第 27 卷 第 6 期
2001 年 11 月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 27, No. 6
Nov., 2001

基于机器学习的普通话韵律规则提取[1)]

朱廷劭 高 文

(中国科学院计算技术研究所 北京 100080)

(哈尔滨工业大学计算机科学系 哈尔滨 150001)

(E-mail: tszhu@cti.com.cn)

摘 要 韵律规则对于语音识别和语音合成研究具有重要意义. 目前的韵律规则大多是根据语言学的研究得出的一些定性的描述. 为了提取出更精确的定量描述的韵律规则,利用聚类分析提取出句子中音节的基频模式,在此基础上使用决策树进行韵律规则的学习,获得了较好的实验结果. 文中首先讨论韵律规则和聚类分析及决策树,然后给出数据预处理技术及所采用的学习算法,最后给出实验结果.

关键词 韵律规则,聚类分析,决策树.

EXTRACTING MANDARIN PROSODIC PATTERNS BY MACHINE LEARNING

ZHU Ting-Shao GAO Wen

(*Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100080)

(*Department of Computer Science, Harbin Institute of Technology, Harbin* 150001)

(E-mail: tszhu@cti.com.cn)

Abstract The pitch models play an important role in speech recognition and synthesis. Most models in using are extracted by linguistics experts, some of which are described qualitatively and of low precision. To acquire more accurate prosodic rules, clustering analysis and decision tree are employed to extract prosodic rules from actual speech, and the result is encouraging. This paper introduces prosodic rules, clustering analysis and decision tree firstly, then each stage of the process is discussed in detail, finally experiments are given.

Key words Prosodic rule, clustering analysis, decision tree.

1 引言

1.1 韵律规则

合成高自然度高清晰度的语音必须有完善的韵律规则,而且韵律规则的总结应该采用

1) 国家自然科学基金(69789301)、国家"八六三"高技术研究发展计划(863-306-ZD03-01-2)和中科院百人计划资助课题. 通信作者高文,中国科学院计算技术研究所.

收稿日期 1999-04-22 收修改稿日期 2000-05-22

定量的方法[1]. 汉语的韵律规则主要表现在音节的时长分布、音高的变化、能量的变化及适当的停顿等几个方面，其中音高和音长的变化对自然度的影响最显著. 国内一些专家学者对这方面进行了研究，但结果比较零散，基本上是一些定性描述.

吴宗济[2]曾对汉语的两字组的声调模式做过很多研究，给出了一些定性的描述. 林茂灿等[3]通过对普通话两字组正常重音的声学分析，得到关于普通话两字组中前后音节的音长和音高的关系描述. 冯隆[4]对北京话中声韵母对时长的影响进行了比较全面的研究，得到了声母时长及单韵母时长在不同情况下的比例关系，并给出了词、句子及说话速度对时长的影响的定性描述. 初敏[5]通过对一个女发音人的音高曲线分析统计归纳出 14 种单音节的音高模式和 22 种两字词声调模式，其中单音节的音高模式包括阴平和上声的声调模式各两种、阳平和轻声的声调模式各三种、去声的模式四种. 每条标准音高曲线都取 30 个样点，经过归整后存入音高模式数据库中. 在进行合成时，按照规则库中的音高曲线修改语音.

通过上面的介绍可以看出，目前的音高变化规律大多是根据语言学的研究得出的一些定性的描述，这在使用计算机进行语音合成时只能提供一些参考，无法在合成过程中直接使用这些规则. 此外，即使采用了定量描述，由于这些规则是由人工进行统计得到，并不能较全面地描述音高变化规则，而且对这些规则的维护和完善很困难. 韵律规则主要描述了音节在不同情况下的音高即基频的变化情况，它对合成语音的自然度和清晰度至关重要，因此如何提取较全面的基频变化规律就成为语音合成技术必须解决的一个问题.

1.2 聚类分析和决策树

聚类分析是一种应用广泛的技术，目标是将数据聚集成类，使得类间的相似性尽量小，而类内的相似性尽量大[6]. 在聚类过程中存在两个基本问题，即如何计算距离以及如何修改聚类中心. 传统的聚类方法主要有 K 均值和自组织映射神经网络等.

