- [69] Johansson R, Nugues P. Dependency-based syntactic-semantic analysis with PropBank and NomBank[C]. CoNLL 2008: proceedings of the twelfth conference on computational natural language learning. 2008: 183-187.
- [70] Stanford-corenlp[EB/OL]. <https://pypi.org/project/stanford-corenlp/>
- [71] Wolff P H. The developmental psychologies of Jean Piaget and psychoanalysis[J]. Psychological issues, 1960.
- [72] Johnson M. The body in the mind: The bodily basis of meaning, imagination, and reason[M]. University of Chicago Press, 2013.

- [73] Lakoff G. Women, fire, and dangerous things: What categories reveal about the mind[M]. University of Chicago press, 2008.
- [74] Neo4j[DB/OL]. <https://neo4j.com/>
- [75] Webber J. A programmatic introduction to neo4j[C]//Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity. 2012: 217-218.
- [76] Sidney S. Nonparametric statistics for the behavioral sciences[J]. The Journal of Nervous and Mental Disease, 1957, 125(3): 497.
- [77] 束定芳. 隐喻学研究[M]. 2000.
- [78] 张旭明. 客体相似性对不同认知方式个体视觉工作记忆表征的影响——来自行为和 ERP 的证据[D]. 2013.
- [79] Kittay E F. Metaphor: Its cognitive force and linguistic structure[M]. Oxford University Press, 1989.
- [80] 李斌. cogbank 属性库[EB/OL]. <http://www.cognitivebase.com/>
- [81] 周昌乐. 认知逻辑导论[M]. 2001.
- [82] 张敏. 带约束的认识缺省理论的动态特性研究[D]. 厦门大学.
- [83] 董明楷, 蒋运承, 史忠植. 一种带缺省推理的描述逻辑[J]. 计算机学报(6):90-97.
- [84] Hintikka J, Hintikka K J J, Hintikka M B P. The logic of epistemology and the epistemology of logic: selected essays[M]. Springer Science & Business Media, 1989.
- [85] 吴梅红. 基于动态认知逻辑的多主体系统知识推理研究[D]. 厦门大学.
- [86] 胡壮麟. 认知隐喻学[M], 北京: 北京大学出版社. 2004.
- [87] 史鑫. 隐喻的思维本质与认知功能[J]. 解放军外国语学院学报, 2002, 25(4):41-44.
- [88] 彭文钊. 隐喻——认知与阐释[J]. 解放军外国语学院学报, 1999(1).
- [89] 李斌. 基于互联网的汉语认知属性获取及分析. 语言文字应用, 2012 年第 3 期: 134-143.
- [90] Jiaba[DB/OL]. <https://pypi.org/project/jieba/>
- [91] 陈宏朝, 李飞, 朱新华, et al. 基于路径与深度的同义词词林词语相似度计算[J]. 中文信息学报, 2016, 30(5):80-88.
- [92] Hilpinen R. Knowing that one knows and the classical definition of knowledge[J]. Synthese, 1970: 109-132.
- [93] Hendricks V F. Mainstream and formal epistemology[M]. Cambridge University Press, 2006.
- [94] 侯友兰. 语境在修辞中的作用[J]. 绍兴文理学院学报:哲学社会科学版, 1994(2):59-62.
- [95] Reiter R. A logic for default reasoning[J]. Artificial intelligence, 1980, 13(1-2): 81-132.
- [96] Palmer M, Wu Z. Verb Semantics and Lexical Selection[J]. Acl Proceedings of Annual Meeting on Association for Computational Linguistics, 2012:133--138.
- [97] 田嘉. 基于深层语义相似的隐喻理解研究[D]. 2015.
- [98] 郭艳华, 周昌乐. 自然语言理解研究综述[J]. 杭州电子工业学院学报, 2000(1):58-65.
- [99] John-Jules Ch. Meyer, Wiebe van der Hoek. Non-Monotonic Reasoning by Monotonic Means[C]. Logics in AI, European Workshop, JELIA '90, Amsterdam, The Netherlands, September 10-14, 1990, Proceedings. Springer, Berlin, Heidelberg, 1990.
- [100] John-Jules Ch. Meyer, Wiebe van der Hoek. A Modal for Nonmonotonic Reasoning[C]. Non-Monotonic Reasoning and Partical Semantics, Ellis Horwood, 1992, 37-77.
- [101] 李永礼, 范荣强. 缺省模态逻辑[J]. 兰州大学学报, 1994(03):77-81.

