一种改进的基于决策树的英文韵律短语边界预测方法

张元平，凌震华，戴礼荣，刘庆峰
（中国科学技术大学 电子工程与信息科学系，合肥 230027）

摘　要： 在英文语音合成系统中，韵律短语边界预测的精度对合成语音的自然度和可懂度有着至关重要的影响。基于决策树的预测方法是现阶段最为常用的韵律短语边界预测方法，但因决策树构建时受到数据平衡性制约，难以针对关键词进行建模，而且在基于决策树进行预测时采用了局部最优的搜索方式无法达到全局最优。所以，为了进一步提升韵律短语边界的预测效果，对基于决策树的预测方法进行了改进，引入韵律短语条件概率，使用 Viterbi 算法同时优化韵律短语边界概率和条件概率，并提出了基于关键词在韵律短语中的位置分布特性的决策树节点概率优化方法。实验表明，在基线系统上使用改进方法后，F-Score 由 68.7% 提升到 77.8%，而不可接受率从 22.4% 降低到 15.2%。

关键词： 语音合成；韵律短语；边界预测；决策树；位置分布
中图分类号： TP391.4　　**文献标志码：** A　　**文章编号：** 1001-3695(2012)08-2921-05
doi:10.3969/j.issn.1001-3695.2012.08.032

Improved decision tree based method for English prosodic phrase boundary prediction

ZHANG Yuan-ping, LING Zhen-hua, DAI Li-rong, LIU Qing-feng
(*Dept. of Electrical Engineering & Information Science, University of Science & Technology of China, Hefei* 230027, *China*)

Abstract: In English speech synthesis systems, the accuracy of prosodic phrase boundary prediction has a critical influence on the naturalness and intelligibility of synthetic speech. Currently, decision tree based prediction is the most popular method for predicting the prosodic phrase boundaries. However, this method can't build models for specific keywords because of the data balance issue. Besides, it wouldn't be possible to achieve the global optimization by the local optimization search method at prediction stage. Therefore, in order to improve the prediction performance, this paper introduced the conditional probability of prosodic phrases, and used Viterbi algorithm to optimize the prosodic phrase boundary probability and conditional probability simultaneously. Furthermore, it proposed an optimization method for probability distribution of the decision tree nodes, based on location distribution characteristics of keywords in prosodic phrases. The experimental results show that F-Score of phrase boundary prediction increases from 68.7% to 77.8% and the non-acceptance rate drops from 22.4% to 15.2% after adopting the proposed method.

Key words: speech synthesis; prosodic phrase; boundary prediction; decision tree; location distribution

0　引言

语音合成技术日趋成熟，已广泛应用于手机、导航、智能家电以及声讯服务等各个领域，为人们的生活带来了便利。但其合成的语音仍然比较机械，无法像自然语音一样能够在句子中间有或长或短的停顿、抑扬顿挫、有快有慢，即还普遍存在着自然度不够高的问题，这与多方面的因素有关，其中韵律层次预测的准确度不高是其重要原因。

一般而言，语音合成系统是由文本分析前端和语音生成后端两大部分组成的。韵律层次预测是语言学前端中必不可少的一个环节，后续停顿、重音、时长的预测，以及基频曲线的生成都与韵律层次的预测结果密切相关。然而准确地预测韵律层次结构并不简单，一方面由于语言的灵活性，很多句子存在着标注多样化的问题，即给定一个句子，不同发音习惯的人给出的韵律层次标注结果可能不同；另一方面韵律层次的划分不仅与语法相关，还与语义、短语长度、讲话风格等因素相关，各因素是综合起来起作用的，需要运用统计方法把各个影响因素较为合理地组织在一起[1,2]。

对于韵律层次，目前还没有一致的划分标准。本文参考 ToBI 标准[3,4]，将语音合成中的韵律层次分为七级，即音素边界、单词边界、隐短语边界、次短语边界、主短语边界、子句边界以及句群边界。各韵律层次由低到高，记为 LP、L0、L1、L2、L3、L4、L5。不同层次的听感各不相同[5]，其中：L3 结束感稍强；L2 结束感要弱于 L3，可以有停顿也可以没有，变化较多，标注人员常常不能准确把握两者的这些细微区别，精确地标注出边界，所以在实际语音合成系统中将 L2 归并到 L3 统一处理，一起视为韵律短语；L4 和 L5 的边界预测比较简单，直接根据文本结构就可以决策出来；LP、L0 边界预测通过查询词典及字母

