基于条件随机场的蒙古语韵律短语预测方法*

刘瑞，飞龙*，高光来，张红伟
(内蒙古大学 计算机学院，内蒙古 呼和浩特 010021 中国)

文 摘：韵律预测是提高语音合成自然度的重要因素。蒙古语语音合成技术的研究仍处于起步阶段，合成语音的自然度较低，韵律预测成为了蒙古语语音合成亟待解决的关键问题。本文结合蒙古语语音学特点，将蒙古语单词、词性作为特征，采用条件随机场模型对蒙古语韵律短语预测进行了实验比较。实验结果表明，本文所采用方法的韵律短语边界预测 F 值达到 83.28%，为蒙古语韵律预测的进一步研究奠定了基础。

关键词：蒙古语；韵律预测；条件随机场；语音合成；
中图分类号： TP393

语音合成系统[1,2]可以通过输入文本，让计算机模拟人类输出具有较高自然度和可懂度的语音的一个人工智能系统，为了合成出高质量的语音，必须要能够对输出语句中的停顿、轻重、长短、速度、升降调等做出合理适当的判断。如何自动生成高质量的语音信息越来越受到研究者的关注。在语音合成研究中，韵律的自动预测是提高语音合成自然度的重中之重。

最早的韵律预测研究方法主要是基于规则的方法。但是由于规则往往具有主观性，所以基于规则的韵律预测系统就具有很大的主观倾向性，规则对语言现象的有限覆盖，是基于规则方法难以克服的弊端。近年来，越来越多的研究者使用基于统计学习的方法来进行韵律边界预测的研究，如隐马尔可夫模型（HMM，Hidden Markov Model）、最大熵模型（MEM，Max Entropy Model）等，但是 HMM 方法没有很好的反映上下文信息，限制了预测时特征的选取；MEM 方法能够处理上下文信息，它对每个节点都进行归一化处理从而找出局部最优值，但并没有找到全局最优值。

基于条件随机场[3]（CRF，Conditional Random Fields）方法能很好的解决上述问题，条件随机场模型在自然语言处理领域得到了广泛关注，该方法提供了一个特征使用灵活、全局最优的标注框架，它既保留了 MEM、HMM 等条件框架的优点，又解决了标记偏置问题。而且条件随机场模型在英语、汉语等语言的韵律预测研究中均获得了较为成功的应用[4]。

在蒙古语的语音合成方面，已经做了一些相关研究。敖其尔、巩政提出了一种波形拼接的蒙古语语音合成方法[5]；萨其容贵用基音同步叠加法建立了多样板语音合成音库，对蒙古语的语音合成进行了研究[6]；田会利对基于词干词缀的有限条词的蒙古语语音合成系统进行了研究[7]；孟和吉雅对基于动词词干词缀的蒙古语语音合成方法进行了研究[8]；敖敏对基于韵律的蒙古语语音合成进行了研究[9]；赵建东对基于 HMM 的蒙古语音合成技术进行了研究[10]。这些工作对蒙古语语音合成做出了很大贡献。但蒙古语语音合成结果自然度不够高，很难达到应用水平。

为了提高蒙古语语音合成结果的自然度，本文结合蒙古语语音学特点，采用基于 CRF 的方法对蒙古语韵律短语预测工作进行了研究。论文结构如下：第 2 部分介绍了蒙古语的特点；第 3 部分详细阐述了基于 CRF 的蒙古语韵律短语预测方法；第 4 部分给出实验结果及实验分析结论；第 5 部分为总结及下一步工作展望。

1 蒙古语特点

蒙古文字是一种拼音文字，共有 35 个字母，其中有 8 个元音、17 个基本辅音和 10 个借词辅音。蒙古语拼写时以词为单位，将空格符作为词之间的分隔符：它从上到下连写，从左到右换行[11]。

蒙古语韵律结构[11]可以包括以下四个层级单元：音节、韵律词、韵律短语、语调短语。音节是蒙古语最小的韵律单元。音节边界处（音节之间）

*基金项目：国家自然科学基金(61263037，61563040)；内蒙古自然科学基金(2014BS0604)；内蒙古大学高层次人才引进科研项目
作者简介：刘瑞（1994-），男（汉族），山西省朔州市人，2014 级硕士研究生。
通讯联系人：飞龙，博士，讲师，E-mail：csfeilong@imu.edu.cn

虽然有无声段，但蒙古语语者能够感知到音节之间有短暂的停延。有研究学者根据可感知的停延层级和合成文本的韵律标注要求，把蒙古语口语韵律层级分为以下三个基本单元：韵律词、韵律短语和语调短语。因为小于韵律词的单位没有特定的韵律特征，因此把韵律词作为最小的具有独立韵律价值的单位。

