

话语意义的 可计算性研究

龙飞◎著

人民日报学术文库

人民日报
出版社

人民日报学术文库

话语意义的 可计算性研究

龙飞◎著

人民日报
出版社·北京

图书在版编目 (CIP) 数据

话语意义的可计算性研究 / 龙飞著. —北京: 人民日报出版社, 2020. 6

ISBN 978 - 7 - 5115 - 6415 - 3

I. ①话… II. ①龙… III. ①算法语言 IV. ①TP312

中国版本图书馆 CIP 数据核字 (2020) 第 091079 号

书 名: 话语意义的可计算性研究

HUAYU YIYI DE KE JISUANXING YANJIU

著 者: 龙 飞

出 版 人: 刘华新

责任编辑: 谢广灼

封面设计: 中联学林

出版发行: 人民日报出版社

社 址: 北京金台西路 2 号

邮政编码: 100733

发行热线: (010) 65369509 65363527 65369846 65369828

邮购热线: (010) 65369530 65363527

编辑热线: (010) 65369533

网 址: [www. peopledaily. com](http://www.peopledaily.com)

经 销: 新华书店

法律顾问: 北京科宇律师事务所 (010) 83622312

印 刷: 三河市华东印刷有限公司

开 本: 710mm × 1000mm 1/16

字 数: 150 千字

印 张: 14

版次印次: 2020 年 6 月第 1 版 2020 年 6 月第 1 次印刷

书 号: ISBN 978 - 7 - 5115 - 6415 - 3

定 价: 78.00 元

此为试读, 需要完整PDF请访问: www.ertongbook.com

作者简介

龙飞女，汉族，1983年5月生，2018年毕业于中国传媒大学文学院，获文学博士学位。主要研究方向：话语语言学、计算话语学。现任职于哈尔滨商业大学外语学院。主持黑龙江省哲学社会课题“面向计算的话语连贯关系的研究”以及校级后备人才课题“面向人工智能的商务话语因果关系的研究”等。

摘 要

智能计算的本质是基于自然语言的计算。在人工智能领域,计算机智能化需要以语言作为计算的媒介,模拟人类的智能。计算的原本意义是数字的处理和操作。语言的计算是通过规则和逻辑对词串和句子进行处理和操作。在人工智能高度发达的今天,话语意义的可计算性研究成为智能计算的重要话题。涉及话语意义计算的领域包括机器翻译、文本摘要、自动会话、机器阅读理解等。

大规模自然语篇的自动处理是话语意义计算的最终目的,如何处理意义问题是话语计算研究的核心任务。本研究从理论和实践层面主要探索以下两个问题:第一,话语连贯关系的可计算性。本研究从显式和隐式关系的识别入手,探讨如何将话语从线性的文字序列转换为可被计算机识别的有结构层次的语义关系;第二,话语意图的可计算性。研究通过话语外部的背景知识约束话语内部信息,从而实现话语意图的理解与计算。

针对第一个问题,本研究将汉语中显式连贯关系和隐式连贯

关系的识别，整合到一个统一的识别框架内，依据优先级进行识别，试图利用层层逼近的方式，提升计算机判定话语语义关系的准确性。

针对第二个问题，本研究从认知的视角探讨话语意图的计算，梳理出面向计算的话语意图理解框架。对于自然语言理解系统而言，模型只能作为一种理论上的可行方案，离算法化的实现还有一定距离。因此，本书进一步简化模型，并未试图覆盖意图推理的全部计算过程。实验的最终目的在于验证“实现框架”的可行性和可计算性，具体实验过程为：对于给定的一段文本或话语，在目标词确定的前提下，通过计算获得目标词的具体语境义或者以目标词为核心的简单句的语境义。

与句法分析和语义分析相类似，话语意义计算的研究过程是一项复杂的长期语言工程，需要语言学、计算机科学及认知科学等学科的协同研究才能逐渐完善。

关键词：话语意义，可计算性，连贯关系，语义识别，意图计算

ABSTRACT

The essence of Intelligent Computing is based on the natural language understanding. Therefore, the intelligence of computer needs to be emulated as human intelligence, and only then can it possess the ability of computing in the field of artificial intelligence (AI). The original meaning of computing is processing and operating numbers, while the research object of language computing is word strings and sentences based on the rule and logic method. With the rapid development of AI, the computability of discourse meaning has become the center of computational process, which has been applied to many fields, such as machine translation, text summarization, automatic session and machine reading comprehension. Meanwhile, it could provide more valuable information for lexicons, phrases and sentences.

