口语对话系统中的
语音理解研究

(申请清华大学工学博士学位论文)

培养单位： 计算机科学与技术系

学　　科： 计算机科学与技术

研 究 生： 孙　辉

指导教师： 吴文虎 教授

副指导教师： 郑　方 教授

二〇〇五年八月

Research on Speech Understanding in Spoken Dialogue Systems

Dissertation Submitted to

Tsinghua University

in partial fulfillment of the requirement

for the degree of

Doctor of Engineering

by

Hui SUN

(Computer Science and Technology)

Dissertation Supervisor: Professor Wen-hu WU

Associate Supervisor: Professor Fang ZHENG

August, 2005

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：（1）已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；（2）为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容；（3）根据《中华人民共和国学位条例暂行实施办法》，向国家图书馆报送可以公开的学位论文。

本人保证遵守上述规定。

（保密的论文在解密后遵守此规定）

作者签名：_____ 导师签名：_____

日　　期：_____ 日　　期：_____

摘 要

语音识别和语义理解是口语对话系统最前端的两个模块，其性能好坏直接影响整个系统的性能。为了提高这两个模块的性能，本文提出基于语义概念的语音理解框架，并围绕语义概念在置信度确认和待登录关键词的发现和自动标注方面提出新的方法和策略。本文主要工作包括：

1．基于语义概念的语音理解框架。针对口语对话系统中识别性能不佳的问题，提出了以语义概念为核心的语音理解策略：在搜索中及早的利用上层语义概念知识，并将语音识别和语义理解两个步骤更加紧密地结合在一起，使得识别器可以直接输出语义概念结果。该框架利用规则描述语义概念知识，避开领域数据收集和标注的困难；另外，它具有良好的可扩展性，可以很方便地在识别中引入对话上下文知识。实验表明：在语音理解框架下，系统的识别性能和理解性能都有较大程度的提高。

2．语义概念的置信度确认方法。语义概念是整个对话系统中的核心内容，为了减少语义概念错误对系统后续模块的负面影响，本文主要从特征方面着手，研究语义概念的置信度确认方法：针对语义概念从分析理解层面提出新的置信度特征，如概念级的语言模型得分；另外，将韵律边界信息作为一种新的特征引入到语义概念的置信度打分中。声学层、分析理解层置信特征和韵律边界特征三者结合取得了很好的语义概念确认效果。

3. 待登录关键词的发现及其语义类属性的自动标注。考虑到对话系统设计初期关键词表不够完善的问题，提出待登录关键词发现及其语义类属性自动标注的策略：通过对语料进行语义分析发现待登录关键词，并根据上下文关系推测该词在词表中可能对应的语义类。待登录关键词发现和自动标注让系统具有一定的自学能力，可以在实际应用中使词表不断完善，使原有规则覆盖更多的语义概念表达方式，进而提高语音理解性能。

关键词：语音理解；语音确认；口语对话系统；语音识别

Abstract

To improve the performance of Speech recognition and language understanding in Spoken Dialogue Systems (SDSs), the speech understanding framework based on semantic concepts is proposed in this dissertation. Under the proposed framework, some other new strategies are investigated involving the confidence measures for semantic concepts and the method for detecting and automatic labeling of new keywords. Main contributions are:

1. The speech understanding framework based on semantic concepts. The performance of speech recognition is not good enough to satisfy the needs of real-life applications. In order to overcome this problem, a novel speech understanding framework based on semantic is proposed, which utilizes the semantic knowledge as early as possible in the recognition process and integrate the speech recognition and language understanding seamlessly. In the framework, semantic knowledge is described by rules that are transformed to Finite State Machine (FSM) later to direct the recognition searching, and this process avoids the difficulties of collecting and labeling domain-specific data. In addition, the framework is easy to extend to introduce much more upper layer knowledge, such as dialogue context knowledge. Experimental results show that the proposed framework can achieve high recognition and understanding performance.

2. Confidence measures for semantic concepts. Semantic concept errors may lead to misunderstanding and bring bad impact on dialogue process. Therefore, confidence measures for semantic concepts are investigated. We propose some new confidence features from understanding layer such as rule probability. And though the analysis of recognition results it is found that some concept errors are caused by boundary error, so according to this the prosodic phrase boundary information are proposed as a new confidence features. Experimental results show that the performance of semantic concepts verification is satisfactory by combining confidence features from acoustic layer, understanding layer and prosodic boundary information.

3. Detecting and automatic labeling of new keywords. At the design stage, the vocabulary of the SDS is maybe not good enough to cover all the keywords in the specific domain. Detecting and automatic labeling of new keywords is proposed to renew the vocabulary automatically during the practice. The data collected by prototype system are parsed to find the new keywords, and then infer their corresponding semantic class according to the context of new keywords. In this way, the SDS has the ability of self-learning, which can renew his vocabulary and make the original grammar cover more expression, so the semantic concepts recognition performance can be improved indirectly.

Keywords: Speech understanding; Confidence measures; Spoken dialogue system; Speech Recognition

目 录

第 1 章 绪论.....1
1.1 对话系统概述.....1
1.1.1 对话系统的组成.....1
1.1.2 口语对话系统研究的意义.....2
1.1.3 口语对话系统的发展.....3
1.2 口语对话系统的研究现状.....4
1.2.1 国内外一些对话系统的简介.....4
1.2.2 口语对话系统研究的难点.....5
1.2.3 研究现状.....6
1.3 研究工作概述.....9
1.3.1 研究目标.....10
1.3.2 研究思路和研究内容.....10
1.4 论文的组织结构.....12
第 2 章 基于语义概念的语音理解框架.....13
2.1 对话系统中的语音识别和语言理解.....13
2.1.1 对话系统中常用的几种识别框架.....13
2.1.2 对话系统中的语言理解.....17
2.1.3 识别器与理解器的接口.....19
2.2 基于语义概念的语音理解框架.....19
2.2.1 问题的提出.....19
2.2.2 基本思想.....20
2.2.3 语义概念的重要性.....21
2.2.4 语义概念的定义.....23
2.2.5 语义概念在识别中的可用性.....24
2.2.6 系统框图.....26
2.3 基于语义概念的语音理解框架的具体实现.....27
2.3.1 语义概念知识在识别中的应用.....28
2.3.2 语音理解框架下对话上下文知识的引入.....32

2.3.3 语义概念的解码过程 33
2.4 实验结果与分析 35
2.4.1 实验背景和实验数据 35
2.4.2 实验设计 35
2.4.3 实验结果与分析 36
2.4.4 讨论 39
2.5 小结 39
第 3 章 语义概念的置信度研究 41
3.1 置信度研究的现状 41
3.1.1 置信特征 42
3.1.2 确认模型 43
3.1.3 评测指标 43
3.2 语义概念的置信度确认 44
3.2.1 问题的提出 45
3.3 分析理解层的置信度特征 46
3.3.1 描述语义概念之间相关性的特征 46
3.3.2 语义分析层的置信度特征 47
3.4 基于韵律信息的置信度特征 48
3.4.1 出发点 48
3.4.2 韵律边界的检测 49
3.4.3 韵律边界信息作为置信度特征 52
3.5 实验结果与分析 52
3.5.1 实验一：不同置信度特征下的语义概念确认性能 53
3.5.2 实验二：加入语义概念确认后对话系统的理解性能 55
3.5.3 分析与讨论 55
3.6 小结 56
第 4 章 待登录关键词的发现及其语义类属性标注 57
4.1 研究意义 57
4.1.1 背景对话系统 57
4.1.2 研究意义 58
4.2 相关研究 60

4.3 待登录关键词的发现.....61
4.4 标注待登录关键词的语义类属性.....62
4.4.1 推测待登录关键词的语义类属性.....62
4.4.2 待登录关键词的语义类属性的确认.....65
4.5 实验结果与分析.....67
4.5.1 实验结果.....67
4.5.2 分析与讨论.....67
4.6 小结.....68
第 5 章 总结与展望.....69
5.1 论文工作总结.....69
5.2 下一步研究的展望.....70
参考文献.....72
致谢与声明.....79
个人简历、在学期间发表的学术论文与研究成果.....80

第1章 绪论

随着信息技术的飞速发展，计算机已经逐渐成为人们日常工作和生活中不可缺少的一部分，人们需要经常与计算机之间进行信息交互。自然语言是人类进行信息交流最主要、最自然的手段，如果人类和计算机之间也能通过自然语言来交流，必将大大提高人机交流的效率和自然度。随着语音识别、语音合成和自然语言理解三项技术的迅速发展，集合这技术于一身的口语对话系统（Spoken Dialogue System）受到国内外研究机构的高度重视，它为人机交互提供了一种更加自然、更加有效的方式：允许人们通过自然语言表达自己的思想，与计算机就某一领域的内容进行信息交互。对话系统的应用必将带来很好的社会效益和经济效益，目前一批初级的应用系统已经面市，常见的比如旅游信息查询、电话客票服务、语音呼叫中心、天气预报信息查询等。

构建一个完善的口语对话系统，需要应用语音信号处理、语音识别、语言理解、对话管理和语音合成等多项技术。本论文的研究工作主要集中在口语对话系统中的语音理解方面，将识别和理解看成一个整体，研究其整体性能的提高。

本章的内容安排如下：首先对口语对话系统的组成和发展做简要介绍，分析口语对话系统研究的重要性和必要性；第二节指出口语对话系统研究的重点和难点，并综述其研究现状；最后给出本文工作的研究思路和具体内容。

1.1 对话系统概述

1.1.1 对话系统的组成

口语对话系统，可以简单地定义为：以自然语音为输入输出接口，通过与用户进行交谈，完成对话任务，实现自动信息服务的系统。要完成人机对话任务，必须综合语音识别、语言理解、对话管理、自然语言生成和

语音合成等多项技术，使之成为一个有机的整体。图 1.1 是典型对话系统的模块结构略图，从图中可以清晰看到对话系统的几个组成部分，包括语音识别器、语义分析器、对话管理器、自然语言生成器和语音合成器。另外，系统的运行还依赖于声学模型、语言模型、句法/语义规则、领域知识、对话模型等五个部分，大多数对话系统还有领域数据库的支持。

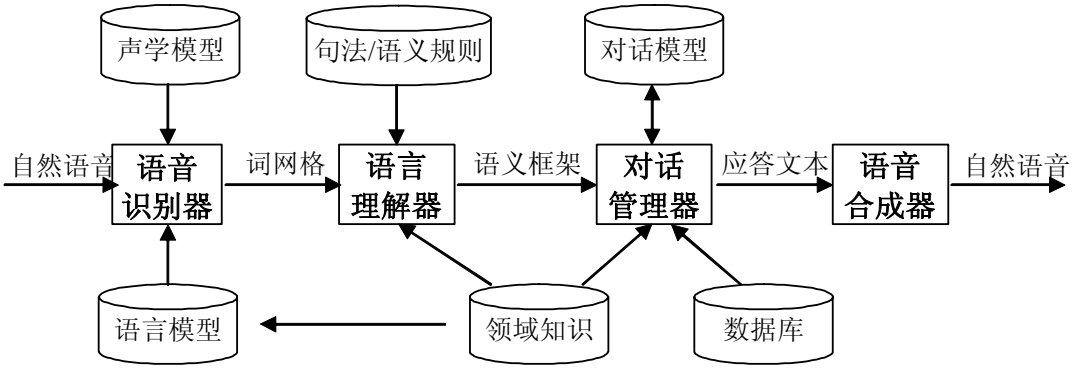

图 1.1 典型对话系统的结构图

口语对话系统中语音识别模块的主要任务是把人的语音转换成文字，这一点类似于其它语音识别应用，如听写机、语音命令系统等，该模块位于整个系统的最前端，其性能高低将直接影响到系统的整体性能。语言理解模块的主要任务是分析用户输入语言的内容，把用户的真实意思用计算机可以理解的内部方式表达出来。对话管理器也是口语对话系统中核心模块之一，它的任务是根据语言理解的结果、对话的上下文知识和历史信息综合分析，确定用户的意图，并根据需要查询后台数据库，组织适当的应答语句，以保证计算机与人的交谈可以有效、友好地继续下去，直到用户的目的得以实现。最后，语音合成模块的任务是将应答文字转换为语音输出给用户。

1.1.2 口语对话系统研究的意义

信息咨询是目前口语对话系统研究的主要热点：人们通过自然语言向计算机表达自己想要咨询的内容，计算机理解之后按照用户的要求查询数据库，并把查询结果反馈给用户。这类人机对话系统的最突出的特点是：面向特定任务（task-oriented）和特定领域（domain-dependent）。

众所周知，我们处在一个信息技术高度发展的时代，人们面对的信息量成几何级数迅速增长，在浩如烟海的数据和信息中，怎样才能更方便地获取到你所需要的信息内容呢？用于信息咨询的口语对话系统为人们提供了一种方便的获取信息的途经，以自然对话的形式与机器进行交流必将大大提高信息交互的速度和自然度。

另外，随着通讯技术和网络技术的发展，手机、掌上电脑等移动通讯设备逐步普及，人们希望随时随地获取信息。但是在这些移动设备上，使用传统的人机输入方式（如键盘、鼠标、手写笔等）很不方便，在这种情况下，具有自然语言理解能力的语音接口自然就成为最方便的人机接口。

综上所述，研究用于信息咨询的口语对话系统有其重要性和必要性。

1.1.3 口语对话系统的发展

口语对话系统是语音技术发展到一定阶段的产物。

语音技术大致可分为语音识别技术和语音合成技术。语音识别使得计算机具有“听”觉，而语音合成技术让计算机具有“说”的能力。从二十世纪六十年代开始，语音识别就一直是一个非常活跃的研究领域。八十年代中期以来，语音识别技术有了实质性的进展，HMM 的广泛研究和应用，使识别技术在大词汇量、非特定人、连续语音识别这三个方面都取得重要发展。许多大词表连续语音识别系统（听写机）在实验室环境中的识别率可以达到 90%以上，卡耐基梅隆大学（CMU）的 SPHINX 系统就是其中一个典型的代表[1][2]。九十年代以来，语音识别系统开始从实验室进入市场，如 IBM 的 ViaVoice 系列。语音合成（Speech Synthesis）的历史可以追溯到 17 世纪，那时候人们用机械装置来模拟人发音。二十世纪 70 年代以后，随着计算机科学的发展，语音合成技术有了较快的发展。80 年代末，基音同步叠加（Pitch-Synchronous Overlap Add，PSOLA）技术成功地应用于普通话语音的编辑合成，大大提高了合成的质量[3]。最近几年，基于数据库的合成方法逐渐成为语音合成的主流方法。在这一方法中，合成语句的语音单元是从一个预先录下的庞大的语音数据库中挑选出来的。由于合成的语音基元都来自自然的原始发音，合成语句的清晰度和自然度都非常高。

语音技术的进步使得人们渴望已久的人机语音接口这一具有重要意义的研究课题成为可能。近十年来，发达国家投入了大量的人力、物力、财力来研究口语对话系统，美国有 DARPA 的 Communicator 计划[4]、欧洲有 ARISE 计划[5]、REWARD 计划[6]、VERBMOBIL 计划[7]等。有许多著名的学府和研究机构都从事这项研究，比如美国麻省理工学院（MIT）的 SLS 实验室、CMU 的 ISL 实验室、Lucent-Bell 实验室、德国的 Erlangen-Nuremberg 大学、日本的 ATR 实验室和 Philips 公司等。在国内，中科院自动化所、清华大学、香港中文大学、香港科技大学、台湾大学等也都投入了相当大的精力进行这方面的研究。

1.2 口语对话系统的研究现状

1.2.1 国内外一些对话系统的简介

经过国内外研究机构一段时间的潜心研究，目前已经出现了不少有实用价值的口语对话系统，以下是一些对话系统的简介。

一、麻省理工学院的 GALAXY 系统[8]，这是一个用于旅游信息查询的系统，能够提供大约 750 个城市的天气预报和 250 个城市的航班情况。它的语音识别器 SUMMIT[9]采用基于分段（Segment-Based）的识别方法，词识别率为 83.9%；语言理解器 TINA[10]用语义框架的结构来描述语义。在 GALAXY 系统的基础上，文[11]提出了开发口语对话系统的参考体系结构 GALAXY-II，成为 DAPAR Communicator 项目的参考体系结构。GALAXY-II 采用客户服务器（Client-Server）结构，其核心服务器是 HUB，语音识别、语言理解、语言生成和语音合成等核心部件均以服务器的形式存在，不同服务器之间通过 Hub 脚本（Hub script）相互交互信息。GALAXY-II 系统作为研究口语对话系统的实验平台，在其基础上又开发了很多不同领域、不同语言的系统。电话天气信息查询系统 JUPITER[12][13]就是其中之一，它的词表规模为 1957，其中包括 650 个城市名和 166 个国家名，声学模型使用上下文相关的 diphone 模型，识别时使用了词类 bigram 模型和词的 trigram 语言模型。

二、卡耐基梅隆大学的 Communicator 系统[14]。作为 DAPRA Communicator 项目的一员，CMU Communicator 也用于旅游信息查询领域。它的前端采用

Sphinx II[2]识别器，并且支持语音插入（Barge-in），词识别率为 85%；该系统采用基于语义文法（Semantic Grammar）的 Phoenix Parser[15]作为语言理解器，对话管理器采用基于 Agenda 的策略[16]，可以处理复杂的任务。

三、德国的 VERBMOBIL 系统[7]。这是一个口语翻译系统，用于会议的安排，它通过一个动态建立的上下文模型和一个建立在语料库上的随机模型，可以预测对话中的下一句将会是什么。

四、由英德法意等国共同开发的 SUNDIAL 系统[17]。它的主要目标是要让用户通过电话以自然的对话来获取例如飞机航班或者火车时刻等信息。系统的词汇量为 1000 左右，是非特定人的系统，而且具有很好的对话管理功能，通过电话进行的对话成功率达到 96%。

五、中国科学院自动化所模式识别国家实验室的 LOADSTAR 系统[18]。该系统向用户提供旅游信息，并可以根据用户的要求计划旅游路线。它采用了大词表连续语音识别的技术，识别结果经过语义项的匹配得到有关的语义概念，对话管理采取人机混合主导的策略，基于模板生成系统应答，系统的应答准确率达到了 90.9%。

六、清华大学智能技术与系统国家重点实验室语音技术中心的 EasyNav 系统[19]，它向用户提供清华大学校园导游服务，包括校园内的建筑物信息和交通信息。该系统考虑了汉语口语中的省略、指代现象，能处理上下文相关的对话。

1.2.2 口语对话系统研究的难点

对比其他的语音识别应用，对话系统有其明显的特点：

第一，口语对话系统都有比较明确的领域限制，一般说来它只需要关心领域相关的内容，对于超出领域限制的用户输入可以不加理会；

第二，不同于语音命令系统中的孤立词和听写机系统中的朗读语音，对话系统面对的是自发语音（Spontaneous Speech），发音比较随意；

第三，对话系统的输入是人们日常生活中的口语，语句中常常包括不流利、不合语法、内容不完整等口语现象；

第四，口语对话系统的应用环境比较多样化，可能是非常安静的实验室环境，可能是充满噪音的正在行驶的汽车中，更有可能是人声嘈杂的商场。

上述对话系统的四个特点中，第一点在一定程度上降低了识别和理解的难

度；而自发语音、口语现象和环境影响却对口语对话系统中的识别和理解提出了较高的要求，成为对话系统实用化的难点问题。

实验室环境下，对于采用标准发音的、仔细朗读的语音，传统的连续语音识别器一般都能够达到很高的识别率。然而，对于无准备的、自然随意的发音，识别器的性能却会急剧下降[20]，这在很大程度上是由于自然语音本身的特殊性决定的。在声学层面，自然语音语速多变，带有各种语气和真实的情绪，还存在严重的协同发音，它们会造成大量音素级的插入、删除和替换现象。此外，不同的人具有不同的口音背景和发音习惯，因此即使说话人努力按照标准的读音去发音，实际的音素序列也不会完全相同[21][22]。汉语相对于西方语言来说，又具有其特殊性。汉语的发音是基于标准音节的声韵结构（包括零声母现象），这种结构非常短，很容易受到口语上下文的影响而发生畸变[23]。在语言层面，自然语音通常伴随着大量的口语现象，例如垃圾、碎片、犹豫、纠正、重复、指代、省略、咳嗽和吹气等现象，这些口语现象在日常生活人与人交谈中普遍存在，它们除了给语义理解和分析带来困难以外，也对识别性能产生较大的负面影响。

口语对话系统的应用背景复杂，输入的语音质量不高，通常伴随着一定的环境噪音，并且语音质量和噪音还对话系统的应用领域紧密相关，不同口语对话系统可能出现的噪音类型也不尽相同，这都给自然语音的识别增添了难度。例如，嵌入式设备上的口语对话系统中，录音设备与一般麦克风的录音质量相差很大；而使用电话作为输入设备的口语对话系统，则面临着电话信道畸变、移动电话和固定电话信道特点不同以及户外使用电话时背景噪音大等一系列问题的困扰。

综上所述，自然语音、口语现象和环境噪音是目前对话系统研究中的主要难点，给研究工作者提出了巨大的挑战。

1.2.3 研究现状

下面，从语音识别、语言理解、对话管理三个方面简单介绍口语对话系统的研究现状。

1.2.3.1 语音识别

语音识别器位于口语对话系统的前端，其性能好坏直接影响到系统能否正确的理解用户语句。在实际的口语对话系统中，用户输入的语音比较随意，协同发音问题突出，当训练语料比较充足时，人们通过训练上下文相关（Context-Dependent）的声学模型来解决协同发音的问题[24]。文[25]定义了广义声韵母集，更全面地覆盖汉语中可能出现的音变现象，并提出了基于广义声韵集的精细建模方法，提高了自发语音的识别正确率。

自发语音中的停顿、语音修改等不连贯现象（Disfluency）也得到了较多的研究[26][27][28]。研究结果表明：声学韵律特征、覆盖隐含事件（Hidden-event）的语言模型、句法或者语义异常等都可以帮助定位语音修改，将这些来自不同知识源的特征结合在一起能够较准确的检测句子中语音修复的插入点。

另外，鲁棒的语音识别也是近来对话系统研究中的热点问题，噪音问题是语音识别鲁棒性研究的主要内容之一。前端去噪（Front-end）算法在声学特征一级研究噪音的处理，通常的做法是在提取特征前先估计噪声的强度，用谱减（Spectral subtraction）的方法降低噪音，然后提取鲁棒的语音特征进行语音识别；后端（Back-end）去噪方法主要在声学模型一级研究噪音问题，通过模型补偿（Model compensation）技术[29][30]，减少测试集和训练集的差别，进而提高含噪语音的识别性能。

