

氏名	米持 幸寿
学位名	博士（システム情報科学）
学位記番号	第61号
学位授与年月日	令和4年3月22日
学位論文題目	音声対話インターフェースにおける自然言語処理の大規模語彙獲得手法
論文審査委員	主査 大場 みち子 副査 新美 礼彦 副査 村井 源 副査 志築 文太郎（筑波大学 教授）

論文要旨

本論文は、2019年から現在まで公立ほこだて未来大学大学院在学中に行ってきた音声対話インターフェースの語彙獲得に関する研究成果である。

音声アシスタントを具備するスマートフォンや家庭用スマートスピーカーの普及に観られるように、音声対話インターフェースが急激に一般的になりつつある。音声対話インターフェースとは音声を使ってアプリケーションプログラムを利用することができるようにする仕組みである。様々な技術の構成が考えられるが、本研究では音声認識によりテキストデータを作り言語処理する構成を対象としており、その言語処理が題材である。アプリケーションプログラムが読み取る必要のある語を入力テキストデータ中から見つけ出す「語の抽出処理」が一回の対話に対する最初の処理である。その処理ではあらかじめ語彙を保持しているが、その語彙に含まれない語を話しかけられた時の、語の抽出処理エラーを削減することを目的として、外部の語彙を活用する際の問題解決手法を提案する。

自然言語処理分野において、テキストデータから属性を考慮して語を抽出する処理を「固有表現抽出」と呼び、音声対話インターフェースに用いることができる。固有表現抽出は自然言語処理として長年研究されてきているが、従来研究ではあまり考慮されて来なかった音声対話インターフェースならではの特別な要件がある。幅広い語彙を獲得すること、常時継続稼働すること、新語を語彙に即時反映すること、応答性能を確保することである。

音声対話インターフェースでは、コマンドプロンプトや GUI などの従来インターフェースと比較すると入力内容の自由度が高く、幅広い語彙を獲得する必要がある。語の抽出

に深層学習を使う場合、訓練データに含まれない語の抽出精度を向上させる一般的な方法として再学習か追加学習が考えられるが、システムの再起動が避けられない。従来研究では大量文書进行处理することが多く、一旦処理を行ったあと抽出に失敗した箇所をリスト化し、語彙として追加して再度処理する方法がとられることがある。しかし、音声対話インターフェースの処理において、一つの入力テキストに対してパラメーターを変更して複数回処理することは認められず、語彙は発生後即時に取り入れられることが求められる。ヒトが音声対話を行う際は、特別にレスポンスタイムに厳しい要求があるため、それを考慮する必要もある。

上述の要件を考慮しつつヒトが使う可能性のある幅広い語彙を獲得するため、1. 複数の言語資源を統合して使うことの有効性を示すこと、2. 深層学習において再学習なしに語彙を利用する手法を提案すること、3. レスポンスタイムを考慮しつつ外部検索を行う手法を提案すること、の三つが本論文の研究課題である。

複数の言語資源を統合して使うことの有効性を示す。音声対話インターフェースにおける語の抽出では「これは映画題名」「これは魚の名前」などと属性を意識する必要がある。これらの情報を保持する言語資源としてシソーラスやオントロジー辞書が必要になることがある。完全無欠な言語資源を作ることは困難であり、補完関係になることが予測されることから、これらの言語資源を同時に使うことでより多くの語彙が獲得できると仮説できる。世の中には多種多様な言語資源が存在するが、学術研究に利用しやすいものとして日本語 WordNet, DBpedia, Wikidata を取り上げる。両語彙セットを統合することで語彙が補完関係にあることを確認し、両語彙セットを同時に利用する価値を二つの検証で示す。一つ目は、対話コーパスの談話テキストを形態素解析によって分解し、隣り合う語の組み合わせが日本語 WordNet や DBpedia にどれくらい見つかるかを調べる。これにより、形態素解析だけでは語の抽出には不十分で、複合語を取り扱う必要があることを示す。二つ目は、日本語 WordNet および Wikidata の概念構造を利用して、食材、魚分類などの名称を双方で検索し、相互補完性があることを明らかにする。