收稿日期： 2011-12-30；**修回日期：** 2012-02-15

作者简介： 张元平(1982-)，男，安徽宣城人，硕士，主要研究方向为语音信号处理（linxi@ustc.edu）；凌震华(1979-)，男，副教授，博导，博士，主要研究方向为语音信号处理；戴礼荣(1962-)，男，教授，博导，博士，主要研究方向为语音识别、语音合成、音视频检索、实时 DSP 技术、自适应信号处理等；刘庆峰(1973-)，男，中国科学技术大学客座教授，博导，博士，主要研究方向为语音信号处理.

发音转换规则即可决策出来,相对而言也比较简单;而 L1 边界使用拆解和拼接规则也能较方便地得到,韵律层次预测的真正难点还是在于 L3 边界预测。

针对 L3 边界预测,国内外学者已经提出了许多不同的方法,在较早期的语音合成系统中使用得较多的是基于规则的预测方法,即专家通过对语音学和语言学的研究总结出一些先验规则,预测 L3 边界只需匹配这些规则即可,但因为 L3 产生的复杂度远远大于有限的先验规则,所以此方法已很少单独使用,仅配合其他方法用于优化预测效果。随着大语料库制作技术的出现以及计算机技术的发展,基于数据驱动的预测方法逐渐占据主导地位,先后出现了基于决策树的预测方法[6]、基于最大熵准则的预测方法[7]、基于神经网络的预测方法[8]、基于 SFC(superposition of functional contours)分层叠加模型的预测方法[9]、基于随机的上下文无关语法的预测方法[10,11]以及基于 CRF(conditional random fields)模型的预测方法[12]等。目前基于决策树的预测方法因训练时间相对较短,分类模型简单直观、易于理解,所以应用最为普遍。但因决策树是通过数据驱动方法得到的,构建决策树时对数据平衡性要求较高,难以针对关键词进行建模,而且在基于决策树进行预测时采用了局部最优的搜索方式无法达到全局最优,所以基于决策树的预测方法仍有进一步改进的余地。

Sanders 等人[13]提出了使用统计词性信息的方法进行 L3 边界预测;牛正雨等人[14]也提出基于边界点词性特征统计的 L3 边界预测方法,统计了韵律短语边界的词性组合模式以及概率信息,据此在词性序列中预测 L3 边界。尽管他们的预测效果一般,但也说明了统计标注数据库得出某些统计特性在 L3 边界预测中是有意义的。

本文通过对较大规模的 L3 人工标注数据库进行统计,发现很多单词出现在 L3 头部或尾部的概率非常高,即具有明显的 L3 头部或尾部倾向性。一方面,因该类单词数量较多,在基于决策树的预测方法中,如果将它们逐一设计到属性问题集中去,则容易导致在构建决策树的过程中训练数据稀疏,反而影响决策树的预测效果。赵晟等人[15]通过实验也得出了词语直接作特征对基于决策树的韵律结构预测没有改进的结论。另一方面,如果不将它们设计到属性问题集中去,则决策树算法无法自动通过其他问题组合充分挖掘如此细致的统计特性。

基于以上考虑,本文对传统的基于决策树的预测方法进行了两方面的改进:a)考虑到对 L3 边界概率进行硬判决,预测结果的正确率不高,本文引入了 L3 条件概率,使用 Viterbi 算法同时优化 L3 边界概率和 L3 条件概率,以选择出最优的 L3 边界;b)提出了基于关键词在 L3 中位置分布特性的决策树节点概率优化方法,利用关键词在 L3 中不同位置的概率分布来调整实际语境下其前、后边界的 L3 边界概率,从而改善了 L3 边界的错判与漏判问题,降低了 L3 边界预测的不可接受率。

1 基于决策树的 L3 边界预测

在本文的英文语音合成系统中,预测 L3 边界前已进行了 L1 边界划分,所以 L3 边界预测就等价于判断某一个 L1 边界是不是 L3 边界。

而决策树应用于 L3 边界预测的基础是把 L3 边界预测看做一个分类任务,即把每个 L1 边界分到不同韵律边界的任务,可能的韵律边界为 L1 边界或 L3 边界。

