

# 基于规则学习的韵律结构预测 \*

赵晟 陶建华 蔡莲红

(清华大学计算机系 北京 100084)

**摘要:**韵律结构的分析和预测作为提高语音合成系统自然度的一个重要核心组成, 日益受到重视。本文提出了一种基于规则学习的汉语韵律结构预测方法, 该方法从人工韵律标注的语料库中抽取语言学特征和两级韵律结构标记, 构建了实例数据库(example database), 再利用规则学习(rule learning)算法从实例中自动归纳韵律短语预测规则。本文通过大量的实验挑选出对于汉语韵律结构预测最有效的特征, 采用和比较了两种典型的规则学习算法。同时, 对于实验结果给出了较为系统的评价参数。实践表明, 规则学习算法用于韵律结构预测达到了90%以上的正确率, 优于目前其他方法的结果, 是一种行之有效的办法。

**关键词:** 韵律结构预测; 规则学习; 韵律词; 韵律短语; 转换规则

**中图分类号:** TP391

## Rule-learning based prosodic structure prediction

ZHAO Sheng TAO Jianhua CAI Lianhong

(Department of Computer Science and Technology of Tsinghua University, Beijing, 100084)

**Abstract** Prosodic structure prediction plays a more and more important role in modern TTS systems. In this paper, a rule-learning based approach is proposed to predict prosodic structure from unrestricted Chinese text. Firstly, a speech corpus is collected, whose text is automatically segmented and tagged and further labeled with two-level prosodic structure and syntactic phrase boundaries. Secondly, features related to prosodic structure are extracted with the corresponding boundary types to establish an example database. Lastly, rule-learning algorithms are applied on the database to induce prediction rules by machine. Various experiments have been conducted to select the best features. Two typical learning algorithms (C4.5 and Transformation-based learning) are experimented and compared with other methods. The paper also suggests general evaluation parameters for prosodic structure prediction. The experiments show that the rule-learning approach can achieve a better accuracy rate of 90% than those of the others. Thus it is justified as an effective way to prosodic structure prediction.

**Keyword** prosodic phrase prediction, rule learning, transformation rules

## 一、引言

文语转换系统(TTS)的目标是让计算机能够像人一样说出具有高度自然和智能的语音。这就需要在系统前端建立从文本到韵律特征(如停顿,重音和语调)的预测模型以指导语音合成。两种常见的合成方法如参数合成和语料库合成都利用韵律特征预测的结果提高合成语音的自然度。韵律特征包括韵律结构和重音等多方面的内容。本文主要考虑的是韵律结构的预测问题。研究表明,人们的话语在韵律上具有层级结构,包括语音词(phonological phrase),语调短语(intonational phrase)等层次<sup>[1]</sup>,这种结构称之为韵律结构。而TTS文本分析的输出结果一般是某种语法结构,与韵律结构不完全相同。韵律结构预测就是要找

---

收稿日期: 2002-03-06

作者: 赵晟,男,1977年生,硕士生,主要研究方向为语音合成和机器学习;陶建华,男,1972年生,博士,主要研究领域为韵律建模的研究和语音合成;蔡莲红,女,1945年生,教授,主要研究领域为多媒体技术和汉语TTS。

\*本文的工作得到了国家863项目的支持(编号:2001AA114072)

到一种有效的办法将语法结构映射为韵律结构。

英语韵律结构预测的工作开始的较早。CART(Classification and Regression Tree)方法曾用于英语的韵律短语( prosodic phrase )预测<sup>[2]</sup>。爱丁堡大学的 Paul Taylor<sup>[3]</sup>等人使用 HMM 对韵律短语建模,正确率达 86.6%。汉语这方面的工作进行的较晚,但已提出不少方法。应宏<sup>[4]</sup>等人总结了结构助词与韵律短语的关系。Intel 中国研究院采用基于反馈神经网络(RNN)的学习方法预测韵律短语达到了 84.7%的正确率<sup>[5]</sup>。微软中国研究院使用基于二元词性文法(bigram)的统计方法预测汉语的韵律词边界<sup>[6]</sup>,给出的召回率和精确率都在 85%以上。此外,也有人采用三元词性文法(trigram)做韵律短语预测<sup>[7]</sup>,但报告的召回率为 60%,精确率为 80%。

