

特定分野の日本語音声対話システムにおける 対話分類精度の比較

鈴木和樹[†] 城裕太郎[‡] 小松川浩[‡]

公立千歳科学技術大学大学院 光科学研究科[†] 公立千歳科学技術大学 理工学部[‡]

1. はじめに

近年、コンピューターの処理速度の加速度的な進化に伴い、AIの技術が進歩している。しかし、技術進歩していく中で企業の開発はクローズドイノベーションで行われることが多い。このため、特定分野毎の対話システムのように裾野の広い応用が可能なサービスに対しても技術は秘匿され、広く普及しない要因となっている。そこで本研究では、日本語音声対話システムに対し、オープンな技術でシステム構築を図るための技術所見の取得を目的とした。

本研究は、レンタカーサービスの日本語音声対話システムの開発を対象とする。開発したシステムは音声認識・音声合成・対話管理の3つで構成されている。対話管理は幾つかアルゴリズム的な検討の余地があることから、本研究ではこの点を中心に議論する。柔軟に対応できるかどうかを検証するためにルールベースや機械学習による対話分類精度を比較した。

2. システムの概要

本研究で開発した日本語音声対話システムの構成を図1に示す。このシステムの構成は①音声認識、②対話管理、③音声合成の大きく3つに分けられる。全体の処理フローとしては以下のように行う。

- I. 入力された音声を①音声認識で文章に変換して②対話管理に受け渡す。
- II. ②対話管理で受け取った文章を各アルゴリズムで解析を行い、適切な応答文を③音声合成に受け渡す。
- III. ③音声合成で受け取った文章を音声に変換して出力する。

①音声認識、③音声合成はMicrosoft Azureの

**Comparison of Speech Classification Accuracy for
Japanese Voice Dialogue System in a Specific Knowledge Domain**
Kazuki SUZUKI[†], Yutaro SHIRO[‡] and
Hiroshi KOMATSUGAWA[‡]

[†] Graduate School of Photonics Science, Public Chitose Inst. Sci. Tech.
066-0012 Chitose Japan

[‡] Faculty of Science and Technology, Public Chitose Inst. Sci. Tech.
066-0012 Chitose Japan

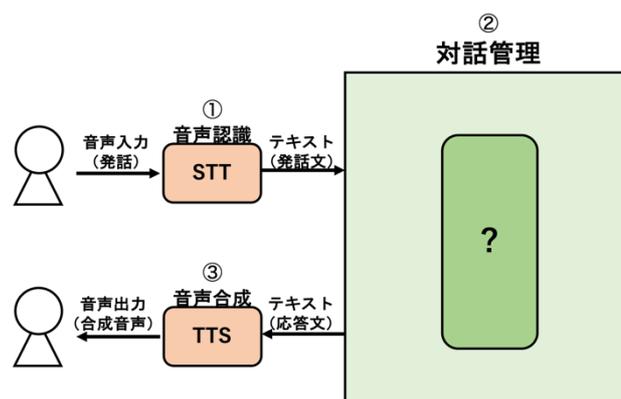


図1. システム構成図

Speech Servicesの中でSpeech to Text (STT), Text to Speech (TTS)をそれぞれ用いた。音声認識・音声合成に関しては、STTやTTSを用いることで遅延が少なく、変換ミスが少ないため、採用した。

②対話管理で用いるアルゴリズムとして、ルールベースによる手法と機械学習による2つの手法を採用した。まず、ルールベースは発話-応答のルールを記述する必要があるため、本研究ではキーワードマッチングによるルールを設定した。キーワードマッチングでは、入力文で期待されるキーワードが2つ以上含まれている場合に、対応する応答文を返すアルゴリズムとした。

機械学習の手法はWord2vec + Random Forest, Bidirectional Encoder Representations from Transformers (以下、BERT)^[1]を採用した。Word2vec + Random Forestでは、Word2vecのWikipedia学習済みモデルを用いて、入力された文章をベクトルに変換し、それを元にRandom Forestで学習・分類を行う。Word2vecは単語のベクトル化しか行えないため、文章ベクトルを「単語ベクトルの相加平均」とした。一方、BERTは自然言語処理の分野でブレイクスルーと言われており、文脈を考慮した文章のベクトル化や文章分類を行うことができる。事前学習モデルに対してFine-tuningを行うことで様々なタスクへの応用が期待されている。本研究では、事前学習モデルとして東北大学で公開されている

Pretrained Japanese BERT models を用いた. このモデルにレンタカーサービスに関する対話データで Fine-tuning を行った.

