様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　６　月　２２　日現在

機関番号：　１３９０４
研究種目：　基盤研究(B)（一般）
研究期間：　2013～2017
課題番号：　２５２８００６２

研究課題名（和文）日本語講義音声の英語字幕付き教材を生成するための音声翻訳に関する研究

研究課題名（英文）A study of automatic English caption generation　for Japanese lecture speech

研究代表者
中川　聖一（nakagawa, seiichi）

豊橋技術科学大学・リーディング大学院教育推進機構・特命教授

研究者番号：２０１１５８９３

交付決定額（研究期間全体）：（直接経費）　　13,000,000円

研究成果の概要（和文）：　まず、日本語と英語の講義音声のDNNによる音声認識システムを開発した。次に、英語講義音声の日本語字幕化の研究を行った。複数音声認識器の複数音声認識結果を統計的手法に基づいて機械翻訳し、言語モデル等で翻訳結果をリスコアリングすることにより、翻訳精度を向上させた。また、音声認識誤りに対処するため、音声認識器のシュミレーションによる学習データを増加させる方法の有効性を示した。予定研究期間を1年間延長し、統計的機械翻訳とニューラル機械翻訳を併用し、翻訳候補の逆翻訳に基づくリスコアリングにより、翻訳精度を大幅に向上させることが出来た。以上の手法を、日本語講義音声の英語への翻訳・字幕化にも適用した。

研究成果の概要（英文）：　First of all, we developed Japanese and English lecture speech recognition systems based on DNN. Next, we studied on English to Japanese translation for lecture speech. We transcribed English lectures by using plural English recognition systems and translated to Japanese based on a statistical machine translation system (SMT). Then, we selected the best translation from plural translation candidates by using a re-scoring technique based on language models and improved the translation result. For attacking mis-recognition errors, we simulated speech recognizers and used them as the training data for the SMT.
By 1-year extension of the research period, we developed a neural machine translation system (NMT), and combine SMT and NMT. By re-scoring based on back translation for these translation candidates, we improved remarkably the translation performance.
Finally, we applied the above techniques to the translation from Japanese lecture speech to English caption.

研究分野：　音声言語情報処理

キーワード：　講義の字幕化　講義音声　日本語講義　音声認識　日英翻訳　英日翻訳　音声翻訳

１．研究開始当初の背景

近年、グローバル化に対応した人材育成の必要の高まりにより留学生は増加する一方である。そのため、留学生に対する効果的な教育支援の提供が喫緊の課題となっている。また、ネットワーク環境の大幅な進捗にともなって、各種教育機関での講義を音声、動画およびスライドの形で大量に保存・公開する動きが広がっている。これらの日本語講義の発話音声に対して、適切な日本語の字幕や英語字幕を付与することができれば、英語に対しては理解度のある留学生に対する教育支援を効果的に行うことができると期待される。

講義や講演に対する日本語字幕付与については、聴覚障碍者や高齢者の理解を支援するため、多くの研究が行われている。しかし、講義音声や講演音声では、話し言葉的な現象（フィラーや言い淀み）が頻出するだけではなく、句点（長いポーズ）がほとんど挿入されない長文が出現したり、前後の文脈とは関係がない接続詞が使用されたりすることがある。そのため音声認識誤りも多く、認識結果をそのまま字幕として使うことは適切でない。適切な位置に改行を挿入して理解を助けるだけでなく、文の整形や要約が必要である。話し言葉である講義音声の機械翻訳の研究は、ほとんど世界的に行われていない。