（1）韵律词：词典词、词干+名词附加成分（文字中不连写的）和时位词、后置词和助动词等与前面的词典词一起构成韵律词。蒙古语单词之间的界定是明确的（空格为界限），不需要特意切分。在韵律词内部不能停顿，在韵律词边界不一定有停顿但是可以有停顿。韵律词和词典词不一定一一对应。

（2）韵律短语：一般由处于同一个节奏群里的两个或几个联系比较紧密的韵律词构成，在韵律短语内部的韵律词之间通常没有可感知的停顿，而在韵律短语之间通常没有比较明显的停顿。

（3）语调短语：由一个或几个韵律短语构成。语调短语一般与简单语法词组或复杂语法词组相对应，句子中可以承担主语部分、谓词部分、宾语部分、定于部分或状语部分。在语调短语后一般有明显的停顿。

三个韵律单元存在着包含关系，即语调短语边界一定是韵律短语边界，韵律短语边界一定是韵律词边界，而韵律短语边界只能落在韵律词边界上。但是韵律短语边界不一定是语法词组边界，词典词边界也不一定是韵律词边界。

在蒙古语韵律层级中，韵律短语是比较重要的韵律单元（Prosodic Unit）。能否从语音学线索准确的预测出韵律短语的边界位置，关系到基频曲线和时长模式的准确拟合，直接影响到合成语音的自然度。韵律短语（Prosodic Phrase）内部一般具有相对稳定的韵律模式。在言语产生过程中，发音人通过音高的不同变化以及在多个韵律短语之间插入长度适当的停顿来增强节奏感。因此，本文研究工作是针对韵律短语边界的预测。

蒙古语是黏着性语言，它没有前缀和中缀，只有后缀，可以由词根或词干后缀接不同的后缀形成大量的新单词，根据这样的构词规则可以生成大规模的蒙古文单词。并且，蒙古文存在“同形不同码”问题，同形不同码是指在一个蒙古语单词中，蒙古语字符具有相同的字形，却有不同的编码。这导致了蒙古语的拼写错误难以被发现，因此蒙古语文本中存在大量的同形但编码错误的单词。这些特点与现象对于蒙古语韵律短语的预测工作带来了很大的困难和挑战。

2 基于 CRF 的蒙古语韵律预测方法

2.1 CRF 模型

CRF 由 Lafferty 等人于 2001 年提出，结合了最大熵模型和隐马尔可夫模型的特点，是一种无向图模型，是在给定待标记的观察序列的条件下计算整个标记序列的联合概率分布。近年来在分词、词性标注和命名实体识别等序列标注任务中取得了很好的效果。它的定义如下：假设 G=（V，E）是无向图，$Y=\{y_v \mid v\in V\}$是以 G 中结点 v 为索引的随机变量y_v构成的集合。如果每个随机变量y_v服从马尔可夫属性，则称（X,Y）是一个 CRF。

设 $C=\{(x_c, y_c)\}$是图 G 中所有的团构成的集合，根据随机场的基础理论，在给定观测序列 x 的条件下标记序列 y 的概率分布 p（y|x）为

$$p_\Lambda(y \mid x) = \frac{1}{Z(x)} \prod_{c\in C} e^{\sum_k \lambda_k f_k(y_c,\ x_c)} \quad (1)$$

式中，$f_k(y_c, x_c)$是特征函数，模型参数是一个实数构成的特征函数权值集合$\Lambda = \{\lambda_k\}$，其归一化因子为

$$Z(x) = \sum_y \prod_{c\in C} e^{\lambda_k f_k(y_c,\ x_c)} \quad (2)$$

当用该模型建模序列数据时，图 G=（V，E）中的状态变量 y 的形状最简单最常用的是一条一阶链。这条链中的团是其中的结点和边。因此，在整个观测序列上可以定义两类特征函数：状态特征函数和转移特征函数。给定训练样本集$\{(x^{(k)}, y^{(k)})\}$和预定义的特征函数，可以从样本中学习一个 CRF 模型。模型参数Λ可以使用极大似然、极大后验等方法估计。

对于 1 个输入测试序列 x，可以使用训练得到的 CRF 模型来推断它对应的标注序列，x 最可能的标记序列$\hat{y}$表示为

$$\hat{y} = \arg\max_y p_\Lambda(y \mid x) = \arg\max_y \sum_{c\in C} \sum_k \lambda_k f_k(y_c x_c) \quad (3)$$