本文采用了 C4.5 决策树算法构建 L3 边界预测决策树,分为两个阶段:a)利用训练样本生成决策树模型;b)通过删除部分节点和子树以避免过度学习,即决策树的剪枝。

在决策树的生成阶段,从决策树的根节点开始,根据信息熵下降最大的原则进行分裂,直至在当前叶子节点中所有测试样本都不能使信息熵下降,或当前叶子节点中所含训练样本的数目小于最小阈值时停止分裂。

在决策树的剪枝阶段,为减小生成树的规模并提升决策树的集外测试正确率,C4.5 算法将对已生成的完整决策树进行剪枝。剪枝采用的是基于规则的后修剪(rule post-pruning)方法[16],它将决策树转变为一组规则集,每一条从根节点到叶子节点的路径即是一条规则,通过调整该规则集来提高精度。

具体的调整方法是将该规则集按照在训练集上的分类精度进行排序,每次删除一条规则,使得该规则集的分类精度得到最大的提高。当删除当前规则不能提高分类精度时,则停止剪枝。

通过上述方法构建好决策树后即可用此决策树来预测 L3 边界。根据输入文本的韵律环境属性,从决策树根节点开始逐层递进查找到相应的叶子节点。该叶子节点中包含的训练样本可能隶属于 L1 边界,也可能隶属于 L3 边界。根据不同类型训练样本的数量比例可得到该叶子节点隶属于 L3 边界的概率,该概率记为 P_{L3},决策树的输出可由 P_{L3}的大小来决定。

$P_{L3}=1$,则决策树输出为 L3 边界;$P_{L3}=0$,则决策树输出为 L1 边界;$0<P_{L3}<1$,简单的硬判决方法是 $P_{L3}>0.5$,输出 L3 边界,$P_{L3}<0.5$,输出 L1 边界,本文的基线系统即采用了该方法。但该方法无法有效利用 P_{L3}的信息,且无法融入 L3 长度分布特性,预测结果的正确率不高。为此,本文引入 L3 条件概率 P_s,使用 Viterbi 算法进行软判决处理,同时优化 P_{L3}和 P_s,以选择出最优的 L3 断点。

2 利用 Viterbi 算法修正结果

一个 L4 如果含有 N 个 L1,则需要确定 $N-1$ 个断点的决策结果,共 2^{N-1}种结果组合,即 2^{N-1}个解空间,如图 1 所示。本文通过 Viterbi 算法在这个解空间中寻找一条最优路径,记为 $(x_1,\cdots,x_i,\cdots,x_N)^*$,其中,$x_i\in\{L1,L3\}$。

本文对较大规模的已标注 L1 边界的英文文本进行了统计,在 N 取不同值时 L4 数目在所有 L4 数目中的占比如图 2 所示。

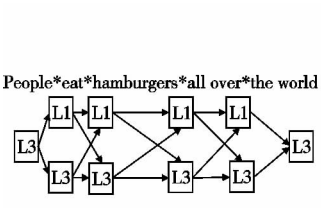

图1 Viterbi解空间示例

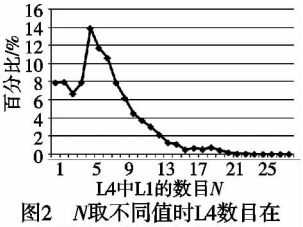

图2 N取不同值时L4数目在所有L4数目中的占比

当 $N>10$ 时,所占比例均不超过 4%,且其决策结果组合数 2^{N-1}不断以指数上升,导致数据较为稀疏,所以仅按式(1)统计出了 $1<N\leqslant 10$ 时的 L3 条件概率为

$$P_{s(N)}(x_i|x_{i-1})=\frac{\text{count}(x_{i-1},x_i|N)}{\text{count}(x_{i-1}|N)} \tag{1}$$

其中:count($x_i|N$)表示在训练语料的所有 L4 中,L1 的总数目为 N 时,且第 i 个 L1 边界的 L3 标注结果为 x_i 所出现的次数。本文以 $P_{L3}(x_i)$作为目标代价,以 $P_{s(N)}(x_i|x_{i-1})$作为连接代价,则

$$(x_1,x_2,\cdots,x_N)^* = \underset{x_1,x_2,\cdots,x_N}{\operatorname{argmax}} P_{L3}(x_1)\times \prod_{i=2}^{N}[P_{s(N)}(x_i|x_{i-1})\times P_{L3}(x_i)] \quad (2)$$