性能再好的语音识别器也会出现识别错误，这些错误会给对话系统后续理解模块带来负面影响。为了减少识别错误对系统性能的影响，人们开始关注对话系统中的语音确认研究[31]。语音确认就是对语音识别结果的正确性进行评价，以此决定接受或者拒绝某个识别结果。在对话系统中加入语音确认，不仅可以提高识别和理解性能，还有助于对话任务的完成，因为对话管理模块可以根据语音确认的结果有效地引导用户，进而提高对话成功率。

集外词的处理是识别中另一个值得研究的方面。对话系统的词表规模有限，在实际应用中，不可避免的会遇到一些词表以外的词，称之为集外

词（Out-of-vocabulary，即 OOV）。为了减少集外词而引起的识别错误，识别器通常会为集外词建立一个通用模型，词表中的某个词如果要想在竞争中胜出，必须要胜过集外词模型的打分，如文[32]中采用音素级 bigram（phone bigram）为集外词进行打分。

1.2.3.2 语言理解及语义分析

语言理解是对话系统的重要组成部分，其功能是通过句法分析、语义分析、语义表示等过程，将用户的语句转化成系统可以理解的形式，并传递给后续的对话管理模块。TINA[10]是 MIT 研究开发的一个著名的自然语言理解器，用于很多对话系统中，如 JUPITER，VOYAGER 等等。TINA 中的句法分析以上下文无关文法（Context Free Grammar, CFG）为基础，具体实现时，TINA 还在基本文法的基础上，从训练语料中自动生成更多的文法，然后将文法按照一定的规则转换为网络，并根据训练语料在文法网络中引入概率。TINA 中的句法分析和语义分析是一致的，文法中的很多非终结符都是有语义的，因此，句法分析产生的分析树可以直接用于填写语义槽。

口语中常常出现的不合语法现象，以及语音识别错误，都会给语言理解带来极大的困难，针对这两个问题，不少研究者提出了部分分析(partial parsing)的想法，让分析器跳过不合文法或者错误的地方，将句子有用信息分析出来。在 TINA 最新实现中，句法分析有两个阶段，第一个阶段进行严格分析，如果不成功则在第二个阶段中使用鲁棒的分析方法对句子进行部分分析，尽可能的获得句子中的有用信息。在天气信息查询系统 JUPITER 中也尝试了基于词捡出（Word Spotting）的语言理解方法[33]，其核心思想是利用有限状态机(Finite State Machine, 即 FSM)将识别器输出的 N-best 结果分析成“键－值”（Key-value）形式的结果，用这一结果直接填写语义槽。这种语言理解方法具有较大的灵活性，在简单应用中效果非常不错。

语言中存在歧义、指代(anaphora)和省略(ellipsis)现象，它们也是对话系统语言理解所必须面对的主要问题。文[34]把对话系统中存在的语言歧义分为三个层次，分别是词法歧义、结构歧义和指代歧义。省略现象是指结构不完整的语句，针对汉语的特点，文[35]提出语境省略概念，并用主题结

构的方法分析省略现象。

1.2.3.3 对话管理

对话管理是对话系统的核心，它使用对话模型来描述对话状态，决定对话状态的转移和应答生成。如何利用恰当的确认策略和混合主导方式，提高对话成功率、任务完成率和用户满意度，是对话管理模块需要解决的问题。

文[36]提出了一种自适应的对话管理方法，它使用层级槽结构（hierarchical slot structure）描述对话，提示问题也是层级结构的，避免用户被动地逐个填写槽值的乏味过程，该方法具有较好的灵活性，其自适应能力使对话过程更加有效和更富智能。文[37]介绍了一种基于表格(forms)的混合主导的对话管理方法，它认为一个对话过程由若干任务构成，用一个表格对应一个任务的形式来描述对话状态，整个应用可以用表格集合来描述。表格内容包含特定任务内的所有语义槽、每个槽对应的属性名，以及槽一级和表格一级的回放消息。回放消息分为 help、prompt 和 back-end 等几种，其中 back-end 类型的消息附有后台任务操作函数，函数的返回值指示各表格的启用和禁用状态，当前要清除的表格和槽的列表，以及报告给用户的当前对话状态。动态调整可容许的表格列表，可以在系统主导和混合主导之间进行切换。文[38]实现了基于主题森林的对话管理方法，并提出了对话管理的心理模型，认为对话管理是一个心理过程，应该直接为理解和应答过程的心理建模，而不是为应答的语言建模。

1.3 研究工作概述

本论文将工作重点放在对话系统前端的识别和理解上，针对口语对话系统中由于自然语音和口语现象而导致的识别性能不佳的问题，尝试通过在识别中更早地利用高层知识，并将语义概念的识别和理解过程更紧密地结合起来的方法提高对话系统的识别和理解的性能。本文在识别框架、语义概念的置信度研究和待登录关键词发现及其语义类属性自动标注三个方面都提出了一些新思路和新方法，并通过实验证明了其有效性。

1.3.1 研究目标

本文的总体目标是改进口语对话系统识别和理解的性能，提高对话系统的实用性，具体研究目标包括：

1) 面对口语对话系统在实用中出现的语音识别性能不佳问题，通过分析几种常用识别框架的优势和不足，提出基于语义概念的语音理解框架：将语音识别和语义概念理解过程更加紧密地结合在一起，以提高识别性能；识别器直接输出用于填写语义槽的语义概念，以提高语言理解的鲁棒性。

2) 在基于语义概念的语音理解框架下，研究语义概念置信度确认的方法：提出新的置信度特征对语义概念进行确认，为后续模块提供更加可靠的语义概念结果。

3) 对话系统在系统设计初期，往往缺乏足够的领域内数据，从而导致设计词表覆盖不够，影响识别性能。针对这一问题，提出待登录关键词发现及其语义类属性自动标注的方法，在对话系统中引入学习机制，使得系统可以在实用中通过词表的完善不断提高性能。

1.3.2 研究思路和研究内容

论文的研究工作涉及口语对话系统识别和理解领域的多个方面，研究思路和工作内容如图 1.2 所示。

具体地说，作者的研究工作包括以下几个方面：

（1） 基于语义概念的语音理解框架

通过分析口语对话系统中现有识别策略的优势和不足，发现识别性能与高层知识在识别中的应用程度有着比较密切的关系，用领域内的实际语料训练的词类 N-gram 模型和根据领域语句特点构建的模板可以有效地提高识别性能。但是这两种方法又有各自的缺点，词类 N-gram 模型对领域数据的依赖性强，而模板方法的灵活性太差。

本文提出了基于语义概念的语音理解框架，将规则形式描述的语义概念知识及早的应用于识别过程，提高了识别性能。新的识别框架避开了领域数据收集和标注的困难，并在语义概念引入识别过程中，充分考虑了对话中可能出现的口语现象，保证了识别的鲁棒性。在新的识别框架下，识别结束时同时可以

输出语义概念结果，使用这一结果直接填写语义槽，能够提高语义理解的鲁棒性，进而提高理解性能。实验结果表明，在基于语义概念的语音理解框架下，识别性能和理解性能都比基线系统有大幅度的提高，最终的音节识别正确率达到 81.57%，语义槽的正确率达到 86.33%。

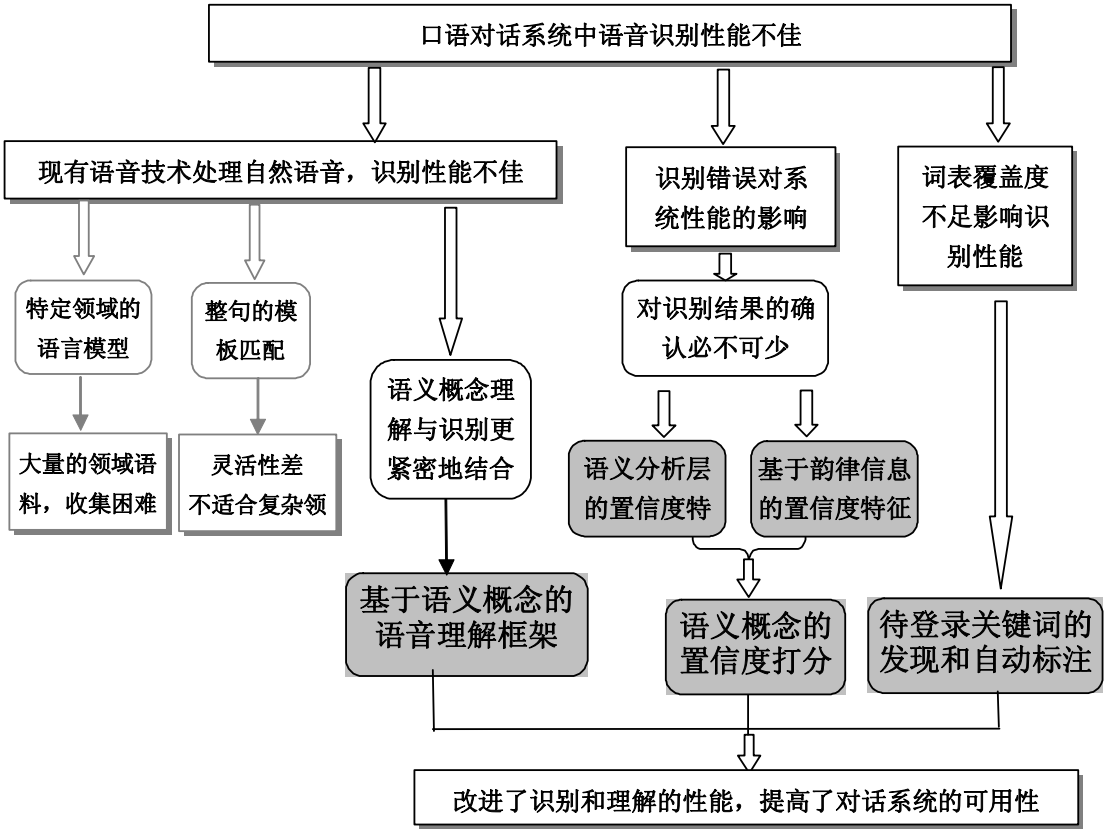

图 1.2 论文的研究思路和研究内容

（2）语义概念的置信度研究

语义概念的插入错误和替换错误会直接影响到后续对话管理模块的性能以及系统的用户满意度，因此，有必要对语义概念的可靠性进行评价。本文在语言层面和分析理解层面都提出了新的置信度特征，适用于基于语义概念的语音理解框架。另外，通过分析识别结果，发现有些识别错误是词边界错误引起的，于是提出将韵律边界作为一种新的特征用于语义概念的置信度打分。实验结果表明，语言层和分析理解层的置信度特征，其确认性能要优于声学层面的置信度特征，将韵律边界特征与声学层、语言层、分析理解层的置信度特征结合后，

语义概念的确认性能进一步提高。

（3）待登录关键词的发现及其语义类属性的自动标注

口语对话系统设计初期，由于经验不足，导致一些领域内的词并不在词表中，这一点对识别性能产生了较大的负面影响。为此，作者提出待登录关键词发现和自动标注的策略：通过对新语料的处理，得到语句的语义分析结果和词法分析结果，根据这两个分析结果，发现待登陆关键词，并结合统计方法推测其语义类属性。实验结果表明，大部分待登录关键词可以正确地对应到词表现有的语义类中；词表更新后，语音识别的性能有大幅提高。

1.4 论文的组织结构

第二章中先介绍本论文涉及领域的相关研究，包括口语对话系统中常用识别框架和语言理解方法，接着针对口语对话系统识别性能不佳的问题，提出基于语义概念的语音理解框架，并给出该框架的具体实现；第三章在基于语义概念的语音理解框架下，研究语义概念的置信度打分方法；第五章从不断完善对话系统的角度，提出对话系统中待登录关键词发现及其语义类属性自动标注的策略；最后在第六章给出全文总结和对相关领域研究的展望。

第2章 基于语义概念的语音理解框架

所谓语音理解（Speech Understanding），主要包括两个方面的涵义——“听到”和“听懂”。首先，要“听到”用户的语音，接着对用户语音进行理解，“听懂”用户的意图。口语对话系统与语音识别其他应用（如听写机和语音命令系统）的主要不同在于它不仅需要“听到”语音，更重要的是能听懂用户的意图，因为对话管理模块只有在理解用户意图之后，才能根据一定的对话策略，帮助用户完成各种交互式任务。因此，对于口语对话系统来说，语音理解的性能是至关重要的，它直接影响着整个系统的性能。

目前，在大多数的口语对话系统中，语音理解过程都是分两步完成的：第一步，语音识别器对输入语音进行识别，输出 N-best 或者词图（Word Graph）形式的识别结果；第二步，语言理解器对识别器的输出进行分析和理解，得到对话管理模块所需要的语义表示形式。这种方法将语音识别和语言理解两个过程完全分隔开来，使得语言理解过程的很多有用信息无法在识别中加以利用。本章中，作者提出基于语义概念的语音理解框架，在这一框架下，识别和理解不再是两个各自独立的模块，它们更加紧密地结合在一起，成为一个统一的语音理解模块，新的语音理解模块完成原来经过识别和理解两个模块完成的工作，实现了语义概念的识别过程。通过在识别过程中及早利用语义知识、对话上下文知识等多种来自于不同知识源的有效信息，提高了语音识别的正确率；识别语音的同时完成对于语义概念的理解能减少错误的识别结果对语义分析的干扰，进而提高了语义理解的正确率。

本章的内容安排如下：第一节简单介绍对话系统中常用的语音识别框架和语言理解方法。第二节分析现有识别框架的不足之后，提出基于语义概念的语音理解框架；第三节介绍基于语义概念的语音理解框架的具体实现；第四节中给出实验结果和分析；最后总结基于语义概念的语音理解框架。

2.1 对话系统中的语音识别和语言理解

2.1.1 对话系统中常用的几种识别框架

语音识别技术的发展使得对话系统的应用成为可能，反过来，对话系统的

特点及其发展也对语音识别技术提出了新的要求。口语对话系统中的语音识别大都采用基于 HMM 的大词表连续语音识别技术，同时也针对口语对话特点进行一定的改进。下面，简单介绍现有系统中常用的语音识别框架。

2.1.1.1 基于特定领域 N-gram 语言模型的连续语音识别

大多数口语对话系统中采用基于 N-gram 语言模型的连续语音识别框架。该识别框架在对话系统中的应用与其在传统大词表、连续语音识别应用中有所不用，主要区别在于：（1）大词表、连续语音识别应用的词表非常大，经常有上万个词汇[39]，而对话系统的词表规模则相对小得多，大都只是包括该领域能够涉及到的词汇，例如 MIT 的天气信息查询系统 JUPITER[13]，其词表规模就只有不到 2000，其中一半左右是城市名和国家名，剩下的主要是查询中常用的词和语言理解器能够理解的词汇。（2）大词表连续语音识别中常用的 tri-gram 统计语言模型是通过大量书面的文本语料训练得到的，这些训练语料在常用词汇和常用表达方式方面与口语对话系统中所用到的很不一致，对话系统中直接使用这种通用的语言模型识别效果会非常不好，因此，对话系统中语言模型通常是领域相关的，需要针对每个特定领域收集语料，并训练一个特定的语言模型。

为每一个对话系统都收集大量本领域的对话语料用于训练语言模型，这是一项非常困难的事情，要花费大量的人力和物力。词类 N-gram（Class N-gram）语言模型[40]的提出，就是为了减少模型参数，以保证在较少训练语料的情况下还能够比较准确地估计出语言模型的参数。词类 N-gram 模型中，根据一定的标准（如语义相同）将所有的词分成一定数目的词类，模型中概率描述的是词类到词类的转移概率，而不是原来的词与词之间的关系，如公式（2-1）所示：

$$P(c(w_i) \mid c(w_{i-N+1}),...,c(w_{i-2}),c(w_{i-1})) \qquad (2\text{-}1)$$

其中，$c(w_i)$表示词 w_i 所属的词类。在口语对话系统的应用中，词类 N-gram 的物理意义更加明确，更加符合实际情况。例如，考察这个句子“我想订一张北京到上海的”，基于词的 Tri-gram 估计的是 P(上海|到, 北京)，而词类 N-gram 模型估计的是 $P(place \mid to, place)$ 和 P(上海 | $place$) 。从语言模型描述上层句法语义限制关系的作用考虑，显然后一种估计更加合理，因为例句在这一位置出现“上

海”这个地点词的概率应该跟它前面的介词“到”和 “到”前面是否是地点词相关，而与前面这个地点词是“北京”还是“广州”并无太大关系。

词类 N-gram 语言模型有较强的预测性，相比词的 N-gram 它的困惑度更低，在其指导下的连续语音识别策略也具有识别率较高的特点，在对话系统中有较多的应用。JUPITER[12]的识别器中结合使用了词类 Bi-gram 和词的 Tri-gram，词类 Bi-gram 包括 200 多个词类。August 系统[41]使用了规模为 500 的词表，以及基于 70 个词类和 229 个词类对的词类 Bi-gram。旅游信息系统 LOADSTAR[18]使用了词和词类的混合模型，词类的规模为 733。

基于领域 N-gram 语言模型的识别方案能够取得较好的识别性能，但是它对于 OOV 问题和口语现象无法很好的处理。另外，尽管词类 n-gram 语言模型对于训练数据的要求相对较小，但是要取得更高的识别性能，训练语料还要达到一定的规模要求，如 JUPITER 就用了 54000 个句子来训练词类 Bi-gram 模型。在很多口语对话系统的设计开发阶段，要得到这样规模足够、覆盖度比较全面的数据库是一件非常不容易的事情。

2.1.1.2 关键词识别

关键词识别（Keyword Recognition），又称为关键词检出（KWS，Keyword Spotting）就是在连续的、无限制的自然语音流中识别出一组给定的词——关键词[42]。与连续语音识别相比，关键词识别策略可以忽略关键词以外的语音部分，这一特点比较适合口语对话系统，因为它可以忽略掉口语语音中无意义的部分，如口头语、插入语以及咳嗽等口语现象，而不影响对语音关键部分的识别。在保证关键词检出率的前提下，如何更好地吸收垃圾词（Garbage Words）成为关键词识别研究的核心内容之一。根据吸收垃圾词所用的补白模型（Filler Model）的不同，关键词识别算法主要分为三类：集外词补白模型方法[43]、子词补白模型方法[44]和在线补白模型方法[45]。

集外词补白模型方法为补白训练专门的声学模型，识别中补白模型与关键词模型相互竞争。补白模型可以是一个也可以是多个，对训练数据中除了关键词以外的其他词进行聚类，每一类对应一个补白模型。

子词补白模型不为关键词以外的词训练专门的声学模型，而是通过拼接子

词模型来形成补白模型，识别中通过调整关键词和补白模型的权重来区分关键词和补白。这种建模方法不需要专门训练的补白模型，灵活性较好，当采用上下文相关的子词模型时可以取得不错的识别效果。

在线补白模型方法不同于前面两种方法：它不是专门为补白建立模型，而是在搜索的过程中动态地形成一个补白，跟关键词竞争，对于每一帧语音，补白模型的似然分是该帧信号对应的 N 个最优匹配成绩的平均分。在这种情况下，补白永远不是得分最高的候选，但是总是排在前几名，当一段语音连续跟某个关键词候选都是最佳匹配的时候，关键词候选才能在与补白的竞争中胜出。在线补白模型方法具有一定的抗噪能力，不足之处是，当关键词数目比较少时在线补白模型的打分不够准确，实验表明：关键词个数较少的情况，在线补白模型的性能会变得很差。

关键词识别的主要特点就是可以忽略非关键的语音部分，具有很高的鲁棒性，但是相比基于 N-gram 语言模型的连续语音识别方法，关键词识别的准确率较差，尤其对于那些含有多个关键词的语句。另外，关键词识别的结果中往往存在较多的误警，会对后面的理解产生负面影响。

2.1.1.3 基于模板匹配的语音识别

模板匹配（Template Matching）[46]的语音识别策略通常用有限状态网络来描述上层的语言知识，并用这个有限状态网络指导识别搜索的过程。基于规则分析的对话系统中一般都有一个上下文无关的领域文法，它描述了特定领域中可能出现的语言表达。领域文法可以生成一个对应的有限状态网络，该网络具有高度的预测性，用它来指导识别，限制搜索空间，能够大大提高识别率。但是，模板匹配的识别策略也存在着非常大的缺陷：鲁棒性差。当遇到不符合有限状态网络的输入时，识别性能会急剧下降。即使输入的句子完全符合有限状态网络，也可能由于句子中包含一些无法预知的口语现象（如咳嗽）而导致识别率下降。

基于模板匹配的语音识别框架对于文法的覆盖程度要求较高，难以在复杂的对话系统中应用，主要用于一些简单领域中，这些领域的文法描述比较简单，可用较少的文法规则最大程度地覆盖领域内所有的表达方式，例如自动总机应

用[47][48]。

2.1.2 对话系统中的语言理解

相比传统的自然语言理解，对话系统中的语言理解有其自身的应用特点。

首先，对话系统中的理解器针对某个特定领域的，从这一点上说，它比自然语言理解简单；第二，对话系统的输入是自然语音，也就是说，语言理解器面对的不是文字，而是语音识别的输出，现有的识别技术决定了识别输出不可避免的会存在一些错误，这些错误给理解带来了很多麻烦；第三，自发语音中存在不少的口语现象，如语序颠倒、重复、错误修改等，这些都对语言理解提出了较大的挑战。

目前，对话系统中的语言理解方法主要分为两类：基于规则的方法和基于统计的方法。

2.1.2.1 基于规则的语言理解方法

基于规则的语言理解方法，建立在 Chomsky 形式语言的体系[49]之上，其核心思想是用文法（Grammar）来描述语言、分析语言。文法和针对文法的分析算法是基于规则的语言理解方法的核心。口语对话系统中，基于规则的理解方法大都使用了上下文无关（Context-free）文法，也就是 Chomsky 的文法体系中的 2 型文法，这种文法要求产生式左端是一个单独的非终结符，它的推导不依赖于特定的上下文。上下文无关文法由一系列描述语言的产生式规则组成，文法分析过程用这些规则判定输入句子相对该文法的合法性，并给出相应的文法结构，一般可以通过句法树来直观的表示。

根据策略不同，文法分析算法[50]分为两类：自顶向下的和自底向上的。例如，TINA[10]采用自顶向下的方法，从文法起始符 S 出发，枚举文法中的规则，对当前状态下的非终结符进行推导，直至所有非终结符都被分析为终结符，且与输入句子匹配成功；自底向上的算法从输入语句的词类出发，对相邻的符号进行归结，生成对应规则的左项符号，直至归结到文法的起始符号 S。自底向上的好处在于无需回溯，对输入语句只进行一次扫描，中间生成的任何成分不会在以后的分析中再次生成，更重要的是该方法除了能够对整个输入结果做出接