式中 $\hat{y}$ 可以用动态规划的 Viterbi 算法来查找。

2.2 特征模板

本文在“单词”这一基本特征的基础上加入了“词性”特征，进行了基于 CRF 的蒙古语韵律短语预测实验。

在 CRF 模型的特征中，“窗口”的内容即以当前单词为中心的上下文信息。窗口长度越大，包含的上下文信息越丰富，对韵律短语预测越有利，但窗口长度太大，可能出现过拟合现象，而窗口长度过小，则加入特征不充分，包含信息有限，会忽略一些有用信息。

3 实验

3.1 实验语料

本文从内蒙古大学计算机学院建立的蒙古语语音合成数据库中选取了 3282 句语音和对应的文本，并根据语音语料库中的语音数据，结合一定的句法语义信息对对应的文本语料库进行韵律短语边界标记。本文标注语料包含 52460 个蒙古文单词，去重后单词个数为 11100 个。实验采用 5 次交叉验证，每次随机采用 90%作为训练集，10%作为测试集。5 次交叉验证语料中测试集的单词个数和集外词单词的分布情况如表 1 所示。

表 1 测试集单词个数和集外词单词的分布情况

测试集	Test1	Test2	Test2	Test2	Test2
单词数	2236	2193	2295	2368	2123
集外词数	560	517	567	587	473

由于蒙古语字母在不同位置有不同的书写形式，鉴于这样的现象，对文本语料库采用拉丁转写的方式，使每一个蒙古语字母都仅有一个拉丁转码与之对应，从而避免了形同音异词难以辨别和区分的问题。

在拉丁转码的基础上进行了韵律短语边界标注[12,13]和词性标注，标注符号说明如表 2 所示。本文的词性标注方式采用一级词性标注，即“第一级标记的命名方式”，如表 3 所示。

表 2 韵律短语边界标注符号说明

标注符号	标注位置	符号意义
B	韵律短语非末尾位置	表示其不是韵律边界
E	韵律短语末尾位置	表示其是韵律边界
S	标点符号位置	表示其代表韵律边界

表 3 词性标注符号说明

词类名称	标注符号	词类名称	标注符号
名词	N	时间词	T
形容词	A	副词	D
数词	M	情态词	H
量词	Q	摹拟词	U
时位词	O	后置词	G
代词	R	语气词	S
动词	V	连接词	C
感叹词	I		

3.2 实验

3.2.1 实验

本文为了确定各种特征在不同情况下对于预测效果的影响，分别做了实验对比。根据不同情况下预测结果的比较确定最佳窗口长度（以单词信息为例，本文中的“窗口长度 n”指当前词、当前词的前 n 个词和当前词的后 n 个词）。因此分别做了 3 组对比实验：第 1 组的训练语料只对其进行人工韵律标注，图 1 所示为语料标注格式样例。在特征模板中只选用“单词”特征，分别采用窗口长度 1、2、3、4 进行实验，不包括标点符号的实验结果如表 4 所示。对第 2 组的训练语料利用词性标注工具进行一级词性标注[14]，并选用单词“词性”作为特征，分别采用窗口长度 1、2、3、4 进行实验比较。不包括标点符号的实验结果如表 5 所示。第 3 组实验为根据前 2 组实验的最优窗口长度结果，采用“单词+词性”作为组合特征进行了实验比较。

```
VYIR_E-YIN B
JIL-dV E
, S
NEHULGEJU B
SILJIHULHU B
AJIL B
VLASIRAGSAN E
, S
NEHULGEJU B
SILJIHULHU B
NI E
AYIL B
ERU/GE B
BVLGA/N/-V B
ASIG B
TVSA B
TEI B
HVLBVGDAGSAN B
VCIR E
. S
```

图 1 语料标注样例

表 4　采用“单词“特征的实验结果

	Test1	Test 2	Test 3	Test 4	Test 5	Ave-F
1	49.25%	41.90%	**43.88%**	**46.42%**	**49.88%**	46.27%
2	**50.42%**	44.31%	43.24%	44.93%	48.94%	**46.37%**
3	48.56%	**45.78%**	41.73%	44.07%	48.59%	45.75%
4	44.44%	42.80%	42.42%	42.70%	44.55%	43.38%

由表 4 可知，当“单词”特征模板长度为 1 时在 5 组测试集得到的结果中有 3 组测试集的 F 值达到了最高，因此“单词”特征的最佳窗口长度确定为 1。

表 5　采用“词性“特征的实验结果

	Test1	Test 2	Test 3	Test 4	Test 5	Ave-F
1	24.41%	23.82%	25.47%	23.11%	23.17%	24.00%
2	28.81%	27.84%	28.49%	28.33%	28.07%	28.31%
3	**30.28%**	28.04%	**30.29%**	**28.45%**	**28.34%**	**29.08%**
4	29.97%	**28.28%**	28.60%	27.58%	28.17%	28.52%

