ISCA Archive

http://www.isca-speech.org/archive

7th **International Conference on Spoken Language Processing**

September 16-20, 2002 Denver, Colorado, USA

Statistical Natural Language Generation for Speech-to-Speech Machine Translation Systems

Bowen Zhou (1), Yuqing Gao (2), Jeffrey Sorensen (2), Zijian Diao (3), Michael Picheny (2)

(1) University of Colorado at Boulder, USA; (2) IBM T.J. Watson Research Center, USA; (3) Texas A&M University, USA

This paper presents a statistical natural language generation scheme for trainable speech-to-speech machine translation (MT) systems. The natural language generation scheme in the translation systems is based on a maximum entropy (ME) statistical model fully trained from a corpus, allowing flexible translation outputs. In this paper, the system architecture and some of its components, including the parsing, information extraction, and translation etc are briefly overviewed, followed by the descriptions of training and search algorithms for ME based sentence level NLG within the MT context. Details of NLG including feature selection and robustness are also addressed. We have implemented the described system for translating between Chinese speech and English speech in an air travel application domain. Encouraging experimental results have been observed and are presented.

Full Paper

Bibliographic reference. Zhou, Bowen / Gao, Yuqing / Sorensen, Jeffrey / Diao, Zijian / Picheny, Michael (2002): "Statistical natural language generation for speech-to-speech machine translation systems", In *ICSLP-2002*, 1897-1900.