一种基于语义分析的汉语语音识别纠错方法*)

韦向峰[1] 张 全[1] 熊 亮[1,2]

(中国科学院声学研究所 北京 100080)[1] (中国科学院研究生院 北京 100049)[2]

摘 要 汉语语音识别的研究越来越重视与语言处理的结合,语音识别已经不是单纯的语音信号处理。N-gram语言模型应用到语音识别系统中,大大增强了系统的正确率和稳定性,但它也有其自身的局限性,使得语音识别出现许多语法和语义的错误结果。本文分析了语音识别产生语音和文字方面的错误的原因和类型,在概念层次网络语言模型的基础上提出了一种基于语句语义分析和混淆音矩阵的语音识别纠错方法。通过三个发音人、5万字的声音语料和216句实验语句的纠错测试,本文的纠错系统在纠正语义搭配型错误方面有比较好的表现,可克服N-gram语言模型带来的一些缺陷。本文提出的纠错方法还可以融合到语音识别系统中,以便更好地为语音识别的纠错处理服务。

关键词 语音识别,纠错,语义分析,语言模型,概念层次网络

An Error-correct Approach in Chinese Automatic Speech Recognition Based on Semantic Analysis

WEI Xiang-Feng[1] ZHANG Quan[1] XIONG Liang[1,2]

(Institute of Acoustics, CAS, Beijing 100080)[1] (Groduate School of Chinese Academy of Science, Beijing 100049)[2]

Abstract Now automatic speech recognition (ASR) is not a simplex signal processing. The natural language processing is more and more regarded in Chinese ASR. As a language model, N-gram improved the accurate rate and stability of ASR remarkably. But there are still many syntactic and semantic errors in ASR because of the inherent limitation of N-gram language model. This paper analysed the reson and the types of the phonetic and literal errors in ASR. An error-correct approach in Chinese ASR was proposed in this paper based on sentence semantic analysis, confusion matrix and a language model constructed on hierarchical network of concepts. The error-correct software system runs well especially in correctting the errors of semantic relationship, tested with vocal corpus of 3 person and 50,000 words and with 216 experimental sentences for error-correct. So the new language model constructed on hierarchical network of concepts can overcome the limitation of N-gram model. The approach in this paper also can be merged into ASR to improve the performance of error-correct in ASR.

Keywords Automatic speech recognition (ASR), Error-correct, Semantic analysis, Language model, Hierarchical network of concepts

1 引言

一个典型的语音识别系统通常可以分为声学处理和语言处理两部分,又称为前处理和后处理。对于大字表非特定人的连续语音识别,主要使用隐马尔可夫模型(HMM)进行前期处理,得到语音的发音单元;后期处理主要使用基于语料库统计的N-gram模型(Bi-gram或Tri-gram),完成音-字转换。人类听觉实验表明,人也只能听清楚连续语音流中70%的音节[1]。因此,连续语音识别的前处理很难也没有必要做到90%以上的首选正确音节识别率。当前语音识别的关键已取决于后处理中的解模糊及纠错能力[1]。

语音识别后处理使用基于语料库统计的N-gram语言模型,大大提高了识别系统的稳定性和正确率,使语音识别系统进入了大众市场。但是,N-gram模型没有使用语言学中语义深层的约束规则,难以解决数据稀疏[6]、远距离搭配和语言递归现象等问题[5]。有研究者[7]用"词汇语义驱动"的方法,以词为中心对音-字转换后的汉字序列进行错误分析,根据词语知识和语法规则自动纠错。

本文基于概念层次网络语言模型[1],以语句语义分析处理为中心,使用句类分析的方法对语音识别系统产生的错误进行纠错处理研究,提高语音识别系统的正确率。本文在语音识别结果的基础上,选取了216句测试语句,其中95句含有语音识别错误,使用本文提出的纠错方法对错误语句进行了纠错处理实验。实验结果表明,该纠错方法具有较大的纠错潜力,还可以和语音识别的前处理相结合,得到更好的纠错效果。