上述方法各有优缺点。HMM 和词性文法是基于统计模型的算法,一般要采用大量的熟语料进行训练,而且容易受数据稀疏(sparse data)的影响。RNN 是一种神经网络的学习算法,推广和容噪能力强,但学习得到的网络权值不易于解释和理解。而 CART 方法中问题集的定义需要较多的专家经验,实现起来比较费时。还需要指出的是,不同研究者实验所用的语料在大小和构成上有很大差别,对结果的评价方法也不一致,使得实验结果之间可比性不强。

本文采取了基于规则学习的方法来同时预测汉语的两级韵律结构:韵律词和韵律短语。基于规则的系统比较简单,便于理解和实现,但专家规则往往难于人工归纳和评价。鉴于自动规则学习在机器学习领域中已有较多的研究,本文分别尝试了用 C4.5 归纳学习和转换规则学习(TBL)算法从实例中自动归纳韵律结构预测规则,并把规则学习方法与 bigram 和 RNN 方法在同一语料上作了比较。本文的组织结构如下,第二节简要介绍和比较了两种规则学习算法,第三节描述了韵律结构预测问题的解决思路和评价方法。第四节具体报告了有关实验,包括语料的准备,特征选择和实验细节等等。第五节给出了实验结果。

## 二、规则学习算法

研究如何从实例数据库中归纳学习得到规则表示是机器学习(machine learning)的一个主要内容。实例数据库中的每个实例是一个广义特征(或属性)向量 $\langle a_1, a_2, a_3, \dots, a_n \rangle$ ,规则就是特征之间的条件关系,比如 *if  $a_j=A$  then  $a_n = C$* 。规则作为一种知识表示相比神经网络等方法更容易理解,也是一种实现“智能”的有效方法。机器学习的规则经过人工修订和增删可以提高它的预测性能。

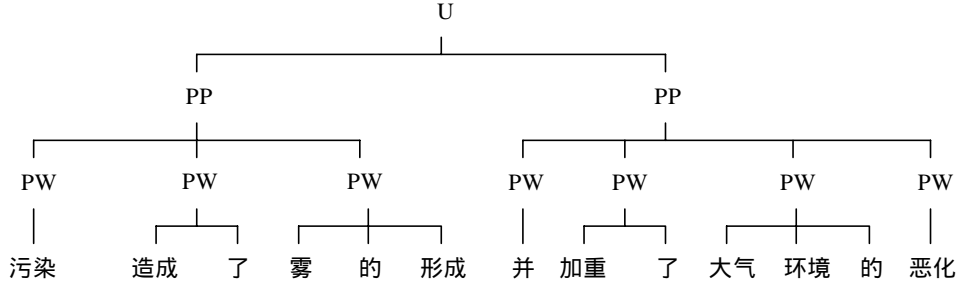
C4.5 归纳学习<sup>[8]</sup>是一种典型的规则学习算法。该算法具有的优点是能够处理连续和离散的属性,对噪声不太敏感,不易受数据稀疏的影响。转换规则学习<sup>[9]</sup>(transformation based learning, TBL)最早用于基于规则的英语词性标注,并获得了和统计方法相当的标注正确率。其基本思想是从问题的一个初始解决方案出发,应用一系列转换规则修改初始的结果,使之逐渐趋于正确。转换规则的学习是一个贪婪搜索过程。首先定义了所有可能的转换规则模板,每一步训练从实例化的模板中选出具有最高有效得分(Score)的规则,如此下去,直到找不出新的有效规则。