２．研究の目的

本研究では、日本語講義の話し言葉音声を認識し、その整形・要約を統一的に扱うことによって、まず日本語の字幕化を行う。次に、それに基づいた統計的音声機械翻訳法によって、留学生の理解を効果的に支援するために英語による字幕を付与する技術を開発する。音声認識自体が非常に難しい上に、間投詞、言い淀み、言い直し等が多くあり、たとえ忠実に音声認識ができても、そのまま字幕化しても読み難い。本研究では、留学生の理解を支援するための字幕付与を行う。留学生の場合、聴覚障害者の場合とは異なり、リスニングによる情報も利用できるから、字幕が完全でなくても良い。重要部分が、正確に日本語や英語で字幕付与されることが重要と考えられる。本研究では、音声認識を行った後、書き言葉に整形し、重要な箇所を抽出し、①技術用語、②重要フレーズ、③重要文、④全発話、を日本語と英語で字幕化する技術を開発する。これらのすべてを確率モデル・統計的翻訳モデルで定式化して解く。

３．研究の方法

(1) 音声認識の研究

音声認識システムにおいて話者の多様性は認識精度を低下させる大きな要因となるため、システムを対象話者に適応させる話者適応に関する研究がこれまで活発に行われてきた。しかし、提案されている多くの話者適応手法は数十秒から数分程度の適応データを想定しており、短時間発話に対する適応は考慮されていなかったので、これを検討した。また音響モデルの特徴パラメータ抽出については、従来、人が設計したパラメータを用いてきたが、これを機械学習により自動設計する手法を検討した。

(2)自動字幕の表示方法の研究

字幕の表示方法として、講義音声の全ての書き起こしを字幕にするのでは、学習者が読みに集中し過ぎて、読解時間が講義の発話時間に追いつかない危険性がある。そこで、全字幕、重要文だけの字幕、重要句だけの字幕、キーワードだけの字幕、字幕なしの効果を比較検討する。

(3)ヒトの講義音声聴き取りと翻訳能力

講義の聴講者の音声聴き取り率と翻訳性能を調査し、機械による音声認識と翻訳性能が役立ちうるか検討する。

(4)機械翻訳の研究

本研究開始時の機械翻訳手法は、統計的機械翻訳手法であったため、この手法をベースに、言い淀みや間投詞などのフィラーの除去などの前処理と高頻出語や専門用語の翻訳を重点的に検討した。なお、本研究の最後には、最近技術進展が目覚ましいニューラル機械翻訳手法も検討項目に加えた。

４．研究成果

(1)音声認識の研究成果

本研究では、短時間発話を対象とした話者適応技術の提案を行った。学習データのクラスタリングを基に話者クラスを定め、この話者クラス群をモデル化した混合ガウス分布と発話との間の対数ゆう度で話者の情報を表現する。これらの対数ゆう度を話者情報として使用し、かつ話者情報推定に使用する発話長を発話先頭 0.5 秒と制限することで、短時間発話認識のための話者適応技術の提案を行った。評価実験の結果、話者情報を音響特徴量とともに DNN(Deep Natural Network) へ入力することで、話者情報を使用しない場合と比較して 7%の相対誤り削減率を得ることができ、短時間発話に対する本手法の有効性が明らかになった。

次に本研究では、DNN の最下層に特徴抽出を行うフィルタバンク層を導入し、ガウシアン形状のフィルタおよびガンマトーン形状のフィルタをもつ DNN を対象とした話者適応において、有効であることを示した。また、フィルタ形状の比較も行い、多くの場合ガンマトーンフィルタ形状が世界的に使用されている 3 角形状よりも良い性能を示した。

(2)字幕表示の研究結果

本研究では、日本語字幕および英語字幕の

様々な表示方法を比較し、講義ビデオにおける字幕表示の有用性について評価した。日本語講義音声および英語講義音声に対する字幕の表示方法として、全文字幕、重要文字幕、重要句字幕、キーワード字幕、および字幕なしを比較・検討した。日本語字幕に対しても英語字幕に対しても全文字幕や重要文字幕が理解度や補助に有用であるが、重要句の字幕も、これらと比べて劣らないことが分かった。字幕の自動化を考慮すると、全文の字幕だけでなく、重要文字幕や重要句幕を開発していくのがよいことを示した。

