Mongolian Prosodic Phrase Prediction using Suffix Segmentation

Rui Liu, Feilong Bao, Guanglai Gao, Weihua Wang

College of Computer Science, Inner Mongolia University

Abstract

Accurate prosodic phrase prediction can improve the naturalness of speech synthesis. Predicting the prosodic phrase can be regarded as a sequence labeling problem and the Conditional Random Field (CRF) is typically used to solve it. Mongolian is an agglutinative language, in which massive words can be formed by concatenating these stems and suffixes. This character makes it difficult to build a Mongolian prosodic phrase predictions system, based on CRF, that has high performance. We introduce a new method that segments Mongolian word into stem and suffix as individual token. The proposed method integrates multiple features according to the characteristics of Mongolian word formation. We conduct the contrast experiment by selecting the following features: word, multilevel Part-of-Speech (POS), multi-level lexical for suffix and the existence for suffix. The experimental results show that our method has significantly enhanced the performance of the Mongolian prosodic phrase prediction system through comparing with the conventional method that treats Mongolian word as token directly. The word feature, level one lexical for suffix feature and existence for suffix feature are effective. The best result is measured by F1-measure as 82.49%.

Characteristics of the Mongolian Word Formation

Mongolian is an agglutinative language. A Mongolian word can be decomposed into a root and several suffixes. Mongolian does not have an infix and prefix, but a suffix. A suffix can be categorized as word-formation suffix, inflectional suffix and ending suffix.

Generally, ending suffixes affect the form only, such as the tense, person and so on. Every root can be followed by many word-formation suffixes and inflectional suffixes to generate the stem. Most of the generated stem can be further followed by an ending suffix and a little of them are followed by two ending suffixes. Moreover, the pronunciation of suffix will be different depending on the stem they follow. Fig. 1 illustrates the relationship among the root, stem and suffix of a Mongolian word.

Approach

A. Segmentation process

For many Mongolian nouns: NNBSP, a special character, is used to concatenate root and suffixes. The NNBSP characters are written as “-” in Latin representations. The ending suffix which is concatenated with stem by NNBSP is called NNBSP-suffix by us. In this study, we segment the Mongolian nouns into stems and NNBSP-suffixes

In order to ensure the accuracy of segmentation, we only segment the NNBSP-suffix in this study. After the segmentation process, we propose two methods to alleviate the large-vocabulary problem.

-- Treat the stem and the NNBSP-suffixes as individual tokens. (SEG)

-- Remove every NNBSP-suffix, and only treat the stems as tokens. (SEG_RM)

B. Selecting Features

According to the characteristics of the Mongolian word formation, nine features are used in this paper. Here, these features are shown as follows:

Word (w): the Mongolian word itself.

Level one POS (p1): level one POS tag of the word.

Level three POS (p3): level three POS tag of the word.

Level one lexical tag for VSES (v1): If a Mongolian word has a VESE, we mark the level one lexical tag for VSES.

Level three lexical tag for VSES (v3): If a Mongolian word has a VESE, we label the level three lexical tag for VSES.

Level one lexical tag for NNBSP-suffix (n1): If a Mongolian word has a NNBSP-suffix, we label the level one lexical tag for NNBSP-suffix.

Level three lexical tag for NNBSP-suffix (n3): If a Mongolian word has a NNBSP-suffix, we label the level three lexical tag for NNBSP-suffix.

NNBSP-suffix (s): NNBSP-suffix itself.

Existence of NNBSP-suffix (t): The tag is showing whether the NNBSP-suffix exist or not in a Mongolian word.

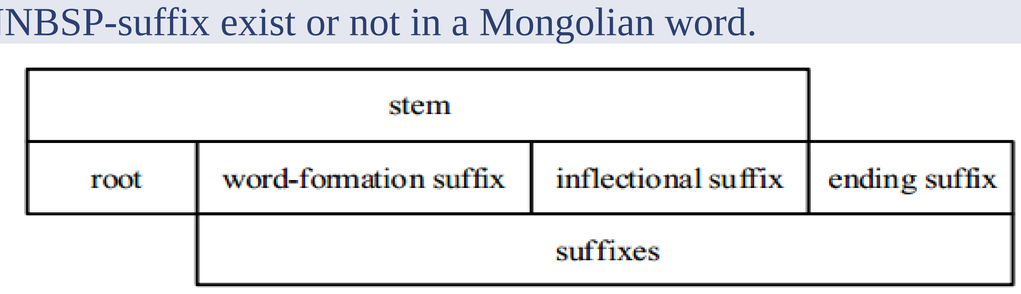

Figure 1. The relationship of root, stem and suffix of a word in mogolian

Conclusions

In this paper, we explore two NNBSP-suffix process methods based on CRF for the Mongolian prosodic phrase prediction. The results of the experiment show that the processing of NNBSP-suffix can improve the performance of the Mongolian speech synthesis system. We find that the performance of the Mongolian prosodic prediction by the SEG method is better than the SEG_RM method and the above methods are better than the method without processing Mongolian word. It indicates that it is helpful to segment the Mongolian word into smaller token. In addition, in the process of feature selection in CRF training, we select nine features: word, level one POS, level three POS, level one lexical tag for VSES, level three lexical tag for VSES, level one lexical tag for NNBSP-suffix, level three lexical tag for NNBSP-suffix, NNBSP-suffix, existence of NNBSP-suffix. Word and NNBSP-suffix are the most effective features among the above features. Finally, by combining the three features: Word, Level three lexical tag for VSES and Existence of NNBSP-suffix, we obtain the best result (F1-a: 82.49, F1-b: 63.49).

Contact

College of Computer Science, Inner Mongolia University,

Hohhot 010021, People's Republic of China

csfeilong@imu.edu.cn