

基于互补模型的汉语韵律间断自动检测

倪崇嘉^{1,2} 刘文举² 徐 波²

(山东财政学院统计与数理学院 济南 250014)¹

(中国科学院自动化研究所模式识别国家重点实验室 北京 100190)²

摘 要 自动韵律间断检测和标注对语音理解和语音合成有十分重要的作用。提出了利用声学、词典和语法相关特征的互补模型方法检测汉语韵律间断。该方法具有下列优点:(1)摒弃了声学相关特征和词典、语法相关特征的独立性假设;(2)互补模型方法不仅在特征层上利用当前音节的上下文信息,而且在模型层次上利用了当前音节的上下文信息。在 ASCCD 语料库上验证了该方法能够获得 90.34% 的韵律间断的检测准确率,较基线系统有 6.09% 的提高。

关键词 韵律间断,互补模型,Boosting 分类回归树,条件随机场,神经网络,支持向量机

中图法分类号 TP319 文献标识码 A

Mandarin Prosodic Break Automatic Detection Based on Complementary Model

NI Chong-jia^{1,2} LIU Wen-ju² XU Bo²

(School of Statistics and Mathematics, Shandong University of Finance, Jinan 250014, China)¹

(National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)²

Abstract Automatic prosodic break detection and annotation are important for both speech understanding and natural speech synthesis. We developed complementary model to detect Mandarin prosodic break by using acoustic, lexical and syntactic evidence. Our proposed method has the following advantages: (1) We do not adopt the independent assumption between the acoustic features and the lexical and syntactic features. (2) The complementary models not only in the features of the current syllable but also in the contextual features of the current syllable at model level realizes the complementarities by taking the advantages of each model. We verified our proposed method in a speech corpus of Chinese discourse(ASCCD), where 90.34% prosodic break detection precision rate can be achieved and 6.09% is improved than the baseline.

Keywords Prosodic break, Complementary model, Boosting classification and regression tree(CART), Conditional random fields(CRFs), Neural network(NN), Support vector machine(SVM)

1 引言

韵律是指在语音中的节奏(rhythm)、语调(intonation)、强调(accent)等模式,通常通过基频轮廓的变化、音节时长的加长以及强调、重读等表现出来。韵律在增加语言表达的自然度和可理解程度方面具有十分重要的作用。我们知道,人们在进行语言交流时,相互传递的不仅仅是语言文字信息,语言所表达的韵律信息也是传递的一个重要信息。所以,韵律特征又称为超音段特征。一方面,合理组织韵律可以使说话者能够清楚明白地表达所要表达的信息;另一方面,对韵律信息的正确理解可为听话者更清楚、准确地理解听到的信息提供重要的帮助。韵律间断或边界是将长句子以韵律词、韵律短语等形式分割成为一个个较短的、更方便人们理解或机器处理的小单元。由于在语音合成领域、语音理解等领域的重

要作用,近年来基于统计机器学习的韵律边界建模的研究引起了广泛的注意。

针对英语, M. Ostendorf 提出了基于决策树和 Markov 序列模型对韵律边界和重音进行建模。她采用决策树和 Markov 序列模型去判断音节的间断类型及其是否重读,是在利用语音识别系统提供的音节或者音素强制切分信息的基础上,抽取时长、基频以及能量的特征,构建决策树模型,去预测间断的类型和音节是否重读。通过确定以机器方法判断的间断类型和音节是否重读能否很好地与经过训练的专业人员标注的结果达到一致来验证模型的性能。实验表明,此方法能够达到重音 83% 和边界 77.0% 的正确率^[1]。S. Ananthakrishnan 构建了韵律识别系统,采用耦合隐马尔可夫模型(Coupled Hidden Markov Model, CHMM)在音节和字层次上对重音和韵律边界进行建模。耦合马尔可夫模型能够对异质

到稿日期:2011-01-09 返修日期:2011-03-21 本文受国家自然科学基金(90820303,60675026,90820011),国家高技术研究 863 计划(20060101Z4073,2006AA01Z194),国家重点基础研究发展计划 973 计划(2004CB318105),山东省优秀中青年科学家科研奖励基金项目(BS2009DX020)资助。