(3)ヒトの講義音声聴き取りと翻訳性能の調査結果

学生による英語講義への理解度を調査するために、講義の書き起こしに対する翻訳実験と講義の音声に対する聞き取りとその翻訳実験を行った。その結果、TOEIC 700 点程度の学生でも、英語講義音声の聞き取り率は、単語単位換算で約 60%程度、TOEIC 500 点程度の学生では、50%以下であった。また、日本語への翻訳性能に関しては、正しく書き起こされた文に対して TOEIC 700 点程度の学生では、我々のシステムの翻訳性能と同等であった。TOEIC 500 点程度の学生では、機械の性能よりも悪く、機械による字幕化が有効であることを示した。

(4)機械翻訳の研究成果

音声翻訳を困難にしている問題点として、自動音声認識(ASR)の出力における音声の誤認識があげられる。我々のベースラインである英語-日本語の話し言葉翻訳システムは、DNN-HMM に基づいた ASR と、対象外ドメインである比較的大規模な講義(TED)と少ない対象ドメインである講義のパラレルコーパスを用いた統計的機械翻訳(SMT)によって構成されている。

初めに、日本語講義音声の英語への翻訳システムの開発を試みた。間投詞やフィラーなどの話し言葉の整形後、専門用語の対訳語の追加などを試みたが、予想以上に困難であったために、まず英語講義音声の日本語への翻訳システムの開発を行うことにした。

本研究では、SMT に対する ASR の認識誤りへの影響を軽減する適応を行った。ASR の誤りの特性を考慮し、認識誤りに適応するために、実際の ASR の認識結果を SMT の学習に利用した。また、書き起こしから疑似的な音声認識誤りを伴った ASR 出力を作成し、同様に学習に利用した。音声認識誤り付きのパラレルコーパスを SMT の学習コーパスに対して追加するか、学習済みのフレーズテーブルに誤り付きのコーパスのみを用いて学習したフレーズテーブルを統合する形で利用した。これらの音声認識誤りに対する適応を行った英日翻訳システムについて MITOCW（MITOpenCourseWare)の講義の書き起こし、および講義音声の ASR の出力を翻訳した結果、翻訳性能が向上することを示した。次に、SMT による翻訳候補分をニューラルネットワークベースの言語モデル等により、リスコアリングする手法を検討した。その結果、複数の音声認識器による音声認識結果に対する翻訳候補をリスコアリングする手法が効果のあることを示した。

近年、ニューラル機械翻訳(NMT)が目覚ましい発展を遂げており、従来の統計的翻訳機械翻訳(SMT)の性能を上回っている。本研究の目標である日本語講義音声の英語への翻訳には、もう一段翻訳性能の改善を要した。そこで、研究期間を 1 年延長し、本研究の終盤に、NMT の検討を行った。NMT は SMT に比べ、学習に必要なパラレルコーパスの量が十分でなければ、翻訳性能を向上させることが難しく、翻訳の語彙サイズについても制限を持つ。本研究では同じパラレルコーパスで学習した NMT と SMT の翻訳文を比較し、この両者が補完的な翻訳候補を出力することを明らかにした。そこで翻訳候補のリスコアリング手法として、文の分散表現ベクトルを利用する方法と翻訳候補結果の原文への逆翻訳に基づく手法を提案した。特に逆翻訳によるリスコアリングでは、比較的小規模なロイター記事の翻訳タスクと比較的大規模な論文抄録の翻訳タスク AEPECT で有効性を確認し、リスコアリングなしの場合のベースラインの翻訳性能を大幅に上回った。また、MIT の英語講義音声の日本語への翻訳においても、逆翻訳によるリスコアリングの有効性を示した。

以上の検討結果に基づいて、再度、日本語講義音声の英語への翻訳を行った。しかしニューラル翻訳を用いても日本語講義音声の英語への翻訳は困難であり、むしろ統計的機械翻訳の方がよかった。今後、なお一層の改善を要する。