决策树方法的起源是概念学习系统 CLS，然后发展到 ID3 方法而为高潮，最后又演化为能处理连续属性的 C4.5[7]. 决策树构造的输入是一组带有类别标记的例子，构造的结果是一棵二叉或多叉树. 二叉树的内部节点(非叶子节点)一般表示为一个逻辑判断，如形式为 $(a_i=v_i)$ 的逻辑判断，其中 a_i 是属性，v_i 是该属性的某个属性值. 树的边是逻辑判断的分支结果，树的内部节点是属性，边是该属性的所有取值，有几个属性值，就有几条边. 树的叶子节点都是类别标记.

为了得到语音基频的变化规律，我们进行了将聚类分析和决策树等机器学习算法应用于汉语语音基频变化规律发现的研究，并取得了较好的实验结果. 这样，一方面使得对大批量语音数据的处理成为可能，另一方面也可以发现较为全面的规则，同时以训练得到的知识形式存储的规律比以列举方式存储的规则占用空间更小，而且结果会更有效.

2 学习过程

语音数据库采用的是语音合成语料库 CoSS-1，它是“八六三”支持项目，由清华大学计算机系、中国科学院声学研究所和社会科学院语言研究所共同完成的. 本文采用了一个女声的全部 126 8 个有调音节和 265 个句子，这些语料尽量涵盖了音段和超音段的音联现象，并且同步录制语音声压波形和声门波阻抗波形. 通过对语料库中句子的每个音节手工进行切

分和基音标注,可以得到句子中每个音节的基频序列.通过音节切分和基频标注后得到的每个音节的基频长短不一,这样的数据无法进行学习,必须对数据进行预处理,使其满足学习算法的要求.图 1 给出了用于韵律规则学习的训练数据获取过程.

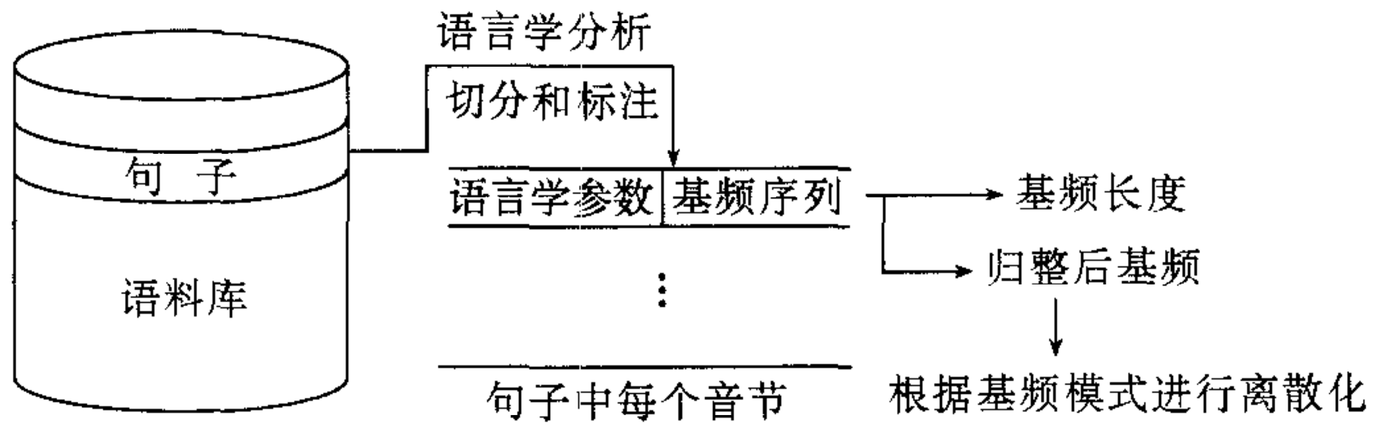


图 1　训练数据获取

首先对每个句子进行语言学分析,得到每个音节的有关语言学参数,同时通过对句子的切分和基音标注得到其基频序列.下面列出了通过语言学分析得到的音节的有关参数:

所在词包含的音节的总数,

音节在词中的序数,

音节所在词的类别,

音节所在词是实词还是虚词,

音节所在词是体词还是谓词,

音节的韵母和声调、前一音节以及后一音节的声调.