2 语音识别错误分析

对于连续语音识别,音节识别的错误可分为3种:吞音、添音和错音。以句子为考察单位,如果识别出的文本音节个数少于句子实际音节个数,这种错误称为吞音,例如"建立残(错案)追究制度";如果识别出的文本音节个数多于句子实际音节个数,这种错误称为添音,例如"中国我(添音)又处在社会主义的初级阶段";如果识别出的文本音节个数与句子实际音节个数相同,但有些音节是错误的,这种错误称为错音,例如"我国今后每年都要(进口)大量小麦"。本文安排3个发音人、使用5万字的声音语料对一个商业语音识别系统进行了测试,结果发现在3种错误类型中,错音占78.2%,添音占

*)本文承国家973项目"自然语言理解的交互引擎研究"(2004CB318104)及中国科学院声学研究所创新项目资助。韦向峰 博士,助理研究员;张 全 研究员,博士生导师。

13.3%，吞音占 8.6%。错音这一错误类型在整个错误中占有很高比例。

对于音-字转换，错误类型可以分为：音对字错和音错字错。音对字错是指音节识别正确但识别出的汉字却是错误的，例如“海浪充饥(冲击)成大片的珊瑚沙滩”；音错字错是指音节识别的结果就是错误的，因此识别的文字结果也是错误的，例如“他们的行程即将姐夫(结束)”。根据对 216 句测试语料的统计(以句子为单位)，在 95 句识别错误的句子中，音对字错的句子占 16.2%，音错字错的句子占 83.8%。这个测试数据再次表明，语音识别中的大部分错误来自于错音，纠错处理首先要解决错音。

错音的纠错处理需要使用音节的混淆音数据。形成混淆音数据有两种方法：一种是使用大量的测试语料收集每个音的混淆音数据，形成音节混淆音数据库。这种方法的困难在于要使用大量的、多个发音人的语音语料，工作量非常大。另一种是考虑汉语的音节由声母和韵母构成，先测试声母和韵母的混淆矩阵，然后根据声韵组合，就可以构造出每个音节的混淆音。这种方法可以大大减少工作量。本文采用后一种方法，首先根据识别的结果构筑声韵混淆矩阵，然后验证音节混淆音是否覆盖正确的结果。根据验证情况，对一些音节进行调整，最终根据错误音节的混淆音给出的正确音候选集中前 10 名包括的正确音大于 95%。

吞音一般出现在句子轻读的音节和非句首零声母音节上，如“的”“了”和“案”“义”等。添音也经常出现在轻读音节和零声母音节上，不过这些音节不是被吞掉而是被添上的。添音、吞音两类错误的纠错要解决轻读音节和零声母音节的问题。

3 基于语义分析的纠错方法

为了纠正语音识别产生的文字错误，需要建立相应的正确语言模型。根据不同的语言模型会有不同的纠错方法，如基于统计语言模型的纠错方法、基于词汇功能语法的纠错方法等等。本文基于概念层次网络语言模型，给出了一种基于语句语义分析的纠错方法，并建立了一个用于语音识别后处理的纠错系统。

根据概念层次网络语言模型，一个语句经过语义分析处理技术(句类分析)[2]，可以得到语句的句类。根据句类知识[4]可以得到构成语句的语义块。语义块是构成语句概念联想脉络的语义单位，分为特征语义块(类似于语法中的述语)和广义对象语义块(类似于语法中的主语、宾语等)。语义块内部又分为核心部分和核心的说明部分。

特征语义块和广义对象语义块之间的概念约束实际上是一种远距离的语义搭配。根据对测试语料的统计，在语音识别的错误句子中，特征语义块-广义对象语义块搭配错误的占到了 58.3%，其中特征语义块核心部分出现错误的占 30%，广义对象语义块核心部分出现错误的占 28.3%。例如“我就一直处以(注意)鱼雷的发展”中的错误属于特征语义块核心错误，而“她身上穿着一件淡褐色倡议(绸衣)”则属于广义对象语义块核心错误。

3.1 语句的语义分析方法

语句语义分析的目的是得到语句的语义类型，给出语句的语义块构成，揭示语义块之间的概念关联性；对于语义块，要分析其内部构成，揭示其构成之间的概念关联性，从而得到整个语句的概念联想脉络。语句语义分析的主要过程如图 1 所示。