The ultimate goal of discourse meaning computing is the automatic processing of large - scale natural language discourse, thus the core mission of discourse computing is to solve the problem of meaning. The

significance of this study lies mainly on a feasible method and systemic interpretation of computability for the meaning computing from discourse level. On the one hand, given the surface level of language, we focus on the transformation from linear text sequences to the structured semantic relations recognized by computer. On the other hand, the computing of discourse intention is realized by the constraints of literal meaning of context. Meanwhile, the recognition of discourse relation is the premise of the intention - understanding, both of which comprise computability of discourse meaning.

To investigate the first question, we integrate the recognition frame of implicit and explicit relations, which has been conducted by patterns as explicit connectives, lexical markers and semantic frame relations, according to its priority. In this way, we try to approach the results of semantic relations recognition more accurately.

To answer the second question, we summarize the framework of computation - oriented discourse intention from cognitive perspective. We also design the model of intention computability, in order to realize natural language understanding (NLU) based on context. This model is regarded as a theoretical and feasible plan at this stage for NLU system. We still have a way to go before reaching algorithmization, and we do not try to cover all the computational process of intention inference. Specifically, if the target word has been identified, the computer could compute the specific meaning of the target word or the intention meaning of the sentence that includes the target word, based on the

context – constrain.

The computation of discourse meaning is a complicated language engineering like syntactic and semantic analysis, which needs the joint efforts of linguistic study, computer science and other relevant subjects to improve gradually.

Keywords: *discourse meaning, computability, coherent relation, semantic recognition, intention computing*

目 录

CONTENTS

1 绪论	1
1.1 引言	1
1.2 选题缘起	2
1.3 本研究的意义和创新之处	3
1.3.1 研究意义	3
1.3.2 创新之处	4
1.4 本研究的内容和组织结构	5
1.5 术语界定	7
2 话语意义计算的相关研究	9
2.1 引言	9
2.2 话语意义研究的角度	10
2.2.1 话语意义的修辞学研究	10
2.2.2 话语意义的语用研究	11
2.2.3 话语意义的认知研究	13
2.2.4 话语意义的计算研究	15

2.2.5	话语意义研究的演进方向	16
2.3	自然语言处理领域对话语意义的相关研究	18
2.3.1	基于词汇的话语意义研究	18
2.3.2	基于话语结构的话语意义研究	20
2.3.3	基于背景知识的话语意义研究	29
2.4	本章小结	33
3	话语意义的研究内容及可计算性问题	35
3.1	引言	35
3.2	话语意义的概念界定	36
3.3	话语意义可计算性研究的内容	37
3.4	基于认知的话语意义的计算	38
3.4.1	话语认知计算的理据性	38
3.4.2	计算的心智观	40
3.5	面向计算的话语意义研究	43
3.5.1	计算的本质	43
3.5.2	话语意义的可计算性	45
3.6	本章小结	47
4	面向计算的话语语义关系的识别与理解	48
4.1	引言	48
4.2	所指判定	49
4.2.1	词汇链的构建	50
4.2.2	事件链的构建	56

