2ZB-01

クラウド音声認識と会議リソースに基づく半自動議事録作成

六浦 由佳† 大平 茂輝‡ 長尾 確†

名古屋大学 大学院情報学研究科† 名古屋大学 情報基盤センター‡

1. はじめに

クラウド音声認識サービスの認識精度は年々向上しており，音声認識を用いた議事録の作成が可能になりつつある．しかし，会議コミュニティに依存する専門用語の認識は難しく，正確な議事録を作成する妨げとなっている．この問題へ対処するため，音声認識テキストから誤り箇所を検出し，正しい単語に変換することが必要である．

本研究では，誤り単語の検出，正解単語の推定，単語の置き換えという 3 つの手順で音声認識結果の修正を行う．誤り単語の検出には音声認識結果と音声書き起こし文を用いた機械学習を利用し，単語の推定と置き換えには書き起こしから得た辞書と，議事録やスライド資料を含む会議リソースを利用する．

2. 会議音声の認識と誤り傾向の分析

2.1. クラウド音声認識サービスの導入

本研究室で運用している会議システム[1]にクラウド型音声認識サービスを導入した．音声認識サービスは Google が提供する Google Cloud Speech-to-Text API（以下，API）を使用した．会議の発表者・参加者の発言音声データを API に渡して，音声認識テキスト（以下，認識結果）を取得した．また，認識結果には誤りが含まれるため，図 **1** の音声を聞きながら認識結果を修正できる音声書き起こしエディタを作成し，正解文を作成した．

2.2. 誤り傾向の分析

2019 年 6 月から 2019 年 12 月までに，参加者と会議の議題の異なる 81 回の会議を行い，14 人のデータを収集した．収集した音声データのうち

図 **1** 音声書き起こしエディタ

認識結果のある時間は 62 時間である．API からは，発言の認識結果・発言開始時刻・発言終了時刻・信頼度が得られ，認識結果中に含まれる単語ごとの開始時刻・単語終了時刻・信頼度も得られる．信頼度は 0.0〜1.0 の範囲で出力され，高いほど認識誤りの可能性が低くなる．

まず，認識結果と正解文から，全データの認識結果の単語誤り率（WER：Word Error Rate）は 13.4%となった．Chiu らの研究[2]では使用した音声認識 API の WER は 5.6%とされている．しかし，これは英語の音声の結果であり，本研究で使用した音声は日本語であったため，認識率が大きく異なると考えられる．

まず，API から取得できる信頼度と誤認識の相関を分析した．図 2 に信頼度ごとの単語数，図 3 に信頼度ごとの単語の正解率を示す．信頼度と正解率は概ね相関しているが，誤り単語数を考慮すると，信頼度の高い単語の数が多いため，信頼度を誤認識判定の指標とすることは難しいと考えられる．誤認識を判断するための，信頼度以外の指標が必要である．

単語の誤り傾向を調査したところ，置換による修正が多いことが分かった．また，認識結果を構成する単語は，延べ数が約 52 万語で，その異なり数は約 1.6 万語であった．図 4 の各単語の正解率を見ると，9 割以上正解になる単語と，9 割以上が誤認識になる単語と二極化していることが分かった．そのため，9 割以上が誤認識になる単語と，中間に位置する単語への処理を検討すればよいと考えられる．

Semi-automatic minute creation based on cloud speech recognition and meeting resources
† MUTSUURA, Yuka (mutsuura@nagao.nuie.nagoya-u.ac.jp)
‡ OHIRA, Shigeki (ohira@nagoya-u.jp)
† NAGAO, Katashi (nagao@i.nagoya-u.ac.jp)
† Graduate School of Informatics, Nagoya University
‡ Information Technology Center, Nagoya University

Copyright ©2020 Information Processing Society of Japan.
All Rights Reserved.