对每个音节的基频序列,首先对其进行长度归整,计算所有基频序列长度的均值,以此为统一归整长度;归整后的基频序列经过滑动平均进行平滑处理.通过上述处理后的基频序列与已得到的基频模式进行距离比较,距离最近的基频模式类作为该基频序列的类别.这样,将句子中音节的原始基频序列转化为两个部分:基频模式类别和原始基频序列的长度.为了学习韵律变化规律,建立两个决策树分别学习序列长度及模式类别的规律,同时学习的结果可以以规则形式给出.

3　长度归整和滑动平均

进行训练的每个音节的基频串的长度有很大差别,为了满足聚类分析的要求,必须将长度不同的各个基频串归整到同一长度,即对于长度为 PitchLen 的基频串 Pitch 将其归整为长度 WrapLen. 假设对于归整后的基频,对于 $0 \leqslant j$(WrapLen 中的 j),其在 PitchLen 中的相应位置应该为 $\mathrm{Loc} = j \times \mathrm{PitchLen}/\mathrm{WrapLen}$,为了计算基频曲线中 Loc 位置的数值,首先确定 Loc 附近的四个基频点,然后使用这四个节点通过插值得到 Loc 对应的数值.在得到每个音节的基频序列后,按照所有基频序列长度的均值对每个音节的基频序列进行长度归整,得到具有统一长度的基频序列以用于后续的分析.

滑动平均的主要目的是滤掉那些短时间内抖动幅度比较大的数据,得到能描述发展趋势的曲线[8].计算方法是,首先在序列尾放置 l 天宽度的窗口,求窗口覆盖到的几个点的平均值,然后窗口向前移动一个单位,当窗口移动到序列起始位置时,计算了 $n-l+1$ 个平均值.实践表明如果滑动窗口的长度相对于序列长度足够小的话,平滑处理后的序列很好地保

存了原序列的趋势.

4 利用聚类结果进行基频序列的离散化

为了进行基频变化规律的学习,本文首先对所有基频序列进行聚类分析,得出一些基频序列模式,然后根据这些模式对每个基频序列进行离散化处理.在对所有音节的基频序列进行长度归整和滑动平均处理后,得到具有统一长度而且较为平滑的基频序列,在此基础上可以进行聚类分析.在本文中采用相似度来度量两个基频曲线的距离,计算公式如下:

$$\mathrm{Distance}(A,B)=\sqrt{\sum_{i=1}^{n}((a_i-\bar{a})-(b_i-\bar{b}))^2}+|\bar{a}-\bar{b}|,$$

$$A=(a_1,a_2,\cdots,a_n),\ \bar{a}=\frac{1}{n}\sum_{i=1}^{n}a_i,\ B=(b_1,b_2,\cdots,b_n),\ \bar{b}=\frac{1}{n}\sum_{i=1}^{n}b_i.$$

聚类过程是在 ISODATA[6](Iterative Self-organizing Data,即迭代自组织数据分析方法)算法的基础上进行的,通过凝聚点的合并分解实现了动态聚类.图 2 给出了经过聚类得到的 24 类基频模式.

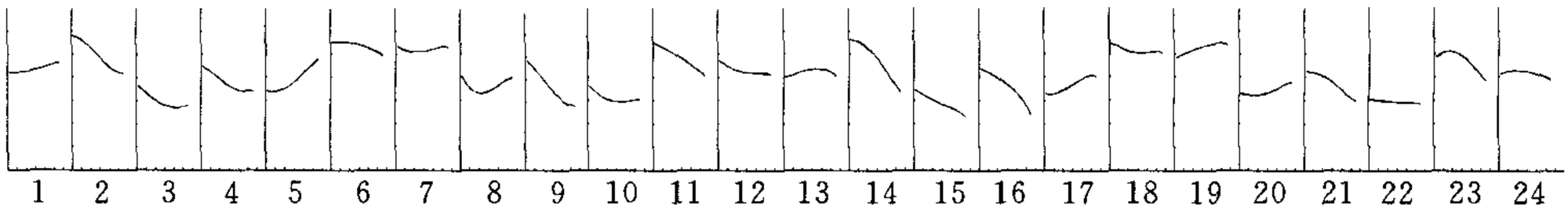


图 2 基频模式(共24类)

离散化就是根据聚类得到的基频模式,将经过上面预处理得到的每个音节的基频序列与所有基频模式计算它们之间的距离,将距离最近的那类赋予该音节的基频序列,这样就可以在更高的层次上对语音变化规律进行学习.