５．主な発表論文等

〔雑誌論文〕（計４件）

①関博史、榎並大輔、朱発強、山本一公、中川聖一、話者クラスタリングに基づく短時間発話音声認識、電子情報通信学会論文誌、査読有、100－D巻、2017、PP81-92

②井佐原均、多言語情報発信シンポジュウム、AAMT ジ　ャーナル、査読無、59 巻、2015、pp.33-39

③井佐原均、国際競争力の強化に今、求められるもの-TKUN の提案-、JAPIO YEARBOOK、査読無、2015 巻、2015、pp.80－81

④Aditra Arie Nugraha, K.Yamamoto, S.Nakagawa,Single-channel dereverberation by feature mapping using cascade neural network for robust distant speaker identification and speech recognition、EuraSip Journal on Audio,

Speech and Music Processing、査読有、2014巻、2014
pp.1-31DOI:10.1186/1687-4722-2014-1

〔学会発表〕（計 33 件）
①佐橋広也、西村友樹、秋葉友良、中川聖一、統計的翻訳とニューラル翻訳による翻訳候補の文の分散表現と逆翻訳に基づくリスコアリングの検討、情報処理学会、音声情報処理研究会、自然言語研究会、2018年

②西村友樹、秋葉友良、塚田元、大規模単語資源を用いた大語彙ニューラル翻訳、言語処理学会、第24回年次大会、2018年

③佐橋広也、西村友樹、秋葉友良、中川聖一、統計的翻訳とニューラル翻訳による翻訳候補の文の分散表現に基づくリスコアリングの検討、言語処理学会、第24回年次大会、2018年

④関博史、山本一公、秋葉友良、中川聖一、大規模データベースCSJを用いたDNNに基づくフィルタバンクの学習の評価、日本音響学会、秋季研究発表会、2017年

⑤V.Ferdiansyah, Seiichi Nakagawa, Captioning methods of lecture videos for learning in English, Proc.25th ICCE, 2017年

⑥T.Nishimura, T.Akiba, Addressing unknown word problem for neural machine translation, ICAICTA,2017年

⑦K.Sahashi, N.Goto, H.Seki, K.Yamamoto, T.Akiba, S.Nakagawa, Robust lecture speech translation for speech misrecognition and its rescoring effect from multiple condidates, ICAICTA, 2017年

⑧後藤統興、山本一公、中川聖一、音声認識誤りを考慮した英語講義音声の日本語への音声翻訳システムの検討、言語処理学会、第23回年次大会、2017年

⑨後藤統興、山本一公、中川聖一、英日講義音声翻訳に対する音声認識誤りを考慮したパラレルコーパスの利用、情報処理学会、音声言語情報処理研究会、2016年

⑩後藤統興、山本一公、中川聖一、対象ドメインの高頻出句に対する人手対訳追加による講義音声翻訳検討、情報処理学会、音声言語情報処理研究会、2016年

⑪R.Minamiguchi, M.Tsuchiya, Developing corpus of lecture utterances aligned to slide components, Proc.COLOING Workshop on Asian Language Resources, 2016年

⑫K.Saito, E.Yamamoto, M.Ueno, K.Kanzaki, H.Isahara, Extraction of phrases useful for machine translation, Proc. ICAICTA, 2016年

⑬H.Mizukami, T.Akiba, Effects of class-based statistical machine translation on unknown names, Proc. ICAICTA, 2016年

⑭H.Seki, K.Yamamoto, S.Nakagawa, A deep neural network integrated with filterbank learning for speech recognition, Proc. IEEE ICASSP, 2017年

⑮N.Goto, K.Yamamoto, S.Nakagawa, Domain adaptation of a speech translation system for lectures by utilizing frequentry appearing parallel phrases in-domain, Proc. APSIPA, 2016年