受或者拒绝的判断外，还能保留部分分析的结果，这样，即使整个句子分析失败了，仍然能够从局部分析中得到一定的信息量。

基于规则的语言理解方法存在两个问题，首先，构建一个好的文法比较困难，需要由了解领域特点、具备语言知识的专家来完成，有时甚至需要收集一定的实际语料，通过分析语料来完成文法；其次，基于规则的方法灵活性不够，尤其是采用自顶向下的分析算法时，只要输入语句中有一点与所有的文法都不匹配，就无法得到这个句子的任何信息，虽然自底向上的分析算法可以保留部分分析的结果，但对于整句层面的语义还是无法准确的把握。为了增强规则方法的鲁棒性，基于语义类的上下文无关增强文法[51]被用于汉语对话分析中，该文法针对口语语音的自发性和口语现象，对规则附加增强属性，使其具有跨成分归结的特性，在一定程度上解决了对话中的口语现象，不足之处是规则的增强属性也引入了更多的分析歧义，使得歧义问题的解决变得更加困难。

2.1.2.2 基于统计的语言理解方法

统计方法在语音识别中有很多应用，如声学识别、统计语言模型。统计方法的最大好处在于统计信息直接来源于真实的语料，由此得到的模型能够反映实际数据的真实情况，在应用中表现出很强的适应性。另外，统计方法与领域无关，因此基于统计的语言理解方法可以很方便地移植到一个新的领域，只要对统计模型进行重新训练即可。

语言理解的过程就是从语音识别结果得到语义概念表示的分析过程。基于统计的语言理解方法就是对这个过程进行概率建模，对识别结果对应的语义表示进行概率打分，从中选取一个概率分最高的语义表示形式。关于统计的语言理解方法有很多研究，方法各不相同。文[52]采用类似 HMM 的方法进行建模，语义概念相当于 HMM 的状态，而识别单元则是状态的输出；文[53]中介绍了基于决策树的统计分析器，决策树的输入是语音识别结果对应的语义特征，输出是概率得分最大的分析树。文[54]改进了[52]的方法，不再使用语义概念作为状态，而是用多个语义概念组成的矢量作为状态，这样做的好处是能描述语句中跨度较大的语义之间的关系。

训练基于统计的语言理解模型不仅需要有足够的领域语料，还需要对这些

语料进行有效的语义标注。对每个领域的语料都进行收集和标注是非常耗时、耗力的，做起来非常困难，限制了基于统计的语言理解方法在口语对话系统中广泛应用。

2.1.3 识别器与理解器的接口

在口语对话系统较早的研究中，大多数语言理解器（Parser）[55][56]的输入是语音识别器输出的 N-best 句子候选列表。识别器对所有候选路径根据其声学得分进行排序，概率分高的路径排在前面，取得分最高的前 N 条路径输出，就是 N-best 句子候选列表。在不同的系统中，N 的选择各有不同，当 $N=1$ 时，识别器只输出一条最佳路径。对话系统 Wheels[55]中的理解器 TINA 对 N-best 候选列表（$N=10$）依次分析，从分析成功的候选中取声学分最高的作为语言理解的结果。

近年来，更多的对话系统采用词图（Word Graph）作为识别器和理解器的接口。词图是语音识别器的另外一种输出形式，声学搜索中保留下来的所有路径组成一个词网格（Word Lattice），对这个网格进行压缩（如去掉相同位置上的重复路径、根据一定的原则剪枝）、后处理（如引入新的知识并对网格进行重打分）即可得到识别结果的词图表示。一般说来，分析 N-best 句子候选比分析词图更加容易，因为可以直接使用传统自然语言理解领域的分析算法，如前面提到的 Chart Parser 算法，而对词图进行语义分析则需要对现有的分析算法加以改进，使之能够分析图结构。相对于 N-best 候选列表，规模相当的词图中包含了更多的信息[57]，因此，在词图的基础上进行语义理解的性能要比基于 N-best 候选列表的为优。

2.2 基于语义概念的语音理解框架

2.2.1 问题的提出

面对自然语音和口语现象，现有语音识别器的性能难以达到理解模块和对话管理模块的要求，大大降低了对话系统的可用性，这也是目前真正实用的对话系统很少的主要原因。

2.1.1 节中我们介绍了对话系统中的三种常用识别框架，它们从不同角度对

提高自然语音的识别性能进行探索：基于 N-gram 语言模型的连续语音识别方法针对领域特点，收集领域语料训练专门的语言模型，并使用了词类 N-gram 语言模型技术，提高识别器对特定领域的适应能力；关键词识别策略通过对关键词加权的方法，提高关键词的识别性能；基于模板的方法在识别搜索中充分发挥模板的预测性，提高识别性能。这三种方法各有其优点，也都有局限性：基于 N-gram 语言模型的连续语音识别方案有较高的识别正确率，但是这种高识别率是建立在经过充分训练的、领域相关的语言模型的基础之上，要收集足够的领域内语料训练特定的语言模型并不容易；关键词检出的语音识别策略在处理自发语音中的口语现象方面比较有优势，但在识别性能方面略差，尤其是在一些比较复杂领域中，另外，它还容易引入误警错误，给后续理解带来较大的负面影响；基于模板匹配的方法充分利用了上层的语言知识，对于符合模板的语句，识别效果非常理想，但是，该方法灵活性较差，一旦输入语句稍不符合模板，识别性能就急剧下降。

通过分析上面三种常用识别框架的优缺点，本文从对话系统的领域特点出发，提出一种新的基于语义概念的语音理解框架，目的是提高口语对话系统中识别和理解的性能，进而提高对话系统的实用性。

2.2.2 基本思想

考虑到现有语音识别技术对自然语音的识别性能远远达不到朗读语音的水平[24]，要想提高对话系统的识别性能，单从特征、模型等识别环节下功夫是不够的。口语对话系统中对人机对话内容的领域限制给我们提供了很多先验知识，可以利用这些知识来降低识别难度，提高识别性能。

本章工作的基本思想是：在语音识别中尽早地引入高层知识作指导，以弥补现有识别技术在识别自然语音时的不足。

高层知识对语音识别结果的影响会是怎样的呢？可以通过分析人类语言感知的过程来推测高层知识对识别性能的影响。日常生活中往往会遇到这样的现象：在一个嘈杂的环境中与别人交谈，如果你不用心去听，往往什么都听不到，而当你“用力”去听的时候，可以听到大部分内容，并理解对方的意思，所谓“用力听”可以理解为在听的同时，大脑对听到的信息进行处理和加工，把主题知识、上下文知识、语义知识以及语法知识与听来的语音信号一起处理，通

过综合判断，把原本并不清晰的信号转变成合乎句法和语义的句子；反之，如果你对交谈的内容缺乏最基本的了解，往往会听错很多内容，因为你缺少必要的先验知识对听来的信号加以处理。微软的研究人员也做过一个实验，让人耳与机器同时来听一批不成句子的词，发现机器识别（使用上下文相关音素模型）的效果比人耳还好些，而实际应用中，语音识别的性能却远远达不到人耳的水平。上述例子都说明高层知识在人耳识别语音时发挥着巨大的作用。

由此我们得到启发：如果在语音识别的过程中能及早的溶入高层知识，是否也有助于提高识别性能呢？事实上，已有研究者进行过这方面的尝试，并取得了不错的效果：语言模型的 Look-ahead[58]技术使得语言模型得以及早引入识别过程中，尽管 N-gram 语言模型所包含的高层知识非常有限，只是提供词一级的概率估计，而且由于计算量和训练数据的原因 N 通常只取 2 或者 3（对应 Bi-gram 或者 Tri-gram），但是它在识别中的及早应用还是对搜索进行了有效的指导，提高了语音识别的性能；另外，前文介绍中提到的基于模板匹配的识别方法对于模板覆盖范围内的语句识别率较高，也是因为该方法能够通过模板的方式将句法、语法知识在识别阶段加以应用。

考虑到 N-gram 语言模型能够引入识别过程的语言层知识过于简单，而且需要大量的训练数据，而模板方法又缺乏灵活性，本文提出基于语义概念的语音理解框架：语义概念是对话系统中很重要的高层知识，希望将语义概念知识及早地引入识别过程，用于指导搜索，以达到提高识别性能的目的。

下面从语义概念的重要性，语义概念的定义，以及语义概念在识别中的可用性三个方面具体介绍基于语义概念的语音理解框架，并第 3 小节中介绍该框架的具体实现。

2.2.3 语义概念的重要性

能否正确的理解用户意图（intention）是完成对话任务的关键。在基于规则的语言理解方法中，通过对句子进行句法分析和语义分析来理解用户意图。传统的分析器大都采用以词性作为终结符的文法规则，但对于汉语口语对话系统来说，基于词性的文法分析并不是最适合的：第一，汉语是表意语言，汉语句子的组成方式更加灵活，不像大多数拼音语言文字那样遵循简单、严格的文法；第二，基于词性的文法规则描述的只是句法层面的表面形式，对理解语言的深

层含义贡献不大。例如，图 2.1 给出了航班信息领域内的一个查询语句——“周五从北京到深圳的航班有哪些？” 在基于词性的文法规则下的句法分析结果（句法树），显然，这一结果对于理解句子含义的作用不大。

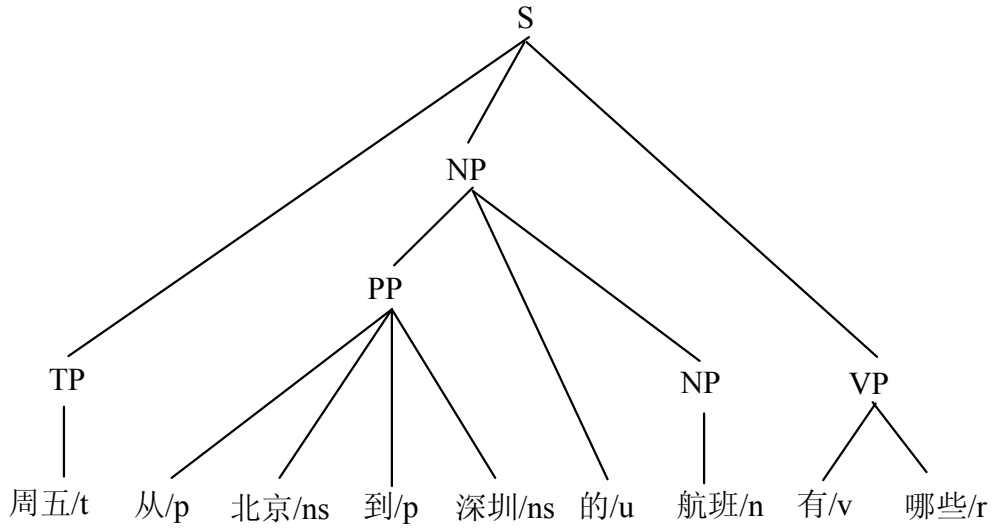

图 2.1 句法分析结果

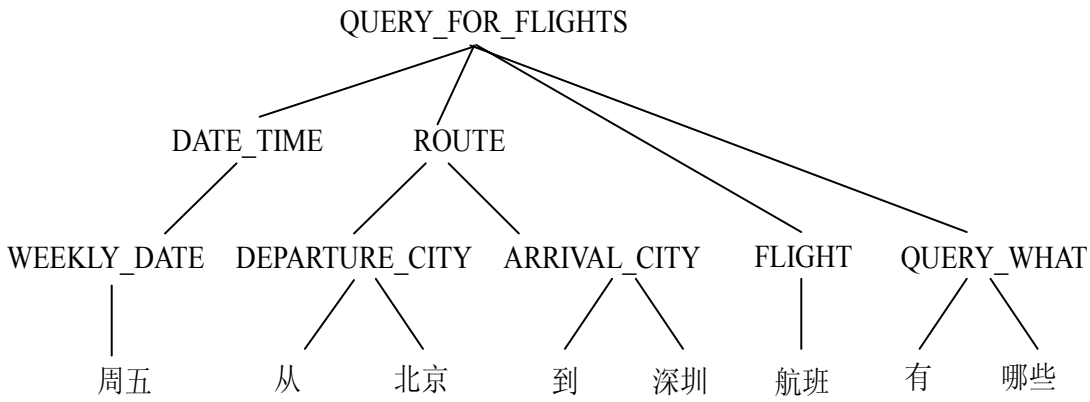

图 2.2 语义文法下的分析结果

相比用词性作为终结符的文法规则，基于语义类的文法规则更加适合汉语口语对话系统，更加有助于语义的理解。图 2.2 给出上面的例句在以语义类为终结符的文法规则下的分析结果，图中除叶子节点外的每个节点都是一个语义单元：根节点（QUERY_FOR_FLIGHTS）表示一个查询航班的意图，子结点 DATE_TIME 和 ROUTE 分别表示时间和航线两个查询条件，其中的航线又分为

起飞城市（DEPARTURE_CITY）和到达城市（ARRIVAL_CITY），根据图 2.2 所示的分析树我们可以很容易的了解用户意图。

从图 2.2 可以看出，语义单元对于理解用户意图起到非常重要的作用。文[59]认为“概念（Concept）是与任务相关的最小语义单元”，很多对话系统的语言分析模块都只对语义概念进行分析。语义概念在整个对话系统中无处不在，它是对话系统的核心内容：识别和理解是为了获得语义概念，对话管理模块则根据语义概念组成的语义表示完成对话状态转移和应答生成的任务。

考虑到语义概念在对话系统中的重要作用，我们提出以语义概念为中心的语音理解框架，用语义概念知识指导声学识别，识别结束的同时得到语义概念。在这一框架下，识别器的识别对象从词语变成了语义概念，理解时也直接根据语义概念以及它们之间的关系来分析语句，了解用户的意图。

2.2.4 语义概念的定义

对话系统中，语义概念是与任务相关的最小语义单元[59]，它们与对话任务中的信息点密切联系。通常，用户的每一句输入包括一个或者多个语义概念，每个语义概念都对应一个与任务相关的语义，这些语义是完成对话任务所必需的信息点。例如，句子“我想订一张明天上午去上海的机票”中有 4 个语义概念，分别是“我想订”、“一张”、“明天上午”、“去上海”，对应的语义如表 2.1 所示。

表 2.1 例句“我想订一张明天上午去上海的机票”中包含的概念及其对应语义说明

概念	语义
我想订	表示订票需求
一张	表示机票数量
明天上午	表示出发时间
去上海	表示目的地点

具体到某个特定的对话系统，语义概念的定义可以从两个方面考虑：

1) 根据对话系统后端数据库中的内容来定义语义概念。信息查询是对话系统最主要的应用之一，大多数提供信息服务的对话系统后端都与本领域的信息数据库相连，如航班信息库、天气数据库等，通过查询数据库，系统反馈给用户所需要的信息。在航班信息系统中，后端数据库中保存着大量的航班信息，每个航班又包括航班号、日期、时间、起始地点、目的地点、价格、票数、机型、航空公司等多项属性，其中每项属性都是与任务相关的语义单元，因此，可以根据这些属性来定义语义概念。

2) 另外，还要根据具体任务的特点，将本领域内常用的具有语义含义的表达定义为语义概念，它们也是正确理解用户语句的关键。表 2.1 中的 “我想订”就属于这类语义概念，假如忽略了这个概念，仅凭剩下的三个概念，系统很难准确地把握用户的意图，不知道用户究竟是想订票，还是想问机票价格或者什么别的信息。

语义概念的定义应该具有这样的特点：第一，概念的定义要满足理解句子语义的需要，也就是说，如果正确地获得了句子中的所有概念，那么一定可以成功地理解整句话的意思；第二，不同的语义概念之间相对独立，即每个概念本身都对应着一个独立的语义，而不需要依靠其它概念。

2.2.5 语义概念在识别中的可用性

相对于句子级表达，语义概念的表达方式较少，可以用一定数量的规则将其全部概括，例如航班信息系统中，与地点相关的语义概念可以用图 2.3 所示的几条规则来描述。语义概念的这一特点使其可以在不影响识别灵活性的前提下，引入识别过程，例如，将描述语义概念的上下文无关规则转变为有限状态自动机（Finite State Machine）用于指导识别搜索，可以提高语义概念单元的识别性能，同时也不会限制到语义概念组成句子时的灵活性。

```
// 城市名或者城市名列表
city_name_list → mat_city_name                                  // 城市名
city_name_list → mat_city_name tag_or mat_city_name             // xx 或者 xx
city_name_list → mat_city_name tag_and mat_city_name            // xx 和 xx
// 表示出发地点，如: (从) xx (出发)
sub_from → mat_city_name                                        // 城市名
sub_from → tag_from_here                                        // 从这里
sub_from *→ tag from mat_city_name                              // 从 xx
sub_from_1 → sub_from                                           // 从 xx
sub_from_1 *→sub_from tag_departure                             // 从 xx 出发
// 表示途径地点，如: 经 xx
sub_stop *→ tag_stop mat_city_name                              // 经过 xx
// 表示到达地点，如: 去 xx (或者 xx)/(和 xx)
sub_to *→ tag_to city_name_list                                 // 去 xx (或者 xx)
// 表示地点概念的顶级规则
info_fromto → mat_city_name                                     // 城市名
info_fromto → sub_from_1                                        // 从 xx (出发)
info_fromto → sub_to                                            // 去 xx
info_fromto → sub_from_1 sub_to                                 // 从 xx (出发) 去 xx
info_fromto → sub_from_1 sub_stop sub_to                        // 从 xx (出发) 经 xx 去 xx

注：1) 未在规则左项出现过的符号均为终结符
    2) “→”前面的“*”表示规则的属性
```

图 2.3 描述地点概念的规则

用语义概念知识来指导识别搜索还有一个优点：可以通过强调某个语义概念，将对话状态信息也引入识别搜索中。对话交互中，对话管理模块会根据当前对话状态，针对完成对话任务还缺少的信息点，主动向用户提出问题，并对用户下回合的语句内容有一个预期（focus expected）。通常情况下，用户会积极配合，根据系统提问做出相应的回答，很少出现答非所问的现象，也就是说，用户的应答与系统预期相一致。因此，可以通过强调与系统预期相关的语义概念让识别器更加关注系统期待的内容，进而提高系统感兴趣单元的识别率。

语义概念只是高层知识的一种，事实上，语句中还包含了其它的高层信息，如句法结构、韵律结构等。从理论上说，这些高层信息对识别搜索也会有指导作用，但是它们的规律复杂、灵活性强、领域特性也更加突出，对它们进行建

模非常困难，即使用复杂的模型描述了句子级的高层信息，也难以引入语音识别过程。

综上考虑，我们选择用语义概念知识来指导识别过程，一方面因为语义概念知识描述起来比较容易，能够在不影响识别灵活性的前提下方便的引入识别过程；另一方面，它具体较好的扩展性，可以通过强调不同语义概念在识别中进一步引入对话上下文信息。

2.2.6 系统框图

由于语义概念知识在识别过程中提前使用，使得声学识别结束的同时能够比较方便的得到语义概念识别结果，这一结果无需再经过语义分析过程，可以直接为后续对话管理模块所用。也就是说，在新提出的基于语义概念的语音理解框架下，对话系统的模块结构有所变化，不同于第一章中图 1.1 所示的典型对话系统的模块结构，而是将原来的识别模块和语言理解模块更加紧密地结合在一起，使之成为一个整体——语音理解模块，如图 2.4 所示。

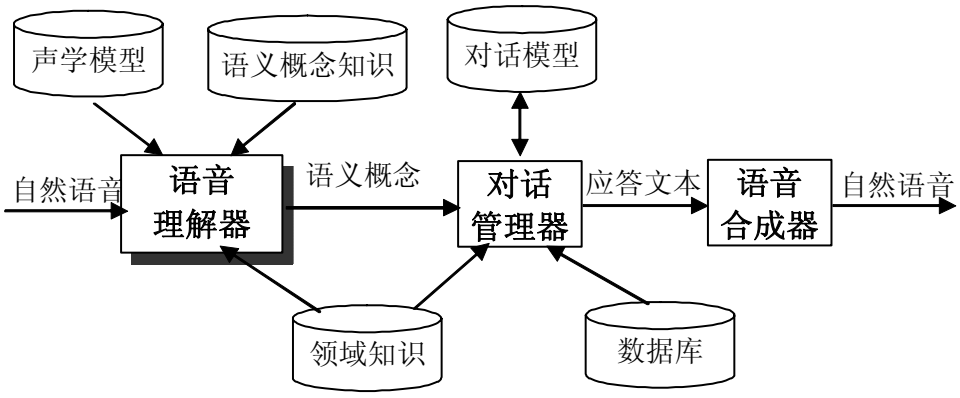

图 2.4 基于语义概念的语音理解框架下对话系统的结构图

在新的语音理解框架下，语言理解的结果以语义概念的形式输出，具有很好的灵活性：当整个句子无法完全分析成功时，可以只输出那些识别成功的语义概念，这样，即使系统只识别出一个语义概念，也有助于对话进程的推进，因为对话管理模块能够根据已知的语义信息，推测用户的意图，引导用户对话，最终达到完成对话任务的目的。

2.3 基于语义概念的语音理解框架的具体实现

本节中，我们以航班信息系统 *EasyFlight* 为例介绍基于语义概念的语音理解框架的具体实现。

EasyFlight 系统是作者所在的课题组研究的对话系统原型，它的数据库中含有国内航线、航班的各种信息，通过数据库查询，系统可以向用户返回以下信息：(1) 有无航线，(2) 有无航班，(3) 航班时间信息，(4) 机型信息，(5) 有无余票及票数，(6) 价格。

实现语音理解框架的前提是系统已经定义了领域词表和领域文法。在介绍语音理解框架的具体实现之前，我们先简单说明一下 *EasyFlight* 的词表和文法。目前，*EasyFlight* 系统的词表规模是 400 左右，400 个词根据词义分成 102 个词类。跟大多数对话系统一样，*EasyFlight* 使用基于规则的分析方法来理解语言，并针对口语对话提出了基于语义类的上下文无关增强文法[51]，通过对规则附加增强属性达到更好地处理语言中口语现象的目的，文法的终结符是词表中的词类，例如，所有地点名的词类都是 mat_city_name，像 mat_city_name 这样的词类作为终结符出现在文法中。

具体说来，上下文无关增强文法中定义了五种规则类型，分别是：

- ✓ 苛刻型（up-tying）规则
- ✓ 跳跃型（by-passing）规则
- ✓ 长程型（long-spanning）规则
- ✓ 无序型（up-messing）规则
- ✓ 交叉型（over-crossing）规则