倪崇嘉(1979-),博士,讲师,主要研究领域为语音识别、韵律模型, E-mail: cjni@nlpr.ia.ac.cn; 刘文举(1960-),博士,研究员,博士生导师,主要研究领域为语音识别; 徐 波(1966-),博士,研究员,博士生导师,主要研究领域为语音识别、数字内容理解。

的音强特征、音高特征、时长特征等声学特征和同质的语法-韵律特征进行建模,又能很好地抓住话语的语法结构和韵律结构之间的联系。通过在波士顿大学广播新闻语料库(Boston University Radio News Corpus, BURNC)上的实验,说明该模型在音节层次上能够达到75%重音正确率和88%韵律边界正确率。该正确率就是将机器标注的结果与人标注的结果进行对比,如果与人标注的一样,就算正确^[2]。而且S. Ananthkrishnan于2008年又在上述研究的基础上,在最大后验(Maximum A Prior, MAP)框架下对重音和韵律边界进行了建模。通过在波士顿大学广播新闻语料库上的实验,说明该方法能够达到86.75%重音正确率和91.61%韵律边界检测的正确率^[3]。另外, S. Ananthkrishnan仅利用RFC特征和韵律语言模型对韵律的边界和重音建模,在BURNC语料库上,分别能够达到67.7%和56.4%的正确率^[4]。Sridhar在最大熵框架下,利用声学、语法的特征对重音和韵律边界进行建模,其结果是在波士顿大学广播新闻语料库和波士顿Direction语料库(Boston Directions Corpus, BDC)上对字的重音和韵律边界分别能够达到86.0%和93.1%的正确率^[5]。Hasegawa-Johnson利用神经网络和高斯混合模型(GMM)在BURNC语料库上对字的重音和韵律边界建模,分别能够达到84.2%和93.0%的识别正确率^[6]。J. H. Jeon利用多层感知机(MLP)对基频、时长等声学信息建模,同时利用SVM对词典和语信息建模,最后将声学韵律模型和词典、语义模型组合获得最终的模型,据检测, BURNC语料库上对音节的重音(Pitch Accent)、间断(break)和语调短语边界(intonational phrase boundary),分别能够达到89.8%, 91.1%和93.3%的正确率^[7]。

对于汉语,胡伟湘利用分类回归树(classification and regression tree, CART)在ASCCD韵律标注的语料库上,使用声学和文本特征对韵律边界建模,实验证明其也能够达到较好的预测正确率^[8]。邵艳秋利用词性特征、依存特征、词长和距离特征,采用CART进行韵律边界建模,使其在测试数据集上能够达到83.9%的正确率^[9]。

在传统的基于统计机器学习的韵律事件建模时,有如下的不足:(1)声学相关特征和词典、语法相关特征的独立性假设。我们知道,实际上声学相关特征和词典、语法相关特征之间是不独立的。(2)利用一种信息来源来训练模型,如建立声学-韵律模型或词典、语法-韵律模型,然后通过加权的方法获得最后的用于韵律事件检测的模型。在建模时,利用单一的信息源获得的模型在检测韵律事件时可能准确率不低,但其召回率可能很低。针对这些问题,我们提出了基于互补模型的韵律间断自动检测方法,该方法不仅能够很好地解决上述在传统的基于统计机器学习的韵律事件建模中的问题,而且通过该模型进行韵律间断自动检测时还有较高的准确率。

本文第2节介绍实验用到的语料库;第3节介绍韵律间断建模时用到的特征;第4节介绍互补模型方法;第5节给出实验的结果;最后给出结论和讨论。

2 语料库

ASCCD语料库由语篇语料、语音数据和语音学标注信息组成,包括18篇文章,体裁覆盖记叙、议论、通讯、散文等常见

文体。语音数据由精选的10位(5女、5男)北京地区标准普通话发音人录制而成。声音文件采用高质量16kHz采样、16位数据、双声道wav格式存储,库容量约1.5GB。语音学标注信息采用人工标注方法完成,内容包括拼音、声韵母、韵律、重音等。

