4S-03

日本語音声言語理解タスクに対する日本語 SLU モデルの活用

末次　拓斗†　　須子　統太‡

早稲田大学社会科学部†　　早稲田大学社会科学部総合学術院‡

1. はじめに

近年，音声対話システムの用途は広がり，特に Amazon Alexa，Google Assistant などの音声アシスタントが広く利用されるようになった．音声対話システムにおいて，音声言語理解(SLU)の自然言語理解(NLU)は，ユーザーの発話を理解する重要なコンポーネントである．本研究では，日本語に特化した SLU モデルの構築を目指し，いくつかの評価実験を行った．

2. 音声言語理解における自然言語理解

2.1. 意図の検出とスロットフィリング

SLU における自然言語理解は通常，テキスト化された発話を入力として，意図の検出とスロットフィリングの 2 つのタスクを行う．例えば，「明日の目覚ましをキャンセル」という発話テキストが与えられた場合，表 1 のように各チャンクにスロット(特定のカテゴリーや属性)を付与し，発話全体に意図を付与する．

表 1　意図の検出とスロットフィリングの例
(スロットの O は Others の略)

発話	明日	の	目覚まし	をキャンセル
スロット	Date	O	alarm_type	O
意図	alarm_remove			

2.2. 従来研究

SLU の分野では，様々な言語に適用できる多言語 SLU データセットと多言語 SLU モデルの開発が進んでいる．これらの研究は多くの言語に対して有望な結果を上げている一方で，一部の言語ではその言語独自の課題に直面している．

特に日本語に対する多言語モデルの性能は低く，[Jack FitzGerald, 2022]における多言語モデルの評価結果では，日本語に対する性能は意図の検出で 51 言語中 45 位，スロットフィリングで 51 位であった．この日本語に対して性能が低い原因として，日本語独自の課題が存在する可能性が考えられるが，それを検証した研究は存在しない．本研究では日本語に対する問題点として，トークン化と多言語モデルの転送能力の低さを検証し，日本語に特化した SLU モデルの構築方法を模索する．

3. 日本語に特化した SLU モデルの構築方法

3.1. 日本語に特化したトークン化の検証

通常，英語などと異なり，日本語でスロットフィリングを行うためには，文章を分割する必要がある．従来研究では，日本語文章を一文字ずつに分割しトークン化を行った[1,2]．しかし，この分割方法では学習モデルが単語や文章の意味を十分に理解できない可能性がある．

そこで本節では，より日本語に適した方法として，形態素解析を用いたトークン化を検証する．検証は，MultiATIS++[1]と Massive[2]の日本語データに対して，従来の一文字ずつのトークン化と，形態素解析を用いたトークン化を行った上で，それぞれの性能を mBERT で評価・比較した(表 2)．形態素解析には MeCab を使用し，辞書は標準的な IPA 辞書と，固有表現が多く登録されている mecab-ipadic-NEologd(NEologd)の 2 つを採用した．NEologd を採用した理由は，発話データに“あいみょん”や“シャーロット”などの固有表現が多く含まれているからである．

表 2　トークン化の方法ごとの性能比較

意図の検出(正解率)		
トークン化の方法	MultiATIS++	Massive
一文字ずつ	93.91	84.03
MeCab(IPA)	**96.61**	**85.98**
MeCab(NEologd)	**96.73**	**86.01**
スロットフィリング(F1)		
トークン化の方法	MultiATIS++	Massive
一文字ずつ	85.34	69.30
MeCab(IPA)	**86.38**	**71.65**
MeCab(NEologd)	**81.30**	**68.68**

実験の結果，トークン化方法に IPA を使用した形態素解析を用いる場合，従来の一文字ずつよりも，意図の検出とスロットフィリング両方の性能が向上することがわかる．一方で，固有表現に強い NEologd を使用するトークン化は，IPA の場合よりも，意図の検出がわずかに向上する一方で，スロットフィリングの性能は大きく低

Utilization of Japanese SLU Model for Japanese Spoken Language Understanding
† Takuto Suetsugu,Tota Suko
School of Social Sciences, Waseda University