每种规则类型的具体含义和举例说明见表 2.2 所示。

表 2.2 上下文无关增强文法的规则类型说明

规则类型 (表示符号)	特点	举例和说明
苛刻型 (*)	传统意义上的规则	dgt_d *→ ato_2 ato_10 ato_1_9 （表示日期的数字 21～29）
跳跃型 (无)	该规则右项间可以跳过一定数目的垃圾词	sub_week_day → ato_week ato_1to6 （如“星期啊三”可以跳过“啊”而归结成“sub_week_day”）
长程型 (~)	该规则右项间距离可以任意长	mark_q_is ~→ tag_is tag_question_mark （例如“是……吗”，可归结为“mark_q_is”）
无序型 (@)	该规则各右项之间不考虑顺序	key_info @→ info_time info_fromto （例如“明天 去上海的”，“北京到深圳 下午的”，都可以归结为“key_info”）
交叉型 (#)	该规则右项间在出现位置上允许交叉	confirm_request #→ mark_q_is key_info （例如，“是 明天上午的 吗”，可以归结为 confirm_request，虽然成分“是……吗”与“明天上午的”在出现位置上相互交叉）

2.3.1 语义概念知识在识别中的应用

要想用语义概念知识指导声学识别，必须先将其转换为识别器可以利用的形式。本文中，作者将语义概念知识用有限状态机（FSM，Finite State Machine）的形式加以表示，并用有限状态机作为识别框架来指导搜索。

将语义概念知识表示为有限状态机的形式需要经过下列步骤：(1) 定义语义概念；(2) 文法划分；(3) 对应每个语义概念生成一个有限状态机，称为子 FSM；(4) 在子 FSM 中标记描述语义概念的规则，并根据规则属性在子 FSM 中添加特殊弧；(5) 将所有的子 FSM 结合成一个 FSM。

1．语义概念定义

按照 2.2.4 节提到的语义概念定义原则，*EasyFlight* 系统中的语义概念定义

如表 2.3 所示，其中前 7 个是根据对话系统后端数据库的内容来定义的，后 3 个是根据领域特点定义的。

表 2.3 *EasyFlight* 系统中的语义概念定义

语义概念	说明	举例
Date_time	表示日期和时间的概念	五月一号上午，晚上八点之前到达的
Place	地点相关的概念	从北京到上海的，到广州或者深圳的
Ticket_num	表示机票数量的概念	五张，二十张
Price_info	机票价格	一千三百八，八百元整，八折
Airway_info	航空公司信息	国航，国际航空公司
Flight_num	航班号信息	CA8376，7158 次
PlaneType	机型信息	波音，747，空客 320
Selection	表示选择的概念	越早越好，最早的那次
QueryExpress	表示询问的概念	有哪些，是什么，多少钱，有……吗
Demand	表示请求的概念	帮我查一下，我想订

2. 文法划分

为了将语义概念表达成有限状态机的形式，先要通过文法划分得到描述每个语义概念的文法规则。所谓文法划分，就是将整个文法分成若干个规则子集，保证每个已定义的语义概念都对应一个规则子集。文法划分时需要注意 2 点原则：首先，表达一个完整语义概念的规则尽量分在一个子集内，以保证规则子集对语义概念的表达具有足够的覆盖度；其次，避免用无序型和交叉型的规则表示语义概念。文法划分的结果可能出现有些规则同时属于两个规则子集的情况，例如语义概念“Ticket_num”和“Price_info”对应的规则子集中都包含描述数量的规则；也可能有些规则不属于任何一个规则子集，如文法中的顶级规则，因为它们往往涉及到多个语义概念。

3．生成与语义概念对应的子 FSM

文法划分后，对每个语义概念，根据其相应的规则子集生成一个有限状态自动机（后文中把每个语义概念对应的有限状态机称为子 FSM）。不考虑规则的增强属性，用文[60]的方法可以将上下文无关文法转换为有限状态机的形式。

4. 对子 FSM 的规则标记

自然语音中的口语现象非常普遍，即使在表达一个较小的语义概念时也是一样，例如“从北京到那个上海的”这个句子中的“那个”显然是一个无意义的口头习语。除此以外，有些概念在空间上有着长程关系，比如“有……吗”、“是……吗”。第 3 步中生成的子 FSM 并不能覆盖上面的口语现象，也不支持描述长程关系的语义概念。因此，我们希望在有限状态机中添加一些特殊的弧，以提高子 FSM 的鲁棒性。

为了在子 FSM 中添加特殊弧，首先对有限状态机进行**规则标记**：将文法规则中非终结符对应的节点在状态机中标记出来，标记的内容包括该节点对应非终结符、生成该符号的规则以及经过路径节点。例如图 2.5 表示地点概念的有限状态机中（描述地点概念的规则参见图 2.3），节点 3 对应非终结符“sub_from”和“sub_from_1”，节点 4 对应非终结符“sub_from_1”，这个两个节点的标记内容如图中注释框所示，其中规则右项成分两边的数字表示该成分在状态机中的开始节点和中止节点，这些数字组成的序列就是规则经过的路径节点。图中节点 6、8、10 也有类似规则标注，因为篇幅的原因图中没有列全或者没有列出。

在规则标注的基础上，通过在某些节点上添加特殊的弧，以支持语义概念表达中的口语现象，并支持具有长程关系的语义概念。仍以图 2.5 为例说明添加特殊弧的原则。根据规则标注，图 2.5 中的节点 8 对应规则“info_fromto -> sub_from_1 sub_to”，得到这条规则的路径需要经历节点 1、节点 3 和节点 8，其中节点 3 对应描述非终结符“sub_from_1”的规则，节点 8 对应描述“sub_to”的规则，规则“info_fromto -> sub_from_1 sub_to”是跳跃型的，也就是说成分“sub_from_1”和“sub_to”之间允许跳过某些无意义的部分，因此它们两个的连接节点 3 上应该添加一条补白（filler）弧，以允许词表中定义的垃圾词跳过。

同样的理由，在节点 4、6、8、9 上也添加 Filler 弧。

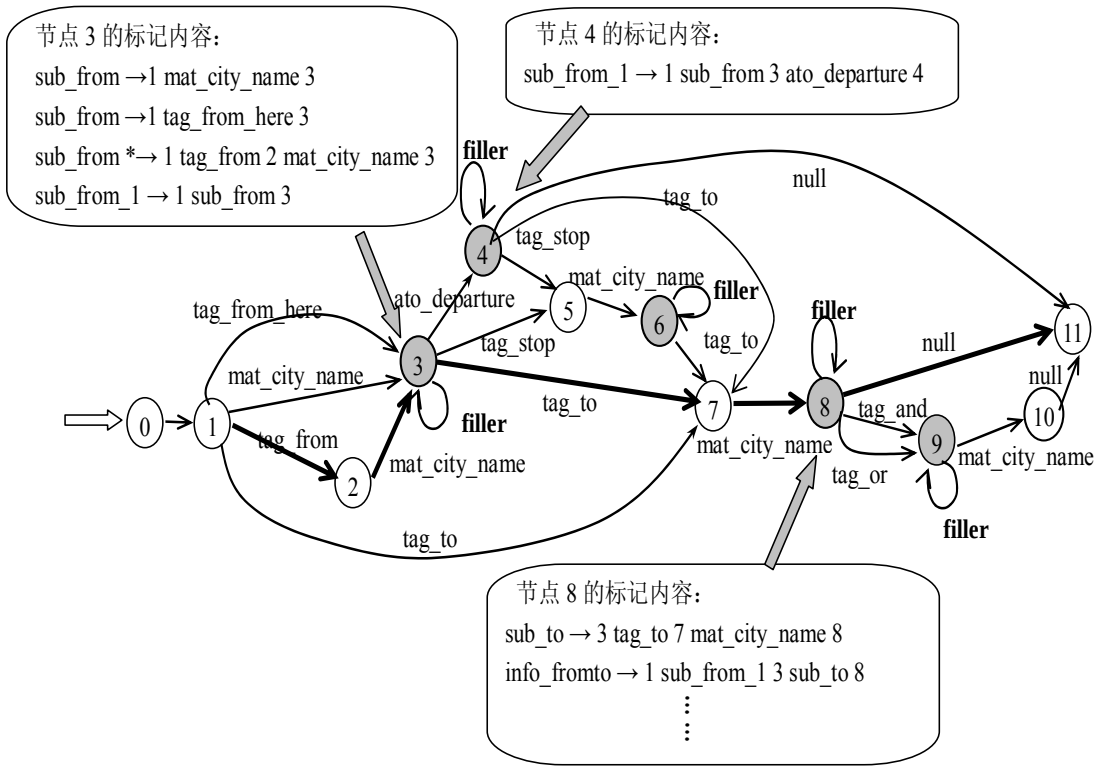

图 2.5 地点概念对应的有限状态机

对于长程型规则（见表 2.2 中的说明），用同样的方法在子 FSM 中找到特定的节点，在该节点上添加两条空弧：一条从当前节点指向 FSM 的外部，另外一条从 FSM 的外部指向当前节点。

在子 FSM 中添加必要的 Filler 弧和空弧后，有限状态机的灵活性大大加强。

5．生成指导搜索的语义概念网络

最后，将语义概念对应的所有子 FSM 并连在一起，并起点和终点之间添加一条 Filler 弧（以便允许语义概念之间出现一些插入语或者垃圾词），这样，形成一个新的 FSM，作为语义概念指导层加入到识别框架中，如图 2.6 所示。识别时，从语义概念指导层中有限状态机的起点开始，根据当前结点出发的弧上的内容确定下层可扩展的词，当下层的识别搜索到达词结尾的时候，回到语义概念层查询接下来可以扩展哪些词。通过这种方式，语义概念知识被用于指导

识别搜索。

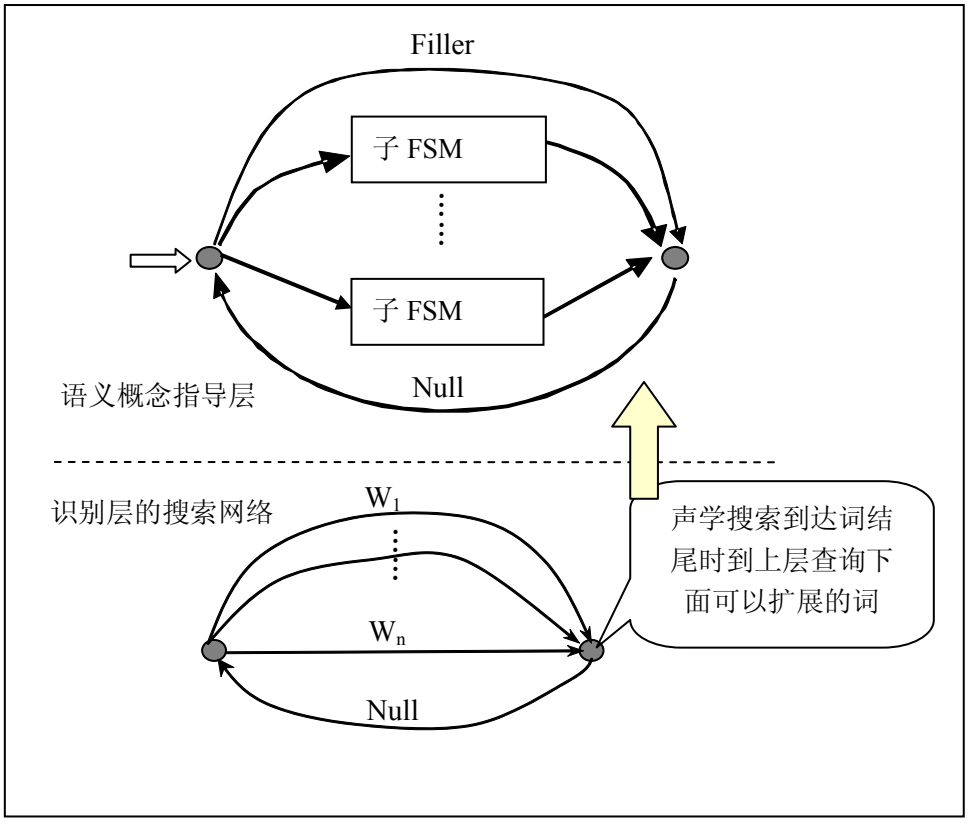

图 2.6 识别中引入语义概念指导的示意图

上面第 4 步中，我们在每个语义概念对应的子 FSM 中添加了一些特殊弧，因此整个 FSM 不但具有高度的预测性，也具有相当的鲁棒性。

2.3.2 语音理解框架下对话上下文知识的引入

这里所谓的上下文知识指的是当前对话状态，对话管理器记录着对话的历史，掌管着数据库的查询，历史记录和查询结果构成当前的对话状态，对话状态决定着对话管理器的应答方式和应答内容。采用混合主导策略的对话系统在必要的时候会主动向用户提出问题，而下一个回合的用户语句通常都是针对系统的提问的，换句话说，在对话管理器的引导下，用户语句中会包含系统期待的内容。当然，这种假设的前提是用户在交互过程中主动配合系统。

一般情况下，对话系统的用户为了获取自己需要的信息，都情愿积极配合

系统，因此，当前对话状态下的系统期待对于用户下一个回合的输入具有很强的提示作用。于是，系统期待的内容就成为识别中可加以利用的有效知识。

在基于语义概念的语音理解框架中，系统期待是通过语义概念引入识别过程的。根据当前状态下系统期待内容，在图 2.6 所示的识别框架中对描述这些内容的子 FSM 进行加权，可以引导搜索识别重点关注系统期待的内容。例如 *EasyFlight* 系统中，当已知出发地和目的地之后，系统期待知道时间信息，主动向用户提问出发时间，此时，通过增大时间概念对应的子 FSM 的权重，让识别器更加关注时间概念，可以提高时间概念的识别性能。

图 2.7 给出对话上下文知识引入识别过程的示意图，其中对最上层的子 FSM 用权重加以强调。

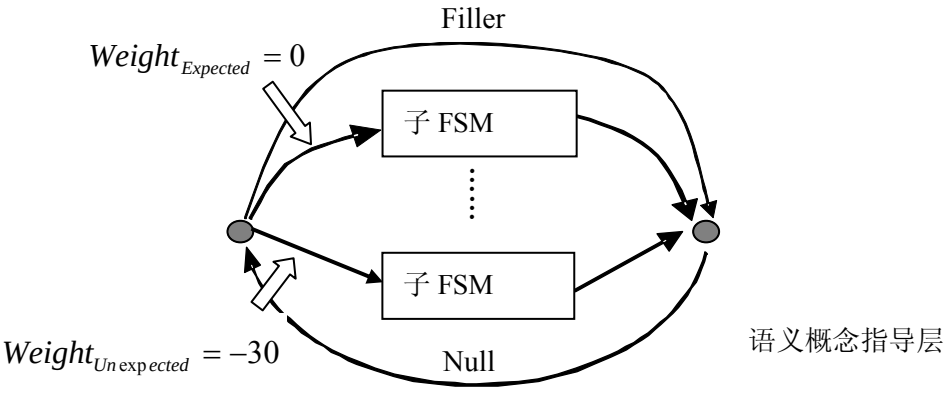

图 2.7 对话上下文知识引入识别过程的示意图

2.3.3 语义概念的解码过程

在 2.3.1 节中，为了提高语义概念对应的有限状态机的鲁棒性，我们对状态机内部的节点加入规则标记，这种规则标记也使语义概念识别和理解的合一过程得以实现，因此整个识别过程可以认为是对语义概念的解码过程。识别器的输出不再是原来的词序列或者词网格，而是语义概念本身，它们可以直接用于填写语义槽。

根据搜索路径和有限状态机节点对应的规则和非终结符，可以在搜索过程中逐步形成语义分析树。

图 2.8 中以语句“从北京飞往乌鲁木齐”为例，给出在识别搜索的同时得到语义概念分析树的过程。

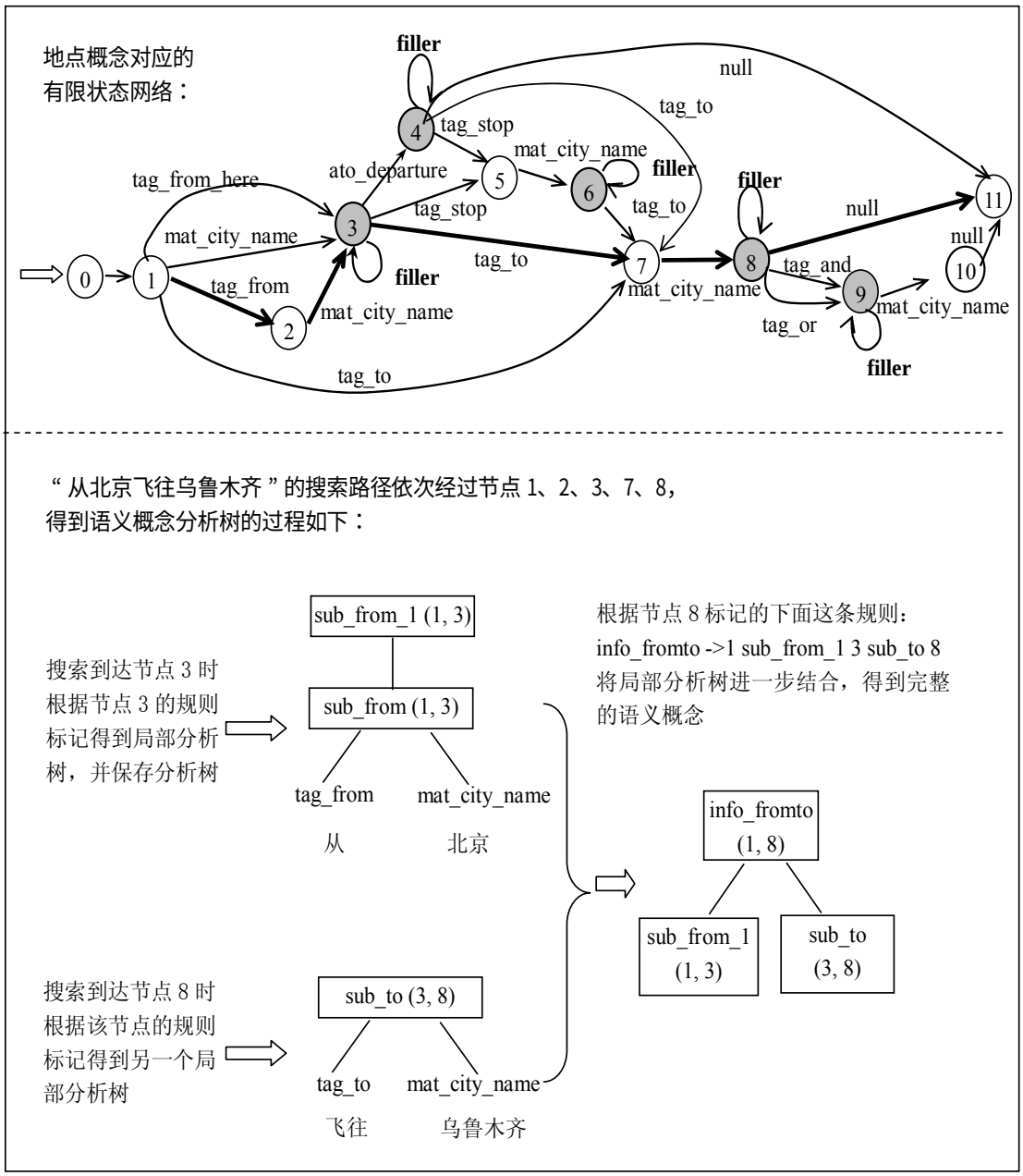

图 2.8 语义概念的解码过程

2.4 实验结果与分析

本节将通过三个实验来评价基于语义概念的语音理解框架。

2.4.1 实验背景和实验数据

口语对话系统 *EasyFlight* 是评价基于概念的语音理解框架的背景对话系统，它的词表定义、文法定义、实验中使用的声学模型和测试数据说明如下：

词表定义：*EasyFlight* 中的词表规模为 398，其中最主要的部分国内城市名和航空公司名称。

文法定义：*EasyFlight* 使用基于语义类的上下文无关增强文法，文法规则有 295 条，其中相当一部分是描述日期、时间概念的规则。

声学模型：*EasyFlight* 识别前端使用的声学模型是使用 863 朗读语音数据训练得到的，声学特征为 42 维 MFCC 特征。

测试数据：本领域内的 500 条语句，是 5 个录音者以自然语音的方式录制的，每人 100 句（16K 采样）。

2.4.2 实验设计

为了更全面的评测本文提出的基于语义概念的语音理解框架，我们设计了三个实验，分别从识别结果、理解结果和对话上下文知识的作用着手。

实验 1：评价语音理解框架下的识别性能，该实验用了两个评价指标，一是常用的音节正确率，二是针对口语对话的特点定义的语义概念内部的音节正确率。

$$\text{音节正确率} = \frac{\text{正确识别音节个数}}{\text{句子中音节的总数}} \times 100\% \qquad (2\text{-}2)$$

$$\text{语义概念内的音节正确率} = \frac{\text{语义概念中正确识别音节个数}}{\text{语义概念内包含的音节数}} \times 100\% \qquad (2\text{-}3)$$

口语对话中可能会含有较多的没有意义的插入语、口头习语等等，系统并不关心这些词，即使识别出来了，也会在分析理解中被忽略；真正对理解语义有用的是属于语义概念内的那些词，因此我们定义了语义概念内的音节识别率，

用来评价系统真正关心的那部分词的识别效果。

实验 2：考察语音理解框架下的理解性能。实验中，以槽正确率（Slot Correction Rate）作为评价标准。*EasyFlight* 中用三元组(*Name*,*Value*,*Type*)定义一个语义槽，其中 Name 是槽名，Value 是槽值，而 Type 表示槽的类型，例如“从北京到上海”对应两个语义槽，分别是(地点, “北京”, 1)和(地点, “上海”, 3)，这两个都是地点槽，类型 1 和 3 分别表示出发地点和到达地点。只有当语义槽的名称、值和槽类型都正确的情况下，我们才认为这个槽是正确的。语义槽正确率的定义如公式（2-4），公式中没有考虑插入错误：

$$\text{槽正确率} = \frac{\text{正确填写的语义槽个数}}{\text{句子中语义槽的总数}} \times 100\% \qquad (2\text{-}4)$$

实验 3：考察对话上下文知识对系统性能的影响。只有在真正的多回合的交互中，对话管理器才能给出期待内容，我们的测试语句是一些前后没有关系的语句，无法应用对话上下文知识。因此，在前两个实验中，识别中没有引入对话上下文知识。本实验中，我们用人工标注的系统期待来模拟对话管理器给出的期待内容，并根据人工标注对识别框架中语义指导层的语义概念进行不同的加权，评价引入对话上下文知识前后的系统识别性能和理解性能的变化。