这就得到了 $1<N\leqslant 10$ 时的预测结果。而当 $N>10$ 时,需要采用分段处理方式[2]。

首先在第 6 ~ 10 个 L1 之间挑选一个 P_{L3}最大的断点,再将此断点之前的部分按照 $1<N\leqslant 10$ 时的情况单独进行 Viterbi 判决。假设在这一部分共找到 k 个 L3 边界,如果 $k=1$,则保留此 L3 边界,并将下一步预测的起始位置设置为当前的 L3 边界;如果 $k>1$,则保留前 $k-1$ 个 L3 边界,并将下一步预测的起始位置设置为第 $k-1$ 个 L3 边界。依次递进,直到处理完整个 L4。

3　基于关键词在 L3 位置分布特性的决策树节点概率优化方法

为进一步改善 L3 边界的预测结果,本文根据相关的统计结果,使用了基于关键词在 L3 中位置分布特性的决策树节点概率优化方法。根据英语语法知识,不能作句首的词基本不应该出现在 L3 头部,不能作句尾的词也很少出现在 L3 尾部。另外,不少高频介词和连词往往出现在 L3 头部。换言之,即部分单词应具有明显的 L3 头尾倾向性,如“me”基本不会出现在 L3 头部,“according” 基本不会出现在 L3 尾部。

因这些单词的数量较多,可能有好几千条,所以在基于决策树的预测方法中,如果将它们逐一设计到属性问题集中去,则容易导致在构建决策树的过程中训练数据稀疏,反而影响决策树的预测效果;但如果不将它们设计到属性问题集中去,则决策树算法又不可能通过其他问题组合挖掘出如此细致的统计特性。为此,本文使用了基于关键词在 L3 中位置分布特性的决策树节点概率优化方法。

首先,对训练数据进行统计,单词出现在 L3 头部,即为 L3 的第一个单词,计入 N_{HEAD};单词出现在 L3 尾部,即为 L3 的最后一个单词,计入 N_{TAIL};其他位置则认为单词位于 L3 的中部,计入 N_{MID}。在 Viterbi 算法修正结果的过程中,可以通过引入 N_{HEAD}、N_{TAIL}以及 N_{MID}之间的比率关系来影响 L3 的预测结果,本文采用了其中的两个比率因子:

$$P_{H/T} = N_{\mathrm{HEAD}}/(N_{\mathrm{HEAD}}+N_{\mathrm{TAIL}}) \quad (3)$$

$$P_{M/\mathrm{ALL}} = N_{\mathrm{MID}}/(N_{\mathrm{HEAD}}+N_{\mathrm{MID}}+N_{\mathrm{TAIL}}) \quad (4)$$

$P_{H/T}$越大,说明该单词出现在 L3 头部的倾向性越大;反之,说明该单词出现在 L3 尾部的倾向性越大。而 $P_{H/T}$接近于 0.5,说明该单词不具有 L3 头尾倾向性。

具有较明显的 L3 头尾倾向性的单词,在输入文本中的实际位置有可能处于 L1 边界,也可能处于 L1 中间。对于处于 L1 边界的单词,基于决策树预测方法可以得到该单词前边界或后边界的 P_{L3},通过调整 P_{L3}就能有效影响最后的 L3 边界决策结果。P_{L3}的调整幅度主要由 $P_{H/T}$来决定,调整方法如下:

a)如果是具有 L3 头部倾向性的单词处于 L1 头部,则调高该单词前边界 P_{L3}。

$$P_{L3} = P_{L3}+\mu\log(1+P_{H/T}) \quad (5)$$

b)如果是具有 L3 头部倾向性的单词处于 L1 尾部,则调低该单词后边界 P_{L3}。

$$P_{L3} = P_{L3}-\mu\log(1+P_{H/T}) \quad (6)$$

c)如果是具有 L3 尾部倾向性的单词处于 L1 头部,则调低该单词前边界 P_{L3}。

$$P_{L3} = P_{L3}-\mu\log(2-P_{H/T}) \quad (7)$$

d)如果是具有 L3 尾部倾向性的单词处于 L1 尾部,则调高该单词后边界 P_{L3}。

$$P_{L3} = P_{L3}+\mu\log(2-P_{H/T}) \quad (8)$$

其中:μ 定义为 $P_{H/T}$的实验参数。据此得到新的 L3 边界概率 P_{L3}。如果 $P_{\mathrm{L3}}>1$,则取 $P_{\mathrm{L3}}=1$;如果 $P_{\mathrm{L3}}<0$,则取 $P_{\mathrm{L3}}=0$。

