

# 语音识别后文本纠错处理

龚媛<sup>1</sup>, 李蕾<sup>2</sup>

(1. 北京邮电大学智能研究中心, 北京 100876; 2. 北京邮电大学智能研究中心, 北京 100876)

**摘要:** 本文采用了全信息自然语言理解的方法论对人机对话系统语音识别结果进行处理, 发现并纠正其中的错误, 从而提高识别的正确率和可读性。实验初步应用于 863 课题“奥运多语言智能信息服务系统”。

**关键词:** 语音识别; 全信息; 自然语言理解; 语用信息

## Error Correction for ASR Text

Yuan Gong; Lei Li

(1. Research Center of Intelligence of Beijing University of Post and Telecommunications, Beijing 100876; 2. Research Center of Intelligence of Beijing University of Post and Telecommunications, Beijing 100876)

**Abstract:** This paper adopts a function of Natural Language Processing (NLP) which based on Comprehensive Information (CI) to analyze the text after Automatic Speech Recognition (ASR) for error detection and correction, in order to improve the accuracy and of recognition. Tests have been done based on the demo terminal for the National 863 Project of “Olympics oriented Multilingual Intelligent Information Service System”.

**key words:** ASR; CI; NLP; pragmatic information

### 1 引言

人们在应用语言的时候, 总是伴随着特定目的。语音识别在使用时也都是有一定的应用目的。如果参与交流的人没有目的相关的共同知识基础, 他们的交流只能是“对牛弹琴”, 不会带来有意义的效果。正是因为现在的机器缺乏人类的知识, 人与机器的语音交流就远未令人满意。语音识别研究者们经过了多年的努力, 终于可以让机器具有了较好的接受人类声音的知识, 但是声音与自然语言之间还有很大距离, 没有足够的自然语言知识, 声音最终还是不能转化成为正确的语言, 成为影响机器与人类顺利沟通的障碍。为此我们提出自然语言理解与语音识别综合的解决方法: 人类的声音首先输入到语音识别引擎, 进行以音频特征分析为主的处理, 进一步根据应用目的不同, 分别经过不同的自然语言理解模块, 这主要是基于语用信息主导的自然语言知识来进行的。

本课题将针对“奥运多语言综合信息服务”项目的典型示范系统“CityGuide”移动终端研究实现语用信息主导的语音识别后语句检错纠错方法。在语音识别引擎以后引入自然语言理解模块, 综合语法、语义和语用信息对语音识别结果进行分析、检错和纠错。“CityGuide”是在智能手机平台上实现的一个信息服务终端, 支持语音输入/输出, 可为奥运期间来北京的参观旅游者提供住宿、交通、旅游等方面的多语言信息服务。目前该演示系统主要支持单句语音输入, 如“今天晚上还有房间吗?”、“请问最近的车站在哪? ”。但是测试过程中的语音输

---

基金资助: 本文得到国家 863 项目“奥运多语言综合信息服务——基于自然语言理解的语音识别后文本处理研究”(2005AA117010)

作者简介: 龚媛(1982-), 女, 扬州, 在读研究生, E-mail:gongyuan001982@yahoo.com.cn

入识别效果很差，达不到实用化的要求[1]。初步实验测试结果表明，增加了语用分析的自然语言理解可以在一定程度上提高语音识别正确率，关键问题就是如何更好的挖掘和使用语用信息。

## 2 理论基础

### 2.1 全信息自然语言理解

为了使机器理解人类自然语言，人们在长期研究的基础上形成了两种基本的方法：基于规则的方法和基于统计的方法。两种方法都难以独立解决问题。基于全信息的自然语言理解方法论，从“全信息理论”的观点出发，同时发挥规则方法和统计方法两者的优势。事物运动状态及状态变化方式的形式是事物的语法信息；一旦这种抽象的“状态及其变化方式”的形式与它的“客体事物”联系起来，就会具有具体的含义，这就是语义信息；而“状态及其变化方式”的形式以及它的含义一旦与特定的“认识主体”联系起来，就会表现出对主体目的的效用，这就是语用信息[1~2]。