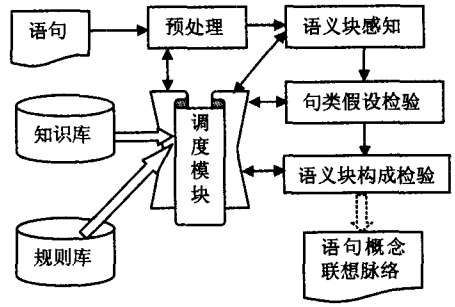

图 1 句类分析系统模块图

在图 1 中，语音识别后的文字语句经过预处理、语义块感知、句类假设检验、语义块构成等分析阶段可得到语句的概念联想脉络。知识库和规则库是整个语句语义分析过程的支撑，在调度模块的控制下根据语句的现场知识将库中的语言知识以规则的形式作用于各个分析阶段。

知识库包括概念知识库、语言知识库和常识知识库。概念知识库包括 108 个高层概念节点以及它们下面延伸的概念节点和对应的知识、57 组基本的语句语义类型以及 3192 组混合语句语义类型和对应的知识，概念节点之间的关联性知识则形成概念关联知识库。语言知识库主要是汉语的字词知识库，字知识库包含 2333 个汉字及相应的语义知识，词知识库包含 41049 个汉语词汇及相应的语义知识。常识知识库需要根据具体的应用领域来制定，例如地理常识知识库。

预处理对文字语句进行粗切分，对于含有切分歧义的文字段作为一个整体保留，在后面的分析阶段再处理。对于可以合并的字串，例如英文字母段、数字段、时间短语、数量短语等，合并为一个处理单位。同时，根据字词的概念知识库，得到语句中各字词的语义知识，为语义块感知分析阶段提供数据准备。

在语义块感知阶段，主要是找出语言逻辑类 l 概念和动态 v 概念。根据规则库的感知规则，对语句的格式和特征语义块构成进行假设，l 概念与语句格式密切相关，而 v 概念常常构成特征语义块的核心。根据特征语义块假设可以进一步给出语句的句类假设，为句类假设检验阶段提供数据准备。

在句类假设检验阶段，采用假设-检验的方法：首先假设某一个句类为语句的语义类型，根据该句类在知识库中的句类知识，对语句的格式和语义块假设进行概念关联性检验。如果检验成功，那么该句类就是语句的语义类型，相应得到构成语句的语义块；如果检验失败，那么取下一个假设句类进行检验。检验的主要依据是知识库中的句类知识、概念关联性知识和规则库中关于句类检验的规则。

对于通过检验的句类，到了语义块构成阶段需要对广义对象语义块的内部构成进行分析，给出语义块的内部构成及其概念关联性。这样，就得到了整个语句的语义类型、语义块构成、语义块之间以及语义块内部的概念关联性。

规则库由一条条的形式化规则构成，规则可以用于语句语义分析的各个阶段。形式化规则已经完全符号化，可以直接由计算机解释执行。例如下面的一条形式化规则就是关于特征语义块假设的形式化规则，表示“如果当前节点(字或词)是 vv 兼类动词，后一个节点是 v 概念动词，那么当前节点假设为特征语义块前部 EQ，后一节点假设为特征语义块核心 E”。

_ConceptCategory((0)). INCLUDE. “vv” , _ObjType

((1),"CC").IS."v" => _SETOBJTYPE((0),
"CHUNK","EQ"),_SETOBJTYPE((1),
"CHUNK","E")

调度模块负责访问知识库和规则库,并将语句语义分析各阶段中的现场信息和库中的约定信息比较,根据分析策略和规则判断是否使用规则库中的规则对语句语义分析结果进行修正,从而得到正确的语句语义分析结果。在实际的分析过程中,有时还需要进行回溯处理,例如全部句类假设检验都失败时需要回溯到语义块感知阶段,重新假设语句的句类和语义块。

以上对于语句语义分析方法的整个流程和各个模块逐一进行了简单介绍。基于这种语句语义分析方法,可以找出语音识别文字中的错误,并对错误进行纠正。具体的错误发现和纠错方法在本文的 3.2 和 3.3 小节中介绍。

3.2 错误的发现

语音识别结果中出现错误的字或词,主要是因为语句不合理或者不合法。出现错误的原因一般有以下几种:(1)不符合语法规则。字词的使用在语句中受到句法规则的约束,不符合语法规则的字词就会产生错误。例如"静止的孤立的研究马克思主义"中的"的"字应该改为"地"。(2)不符合上下文的约束,例如"她即使(既是)妖艳的女人,又是能干的主妇。"(3)不符合语义约束,例如"这些人几乎把整个实践(世界)打得精疲力尽"。

