Linguistic Features Selection in Fundament Frequency Patterns

Yiqiang CHEN, Wen GAO
Institute of Computing Technology, Chinese Academy of Sciences, Beijing
{yqchen,wgao}@ict.ac.cn

Tingshao ZHU
Dept. of Computing Science, University of Alberta, Edmonton
ts Zhu@cs.ualberta.ca

ABSTRACT

The prosodic pattern generation and prediction is more important for synthesizing natural-sounding speech reproduction of input Chinese text. In this paper, the typical pitch models are clustered from a large actual speech database firstly. Then we propose several methods including rough set method and Bayesian relief network on linguistic features selection, which can be directly used to predict pitch, energy, and duration patterns. A comparison between these two methods is proposed and to overcome each disadvantage, we combined the results of these two methods, and coded the most important features to Bayesian relief network firstly. After learning, some experiment shows the F0 model prediction based on the selected features is the same as original one for most pitches.

1. INTRODUCTION

Prosody is one of important component for TTS system. The research now mainly deals with invoking higher-level linguistic features in exploring the prosodic information of Mandarin speech to assist prosodic pattern generation [1][2]. From the Chinese Dictionary and the existence of a text processing model, the lexical information (phonemic representations and lexical stress) and symbolic prosodic markers can be obtained. These linguistic features may have different effect on different prosodic component prediction such as fundament frequency (F0) patterns, F0 duration and F0 energy as well. It’s essential for us to select the suitable linguistic features in the above prediction. In this paper, the typical pitch models are clustered from a large actual speech database firstly. Then we propose several methods including rough set method and Bayesian relief network on linguistic features selection, which can be directly used to predict pitch, energy and duration patterns. A comparison between these two methods is proposed and to overcome each disadvantage, we combined the results of these two methods, and coded the most important features to Bayesian relief network firstly. After learning, some experiment shows the F0 model prediction based on the selected features is the same as original one for most pitches.

2. PROSODIC PATTERNS AND LINGUISTIC FEATURES

2.1 The Prosodic Patterns

We assume that the F0 contour in the continuous speech data are not variety randomly but can be obtained through modifying some classic F0 model with duration and mean. These classic F0 models can be obtained from the preprocessed actual F0 contours. For the F0 pattern classification, the original prosody pattern is divided into three parts: zero-mean F0 pattern, the duration, and the mean as our output.

The classification preprocessing mainly deals with the data from speech database directly, which extracts pitch, wraps the duration and normalizes and smooth and zero mean the pitch values to meet the requirement of cluster algorithm. The ISODATA [3] (Iterative Self-organizing Data) algorithm is chosen for our clustering. After clustering, there are 18 F0 patterns are classified, Figure 1 shows them:
The original pitch from sentences is discrete with extracted classic F0 models, and at the same time the original length and mean should be kept for future learning. Rough set is employed for linguistic features selection.

2.2 The Linguistic Features

Our aim is to explore the relationship between the prosodic pattern of Mandarin speech and the linguistic features of the input text to simulate human’s prosody pronunciation mechanism. The Chinese Dictionary that we are using includes the spell, vowel, constant, tone, part of speech (POS), and some word syntax and semantic. From this dictionary and the existence of a text processing model, the lexical information (phonemic representations and lexical stress) and symbolic prosodic markers can be obtained. In this paper, after parsing, the linguistic features including the following:

- The number of pitch in word (len)
- The sequence number serial number of pitch in word (wordno)
- Word class and POS (pos)
- Is substantive or function word? (xs)
- Is prediction or noun word? (tw)
- The vowel, constant and tone of current pitch
- The vowel, constant and tone of prior and post pitch

3. Linguistic Features Selection

3.1 The Linguistic Features Selection Based on Rough Set Method

The Rough set is proposed to find the minimum attribute set. The rough set theory [4] is based on indiscernibility relation. Suppose four finite, non empty sets R, A, V and f, where R is the universe, and A is a set of attributes, V is the value set of each attribute and f is a function map $f : U \times A \rightarrow V$. The indiscernible relation I is associated with every subset of attributes $P \in A$ and defines as:

$$I(P) = \{(r_i, r_j) \in U \times U : f(r_i, attr) = f(r_j, attr), \forall attr \in P\}$$  \hspace{1cm} (1)

Where $f(r_i, attr)$ is the value of attribute $attr$ in object $r_i$. If $(r_i, r_j) \in I(P)$, then $r_i$ and $r_j$ are P-indiscernible. Rough set can remove unnecessary attributes from the set A by considering redundancies and dependencies between attributes [5]. Let P be a subset of A, and the initial P is the set A. If $I(P) \neq I(P - \{attr\})$, then we say that the $attr$ can be moved from the set A. Thus the main features are selected by Rough set. Such as for predicting F0 model, the linguistic features selected through rough set are the sequence serial number of pitch in word, the vowel, constant and tone of current pitch, prior and post pitch.