下することがわかる．

この原因として，NEologd では IPA よりもスロットの誤分割が増加したことが考えられる．スロットの誤分割とは，トークン化の際にスロット部分を誤った形で分割してしまうことである．例えば「メキシカンレストラン」という発話が与えられた場合，正しくは「メキシカン」に food_type，「レストラン」に business_type のスロットを割り当てる．しかし，誤って「メキシカンレストラン」という一つの単語で分割した場合，その後のタスクで正しくスロットを割り当てることができない．こういったスロットの誤分割が，IPA よりも NEologd では高い割合で起きていることが確認できた(表 3)．

表 3　トークン化の方法ごとのスロット F1 とスロット誤分割率（誤分割数/スロット合計数）

	MultiATIS++		Massive	
トークン化	F1	誤分割率 (%)	F1	誤分割率 (%)
一文字ずつ	85.34	-	69.30	-
MeCab(IPA)	86.38	**0.59**	71.65	**0.73**
MeCab (NEologd)	81.30	**8.67**	68.68	**1.79**

つまり，日本語における発話テキストのトークン化は，形態素解析が性能向上に有効だが，形態素解析などで複数文字にトークン化した場合，「トークン化の段階でスロットを誤分割してしまう」という新たな問題が発生すると考えられる．このスロット誤分割はエラー伝搬を起こし，モデルのスロットフィリング F1 低下を招く．形態素解析や，より発展的な方法で日本語のトークン化器を構築する際にはスロット誤分割率は考慮すべき指標と言えるだろう．

3.2.　単言語モデルと多言語モデルの検証

SLU の分野では，多言語モデルの研究が進んでおり，英語以外の言語，特にデータが不十分な言語で有効な手法だとされている．一方で，日本語や韓国語など一部の言語に対しては多言語モデルの性能は低く，異なる言語間での転送能力が低いことが指摘されている[3]．

そこで本節では，どのような SLU モデルが日本語に対して有効なのかを検証するために，多言語モデル(mBERT, XLM-R)と日本語単言語モデル(東北大 BERT, 日本語 RoBERTa)の 4 つの SLU ジョイントモデルを構築し，性能比較を行った(表 4)．トークン化には MeCab(IPA)を使用した．実験結果から，ほぼ全ての評価値で日本語単言語モデルが，多言語モデルよりも性能を上回り，東北大 BERT が最も良い性能を示した．

表 4　多言語モデルと日本語モデルの比較

意図の分類(正解率)		
モデル	MultiATIS++	Massive
mBERT	96.61	85.98
東北大 BERT	**96.95**	**86.82**
XLM-RoBERTa	92.66	82.25
日本語 RoBERTa	**93.57**	**82.95**
スロットフィリング(F1)		
モデル	MultiATIS++	Massive
mBERT	86.38	71.65
東北大 BERT	**85.49**	**72.84**
XLM-RoBERTa	77.35	54.94
日本語 RoBERTa	**79.82**	**56.18**

4.　まとめ

本論文では，日本語に特化した SLU モデルの構築を目指し，従来モデルの課題点を検証した．検証の結果，本論文で構築した，形態素解析(MeCab, IPA)と東北大 BERT を用いたモデルでは，従来手法より意図の検出では MultiATIS++で 3.04%, Massive で 2.79%, スロットフィリングでは MultiATIS++で 0.15%, Massive で 3.54%の性能向上を達成した．また，形態素解析などによる複数文字へのトークン化は，意図の検出とスロットフィリング両方の性能向上に貢献するが，スロットの誤分割という新たな問題を引き起こすことを指摘した．

参考文献

[1] FitzGerald Jack, et al, MASSIVE: A 1M-Example Multilingual Natural Language Understanding Dataset with 51 Typologically-Diverse Languages, arXiv preprint arXiv:2204.0858, 2022.

[2] Weijia Xu et al, End-to-End Slot Alignment and Recognition for Cross-Lingual NLU, In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 5052-5063, 2020.

[3] Wasi Ahmad, et al. Syntax-augmented Multilingual BERT for Cross-lingual Transfer. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 4538-4554, 2021.