对于不符合语法规则产生的错误,可以通过制定确定的纠错规则库,然后根据形式化的规则去纠正。对于不符合上下文产生的错误,需要获取关于上下文的语境知识,有的需要相应的常识知识,目前还没有形成很好的解决办法。对于不符合语义约束而产生的错误,本文给出一种依据语义关联性的错误发现方法,称为孤魂发现。

"孤魂"是指语句中出现的孤立的不合理的字或词语,与语句中其他词语没有任何概念关联。根据语句语义分析方法,可用句类知识和概念关联性知识判断语句中语义块以及词语之间是否存在概念关联。因此,对于存在语音识别错误的语句,可以找出与其他词语没有关联的词语,即发现孤魂。

孤魂有可能就是语句中错误的地方。通过混淆音矩阵可以给出孤魂的可替换词语候选集,尝试候选集中的所有词语,如果使用语句语义分析方法找到正确的具备语句联想脉络的词语,就完成了纠错处理。

孤魂发现会遇到"其他"词语也是"孤魂"的情况,此时形成"孤群"。例如"为称波兰是名跃过编辑性宫殿台(伪称波兰士兵越过边境进攻电台)"。测试表明,语音识别的文字错误中多数是孤魂而不是孤群。如果孤魂候选集中没找到正确的词语,纠错处理系统与用户交互,由用户修改处理。

3.3 纠错处理系统

纠错处理系统的流程如图 2 所示。首先,通过"语音识别软件控制"获取语音识别的结果数据,利用结果数据构造构筑声韵混淆矩阵,形成混淆音数据库。

结果数据中的文字则进入"接口"。"接口"负责保存文字语句中的词语,并查找语句语义分析系统中的词语知识库,得到语句中和词语有关的概念知识,形成"临时词知识库"。

"预处理"根据词语的概念符号,计算输入语句中各词语之间概念的相关程度,把没有关联性的词语记录到"孤魂可疑集合"。对于预处理中连续出现的多个孤魂-孤群,目前不作任何处理,而是将包含孤群的语句直接送到"结果显示和交互处理",由用户进行处理。

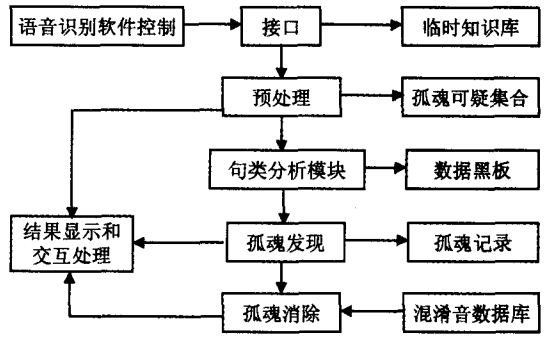

图 2 纠错处理系统流程图

"句类分析模块"依据语句的语义分析方法,对"预处理"后的语句,利用语句语义分析系统中的句类知识、概念关联知识和规则库中的规则,对语句进行语义分析处理,并将结果(含中间结果)保存到"数据黑板"。

"孤魂发现"综合预处理形成的"孤魂可疑集合"和句类分析模块形成的"数据黑板"中的数据,确定可能的"孤魂",并将其记入"孤魂记录"。

"孤魂消除"则根据"孤魂记录"中记录的孤魂,结合根据语音识别结果数据形成的混淆音数据库,得到候选词语集合。然后依据语句语义分析方法中的概念关联检验和句类假设检验判断候选集中是否有适当的词语,完成孤魂消解。最后,纠错系统将纠错处理的结果显示给用户。

可见,基于语句语义分析的语音识别纠错系统的主要过程是:接收来自语音识别的数据,先进行预处理,然后进行语义分析,根据预处理和语义分析的结果进行孤魂发现,利用混淆音矩阵完成孤魂消除和纠错处理。