### 2.2 自然语言理解与语音识别

语音作为一种理想的人机通信方式具有自然、方便、快速的特点，让机器能够理解人的语音一直是人们追求的理想，传统语音识别的方法无论是基于统计的模型还是基于规则的模型，对识别的内容并不进行正确性分析。导致这种错误的原因主要是语音模糊。例如将“call a doctor”识别成“call a salt” [3~4]。如果换一种处理思路，采用自然语言理解的方法分析结果的内容，则很容易判断出“call a salt”不符合人们的用法。[5~6]为此我们提出如图1所示的自然语言理解与语音识别综合模型。

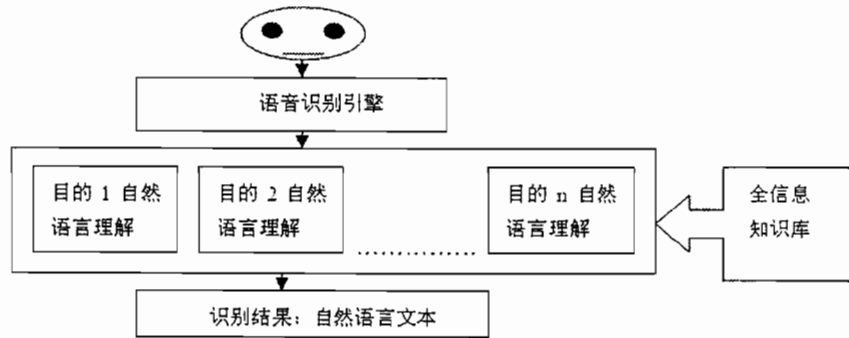


图1 自然语言理解与语音识别综合模型

## 3 全信息纠错算法

本文研究的主要内容：采用全信息自然语言理解的方法论对人机对话系统语音识别结果进行处理，发现并纠正其中的错误，从而提高识别的正确率和可读性。具体包括以下四点：

- 1) 建立基于全信息语音识别文本常识知识库。
- 2) 语音识别输出结果的语法、语义和语用错误识别。
- 3) 语音识别输出结果的错误纠正。

4) 应用于“奥运多语言智能信息服务系统关键技术及其示范系统研究”项目的终端部分，提高终端语音识别结果的正确性。

### 4 语法信息

语法信息主要考虑了语音识别稳定度，即语音识别引擎目前的输出结果的统计特性，通过比较正确的语句和大量的错误训练语句，我们可以集中发现容易出错的部分，举例如图2所示。

可见，其中“给我一个”、“的”、“双人间”的语音识别正确率较高，而“带淋浴”、“请”的识别正确率较低，说明“双人间”作为输入时语音识别稳定度高。如果语句中出现了稳定度很高的词语，那么我们有充足的信心相信他们是正确的词语，由此可以以他们为正确的出发点去为其他可信度较低的词语找错和纠错。