2.4.3 实验结果与分析

实验 1：评价基于语义概念的语音理解框架下的识别性能

本实验中，基于语义概念的语音理解框架与两种对话系统中两种常用的识别策略进行对比：一是关键词识别的方法，二是模板匹配的方法。

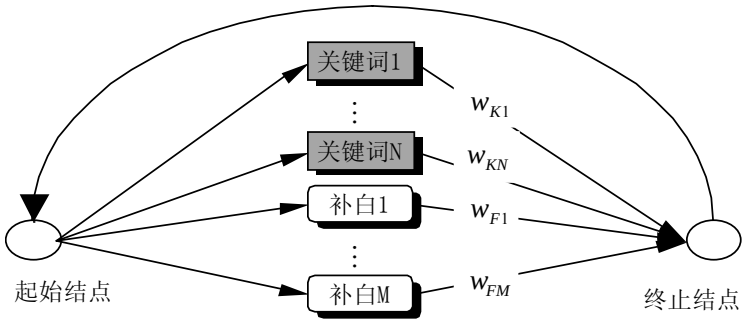

图 2.9 关键词方法的识别框架[42]

下面，简单说明一下关键词识别方法和模板匹配方法的实现。图 2.9 给出了关键加权的捡出框架，词表内的关键词和补白被加以不同的权值以突出关键词，本文中用使用补白模型的定义为汉语中出现的 418 个音节。模板匹配的识别策略中，将全部文法组织成一个有限状态机，有限状态机经过最小化和确定化[61]后用于指导识别搜索过程。

表 2.3 给出了基于语义概念的语音理解策略与另外两种识别方法的比较结果。对比关键词识别策略和模板方法，本文提出的方法在音节识别正确率上至少提高了 5 个百分点的绝对值，错误率下将分别达到 36.8%和 26.9%。新方法的识别性能高于基于模板匹配的方法，这主要是因为在识别中引入语义概念知识的时候，充分考虑了鲁棒性问题，在指导搜索的有限状态网络中添加一些 Filler 弧，处理自然语音中的口语现象。

表 2.3 的最后一列给出了语音概念内部音节识别的性能：相比基于关键词方法和模板匹配方法的两个基线系统，语义概念内的音节正确率提高 10 个百分点左右，错误率下降分别达到 58.0%和 47.1%。在这一评价指标下，基于语义概念的语音理解框架的识别性能比两个基线系统有大幅提高，这一结果说明语义概念知识可以有效的指导识别搜索，同时也说明新方法对音节识别性能的提高主要来自于对语义概念部分识别能力的加强。

表 2.3 基于概念的语音理解策略的识别结果(%)

方法	音节正确率	语义概念内的音节正确率
关键词方法（基线系统 1）	70.82	76.47
模板方法（基线系统 2）	74.80	81.29
基于语义概念的语音理解方法	**81.57**	**90.11**

实验结果表明，基于概念的语音理解框架通过将规则描述的语义概念知识及早地引入识别搜索过程大大了提高对话系统中的语音识别性能。

实验 2：理解性能的评价

基于语义概念的语音理解框架的一大特点就语义概念的理解和识别过程是一体的，识别完成的同时可以得到语句中包含的语义概念，这样做的好处是不必再对识别结果（词序列或者词网格）进行语义分析，识别结束后即可根据同时得到语义概念填写语义槽，达到理解用户的意图的目的。表 2.4 给出了语音理解的性能，与之对比的是先识别、后根据识别结果进行理解的方法。这两种方法的前端识别都是采用基于语义概念的语音理解策略，也就是说，先识别、后理解的方法在识别过程中也同样利用了语义概念知识。

表 2.4 基于概念的语音理解框架的理解性能(%)

方法	**槽正确率**
先识别、后理解方法	75.97
语音理解方法	86.33

从表 2.4 的结果看出识别理解一体化的方法具有更高的槽正确率，相比先识别、后理解方法，槽错误率下降 43.1%。对结果分析后发现，理解性能的提高主要是因为语音理解方法的鲁棒性较高。识别理解一体化的方法根据识别得到的语义概念直接填写语义槽，而先识别、后理解的方法需要根据文法分析的结果完成语义槽，因为文法覆盖度不够而导致的分析失败影响了槽正确率。例如，语句“我想买一张机票到深圳，从北京飞，有吗？”，因为文法没有覆盖“到深圳从北京飞”这样的表达，导致了分析失败，只能得到语句一部分内容（我想买一张机票到深圳，有吗？）对应的语义槽。以上分析说明，先识别、后理解的方法的理解效果除了与声学识别性能相关外，还与文法的覆盖程度有关；而识别理解一体化的方法对句子一级的规则没有过多的要求，只要描述语义概念的规则具有足够的覆盖度就能够取得较好的理解效果，因此具有较强的鲁棒性。

识别理解一体化的方法语义槽的正确率方面优于传统的先识别、后理解的方法，但是也存在不足，它的语义槽插入错误率为 24.50%，高于传统方法 16.97% 的插入错误率。下一章中我们将针对语义概念的插入错误和替换错误研究语义概念的置信度打分。

实验 3：对话上下文知识的作用

测试数据中有 18%的句子是用户回答系统关于时间概念的提问的，也就是说在用户说出这些语句之前，系统期待就它们与时间概念有关。用人工标注的期待内容来模拟系统给定的期待后，对这 18%的数据重新进行了识别实验，结果如表 2.5 所示。

表 2.5 对话上下文知识在识别和理解中的作用(%)

	音节正确率	语义槽的正确率
没有上下文指导	85.7	71.8
识别中引入上下文知识	**88.6**	**80.0**

对比没有系统期待内容提示的情况，引入上下文知识后，系统的识别性能和理解性能都有一定程度的提高，音节错误率下降和语义槽错误率下降分别是达到 20.3%和 29.1%，这一结果说明对话上下文知识对前端的声学识别具有一定的指导意义。

2.4.4 讨论

评价基于语义概念的语义理解框架时，没有设计与基于 N-gram 的连续语音识别策略的比较实验，主要是基于两个方面的原因：

第一，基线系统实现困难，因为没有足够的领域内语料，无法训练特定领域的 N-gram 语言模型，如果一般领域的语言模型，识别性能比较差，不能说明问题。

第二，提出基于语义概念的语音理解框架的目标之一就是减少对话系统开发初期对领域数据的依赖程度，而是从通过基于知识的方法提高识别性能。从这个角度出发，基于概念的语音理解策略可以不与使用领域 N-gram 模型的连续语音识别策略进行比较。

2.5 小结

本章针对口语对话系统中识别性能不佳的问题，提出了基于语义概念的语

音理解框架，将上层语义概念知识及早的引入识别搜索中，该框架的主要特点如下：

第一，语义概念知识以规则的形式表述，并转换成有限状态机的形式在识别中加以利用，避开了特定领域数据收集和标注的困难；

第二，通过在有限状态机中添加特殊弧（如补白弧），以支持表达中的口语现象，保证了基于语义概念的语音理解框架具有足够的鲁棒性；

第三，语义概念知识在识别中提早应用后，实现了语义概念解码过程，识别器直接输出语义概念，也就是说，在新的框架下，没有单独的语音识别模块和语言理解模块，取而代之的是语音理解模块，它一次完成前面两个模块分两步完成的任务；

第四，新的框架具有良好的可扩展性，可以很方便在框架中继续引入新的知识以提高语音理解的性能，如对话上下文知识。

实验结果表明：基于语义概念的语音理解框架通过将上层语义概念知识及早的引入识别搜索中，大大提高了核心语义概念的识别性能，使得语义概念中音节识别正确率达到 90.11%；语义概念理解和识别的合一过程提高了理解的鲁棒性，最终的语义槽正确率达到 86.33%；另外，在框架中引入对话上下文知识后，语音理解的性能进一步得到提高。

总之，基于语义概念的语音理解框架有效地提高特定领域下对话系统的识别性能和理解性能，具有理论价值和实用价值。

第3章 语义概念的置信度研究

置信度测量（Confidence Measure）是语音识别中一个相对较新的研究课题，它是一种语音确认技术，可以给出语音识别结果的可靠性评价。语音确认最初主要用在关键词识别领域[62]，目标是在几乎不影响关键词正确捡出率的前提下，尽可能地降低误警率。随着对话系统的深入研究，语音确认技术逐渐被引入其中。对话系统中，语音识别和语言理解都可能会出现错误，因此，在缺乏语音确认的情况下，系统一般会采取信息确认的策略，对每一个信息点逐一确认，这不可避免地增加了很多乏味、低效率的对话回合，导致完成任务所需的对话回合数的增多，影响了用户对系统的满意度；而如果系统不对信息点逐一确认，结果可能会更糟，因为对于那些直到对话过程快要结束时才发现的理解错误，系统恢复起来更加困难，会影响到整个对话系统的任务完成率。语义概念置信度研究的目的就是通过置信度估计，接受识别正确的语义概念，拒绝那些不正确的识别结果，起到减少对话回合、保证整个对话过程更加有效的作用，从而提高对话系统的性能[31]。

置信度的研究可以从两个方面着手——提出新的具有区分度的确认特征和提出区分能力更强的确认模型。本章主要介绍作者在确认特征方面的研究工作：针对语义概念，提出了理解分析层面和韵律层面新的置信度特征。

本章的内容安排如下：第一小节简单介绍置信度研究的基本原理和研究现状；第二小节分析基于语义概念的语音理解框架下的典型语义概念错误，提出理解分析层面的置信度特征；第三小节提出使用韵律边界作为一种新的置信度特征；第四小节给出实验结果和分析；最后一节是对本章内容的总结。

3.1 置信度研究的现状

置信度研究涉及以下三方面的内容：（1）提取有效的置信特征；（2）联合多种置信度特征的确认模型；（3）置信性能的评测指标。下面三小节逐一对其进行介绍。

3.1.1 置信特征

最早研究的置信特征来自于声学层面[63]。传统语音识别算法一般利用最大后验概率决策规则进行识别，识别结果应该满足下面的公式：

$$\begin{aligned}\hat{W} &= \arg\max_{w} p(W \mid X) \\ &= \arg\max_{w} \frac{p(X \mid W)\,p(W)}{p(X)}\end{aligned} \qquad (3\text{-}1)$$

理论上，给定 X、识别结果为$\hat{W}$ 时，根据后验概率 $P(\hat{W} \mid X)$ 就可以评价识别结果的可靠性，后验概率本身就是声学层中很好的置信度特征。然而，由于对给定的 X，$P(X)$是常量，在实际识别中通常被忽略不计，即识别结果给出的是词表中相对最匹配的词，而不是置信度足够大的词。但是在计算代表置信度水平的后验概率时，必须估计 $P(X)$。

目前有多种估计 $P(X)$的计算方法。All-phone 的方法[64]对词表中所有模型得分累加计算 $P(X)$

$$p(X) = \sum_{W} p(W)\,p(X \mid W) \qquad (3\text{-}2)$$

这种方法的计算量非常大，尤其是在大词表的连续语音识别中，必须计算每一个词模型 W 对应的概率值 $P(X|W)$。Catch-all 方法[65][66]将搜索空间的语法限制去掉，任何一个词都可以连接其他所有的词，任意一个词序列都可以被识别出来，搜索出来的最佳路径的似然值即可近似为 $P(X)$。也有文献仅仅根据识别结果的前 N 个词候选近似计算 $P(X)$[67]。

除了后验概率，搜索结束后得到的词网格中也包含着很多有用信息，可以作为置信度特征[68][69]，例如与候选词在时间处于并列竞争位置的其它词的个数，显而易见，竞争词越多说明候选词越不可靠。

语言模型回退(Back-off)的情况和语言模型分也可以作为一类置信度特征[70][71]，从另一个角度评价候选结果的可靠性。文[72]的结果表明，在对话系统应用中，语言模型层面的置信度特征其确认性能要优于基于后验概率的声学层置信特征。

3.1.2 确认模型

在得到多种置信特征以后，所面临的问题就是如何联合多种置信特征给出最后的确认结果。语音确认要解决的实际上是一个分类问题，即将待确认的识别结果分成两类(“对”或“错”)即可，当然在一些系统中还需要标注出错误识别的错误类别，在这里我们暂不考虑这种情况。大多数分类技术和方法在这里都是适用的，例如线性分类方法、决策树、神经元网络方法以及支持向量机法[73]等，其中一些分类方法还可以给出[0, 1]区间之内的分数，作为待确认词可靠程度的概率值。本文的实验采用 Fisher 线性分类方法作为确认模型，下面简单介绍其工作原理[74]。

$$g(x) = w^T \cdot x + w_0 \qquad (3\text{-}3)$$

Fisher 线性分类法通过公式(3-3)把 d 维空间的样本投影到一条直线上，形成一维空间，这样可将 d 维分类问题转化为一维分类问题，式中 $x=[x_1, x_2, \ldots, x_d]^T$ 是 d 维特征向量，$w=[w_1, w_2, \ldots, w_d]$为权向量。在所有投影方向中，对样本分类效果最好的那个方向是最优线性投影，它应该尽可能达到两个标准：第一，d 维样本投影后在一维空间里各类样本尽可能分得开一些，即两类均值之差越大越好；第二，各类样本内部尽量密集，即类内离散度越小越好。因此，Fisher 准则函数定义为上述两项的商，只要找到使 Fisher 准则函数取极大值的投影方向即找到了最优的投影方向。

3.1.3 评测指标

具体介绍评测指标之前， 表 3.1 先给出一些符号的定义。

表 3.1 评测指标中所使用符号的定义

符号	意义及说明
N	需要识别的结果总数
N_c	识别正确的结果个数
N_e	识别错误的结果个数
$N_{c\text{-}E}$	识别正确却被系统认为是错误的结果个数
$N_{e\text{-}C}$	识别错误却被系统认为是正确的结果个数

语音确认的评测指标有以下几种：

错误拒绝率(FRR, False Rejection Rate)，表示错误拒绝比率，即：

$$FRR = \frac{N_{c-E}}{N_c} \quad (3\text{-}4)$$

错误接受率(FAR, False Acceptance Rate)，表示错误接受比率，即：

$$FAR = \frac{N_{e-C}}{N_e} \quad (3\text{-}5)$$

错误拒绝率和错误接受率两者之间是有关联的，错误拒绝率越高错误接受率越低，反之亦然。用 ROC（Receiver Operating Characteristic）曲线[75]可以很好地描述两者之间的关系，在坐标系中，ROC 曲线越靠近坐标轴表示语音确认的性能越好。

等错误率(EER, Equal Error Rate)：在坐标系中 ROC 曲线与从左下角到右上角的对角线的交点，可以认为是错误拒绝率和错误接受率的最佳折中方案。等错误率越小代表语音确认的性能越好。

确认错误率（Confidence Error Rate）[76]同时考虑错误拒绝和错误接受的情况，即：

$$CER = \frac{N_{c-E} + N_{e-C}}{N} \quad (3\text{-}6)$$

在多数语音确认的应用中，对 FRR 要求比较严格，不能过大，否则正确候选损失较大，会带来很坏的负面影响。因此，也将 FRR 固定时的 FAR 的值作为语音确认的评测标准。

3.2 语义概念的置信度确认

最初，对话系统中的置信度研究较多地集中在词一级（Word Level）和句子一级（Utterance Level）[31][77][78]，近几年，人们开始研究语义概念方面的置信度打分[72][79]。事实上，语义概念才是对话过程中的核心内容，是系统了解用户意图的关键：如果语义概念识别错误，即使句子中大部分识别结果都是正确的，也会造成系统理解错误，影响对话进程；从另一个角度说，即使一个语句中大部分的识别结果都不可靠，但只要有一、两个语义概念识别正确，系统就可以

或多或少的了解用户的意图，有助于对话任务的完成。因此，对话系统中，对语义概念的正确性进行评价十分必要，有了语义概念的置信度评价，甚至可以不考虑词和整个句子的置信度。本节将介绍作者在语义概念置信度打分方面的工作，研究重点在提出新的置信度特征方面。

3.2.1 问题的提出

识别错误包括替换错误、插入错误和删除错误三种类型，要正确理解用户意图，系统必须通过多回合的提问与确认来纠正替换和插入错误，这使得对话回合数大大增加，极大的影响了系统性能。替换错误和插入错误是语义概念置信度研究的主要目标，即在保证大部分正确识别结果被接受的前提下，通过置信度打分拒绝替换错误和插入错误，以减少不必要的对话回合。

EasyFlight 中语义槽识别错误的具体情况见表 3.2（实验说明见 2.4.3 的实验二），其中语义槽的总错误率定义如下：

$$\text{总错误率} = \frac{\text{槽替换错误}+\text{槽插入错误}+\text{槽删除错误}}{\text{句子中语义槽的总数}} \times 100\% \qquad (3\text{-}7)$$

表 3.2 *EasyFlight* 中语义概念错误的具体情况(%)

	替换错误率	**插入错误率**	**删除错误率**	**总错误率**
语音理解	6.83	24.50	6.83	38.16
先识别、后理解	9.40	16.97	14.79	41.01

从上表中看出，不考虑语义槽插入错误的情况下，基于语义概念的语音理解框架具有较好的性能，槽正确率达到 86.34%，但是语义槽插入错误率较高，这对后续对话过程造成了一定的干扰，因此有必要对语音理解的结果进行置信度评价，以便及早地发现错误。

句子一级和词一级的置信度打分中多使用来自声学层面的确认特征，主要以后验概率为主。基于语义概念的语音理解框架下，以后验概率为主的声学置信特征在评价语义概念的可靠程度方面，性能不是很理想，这主要有两个原因：第一，计算后验概率时，需要估计 $P(X)$，出于计算量的考虑一般不使用 Catch-all

的方法估计 $P(X)$，而是用前 N 个词候选来近似计算。在基于语义概念的语音理解框架下，高层语言知识及早的引入识别搜索过程中，限制了搜索空间，不符合搜索限制的候选不参与打分，在这种情况下，前 N 个候选不再是声学上得分最高的 N 个，而是符合限制条件的候选中得分最高的 N 个。也就是说，在基于语义概念的语音理解框架下，利用前 N 个候选估计的 $P(X)$不够准确，这会影响语音确认的性能。第二，语义概念的错误并不完全是由语义概念内部的识别错误引起的，有些语义概念错误是因为受到前面识别错误的影响，例如“十二月二十四日的”被错误识别成两个概念“十二元”（对应语义槽插入错误）和“二十四日”（对应语义槽替换错误），对于后面一个概念来说，其中的每一个词都识别正确了，但是语义上不完整。显然，基于后验概率的声学层置信度特征对于这种错误无能为力。

综上所述，为了更好地对语义概念进行置信度确认，必须引入新的置信度特征。

3.3 分析理解层的置信度特征

本节提出分析理解层置信度特征，主要包括两个方面：一是描述连续语义概念之间相关性的特征；二是评价语义分析结果可靠性的特征。

3.3.1 描述语义概念之间相关性的特征

文[71]的研究表明，即使在词一级的语音确认中，来自语言模型的置信特征其确认性能也要优于声学层的置信度特征。受此启发，我们认为句子中连续几个概念之间的相关信息应该也是一种有效的置信度特征。因此，参考 N-gram 语言模型，可以用下述三个条件概率来估计连续语义概念之间的关系（可以认为是概念语言模型，即 Concept Language Model）：

$$p(C_i \mid C_{i-1}) \tag{3-8}$$

$$p(C_i \mid C_{i+1}) \tag{3-9}$$

$$p(C_i \mid C_{i-1}, C_{i+1}) \tag{3-10}$$

其中 C_i 表示语义概念，公式（3-8）考虑当前语义概念与前一个概念的关系，

（3-9）考虑当前概念与其后面概念的关系，而最后一个公式则考虑了以当前语义概念为中心的概念三元组出现的情况。特定领域的对话系统中涉及到的语义概念并不多，一般只有几十个，因此，要得到上面三个概率分并不困难，可以用实际语料训练，也可以用文法规则生成的伪数据来训练得到。

3.3.2 语义分析层的置信度特征

基于语义概念的语音理解框架的核心思想是将语义概念知识尽早地用于指导搜索过程，识别中语句中语义概念相关的部分被充分强调，从而提高了语义概念的识别正确率。在提高识别正确率的同时，这种强调也不可避免地会引入一些插入错误。地点和时间是航班信息领域最重要的两个概念，很多插入错误都出现在这两个概念上，占所有插入错误的 67.3%。分析这部分错误后发现：对于插入错误的语义概念来说，分析得到它们的语义规则在实际语料中并不常用。因此，我们将**得到当前语义概念的那条规则的概率分**作为评价语义可靠程度的特征之一：分析得到语义概念的规则越不常用，该语义概念的可靠性就越差。

图 3.1 中给出了规则概率起作用的一个例子，这个例子中有一个插入错误“深圳”， 但从语义概念语言模型分上看，这个错误难以发现，但是，得到“深圳”这个地点概念的规则是图 3.2 中的所列的第一条规则，该规则的概率分只有 0.07，也就是说训练语料中直接用图 3.2 规则 1 得出地点概念的情况比较少，因此，“深圳”这个地点概念的可信度较低。

标注结果：	本	周	五	的	三张	票价	多少
识别结果（拼音）：	ben zhou	wu filler		shen_zhen	piao_jia	duo_shao
（文本）：	本	周	五		深圳	票价	多少

图 3.1 语义概念错误举例

1. info_fromto → mat_city_name (0.07)	如：“北京”
2. info_fromto → sub_to (0.47)	如：“去上海的”
5. Info_fromto → sub_from_1 (0.05)	如：“从上海起飞”
3. info_fromto → sub_from_1 sub_to (0.33)	如：“从北京飞上海”
4. info fromto → sub from 1 sub stop sub to (0.08)	如：“北京经广州到湛江”

图 3.2 描述地点概念的顶级规则

另外，在语义分析层，还考虑了下面两个特征，它们从整个句子的全局出发，考虑语义概念的置信程度。

- ✧ 相同语义槽的数目：该特征反映分析结果中是否存在**语义槽冲突**，一般情况下，一个句子中表示同一概念的语义槽应该只有一个，如果出现两个或者两个以上，那么其中很可能有识别错误的。
- ✧ 在整句分析中，当前的语义概念是否与句子中其它概念一起可以被更高一级规则成功分析：这一特征考察当前语义概念能否与其它语义概念互相配合，得到更高层的语义。

3.4 基于韵律信息的置信度特征

3.4.1 出发点

在语音识别中，识别错误往往会伴随着时间边界错误，并引起识别结果互相干扰的现象：如图 3.3 所示，如果某一个词的识别发生替代和插入错误，那么很可能会发生识别结果的前后时间边界与正确的边界不一致的现象，边界的不准确会影响到后续词的识别准确度。