深層学習固有表現抽出器で再学習なしに外部語彙を利用する手法として、入力に語彙特徴量を追加することを提案し、その有効性を示す。訓練データに含まれない語彙を追加する手法として追加学習や再学習が一般的であるが、大きな計算コストと再起動が必要である。固有表現抽出器には周辺語を利用して未知語を抽出する能力もあるが精度は低い。本研究では、システムに対して外部から語彙を与えることで問題の軽減を図る方法を提案する。最新の技術として Google 社が発表している言語処理モデル BERT を取り上げる。深層学習を利用する未知語としての語の抽出において、訓練データに含まれる語より長い語の抽出エラーが多いことに着目し、実際にその現象が発生することを確認する。現象が発生している状況に対して語彙情報を与えることで問題が軽減することを確認するため、入力に語彙情報としてバイナリ特徴量を追加し、効果を明らかにする。これらの状況が現実社会で起こり得ることを示すため、Wikidata から著作物名を収集し、その長さが

年々増加していることを確認し、本提案で示すような対策が現実に必要な場合があることを示す。特に日本の日本語のマンガとアニメの題名は、米国の英語のものよりも優位に長いことを示す。

レスポンスタイムを考慮しつつ外部検索を行う手法を提案し、有効性を示す。商品データベースなどは日々更新されるため、その度に深層学習の抽出器を再訓練することは現実的でない。音声対話インターフェースを提供するシステムから外部にあるデータベースを即時検索して語彙を利用する場合を考える。入力文からデータベースにある語を見つけ出すには、非常に多くのクエリを実行する必要があるため、レスポンスタイム要求に対して大きな障壁となる。即時検索する際に発生するクエリの回数を削減するために、日本語形態素文字種境界（JMCTB）法を提案する。音声対話インターフェースを対象とするため、話し言葉コーパスを使って実験を行い、効果があることを示す。

三つの研究成果を説明した後、それらを総合的に活用する音声対話インターフェースのシステムアーキテクチャーを提案する。外部の語彙データや語彙に使える外部システムをオンラインで結合し、音声対話インターフェースの要件を満たしつつ語彙を増強することができるようになることを説明する。提案するアーキテクチャーにより、あらかじめシステム語彙に含まれていない外部語彙を直接活用でき、新語にも即時対応が可能であることを示す。これにより、音声対話インターフェースが外部語彙を獲得する技術に大きく寄与するものであると考える。

審査結果の要旨

音声対話インターフェースの自然言語処理において語彙不足によるエラーの発生を抑制する方法の研究について述べている。幅広い語彙を獲得する、常時継続稼働する、新語を即時反映する、応答性を確保する音声対話インターフェース特有の要件を満たすため、3つの研究を行った。これらの3つの研究を総合的に活用する音声対話インターフェースのシステム・アーキテクチャーを提案することで、あらかじめシステム語彙に含まれていない外部語彙を直接活用でき、新語にも即時対応が可能であることを示した。

本論文では第1章を序論として本研究の背景と音声対話における要件、研究目的と課題を示している。第2章では当該分野の関連技術、第3章では当該分野の関連研究についてまとめている。

第4章では日本語WordNetおよびDBpedia/Wikidataの利用価値検証を、形態素解析結果に対する複合語の発見数と、魚介類、食材、料理名の相互補完性という二つの観点で検証した結果を述べている。

第5章では語彙特徴量を付与することで長い未知語の抽出エラーを改善する手法を提案し、効果を示している。Wikidataから収集した著作物題名において年々長くなっていること、

日本語の漫画および映画の題名が長いことを示している。

第6章では、日本語形態素文字種境界（JMCTB）法を提案し、N-gram形態素を検索する場合に比べ96%の検索量が削減できることを示している。

第7章ではこれら第4章から第6章の三つの研究成果を総合して使う統合アーキテクチャーを説明し、語彙エラーの削減に貢献することを示している。

以上により、本論文は音声対話インターフェースにおいて、語彙不足によるエラー発生問題を解決する新しい語彙獲得手法を開発し、音声対話インターフェースの発展や普及に大きく寄与するものである。よって、本学の学位論文として十分値するものと判断する。