这两种学习算法都是有指导的学习,可以用于从大量实例中自动挖掘有效的规则,并给出规则的评价参数。但这两者也有显著的不同。一方面,C4.5 归纳学习是完全自动的学习算法,所需要做的是选取有用的特征,构建实例数据库供它学习。而 TBL 学习需要人工定义转换规则模板,这些模板将直接决定规则学习的有效性。这样看来,TBL 学习只能看作是半自动的。另一方面,C4.5 学习采用的是分治的策略,速度较快。而 TBL 的贪婪学习过程比较慢。考虑到这些差别,本文的研究方法是先用 C4.5 归纳学习做实验,选取最有用的特征来归纳 C4.5 规则,然后利用这些规则来帮助设计转换规则模板。

## 三、韵律结构的预测

### 3.1 韵律结构预测方法

语音学和语言学研究表明，语音材料在韵律上具有树状层次结构。韵律层级(prosodic hierarchy)主要分为三个基本层次：韵律词、韵律短语和语调短语<sup>[10]</sup>。对于语音合成来说，语调短语往往被认为是由标点符号分开的，比较容易识别。因而本文主要研究的是韵律词和韵律短语的预测问题。图 1 是两级韵律层级结构的一个例子。



U：语调短语, PP：韵律短语, PW：韵律词

图 1: 两级韵律结构

这样的树状结构中，语法词是最小的单位，位于叶子节点。韵律词的边界一定是语法词的边界，而韵律短语的边界一定是韵律词的边界。

假设有一串语法词  $W = w_1, w_2, \dots, w_n$ ，两个相邻语法词  $w_i$  和  $w_{i+1}$  之间的边界表示为  $\langle w_i - w_{i+1} \rangle$ 。边界有三种类型： $B_0$  ( $w_i$  和  $w_{i+1}$  在一个韵律词里)， $B_1$  (它们在一个韵律短语里，但不在一个韵律词里)， $B_2$  (它们在不同的韵律短语里)。由此韵律结构的预测问题就转化为对边界类型的分类问题。假设边界类型是由边界所处的语言学信息决定的。本文利用了一个具有韵律结构标注的语料库，提取语料中每个边界的类型和其上下文相关的语言学特征，组成实例数据库，最后利用规则学习算法自动归纳出对边界类型的预测规则。

### 3.2 预测结果评价

韵律结构预测作为对边界类型分类的问题，对它的结果评价不能简单的只考虑某一类的分类情况，比如只考虑韵律短语的召回率和精确率等，而应该全面地考虑每个类别的分类情况以及整体的正确率。本文则把学习到的规则应用在测试集上，对每个边界得到一个预测的边界类型。和人工标注的类型比较，得到一个分类混淆矩阵(confusion matrix)如下：

人工标注的类型	预测得到的类型		
	$B_0$	$B_1$	$B_2$
$B_0$	$C_{00}$	$C_{01}$	$C_{02}$
$B_1$	$C_{10}$	$C_{11}$	$C_{12}$
$B_2$	$C_{20}$	$C_{21}$	$C_{22}$

$C_{ij}$  表示人工标注类型为  $B_i$  的边界被预测为  $B_j$  的数目，由此定义评价参数如下：

$$Recall_i = \frac{C_{ii}}{C_{i0} + C_{i1} + C_{i2}} \quad (i = 0,1,2) \quad Precision_i = \frac{C_{ii}}{C_{0i} + C_{1i} + C_{2i}} \quad (i = 0,1,2)$$

$$FMeasure_i = 2 * Recall_i * Precision_i / (Recall_i + Precision_i) \quad (i = 0,1,2)$$

$$Accuracy_1 = \sum_{i=0}^2 C_{ii} / \sum_{j=0}^2 \sum_{i=0}^2 C_{ij} \quad Accuracy_2 = (\sum_{j=1}^2 \sum_{i=1}^2 C_{ij} + C_{00}) / \sum_{j=0}^2 \sum_{i=0}^2 C_{ij}$$