附录 A 特征相似型汉语隐喻理解结果

汉语隐喻句	汉语隐喻理解结果	认可度评分
时间是金钱	时间是重要的	4.6
人生是美梦	人生是精彩的	3.4
人生如浮萍	人生是随波逐流的	5
肤若凝脂	肌肤是光滑的	4.8
友谊是阳光	友谊是纯粹的	3.8
君子之交淡如水	君子的友谊是纯粹的	3.2
友谊是美酒	友谊是深厚的	4.2
友谊是鲜花	友谊是美好的	4.8
石油是黄金	石油是宝贵的	5
散文是花	散文是优美的	4.6
眼睛是星星	眼睛是明亮的	4.8
生命是小草	生命是脆弱的	3.6
生活是杯白开水	生活是平淡的	5
生命是花	生命是绚丽的	4.8
爱情是根线	爱情是缠绵的	3.8
爱情是甘露	爱情是甜的	4.8
山水如画	山水是美的	4.8
市场是草原	市场是广大的	5
青春是匹野马	青春是奔放的	4.4
青春是烟火	青春是美的	4.4
原子是太阳系	原子是繁杂的	4.2
母爱是水	母爱是细腻的	4.8
生活是赌博	生活是精彩的	2.4
洪水是猛兽	洪水是肆无忌惮的	4.8

你这个 狗奴才	奴才是百依百顺的	4.2
女人是水做的骨肉	女人是轻的	2.4
男人是泥做的骨肉	男人是厚实的	2.2
她真是一只铁公鸡	她是抠的	4.4
忽如一夜春风来， 千树万树梨花开	雪是白的	4.4
乔公二 女 秀所钟，秋水并蒂开 芙蓉	女子是娇艳的	5
千峰笋石 千株玉， 万树松萝 万朵云	山峰是高的	3.8
女人是白月光	女人是柔和的	4.4
思念是药	思念是苦的	4.6
思念是杯浓茶	思念是深沉的	4.8
思念是河	思念是长的	4.4
手指是熊掌	手指是粗的	4.8
幸福是毛毯	幸福是温软的	5
婚姻是城堡	婚姻是经久的	3.6
生活是个刽子手*	生活是残忍的	5
朱门酒肉臭，路有冻死骨*	朱门是显赫的	4.8
爱是人间四月天*	爱是温暖的	4.6
光阴如流水*	光阴是飞快的	4.6
人生是盛宴*	人生是盛大的	4.8
正义是人类良知的声音	正义是不屈不挠的	3.2
月亮是遗孀*	月亮是孤独的	4.8
内心那处清泉	内心是纯洁的	4
老子《 道德经 》，是一块珍藏了三千年的 古茶饼*	道德经是深远的	4.8
光阴如流水*	光阴是飞快的	4.4

还有几家 红白大礼 , 至少 还得三二千两银子用*	红礼是喜庆的	5
还有几家 红白大礼 , 至少 还得三二千两银子用*	白礼是悲伤的	5

“*”代表源域或者目标域属性缺失；斜线代表与文化相关的汉语隐喻句。

附录 B 关系相似型汉语隐喻理解结果

类别	汉语隐喻句	汉语隐喻理解结果	评分
名词隐喻	老师是园丁	老师教学生就像园丁培育花朵	5
名词隐喻	老师是蜡烛	老师讲课就像蜡烛点燃	3
名词隐喻	老师是灯塔	老师引导学生就像灯塔指引船只	5
名词隐喻	学生是种子	老师培养学生就像人播种种子	5
名词隐喻	船长是父亲	船长离开船就像父亲离开家	4.8
名词隐喻	黄河是母亲	黄河泛滥就像母亲离开	2.2
名词隐喻	总统是母亲	总统呼吁人民就像母亲抚养孩子	4.2
名词隐喻	流言是瘟疫	流言纷飞就像瘟疫流行	5
名词隐喻	婚姻是监狱	人开始婚姻就像人关押监狱	3.6
名词隐喻	知识是灯塔	知识学习人们就像灯塔引导船只	4.6
名词隐喻	医生是屠夫	医生治疗病人就像屠夫杀猪	4.2
名词隐喻	灵感是火花	引发灵感就像迸发火花	4.8
名词隐喻	理念是建筑	保护理念就像保持建筑	4.2
名词隐喻	理想是种子	理想出现就像种子萌发	5
名词隐喻	南有樛木，葛藟 累之*	葛藤缠绕树枝就像女子依靠丈夫	4.8
动词隐喻	船犁大海	船驶入大海就像犁犁田	4.4
动词隐喻	汽车喝汽油	汽车使用汽油就像人喝水	5
动词隐喻	人吸收经验	人获取经验就像海绵吸收水分	5
动词隐喻	马吃炮	马打炮就像人吃饭	2.6
动词隐喻	吃透汉字	人使用汉字就像人吃饭	4.6
动词隐喻	猎杀财富	抢夺财富就像猎杀动物	5
动词隐喻	集团转身	集团展开就像人转身	2
动词隐喻	股价跳水	股价下降就像运动员跳水	5
动词隐喻	品尝人生*	人享受人生就像人品尝食物	5