当具有 L3 头尾倾向性的单词处于 L1 边界时,通过 $P_{H/T}$及 $P_{H/T}$的实验参数 μ 来调整 P_{L3}的大小以优化预测结果。但当这些单词不在 L1 边界而处于 L1 中间时,因其前后边界都不是断点,所以无法直接使用 $P_{H/T}$的相关信息,为此,需要使用 $P_{M/\mathrm{ALL}}$来进一步优化预测结果。

$P_{M/\mathrm{ALL}}$越小,说明该单词出现在 L3 前后边界的概率越大。当 $P_{M/\mathrm{ALL}}$小到一定的范围时,就认为输入文本的 L1 划分可能出错了,可以强行地在该单词前边界或后边界划分出一个断点,与其他断点一起交由 Viterbi 算法选择出最优断点。具体分为下面两组来处理:

a)如果该单词为 L3 头部倾向性单词,则在其前边界增加一个新的决策断点,且

$$P_{L3} = \delta\times P_{H/T}\times(1-P_{M/\mathrm{ALL}}) \quad (9)$$

其中:δ 为 $P_{M/\mathrm{ALL}}$的实验参数。

b)如果该单词为 L3 尾部倾向性单词,则在其后边界增加一个新的决策断点,且

$$P_{L3} = \delta\times(1-P_{H/T})\times(1-P_{M/\mathrm{ALL}}) \quad (10)$$

4　实验

4.1　标注数据库

为了开展英文 L3 边界预测研究,本文建立了一个英文短语边界标注数据库,共 56 079 句,含有 137 953 个 L3 边界,283 032 个 L1 边界。

该数据库不仅包含文本数据,还包含相匹配的语音数据,因此本文采用了人工与自动方法相配合的方式对数据库进行标注。其中自动方法为基于声学和文法信息共同检测 L3 边界的方法[17],它作为一种辅助手段,能够正确检测出 70% 以上的 L3 边界,大大减少了人工标注的工作量,在提高效率的同时,也较大幅度地提高了标注的一致率。本文所使用的标注数据库的一致率为 83.5%。

为充分利用标注数据库,本文将其随机地分为三部分:a)训练集,含有整个标注数据库 80% 的标注语句,共 44 879 句,用于训练 C4.5 决策树;b)开发集,含有整个标注数据库 10% 的标注语句,共 5 600 句,用于调优 $P_{H/T}$与 $P_{M/\mathrm{ALL}}$的实验参数 μ 和 δ;c)测试集,含有剩余 10% 的标注语句,共 5 600 句,用于验证利用 Viterbi 算法修正结果。以及基于关键词在 L3 中位置分布特性的决策树节点概率优化方法的效果。

为区别主观测试与客观测试,本文在 5 600 句测试集中按领域抽取了 500 句作为主观测试集,其中行业信息播报领域文本(包含导航路径播报、手机信息提示、呼叫中心业务介绍以及电子词典中英互译文本)共 300 句,另有疯狂英语文本 50 句、中国日报文本 50 句、美国之声文本 50 句、新概念英语文本 50 句。客观测试使用测试集中剩余的 5 100 句文本。