其中  $Recall_i$  是边界类型  $B_i$  的召回率,  $Precision_i$  是对应的精确率。  $FMeasure_i$ <sup>[11]</sup> 是同一类型  $B_i$  的召回率和精确率的综合参数。  $Accuracy_1$  是边界类型的总预测正确率。如果把  $B_2$  都当作  $B_1$  处理, 即预测一级韵律结构, 总正确率可估算为  $Accuracy_2$ 。其中两个总正确率是比较关键的参数。

## 四、实验

### 4.1 语料的准备

实验采用的语料共有 3167 个句子, 文本全部来自人民日报, 由新闻播音员用正常语速朗读。韵律结构的标注由人工完成。两名有经验的标注人员通过听录音给文本标上两级韵律边界信息, 标注的一致率达到 98.5%。语法结构的标注用程序自动完成, 主要对语料文本进行了自动分词和词性标注, 此外手工标注了 500 个句子的语法短语边界。

该语料覆盖了汉语所有的有调音节和词性, 总共有 56446 个汉字, 37669 个语法词(语法词长为 1-4 个汉字)。韵律词边界共有 16194 个, 而韵律短语边界有 7231 个。表 1 列出了这些语法或韵律单元的平均长度。图 2 则反映了语料中韵律短语长度的分布。

	平均长度(汉字数)
语法词	1.5
韵律词	2.4
韵律短语	7.8
句子	17.0

表 1 平均长度信息

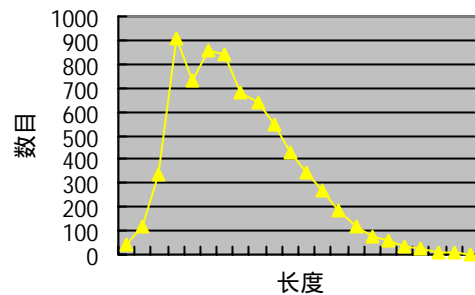


图 2.韵律短语的长度分布

假定所有的语法词边界都是韵律词边界,  $Accuracy_1$  为 42.99%; 如果都是韵律短语边界,  $Accuracy_1$  仅为 19.20%。因而, 选取更有效的语言学特征则成为提高韵律结构预测正确率的重要保证。

### 4.2 语言学特征

语言学特征来源于韵律边界所在的上下文语言学信息, 本节试图列举出与韵律结构预测有关的所有可能的特征, 为特征选择实验做准备。特征按不同类型分为: 语音学特征, 词法特征, 语法特征和其它特征。具体描述如下:

#### I. 语音学特征

汉语是单音节的有调语言。语音学研究表明在连续的汉语语流中因上下文或韵律结构变化存在音变现象<sup>[10]</sup>, 如儿化、轻声或变调等, 其语音学信息与韵律结构有所关联。汉语中有调音节有 1278 个, 而声韵母只有 61 个。为避免以音节作为特征造成特征空间过大, 采用的语音学特征是边界周围的汉字的声韵母  $SYIF$  和所在拼音的声调  $TONE$ 。(  $SYIF$  ,  $TONE$  是特征的代号, 以下同 )

#### II. 词法特征

词语是语法结构中的基本单位, 它与韵律结构有着必然联系, 但每个词语又有自身的

特点，具有各自的出现频度和长度，对韵律结构预测的作用不能一概而论。本文选取词语 *WORD*，词语长度 *WLEN* 和词语频度作为词法特征。

### III. 语法特征

词性信息 *POS* 是最容易得到的语法信息。边界预测模型大多采用了词性作为特征。不同的词性集合对韵律边界的预测能力也有差别。文献<sup>[12]</sup>指出，较深层语法信息对于提高预测正确率有较大作用。利用语法短语的信息给句子中的每个词都标上短语标记 *SPT*，指出它是否在某种短语内部或头部。这些 *SPT* 标记在实验中作为较高层的语法特征使用。表 2 列出了部分实际使用的 *SPT* 标记。

