（様式５）

学 位 論 文 要 旨

西暦　2023年　1月　10日

学位申請者
（　嶋　和明　）　印

学位論文題目

タスク指向型対話システムにおける多様な自然発話を収集する手法に関する研究

学位論文の要旨

音声による機器操作は，手動による複雑な機器操作を音声で実行でき，操作性を向上させる手段の一つである．特に車載機器では human machine interface (HMI) として重要である．昨今では in-vehicle infotainment (IVI) とも呼ばれる機器は電話機能，メール送受信，天気予報など様々な機能を提供するようになった．その一方，多機能化とともに操作の複雑さが増す課題に直面している．また，車室内では操作部を設けるスペースに制約があるだけでなく，走行中の機器操作は安全面において課題がある．音声による IVI の操作が可能であれば，機能名を発話するのみで求める機能を呼び出すことが可能となり，ドライバーはハンドルから手を放さず，視線も前方から逸らす必要がなくなり，HMI の向上のみに留まらず安全面においても寄与できる．さらに，機器が受け付けられる発話が定型音声コマンドに限らず，自然な発想による発話を受け可能であれば，ドライバーの認知負荷低減にも繋がる．

昨今ではクラウドを活用した大語彙連続音声認識技術の躍進により，多様な発話も高精度にテキスト化可能となった．これにともない，話者の期待は，簡潔な定型音声コマンドベースの発話のみならず，より自然で自由な発想による多様化した発話でも音声操作可能という期待へと変化することが予想される．また，既存の機器においても技術発展に伴い「単純な機械」と認識されていたものが「高度な言語処理技術を搭載した機器」と話者が認識すれば，より自然で自由な発想による多様化した発話がなされることも予想される．

このような多様な言い回しに対応させるには，音声認識結果のテキストから話者の意図を推定する，言語理解器用の学習データ（学習用コーパス）や，その評価用データ（評価用コーパス）にも多様な言い回しが含まれている必要がある．

代表的なコーパス収集手法として Wizard of Oz (WOZ) 手法が従来から用いられている．この手法では，WOZ システムと呼ばれる実際のシステム（対話ロボットなど）を模倣した実験環境を用意する．このシステムの特徴は，被験者にはプログラムなどで制御された対話システムであるように思わせ，実際には，人間のオペレータが別室などでシステム制御する仕組みである．WOZ システムを介して被験者がシステムを操作した記録を取ることにより，未開発のシステムであっても，被験者がシステムをどのように操作するか調査可能な手法である．一方，WOZ システムは被験者に発話対象は機械と認識させる手法である．そのため，単純で機械に理解されやすい発話や，定型音声コマンドのような発話が収集されることが予想され，より自然で自由な発想による多様化した発話の収集には向かない懸念がある．

また，WOZ 手法に代表される「疑似的な環境を用いた発話収集手法」は，目的とするドメ

インの発話を収集するための手法であるため，高品質な発話例を収集できるものの，発話ログデータから収集する手法比べ，WOZ システムなどの実験環境の準備や，被験者に直接コンタクトする作業がある分，手間を要する課題がある．

以上を踏まえ，本研究では**「多様化した発話でも操作可能なシステムの開発・評価のため，より自然で自由な発想による多様化した発話例を整備すること」を目的**としている．この目的を達成するためには，発話ログデータを活用するというよりは，発話ログデータより自然で自由な発想による多様化した発話を収集する方が本研究の目的に即している．そのため，収集環境やプロセスに工夫の余地がある「疑似的な環境を用いた発話収集手法」を用いることを本研究の方針とし，以下の特徴を持つ手法を提案することを目標と定めた：

- より自然で自由な発想による多様化した発話を収集可能
- 手法の実施負荷低（時間的負荷低減）

以上を踏まえ，本論文は以下のような構成となっている．

第1章では，車載機器を例に，音声認識操作が寄与する操作性の向上について紹介するとともに，多様な発話により音声操作するためには，言語理解器の学習データ，評価データにも多様な発話を含んでいる必要があることから，本研究の目的は「多様化した発話でも操作可能なシステムの開発・評価のため，より自然で自由な発想による多様化した発話例を整備すること」と定めている．

発話例の収集手法として，従来手法の特徴をもとに分類化し，それぞれの長所短所より最終的に「疑似的な環境を用いた発話収集手法」を用いることを本研究の方針とした．また，本研究の目的とする発話を収集可能な新たな収集手法の提案と，「疑似的な環境を用いた発話収集手法」の課題である収集作業の手間を改善することを目標と定めた．

第2章では，コーパスに関する調査研究とともに，コーパス関連研究における本研究の位置づけと，本研究における「発話の多様性」の定義について述べている．

コーパスとは元来，言語資料であることから，文学，言語学分野で広く活用されている．しかし，昨今では開発資源として text to speech (TTS)，音声認識技術，言語理解器の開発にも活用されている．本研究も，コーパスを開発資源として活用する研究に分類され，本研究で扱うコーパスとは「タグ付き音声言語テキストコーパス」を意味することを示した．また，本研究における「発話の多様性」は，一発話あたりの形態素数，異なり形態素数と，その手法が多様な形態素を収集可能な手法であるかを，一発話あたりの異なり形態素数比 token / utterance ratio (TUR) という指標を定義し確認することとした．

第3章では，新たな「疑似的な環境を用いた発話収集手法」として，インタビュー形式による多様性の高い機器操作発話収集手法を提案している．ドメインはカーナビの「目的地検索」である．WOZ システムのような実験環境を用いないことで手法の手間を抑え，オペレータは被験者にインタビューする中でシチュエーションを提示し，そのシチュエーションにて用いる被験者の発話を聞き出し，書き起こしたものがコーパスとなる．