5 利用决策树进行基频变化规律学习

本文采用了 C4.5 进行决策树的建立及规则生成.构造决策树的方法是采用自上而下的递归构造[9],其构造思路是,如果训练例子集合中的所有例子是同类的,则将之作为叶子节点,节点内容即是该类别标记;否则,根据某种策略选择一个属性,按照属性的各个取值,把例子集合划分为若干子集合,使得每个子集上的所有例子在该属性上具有同样的属性值;然后再依次递归处理各个子集.

为了学习基频变化规律,建立两个决策树,分别用于基频类别和长度变化规律的学习,基频序列长度指基频曲线包含的基音周期个数.表 1 和表 2 分别给出了决策树数据的定义及生成的部分规则式,其中声母的编码采用零声母、唇阻声、舌尖前阻声、舌尖阻声、舌尖后阻声、舌面阻声和舌根阻声.韵母的编码采用单韵母、复韵母(前响复韵母、后响复韵母和中响复韵母)以及鼻韵母(舌尖鼻韵母和舌根鼻韵母).声调 0,1,2,3,4 分别表示轻声、阴、阳、上、去声调.词的类别主要是指该词的词性,如形容词、名词、动词等等.规则式的后件与图 2 中基频模式编号相对应.

表 1　基频类别决策树数据的定义

条件属性	所在词包含的音节的总数(len)
	音节在词中的序数(wordno)
	音节所在词的类别(type)
	音节所在词是实词还是虚词(xs)
	音节所在词是体词还是谓词(tw)
	音节、前一音节及后一音节的声调(tone, pretone, posttone)
决策属性	基频类别

表 2　长度决策树数据的定义

条件属性	所在词包含的音节的总数(len)
	音节在词中的序数(wordno)
	音节所在词的类别(type)
	音节所在词是实词还是虚词(xs)
	音节所在词是体词还是谓词(tw)
	音节的韵母和声调(pycon, tone)
	前一音节及后一音节的声调(pretone, posttone)
决策属性	基频序列长度值均分分段离散化

部分基频模式变化规则:

tone＝2 and pretone＝1 → class 7,

type＝1 and tone＝2 and pretone＝4 and posttone＝5 → class 7,

type＝2 and tone＝2 and pretone＝2 and posttone＝3 → class 18,

wordno＝1 and type＝4 and tone＝2 and pretone＝4 and posttone＝3 → class 7.

部分时长规则:

wordno＝1 and type＝14 and pycon＝2 and tone＝2 → class 7,

type＝14 and pycon＝2 and pretone＝3 → class 7,

wordno＝2 and len＝2 and type＝14 and pycon＝2 and pretone＝5 → class 5,

wordno＝1 and type＝22 and pycon＝2 and tone＝2 and posttone＝2 → class 6.

6　实验结果

按照传统的计算调值方法,可以对图 2 给出的典型基频模式进行调值计算.通过对图 2 所显示的聚类得到的基频曲线进行简单的调值计算,可以将 24 类经过聚类得到的基频曲线进行分类如下:

阴平　2, 6, 7, 12, 13, 18, 22, 24;

阳平　1, 5, 17, 19;

上声　8;

去声　4, 9, 14, 16, 23;

半上　3, 10, 15, 20, 21.

出现半上声是由于我们所进行的基频模式聚类是在实际句子音节的基础上进行的,而根据所给出的变调规则:上声＋上声⟶阳平＋上声, 上声＋非上声⟶半上＋非上声,句子中音节在实际输出时,会产生半上声的.此外,通过上面的分析也可以看出,聚类出的基频曲线与公认的声调调值比较吻合.