### IV. 其他特征

从语料的统计结果看，韵律词和韵律短语都有一定的长度限制，它们所在句子的长度 *SLENC*(以字符计)或 *SLENW*(以词数计)被用作长度特征。基于 HMM 的统计模型假定韵律边界序列符合 Markov 假设，前面的边界类型对当前的边界类型会有影响。在训练语料中，当前边界所在位置前后的边界类型都是已知的，而在测试时，前后类型都是未知的。因此只有把前面预测的类型作为特征使用。

B-NP	名词短语左边界
I-NP	名词短语内部
B-VP	动词短语左边界
I-VP	动词短语内部
B-PP	介词短语左边界
O	不属于任何短语

表 2 短语标记

边界 特征	<污染—造成>	<形成—并>
<i>SYIF_0</i>	an	eng
<i>SYIF_1</i>	z	b
<i>TONE_0</i>	3	2
<i>TONE_1</i>	4	4
<i>WORD_0</i>	污染	形成
<i>WORD_1</i>	造成	并
<i>POS_0</i>	vn	v
<i>POS_1</i>	v	c
<i>POS_-1</i>	w	u
<i>SPT_0</i>	B_NP	I_NP
<i>BTYPE_0</i>	$B_1$	$B_2$

表 3 实例数据库条目

### 4.3 实例数据库

对于训练语料中的每个边界，提取与它相关的特征，组成实例数据库。数据库的每个条目就对应于语料中的一个边界。表 3 列出了图 1 中的两个边界对应的数据库条目。第一列是特征名称，后面的列是特征的取值。特征名称是由上节提到的特征代号加上一个数字。该数字指示该特征来自于边界附近的哪个词。举例来说，*POS\_0* 指的是边界左边第一个词的词性，*POS\_-1* 是指边界左边第二个词的词性，而 *POS\_1* 是右边第一个词的词性...依次类推。数字的范围由一个滑动窗口决定。*BTYPE\_0* 是指当前边界的类型，也就是要预测的韵律边界类型。

### 4.4 特征选择实验

数据库中的实例都有很多的特征，通过研究不同特征对韵律结构预测的作用，可以简化特征集合，提高预测的准确率。按照特征类型选出特征的组合，用 C4.5 算法在数据库上归纳学习，比较规则在测试集上的表现，从而确定对韵律预测有效的特征组合。

#### I. 原型系统 (实验 1)

原型系统采用词性 *POS* 作为特征。具体来说，对每个边界有两个特征，即 *POS\_0* 和 *POS\_1*。词性标记集为自动标注程序的基本词性集合 *BSET*，总共 30 个词性标记。原型系统

的结果用作特征选择实验的基准。其他实验的结果都和原型系统比较以确定所使用特征的有效性。

### II. 词性窗口大小 (实验 2-9)

假设窗口大小为  $L+R$ ，表示与边界左边  $L$  个词和右边  $R$  个词有关的特征都要考虑。让  $L$  和  $R$  值在一定范围内自由变化，通过实验来选择最优的窗口大小和位置。对词性特征  $POS$  进行了实验，所用特征简单表示为  $POS\{-L+1, R\}$ 。当然对其他类型的特征也可做类似的实验。

### III. 词性集合 (实验 10-11)

不同词性集合的韵律预测能力也有所不同。除了基本词性集  $BSET$  外，设计了另外两种词性集合并用它们作了实验。第一种是在基本词性集基础上，把出现频次最高的 100 个词本身作为一个词类，构成了扩大的词性集合  $LSET$ ，它共有 130 个标记。第二种是对基本词性集合进行聚类处理。对于每个词性，统计它前后边界类型的分布概率，组成代表该词性的一个 6 维特征向量。对这些向量进行层次聚类，得到了一个大小为 10 的词性集合  $CSET$ 。使用词性特征  $POS$  和大小为  $1+1$  的窗口进行了实验。