由表 5 可以看出，当“词性”特征模板窗口长度为 3 时，5 组测试集的得到的结果中有 4 组测试集 F 值达到了最高值。因此选用“词性”特征的最佳窗口长度为 3。

所以本文确定上下文单词窗口长度为 1、上下文单词词性窗口长度为 3 作为最佳特征模板“T”，进而进行第 3 组实验得到不包括标点符号的实验结果如表 6 所示和包括标点符号的实验结果如表 7 所示：

表 6　采用“单词”加“词性”的组合特征的实验结果

	Test 1	Test 2	Test 3	Test 4	Test 5	Ave-F
T	59.38%	59.18%	58.56%	58.05%	61.39%	59.31%

表 7　采用“单词”加“词性”组合特征的实验结果

	Test 1	Test 2	Test 3	Test 4	Test 5	Ave-F
T	84.19%	83.42%	82.06%	82.34%	84.37%	83.28%

从表 7 可以看出，选用“单词+词性”作为组合特征的韵律短语预测结果达到了 83.28%。从表 6 可以看出，不包括标点符号的韵律短语预测结果并不理想，主要原因可能源于以下两点：（1）蒙古语词汇量规模很大，实验语料中涉及的词汇量有限，并且测试集中的集外词所占比例很大；（2）在实验中我们只选用了单词和一级词性两个最基本的特征来构建 CRF 模型，包含特征信息有限。另外，在语料预处理阶段的词性标注工作存在一定的错误率也会对预测结果造成影响。

3.2.2　评价标准

实验的评价标准是韵律短语边界 F 值。由于标点符号可以看作比较明显的韵律边界特征，所以实验结果 F 值的计算考虑包含标点符号和不包含标点符号两种情况。它的值通过正确率（P，Precision）和召回率（R，Recall）计算得到：

$$P = \frac{\text{实验得到的正确韵律边界标注}}{\text{实验得到的所有韵律边界标注}}$$

$$R = \frac{\text{实验得到的正确韵律边界标注}}{\text{人工标注的所有韵律边界}}$$

从而 F 值的定义为：　$F = \frac{2 * P * R}{P + R}$

4　结束语

本文结合蒙古语的语音学特点，采用基于 CRF 的方法对蒙古语韵律短语边界进行预测，实验结果表明，韵律短语边界预测 F 值达到了 83.28%，对于增强语音合成系统的输出语音自然度起到了一定的作用，并为今后深入研究蒙古语韵律预测起到了参考作用。

在下一步的工作中，我们将围绕以下几点开展研究工作：（1）标注更大规模的文本语料当作实验数据；（2）针对蒙古文词汇量规模大这一特点，采用词干-后缀的切分方式对蒙古语韵律进行进一步预测；（3）针对蒙古文的“同形不同码”现象，采用校正和中间字符表示等方法进行处理从而克服这一现象对预测结果的影响；（4）进一步增加特征数量，如二级或三级词性、蒙古文单词的词干、后缀、音节和音素等特征。

参 考 文 献

[1] Zen Hei-ga，Takashi N，Junichi Y，et al. The HMM-based Speech Synthesis System(HTS)Version 2.0 [C] // 6th ISCA Workshop on Speech Synthesis. Bonn，2007:294-299

[2] 井晓阳，罗飞，王亚棋. 汉语语音合成技术综述 [J]. 计算机科学，2012,39(11A),336-390
JING Xiao-yang,，LUO Fei，WANG Ya-qi. Overview of the Chinese Voice Synthesis Technique [J]. Computer Science，2012,39(11A),336-390（in Chinese）

[3] 董远， 周涛， 董乘宇， 王海拉等. 条件随机场模型在韵律结构预测中的应用[J]. 北京邮电大学学报,2009,32(5):36-40.
DONG Yuan, ZHOU Tao, DONG Chengyu, WANG Haila. Prosodic Structure Prediction Based on Conditional Random Field Model[J]. Journal of Beijing University of Posts and Telecommunications, ,2009,32(5):36-40. （in Chinese）