⑯野村高弘、塚田元、秋葉友良、ベトナム語翻訳への教師なしバイリンガルトークナイザの適用、言語処理学会年次学会、第22回年次大会、2016年

⑰後藤統興、山本一公、中川聖一、対象ドメイン内高頻出句の対訳作成による講義音声翻訳の検討、日本音響学会、春季研究発表会、2016年

⑱中川聖一、音声処理技術がヒトの能力を超える日、電子情報通信学会、音声研究会（招待講演）、2015年

⑲Takahiro Nomura, Hajime Tsukada, Tomoyoshi Akiba、Improvement of word alignment models for Vietnamese-to-English translation、Int. Workshop on Spoken Language Translation、2015年

⑳Norioki Goto, Kazumasa Yamamoto, Seiichi Nakagawa, English to Japanese spoken lecture translation system by using DNN-HMM and phrase-based SMT、ICAICTA、2015年

㉑後藤統興、山本一公、中川聖一、英語講義音声の認識と日本語への翻訳への検討、日本音響学会、春季研究発表会、2015年

㉒関博史、山本一公、中川聖一、年齢性別クラスタリング情報を考慮したDNN-HMMによる音声認識の検討、日本音響学会、春季研究発表会、2015年

㉓川口亮、山本一公、中川聖一、講義音声の

前処理と2段階翻訳に基づく日英音声翻訳、情報処理学会、音声言語処理研究会、2015年

㉔A.Abe, K.Yamamoto, and S.Nakagawa, Robust speech recognition using DNN-HMM acoustic model combining noise-aware training with spectral subtraction, Proc. Interspeech, 2015年

㉕Takahiro Nomura, Tomoyoshi Akiba、Pivot Translation using source-side dictionary and target-side parallel corpus towards MT from resource-limited languages、ICAICTA、2014年

㉖Hitoshi Isahara, Natural Language Processing and Language Resources、ICAICTA、2014年

㉗Hitoshi Isahara, Machine Translation for Intensification of the International Competitiveness、CJNLP、2014年

㉘Hiroshi Seki, Kazumasa Yamamoto, Seiichi Nakagawa, Comparison of syllable-based and phoneme-based DNN-HMM in Japanese speech recognition, Proc. ICAICTA,2014年

㉙Veri Ferdiansyah , Seiichi Nakagawa, English to Japanese spoken language translation system for classroom lecture, Proc. ICAICTA,2014年

㉚福島太喜、秋葉友良、講義音声翻訳における話し言葉の整形と翻訳の同時最適化法の検討、日本音響学会、春季研究発表会、2014年

㉛ Hitoshi Isahara, Toward practical use of machine translation, TAUS Tokyo Executive Forum 2013、2013年

㉜土井佑也、フェルディアンシャーヴェリ、中川聖一、留学生のための日本講義ビデオの字幕表示方法の比較、日本音響学会、秋季研究発表会、2013年

㉝フェルディアンシャーヴェリ、中川聖一、外国語（英語）講義映像に対する字幕提示の理解度効果、情報処理学会、音声言語情報処理研究会、2013年

〔図書〕（計２件）

① 寺嶋一彦（監修）、中川聖一、他97名、情報機構、今後の超高齢化社会に求められる生活支援ロボット技術、2015、622

② 井佐原均、JAPIO YEARBOOK、我が国の国際競争力強化のための機械翻訳の実活用、2014、583

６．研究組織

(1)研究代表者

中川　聖一(NAKAGAWA, Seiichi)

豊橋技術科学大学・リーディング大学院教育推進機構・教授

研究者番号： 20115893

(2)研究分担者

井佐原　均 （ISAHARA, Hitoshi）

豊橋技術科学大学・情報メディア基盤センター・教授

研究者番号：20358881

(3)研究分担者

秋葉　友良（AKIBA, Tomoyoshi）

豊橋技術科学大学・大学院工学研究科・准教授 、

研究者番号：00356346

(4)研究分担者

土屋　雅稔（TSUCHIYA Masatoshi）

豊橋技術科学大学・情報メディア基盤センター・准教授

研究者番号：70378256

(5)研究分担者

山本　一公（YAMAMOTO, Kazumasa）

中部大学・工学研究科・准教授

研究者番号：40324230