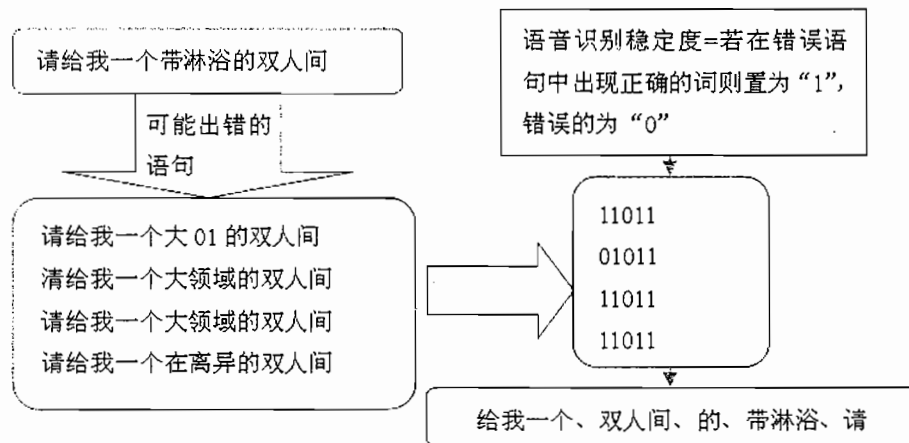


图2 语音识别稳定度算法流程

#### 4.1 语义信息

语义信息主要考察当前词语揭示当前所属类的含义的能力，有核心词和普通词之分。我们用[0, 1]区间内的数值来表示这种语义度量值。其数值的获得可以通过统计方法原理类似于信息检索中常用的tf-idf算法，即考虑文本中的词频和词语的倒排文档频率。就目前项目的情况而言，这部分内容所需要的大规模的语料库条件不具备，所以从长远来讲，应当采用这种方法，但是目前我们将主要精力放在算法本身，对于这块内容就采用简单的人工构建的方法完成。我们根据场景把对话内容分成了饭店、购物、打车、讨论价钱、看病、旅馆、问路、修理、其他9类。

#### 4.2 语用信息

作为对于自然语言处理中语用信息的基础性研究，本课题结合语音识别后文本的检错纠错，通过真实语料的测试来检验和提高语用分析的效果，增强语用分析的实用性。在此过程中，密切关注语用信息与语法信息、语义信息的关系，希望能够总结出三者间较好的结合途径。

本文语用分析主要考察在某种应用目的下，当前语句中一个词语与周围环境的协调适应能力。从最简单的二维环境入手，采用统计方法考察两个词语的协调能力，目前，我们主要考虑对长句子进行分类，建立核心词、普通词及其协调性的语用知识库，经过统计计算找出核心词及关联性。再在识别与纠正过程中运用知识库的知识，得到最优解。例如：