3.2 The Linguistic Features Selection Based on Bayesian Relief Network Method

Bayesian belief networks specify joint conditional probability distributions. They allow class conditional independence to be defined between subsets of variables. They provide a graphical...
model of causal relationships, on which learning can be performed. A belief network is defined by two components. The first is a directed acyclic graph, where each node represents a random variable, and each arc represents a probabilistic dependence. If an arc is drawn from a node $Y$ to a node $Z$, then $Y$ is a parent or immediate predecessor of $Z$, and $Z$ is a descendent of $Y$. Each variable is conditionally independent of its nondescendants in the graph, given its parents. In Bayesian relief network constructing, if network structure is given, then learning the network is straightforward and simple [6]. In recent years, many Bayesian network structure learning algorithms have been developed [7][8]. These algorithms generally fall into two groups, search & scoring based algorithms and dependency analysis based algorithms. For linguistic features selection, we only need to know the dependency relationship among the features and predicted variables. Thus, our algorithm is based on dependency analysis. The mutual information and condition mutual information measurement are used in our Bayesian belief network constructor. In information theory, mutual information is used to represent the expected information gained on sending some symbol and receiving another. In Bayesian networks, if two nodes are dependent, knowing the value of one node will give us some information about the value of other node. This information gain can be measured using mutual information. Therefore, the mutual information between two nodes can tell us if two nodes are dependent and how close their relationship is. The mutual information of two nodes $X_i, X_j$ is defined as:

$$I(x_i, x_j) = \sum_{x_i, x_j} p(x_i, x_j) \log \frac{p(x_i, x_j)}{p(x_i)p(x_j)}$$

In addition, the conditional mutual information is defined as:

$$I(x_i, x_j | c) = \sum_{x_i, x_j} p(x_i, x_j, c) \log \frac{p(x_i, x_j | c)}{p(x_i | c)p(x_j | c)}$$

Where $C$ is a set of nodes. When $I(x_i, x_j)$ is smaller than a certain threshold $\varepsilon$, we say that $x_i, x_j$ are marginally independent. When $I(x_i, x_j | c)$ is smaller than $\varepsilon$, we say that $x_i, x_j$ are conditionally independent given $C$. We treat each linguistic feature as a random variable, which is represented by a node in a Bayesian network. The experiment for feature selection of pitch model is showed in Figural 1:

![Figure 1: feature selection of pitch model](image)

For Bayesian relief network classification, the selected features are post vowel, current tone, current consonant and prior tone. The NN and Decision tree are constructed using these features as input or condition attributes and the F0 model as output or decision attributes.

### 3.3 The Combination of Two Methods

The advantage of Bayesian relief network classification is that some background knowledge can be coded to network firstly, so we combined the results of these two methods, and coded the most important features to our Bayesian relief network firstly. After learning, the linguistic features for F0 model, F0 duration and F0 mean are gained, and table 1 shows them:
3.4 Experiment
The NN and Decision tree are constructed using these features as input or condition attributes and the F0 model as output or decision attributes. The result shows in figure 2:

![Figure 2](image)

For most sentence, our experimental results shows the F0 model prediction based on the selected features is the same as original one for most pitches.

4. CONCLUSION
Linguistic features may have different effect on different prosodic component prediction such as fundament frequency (F0) patterns, F0 duration and F0 energy as well. It’s essential for us to select the suitable linguistic features in the above prediction. In this paper, the typical pitch models are clustered from a large actual speech database firstly. Then we propose several methods including rough set method and Bayesian relief network on linguistic features selection, which can be directly used to predict pitch, energy, and duration patterns. A comparison between these two methods is proposed and to overcome each disadvantage, we combined the results of these two methods, and coded the most important features to Bayesian relief network firstly. After learning, some experiment shows the F0 model prediction based on the selected features is the same as original one for most pitches.

5. REFERENCES