### IV. 其他实验 (实验 12-18)

实验 12-13，词语长度  $WLEN$  和句子长度  $SLEN$  加入到原型系统中，研究长度信息的预测作用。实验 14 是关于声韵母  $SYIF$  和声调  $TONE$  的。实验 15 考虑语法短语边界  $SPT$  特征的作用。实验 16 和 17 用前面预测的边界类型来帮助预测当前类型。实验 18 使用了  $WORD$  特征。

这些实验的结果都列在表 4 中。其中  $R_i$  表示  $Recall_i$ ， $P_i$  表示  $Precision_i$ ， $A_1$  表示  $Accuracy_1$ ， $A_2$  表示  $Accuracy_2$ 。

No.	特征	词性集	$R_0$	$R_1$	$R_2$	$P_0$	$P_1$	$P_2$	$A_1$	$A_2$
1	$POS\{0,1\}$	$BSET$	0.64	0.80	0.66	0.75	0.65	0.90	0.72	0.79
2	$POS\{0,0\}$	$BSET$	0.66	0.57	0.08	0.50	0.50	0.51	0.50	0.64
3	$POS\{-1,0\}$	$BSET$	0.54	0.66	0.29	0.57	0.54	0.50	0.54	0.68
4	$POS\{0,2\}$	$BSET$	0.69	0.77	0.66	0.72	0.67	0.89	0.72	0.79
5	$POS\{-1,1\}$	$BSET$	0.71	0.74	0.70	0.72	0.68	0.83	0.72	0.79
6	$POS\{-1,2\}$	$BSET$	0.70	0.73	0.71	0.72	0.68	0.80	0.71	0.79
7	$POS\{-2,1\}$	$BSET$	0.72	0.73	0.68	0.71	0.67	0.83	0.71	0.79
8	$POS\{-2,2\}$	$BSET$	0.69	0.73	0.70	0.72	0.67	0.80	0.71	0.79
9	$POS\{-3,3\}$	$BSET$	0.71	0.72	0.70	0.71	0.68	0.79	0.71	0.79
10	$POS\{0,1\}$	$LSET$	0.64	0.84	0.69	0.81	0.66	0.86	0.74	0.81
11	$POS\{0,1\}$	$CSET$	0.70	0.70	0.61	0.65	0.64	0.91	0.68	0.75
12	$POS\{0,1\},WLEN\{0,1\}$	$BSET$	0.83	0.80	0.68	0.80	0.75	0.85	0.79	0.86
13	$POS\{0,1\},WLEN\{0,1\},SLEN$	$BSET$	0.86	0.75	0.71	0.79	0.78	0.77	0.78	0.87
14	$POS\{0,1\},TONE,SYIF$	$BSET$	0.70	0.77	0.65	0.72	0.67	0.89	0.72	0.79
15	$POS\{0,1\},SPT\{0,1\}$	$BSET$	0.65	0.81	0.66	0.75	0.66	0.90	0.73	0.79
16	$POS\{0,1\},BTYPE_{-1}$	$BSET$	0.75	0.78	0.67	0.75	0.70	0.89	0.75	0.82
17	$POS\{0,1\},BTYPE_{\{-1,-2\}}$	$BSET$	0.75	0.75	0.72	0.75	0.71	0.81	0.74	0.82
18	$POS\{0,1\},WORD\{0,1\}$	$BSET$	0.54	0.91	0.57	0.8	0.6	0.99	0.70	0.78