语音学标注信息采用C-ToBI符号集符号,利用语音学软件Praat标注完成^[10,11]。标注文件中标注4层信息,分别是声、韵母层(DE)、拼音层(PY)、间断指数层(BI)和重音层(ST)。间断指数的数值划分为:0(缺省值,未标出):韵律词内的音节边界;1:韵律词间断;2:次要韵律短语(minor phrase)间断;3:主要韵律短语(major phrase)间断;4:语调组间断。实验中,我们不对不同类型的间断,即韵律词间断、次要韵律短语(minor phrase)间断、主要韵律短语(major phrase)间断和语调组间断做进一步的细分,而是把它们看作一类,统称为韵律间断或韵律边界。表1列出了ASCCD语料库中韵律间断的分布。

表1 ASCCD语料库中韵律间断的分布

总共	非间断	间断
87586	54614	32972
100%	62.35%	37.65%

3 声学特征和词典、语法方面的特征

3.1 声学特征

3.1.1 时长相关特征

语言学理论表明,时长相关特征对于韵律边界的建模有很大的帮助。我们之前的工作也表明,在韵律边界处的很多地方出现停顿;韵律边界处音节的时长比其他位置要长等^[12]。因此,我们计算时长方面的特征,用于韵律边界建模。

对于每一个音节,计算下列时长相关特征:

SilD:音节之后的静音段的长度;

SylDur:音节的时长;

SylDurRatio:当前音节的时长与其后面一个音节时长的比值;

PDur:当前音节的基频与其后面一个音节基频中断的时长。

共4个时长方面的特征。

3.1.2 基频相关特征

韵律边界处基频或音高的变化非常复杂。在韵律边界处,常常伴随着音高高线的提升以及音高曲线的中断。在高级别的韵律边界处,音高中断与韵律级别的高低有着很重要的关系。在低级别的韵律边界处,当静音段不是很明显的时候,音高曲线的中断确实比较明显。

利用Praat语音分析软件中的“To Pitch…”命令,同时设置时间间隔为0.01s,最小基频(Pitch floor)为50Hz,最大基频(Pitch ceiling)为500Hz来抽取基频^[13]。为了减少不同说话人以及同一说话人对基频的影响,用z-score算法对抽取的基频进行正则化。对于每一个音节,不仅提取其统计方面的特征,还计算它的轮廓特征。对于相邻的两个音节,进行其统计特征方面的比较。

设 $f(t)$ 是基频曲线, t 是时间,则 $f(t)$ 可以近似表示为

$$f(t) \approx \sum_{n=0}^M a_n P_n(t) \quad (1)$$

$$式中, P_n(t) = \begin{cases} 1, & n=0 \\ t, & n=1 \\ \frac{2n-1}{n}tP_{n-1}(t) - \frac{n-1}{n}P_{n-2}(t), & n \geq 2 \end{cases} \text{是第 } i$$

阶 Legendre 多项式。在 $f(t)$ 的展开式中, 每一个系数 a_n 表示一定含义。例如, a_0 表示均值, a_1 表示斜率等。

对于每一个音节, 计算如下的音高相关特征:

PthMax: 音节音高的最大值;

PthMin: 音节音高的最小值;

PthRange: 音节音高的范围, 即音节音高的最大值减去最小值;

PthMean: 音节音高的均值;

ConPth $_a_i, i=0, 1, 2, 3, 4, 5$: 音节音高曲线 5 阶 Legendre 展开式的系数;

PDlt: 音节音高曲线的最后一个非零值与该音节之后音节的音高曲线的第一个非零值之间的差值;

TPDlt: 音节音高最大值和其后音节音高最大值之间的差值;

BPDlt: 音节音高最小值和其后音节音高最小值之间的差值;

PMDlt: 音节音高均值和其后音节音高均值之间的差值;

PRatio: 音节音高曲线的最后一个非零值与该音节之后音节的音高曲线的第一个非零值之间的比值。

共 15 个音高方面的特征。

3.1.3 能量相关特征

与计算基频相关特征类似, 计算能量相关特征。首先, 利用 Praat 软件中的“To Intensity...”命令, 并设置最小基频值是 50Hz, 时间间隔是 0.01s, 抽取语音的能量^[13]。

对于每一个音节, 计算如下的能量相关特征:

EngMax: 音节能量的最大值;

EngMin: 音节能量的最小值;

EngRange: 音节能量的范围, 即音节能量的最大值减去最小值;

EngMean: 音节能量的均值;

