JAIST Repository
https://dspace.jaist.ac.jp/

Title	話し言葉音声認識のためのトリガーペアに基づく言語モデルの適応
Author(s)	Troncoso Alarcon, Carlos
Citation	
Issue Date	2006-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/969
Rights	
Description	Supervisor:党　建武，情報科学研究科，博士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Trigger-Based Language Model Adaptation for Conversational Speech Transcription

Carlos Troncoso Alarcón
School of Information Science,
Japan Advanced Institute of Science and Technology

March 2006

Abstract

Recently, the major target of automatic speech recognition research has shifted from dictation of document-style sentences to transcription of spontaneous conversational-style speech. Research in this field is still immature, and the current recognition accuracy is low. Language models play a crucial role in automatic speech recognition since they provide effective constraint and preference for possible word sequences. The most widely used language model is the n-gram model, which is powerful in modeling short-distance dependencies between words, but cannot capture long-distance dependencies because it relies on a word history limited to $n-1$ words. This thesis addresses the trigger-based language model, which incorporates long-distance topic constraints by means of related keywords, called *trigger pairs*. Meetings and conversations, which are the main target of this study, are centered in a topic in many cases, so the trigger pairs could capture long-distance topic constraints. The trigger-based language model is also insensitive to disfluencies, which are prominent characteristics in conversational speech, because it focuses on the co-occurrence of topic keywords.

However, reliable statistical estimation is the most critical problem for this kind of long-distance language model, especially for spontaneous speech, where only a small amount of training data is available compared with document-style language. This work proposes two methods to fully exploit the available in-domain data to adapt the trigger-based language model to conversational speech. Here, task-dependent trigger pairs are extracted that match more closely the addressed task. In addition, to enhance the reliability of probability estimates derived from the small amount of data, a back-off scheme that incorporates the statistics from a large corpus is proposed.

Chapter 1 introduces the two main approaches to language modeling and the application of statistical language modeling to automatic speech recognition.

Chapter 2 reviews the major language modeling techniques and presents the concept of the proposed approach. Then, the evaluation measures for language model performance and the different ways of incorporating long-distance language models are explained.

Chapter 3 presents a trigger-based language model for the transcription of travel expressions and extemporaneous speeches on given topics. Generally in language modeling, when the training corpus matches the target task, its size is typically small, and therefore insufficient to provide reliable probability estimates. On the other hand, large corpora are often too general to capture task dependency. The proposed approach tries to overcome this generality-sparseness trade-off problem by constructing a trigger-based language model in which task-dependent trigger pairs are first extracted from the corpus that matches the task, and then their occurrence probabilities are estimated from both the task corpus and a large text corpus to avoid the data sparseness problem. In the experiments, the perplexity by the proposed model was lower than that by the conventional trigger-based model constructed from one single corpus, and 12.8% lower than the baseline.

Chapter 4 addresses the trigger-based language model for the transcription of panel discussions on political and economic issues. In meetings, the topic is focused and consistent throughout the whole session, therefore keywords can be correlated over long distances. The trigger-based language model can capture such long-distance dependencies, but the derived trigger pairs are not task-dependent if it is typically constructed from a large general corpus. The proposed method makes use of the initial speech recognition results to extract task-dependent trigger pairs and to estimate their statistics. Moreover, the back-off scheme is introduced to exploit the statistics estimated from a large corpus. The proposed model reduced the perplexity considerably more than the typical trigger-based language model constructed from a large corpus, and achieved a remarkable perplexity reduction of 44% over the baseline when combined with an adapted trigram language model. In addition, a reduction in word error rate was obtained when using the proposed language model to rescore word graphs.

Chapter 5 concludes the thesis with a summary of contributions and future directions.