

平成30年度 情報工学コース卒業研究報告要旨

長尾 研究室	氏 名	川 瀬 卓 也
卒業研究題目	会議音声からの議事録要約生成に関する研究	

企業や研究室など、議論を行う会議は様々な場所で行われている。会議で取り上げられる話題には、進捗報告や研究の指針など今後の活動に影響を与えるものが多く、会議後に内容を確認できることが望ましい。そのため、議事録の作成を必要とするが、会議における発言のすべてを記録する人的コストは非常に大きいため、多くの場合、要約した議事録を作成する。しかし、大学の研究室で行われる会議では、発言を要約した議事録の作成に必要とされる知識や経験に差があり学生には比較的困難なタスクであるため、自然言語処理技術を用いて、自動化することを目指す。会議音声から自動的に議事録を生成する基盤技術として、音声認識や自動文章要約などの研究も行われているが、要約された文章を自動的に生成することは非常に難しい。

本研究では人手に依存しない、会議音声からの議事録生成の第一歩として、RNNの sequence-to-sequence モデルを利用した、会議における発言に含まれる重要単語を抽出する手法を提案する。演算部分には Long-Short Term Memory を使用し、系列長の長いデータの情報が損なわれないようにした。図1に Encoder、図2に Decoder の構造を示す。学習データには実際に行われた会議で取得した、会議中の発言音声とその発言を人手で要約した書記テキストを用いる。発言音声は音声認識により音素列へと変換し、入力データとして用いる。書記テキストは、形態素解析によって内容語のみの単語列へと変換し、正解データとして用いる。本研究で使用した、人手によって作成された議事録には、発言内の重要な部分が要約され記述されていると考えられるため、重要単語の抽出が期待できる。

テストデータ用の音素列から抽出された単語を、対応する書記テキストの単語列と比較し、発言単位での適合率を計算することで、本研究で生成した重要単語抽出モデルの性能を評価した。適合率は、抽出した単語のうち書記テキストに含まれている単語の個数を、抽出した単語の個数で割り、百分率で表したものである。テストデータとして2,915 発言から重要単語を抽出したところ、適合率の平均値は9.61%という結果であった。学習に用いた音素系列と単語系列の長さの違いや、音声認識の精度の影響で音素列から正しい隠れ状態ベクトルが得られなかったことが低い抽出性能の原因の一つであると考えられる。

本手法を改善することにより、書記テキストには記述されていない重要な単語を抽出することや、重要単語を利用した議事録内発言の検索手法などの応用を考えており、今後は、重要単語抽出の精度向上のため、議事録における単語の出現頻度や議論構造を考慮した手法の検討、単語の重要性のより厳密な定義などを行う予定である。

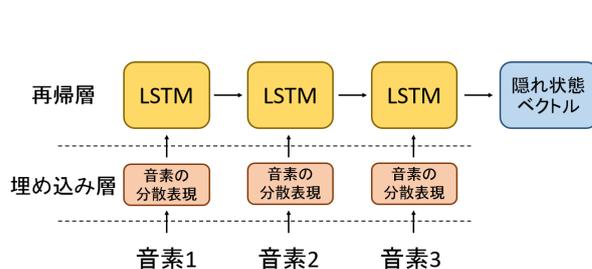


図1: Encoderの構造

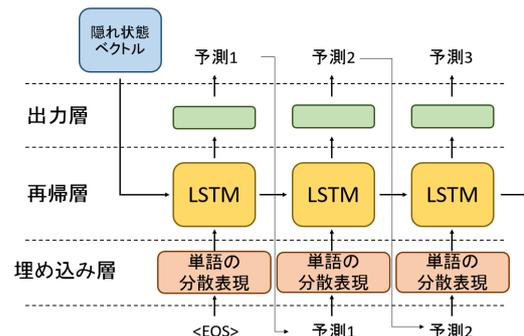


図2: Decoderの構造