

論文内容の要旨

博士論文題目

DPGMM-RNN hybrid model: towards universal acoustic modeling to ASR at
different supervised levels

DPGMM-HMM: 異なる学習データ下における一般性の高い音響モデルの構築に
向けて

氏 名 WU BIN

(論文内容の要旨)

The independent development of methods for unsupervised and supervised learning induces the different treatments to the unsupervised phoneme discovery and the supervised speech recognition; the two tasks both need acoustic modeling to find patterns that form the perceptual units such as phonemes and words; the only difference is at different supervised levels. So it is reasonable to regard the unsupervised phoneme discovery as the unsupervised ASR (that finds units from speech without text). We propose to use universal acoustic modeling (instead of separated ones) of supervised and unsupervised ASR for the whole process from acoustic waveform to speech units.

The study aims to construct universal acoustic modeling for speech recognition at different supervised levels. Specifically, the work proposes the hybrid model, which combines the Dirichlet process Gaussian mixture model and recurrent neural network (DPGMM-RNN). Furthermore, the proposed approach is utilized (1) to improve phoneme categorization by relieving the fragmentation problem; (2) to extract perceptual features to improve ASR performance.

氏 名	WU BIN
-----	--------

(論文審査結果の要旨)

The automatic extraction of acoustic language unit is a fundamental problem of the speech processing. In previous research the independent development of methods for unsupervised and supervised learning induced the different treatments to the unsupervised phoneme discovery and the supervised speech recognition; the two tasks both need acoustic modeling to find patterns that form the perceptual units such as phonemes and words; the only difference is at different supervised levels.

The thesis research aims to construct universal acoustic modeling for speech recognition at different supervised levels, including supervised and unsupervised ASR for the whole process from acoustic waveform to speech units. Specifically, the work proposes the hybrid model, which combines the Dirichlet process Gaussian mixture model and recurrent neural network (DPGMM-RNN). Furthermore, the proposed approach is utilized (1) to improve phoneme categorization by relieving the fragmentation problem; (2) to extract perceptual features to improve ASR performance.

The proposed research provides a general framework for unsupervised phoneme discovery and supervised speech recognition and the modeling and implementation in various languages. A series of his research resulted in two high-quality peer-reviewed journal papers, two peer-reviewed international conference papers. As a result, the thesis is sufficiently qualified as a Doctoral thesis of Engineering.