动词隐喻	爱情绽放	爱情发生就像花朵绽放	4.8
动词隐喻	经济滑坡	经济下降就像山体滑坡	5
动词隐喻	经济回暖	经济回复就像天气回暖	5
动词隐喻	进程起飞	进程开始就像飞机起飞	4.2
动词隐喻	大地颤栗	大地颤抖就像人颤栗	5
动词隐喻	大脑生锈	大脑麻痹就像铁生锈	5
动词隐喻	需求退潮	需求下降就像海水退潮	5
动词隐喻	咀嚼笑话	知道笑话就像咀嚼食物	4.6
动词隐喻	情感沸腾	情感激发就像水沸腾	4.6
动词隐喻	拥抱生活	享受生活就像拥抱人	5
动词隐喻	生育思想	产生思想就像生育孩子	5
动词隐喻	编织童话	人编童话就像人编织毛衣	4.6
动词隐喻	资产减肥	资产收缩就像人减肥	5
动词隐喻	消费增长跑过 储蓄*	消费增长超过储蓄就像人跑过人	5
动词隐喻	市场气氛在商业领域逐渐萌芽。	市场气氛增加就像种子萌芽	4.6
动词隐喻	引爆债务危机	引起危机就像引爆炸弹	5
动词隐喻	公司迁徙	公司搬迁就像候鸟迁徙	5
动词隐喻	资产触电	资产收缩就像人触电	5
动词隐喻	金融危机肆虐 全球市场	金融危机控制全球就像狂风肆虐 树林	5
动词隐喻	多边主义正在 走向死亡*	多边主义垮台就像人死亡	4.8
动词隐喻	货币涌入	货币增加就像河流涌入	4.8
动词隐喻	公司复活	公司失败就像人复活	1.8
动词隐喻	扼杀创新力*	抑制创造力就像扼杀生命	5
动词隐喻	杀入院线市场*	进入院线市场就像杀入军营	4.4

动词隐喻	包装歌星	打造歌星就像包装物体	5
动词隐喻	酝酿想法	萌生想法就像酝酿美酒	5

“*”表示动作信息缺失；斜线表示与文化相关的汉语隐喻句。

附录 C 形容词联想网代表性部分

目标形容词	一次扩展	两次扩展	三次扩展
白	洁白	素	淡雅
	反动	逆	叛逆
	纯洁	天真	纯真
	纯洁	高洁	—
	干净	清新	—
	无偿	无用	无效
矮	低	卑微	低人一等
	低	细微	轻微
	小	狭隘	—
黑	黑黝黝	昏黄	—
	黑黝黝	阴	阴险
	秘密	神秘	玄妙
	秘密	地下	非法
冷	寒冷	冰冷	冷漠
温暖	温和	温柔	和善
旧	老	悠久	遥远
黄	熟	烂熟	娴熟
	熟	成熟	—
红	热	热门	—
	吉利	吉庆	喜庆
	革命	—	—
甜	甘	甜美	舒适
	香甜	熟	成熟
	幸福	—	—

