

未知語を理解可能な音声対話システム

Spoken Dialogue System Which Understands Out-of-Domain Keywords

小林優佳^{1*} 岩田憲治¹ 渡辺奈夕子¹ 吉田尚水¹ 久島務嗣¹ 藤村浩司¹
Yuka Kobayashi¹, Kenji Iwata¹, Nayuko Watanabe¹,
Takami Yoshida¹, Tsuyoshi Kushima¹, Hiroshi Fujimura¹

¹ 株式会社東芝 研究開発センター

¹ Corporate Research & Development Center, Toshiba Corporation

Abstract: This paper describes a spoken dialogue system which understands out-of-domain (OOD) slot values. The spoken dialogue system has to understand some keywords which are not included in training data to reject them and get new keywords. Context information is an important clue for detecting values because the values in a given slot tend to appear in similar contexts. This system detects values from user utterances with context information. After that, this system estimates slots of the values by a proper use of either or both keyword and context information which are divided by value detection.

1 はじめに

近年、携帯端末やウェブサイト上で自然な発話でやりとりができる音声対話システムが注目されている。ユーザの多様な発話に対応できるようにするため、発話理解モデルにはニューラルネットワークなどの統計的な手法が用いられるようになったが、統計的な手法は学習データに含まれない未知語を正しく理解することが難しい。そのため図1左のようにユーザ発話に含まれるキーワードを無視してしまう。ユーザはなぜ伝わらないのかわからないので、対話が進まない。ユーザの発話には新しい商品名、流行語など、日々生まれていく新しいキーワードが含まれるため、新規キーワードを逐次学習データに追加していくのは現実的ではない。そのため、音声対話システムにとって未知語を正しく理解するという事は非常に重要である。そこで、我々は学習データに含まれない未知語を理解できる音声対話システムを提案する。未知語を正しく理解することで、図1右のように未知語を「知らない」ということをユーザに正しく伝えることができ、ユーザは別のキーワードでシステムに伝えることができ、対話が円滑に進められる。

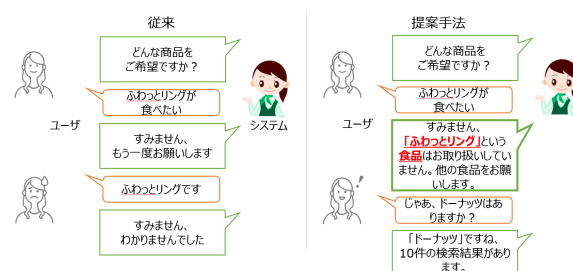


図 1: 対話例

2 提案手法

2.1 バリュエ検出器とスロット推定のパイプライン

本システムの処理の概要を図2に示す。

音声対話システムでは検索や登録処理に必要なスロット・バリュエのペアをユーザ発話から抽出するスロットフィリングを行う。スロットフィリングは発話からバリュエを検出するタスクとバリュエのスロットを推定するタスクに分解することができる。バリュエ検出に失敗するとキーワードの存在を無視してしまうので、図1右のような応答を返すことができない。しかし、スロット推定については信頼度が低ければスロットに言及せずに応答することで推定に失敗しても応答文でリカバーすることができる。そこで、本手法ではバリュエ検出の性能を最適化するために、バリュエ検出器とス

*連絡先：(株)東芝 研究開発センター
メディア AI ラボラトリー
〒212-8582 神奈川県川崎市幸区小向東芝町1
E-mail: yuka3.kobayashi@toshiba.co.jp