[4] 包森成. 基于统计模型的韵律结构预测研究[D]. 北京：北京邮电大

学, 2009.
BAO Sen-cheng. Research on Prosodic Strucutre Prediction Based on Statical Model[D].Beijing: Beijing University Of Posts And Tele-communications,2009 （in Chinese）
[5] 敖其尔，巩政．一种波形拼接的语音合成实验 [C] // 第三届全国人机语音通讯学术会议．重庆，1994:408-412
Ochir, Zheng Gong. A Test of The Speech Synthesis With The Wave-form Cancatenation [C] // 3th NCMMSC. Chongqing, 1994:408-412（in Chinese）
[6] 萨其容贵． 蒙古语语音合成技术的研究 [D]. 呼和浩特: 内蒙古大学，2005.
The Research of Mongolian Speech Synthesis Technology [D]. Hohhot: Inner Mongolia University, 2005.（in Chinese）
[7] 田会利．基于词干词缀的有限条词的蒙古语语音合成系统的研究 [D]. 呼和浩特: 内蒙古大学，2007
TIAN Hui-li. The Research on Mongolian Speech Synthetical-System Bsaed on ETYMA and AFFIX For Tinite Words. [D]. Hohhot: Inner Mongolia University, 2007.（in Chinese）
[8] 孟和吉雅． 基于动词词干词缀的蒙古语语音合成方法 [J]. 内蒙古大学学报,2008,39(6):693-697
Monghjaya. A Research on Mongolian Speech Synthesis-system Based on Stems and Affixes [J]. Journal of Inner Mongolia University, 2008,39(6):693-697（in Chinese）
[9] 敖敏．基于韵律的蒙古语语音合成研究[D]. 呼和浩特: 内蒙古大学，2012.
Aomin. Research on Mongolian speech synthesis based on prosody [D]. Hohhot: Inner Mongolia University, 2012. （in Chinese）
[10] 赵建东，高光来，飞龙. 基于 HMM 的蒙古语语音合成技术研究[J]. 计算机科学，2014,41(1):80-104.
ZHAO Jian-dong,GAO Guang-lai,BAO Fei-long. Research on HMM-based Mongolian Speech Synthesis [J]. Computer Science 2014,41(1):80-104. （in Chinese）
[11] 敖敏，熊子瑜，呼和. 蒙古语标准话朗读话语韵律短语研究[J]. 中央民族大学学报,2012,39(4):143-148.
Aomin, XIONG Ziyu, Huhe. Research on Prosodic Phrase of Reading Speech in Standard Mongolian [J]. Journal of the Central University for Nationalities, 2012,39(4):143-148. （in Chinese）
[12] 赵建东，高光来，飞龙. 蒙古语语音合成语料库标注规则的设计[J]. 内蒙古大学学报，2013,44(3):324-328.
ZHAO Jian-dong,GAO Guang-lai,BAO Fei-long. Designing a Rule Annota tion of Corpus Data in Synthesis of Mongolian Speech. Journal of Inner Mongolia University, 2013,44(3):324-328. （in Chinese）
[13] 赵建东，高光来，飞龙. 基于历史模型的蒙古文自动词性标注研究[J]. 中文信息学报，2013,27(5):156-165.
ZHAO Jian-dong,GAO Guang-lai,BAO Fei-long. Research on Histo-ry-based Mongolian Automatic POS Tagging [J]. Journal of Chinese Information Processing, 2013,27(5):156-165. （in Chinese）
[14] 斯•劳格劳，华沙宝，萨如拉. 基于 NFA 的蒙古语词法分析算法研究 [A]. 少数民族青年自然语言处理技术研究与进展——第三届全国少数民族青年自然语言信息处理、第二届全国多语言知识库建设联合学术研讨会论文集[C].2010
S．Loglo，HuaShabao，Sarula. Research on Mongolian Lexical Parser Algorithm Based on NFA [A]. Research and Development of natural language processing technology of Minority youth——The third session of the national minority youth natural language information processing、The second session of the National Multi-language knowledge base construction joint Symposium [C]，2010

Approach to Predicition Mongolian Prosody Phrase Based on CRF Model

LIU Rui , BAO Feilong, GAO Guanglai, ZHANG Hongwei
College of Computer Science, Inner Mongolia University, Hohhot Inner Mongolia 010021, China；

Abstract: Mongolian prosody prediction is a key factor to improve the Naturalness of the Speech and the research of Mongolian speech synthesis technology is still at the initial stage . However, the low naturalness of synthesized speech lead to the prosody prediction has become a key issue to be solved Mongolian speech synthesis. According to the Phonetics Features of Mongolian language. this paper takes the word, part-of-speech as features to do comparison based on conditional random fields. The experimental results have shown that a good performance achieved by the CRF and the prosodic phrase boundary prediction F value reached 83.28%. It laid the foundation for further study of Mongolian prosody phrase prediction.

Key words: Mongolian; prosody prediction; CRF; speech synthesis;