EngRatio: 音节能量的平均值与其后音节能量平均值之间的比值;

ConEng $_a_i, i=0, 1, 2, 3, 4, 5$: 音节能量曲线 5 阶 Legendre 展开式的系数。

共 11 个能量方面的特征。

3.2 词典、语法方面的特征

由于韵律间断模型在语音合成系统中的作用, 从文本方面获得词典和语法方面的特征进行韵律边界建模被广泛研究。首先, 我们利用 Stanford 分词工具进行分词^[14,15], 然后利用 Stanford PosTag 进行词性标注, 获得 POS (part-of-speech) 信息^[16]。利用这些信息计算词典、语法方面的特征。

对于每一个音节, 首先计算如下特征:

BSeg: 该音节是否是分词以后的词的边界;

Tone: 该音节的声调;

ID: 该音节的 identification;

PosTag: 该音节的词性;

Wen: 该音节所在词的长度;

Hdis: 该音节到所在词开始的音节个数;

Tdis: 该音节到所在词结束的音节个数。

然后, 在当前音节的前一个音节和后一个音节所组成的上下文窗口中计算上述词典和语义方法的特征。共 21 个词典和语法方面的特征。

至此, 我们列举了在汉语韵律边界建模中所用的全部特征, 共 51 个。

4 韵律边界的互补模型

文献[7]的式(2)一式(5)经常被用于韵律事件的建模。我们引用如下:

假设 $P^* = \{p_1^*, p_2^*, \dots, p_n^*\}$ 表示最有可能的韵律事件序列, 则

$$P^* = \arg \max p(P|A, S) \quad (2)$$

$$\approx \arg \max p(P|A)p(P|S) \quad (3)$$

$$\approx \arg \max \prod_{i=1}^n p(p_i | a_i) \lambda p(p_i | \phi(s_i)) \quad (4)$$

$$\approx \arg \max \lambda \sum_{i=1}^n \log(p(p_i | a_i)) + \sum_{i=1}^n \log(p(p_i | \phi(s_i))) \quad (5)$$

式中, $A = \{a_1, a_2, \dots, a_n\}$ 表示声学特征序列, $a_i = (a_i^1, a_i^2, \dots, a_i^k)$ 是音节的声学特征向量, $S = \{s_1, s_2, \dots, s_n\}$ 表示词典和语法特征序列, $\phi(s_i)$ 表示该音节的上下文窗口中的词典和语法特征, $\log(p(p_i | a_i))$ 表示声学-韵律模型得分; $\log(p(p_i | \phi(s_i)))$ 表示语法-韵律模型得分; λ 表示声学-韵律模型和语法-韵律模型之间的权重。式中的声学-韵律模型和语法-韵律模型可以通过统计机器学习的方法获得, 例如决策树、神经网络、支持向量机等。

我们知道, 声学相关特征和词典、语法相关特征不是独立的, 为了简化计算, 减少计算的复杂度, 式(2)将 $p(P|A, S)$ 简化为 $p(P|A)p(P|S)$ 。事实上, 没有必要这样做。

首先, 可以将式(2)变形为

$$\begin{aligned} P^* &= \arg \max p(P|A, S) \\ &= \arg \max (\lambda \cdot p(P|A, S) + (1-\lambda) \cdot p(P|A, S)) \quad (6) \\ &= \arg \max (\lambda \cdot p_1(P|A, S) + (1-\lambda) \cdot p_2(P|A, S)) \quad (7) \end{aligned}$$

式(6)、式(7)仅仅是式(2)的一个变形。在这里, 我们给 $\lambda \cdot p(P|A, S)$ 一个新的符号 $\lambda \cdot p_1(P|A, S)$, 给 $(1-\lambda) \cdot p(P|A, S)$ 一个新的符号 $(1-\lambda) \cdot p_2(P|A, S)$ 。