提案手法により，被験者100人から収集したコーパスと，商用カーナビの目的地検索の発話ログデータにて形態素解析による比較を行い，提案手法のコーパスの方が，一発話あたりの形態素数，異なり形態素数が多く，TUR も高いことから多様性の高い発話を収集できており，提案手法自体も多様な形態素を出現させやすい手法であることを示した．また，言語理解器を用いた推定精度の検証では，提案手法のコーパスと発話ログデータを混ぜた学習データは，発話ログデータのみの学習データより言語理解精度が向上することを示し，本手法の有用性を示した．

第4章では，3章のインタビュー形式による収集手法にて用いた，probing の有用性について検証している．本論文における probing とは簡単に言えば「他に思いつく言い方はありますか？」と，被験者に同じ質問を再度行い，1シチュエーションから複数の回答を得る手法である．本手法は，発話数を多く集める点において有効と考えられるが，被験者が1シチュエーションから想起した，第二，第三発話は，その被験者にとって，第二，第三候補の発話とも言えるため，実際には用いられない，または使用頻度の低い発話の可能性がある．

この懸念を払拭するため，従来手法にあたる被験者100人の第一発話のみで構成されたデータセットの異なり形態素数を調べ，同一の値となる提案手法のデータセットの被験者数を調べた結果，被験者59人となった．次に，発話ログデータ上に出現する形態素が，各データセットに含まれているか検証し，発話ログデータに対する形態素のカバレッジを確認することにより，有用性を検証した．その結果，提案手法の被験者59人のデータセットと，従来手法の被験者100人のデータセットのカバレッジが，同程度であることを確認した．これらにより，提案手法の被験者59人のデータセットと，従来手法の被験者100人のデータセットが同程度の品質であることが示された．

さらに，提案手法による作業負荷の低減度合いを，一発話あたりの作業時間に基づく時間的負荷指数により，従来手法と提案手法を比較した結果，提案手法の方が時間的負荷は37%少ないことを示した．

第5章では，発話対象を人間と想定した Web アンケートによるコーパス収集手法を提案している．本提案では，被験者に Web アンケート上でシチュエーションを提示し，そのシチュエーションにて用いる発話をアンケートの回答文として記載させる．この際，被験者に発話対象が人間だと教示することが提案手法の特徴である．本提案手法を「Web アンケートを活用した GiFT (gift for the Tin man) 手法」と名付けた．

収集実験では「介護を受ける」をドメインとし，まず被験者30人による小規模な実験を行った．収集したコーパスを，3章のコーパスと，形態素解析による比較を行った結果，一発話あたりの形態素数は多いものの，品詞の分布に大きな差があり，提案手法による向上ではなく，ドメイン違いによる影響が懸念された．この知見を活かした続く実験では，ドメインを「介護を受ける」に統一し，被験者も200人に増やした上で，「提案手法である発話対象を人間の介護士とした100人」のグループAと，「発話対象を介護ロボットとした100人」のグループBに分け，再度実験を行った．

時間的負荷の検証では，4章で算出した時間的負荷指数と，提案手法の時間的負荷指数を比較し，約半減していることを示した．また，形態素解析により，グループA, B のコーパスを比較した結果，一発話あたりの形態素数，異なり形態素数は，提案手法のグループAの方が多く，TUR もグループAの方が大きいことから，提案手法は多様性の高い発話を収集できているとともに，収集手法自体も多様な形態素を効率的に収集可能な手法であることが示された．

第6章では，本研究のまとめと成果，また，今後の課題について述べている．

（様式６）

Ｓｕｍｍａｒｙ

Applicant for degree ：
Kazuaki Shima

Title of thesis ：

A Study on Methods for Collecting Diverse Natural Utterances in Task-Oriented Dialogue Systems

This dissertation presents methods for collecting utterances in task-oriented dialogue systems. Utterances, which are uttered by people, change depending on a dialogue target. For example, the number of morphemes per utterance is less when a user speaks to a machine than when a user speaks to a person. In addition, Advances in cloud speech recognition technology are anticipated to further improve the accuracy of that, and more natural and diverse utterances will be accepted by equipment. Considering these factors, we devised some new methods to collect more natural and diverse utterances for future task-oriented dialogue systems. This study verified the diversity of utterances by some metrics including the number of morphemes per utterance, the number of different morphemes, and a type/utterance ratio (TUR).

First, we devised a method of collecting high-diverse utterances corpus for equipment operation in the form of interviews using probing. The method was validated by comparing the corpus of this method and voice operation log data of a point-of-interest searching service for car navigation systems. The results showed that the corpus of the devised method has a higher diversity of utterances than log data. Furthermore, the corpus was also useful training data for a language understanding module.

Second, we validated the usefulness of the probing used in the interview of the first method. The corpus using the probing had roughly the same coverage as the conventional one that collects one utterance per situation despite the small number of subjects. Furthermore, the probing had been shown to collect many useful utterances easily with less workload time.

Third, we devised the Web-questionnaire-based corpus collection method. The feature is that subjects are instructed a dialogue target is a person instead a machine. We named the method the GiFT (gift for the Tin man) method. In the validation phase, we defined receiving nursing care as a domain. Utterances were collected as text from group A, who were instructed a dialogue target is a caregiver, and group B, who were instructed a dialogue target is a nursing robot. The results showed that group A of the devised method had more diverse utterances than group B, and decreased workload time compared with the first method.