在通过聚类得到的 24 类典型基频曲线的基础上的韵律变化规则学习也完全符合上述的变调规则.由于篇幅所限,下面仅给出其中的部分规则,其中各参数的解释可参见表 1,规则式的后件为图 2 所列出的基频模式的序号:

type＝2 && tone＝3 && posttone＝3 → class 1,

type＝3 && tone＝3 && posttone＝3 → class 5,

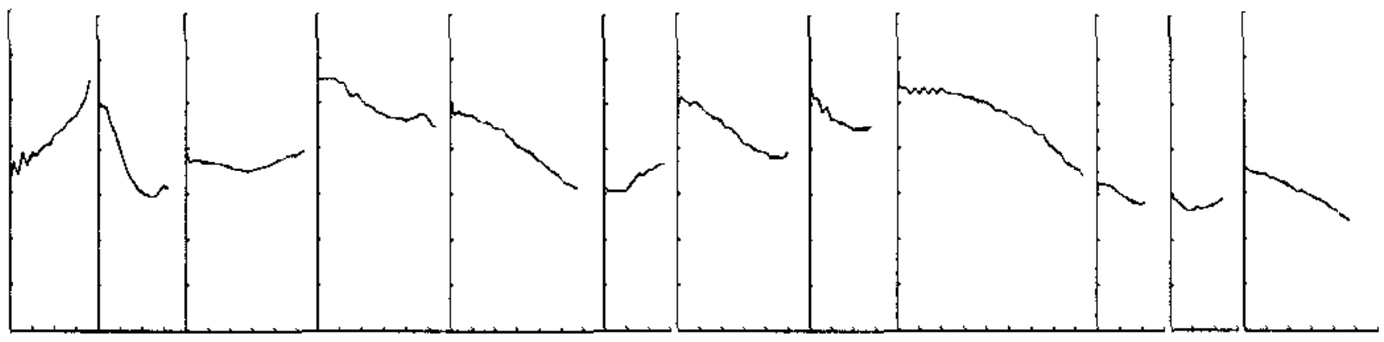
wordno＝1 && tone＝3 && pretone＝3 && posttone＝1 → class 3,

wordno＝2 && type＝18 && tone＝3 && pretone＝4 && posttone＝2 → class 10,

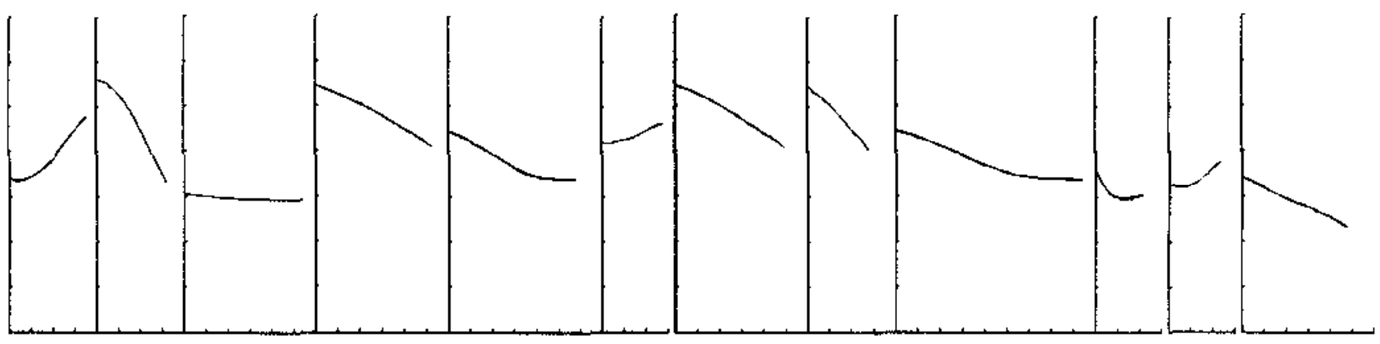
wordno＝1 && len＝2 && type＝4 && tone＝3 && posttone＝4 → class 3.

通过上面的分析可以看出,利用数据挖掘的方法进行韵律规则的学习是完全可行的,通过学习得到的基频曲线符合普通话调类,基频变化规则也完全包含了公认的变调规则,而且可以得到更多更全面的规则.