如果用同样的方法建模 p_1 和 p_2 , 则式(2)是传统的方法, 例如集成机器学习的方法^[17]; 如果不用同样的方法建模 p_1 和 p_2 , 并假设声学特征和词典、语法特征之间独立, 则式(7)就可以写成式(2); 如果不用同样的方法建模 p_1 和 p_2 , 并放弃声学特征和词典、语法特征之间的独立性假设, 则可以得到一种新的韵律事件建模的方法, 我们称之为互补模型方法, 并称 p_1 和 p_2 是一对互补的模型。互补的含义是一种模型能够检测韵律事件, 另外一种方法也可以检测韵律事件, 但是它们的分布不同。虽然检测的结果有重叠, 但它们是不完全重叠, 之间存在互补。该思想来源于系统融合, 不同的系统可以进行融合, 以提高精度。我们知道, 一个韵律事件的出现不仅与其当前音节有关系, 而且与其周围的音节有关系。比如, 一个音节是否被感知重读, 通常通过与它附近的音节对比而

获得。再比如音节后是否有韵律间断问题,也是通过其前后的音节对比来确定的。虽然在计算特征时也考虑到上下文的影响因素,但那是在特征层次上。现在我们在模型层次上,即能够在更高的层次上建模上下文的影响。我们分别用 Boosting 方法和 CRFs 方法建模 p_1 和 p_2 。Boosting 是近年来流行的一种用来提高学习算法精度的集成机器学习算法,是一种提高任意给定学习算法准确度的方法^[18]。条件随机场(Conditional Random Fields, CRFs)是一个无向图模型,通过训练该模型来最大化一个条件概率。CRFs 的一个比较通用的特殊情况是一个线性链结构的图,对应于一个有限状态机,这种结构非常适合用于标注顺序数据^[19]。

5 实验

5.1 实验环境

对于 ASCCD 语料库,从每个人的 75 句话中,随机选择 50 句作为训练集,其余的 25 句作为测试集,在句子层次上训练集与测试集的大小是 2:1,在音节层次上训练集共包含了 58949 个音节,测试集上共包含了 28637 音节。

在模型训练时,对于分类回归树模型,采用 WEKA 的 C4.5 算法和 WEKA 的默认设置训练得到。对于神经网络模型,采用 WEKA 的多层感知器(Multi-layer perception)算法,设置了 1 个隐层,隐层所含节点的个数等于输入特征的一半。对于 Boosting 分类回归树模型,采用 WEKA 的 MultiBoost-AB 作为“强”分类器,C4.5 分类回归树作为“弱”分类器,训练 Boosting 分类回归树模型^[20]。对于支持向量机模型,采用 LibSVM 工具及 RBF 核训练 SVM 模型^[21]。对于 CRFs 模型,采用 CRF++ 0.53 工具训练得到^[22]。

5.2 实验结果与分析

5.2.1 声学韵律模型

利用决策树和神经网络对声学特征分别进行建模,其模型分类结果在表 2 中列出。

表 2 在 ASCCD 语料库上不同的声学韵律模型的性能

		准确率(%)	召回率(%)	F 值
决策树 Decision tree	非间断	79.19	93.73	85.85
	间断	84.77	58.61	69.30
	平均	81.27	80.63	80.95
神经网络 Neural network	非间断	80.99	93.34	86.73
	间断	84.95	63.18	72.46
	平均	82.46	82.09	82.27

从表 2 可以看到,决策树模型和神经网络模型分类效果差别不大,但神经网络模型对于声学特征要好一些。

5.2.2 词典、语法韵律模型

表 3 在 ASCCD 语料库上不同的词典、语法韵律模型的性能

		准确率(%)	召回率(%)	F 值
决策树 Decision tree	非间断	90.60	77.72	83.67
	间断	69.78	86.45	77.22
	平均	82.83	80.98	81.89
支持向量机 SVM	非间断	92.8	90.9	91.8
	间断	85.2	88.1	86.6
	平均	89.9	89.8	89.9
条件随机场 CRFs	非间断	91.44	87.50	89.43
	间断	80.41	86.24	83.22
	平均	87.33	87.03	87.18

对于词典、语法特征,利用 3 种不同的建模方法,即决策树、神经网络和支持向量机对其进行建模。表 3 列出了 3 种模型分类效果。从表 3 可以看到,CRFs 和 SVM 模型获得较好的分类效果。对比表 2 和表 3 可以发现,基于词典、语法特征的模型分类效果要好于相应的基于声学特征的模型。