表 4 特征选择实验

## 4.5 特征选择结论和 C4.5 学习实验

根据特征选择实验的结果有如下结论：(1) 词性特征 *POS* 是最基本也是最有用的特征。一般大小为  $2+1$  的窗口已经足够了。再扩大窗口，将使学习时间变长，而正确率的提高有限。(2) 较大的词性集合预测能力较强。因为词性集合变小，将导致特征空间也变小，也不利于对训练样本的分类。(3) 长度信息对预测正确率的提高很有帮助。(4) 语音学特征对预测没有明显作用，这与前面设想的不同。使用语法短语信息后，正确率确有提高，但不是很显著。可能的原因是标注语法短语的句子数目太少。(5) 使用前面预测的边界类型 *BTYPE* 作为特征提高了预测正确率，这体现了前后边界类型的相关性。(6) 词语 *WORD* 直接作为特征，没有明显改进，反而导致某些类别的召回率和精确率下降。

把那些有效的特征组合起来，用 C4.5 算法学习得到最终的预测规则。这些特征是词长  $WLEN\{-1, 1\}$ ，边界类型  $BTYPE\{-1\}$ ，句子长  $SLEN$  和词性集合  $LSET$  上的词性  $POS\{-1, 1\}$ 。

#### 4.6 转换规则学习实验

C4.5 算法选择得到的有效特征集合同样可用于 TBL 学习。转换模板的设计将主要采用这些特征。此外 C4.5 算法习得的规则也用来帮助设计转换规则的模板。比如，假定 C4.5 规则中有一条： $if (POS_{-1} == \text{的} \text{ and } POS_0 == n \text{ and } POS_1 == d) \text{ then } BTYPE_0 = B_2$ ，该规则的正确率很高，那么就应该有模板能够包含这条规则。下面列出了部分实验中使用的转换规则模板的简写形式。

$POS_0 POS_1 \Rightarrow BTYPE_0$   
 $POS_{-1} POS_0 POS_1 \Rightarrow BTYPE_0$   
 $BTYPE_0 POS_0 POS_1 \Rightarrow BTYPE_0$   
 $BTYPE_0 POS_{-1} POS_0 POS_1 \Rightarrow BTYPE_0$   
 $BTYPE_{-1} POS_0 POS_1 \Rightarrow BTYPE_0$   
 $BTYPE_{-1} POS_{-1} POS_0 POS_1 \Rightarrow BTYPE_0$

每一条规则模板分为两部分，左边是规则的条件部分，列出了若干与边界预测有关的特征。右边是规则满足后执行的动作，即对边界类型  $BTYPE_0$  的修改。例如， $POS_{-1} POS_0 POS_1 \Rightarrow BTYPE_0$  写成条件表达式就是  $if (POS_{-1} == X \text{ and } POS_0 == Y \text{ and } POS_1 == Z) \text{ then } BTYPE_0 = B$ ，其中  $X, Y, Z, B$  是模板变量。令  $X=\text{的}, Y=n, Z=d, B=B_2$ ，把模板实例化为具体的规则，该规则正是上面列出的那条 C4.5 规则。转换学习过程需要对问题给出一个初步的解决方案，然后应用转换规则对初始结果进行修正以提高正确率。对于边界预测问题，采用的初始化方案是很简单的，即将所有的边界类型都设为  $B_1$ 。学习算法在模板变量所有可能的取值范围内穷举搜索，每一次循环都试图找到一条得分最高的规则，使得该规则正确预测的数目减去错误预测的数目之差为最大。

## 五、实验结果与评测

本文采用交叉检验的方法在语料上对 C4.5 算法和转换学习算法 TBL 进行了实验。把整个语料分为 5 份，每一次留出一份用作测试，其余的用来训练，通过对 5 次测试参数的平均估计规则学习算法的推广能力。同时在所用语料上实现了基于词性的二元文法 bigram 统计方法、基于反馈神经网络 RNN 的方法(各层节点数分别为 30、60、2，输入时延为 3，输出层到隐层的反馈时延为 2)，以便与规则学习算法比较。此外还将 C4.5 规则和 TBL 规则结合起来实验，即先用 C4.5 规则预测初始类型，再用 TBL 规则来提高正确率。