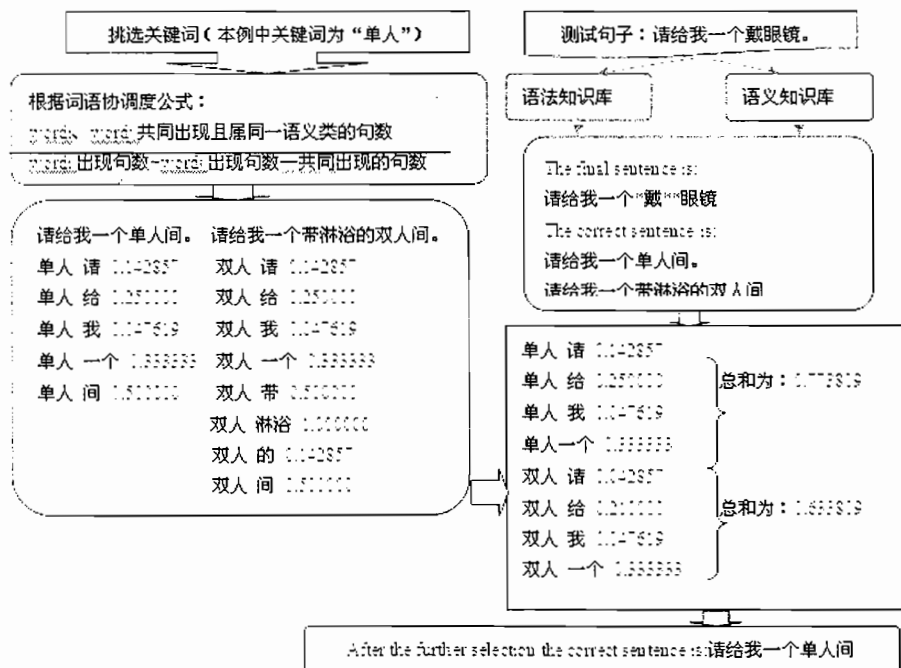


图3 语用知识库算法流程

### 4.3 综合纠错系统组成

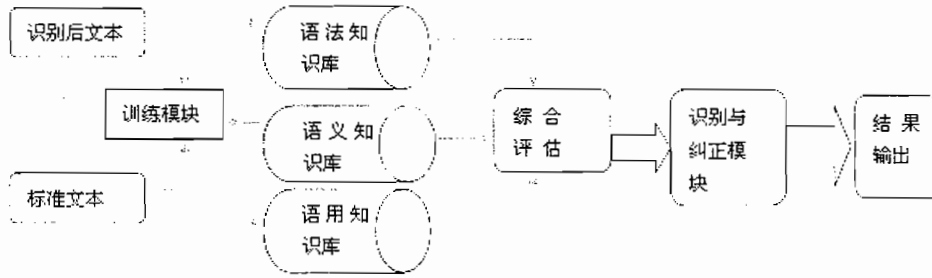


图4 系统组成

### 4.4 算法流程

在训练部分我们主要是建立获得语音识别各类文本常识知识库，包括语法知识库，语义知识库，语用知识库。在测试模块主要调用语法、语义和语用知识库对输入测试句子进行判错和纠错。首先创建 CSentence 与 CTestsentence 类，建立词典，读入句子，切词，与标准答案进行对比，得到每个词语的辨析结果，对应存储到各个库中，其中 CTestsentence 公共继承 CSentence，具体算法流程如下：

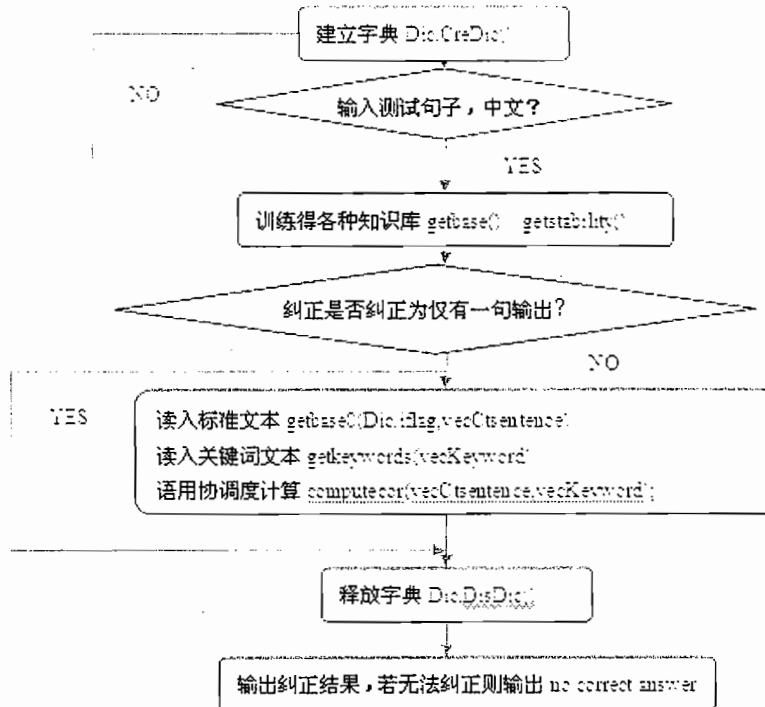


图5 语法、语义和语用算法结合流程图

## 5 实验与结果分析

目前我们主要采用大规模真实语音识别后文本为测试用例，测试通过此系统对语音识别后文本的语法、语用、语义错误进行纠正，判断得到的正确率是否能得到提高，提高程度如何。通过多人运用语音录入软件—IBM viavoice 中文版进行真实语音录入，生成语音识别后文本，然后以此作为系统输入，采用灰盒测试法，得出系统在大规模环境下的实际结果。程序加入文本自动循环输入、判别（纠正结果与标准文本对比）及输出（纠正后文本、纠正正确率，正确率提高百分比），考虑数据庞大，采取逐个文档（单个人录音，67句）统计，个数统计是以句为计量单位。我们采用中文测试用例 1206 例，得出系统在大规模环境下的实际结果。测试结果如表一所示：测试结果举例：