请看一个具体的语音识别纠错例子。语音识别系统输出的句子为"我国今后每年都要今后许多小麦"。

(1)预处理:根据词语概念的关联性,进入孤魂可疑集合的词语是:我国、今后(第二个)。

(2)语义分析:语句中没有可以假设为全局特征语义块的动态概念,同时也没有词语表明语句属于无特征语义块的句类。

(3)孤魂发现:由于"今后(第二个)"前面有时间概念("今后每年"),同时有特征语义块的逻辑说明概念("都要"),这两类概念合起来经常充当特征语义块核心的前修饰部分。这些表明这个位置应当出现一个动态概念,因此"今后(第二个)"可能是孤魂。

(4)孤魂消除:根据孤魂发现给出的孤魂记录和对该位置上的概念预期情况,利用混淆音数据可给出这个位置上的候选词语集,得到"浸透""进口"两个动态概念。用它们对应的句类知识对语句进行检验,否定掉"浸透",认可"进口"。

(5)完成纠错处理,给出显示结果"我国今后每年都要进口许多小麦"。

4 实验结果和讨论

本文选取来自《人民日报》、《战争风云》等处的测试语料数据,文体涉及表述文、叙述文、评述文和论述文 4 种文体。使用 5 万字的语料,由 3 个不同的发音人对一个商业语音识别系统进行了测试。对语音识别系统产生的吞音、添音和错音等错误进行了统计,其中吞音占 8.6%,添音占 13.3%,错

音占 78.2%。

基于该商业语音识别系统及其接口，本文搭建了一个汉语语音识别纠错系统，通过构筑声韵混淆音矩阵和基于语句语义分析的纠错方法，可以对语音识别结果中的错误进行纠错处理。使用 216 句测试语句对纠错处理系统进行测试，其中有 95 句含有语音识别错误。纠错处理系统发现 74 句错误语句，错误发现率为 78%。在发现错误的语句中，纠错系统纠错后语句正确的有 53 句，纠错的正确率为 72%。

测试结果表明，纠错系统可以提高语音识别系统的识别正确率，尤其对由于错音引起的识别错误或语义搭配型错误纠错，效果较好。这主要是因为该商业语音识别系统采用了 Tri-gram 统计语言模型，这种语言模型将人类的语言感知过程作为一个黑箱来处理，用概率(或频度)的方法来模拟。如果遇到的词串是训练语料中出现过的、高频的，会给出很好的结果；否则，只能根据概率最佳给出结果。这种模型和方法即使在最好的情况下，也仅仅只是给出了词语紧邻的搭配知识，得到的数据不能全面反映词语间的语义关联知识。

N-gram 语言模型无法解决非常用词的数据稀疏问题。语句中词语之间的关联不是线性连续的关系，而是一种多层级的关系。N-gram 语言模型无法具体区分这些不同，都简单地用概率给予表达，无法给出符合概念联想脉络的词语，也无法判断给出的结果是否正确，只能得到一个概率最优的结果。

本纠错系统所依据的概念层次网络模型可以克服 N-gram 语言模型所固有的缺陷，揭示语句中词语之间的概念关联性知识，通过知识库和规则库对词语进行正确有效的语义约束。这种语义约束关系正好可以用来对语音识别系统产生的错误进行语义纠错。

结束语 语音识别系统所处理的语音信号具有两个特点：一个是存在环境噪声，而且环境噪声的声学特性与人的语音相近；第二个是不同的发音人语音具有较大的差异，即使是同一个人，在不同时间的语音都存在一定的差异。对于连续语音识别，由于发音的连贯性，音位和音位之间的连续平滑过渡必然造成识别中音与音、字与字、词与词之间的分割困难。目前，仅凭单纯的信号处理已经很难提高语音识别系统的正确率，语音识别正确率的提高越来越取决于语音识别的后期处理和语言理解模型。

基于 N-gram 的统计语言模型便于处理大规模的、经常出现的语言现象，但是对于语言深层次的语义约束等问题却难以解决。本文基于概念层次网络语言模型，提出了一种基于语句语义分析的语音识别纠错方法。实验表明，该方法在纠正因语义约束而产生的错误方面有很好的效果。如果能够把这种方法和语音识别系统相融合，将可以提高语音识别的正确率。关于如何根据上下文语境、利用超语句的更大范围的语言知识来对语音识别的错误进行纠正，则是本文进一步的研究方向。