为了显示学习的结果,本文分别给出了一组基频模式和序列长度的测试结果. 图 3(a)显示的是句子中每个音节的原始基频序列,图 3(b)显示的是根据决策树计算出的类别,按照聚类出的基频模式曲线生成的基频序列. 表 3 给出了一个基频序列长度的测试结果.



(a) 原始基频



(b) 测试结果

图 3 “我请她介绍一下客店的情况”

表 3 基频序列长度测试结果

	从	今	年	四	月	起	封	淀	两	年
原始基频序列长度	44	44	53	38	29	12	34	42	28	41
决策树计算结果	47.5	47.5	52.5	22.5	33.5	12.5	31.5	47.5	27.5	42.5

7 结束语

韵律规则对于语音合成和语音学研究具有重要意义,而目前的韵律规则大多是根据语言学的研究得出的一些定性的描述. 为了有效地进行规则获取及进行更深入的研究,本文利用聚类分析通过对实际句子中音节基频序列的迭代聚类,得出典型的基频模式,并在此基础上采用决策树获得显式的韵律规则. 通过实验,由决策树的分类结果生成的基频序列与测试的基频序列相吻合程度较好,获得了较好的实验结果. 从大量的实际语料中提取韵律规则,使得对语音韵律规律的定量研究成为可能,从而可以采用机器学习算法进行韵律规则的提取,并且由于韵律规则是从实际语料中获取,对提高合成语音的连续度和自然度会有很大作用. 我们关于语音韵律规则学习的研究仍在进行中,希望利用机器学习方法得到更全面更系统的语音变化规律,以提高合成语音的质量.

参 考 文 献

1　刘庆峰,倪晋富,王仁华. 提高合成语音自然度的研究. 见:第三届中国计算机智能接口与智能应用学术会议论文集, 1997. 163～168

2　吴宗济. 普通话语句中的声调变化. 中国语文, 1982, (6):439～449

3　林茂灿,颜景助,孙国华. 北京话两字组正常重音的初步实验. 方言, 1984, (1)

4　冯　隆. 北京话语流中声韵调的时长. 见:北京语音实验录. 北京:北京大学出版社, 1985. 131～195

5　初　敏. 高清晰度高自然度汉语文语转换系统的研究[学位论文]. 北京:中国科学院声学研究所, 1995

6　王碧泉,陈祖荫. 模式识别:理论、方法和应用. 北京:地震出版社, 1989

7　Quinlan J R. Induction of decision trees. *Machine Learning*, 1986, (1):81～106

8　Davood Raflei, Alberto Mendelzon. Similarity-based queries for time series data. In: ACM SIGMOD Conf. on the Management of Data (Sigmod'97), 1997

9　Quinlan J R. Simplifying decision trees. *Internat. Journal of Man-Machine Studies*, 1987, **27**:221～234

朱廷劭　1971 年生,博士. 1993 年、1996 年分别于南京航空航天大学计算机应用专业和中国科学院计算技术研究所获工学学士和工学硕士. 主要研究方向为数据挖掘、文语转换及 Web Mining. 已发表论文 10 余篇.

高　文　1956 年生,教授、博士生导师. 1988 年获哈尔滨工业大学计算机应用博士学位,1991 年获日本东京大学电子学博士学位. 主要研究领域为多媒体数据压缩、图像处理、计算机视觉、多模式接口、人工智能、虚拟现实等.

(上接第 762 页)

参考自动化学报;5. 投稿时请注明文章所属的方向(见征文范围);6. 请说明联系作者的详细通讯地址、电话和电子邮件信箱;7. 因版权等引起的纠纷,作者自负.

三、重要日期

论文截稿时间为 2002 年 3 月 15 日;录用日期:2002 年 5 月 15 日前发录用与否通知.

四、投稿地址

秦皇岛燕山大学电气工程学院(YAC'2002)组委会

邮政编码:066004

联系人:关新平　罗小元

电话:0335－8057041 或 8057034　　传真:0335－8051148

E-mail: xpguan@ysu.edu.cn