4.2.3	同指消解的方法	63
4.3	面向计算的话语连贯关系的识别	65
4.3.1	话语标记语标示局部连贯与整体连贯关系	66
4.3.2	话语单元识别	74
4.3.3	话语连贯关系的识别方法	84
4.4	话语语义的表征方法	106
4.4.1	基于线状图语义表征方法	107
4.4.2	基于树状图的语义表征方法	110
4.4.3	基于盒状图的语义表征方法	115
4.5	本章小结	119
5	话语意图的计算	120
5.1	引言	120
5.2	话语意图计算的途径与方法	121
5.2.1	话语意图的认知计算	122
5.2.2	话语意图计算的方法	142
5.3	话语意图计算的实现框架	144
5.3.1	话语意图计算中语境的形式化研究	144
5.3.2	话语语义空间的构造	154
5.3.3	话语知识库的构建	155
5.3.4	动态语境知识库的构建	156
5.3.5	话语意图的推导	159
5.3.6	话语意图理解实现流程	160
5.4	话语意图计算模型	161

5.4.1	话语意图理解模型的设计	162
5.4.2	语料处理和知识库的构建	165
5.4.3	确定目标词的意图	169
5.4.4	实验结果及分析	175
5.5	本章小结	183
6	结论与展望	184
6.1	本研究的主要发现	184
6.1.1	话语语义关系的识别理解	184
6.1.2	以语义框架作为背景语境约束话语意图	185
6.1.3	话语意图计算的模型	186
6.2	本研究的局限性	186
6.2.1	隐式连贯关系的识别判定的盖然性	186
6.2.2	话语意图计算模型的局限	187
6.3	研究展望	188
	参考文献	190

表目录

表 1	RST 中常见的修辞关系类型	22
表 2	显式关系分类	86
表 3	隐式关系分类	94
表 4	目标词语义词典	167
表 5	框架语义知识词典	167
表 6	短语规则	168

图目录

图 1	本研究的内容概览	5
图 2	认知背景知识在话语理解中的动态过程	14
图 3	RST 分析结果：篇章修辞结构树示意图	23
图 4	篇章图树库 (Discourse Graph Bank)	24
图 5	修辞结构理论标注实例	25
图 6	心智计算理论的基本思想	42
图 7	语篇连贯中的事件链	60
图 8	篇章连贯关系强弱度	62
图 9	基于短语结构分析的语义单元树	77
图 10	语义组块识别流程图	84
图 11	显/隐式关系识别流程图	105
图 12	例句的话语结构	106
图 13	依存关系图 Dependency Tree	109
图 14	语篇宏观结构示意图	111
图 15	示例中修辞结构树	112

图 16	示例中宾州篇章树库 (PDTB)	114
图 17	语篇表征结构 (DRS)	116
图 18	分段式语篇表征结构 (SDRS)	118
图 19	认知计算流程图	141
图 20	语境构成的集合论表示	148
图 21	词、句语义构成的集合论表示	150
图 22	话语世界到认知空间与语义空间的映射	154
图 23	动态上下文 & 背景语境词典的构建	156
图 24	现场语境知识词典的构建	157
图 25	基于语境的话语含义的推导	159
图 26	话语意图理解实现流程	160
图 27	目标词语义词典与框架语义词典的对应关系	163
图 28	ICTCLAS 主界面	164
图 29	话语意图理解模型	165
图 30	意图理解实现流程	171
图 31	例 (1) 解析结果	177
图 32	例 (2) 解析结果 1	178
图 33	例 (2) 解析结果 2	178
图 34	例 (2) 解析结果 3	179
图 35	例 (3) 解析结果	179
图 36	例 (4) 解析结果	180
图 37	例 (5) 解析结果	180
图 38	例 (6) 解析结果	181

1 绪论

1.1 引言

在词汇、句法层面，自然语言处理的一些理论模型和研究方法已经应用于计算机语义计算处理中。（宗成庆，2013）但是，话语层面的自然语言处理还未成熟。原因有两点，首先，语言的使用单位是话语，相对于词汇、语法单位，其结构以及意义的处理过程更加复杂。词汇、短语和句子构成话语意义计算的基础，除此之外，构成话语的各个单元之间的语义关系以及相互联系的方式也是话语意义计算必不可少的研究内容。其次，自然语言处理的话语基础理论研究和涉及背景知识的语用计算研究还较为薄弱。对于计算机自然语言处理而言，汉语话语语义关系的识别以及语境的计算还处于初级阶段。

计算机自然语言处理水平还远远不能满足信息化社会的需