5.2.3 声学韵律模型和词典、语法韵律模型的结合

表 4 列出了利用式(5)将不同声学韵律模型与词典、语法韵律模型结合以后的分类结果。式(5)中结合的权重 λ 的范围为 0.4~0.9。从表 4 可以看到,利用不同知识源的特征所建立的模型的结合比基于单一信息源的特征所建立的模型分类效果要好。我们应该注意到在表 4 中,CRFs* 和 Boosting CART* 不是通过式(5)加权获得的,而是通过利用全部的特征分别采用 CRFs 和 Boosting CART 方法建模获得模型。

表 4 在 ASCCD 语料库上不同的声学韵律模型和词典、语法韵律模型结合的性能

		准确率(%)	召回率(%)	F 值
神经网络/决策树 NN/Decision tree	非间断	90.44	82.00	86.01
	间断	73.85	85.43	79.22
	平均	84.25	83.28	83.76
神经网络/支持向量机 NN/SVM	非间断	92.76	90.89	91.82
	间断	85.20	88.07	86.61
	平均	89.94	89.84	89.89
神经网络/条件随机场 NN/CRFs	非间断	93.18	90.28	91.71
	间断	84.48	88.90	86.63
	平均	89.93	89.77	89.85
Boosting 分类回归树* Boosting CART*	非间断	91.35	90.10	90.72
	间断	83.74	85.66	84.69
	平均	88.51	88.45	88.48
条件随机场* CRFs*	非间断	92.15	90.26	91.20
	间断	84.17	87.08	85.60
	平均	89.18	89.07	89.12

现在将 Boosting CART* 和 CRFs* 作为一对互补模型,通过式(7)获得新的模型,记作 Boosting CART* + CRFs*,并将该模型用于韵律间断的检测和分类。在式(7)中, λ 的值为 0.5。表 5 列出了互补模型的韵律间断的检测结果。

表 5 在 ASCCD 语料库上互补模型的性能

		准确率(%)	召回率(%)	F 值
基线系统神经网络/ 决策树 (NN/Decision tree)	非间断	90.44	82.00	86.01
	间断	73.85	85.43	79.22
	平均	84.25	83.28	83.76
Boosting 分类回归树+ 条件随机场* Boosting CART* + CRFs*	非间断	93.20	91.05	92.11
	间断	85.53	88.83	87.15
	平均	90.34	90.23	90.28

从表 5 可以看到,Boosting CART* + CRFs* 模型的性能比单个 Boosting CART* 和 CRFs* 要好。CRFs* 在模型层次上提供了上下文属性,与 Boosting CART* 形成了互补的效应。因此,Boosting CART* 和 CRFs* 的结合获得了较好的检测结果。

总之,通过上面的实验可以看到,互补模型方法摒弃了声学和词典、语法特征之间的独立性假设,不仅在特征层上利用了音节的上下文特性,而且在模型层上建模了上下文的性质,最终能够获得较高的识别正确率。

结束语 本文提出了利用互补模型对汉语连续语音中的韵律间断进行自动检测。本方法摒弃了声学和词典语法特征之间的独立性假设,并且不仅在特征层上利用了音节的上下

文的特性,而且在模型层上建模了上下文的性质。通过在ASCCD语料库上的实验说明,本方法能够获得90.34%的检测准确率,比基线系统高了6.09%。今后,我们将进一步探索用于韵律间断检测的特征和建模方法,并利用本互补模型方法对实际的连续语音进行韵律间断的检测。