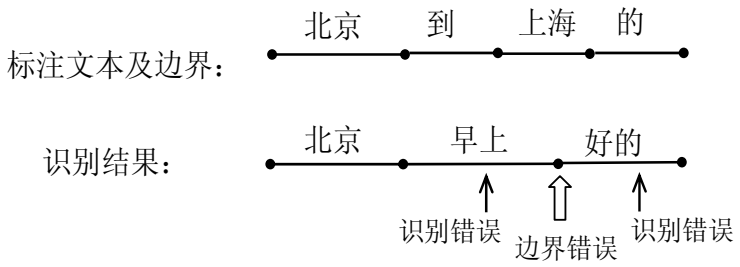

图 3.3 识别结果互相干扰现象示例

识别错误会导致边界错误，反过来想，如果能够检测到边界错误，那么可以推测错误边界前后的识别结果有可能不准确。于是，我们考虑将边界信息作为一种置信度特征用于语音确认。

可以用韵律边界与识别得到的语义概念边界进行对比，以此判断语义概念的边界是否正确。在日常会话中，人们说的每一句话并不是字和词的简单组合，而是融入了很多韵律（Prosody）信息（如停顿、重音等），反

映出说话人的情绪和对所说问题的态度。这些字面外的信息对人们正确理解说话人意图起到非常重要的作用。从韵律在语言表达中的作用出发，可以这样理解它的本质[80][81]：韵律是语音产生过程中的系统组织者，它将各种语言单位组织成一句话或者一组连贯的话语。对于如何定义韵律，目前并没有一个统一的看法，但是研究人员普遍认同韵律同基频、时长、音高、强度等声学特征相关[81][82]。韵律边界信息是众多韵律信息的一种，它主要表现在短语或者句子之间的停顿上，无法在文字上反映，但是可以从听觉上感知。韵律边界可以将一个长长的句子按照一定的语义分段，帮助人们理解语义[83]。根据韵律边界的特点，我们可以将识别得到的语义概念边界与韵律短语边界相比较，看两种方法得到的边界是否一致，以此来评价识别边界的正确性，进而推测识别结果的可靠程度。

韵律边界检测时主要使用韵律相关的声学特征，主要包括基频、能量、时长等。这些特征不同于语音识别中采用 MFCC 特征，在语义概念确认过程中，引入韵律边界特征，相当于在确认中引入了新的知识源，有利于确认性能的提高。

3.4.2 韵律边界的检测

下面简单介绍一下韵律边界检测的方法和实验结果[84]。

3.4.2.1 韵律边界的定义和标注

一般认为，韵律是存在级层结构的，但是究竟应该如何划分层级还存在着一些争议，比较公认的一种分为音节、音步、韵律词、韵律短语、语调短语和句子[85]。针对口语对话系统中的句子相对较短的特点，我们定义下面 4 类韵律边界：

- ✧ **音节边界**：字与字之间的边界，默认每个字后面都是一个音节边界。
- ✧ **韵律词边界**：韵律词与韵律词之间的边界。注意：此处说的韵律词可以是一个或者几个语法词，也可以是一个语法词的某一部分，在韵律词中间没有可感知的停顿。

- ✧ **韵律短语边界**：韵律短语之间的边界，韵律短语边界后面有可感知的停顿，但有些情况下不是很明显。
- ✧ **语调短语和句子边界**：语调短语或者句子之间的边界。由于口语对话中的句子相对较短，而且基本上每个句子只有一个分句，语调短语边界和句子边界基本上一致，所以对于这两者就不加区分了。此类边界后面有明显的停顿。

上面这四种韵律边界之间存在包含关系，也就是说语调短语和句子边界一定是韵律短语边界，韵律短语边界必然是韵律词边界，而韵律词边界又必然是音节边界。

我们采用基于语义的韵律边界标注方法[86] [87]标注韵律边界，标注举例如图 3.4 所示（用 0、1、2、3 分别表示上面的四种韵律边界，图中没有标明音节边界）。

八点 1 左右 2 有哪些 3?
从北京 1 到乌鲁木齐的 2 票价 1 是多少 3?
到北京 2 都有 1 哪些航班 3?
你好 3 请问 2 到上海的 1 航班 2 都有 1 几点的 3?
太早了 3 有没有 1 中午的 3。

图 3.4 韵律边界标注示例

3.4.2.2 韵律边界对应的特征

一般说来，语音识别可以同时给出识别结果和结果对应的音节边界，在已知音节边界的情况下，韵律边界的检测问题可转化为韵律边界分类问题，即对每一个音节边界进一步判断它的边界类型。用于边界分类的韵律特征主要包括以下三个方面：

时长相关特征：时长相关特征主要是指边界处的停顿现象，以及边界前后音节时长的变化。这些特征在韵律短语边界，尤其是语调短语边界表现得比较突出。具体说，时长特征包括候选边界后静音的长度、候选边界之前一个音节前面的静音长度、候选边界前后音节的时长、候选边界前后

音节时长之差。

基频相关特征：基频变化是韵律信息重要内容，也是韵律边界的一个重要标志。一般认为，韵律边界（尤其是韵律短语边界和语调短语边界）都有基频重置（F0 reset）现象。除了基频重置特征，还包括候选边界前后音节的基频均值、基频最小值、基频最大值等。

能量相关特征：能量对于韵律边界分类作用不如前两方面特征那么重要，只用到了候选边界前后音节能量的均值、最大值，以及前后音节能量均值和最大值的差值。

除了上述三类特征，还用到了候选边界在句子中的位置特征，包括候选边界离句首、句尾的距离。

3.4.2.3 韵律边界分类的结果

将基频、能量相关的特征根据说话人进行归一化后，用 CART（Classification And Regression Trees）决策树对韵律边界进行分类。

本实验用到的数据来自航班订票查询领域，包括 12 人以自然语音的方式录制的数据，每人 100 句，共 1200 句，其中 1000 句作为训练集，100 句作为开发集，剩下的 100 句是测试集，测试集的分类结果如表 3.3 所示。

表 3.3 韵律边界分类的结果

分类结果 / 标注类型	**0**	**1**	**2**	**3**	**正确率(%)**
0	583	4	6	0	98.314
1	3	75	86	0	45.732
2	2	0	77	0	97.468
3	1	0	0	121	99.180

从表 3.3 的结果看，整体的分类效果比较理想，但是韵律边界类型 1 和类型 2 混淆度较高。事实上，标注的过程中我们就考虑到这两种边界类型可能会比较难以区分，如句子“八点 1 左右 2 有哪些”，按照我们标注

的原则，“八点左右”后面应该是一个韵律短语边界，“八点左右”省略了中心词“航班”（完整的句子应该是“八点左右的航班有哪些”）做句子的主语，从句法上说应该可以与后面的成分分开。但是由于整个句子比较短，很难从声学上区分“八点”后面的韵律词边界类型 1 与“左右”后面的边界类型 2。

3.4.3 韵律边界信息作为置信度特征

在使用韵律边界类型作为语义概念的置信度特征时，不考虑边界类型 1 和 2 的区别，把这两种边界当成一类。如果语义概念识别正确，那么概念的边界至少应该是一个韵律词边界，不能仅仅是音节边界。因此，我们将语义概念的开始和结束位置的韵律边界是否音节边界作为一种置信度特征。如果概念边界是韵律词、韵律短语或者语调短语边界，那就说明识别结果对应的边界与韵律边界的要求相符；反之，说明识别结果的边界不是十分准确，识别结果的可信程度也就大大降低了。具体说来，用到的特征包括：

- ✧ 语义概念的开始和结束位置的韵律边界类型
- ✧ 语义概念中韵律短语边界的个数
- ✧ 语义概念中语调短语边界的个数

3.5 实验结果与分析

针对语义概念的置信度打分，本章从语义概念分析理解层和韵律层面提出了新的特征，下面通过实验来验证所提特征的性能。

实验中，采用 3.1.2 节介绍的 Fisher 线性分类器作为确认模型。为了训练置信度模型，在第三章的实验数据的基础上，又收集了 900 句话作为置信度确定的训练集，第三章实验中使用的 500 句作为测试集。

本节主要设计了两个实验：

实验一：比较不同置信度特征的确认性能，特征主要分成三类：

1) 声学层置信特征

2) 语义概念分析理解层的置信特征

3) 韵律边界特征

实验二：测试加入语义概念确认后对话系统的理解性能。

3.5.1 实验一：不同置信度特征下的语义概念确认性能

本实验中使用的声学层的置信度特征具体包括以下几种：

- ✧ 词的后验概率得分，它反映待确认候选与模型匹配的平均情况；
- ✧ 词中最小的音素级后验概率得分，该特征反映待确认序列中置信度水平最低的部分；
- ✧ 词图中与词候选处于并列位置的其它词候选个数，一般说来，竞争词的数目越多，候选词的可靠性就越低；
- ✧ 词图中，在并列位置包含当前候选词的路径的条数，候选词在不同路径的并列位置中出现的次数越多，说明它越可靠；
- ✧ 词结束时搜索空间中的活动路径的数目。

不同置信度特征的确认性能如图 3.5 中的 ROC 曲线所示，曲线越靠近坐标轴说明确认性能越好。图中共有 4 条曲线 A、B、C、D：曲线 A 是仅使用声学层置信度特征的结果；曲线 B 仅使用了分析理解层的置信度特征；曲线 C 结合了声学层和分析理解层的特征；而曲线 D 则是综合了声学层、分析理解层和韵律边界信息三类特征后的确认结果。明显地可以看出：曲线 B 比曲线 A 更加靠近坐标轴，说明分析理解层的置信特征其确认性能要优于声学层特征；曲线 D 最接近坐标轴，说明了综合三类置信度特征可以取得更优的语义概念确认性能。

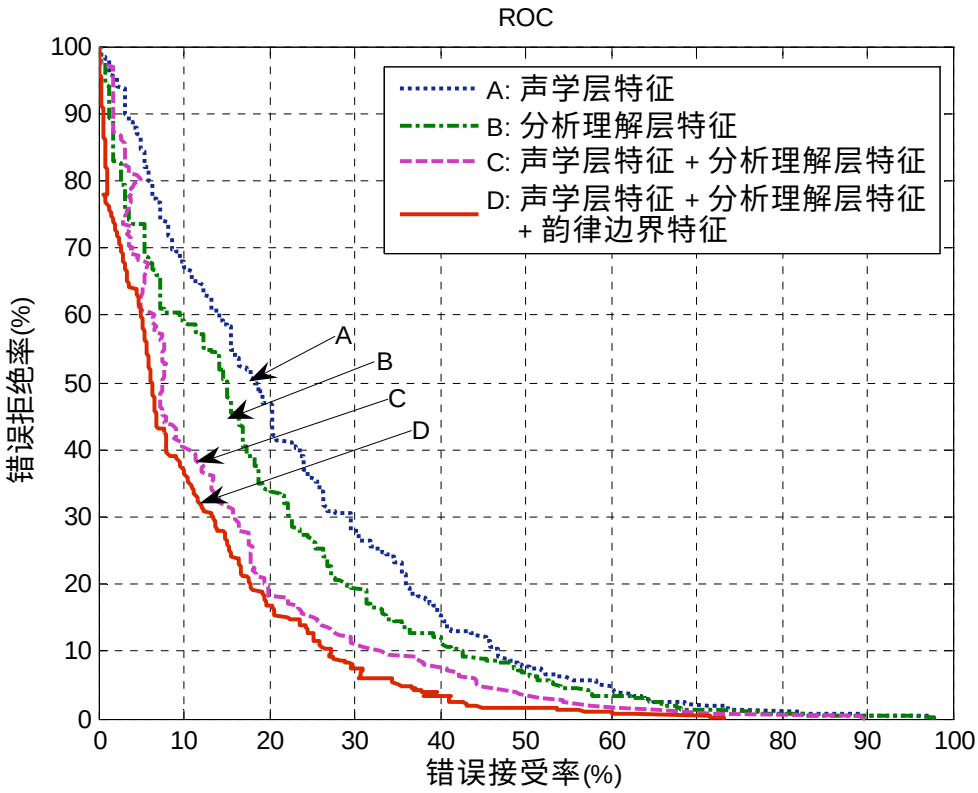

图 3.5 不同置信度特征下的语义概念确认性能

为了进一步的比较，再使用特定错误拒绝率下的错误接受率来评测不同置信度特征的确认性能。一般情况下，在语音确认任务中，对错误拒绝率（FRR）要求较高，FRR 不能太大，否则正确候选的损失过大，会带来很坏的负面影响，因此我们更加关心 FRR 较低的情况下，错误接受率（FAR）的大小，表 3.4 给出 FRR 为 5%和 10%时 FAR 的情况。

表 3.4 错误拒绝率为 10%和 5%时语义概念的错误接受率(%)
（其中 1 表示声学层特征，2 表示分析理解层特征，3 表示韵律边界特征）

置信度特征	错误接受率（错误拒绝率=10%）	错误接受率（错误拒绝率=5%）
(1)	46.8	60.0
(2)	42.7	54.1
(1, 2)	33.8	43.6
(1, 2, 3)	27.2	35.5

从表 3.4 中的结果表明，分析理解层的置信度特征在语义概念确认方面比声学层特征更加有效，而两种特征结合后大大了提高确认性能，如果在这两类特征的基础上在加入基于韵律信息的置信度特征可以进一步提高语义概念的确认能力。

3.5.2 实验二：加入语义概念确认后对话系统的理解性能

表 3.5 加入语义概念置信度确认后的理解性能(%)

	替换错误率	插入错误率	删除错误率	总的槽错误率
无确认	6.83	24.50	6.83	38.16
加入语义概念确认	3.33	8.07	14.65	26.05
错误率下降	—	—	—	31.7

在语音理解模块之后，加入语义概念进行置信度确认能够提高系统的整体理解性能，实验结果如表 3.5 所示。通过调整置信度阈值，保证语义概念错误拒绝率为 5%时得到表 3.5 的结果，从中看出经过语义概念确认后，替换错误和插入错误大大减少，尽管删除错误有所提高，但总的语义槽错误率还是有了较大的改进，错误率下降达到 31.7%。

3.5.3 分析与讨论

本章的主要工作是针对语义概念的特点，提出新的置信度特征——分析理解层的置信度特征和韵律边界特征。实验结果表明：新特征具有较好的语义概念的确认性能；另外，新特征与原有的声学层置信度特征来自不同的知识源，从不同的方面评价待的语义确认概念的可靠程度，因此，它们的作用可以相互叠加，综合使用三类特征后的语义概念的确认性能进一步提高。

3.6 小结

语义概念的置信度确认在口语对话系统中非常重要，它能够减少识别错误对后续对话过程的负面影响，是近年来口语对话系统研究的一个热点。本章在基于语义概念的语音理解框架之上，针对语义概念，从分析理解层面提出新的置信度特征，并将韵律边界信息作为一种新的特征引入语义概念的置信度打分中。实验结果表明分析理解层的置信度特征优于声学层置信度特征；将声学和分析理解层的置信度特征结合后，确认性能要优于单独使用一种特征；在上面两类置信度特征的基础上加入韵律边界信息后，进一步提高了语义概念的确认效果。另外，实验结果还表明，在语音理解之后加入语义概念的确认，可以提高系统的理解性能。

第4章 待登录关键词的发现及其语义类属性标注

待登录关键词是指对话系统词表没有覆盖的，但又确实是该领域对话中常用到的那些关键词。对于对话系统的词表而言，待登录关键词可以认为是“新词”，但相对于一般领域的字典而言，这些词并不新，因为它们大部分都被字典收录。因此，这里所谓的待登录关键词与自然语义理解领域的新词意义有所不同：在自然语义理解领域，新词指那些在字典中都没有收录过，但又确实能称为词的那些词，最典型的代表就是人名。本章中下文其后部分为了表述方便可能会用到 “新词”这一概念，如无特殊说明均指待登录关键词，而不是自然语义理解领域的新词概念。

在很多对话系统中，词表不仅仅起到字典的作用，还定义了词语的语义信息，例如，*EasyFlight* 的词表不仅规定了航班信息领域常用的词语，还将词语按照一定的语义类属性加以组织，这样便于在文法中直接使用语义类作为终结符，编写规则[51]。本章的主要目标是通过发现待登录关键词并自动标注其语义属性，将其添加到词表相应的语义类中，不断完善对话系统的词表，以使现有文法可以分析理解更多的语句，避免因为出现词表以外的词而引起的识别错误和分析失败。

本章的内容安排如下：第一节说明待登录关键词发现及其语义类属性自动标注的研究背景和研究意义；第二节简单介绍相关方面的研究；第三节介绍待登录关键词发现的方法；第四节中重点说明如何确定待登录关键词的语义属性；第五节给出实验结果和分析；最后，第六节总结本章内容。

4.1 研究意义

4.1.1 背景对话系统

本章内容以口语对话系统 *EasyFlight* 为研究背景。*EasyFlight* 系统定义词表时充分考虑了词的语义含义，目前词表规模为 400 左右，初分成 100 左右的语义类，每一个语义类都有对应的语义解释。换句话说，词表中的词一旦被识别出来，就可以根据它所属的语义类了解其语义含义。

表 4.1 给出了一些语义类的定义和属于该类的词语举例。

表 4.1 *EasyFlight* 中词类的定义和举例

语义类名	举例
mat_city_name	北京
mat_airline_abbr	南航
mat_time_of_the_day	上午
tag_to	到
tag_exist_or_not	有没有
tag_how_many	多少
ato_week	礼拜
ato_1to6	六 （表示周日期的数字）

待登录关键词发现及其语义类属性自动标注就是要确定待登录词的语义类属性，以便将它们直接添加到词表中正确的语义类下。

4.1.2 研究意义

待登录关键词发现和语义类属性标注的主要目的可以概括为以下两个方面：

第一，集外词对识别有较大的负面影响，更何况很多待登录关键词是系统理解用户意图时不可忽略的部分，直接影响到能否正确地获取句子语义，因此，将待登录关键词添加到词表中正确的语义类中可以改进对话系统的性能；

第二，对话系统应该通过实际应用中不断完善，发现待登录关键词并确定其语义类属性是其中的一个方面。

4.1.2.1 影响语音理解性能的关键因素

对话系统的领域特点决定了它的词表规模有限，而且在词表设计阶段，很难考虑到所有可能的情况，设计出来的词表难免出现覆盖度不够的问题：这一方面是由于设计人员的经验不足，另一方面也是口语的灵活性和多样性造成的。

于是，在实际应用中，识别器不可避免地会遇到词表中不存在的词语，这些集外词本身不能被系统理解，还会还影响到与它相邻的集内词的识别效果。识别时加入 OOV 模型可以解决集外词对集内词的识别影响，但是要从根本上解决集外词引起的理解错误还需要依靠词表的不断完善，这里所谓的词表完善，不仅是将新词加入词表，还要将其加入正确的语义类中，以保证后续语言理解模块可以正确理解其语义。

在基于语义概念的语义理解框架下，词表的完善还有另外一层更重要的意义，那就是保证语义概念知识能在识别搜索中起到更好的指导作用。例如，“从北京至广州”中的“至”是集外词，这使得“从北京至广州”不符合描述地点概念的有限状态网络，识别搜索到达“至”这个词结尾后，后面可以扩展哪些词，没有任何指导信息；如果将“至”加入词表中“tag_to”类下，那么“从北京至广州”就完全符合地点概念的有限状态网络，该候选路径在所有搜索路径中将处于更加有利的位置。

另外，在对话系统中，待登录关键词的发现有助于文法规则的完善。一些新词在原有的词表中找不到合适的语义类，这些词的发现对于新规则的产生有着积极作用。

4.1.2.3 对话系统不断完善的需要

一般说来，对话系统刚刚设计完成时，性能并不理想，这是因为对话系统在开发设计初期无法模拟真实的运行环境，只能根据经验和有限的领域数据设计词表、文法，声学模型的训练也缺少足够的真实数据。对话系统想要真正达到理想的效果，需要在实际应用的过程中，不断地完善系统、提高性能。图 4.1 是对话系统在实际应用的过程中不断完善的示意图，系统原型（prototype）实际使用后，可以收集很多极有研究价值的真实语料和对话实例，从中可以发现影响系统性能的主要问题。通过分析系统存在的不足，研究相应的解决方案，并利用真实数据对声学模型、语言模型、词表、文法进行更新，用改进后的对话系统替代原来的系统。重复上面的过程，可以逐渐提高对话系统在实际应用中的性能。待登录关键词的发现并标注其语义类属性是改善对话系统性能的一个重要举措。

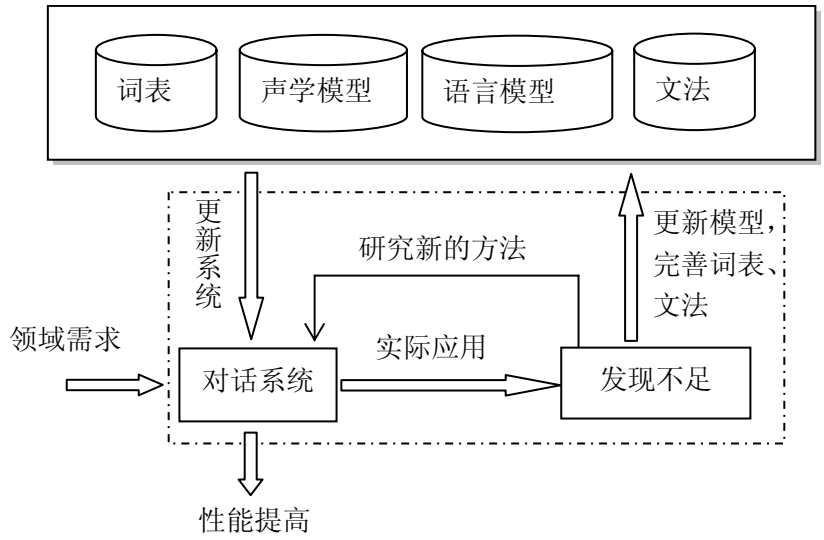

图 4.1 对话系统在应用中不断完善的示意图

4.2 相关研究

在详述本文提出的待登录关键词发现及其语义类属性标注方法之前，先对相关方面的研究作一个简单介绍。

在基于规则的语言理解方法中，规则的设计是非常重要的一步，也是需要花费较多时间和精力的一步。为了更快的设计领域文法，研究人员希望利用数据驱动的方法，从实际语料中自动获取词类、短语的表达方式，甚至全部的文法规则。McCandless 和 Class 提出根据 bigram 分布的相对熵（relative entropy）来合并词类的想法[88]。Helen 等人在这个想法的基础上进一步深入研究，提出根据 bigram 分布的 K-L 距离（Kullback-Liebler Distance）合并词类，并根据词与词之间的互信息（Mutual Information，即 MI）来获取短语表达方式，进而得到文法规则的方法[89][90]。下面，简单介绍这一方法。