4.2 基线系统的效果

本文采用的基线系统为传统的基于决策树的预测方法。其决策树是由 C4.5 决策树算法构建出来的,输出属性为枚举量{L1,L3},输入属性包括:

a)当前词的词频,前一个词的词频,后一个词的词频,前前一个词的词频,后后一个词的词频。

b)当前词的词性,前一个词的词性,后一个词的词性,前前一个词的词性,后后一个词的词性。

c)当前词含有的 L0 个数,前一个词含有的 L0 个数,后一个词含有的 L0 个数。

d)当前 L1 的长度,前一个 L1 的长度,后一个 L1 的长度,前前一个 L1 的长度,后后一个 L1 的长度。

另外需要指出的是,基线系统在使用决策树输出的 P_{L3} 时采用了硬判决的方法。

为衡量 L3 边界的预测效果,本文在客观测试中选用的指标为 F-Score,其与正确率和召回率的关系如下:

$$\text{F-Score} = \frac{2 \times \text{正确率} \times \text{召回率}}{\text{正确率} + \text{召回率}} \tag{11}$$

而在主观测试中选用的指标为不可接受率。其测试方法为:三名实验员分别独立地对 500 句主观测试集的效果进行可接受或不可接受的判断,然后采用投票制确定测试结果,即如果有两人或三人认为某句的预测结果不可接受,则该句测试结果为不可接受,否则视为可接受。需要指出的是,这三名实验员经过专业的标注训练,长期从事韵律标注、浊浊切音、基频修正等相关工作,她们的发音习惯比较相近,对英文语法的掌握程度也基本相当,标注一致性相对较高。

经测试,基线系统在 5 100 句客观测试集上的正确率为 69.8%,召回率为 67.7%,F-Score 为 68.7%。另外,基线系统的平均不可接受率为 22.4%。具体测试结果如表 1 所示。

表 1 基线系统的不可接受率测试

文本领域	总句数	不可接受的句子
行业信息播报	300	52
疯狂英语	50	14
中国日报	50	19
美国之声	50	14
新概念英语	50	13
总计	500	112

4.3 利用 Viterbi 算法修正结果的效果

加入 Viterbi 算法修正结果后,后面的判决结果可能会对前面的误判进行纠正,避免了错误的传递。并且,因 Viterbi 算法从全局出发,整句的预测结果更符合人的朗读习惯,所以预测结果趋于平滑。系统在客观测试集上的正确率提升到74.8%,召回率提升到 70.2%,F-Score 为 72.4%,系统改善较为明显。经主观评测,系统平均不可接受率也降低到了 19.6%。具体测试结果如表 2 所示。

表 2 利用 Viterbi 算法修正结果后不可接受率测试

文本领域	总句数	不可接受的句子
行业信息播报	300	47
疯狂英语	50	12
中国日报	50	16
美国之声	50	13
新概念英语	50	10
总计	500	98

4.4 基于关键词在 L3 中位置分布特性的决策树节点概率优化方法的效果

本文截取了单词头尾倾向性较为明显的 $P_{H/T}$ 区间[0, 0.15]∪[0.85, 1]。$P_{H/T} \in [0, 0.15]$,共有 1 077 个单词,具有较明显的 L3 尾部倾向性;$P_{H/T} \in [0.85, 1]$,共有 9 295 个单词,具有较明显的 L3 头部倾向性。

表 3 和 4 分别为 L3 头部倾向性概率最大和 L3 尾部倾向性最大的 20 个单词。

表 3 L3 头部倾向性概率最大的 20 个单词

单词	$P_{H/T}$	单词	$P_{H/T}$
According	0.999 30	Including	0.993 49
Untitled	0.999 08	And	0.993 38
Thank	0.999 04	We	0.993 22
Let	0.998 59	Which	0.992 57
Despite	0.997 97	If	0.992 18
But	0.997 29	She	0.992 09
Whose	0.997 20	Among	0.991 22
I	0.996 31	Copyright	0.990 91
Joining	0.995 82	He	0.990 59
Although	0.994 60	My	0.990 14

表 4 L3 尾部倾向性概率最大的 20 个单词

单词	$P_{H/T}$	单词	$P_{H/T}$
Clock	0.002 97	Ago	0.001 84
Birthday	0.002 92	Mistake	0.001 32
Minds	0.002 85	Century	0.000 99
Hour	0.002 81	Prohibited	0.000 97
Task	0.002 61	Angeles	0.000 81
Row	0.002 57	York	0.000 79
Sites	0.002 46	Afternoon	0.000 71
Story	0.002 36	Updated	0.000 52
Statement	0.002 29	Translator	0.000 43
Moment	0.002 05	Transcript	0.000 14

