Chapter 62
Feedback-Driven Refinement of Mandarin Speech Recognition Result Based on Lattice Modification and Rescoring

Xiangdong Wang
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

Hong Liu
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

Yang Yang
Jiangsu Enterprise Information Operation Center, China Telecom Corporation Limited, Nanjing, China

Yueliang Qian
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

Duan Jia
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

ABSTRACT

In real world applications of speech recognition, recognition errors are inevitable, and manual correction is necessary. This paper presents an approach for the refinement of Mandarin speech recognition result by exploiting user feedback. An interface incorporating character-based candidate lists and feedback-driven updating of the candidate lists is introduced. For dynamic updating of candidate lists, a novel method based on lattice modification and rescoring is proposed. By adding words with similar pronunciations to the candidates next to the corrected character into the lattice and then performing rescoring on the modified lattice, the proposed method can improve the accuracy of the candidate lists even if the correct characters are not in the original lattice, with much lower computational cost than that of the speech re-recognition methods. Experimental results show that the proposed method can reduce 24.03% of user inputs and improve average candidate rank by 25.31%.

DOI: 10.4018/978-1-7998-2460-2.ch062

1. INTRODUCTION

In recent years, considerable progress has been made in automatic speech recognition (ASR) technology, and applications such as speech assistants and speech input systems are becoming popular. However, in the state-of-art systems, recognition errors remain inevitable, due to environmental noise, accent, specific domain or topic, etc. In many cases, only a few errors can change the meaning of the sentence completely, which greatly affect the user's experience and the feasibility of the ASR technology.

To improve the feasibility of ASR systems, some researchers try to incorporate human-computer interaction technologies into ASR systems and allow the user to provide feedbacks (such as verification and correction) of the recognition results through human friendly interfaces. Many interaction methods for user feedback have been proposed, such as multi-modal interaction combining keyboard, re-speaking and handwriting (Oviatt, Cohen, et al., 2000) and candidate list (also known as alternative list) selection (Ogata and Goto, 2005; Nanjo, Akita and Kawahara, 2006; Cardinal, Boulianne, et al. 2007; Vertanen and Kristensson, 2011). In recent years, word candidate list has become the most popular interface for user feedback. In the interface, a candidate list is provided for each word in the recognition result, and when the 1-best result (namely, the top-1 candidate) is not correct, the error may be corrected by selecting candidate words in the candidate list. This correction method is user friendly and can greatly improve the efficiency of error correction.

For the generation of candidate lists, word confusion network (CN) [Xue and Zhao, 2005] extracted from the N-best lattice is widely used for languages such as English [Vertanen and Kristensson, 2011] and Japanese [Ogata and Goto, 2005]. For an utterance, a sequence of candidate lists can be obtained directly from the CN, with each candidate list providing alternative words (if any) besides the top-1 word. However, for the Chinese language, the word CN is not the best choice. In Chinese, words are formed by characters and most characters can be words by themselves while they are also included in multi-character words. Therefore, in candidate lists obtained from the word CN, a character may be repeatedly included in different candidate words in a candidate list or even in different candidate lists. This makes the candidate lists redundant and sometimes confusing to the user. To solve this problem, in our earlier work, candidate lists based on Chinese characters is introduced and a method for generation of the candidate lists is proposed [Li, Wang, et al., 2009]. In the candidate lists generated, each candidate is a Chinese character, and characters competing for each other is organized in one candidate list. This makes the interface present more information with limited candidates and be much friendlier to Chinese users.

Although efficiency can be improved by using the interface based on candidate lists, the procedure of error correction is still labor intensive and time consuming. That is because the user need to check the candidates carefully before making selection, and to input by keyboard when the correct words are not in the candidate lists. Therefore, improving the accuracy of speech recognition and reducing user corrections, in particular keyboard input operations, is currently the main focus of the research. For this purpose, besides improving the speech recognition engine, learning from the user feedback to refine the recognition result is also a promising approach (Nakashima, Zhang and Naka, 2009; Senay, Linares, et al. 2010; Laurent, Meignier, et al., 2011; Rodriguez, Garcia-Varea, and Vidal, 2010). In this paper, this scheme is referred to as feedback-driven refinement of recognition result. Under the scheme, when a user provides feedback (keyboard input, candidate selection, marking the result as correct or incorrect, etc.) to the recognition result of an utterance, the correct part in the result that has been confirmed by the user can be used as context constraints, and a refined recognition result of the part without feedback can be generated using the constraints. This method can automatically correct errors by exploiting user