自动获取词类和文法规则的过程是一个聚类过程：初始状态时，将词表中每个词单独看作一类，然后，从时间和空间两方面对初始的词类不断进行合并，称为时间聚类（temporal clustering） 和空间聚类（spatial clustering）。合并过程进行一定次数后，可得到描述短语表达和句子表达的规则。

1) Temporal clustering 合并那些总是一起出现的词，如 “show”和“me”被合并成“show me”。Temporal clustering 时考虑两个词（e_1, e_2）之间的互信息，即计算

$$MI(e_1,e_2)=P(e_1,e_2)\log\frac{P(e_2\mid e_1)}{P(e_2)}$$

每次聚类将 MI 分最高的两个词聚在一起，认为这它们经常一起出现。

2) Spatial clustering 会合并那些具有相同上下文的词，例如“Toronto”和“Tampa”（两个城市名），Spatial clustering 时考虑两个词的 bigram 分布之间的 K-L 距离，即

$$Dist(e_1,e_2)=Div(p_1^{left},p_2^{left})+Div(p_1^{right},p_2^{right})$$

其中

$$Div(p_1,p_2)=D(p_1\|p_2)+D(p_2\|p_1),$$

而 $D(p_1\|p_2)$就是两个词 bigram 分布的 K-L 距离[90]，每次聚类时将距离最小的两个词聚在一起，因为它们出现的上下文环境非常相近。

本章提出的待登录关键词的语义类属性标注是为新词在原有的词表中找到相应的语义类，从这一点上说，与 Spatial clustering 的目的有些类似，spatial clustering 也是将具有相同上下文的词合并在一起。但是，对于那些出现次数比较少待登录关键词，Spatial clustering 的效果不太理想，因为 Spatial clustering 完全根据上下文来聚类，而出现次数少的关键词的上下文偶然性比较大。

4.3 待登录关键词的发现

本章提出的待登录关键词发现及其语义类标注方法应用于实际语料的标注文本之上。识别器对自然语音的识别性能本来就不够理想，再加上集外词的干扰，含有集外词的语句识别效果普遍较差，因此，我们无法直接使用识别器输出的文本作为待登录关键词发现的依据，而是在标注文本的基础上发现新词，并确定其语义类属性。但是，为了减少标注的工作量，可以根据句子的置信度打分，有选择性标注数据，只对那些整句置信度得分较低的句子进行标注，从中发现待登录关键词。

待登录关键词的发现主要依靠现有文法对语句的分析结果，并借助通用汉语词法分析系统 ICTCLAS[91]给出的词法分析结果。*EasyFlight* 中采用改进的自底向上的分析算法[51]，能够输出部分分析结果，没有输出的部分对于系统来说就是待登录关键词或者它们的组合。文法分析给出的待登录关键词是否准确还要进一步结合通用词法分析器的分词结果。

下面举例说明待登录关键词发现的过程，如图 4.2 所示。分析器对句子“五月一号飞往北海的有没有”的分析结果表明“往”是一个新词，但是词法分析系统的分词结果认为“飞往”是一个词，此时，我们认定这句话中的待登录关键词是“飞往”而不是“往”。

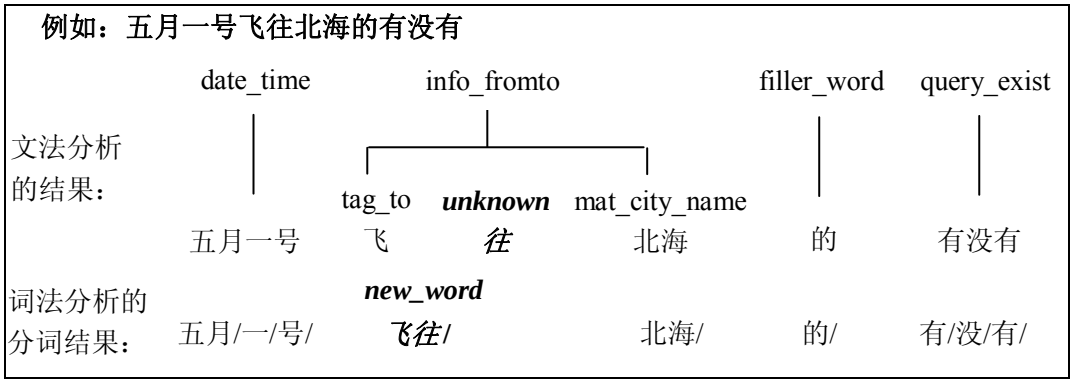

图 4.2 待登录关键词发现举例

4.4 标注待登录关键词的语义类属性

确定待登录关键词的语义属性包括三个步骤：首先，根据待登录关键词在句子中的上下文关系，推测新词最有可能的词类属性；然后，根据对推测得到的词类属性进行可靠性评价；最后将语义属性足够可靠的待登录关键词添加到词表中。

4.4.1 推测待登录关键词的语义类属性

推测待登录关键词的语义类属性的前提是词表中存在与新词同属于一个语义的词。如果新词在原有的词表中找不到一个与它意义相近的词，那就无从知道新词的语义含义。

被语义文法成功分析的语句是标注待登录关键词语义属性的基础。这里所谓的“成功分析”是指句子中所有的成分都可以归结成语义概念，不存在无法归结的部分。对于那些被成功分析的语句，保存它们的分析结果，得到一个知识库，利用这个知识库可以推测待登录关键词的语义类属性。知识库中的内容如图 4.3 所示，包括两个层次的内容：一是句子级的模板，即语义概念组成的序列，句子模板前的 n_i 表示该句子模板在知识库中出现的次数；二是语义概念级的模板，即词表中的语义类组成的序列，概念模板前的 n_i 表示当前的语义概

念模板在知识库中出现的次数。

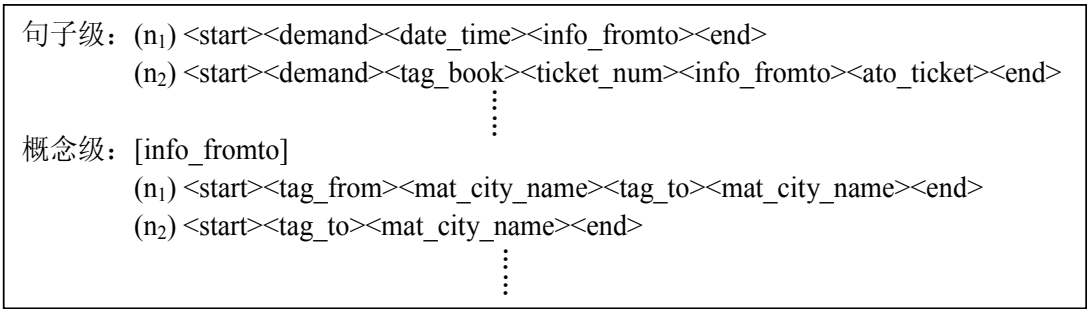

图 4.3 知识库中的内容

对含有待登录关键词的语句进行语义分析，可以得到待登录关键词所处的上下文关系，如下：

<$Concept_{pre}$><Unknown><$Concept_{next}$>

Unknown 表示待登录关键词，$Concept_{pre}$ 和 $Concept_{next}$ 是待登录关键词前后的两个语义概念。这里所谓的语义概念在第三章 3.3.1 节定义的基础上，又增加了两种，分别是句子开始<start>和句子结尾<end>。根据新词所处的语义概念级上下文关系推测新词的语义类属性，分为两种情况：

(1) 如果 $Concept_{pre}$ 与 $Concept_{next}$ 表示同一个概念 A，则新词是描述概念 A 用到的语义类；

(2) 如果 $Concept_{pre}$ 和 $Concept_{next}$ 是两个不同的概念，那么新词可能是描述 $Concept_{pre}$ 用到的语义类，也可能是描述 $Concept_{next}$ 用到语义类，还可能对应一个独立的语义概念。

确定待登录关键词所属的语义概念后，将含该词的概念分析结果与知识库中的概念级模板进行匹配，可以推测待登录关键词的候选语义类属性。

下面用简单的例子说明推测语义类属性的过程。在句子“……星期六自西安回北京……”中，“自”是一个待登录关键词，对这句话的语义分析结果如图 4.4 所示。词“自”是处于概念“date_time”和“info_fromto”之间，从下面三种情况考虑这个新词的语义类属性：

1) 新词本身就对应一个独立的概念 A，它是描述概念 A 用到的某个语义类；

2) 新词是描述概念“date_time”用到某个语义类；

3) 新词是描述概念“info_fromto”时用到的某个语义类。

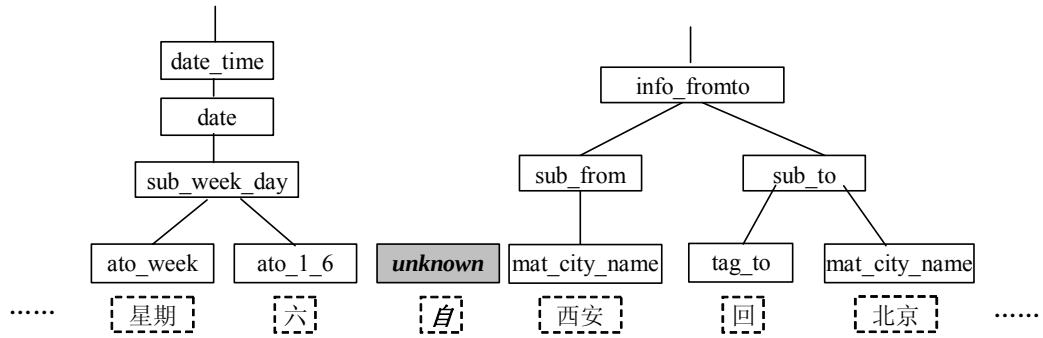

图 4.4 含有新词的语句的分析结果

对于第一种情况，先根据公式（4-1）推测待登录关键词所属的概念

$$Concept_{new\_word} = \arg\max_{Concept} p(Concept \mid Concept_{pre}, Concept_{next}) \qquad (4\text{-}1)$$

其中 $p(Concept \mid Concept_{pre}, Concept_{next})$ 是通过知识库中句子级模板得到的，如果（$Concept_{pre}$，Concept，$Concept_{next}$）（其中 Concept 可以除了概念“date_time”和“info_fromto”以外的任意语义概念[1]）这样的三元组在句子模板中不存在，那么新词不属于任何概念，从而推测它是一个垃圾词。

确定待登录关键词所属的语义概念后，将含有该词的概念分析结果与知识库中的语义概念级模板进行匹配，可以得知新词对应语义类属性。例如，在第三种情况下，新词“自”是描述概念“info_fromto”用到的语义类，含有新词的部分分析结果与知识库中“info_fromto”的概念模板匹配，最佳匹配结果如下：

<start><***unknown***><mat_city_name><tag_to><mat_city_name><end>
<start><***tag_from***><mat_city_name><tag_to><mat_city_name><end>

根据最佳匹配即可推测出新词对应的语义类属性，例如上例中“自”对应语义类<tag_from>。如果存在多个最佳匹配，对一个新词可能会推测出多个语义类属性，此时根据下面的公式选取可能性最大的语义类

$$C_{new\_word} = \arg\max_{C} p(C \mid C_{pre}, C_{next}) \qquad (4\text{-}2)$$

[1]在航班信息领域，概念“date_time”和“info_fromto”出现的频率比其他概率要高很多，并且它们几乎可以出现在句子的任何位置，因此，如果三元组($Concept_{pre}$, Concept, $Concept_{next}$)中 Concept 考虑这两个概念，那么公式(5-1)在大多数情况下会得到“date_time”或者“info_fromto”，而真正的正确结果无法在竞争中胜过这两个概念。当然，这样做会带来另一个问题：如果新词是 “date_time”或者“info_fromto”概念用到的语义类，并且不需要结合其他词，自己本身就是一个完整的语义，那么可能会无法正确推测这些词的语义类属性。庆幸的是，这类词比较容易总结，原有词表中已经相当完备了。

式中 C_{new_word} 表示推测出的语义类属性，C_{pre} 表示新词前面一个词的词类属性，C_{next} 表示新词后面那个词的词类属性，$p(C \mid C_{pre}, C_{next})$ 从前文所说的知识库中（考虑概念级的模板）获取，计算方法如公式（4-3）所示，N 表示词表中语义词类的数目，$\#(C_{pre}, C, C_{next})$ 表示“C_{pre} C C_{next}”出现的次数。

$$p(C \mid C_{pre}, C_{next}) = \frac{\#(C_{pre}, C, C_{next})}{\sum_{i=1}^{N} \#(C_{pre}, C_i, C_{next})} \quad (4\text{-}3)$$

在上面的三种情况下，都可以为待登录关键词推测出一个候选语义类属性，下一节的语义类属性确认步骤中将去掉那些不可靠的候选属性。

4.4.2 待登录关键词的语义类属性的确认

上一小节中根据待登录关键词所处的上下文位置推测出它们可能的语义类属性，本节对推测的结果进行可靠性评价，从中选择置信度分数足够高的结果。这一步非常重要，因为只有将待登录关键词添加到正确的语义类中，才能起到提高系统性能的作用，否则反而会给系统带来负面的影响。

受到语音确认的启发，我们可以通过对推测结果的确认来保证其正确性。确认时，主要考虑两个方面的**特征**：

1) 推测结果的概率分：公式（4-2）中的 $p(C \mid C_{pre}, C_{next})$ 表示在上下文是 C_{pre} 和 C_{next} 时，待登录关键词属于词类 C 的概率，这个概率分可以用来评价推测结果的可靠程度，分数越大，推测结果就越可靠。实际中，我们把这个概率分的阈值设为 0.5，得分小于这一阈值的推测结果认为是不可靠的。

2) 待登录关键词的词性标注是否与它所属的语义类中其他词的词性标注一致。用汉语词法分析工具 ICTCLAS 对原有词表进行词性标注，由此可知每个语义类的词性分布，例如在词类“mat_city_name”中，所有的词都是名词，如果将某个词性是动词的待登录词推测为“mat_city_name”类，那么这个结果就是非常不可靠的。

上述两个特征都符合要求的情况下，待登录关键词的候选语义类属性才能被保留，如果所有的候选语义类属性都不能通过确认，那么这个待登录关键词的词类属性为未知（Unknown）。

4.4.3 在词表中添加待登录关键词

当某个待登录关键词在语料文本中出现的次数大于等于 N 时，考虑将该词加入词表中。待登录关键词每出现一次，都可以根据其出现的上下文关系推测出它的候选语义类属性，相同的词出现 N 次出现后，通过投票的方式，选出票数最多的候选语义类属性作为新词的语义类属性。后面的实验中，N 取值为 3。

待登录关键词发现及其语义类属性自动标注的整个过程如图 4.5 所示。

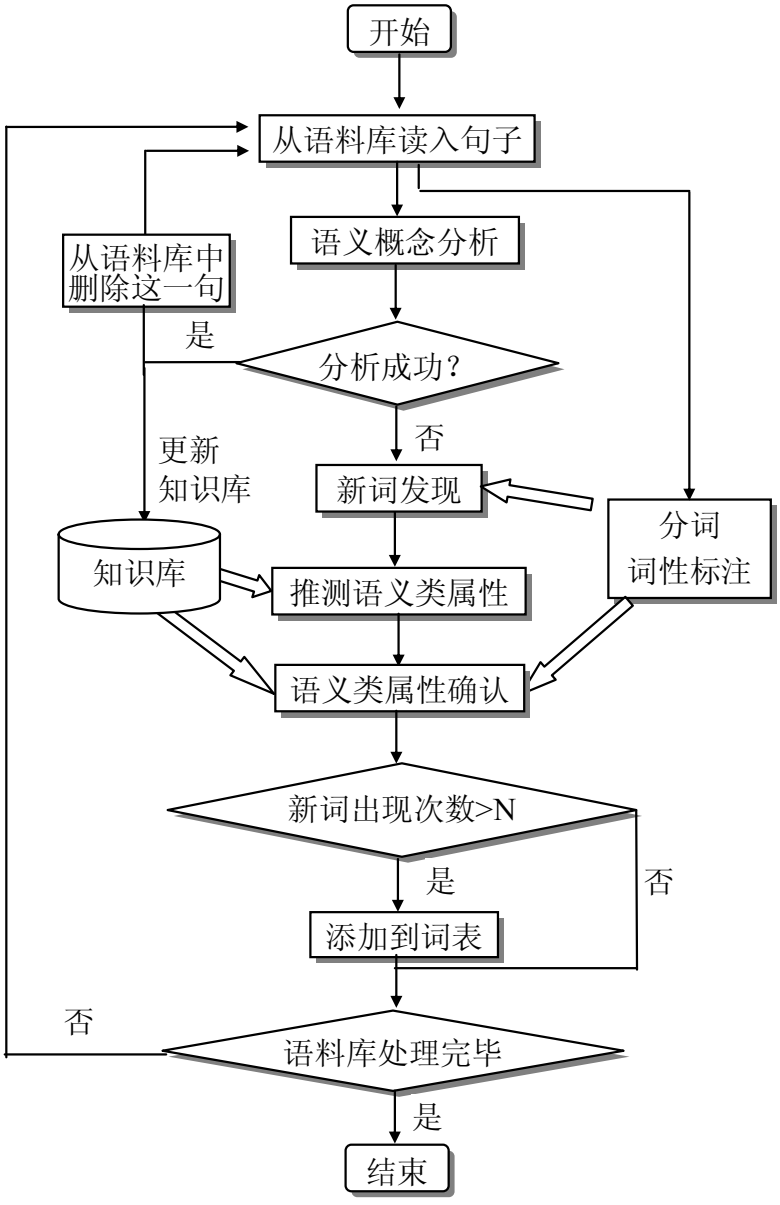

图 4.5 待登录关键词的发现及其语义类属性标注的过程

4.5 实验结果与分析

4.5.1 实验结果

实验数据：

EasyFlight 原型系统完成之后，为了进一步测试系统性能，我们又收集了一部分领域内的人人对话语料，并从中选取 900 个用户的语句，这部分数据是待登录关键词发现及其语义类标注的实验数据。

分析这批数据发现，900 个句子中有 42.5%的句子都含有待登录关键词，待登录关键词共有 109 个，其中包含地点名词 38 个（城市名、省名或者国家名）。对于地点名，对照地名字典可以比较容易地确定其语义类属性，这些词的语义类属性自动标注结果可以认为是百分之百正确，所以在评价待登录关键词语义类属性的自动标注结果时，把地点名词排除在外。这样，测试数据中的待登录关键词的数目有 71 个。

评价标准和实验结果：

对于系统现有的词表来说，待登录关键词可以分为两种情况：

1) 可以被直接划分到词表中已有的语义类中；

2) 不属于当前词表中的任何语义类。

81.7%的待登录关键词属于情况 1，评价这些词的语义类属性的标注结果是否正确直接看它是否被划分到正确的词类中即可；而对于情况 2，如果待登录关键词被划分到未知类，就可以认为该词的语义类属性标注正确，否则认为标注错误。

经过语义类属性的自动标注，91.5%的待登录关键词可以正确地添加到词表中相应的语义类中。在词表完善前后，分别对 900 个句子进行识别测试，以此考察待登录关键词发现和自动标注策略对系统性能的影响。结果表明：词表更新前，音节识别的正确率只有 65.9%，而通过发现和标注待登录关键词自动更新词表之后，音节识别的正确率提高到 74.8%，这说明待登录关键词的发现及其语义类属性的自动标注有利于对话系统识别性能的提高。

4.5.2 分析与讨论

尽管待登录关键词语义类属性标注的正确率达到 91.5%，我们还是担心那些

错误添加的关键词会不会给识别带来较大的负面影响。通过分析结果，发现出错的 6 个待登录关键词都可以被原词表的语义类覆盖，但是标注过程没有正确地得到其对应语义类，其中有 2 个词划分到垃圾词类中，另有 3 个词划分到未知类，只有 1 个词错分到其它语义类中。划分到垃圾词类和未知词类的新词不会出现在描述语义概念的文法规则中，不会干扰语义概念对识别搜索的指导作用，因此不会对识别性能带来负面影响，也不会对语义理解造成什么负面影响，只是无法理解这些新词的语义而已。

对于那些当前词表无法覆盖待登录的关键词，本章提出方法将它们直接划分到未知类。事实上，这类词中很大一部分对于正确理解句子语义还是非常重要的，它们也是有语义的，例如“每天”、“回程”、“别的”，这些词的发现提示我们要在词表中添加新的语义类，并在文法中添加相应的规则。

4.6 小结

本章针对口语对话系统设计初期关键词表不够完善的问题，提出待登录关键词的发现及其语义类属性自动标注的方法。对含有待登录关键词的句子进行语义分析，得到部分分析的结果，结合句子的词法分析可以确定待登录关键词，随后根据待登录关键词出现的上下文关系推测该词在词表中可能对应的语义类，经过语义类属性的确认后得到待登录关键词的语义类属性，并将其添加到词表。待登录关键词发现和自动标注的策略让系统具有一定的自学能力，使系统可以经过实际测试和应用不断完善词表；更重要的是，词表的完善使得原有规则可以覆盖更多的语义概念表达方式，从而提高语音理解的性能。

第5章 总结与展望

5.1 论文工作总结

本文针对口语对话系统中语音识别任务的若干难点，以语音理解作为研究对象，在基于语义概念的语音理解框架、语义概念的置信度确认及待登录关键词的发现和自动标注方面进行了初步的探索和研究，提出了一些新方法、新策略，并通过实验证明了其有效性，同时也为对话系统中语音理解领域的深入研究奠定了一定的基础。

概括来说，本文的工作重点和贡献主要体现在如下几个方面：

(1) 提出基于语义概念的语音理解框架

针对口语对话系统中语音识别性能不佳的问题，提出了以语义概念为核心的语音理解框架。新的语音理解框架下，上层语义概念知识机及早地引入到识别搜索中，而识别和理解两个步骤也更加紧密地结合在一起，成为合一的过程。该框架利用规则描述语义概念知识，避开领域数据收集和标注困难的问题；另外，它具有良好的可扩展性，可以很方便地在识别中引入对话上下文知识，进一步提高识别性能。实验结果表明：在基于语义概念的语音理解框架下，对话系统前端识别的性能大大提高，语义概念内部的音节识别正确率达到 90.11%；语义概念识别和理解和一体化过程提高了理解的鲁棒性，使得最终的语义槽正确率达到 86.33%。