在上述 1 077 个具有较明显的 L3 尾部倾向性的单词和 9 295个具有较明显的 L3 头部倾向性的单词中,本文依据 $P_{M/\text{ALL}}$ 值的大小选取了 $P_{M/\text{ALL}} < 0.25$ 的所有单词,共 322 个。表 5 列出了 $P_{M/\text{ALL}}$ 值最小的 10 个单词。

表 5 $P_{M/\text{ALL}}$ 值最小的 10 个单词

单词	$P_{M/\text{ALL}}$	$P_{H/T}$
Transcript	0.003 61	0.000 14
Voiceover	0.005 28	0.002 30
Updated	0.005 74	0.000 52
Prohibited	0.007 83	0.000 97
Huh	0.008 45	0.004 58
Translator	0.011 22	0.000 43
Mailbag	0.011 61	0.001 39
Newswire	0.021 78	0.003 47
Clip	0.028 03	0.000 04
But	0.032 55	0.997 29

在筛选出上述单词后,首先需要确定 $P_{H/T}$ 与 $P_{M/\mathrm{ALL}}$ 的实验参数 μ 与 δ 的值。因不可接受率是主观测试指标,要获取其与 μ、δ 的变化关系,人工工作量巨大,且不一定准确,所以通过构建简单的自动化测试流程,以客观测试指标 F-Score 来选择合适的 μ 与 δ 值。在标注数据库的开发集上,分别获得了 μ、δ 与 F-Score 的变化关系图,如图 3、4 所示。

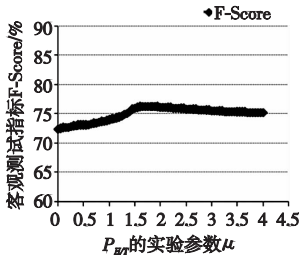

图3　μ与F-Score的变化关系

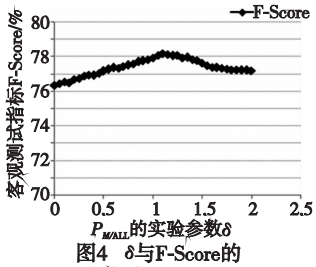

图4　δ与F-Score的变化关系(μ=1.75)

由图 3 可知,当 $\mu=0$ 时,F-Score = 73.1%。随着 μ 值的逐步增大,F-Score 逐步上升,说明 $P_{H/T}$ 的引入是有效的。当 $\mu=1.75$ 时,F-Score = 76.4%,取得最大值。但随着 μ 值的进一步增大,具有头尾倾向性单词的 L3 边界概率 P_{L3} 接近于 0 或 1,而 P_{L3} 接近于 0 时,因权重过低会被 Viterbi 算法忽略视为 L1 边界;P_{L3} 接近于 1 时,因权重过高又往往会被视为 L3 边界的最优解。所以 Viterbi 算法起到的平滑作用越来越小,F-Score 也逐步下降。因此,本文取 $\mu=1.75$。

由图 4 可知,加入 $P_{M/\mathrm{ALL}}$ 的调整后,决策断点增多,因而 L3 边界的召回率得到有效提升,并且因为仅选取了 $P_{M/\mathrm{ALL}}<0.25$ 的 322 个单词,所以正确率的下降是有限的。当 $\delta=1.1$ 时,F-Score = 78.2%,取得最大值。但随着 δ 的进一步增大,$P_{M/\mathrm{ALL}}$ 值较大的单词边界前后新增 L3 边界的概率较大,强行划分出来的 L3 边界越来越多,这导致了正确率快速下降,F-Score 也将随之降低。因此,本文选取 $\delta=1.1$。

当 $\mu=1.75$、$\delta=1.1$ 时,本文在测试集上对加入 $P_{H/T}$ 和 $P_{M/\mathrm{ALL}}$ 调整,且使用 Viterbi 算法修正决策结果的新系统进行了测试,测试结果如表 6 所示。新系统 F-Score 为 77.8%,与开发集上的测试结果基本吻合。平均不可接受率降低到 15.2%,如表 7 所示。

表6　各系统在测试集上的测试结果

测试指标	基线系统	Viterbi 修正系统	新系统
正确率/%	69.8	74.8	80.2
召回率/%	67.7	70.2	75.5
F-Score/%	68.7	72.4	77.8
不可接受率/%	22.4	19.6	15.2

