Abstract

One of the enduring problems in developing high-quality Chinese TTS (text-to-speech) systems is accurate text-to-Pinyin conversion. To solve the problem, identification of words and assignment of correct POS (Part-of-Speech) tags for an input sentence are very important tasks. Also, determining the correct Pinyin of polyphonic characters in unknown words is an important problem in a Chinese TTS system.

The unknown word problem has significant effects on the accuracy of the synthesized sound, so accurately predicting the category of unknown words can help a Chinese TTS system to pronounce more naturally. In this paper, we present an SVM (support vector machine)-based method that predicts the unknown words for the results of Chinese word segmentation and POS tagging. For high speed SVM processing to be used in TTS, we pre-detect the candidate boundary of the unknown words before starting the actual prediction. Results of the experiments are very promising by showing high precision and high recall with also very high speed.

1. Introduction

It is well known that the pronunciation and spelling of a word is generally different. The text-to-Pinyin conversion can be described as a function mapping from the spelling of Chinese words to their phonetic symbols. Because pronunciation of a word cannot always be determined only from the spelling of the word, the function needs some linguistic knowledge, especially morphological and phonological, but often also semantic knowledge.

Since there is no blank to delimit the words in Chinese sentences, we need a high quality word segmentation and POS tagging process for high quality text-to-Pinyin conversion. However, we cannot include all the new words for segmentation and tagging in a dictionary even if we generate a word dictionary from very large corpus. Therefore, unknown word handling for correct pronunciation processing should be essential for a more accurate and natural TTS sound.

2. Overall architecture of the text-to-Pinyin converter using unknown word prediction

Figure 1 shows the Chinese text processing steps for a Chinese TTS system. The system is composed of three steps. First, we segment input texts into words and assign POS tags to each word automatically. Next, we predict the category of unknown words for tagged texts by our proposed two-phase SVM-based method. Finally, a text-to-Pinyin conversion module generates Pinyin strings for correct pronunciation.

In our research, we used previously developed word segmentation and POS tagging system called POSTAG/C [1]. POSTAG/C is a system which combines word segmentation module based on rules and a dictionary with a POS tagging module based on HMM (Hidden Markov Model). The word dictionary was automatically acquired by POS tagged corpus and the system has high portability.

The word segmenting module is composed of two parts. Input Chinese texts go through a pre-segmenting process at first. The pre-segmenting process splits input texts into short phrases using some special characters that are symbols, numbers, foreign language characters, and mono-character words. Next, we segment in more detailed using four heuristic rules which [2] proposed.
3. Two-phase unknown word category prediction

The two-phase unknown word prediction process extracts candidate boundaries where unknown words possibly occur in a sentence at first. Next, we predict the category of the words with these boundaries using an SVM-based method. Figure 2 explains the entire procedure that extracts possible boundaries and predicts the words using a SVM-based method.

3.1. Detection of the candidate boundary

Each module that is a part of voice synthesis systems should operate in real time. However, if we check all the texts to predict the unknown word from the beginning of input texts to the end, the speed may become very slow. Moreover, we need a more efficient method if we take into account the slow speed of SVM that will be used in our research. SVM is one method that exhibits the best performance among all the machine learning methods, but slow learning and prediction time is its major shortcoming. To overcome speed problems while not losing accuracy, instead of examining the whole sentence, we detect the candidate boundaries where occurrences of unknown words are possible.

As a general Chinese word segmentation system, POSTAG/C also outputs contiguous single Chinese characters for the unknown words. Therefore, we can use the boundary where single Chinese characters appear consecutively as the first candidates of unknown words. Studies that show more than 90% of the unknown words are actually included in this boundary in Chinese support our approach [3]. Without stopping here, we extend our target boundary to increase the recall of the boundary detection by including 2-character words. So, our system can cover the bottom case as well as the top one in Figure 3.

We cannot use the sequence of all the single Chinese characters as the candidate boundaries because a single character can frequently be used as a word. In our own statistics using Chinese news articles, the number of total boundaries that are the series of a single character in person names was 128,410 cases, but only 16,955 cases among them actually include unknown words.

To cope with these spurious cases, we select the candidate boundaries for a series of single characters by matching to the pre-learned hanzi bigram patterns. These patterns are learned by person names and location names that are extracted from training data. We generated the patterns by combining two characters which are adjacent in person names or location names. Next, we generated a dictionary with patterns and their frequencies. There are 34,662 person patterns and 15,958 location patterns used in our system. We extract the boundaries that match with more than one bigram pattern.