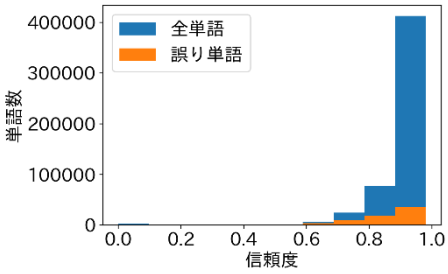

図 2　信頼度別の単語数

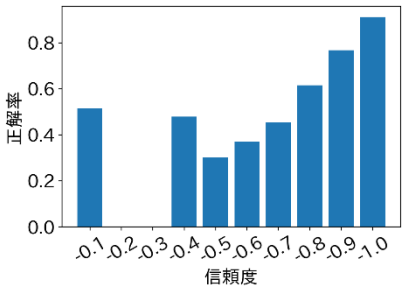

図 3　信頼度別の単語の正解率

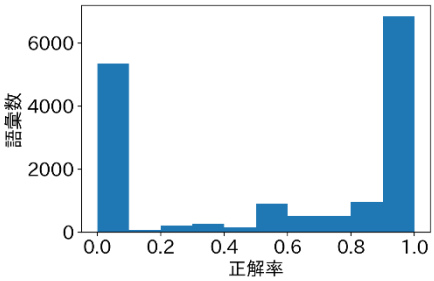

図 4　正解率別の語彙数

3.　音声認識結果の修正

3.1. 会議音声認識結果からの誤り単語検出

誤り単語検出はルールベースによる検出と BERT[3]の Fine Tuning の 2 つの手法を用いた．

BERT の Fine Tuning では，正解を一般表現，誤認識を固有表現と仮定した固有表現抽出によって，誤認識の判定をした．ルールベースでは，まず 2.2 項の知見に基づいて，正しく認識される単語，誤認識される単語，どちらか不明な単語に分類した．正しく認識される単語はそのままにし，誤認識される単語は辞書で変換し，どちらか不明な単語は会議リソースを利用して誤認識であるか判定した．

3.2. 会議リソースと音声認識結果に基づく正解単語推定

スライド等の資料を使用した会議の認識結果の修正を行うため，正解となる単語は会議リソース（現在および過去の会議分の資料を含む）に含まれると仮定する．河野ら[4]の発音記号間の距離を利用して，発音の類似性から正解単語を推定した．ただし，ここで述べる発音記号は実際の発音記号ではなく，形態素解析エンジン MeCab を用いて単語を平仮名に変換し，さらに平仮名から IPA 発音記号との照合によって，発音記号列に変換したものである．会議リソースに含まれる単語も平仮名から変換した発音記号列に変換し，発音記号列同士の距離を基に正解単語を推定した．

3.3. 単語の置き換え

単語の置き換えは，辞書を用いた変換と，3.2 項の手法により推定した単語から，適切な単語を選択して置き換える手法をとった．後者の単語の判定には BERT によって算出される尤度を使用した．表 1 に 1300 文に対して修正処理を加えた結果を示す．

表 1　修正の結果

誤り判定手法	WER 低下文数	WER 上昇文数
辞書のみ	9	1
ルールベース	6	138
機械学習	7	33

4.　おわりに

本研究では認識誤り単語の検出，正解単語の推定，単語の置き換えの 3 つの手順で音声認識結果の修正を行った．提案手法によって，辞書に含まれる単語のみを変換した結果が最も WER が改善され，他の手法は WER が悪化した．今後の課題として，精度低下の原因である，正解単語の推定と，誤認識と推定した単語の置き換え判定の改善，また，提案手法を組み込んだ自動議事録生成システムの開発を行う予定である．

参考論文

[1] K. Nagao, K. Kaji, D. Yamamoto and H. Tomobe, Discussion Mining: Annotation-Based Knowledge Discovery from Real World Activities. In Proceedings of the Fifth Pacific-Rim Conference on Multimedia, pp.522‐531, 2004.

[2] C. Chiu et al., State-of-the-art speech recognition with sequence-to-sequence models. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018.

[3] J. Devlin et al., BERT: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805, 2018.

[4] 河野宏志，城塚音也，高木徹．国際音声記号を用いた発音類似度算出アルゴリズムの検討．情報科学技術フォーラム講演論文集 13.2, pp.261-262, 2014.

Copyright ©2020 Information Processing Society of Japan.
All Rights Reserved.