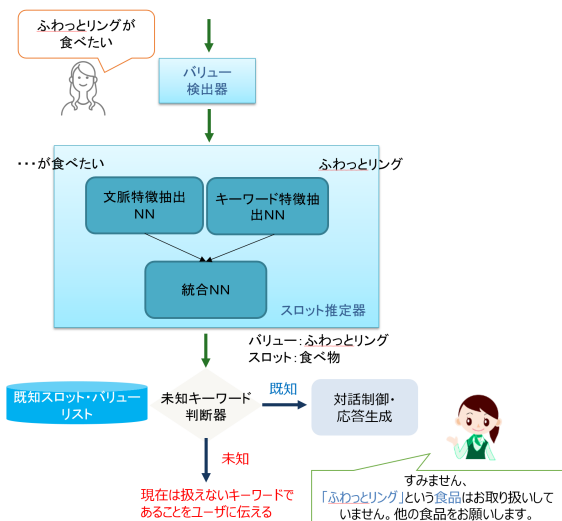


図 2: 音声対話システム

ロット推定器をわける。

2.2 バリュースロット検出器

バリュースロット検出器のモデルは RNN Encoder-Decoder モデルを使用し、時系列情報にラベルを付与するタスクとして解く [1]。未知語であっても、既存のスロットのバリューであればユーザは同じ文脈で使うことが多い。たとえば食べ物名であれば、既知バリュー「和食」の文脈「和食が食べたい」は未知語「ふわっとリング」に対しても「ふわっとリングが食べたい」のようにそのまま使うことができる。そこで、モデルが文脈を重視するような学習方法を適用する。学習データ中のキーワードをランダムに別の単語に置き換える。こうすることで、学習データ中のキーワードには共通の特徴がなくなるため、学習されたモデルはキーワードに依存しないモデルとなる。モデルはキーワード以外の部分、すなわち発話文の文脈の特徴量を重視したモデルとなる [2]。

2.3 スロット推定器

バリュースロット検出器によってどの部分がバリューかは特定できているので、スロット推定器は検出したバリューのスロットを推定することがタスクとなる。バリュースロット検出は文脈に注目することで未知語を検出することができる。しかし、スロット推定は文脈だけで解決できるとは限らない。たとえば「～が食べたい」という文脈からスロットが食べ物であることがわかる。しかし、「～がいい」という文脈からは食べ物なのか場所なのかを特定することはできない。そのためスロット推定

は文脈情報・キーワード情報を使い分ける必要がある。バリュースロット検出器によってユーザ発話はキーワード部分とそれ以外の文脈部分に分けることができる。そこで、本手法ではキーワード特徴量、文脈特徴量を別々にエンコードし、それぞれのエンコーダの出力を重みつきで使用 [3]。発話文によって重みの値を変えることで、キーワード・文脈特徴量を使い分け、既知・未知語のどちらも正しく推定できるようになる。

2.4 未知語かどうかの判定

既知スロット・バリューリストを使用して発話文から検出されたスロット・バリューが未知語かどうかの判定を行う。検出したスロット・バリューと該当するスロットの各バリューの類似度を算出し、どの既知バリューとの類似度も閾値以下の場合、未知語と判定する。類似度の算出方法には単語分散表現のコサイン類似度を使用する。

未知語と判定された際は図 2 のような応答文を返す。応答文はあらかじめテンプレートを作成しておき、テンプレートに未知語のスロット・バリューを埋めて発話し、ユーザに未知語があったことを伝える。

3 おわりに

学習データに含まれない未知語を理解し、ユーザが未知語を発話した際に、システムが理解できないことをユーザに通知し、ユーザに適切な発話を促すことができるシステムを開発した。今後は基本性能の改善に加え、より複雑なユーザ発話を理解できるよう改良を進めていく。

参考文献

- [1] B. Liu and I. Lane. Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling. In *INTERSPEECH*, pages 685–689, Sep 2016.
- [2] Y. Kobayashi, T. Yoshida, K. Iwata, H. Fujimura, and M. Akamine. Out-of-Domain Slot Value Detection for Spoken Dialogue Systems with Context Information. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 854–861, Dec 2018.
- [3] Y. Kobayashi, T. Yoshida, K. Iwata, and H. Fujimura. Slot Filling with Weighted Multi-Encoders for Out-of-Domain Values. In *INTERSPEECH*, pages 854–858, 2019.