加粗表示该词与文化相关。

附录 D 动词联想网代表性部分

动词	动词扩展词	动词图式	动词逻辑结构	例子
吸收	吸收	1	在外转在内 (B/D, A/B)	在外转在内(养分, 植物)
	收	1	在外转在内 (B/D, A/B)	在外转在内(税, 国家)
	接受	1	在外转在内 (B/D, A/B)	在外转在内(批评, 人)
	回收	1	在外转在内 (B, A/B)	在外转在内(废品, 人)
下降	下降	2	(C, C, B, 向下) (D, D, D, 向下)	(天空, 陆地, 飞机, 向下) (-, -, 体重, 向下)
	降落	2	(C, C, B, 向下)	(天空, 陆地, 飞机, 向下)
	跌落	2	(D, D, D, 向下)	(高位, 地位, 大盘, 向下)
	跌	2	(D, D, D, 向下)	(高位, 地位, 大盘, 向下)
	贬值	2	(D, D, D, 向下)	(高位, 地位, 人民币, 向下)
集合	集合	3	集合(整体(A, A))	集合(整体(同学, 同学))
	合并	3	合并(整体(B, B))	合并(整体(公司, 公司))
	联合	3	联合(整体(A, A)) 联合(整体(B, B))	联合(整体(人, 人))
攻击	攻击	4	攻击(A/D, A/D)	攻击(他, 我)
	进攻	4	进攻(A/D, A/D)	进攻(军队, 国家)
	攻打	4	攻打(A/D, D)	攻打(军队, 国家)
	抨击	4	抨击(A, A/D)	抨击(人, 事)

“1”表示容器图式；“2”表示起点-目标-路径图式；

“3”表示整体-部分图式；“4”表示链接图式。

附录 E 词语 Cogbank 属性库与自建属性库对比

词语	Cogbank 属性库	自建属性库
思念	悠久、长	微痛、孤寂、甜美、苦、悲哀、深沉、悠久…
友谊	纯洁	真诚、深厚、真挚、亲密、牢固、短暂、纯洁…
市场	热闹、鱼龙混杂	庞大、大、广阔、巨大、稳定、紧张、热闹…
肌肤	水灵、温暖	完美、柔滑、幼嫩、水灵…
盒子	单调	漂亮、精致、宽、大、方…
心脏	默默无闻、重要、完美无缺、勤奋	重要、强壮、珍贵、健康…
芙蓉	大红大紫	娇艳、美丽
歌	优美、动人、动听、好听	快、慢、悲壮、忧郁、优美、欢快…
太阳系	古老	复杂、巨大
书	软、厚、结实、博大、杂	精美、精深、枯燥、有趣、广…
路	长、坦荡	远、长、荒凉、曲折、孤独…
灯	明亮	暖、柔和、昏暗、清晰、明亮…
钥匙	灵便	成功、合适、好、小…
声音	美好、漫无边际、轻快、稍纵即逝	嘶哑、微弱、古怪、响亮、优美、悦耳、清脆…
武器	可以、锐利	强大、锋利、锐利、厉害、先进…
房子	高大、大、高、朴实无华	暖和、漂亮、旧、小、豪华、舒适、简陋…
夜	黑、深沉、安静、神秘、孤独、宁静…	冷、寂静、深、黑、浩瀚、宁静、安静、神秘、孤独…

附录 F 部分词语获取的动作集合

场景信息	动作集合
老师	教, 讲, 学习, 鼓励, 讲课, 批评…
医生	诊断, 发现, 治疗, 抢救, 注射…
青春	飞扬, 洋溢、逝去、叛逆、萌动…
爱情	开始、发生、燃烧、幻灭、相信、破裂…
思路	创新、开阔、进行、转变、展开、阻滞…
流星	滑落、闪烁、划过、闪过、出现…
骆驼	驮、睡眠、跪下、运…
人生	享受、面对、改变、思考、开始、度过…
大脑	负责、控制、损伤、思考、分泌、进化…
危机	出现、解决、发生、处理、化解、加剧…
想法	改变、说出、提出、了解、坚持、表达…
精神	创新、奋斗、发扬、学习、重视、开拓…
局面	对峙、出现、分离、改变、控制、僵持…
事实	接受、承认、无视、否认、了解、相信…
文章	写、发表、读、喜爱、刊登、承担、剽窃…
理性	保持、丧事、具有、失去、需要、缺少…
阳光	刺眼、享受、喜欢、穿过、吸收、反射…
品牌	保护、推广、进入、塑造、增加、建立…
进程	加快、继续、促进、推动、缩小、受阻…
货币	发行、兑换、增长、贬值、升值、紧缩…
气氛	进行、融洽、感染、笼罩、改善、进入…
(人, 财富)	拥有、发现、得到、享受、争夺、占有…
(经济, 储蓄)	需要、下调、改善、帮助、增长、减少…
(船, 大海)	沉没、行驶、进入、前进
(汽车, 汽油)	有, 使用, 没有, 需要

