

会議も！ 談話も！ 実シーンの会話の文書化

— 会話音声認識・検索技術 —

どんな研究？

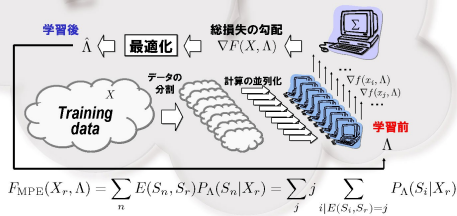
- 会議や打合せの内容を効率的に把握可能
 - 音声認識技術により会話を高精度に文書化
 - 検索により
 - ✓ 特定的话题
 - ✓ 特定の話者（話者間）
- の会話シーンを検出

もたらされる変革

- 会社で行われる把握しきれない数の打合せの中から優れたアイデアを発掘
- 多数の動画コンテンツの中から、あの言葉のあのシーンを一発検索
- 翻訳技術を併用し世界中の動画コンテンツに自動音声字幕

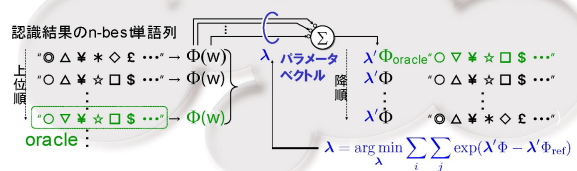
音響モデルの識別的学習

認識誤りを最小にする学習方法により、極めて高精度な音響モデルを構築



識別的言語モデルによる誤り訂正

単語・品詞・読み・音素など様々な観点から 正解 - 誤りのパターンを獲得し誤りの少ない単語列を見つけ出す



高速探索技術

1000万語の認識が可能
多くの単語列を評価することで精度も向上

$$\Omega(X \rightarrow Y_1 | T_1, T_2, T_3, \dots, T_M) = \min_{Y_1} \{ \Omega(X \rightarrow Y_1 | T_1) + \min_{Y_2} \{ \Omega(Y_1 \rightarrow Y_2 | T_2) + \min_{Y_3} \{ \dots \min_{Y_M} \{ \Omega(Y_{M-1} \rightarrow Y_M | T_M) \} \dots \} \}$$

最適なモデル構造推定

音声認識開発者の経験に頼っていたモデル構造選択を自動で最適化

モデル適応技術

用途に即したモデルを生成



従来の音声認識

一人の発話

読み上げ口調

用途が限定的になってしまう



世界最高水準の
音声認識技術



実践的な音声認識

複数人の発話

より自然な口調

講演音声を高精度に認識
会議や打合せ記録の作成も可能



関連文献

E. McDermott, A. Nakamura, "Flexible Discriminative Training Based On Equal Error Group Scores Obtained From An Error-Indexed Forward-Backward Algorithm," Interspeech'08, pp. 2398-2401, 2008.
大庭, 堀, 中村, "識別的誤り訂正学習における対立単語列と素性の選定," 電子情報通信学会技報報告, 2007-SLP-69, pp. 235-240, 2007.

連絡先: 小川厚徳 (Atsunori Ogawa), 大庭隆伸 (Takanobu Oba)

メディア情報研究部 信号処理研究グループ