参考文献

- [1] Wightman C W, Ostendorf M. Automatic labeling of prosodic patterns [J]. IEEE Trans. on Audio and Speech Processing, 1994, 2(4): 469-481
- [2] Ananthakrishnan S, Narayanan S S. An Automatic Prosody Recognizer Using a Coupled Multi-stream Acoustic Model and a Syntactic-Prosodic Language Model[C]// IEEE Proceedings of International Conference on Acoustics, Speech, and Signal Processing, 2005: 269-272
- [3] Ananthakrishnan S, Narayanan S. Automatic prosodic even detection using acoustic, lexical and syntactic evidence [J]. IEEE Trans. on Audio, Speech, and Language Processing, 2008, 16(1): 216-228
- [4] Ananthakrishnan S, Narayanan S. Fine-grained Pitch Accent and Boundary Tone Labeling with Parametric F0 Features[C]// IEEE Proceedings of International Conference on Acoustics, Speech, and Signal Processing, 2008
- [5] Sridhar V K R, et al. Exploiting Acoustic and Syntactic Features for Automatic Prosody Labeling in a Maximum Entropy Framework[J]. IEEE Trans. Audio, Speech, and Language Processing, 2008, 16(4): 797-811
- [6] Hasegawa-Johnson M, Chen K, Cole J, et al. Simultaneous Recognition of Words and Prosody in the Boston University Radio Speech Corpus[J]. Speech Communication, 2005, 46(3/4): 418-439
- [7] Jeon J H, Liu Yang. Automatic Prosodic Events Detection Using Syllable-based Acoustic and Syntactic Features[C]// IEEE Proceedings of International Conference on Acoustics, Speech, and Signal Processing, 2009: 4565-4568
- [8] Hu W, Huang T, Xu B. Study on Prosodic Boundary Location in Chinese Mandarin[C]// IEEE Proceedings of International Con-

ference on Acoustics, Speech, and Signal Processing. Vol. 1. 2002: I-501-I-504

- [9] 邵艳秋, 韩纪庆, 刘挺, 等. 自然风格言语的汉语句重音自动判别研究[J]. 声学学报, 2006, 31(3): 203-210
- [10] Chen Xiao-xia, Li Ai-jun, Sun Guo-hua, et al. An Application of SAMPA-C in Standard Chinese[C]// Proc. ICSLP, Beijing, 2000
- [11] Li Ai-jun. Chinese Prosody and Prosodic Labeling of Spontaneous Speech[C]// Speech Prosody 2002. Aix-en-Provence
- [12] 倪崇嘉, 刘文举, 徐波. 汉语韵律短语的时长与音高研究[J]. 中文信息学报, 2009, 23(4): 82-87
- [13] Boersma P, Weenik D. Praat: doing phonetics by computer[CP/OL]. <http://www.praat.org>
- [14] Tseng H, Chang P, Andrew G, et al. A Conditional Random Field Word Segmenter[C]// Fourth SIGHAN Workshop on Chinese Language Processing, 2005
- [15] Chang Pi-chuan, Galley M, Manning C. Optimizing Chinese Word Segmentation for Machine Translation Performance[C]// ACL Third Workshop on Statistical Machine Translation, 2008
- [16] The Stanford Postagger [CP/OL]. <http://nlp.stanford.edu/software/tagger.shtml>
- [17] Sun Xue-jing. Pitch Accent Prediction Using Ensemble Machine Learning[C]// Proceedings of the International Conference on Spoken Language Processing, 2002: 963-956
- [18] Freund Y, Schapire R E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting[J]. Journal of Computer and System Sciences, 1997, 55(1): 119-139
- [19] Lafferty J D, McCallum A, Pereira F C N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]// Proc. International Conference on Machine Learning (ICML'01). 2001: 282-289
- [20] Garner S. Weka: the waikato environment for knowledge analysis[C]// Proc. the New Zealand Computer Science Research Students Conf. 1995
- [21] LIBSVM—A Library for Support Vector Machines [CP/OL]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [22] CRF++: Yet Another CRF toolkit[CP]. <http://crfpp.sourceforge.net/>

(上接第 231 页)

- [4] Ahuja R K, Orlin J B. Distance-directed augmenting path algorithms for the maximum flow problem[J]. Naval Research Logistics Quarterly, 1991, 38: 413-430
- [5] Glodberg A V, Rao S. Beyond the flow decomposition barrier [J]. J Assoc Comput Mach, 1988, 45(5): 783-797
- [6] 邹豪思, 王志远. 网络最大流的矩阵算法[J]. 内蒙古大学学报, 2001, 32(4): 465-469

- [7] 孙泽宇, 丁国强, 程志谦. 网络最大流求解算法的研究[J]. 微机计算机信息, 2010(3)
- [8] 党耀国, 刘思峰, 方志耕. 网络最大流的割集矩阵算法[J]. 系统工程理论与实践, 2003, 9: 125-128
- [9] Tutte W T. Graph Theory [M]. Cambridge: Cambridge university press, 2001
- [10] 高随祥. 图论与网络流理论[M]. 北京: 高等教育出版社, 2009