专用机器翻译系统SCDOS的技术研究与实现

张琳琳　王能忠
(西南师范大学计算机科学系，重庆，630715)

随着计算机的普及和发展，在运算速度和存储能力迅速提高之后，输入输出的人机界面变得越来越重要。语音I／O成为计算机最好的接口方式之一。DOS（Disk Operating System）是目前使用较普遍的操作系统之一。利用语音识别与合成技术，我们开发了面向DOS操作系统的语音机器翻译系统——SCDOS（Speech Control DOS Executing System)，系统将语音技术与人机界面技术结合起来实现语音控制DOS命令的执行，用户通过麦克风向计算机用自然语言口述自己的要求，SCDOS系统将此要求翻译为DOS命令并进行相应的处理，这种人机接口是轻松愉快的。

SCDOS系统由三大部分组成：语音识别、语言处理和语音合成。本文讨论第二部分即专用机器翻译部分，该部分利用机器翻译的技术和方法，对输入句子进行语法分析、句法分析和语义分析得到DOS命令，然后给予执行。

我们提出了基于规则库的语义分类方法来提高规则推理的效率。利用语义相似度作为分类依据，对规则库中互联的规则，通过分类词典概念阶层中的语义概念间的位置关系和DOS领域知识计算出相关联词汇的语义相似度，如果他们的语义相似度属于同一类别（即词汇语义相近），则找出它们中的一个语义相似度适中的词汇——中心意义词,并用该中心意义词去代替其余词汇，这样，各条互联规则中的不同词汇就减少一个。利用语义相似度来判别规则的类别，可计算出规则间的语义距离，从而确定了规则的优先级初值。这种分类符合语义体系的分类思想，而且经实践证明是行之有效的。SCDOS系统采用动词分类为主，名词分类为辅的语义分类法。

我们提出了基于语用环境知识、结合分词标志的逆向最大匹配分词法，并在分词过程中同时进行英文组块。SCDOS系统中的词条分为三大类：特殊词汇、结构词汇、非汉字字符。参照语言学中的组块理论，在分词过程中利用非汉字字符间的汉字作为分隔标志，采用自然语言中的“朝前看一词”（Look Ahead of One Word）思想，扫描下一个位置以确定前面的待定状态。在SCDOS系统中，DOS命令与相应汉语之间不是简单的一一映射关系，而是一种多对多的关系。系统中词典不提供直接译文，而存放对口语DOS命令进行翻译时所需要的词条信息，再利用关键字智能匹配法进行翻译。首先，SCDOS系统利用分词结果得到的词素提供词类和语义信息，抽取出其中的名词词链和动词词链，检索命令字规则库，以语义值为关键值查出规则库中该类规则的起始地址，在相应记录段中用折半二分法推导出DOS命令字；其次，SCDOS系统再抽取出分词结果中的非汉字串链，检索命令参数规则库，测试上下文相关匹配函数，得出非汉字串的类型，求出DOS命令的参数；最后将DOS命令字与参数合并在一起就构成了一个完整的DOS命令。

SCDOS系统用Borlandc C++开发而成，由输入模块、分词模块、非汉字字符组块模块、规则库语义分类模块、关键字智能匹配模块、命令执行模块等组成。系统的运行结果证明我们采用的技术方法是正确的。