

# 基于类比的计算语义处理机制

## Computational Semantics based on Analogy Principle

刘海涛 LIU Haitao

(青海铝厂计算机中心, 西宁810108)

自然语言的计算机处理是人类语言处理机制的仿真。对于语义的研究导致了计算语义学的产生, 它的目的在于研究自然语言语义形式化的理论和方法, 狭义说计算语义学是将语义分析看作为一种演算过程, 它通过逻辑的方法处理语义问题, 广义上讲, 计算语义学研究利用计算机来处理 and 仿真人类语义处理机制的方法和理论, 特别是歧义问题的处理和消解。作者认为计算语义学的广义概念可能更适宜于大规模、真实文本处理的应用。

在现有的理论和计算语言学中语义的处理一般是建立于语义可分解性前提之上的, 两种常用的方法是“义素”分析法和“语义场理论”, 二者由于其固有的问题, 都难以满足处理大规模真实文本的要求。

我们说语言在各个层次充满了模糊和不确定, 指的正是语言语义方面的问题, 因为语言的句法方面是不难通过形式化的方法来描述的, 乔姆斯基的生成语法理论和特思尼耶尔的依存语法理论都是这一方面的杰出代表。就语言符号的理解和分析而言, 内容是一种隐含的、模糊的东西, 它只有通过形式才能被感知。不幸正在于此, 一种形式可能表示多种内容, 这就是计算机在处理自然语言时遇到的最大难题——歧义问题。

对于有歧义的语句, 理解的任务就是从多种结构中选出最适宜的和最可能的结构, 注意我们在这里使用了“适宜”和“可能”这两个非绝对的词, 目的在于说明在语言理解领域没有什么绝对的正确, 而只有相对的“可能”, 这是我们引入“类比语义”处理机制的理据之一。语义的不可分解性和隐含性、歧义问题的复杂性、语言理解的无限性、语义的关联性、大规模真实文本处理的迫切性等, 所有这些因素交织在一起, 使得我们必须寻求新的语义处理方法和机制。

类比是人类智能的重要组成部分, 它在认知、科学发现、问题求解、推理等方面具有重要作用。所谓类比, 就是通过寻找与确定两个或两类事物之间的对应关系, 根据对其中一方的特征和规律的认识来推测另一方的特征和规律。在语言理解中, 类比方法也是一种主要的手段。我们根据维特根斯坦的语言哲学“用法论”曾提出“某一语言单位的意义即它的全部语境关系”的看法, 这一是进一步建立类比语义处理机制的基础。在我们的方法中语义是隐含于其语境关系之中的, 这种看法体现了语义的不可分割性, 同时也可以满足目前大规模真实文本处理的需要。类比推理的基础是建立对象的形式化描述, 我们采用句法语义标注的手段去分析真实文本, 从而使得真实文本语料库变为无结构性歧义的语言知识库。通过基于模糊集和概率统计原理的相似性度量方法, 将源文和语言知识库中的树结构进行比较和对应, 进而达到分析理解文本的要求。可以说类比语义学正是利用语料库方法构建自然语言处理系统的理论基础之一。类比语义处理机制或理论的提出是与目前理论及计算语言学界越来越重视语料库语言学的潮流相适应的, 作为一种新的语义处理方法和理论值得进一步研究, 提高其可操作性和可实现性。