3.2. SVM training features

We use 10 features for SVM training for unknown word category prediction. Each character in the boundary predicts its own position tag (described in section 3.3) using lexical and position tag features of the previous two and the next two characters. Moreover, we use additional features such as a possible character in a family name of a Chinese person and foreign transliteration, and the last character of a location name. The number of features of a family name is taken from the top 200 family names, which are most frequently used in China, and the number of features of foreign transliteration is 520. We also use high frequency 100 features of the last character of location names from our corpus.

Using the individual characters as features for prediction is useful because we must deal with unknown words that are contiguous single characters. The character-based features allow the system to predict the unknown words more effectively in this case, as shown in [4].

3.3. SVM-based prediction of unknown word

We develop an SVM based unknown word prediction method for the output of the candidate boundary detection. We use a library for support vector machines, LIBSVM [5] for our experiments. The kernel function is a RBF that can achieve the best parameters for training in generating the final model. We give a position tag for each character and create features that are used in training and testing.

The prediction first assigns the presumed position tags to the characters in a candidate boundary. Then we combine those characters according to the information of the assigned position tags, and finally identify the entire unknown word.

During the unknown word prediction step, we use 4 different classes of position tags to classify the characters. These classes are \([B-POS]\), \([I-POS]\), \([E-POS]\) and \([O]\), where POS is a POS tag of a word, such as a person name or a location name, and \(B\), \(I\) and \(E\) are the classes of characters according to the positions in the word (\(B\): Begin Position; \(I\): Inside Position; \(E\): End Position). \(O\) is the class of outside characters which are not classified into the previous three position classes. After the prediction step, we combine these characters into a single word. Finally, we carry out rule-based postprocessing using error cor-
retraction rules as the following:

\[ PT_i : \{ \emptyset \}, PT_{i-1} : [B - NR], PT_{i+1} : [E - NR] \]
\[ \rightarrow PT_i : [I - NR] \]

where \( PT_i \) is a current position tag, \( PT_{i-1} \) is a previous position tag, \( PT_{i+1} \) is a next position tag, and \( NR \) is a POS for a person name.

Figure 4 shows an example of the final result of our unknown word prediction.

4. Text-to-Pinyin conversion with predicted unknown words

Our automatic text-to-Pinyin conversion module is shown in Figure 5. The method of our conversion is a dictionary-based and rules-based hybrid approach. First, we examine the input words (including category predicted unknown words) whether some polyphonic characters exist in those words. If no polyphonic character exists in the word, the conversion module simply outputs the sequence of Pinyin of each character using a character-Pinyin dictionary. If one or more polyphonic characters do exist, we examine whether the input word exists in a polyphonic word-Pinyin dictionary which was automatically generated by POS tagged corpus and refined by ourselves. If the pronunciation of the word exists in this dictionary, our module outputs the corresponding category’s Pinyin string. During this processing step, the predicted category of the unknown words is greatly used in determining the correct pronunciation. However, if the input word does not exist in the dictionary, our module has to decide the approximate Pinyin using a polyphonic character priority(frequency) dictionary, which has been achieved by a commercial Chinese speech script. In this process, our module converts the decided Pinyin strings into more natural forms with standard Chinese phonetic rules. For example, when Pinyin which have tonemark 3 occur successively in a word, the front tonemark is converted into tonemark 2.

As mentioned above, a general POS tagger splits the unknown words into individual single characters. In that case, an automatic text-to-Pinyin generator has a great chance to pick up an incorrect Pinyin for a polyphonic character in unknown words. Therefore, our unknown word prediction method can give the valuable category information for correct text-to-Pinyin conversion. In practice, we obtained more increased accuracies for the person name and location name Pinyin generation, which cannot rely on the character-Pinyin dictionary in general, using our two-phase SVM-based unknown word prediction method.

5. Experiments

In this section, we show the prediction results of Chinese person names, foreign person transliterations, and Chinese location names. We also show that the rate of the correct text-to-Pinyin conversion will increase by our proposed unknown word prediction method.

5.1. Data and preprocessing

The corpus in our experiments is composed of news articles from the People’s Daily, during a one-month period. We divide the corpus into 5 parts and conducted a 5-cross validation. We delete all person names and location names from the dictionary to test the unknown word prediction performance. There are 17,620 person names and 24,992 location names. For more efficient experiments, we pre-processed the corpus; Chinese person names were originally split into the family name and the first name in the original the People’s Daily corpus, and the compound words were also split into each component word. Therefore, we combined those split words into a single word. Then dictionaries were automatically generated from the pre-processed corpus.