(2) 提出分析理解层的置信特征和韵律边界特征用于语义概念确认

语义概念是整个对话系统中核心内容，为了减少语义概念错误对系统后续模块（主要是对话管理和应答生成模块）的负面影响，本文主要从特征方面着手，研究语义概念的置信度确认方法。在语音理解框架下，针对语义概念从分析理解层面提出新的置信度特征，并将韵律边界信息作为一种新的特征引入语义概念的置信度打分中。实验结果表明：分析理解层的置信度特征优于声学层置信度特征；将声学层和分析理解层的置信度特征结合后，确认性能要优于单独使用一种特征；在声学特征和分析理解层特征的基础上再加入韵律边界信息后，语义概念的确认效果得到进一步提高。

(3) 提出待登录关键词的发现及其语义类属性的自动标注

考虑到对话系统设计初期关键词表不够完善的问题，提出发现待登录关键词并自动标注其语义类的策略。对含有待登录关键词的句子进行语义分析，得到部分分析的结果，并结合句子的词法分析确定待登录关键词，随后根据待登录关键词出现的上下文关系推测该词在词表中可能对应的语义类，经过语义类属性的确认最终确定待登录关键词的语义类属性，并将其添加到词表。待登录关键词发现和自动标注让系统具有一定的自学能力，使系统可以经过实际测试及应用不断完善词表；更重要的是，词表的完善使得原有规则可以覆盖更多的语义概念表达方式，从而提高了系统的语音理解性能。

5.2 下一步研究的展望

本文虽然在对话系统的语音理解方面进行了一些初步研究，提出了一些新方法和新思路，取得了一定的成果，但同时也发现了一些不足之处。下面将指出这些不足点，以及计划进一步深入开展研究的若干方向。

(1) 统计方法在语音理解框架中的应用

在本文提出的基于语义概念的语音理解框架中，语义概念知识是通过规则的方式引入识别过程并指导搜索的。事实上，统计的方法应该可以跟规则方法相结合，进一步提高语音理解的性能。对话系统设计人员在深入了解领域特点的基础上，总结出描述语义概念的文法。尽可能地覆盖所有的表达方式是设计人员在文法设计时的目标之一，因此，描述某个语义概念的规则可能有很多，而每条规则的重要程度不尽相同。在实际数据中，往往只有某些文法规则是概念表达中常用的，而另外一些规则用到的可能性要小很多，规则的重要程度对于指导识别搜索也是十分有用的。遗憾的是，传统的规则方法无法刻画这种规律，统计方法描述平均规律的特征应该能够弥补规则方法的这一不足。针对实际数据的特点，如何将统计方法应用到现有的语音理解框架中是进一步提高语音理解性能的一条重要途经。

(2) 置信度得分在识别过程中的应用

本文研究的语义概念确认将识别和确认分成两个步骤，即先识别后确认，这种后确认的方法只能拒绝识别错误，无法提高识别率。如果能将语音确认技术融入识别过程，以置信度得分为依据进行剪枝，及早地将错误候选减掉，可以避免错误候选对原本应正确识别的候选所带来的负面影响，保障原本受到干扰的正确候选最终可以被识别出来，从而起到提高识别率的效果。如何将语音确认与语音识别过程更好的结合起来是今后语音识别的研究方向。

(3) 语义规则的自动学习

对话系统中的语言理解以基于规则的方法为主，要设计出一个较好的文法往往要花费较多的时间和精力，即使这样专家设计的文法也很难覆盖领域内所有的表达方式。从语料库中自动或者半自动的学习领域规则能够从实际的数据出发，经过对语料的处理逐步发现规律，这种方法得到文法可以弥补人工设计文法的不足，甚至替代人工设计的文法，加速对话系统的设计过程。本文提出了待登录关键词发现并对其语义类属性进行标注的方法，如何将这一方法扩展到规则的（半）自动学习是下一步研究的方向。

参考文献

[1] Huang,X.D., Lee,K.F., Hon,H.W., and Huang,M. Improved acoustic modeling for the SPHINX speech recognition system. In: Proceedings of ICASSP91, Toronto, Canada, 1991. 345-348

[2] Huang, X.D., Alleva, F., Hon, H-W., at el. The SPHINX-II speech recognition system: an overview. Computer Speech and Language. 1993(2):137-148

[3] 杨行峻, 迟惠生. 语音信号数字处理. 北京, 电子工业出版社. 1995

[4] Goldschen A, Loehr D. The role of the DARPA communicator architecture as a human computer interface for distributed simulations. Technical report, MITRE Corporation, 1999

[5] Os D E, Boves L, Lamel L, et al. Overview of the ARISE project. In: Proceedings of the 6th European Conference on Speech Communication and Technology. Budapest, Hungary, 1999. 1527-1530

[6] Failenschmid K, Thornton J H S. End-user driven dialogue system design: The REWARD experience. In: Proceedings of the 5th International Conference on Spoken Language Processing. Sydney, Australia, 1998. 37-40

[7] Wahlster W. VERBMOBIL: Translation of face-to-face dialogs. In: Proceedings of the 3rd European Conference on Speech Communication and Technology. Berlin, Germany, 1993. 29-38

[8] Goddeau D., Brill E., Glass J., et al. Galaxy: A human-language interface to on-line travel information. In Proc. ICSLP. Yokohama, Japan, 1994. 701-710

[9] Zue V., Phillips M, Seneff S., The MIT summit speech recognition system: A progress report. In Proceedings of DARPA Speech and Natural Language Workshop. Philadelphia, Pennsylvania, 1989. 179-189

[10] Seneff S. TINA: A natural language system for spoken language applications. Computational Linguistics. 1992. 18:61–86

[11] Seneff S., Hurley E., Lau R., et al. GALAXY-II: A reference architecture for conversational system development. In: Proceedings of ICSLP'98. Sydney, Australia, 1998. 931-934

[12] Glass J., Hazen T. Telephone-based conversational speech recognition in the Jupiter domain. In: Proceedings of ICSLP'98 , Sydney, 1998.

[13] Zue V., James R., Glass J., et al. JUPITER: A Telephone-Based Conversational Interface for Weather Information. IEEE Transactions on Speech and Audio Processing, 2000. 8(1):100-112

[14] Rudnicky A., Bennett C., Black A., at el. Task and domain specific modelling in the Carnegie Mellon Communicator system. In: Proceedings of ICSLP2000. Beijing, China, 2000. 2:130-134

[15] Ward W. Understanding spontaneous speech: the Phoenix System. In: Proceeding of ICASSP91, Toronto, Canada, 1991. 365-367

[16] Xu W., Rudnicky A. Task-based dialog management using an agenda. In: Proceedings of ANLP/NAACL Workshop on Conversational Systems. 2000. 42-47

[17] Simpson A., Fraser N., Blackbox and glass box evaluation of the SUNDIAL system. In: Proceedings of Eurospeech93. Berlin, Germany, 1993. 2:1423--1426

[18] Huang C, Xu P, Zhang X et al. LODESTAR: A mandarin spoken dialogue system for travel information retrieval. In: Proceedings of the 6th European Conference on Speech Communication and Technology. Budapest, Hungary, 1999. 1159-1162

[19] 黄寅飞, 郑方, 燕鹏举, 等. 校园导航系统 EasyNav 的设计与实现. 中文信息学报, 2001. 15(4): 35-40

[20] Lussier F E, Morgan N. Effect of speaking rate and word frequency on pronunciations in convertional speech. Speech Communication, 1999. 29: 137-158

[21] Decker A M, Lamel L. Pronunciation variants across system configuration, language and speaking style. Speech Communication, 1999. 29: 83-98

[22] Greenberg S. Speaking in shorthand – a syllable-centric perspective for understanding pronunciation variation. Speech Communication, 1999. 29: 159-176

[23] 林焘, 王理嘉. 语音学教程. 北京: 北京大学出版社. 1992

[24] 宋战江. 汉语自然语音识别中发音建模的研究: [博士学位论文], 北京: 清华大学, 2001

[25] Zheng F, Song Z J, Fung P, et al. Mandarin pronunciation modeling based on CASS corpus. Sino-French Symposium on Speech and Language Processing, Beijing, China, 2000. 47-53

[26] Spilker J., Weber H., Gorz G. Detection and correction of speech repairs in word lattices. In: Proceedings of EuroSpeech'99, Budapest, Hungary, 1999. 2031-2034

[27] Hakkani-Tur D., Tur G., Stolcke A., et al. Combining Words and Prosody for Information Extraction from Speech. In: Proceedings of EuroSpeech'99, Budapest, Hungary, 1999. 1991-1994

[28] Liu Y., Shriberg E., Stolcke A. Automatic disfluency identification in conversational speech using multiple knowledge sources. In: Proceedings of Eurospeech2003. Geneva, Switzerland, 2003. 957-960

[29] Gales M.J.F., Young S.J. Robust continuous speech recognition using parallel model combination. IEEE Transactions on Speech and Audio Processing, 1996, 4(5): 352–359

[30] Renevey P., Drygajlo A. Statistical estimation of unreliable features for robust speech recognition. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP), Istambul, Turkey, 2000, 1731–1734

[31] Hazen T., Seneff S and Polifroni J. Recognition confidence scoring and its use in speech understanding systems. Computer Speech and Language, 2002. 16: 49–67

[32] Bazzi I., Glass J., Learning units for domain-independent out-of-vocabulary word modeling. In: Proceedings of Eurospeech. Aalborg, Denmark, 2001. 61–64

[33] Theresa K. Burianek. Building a Speech Understanding System Using Word Spotting Techniques. MIT Master Thesis, 1999

[34] Ocelikova J., and Matosek V. Processing of Anaphoric and Elliptic Sentences in a Spoken Dialog System. In: Proceedings of EuroSpeech'99, Budapest, Hungary, 1999. 1407-1410

[35] 黄寅飞, 口语对话系统 EasyNav 的研究与实现: [博士学位论文], 北京: 清华大学, 2002

[36] Zanten G V. User Modelling in Adaptive Dialogue Management. In: Proceedings of the 6th European Conference on Speech Communication and Technology (EuroSpeech), Budapest, Hungary, 1999. 1183-1186

[37] Papineni K A, Roukos S, Ward R T. Free-flow Dialog Management Using Forms. In: Proceedings of the 6th European Conference on Speech Communication and Technology (EuroSpeech). Budapest, Hungary, 1999. 1411-1414

[38] 邬晓钧, 对话管理和可定制对话系统框架的研究: [博士学位论文], 北京: 清华大学, 2003

[39] 郑方, 牟晓隆, 徐明星, 武健, 宋战江, (Zheng99) "汉语语音听写机技术的研究与实现", 软件学报, 1999, 10(4): 436-444

[40] Brown P.F., Pietra V.J.D., Souza P.V., et al. Class-based N-gram models of natural language. Computational Linguistic. 1992. 18(4): 467–479

[41] Gustafson J, Lindberg N, Lundeberg M. The August Spoken Dialogue System. In: Proceedings of the 6th European Conference on Speech Communication and Technology. September, Budapest, Hungary, 1999. 1151-1154

[42] 郑方, 连续无限制语音流中关键词识别方法研究: [博士学位论文], 北京: 清华大学, 1997

[43] Renals S, Morgan N, Bourlard H M, et al. Connectionist probability estimators in HMM speech recognition. IEEE Trans on Speech and Audio Processing. 1994, 2(1): 161-174

[44] Rose R, Paul D. A hidden Markov model based keyword recognition system. In: Proceedings International Conference on Acoustics, Speech and Signal Processing (ICASSP), Albuquerque, USA, 1990. 129-132

[45] Bourlard H, Hoore B D, Boite J M. Optimizing Recognition and Rejection Performance in Word- Spotting System. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP). Adelaide, Australia, 1994. 373-376

[46] 陆正中. 口语对话系统中的语音识别研究: [硕士学位论文], 北京: 清华大学, 2002

[47] Guo Q., Yan Y.H., Lin Z.W., Yuan B.S., Zhao Q.W., Liu J. Keyword Spotting in Auto-Attendant System. In: Proceedings of ISCSLP2000, Beijing, 2000. 223-225

[48] Zhang G.L., Sun H, Zheng F., et al. Robust Speech Recognition Directed by Extended Template Matching in Dialogue System. In: Proceedings of the 5th World Congress on Intelligent Control and Automation. Hangzhou, China, 2004. 4207-4210

[49] [美]诺姆 乔姆斯基. 句法结构. 刑公畹等据 1957 年本译, 中国社会科学院出版社, 1979 年版

[50] Allen J. Natural Language Understanding. 2nd edition, Bennjammin/Cummings Publishing Company, Redwood City, California, 1995

[51] 燕鹏举, 对话系统中的自然语言理解研究: [博士学位论文], 北京: 清华大学, 2002

[52] Pieraccini, R., Tzoukermann, E., Gorelov, Z., Progress report on the CHRONUS system: ATIS benchmark results. In: Proceedings of the DARPA Speech and Natural Language Workshop. Morgan Kaufman, Harriman, NY, 1992. 67-71

[53] Magerman, David M. Statistical Decision-Tree Models for Parsing. In: Proceedings of the ACL Conference, 1995. 276-283

[54] He Y., Young S. Semantic processing using the Hidden Vector State model. Computer Speech and Language. 2005. 19: 85-106

[55] Meng H., Busayapongchai S., Glass J., et al. WHEELS: A conversational system in the automobile classifieds domain. In: Proceedings of ICSLP. 1996. 542-545

[56] Wang C., Glass J., Meng H., et al. YINHE: A Mandarin Chinese Version of the Galaxy System. In: Proceedings of the European Conference on Speech. Communication and Technology. Rhodes, Greece, 1997. 351-354

[57] Michael T. Johnson, Mary P. Harper, and Leah H. Jamieson. Interfacing Acoustic Models with Natural Language Processing Systems. In: Proceedings of the International Conference on Spoken Language Recognition, Sydney, Australia, 1998

[58] 武健. 汉语语音识别中统计语言模型的构建及其应用: [硕士学位论文], 北京: 清华大学. 2000

[59] Pieraccini, R. and Levin, E. Stochastic representation of semantic structure for speech understanding. Speech Communication, vol. 11, 1992, pp. 238-288

[60] Mehryar Mohri. Weighted Grammar Tools: the GRM Library. In Jean claude Junqua and Gertjan van Noord, editors, Robustness in Language and Speech Technology. Kluwer Academic Publishers, The Netherlands, 2001. 165-186

[61] Hopcroft, J. E. & Ullman, J. D. Introduction to Automata Theory, Languages, and Computation. Addison Wesley, Reading, MA, 1979

[62] Feng M W, Mazor B. Continuous word spotting for applications in telecommunications. In: Proceedings 2nd International Conference on Spoken Language Processing (ICSLP), 1992. 21-24

[63] S. Kamppari and T.J. Hazen. Word and Phone Level Acoustic Confidence Scoring. In: Proceedings of ICASSP2000, Istanbul, Turkey, 2000. 1799-1802

[64] S. Young. Detecting misrecognitions and out-of-vocabulary words. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP). Adelaide, Australia, 1994. II: 21-24.

[65] Williams G, Renals S. Confidence measures from local posterior probability estimates. Computer Speech and Language, 1999, 13: 395-411

[66] Cox S, Rose R. Confidence Measures for the Switchboard Database. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1996. 511-514

[67] F. Wessel, K. Macherey, H. Ney. Using posterior word probabilities for improved speech recognition. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP), Istambul, Turkey, 2000. 1587-1590

[68] Schaaf T, Kemp T. Confidence measures for spontaneous speech recognition. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP). Munich, Germany, 1997. 875-878

[69] K. Hacioglu and W. Ward. A Concept Graph Based Confidence Measure. In: Proceedings of ICASSP, Orlando, Florida, May 2002, pp. 225–228

[70] Uhrik C., Ward W. Confidence metrics based on n-gram language model backoff behaviors. In: Proceedings of the 5th European Conference on Speech Communication and Technology (EuroSpeech). Rhodes, Greece, 1997. 2771-2774

[71] Rub´en, S., Pellom, B., Hacioglu, K.,et al. Confidence measures for spoken dialogue systems. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, 2001, 1: 393–396

[72] Sameer S. Pradhan and Wayne H. Ward. Estimating semantic confidence for spoken dialogue systems. In: Proceedings of ICASSP2002, Orlando, Florida, 2002. 233–236.

[73] Ma C X, Randolph M A, Drish J. A Support Vector Machines-Based Rejection Technique for Speech Recognition. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP), Salt Lake City, USA, 2001.

[74] 边肇祺, 张学工等. 模式识别. 北京: 清华大学出版社, 2000

[75] Rohlicek, J.R., Russel, W., Roukos, S., Gish, H. Continuous hidden Markov modeling for speaker-independent word spotting. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP). Glasgow, UK, 1989, 627-630

[76] Siu M, Gish H. Evaluation of word confidence for speech recognition systems. Computer Speech and Language, 1999, 13: 299-319

[77] Paul C, Chun J, Daniel W, et al. Is this Conversation on Track? In: Proceedings of the 7th European Conference on Speech Communication and Technology (EuroSpeech). Alborg, Denmark, 2001. 2121-2124

[78] C. Pao, P. Schmid, and J. Glass. Confidence Scoring for Speech Understanding Systems. In: Proceedings of ICSLP 98, Sydney, Australia, 1998

[79] D. Guillevic, S. Gandrabur, and Y. Normandin. Robust Semantic Confidence Scoring. In: Proceedings of ICSLP2002, Denver, Colorado, 2002. 853–856

[80] Shattuck-Hufnagel S, Turk A E. A prosody tutorial for investigators of auditory sentence processing. Journal of Psycholinguistic Research, 1996, 25(2): 193-247

[81] Hiroya Fujisaki. Prosody, Models, and Spontaneous Speech. Computing Prosody: Computational Models for Processing Spontaneous Speech. Springer, 1997

[82] Aylett M P. Stochastic suprasegmentals: relationships between redundancy, prosodic structure and care of articulation in spontaneous speech. [Ph. D thesis]. University of Edinburgh, 2000

[83] H. Niemann, E. Noth, A. Kießling, R. Kompe, A. Bathliner. Prosodic Processing and Its Use in VERBMOBIL. In: Proceedings of Icassp'97, 1997. 1: 75-58

[84] Hui SUN, Mingxing XU, Wenhu WU. Study on Detection of Prosodic Phrase Boundaries in Spontaneous Speech. International Symposium on Chinese Spoken Language Processing (ISCSLP'2002). Taibei, 2002. 281-284

[85] 王蓓, 杨玉芳, 吕士楠. 汉语韵律层级边界结构的声学相关物. 第五届全国现代语音学学术会议论文集, 2001. 161-165

[86] A. Batliner, R. Kompe, A. Kießling, H. Niemann, E. Noth. Syntactic Prosodic Labeling of Large Spontaneous Speech Databases. In: Proceedings of Icslp'96, 1996

[87] Price, P., M. Ostendorf, S. Shattuck-Hufnagel and C. Fong. The use of prosody in syntactic disambiguation. Journal of the Acoustic Society of American, 1991. 90: 2956-2970

[88] McCandless M. and J. Glass. Empirical Acquisition of. Word and Phrases Classes in the ATIS Domain. In: Proceedings of Eurospeech93. Berlin, Germany, 1993. 2: 981–984.

[89] Siu K.C., and Meng H.M. Semi-Automatic Acquisition of Domain-Specific Semantic Structures. In: Proceedings of EuroSpeech 1999, Budapest, 1999. 5:2039-2042

[90] Meng H., Siu K.C. Semiautomatic acquisition of semantic structures for understanding domain-specific natural language queries. IEEE Transactions on Knowledge and Data Engineering. 2002. 14(1): 172-181.

[91] Hua-Ping Zhang, Hong-Kui Yu, De-Yi Xiong and Qun Liu. HMM-based Chinese Lexical Analyzer ICTCLAS. In: Proceedings of 2nd SigHan Workshop, 2003. 184-187

致　谢

衷心感谢我的导师吴文虎教授和郑方教授五年来对我的悉心指导和关怀。两位导师不仅专业上给我很好的指导，而且其严谨求实的治学作风、平易近人的待人原则和忘我的工作精神都将是我长期学习的榜样。在此，谨向两位恩师致以最诚挚的谢意！

感谢语音技术中心的其它老师，包括方棣棠教授、李树青教授、徐明星老师邬晓钧老师，以及实验室的全体同窗，他们与我进行了许多有益的讨论，同时给予我很多工作上的支持，在此一并向他们表示感谢。

最后，衷心感谢我的家人和朋友，他们无私的爱和默默的关怀，一直伴随着我的奋斗过程。

声　明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签　名：_____日　期：_____

个人简历、在学期间发表的学术论文与研究成果

个人简历

1977 年 10 月 6 日出生于山东省平等市。

1996 年 9 月考入清华大学计算机科学与技术系，2000 年 7 月本科毕业并获得工学学士学位。

2000 年 9 月免试进入清华大学计算机系攻读博士至今。

发表的学术论文

[1] **Sun H**, Zhang G L, Zheng F, et al., Using word confidence measure for OOV words detection in a spontaneous spoken dialog system, in Proceeding of the 8th European Conference on Speech Communication and Technology (EuroSpeech). Geneva, Swiss. 2003. 2713-2716

[2] **Sun H**, Xu M X, Wu W H. Classification of dialogue acts using prosodic features in Chinese spontaneous speech. In: Proceedings of the First International Conference on Machine Learning and Cybernetics. Beijing, China, 2002. 3: 1163-1166 (EI indexed, accession number 7598019)

[3] **Sun H**, Xu M X, Wu W H. Study on detection of prosodic phrase boundaries in spontaneous speech. In: Proceedings of 3rd International Symposium on Chinese Spoken Language Processing (ISCSLP), Taibei. 2002.281-284

[4] **孙辉**, 徐明星, 燕鹏举, 吴文虎. 电话语音数据库的收集和标注. 第六届全国语音通讯技术会议论文集, 2001 年. 325-328

[5] Zhang G L, **Sun H**, Zheng F, et al., Robust speech recognition directed by extended template matching in dialogue system. In: Proceedings of the 5th World Congress on Intelligent Control and Automation, Hangzhou, China, 2004. 5: 4207-4210 (EI indexed, accession number 8127463)

[6] Yan P Y., Zheng Fang., **Sun H**., at el. Spontaneous speech parsing in travel information inquiring and booking systems. Journal of Computer Science and Technology. 2002, 17(6): 924-932

被录用的学术论文

[1] **孙辉**, 郑方, 吴文虎. 基于上下文相关置信度打分的语音确认方法. 清华大学学报（自然科学版）. 已录用. (To be EI indexed)