表7　新系统的不可接受率测试

文本领域	总句数	不可接受的句子
行业信息播报	300	39
疯狂英语	50	8
中国日报	50	10
美国之声	50	10
新概念英语	50	9
总计	500	76

5　结束语

本文在基于 C4.5 决策树的 L3 边界预测方法的基础上,引入 L3 条件概率,使用 Viterbi 算法同时优化 L3 边界概率和条件概率,从而实现了决策树结果的软判决,F-Score 由 68.7% 提升到 72.4%,而不可接受率则从 22.4% 降低到 19.6%。在此基础上,本文又提出了基于关键词在 L3 中位置分布特性的决策树节点概率优化方法。通过对较大规模标注语料的统计,筛选出了具有明显的 L3 头尾倾向性的单词,由此在系统中加入了 $P_{H/T}$ 和 $P_{M/\mathrm{ALL}}$ 的调整,F-Score 也由 72.4% 提升到 77.8%,而不可接受率又从 19.6% 降低到 15.2%,充分说明了该方法的有效性。

但因为 L3 训练集在标注过程中存在答案多样化的问题,继续大幅提升 F-Score 并不容易,下一步的工作重点在于细致分析主观评测中发现的不可接受的 L3 预测结果,进一步降低不可接受率。

参考文献:

[1] 陈志刚. 中文语音合成系统中文本分析的若干关键技术研究[D]. 合肥:中国科学技术大学, 2003.

[2] 李剑锋. 韵律层次预测中基于统计模型的机器学习方法研究[D]. 合肥:中国科学技术大学, 2005.

[3] SILVERMAN K E A, BECKMAN M E, PITRELLI J F, *et al*. ToBI: a stan dard for labeling english prosody[C]//Proc of International Conference on Spoken Language Processing. 1992:867-870.

[4] 杨军. ToBI 韵律标注体系及其运用[J]. 现代外语, 2005, **28**(4): 360-366.

[5] LI Wei-jun, YANG Yu-fang. Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences[J]. Neuroscience, 2009, **158**(4): 1416-1425.

[6] 苟恩东,钱揖丽,郭庆,等. 应用二叉树剪枝识别韵律短语边界[J]. 中文信息学报, 2006, **20**(3): 1-5.

[7] 李剑锋,胡国平,王仁华. 基于最大熵模型的韵律短语边界预测[J]. 中文信息学报, 2004, **18**(5): 56-63.

[8] YING Zhi-wei, SHI Xiao-hua. An RNN-based algorithm to detect prosodic phrase for Chinese TTS[C]//Proc of International Conference on Acoustics, Speech, and Signal Processing. 2001: 809-812.

[9] BAILLY G, HOLM B. SFC: a trainable prosodic model[J]. Speech Communication, 2005, **46**(3-4): 348-364.

[10] FUJIO S, SAGISAKA Y, HIGUCHI N. Prediction of prosodic phrase boundaries using stochastic context-free grammar[C]//Proc of the 3rd International Conference on Spoken Language Processing. 1994: 18-22.

[11] READ I, COX S. Stochastic and syntactic techniques for predicting phrase breaks[J]. Computer Speech & Language, 2007, **21**(3): 519-542.

[12] DONG Yuan, ZHOU Tao, DONG Cheng-yu, *et al*. A two-stage prosodic structure generation strategy for mandarin text-to-speech systems[J]. Acta Automatica Sinica, 2010, **36**(11): 1569-1574.

[13] SANDERS E, TAYLOR P. Using statistical models to predict phrase boundaries for speech synthesis[C]//Proc of European Conference on Speech Communication and Technology. 1995: 1811-1814.

[14] 牛正雨,柴佩琪. 基于边界点词性特征统计的韵律短语切分[J]. 中文信息学报, 2001, **15**(5): 19-25.

[15] 赵晟,陶建华,蔡莲红. 基于规则学习的韵律结构预测[J]. 中文信息学报, 2002, **16**(5): 30-37.

[16] 屈俊峰. 决策树的节点属性选择和修剪方法研究[D]. 北京:中国地质大学, 2006.

[17] 吴晓如. 多语种语音合成中的关键技术研究[D]. 合肥:中国科学技术大学, 2009.