5.2. Experiments and the results

The first experiment is to show how exactly our method selects the candidate boundary of an unknown word. The reduced amount of total boundaries to be recognized by SVM and the possible loss of unknown word candidates after applying our boundary detection step are shown in Table 1 and 2, for person and location, respectively.

Table 1: Reduction of the candidate boundaries (person).

<table>
<thead>
<tr>
<th></th>
<th>before</th>
<th>after</th>
<th>reduction rate</th>
</tr>
</thead>
<tbody>
<tr>
<td># of total</td>
<td>128,410</td>
<td>20,434</td>
<td>84.09%</td>
</tr>
<tr>
<td>boundary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of boundary</td>
<td>16,955</td>
<td>14,712</td>
<td>13.23%</td>
</tr>
<tr>
<td>including actual unknown words</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As shown in each table, even if a few real person names and location names are excluded from the candidates (13.23% provided by Voiceware Inc. (http://www.voiceware.co.kr)
and 3.05%), the number of total boundaries for SVM to predict is drastically reduced by 84.09% and 76.75%, respectively. We confirmed through our experiments that those missing real candidates do not much affect the overall performance for final SVM-based prediction.

Secondly, we tested the overall performance of the SVM-based unknown word prediction on the result of the candidate boundary detection. We divided the test corpus into 5 parts and evaluated them by 5-cross validation. Experiment results are measured in terms of precision, recall and F-measure.

Table 4: The performance increase of an automatic text-to-Pinyin conversion.

<table>
<thead>
<tr>
<th></th>
<th>before</th>
<th>after</th>
<th>reduction rate</th>
</tr>
</thead>
<tbody>
<tr>
<td># of total Pinyin</td>
<td>2,214</td>
<td>2,214</td>
<td>93.95%</td>
</tr>
<tr>
<td># of error</td>
<td>134</td>
<td>79</td>
<td>96.43%</td>
</tr>
<tr>
<td>correct conversion rate</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Prediction performance for person names and location names.

<table>
<thead>
<tr>
<th></th>
<th>precision</th>
<th>recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>person name</td>
<td>88.06%</td>
<td>90.96%</td>
<td>89.49%</td>
</tr>
<tr>
<td>location name</td>
<td>90.93%</td>
<td>91.34%</td>
<td>91.14%</td>
</tr>
</tbody>
</table>

Table 3 shows the final results of the SVM based prediction for person names and location names. The result of the prediction is quite promising; Recall as well as the precision is very high compared with the previous results in similar environments. So, we can verify that SVM-based method using character features is a good approach for Chinese unknown word prediction. And such additional features as Chinese family names, transliterated foreign names and the last characters of the location names, help to increase the performance of the prediction. Since our SVM was trained by somewhat unbalanced data, there were some over-predicted results in the initial output. In practice, the data of class O are about three times than other classes in the prediction step. So, in the result of the prediction of the position tag for each character, the F-measure of class O is about 97%; But the F-measures of class B and E are about 92% and 94%, respectively. That is, other classes other than class O are somewhat mis-classified into class O. These over-predicted results for each character have bad effect on the prediction result for real words. Therefore, our heuristic rule-based post-processing also plays a major role to increase the final performance.

Finally, we present how much the performance of the text-to-Pinyin conversion is increased by predicting the unknown words. Table 4 shows the increment of the correct rate of Chinese text-to-Pinyin conversion by our proposed method. The target data is above-mentioned speech script which is composed of 2,227 sentences. The number of person names and location names in the data is 860 and 631, respectively. As shown in Table 4, the correct rate of conversion is increased by about 3% and 2%, respectively. Even though there are very small amount of person names and location names in our speech script data, nevertheless we achieved better conversion results by using our proposed method. If we apply our unknown word category prediction method to the corpus which includes a large number of person names and location names, we will be able to obtain a more significant better result.

6. Conclusions

The unknown word problem has remarkable effects on the accuracy and the naturalness of the sound in Chinese TTS systems. In this paper, we present a two-phase method for high speed unknown word prediction to be practically usable in a TTS system. We first pre-detect the candidate boundary of the unknown words from the result of Chinese segmentation and tagging. Next, we predict the category of unknown words using the support vector machine. Experimental results are very promising by showing high precision and high recall along with high speed. Moreover, we can obtain the increment of the correct rate of our automatic text-to-Pinyin conversion in practical commercial data.

7. Acknowledgements

This research was supported by grant No. (R01-2003-000-10181-0) from the Basic Research Program of the Korea Science and Engineering Foundation.

8. References