有关实验结果列在表 5 中。其中  $F_1$  表示  $FMeasure_i$ ， $R_i$ 、 $P_i$ 、 $A_1$  和  $A_2$  意义同表 4。

实验方法	$R_0$	$P_0$	$F_0$	$R_1$	$P_1$	$F_1$	$R_2$	$P_2$	$F_2$	$A_1$	$A_2$
C4.5	0.914	0.837	0.874	0.814	0.822	0.818	0.712	0.829	0.766	0.829	0.904
TBL	0.849	0.884	0.866	0.782	0.848	0.814	0.851	0.613	0.713	0.818	0.895

bigram	0.653	0.746	0.696	0.874	0.816	0.844	-----	-----	-----	-----	0.793
RNN	0.764	0.803	0.783	-----	-----	-----	0.883	0.857	0.870	-----	0.837
C4.5+TBL	0.877	0.906	0.891	0.822	0.870	0.845	0.859	0.684	0.762	0.849	0.915

表 5 各种方法实验结果比较 (-----表示该项无效)

由上表可见，两种规则学习的办法都获得了高于简单统计方法和 RNN 方法的正确率，其中 C4.5 算法和 TBL 办法的结果相差不多，而且 C4.5 算法和 TBL 方法的组合进一步提高了正确率。实验结果说明韵律结构预测作为典型的符号预测问题非常适合用规则学习的办法解决。

## 六、总结与展望

本文提出了一种基于规则学习的韵律结构预测方法。其主要思想就是提取与韵律结构有关的语言学特征，然后用规则学习算法自动归纳预测规则。同时本文给出了对预测结果较为系统的评价参数，通过大量实验比较了不同特征的预测能力。实验结果表明，规则学习算法对于一级韵律结构预测正确率可达 91.5%，两级韵律结构预测可达 84.9%，优于目前其他方法的结果，是一种行之有效的办法。

但应该看到，实验得到的最好预测正确率离人工标注的结果还有一段距离，研究需要更加深入。主要有三个方面值得考虑。其一，对于与韵律结构有关的语言学特征要进一步挖掘，要考虑更深层语法信息，甚至语义信息的作用。其二，采用其他的规则学习算法进行实验。其三，语料库的规模需要扩大，使之能够覆盖更多类型的文本和包含不同的说话风格。

### 参考文献

- [1] Abney Steven. Chunks and dependencies: bringing processing evidence to bear on syntax. Computational Linguistics and Foundations of Linguistic Theory, CSLI, 1995
- [2] Michelle Wang and Julia Hirschberg. Automatic classification of intonational phrase boundaries. Computer Speech and Language 6:175-196.
- [3] Paul Taylor, Black and Alan.W.Black. Assigning phrase breaks from part-of-speech sequences, Computer Speech and Language v12
- [4] 应宏,蔡莲红. 基于结构助词驱动的韵律短语界定的研究.中文信息处理学报,1999,13(6)
- [5] Zhiwei Ying and Xiaohua Shi. An RNN-based algorithm to detect prosodic phrase for Chinese TTS, ICASSP2001
- [6] Yao Qian, Min Chu, Hu Peng. Segmenting unrestricted Chinese text into prosodic words instead of lexical words, ICASSP2001
- [7] 牛振雨,柴佩琪.基于边界点词性特征统计的韵律短语切分.中文信息处理学报, 2001,15(5)
- [8] Quinlan,J.R. Induction of decision trees. Machine Learning, 1(1):81-106
- [9] Brill, Eric. Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging. Computational Linguistics. 21(4):543-565
- [10] Li Aijun, Lin Maocan.. Speech corpus of chinese discourse and the phonetic research. ICSLP2000
- [11] C.J. van Rijsbergen, Information Retrieval. Butterworths, London, 1979.
- [12] Julia Hirschberg, Owen Rambow. Learning Prosodic Features using a Tree Representation, Eruospeech 2001