例一: the final sentence:  
\*进\*的\*又\*一个单人\*接\*  
The correct sentence is:  
请给我一个单人间

例二: the final sentence:  
请给我一个\*大\*\*厘米\*的双人间  
The correct sentence is:  
请给我一个带淋浴的双人间

表 1. 测试结果

项目	数值
总测试句子数	1269
(人工判定) 语音识别结果正确的句子数	448
语音识别正确率	35.3%
(语法语义模块加入后) 纠正正确的句子数	331
纠正正确率	52.6%
(加入语用模块后) 纠正正确的句子数	388
纠正正确率	53.7%
语音识别+本模块正确率	65.9%
比未使用本模块时的正确率提高	30.6%

综上所述, 本系统实现了需求规定的功能及性能, 能够较好对语音识别后文本进行分类, 并以较高正确率发现及纠正文本中出现的语法、语义和语用三个层次上的错误。

## 6 总结

本文提出了提高语音识别正确率的一种新思路——在语音识别引擎后增加一个基于全信息的自然语言处理和理解模块, 对语音识别结果进行分析评估, 发现可能的错误并将其纠正。经过实现与测试, 验证了该思路的可行性。我们相信语用信息不是空中楼阁, 它与语法、语义信息是密切联系在一起的, 换句话说, 语用信息是在语法、语义信息基础之上才能存在、描述和使用的。因此, 明确他们之间的关系和互动方式, 充分发挥整体大于部分和的作用, 才能最终实现对于自然语言的理解。

### 参考文献:

- [1] 钟义信. 信息科学原理第三版, 北京邮电大学出版社, 2002.10
- [2] 钟义信.周延泉.李蕾. 信息科学教程, 北京邮电大学出版社, 2005.4
- [3] 李蕾.钟义信. 自动文摘系统中基于全信息词典的复杂语句分析方法及其实现, 2000.8, 电子学报
- [4] 刘建毅.张鹏飞.王枫.郭燕慧.李赞. 北京邮电大学智能研究中心, 《高性能电子邮件过滤系统的设计与实现》, 计算机应用研究, 第四期, 2005 年, Page224-225, Page233.
- [5] 李蕾.周延泉. 一种新型智能移动信息服务探讨, CAAI-11, 2005.10
- [6] 李蕾.周延泉.王菁华. 北京邮电大学 智能科学与技术研究中心, 《基于全信息的中文信息抽取系统及应用》, 北京邮电大学学报, 第 28 卷, 第 6 期, 2005 年 12 月, Page48-51.
- [7] 刘建毅. 基于统计与规则相结合的口语语言理解研究, 北京邮电大学博士论文, 2005.7
- [8] 胡敬东. 语音识别后处理系统的研究, 华南理工大学硕士学位论文, 2000.3.
- [9] 张建平.王作英.赵庆卫.清华大学电子工程系, 语音理解中的容错技术的研究, 电子学报[J], 2000 年 3 月第 3 期, Page 84-86
- [10] 江铭虎, 袁保宗, 林碧琴, 北方交通大学, 噪音环境下语音识别理解系统的研究[J], 铁道学报, 第 19 卷第 6 期, 1997 年 12 月, Page 73-77
- [11] 张瑞强, 王作英, 张建平, 清华大学电子工程系, 带拼音纠错的汉语音字转换技术[J], 清华大学学报(自然科学版), 1997 年第 37 卷第 10 期, Page 9-11
- [12] 李明琴 王作英 陆人衿, 清华大学电子工程系, 语音识别音字转换中的快速容错算法[J], 中文信息学报, 第 16 卷第 5 期, Page 38-43