参 考 文 献

1 黄曾阳. HNC(概念层次网络)理论[M]. 北京：清华大学出版社，1998
2 晋耀红. 基于 HNC 理论的句类分析系统的设计与实现[D]：[硕士学位论文]. 北京：中国科学院声学研究所，1998
3 孙伟峰. 基于句类的指代解析及其在语音识别中的应用[D]：[硕士学位论文]. 北京：中国科学院北京软件工程研制中心，2000
4 苗传江. HNC 句类知识研究[D]：[博士学位论文]. 北京：中国科学院声学研究所，2001
5 王轩，等. 语音识别中统计与规则结合的语言模型[J]. 自动化学报，1999，25(3)：309～315
6 关毅，等. 现代汉语计算语言模型中语言单位的频度-频级关系[J]. 中文信息学报，1999，13(2)：8～15
7 赵力，等. 汉语连续语音识别中语音处理和语言处理统合方法的研究[J]. 声学学报，2001，26(1)：73～78.

(上接第 151 页)

5.3 HMMs，MEMMs 和 CRFs 的比较

第 3 节从理论上说明了 CRFs 的优越性。为了验证这一结论，本文在相同的数据集上分别采用 HMMs，MEMMs 和 CRFs 进行了词性标注，所有的标注均遵照规范[4]。其中，HMMs 使用一阶的马尔科夫模型，MEMMs 和 CRFs 分别使用了模板 A 和模版 B 进行实验。开放测试实验的结果如表 6 所示。

表 6 HMMs，MEMMs 和 CRFs 词性标注结果比较

模型	HMMs	MEMMs		CRFs	
		模板 A	模板 B	模板 A	模板 B
总准确率	92.03%	91.71%	95.65%	92.31%	96.60%

由表 6 可以看出，当采用模板 A 时，标注的效果按照高到低的顺序依次为 CRFs，HMMs，MEMMs。这是因为 MEMMs 存在着 Label Bias 问题，因此效果最差，而 CRFs 效果最好。当采用模板 B 时，MEMMs 的效果好于 HMMs，而 CRFs 的效果比 MEMMs 好，这主要得益于 CRFs 和 MEMMs 都能够采用任意的特征来构造模板，从而为模型标注效果的提高提供了可能性和方便性。

结论和展望 本文使用 CRFs 对中文进行词性标注。除了使用词的上下文信息之外，对于兼类词，利用词在训练集中的统计信息，得到词最可能的词性，为特征模板添加新的特征；对于未登录词，利用中文词的构词特点以及成语词典列表生成新的特征。在实验中，对实验数据和实验结果进行了统计分析。实验结果表明，使用 CRFs 的中文词性标注方法能够取得很好的标注效果。在相同的条件下标注效果优于 HMMs 和 MEMMs。

目前，对未登录词的标注效果还不是非常理想，今后的工作将针对未登录词，深入地调研汉语的构词特点对词性的影响，以及不同词性的词各自的特点，使用更多的特征，进一步提高标注的正确率。

参 考 文 献

1 Lafferty J，McCallum A，Pereira F. Conditional Random Fields：Probabilistic Models for Segmenting and Labeling Sequence Data. In：Proceedings of the 18th International Conf on machine Learning，2001. 282～289
2 周明，吴进，黄昌宁. 用于词性标注的一种快速学习算法对 Brill 的基于变换算法的一项改进. 计算机学报，1998(4)：357～366
3 Sha F，Pereira F. Shallow Parsing with Conditional Random Fields. In：Proceedings of Human Language Technology Conference and North American Chapter of the Association for Computational Linguistics (HLT-NAACL)，2003
4 现代汉语语料库加工规范-词语切分与词性标注. 北京大学计算语言学研究所，1999
5 白栓虎. 基于统计的汉语词性自动标注方法. 语文建设，1994(10)：38～40
6 Bai Shuanhu. An Integrated Model of Chinese Word Segmentation and Part-of Speech Tagging. In：Advanced and Applications on Computational Linguistics，Third National Computational Linguistics Meeting，Shanghai. Nov. 1995. 56～61
7 Bai S H，Xia，Y，Huang C N. Automatic Part-of-Speech Tagging System of Chinese：[Technical Report]. Beijing：Tsinghua University，1992