各アルゴリズムを適用した対話管理は, レンタカーサービスで想定される 30 パターンの一問一答 [2] 形式の対話への対応を可能とした. 入力文を分類する際に, ルールベースでは返答可能・不可能の二値分類で行う. 機械学習では 30 パターンの中でどの分類が最も確率が高いかを判断し, 入力文を 30 パターンに分類を行う. 本システムに適用したアルゴリズムごとに対話分類精度を比較する実験を行った.

3. 実験結果

本実験は, ルールベースおよび機械学習によって対話データの分類を行い, 対話分類精度の比較を行った. 対話データ例を表 1 に示す.

表 1. 対話データ例

入力文	応答文
どの車にもカーナビ、ETCはありますか	はい、カーナビ、ETC共に全ての車に搭載しております
レンタル期間を延長できますか	延長される場合は必ず出発店舗に一度ご連絡をお願いしております
他の店舗に返却できますか	返却店舗を変更される場合は必ず出発店舗に一度ご連絡をお願いしております
⋮	⋮
一度に複数台の予約はできますか	一台ずつのご予約となりますのでご了承ください

表 1 のように, 対話データは入力文と応答文が対になっている. 入力文の言い換えを行った文章をルールベースで分類を行った. ルールベースでの分類精度が 50%程度の対話データ 90 件用意した. これらをテストデータとし, 機械学習の手法を用いて正確に分類できるか以下の手順で実験を行った.

- I. 30 パターンに分かれたレンタカーサービスに関する対話データを 1332 件用意した.
- II. 対話データを学習用データ 8 割と検証用データ 2 割に分割した.
- III. Word2vec, BERT を用いてそれぞれのデータを入力し, 文章ベクトルに変換した.
- IV. ベクトル化したデータを Random Forest で学習, BERT で Fine-tuning を行った.
- V. それぞれの学習済みのモデルを用いてテストデータ 90 件を 30 パターンで分類した.

本実験で用いたテストデータ例を表 2, 対話分類の結果を表 3 に示す. テストデータは各文章に正解ラベルが割り振られており, これにより予測ラベルと正解ラベルとの正誤判定を行い, 分類精度を算出した. 表 3 より BERT の精度が Word2vec + Random Forest と比べ 0.10 上がって

いることがわかった. Word2vec + Random Forest では, 文章ベクトルは単語の順序を考慮していない. 一方, BERT で取得した文章ベクトルは文脈を考慮しているため, 精度が上がったと考えられる. 本実験を通して, BERT が他のアルゴリズムと比べ, 正確に対話分類を行うことが可能だと示された.

表 2. テストデータ例

入力文	正解ラベル
カーナビ付いてる?	0
ETC積んでる?	0
カーナビゲーションとETC使える?	0
⋮	⋮
1つの予約で複数台レンタルは可能ですか	29
同時に複数台を予約することはできますか	29
1回に何台まで車を借りれますか	29

表 3. 対話分類結果

classification	accuracy
Word2vec + Random Forest	0.61
BERT	0.71

4. まとめ

本研究は日本語音声対話システムに対し, オープンな技術でシステム構築を図るための技術所見の取得を目的とし, 技術検討を行ってきた. 開発したシステムの音声認識・音声合成は Speech Services, 対話管理は BERT を用いた. 実験結果より, 構築したシステムが一問一答形式の対話を一定程度分類できることが示された. 本研究では, レンタカーサービスを題材としたが, 同様のシステム構築とすることで, 分野の異なるサービスに対しても適用可能だと考える.

今後は, 「頑健性」と「柔軟性」の 2 つの観点でシステムの改善を念頭に置いている. 「頑健性」では, 関係ない話題を省くフィルタリング機能を実装する予定である. 「柔軟性」では, 類義語置換機能を実装する予定である. これらの機能を実装することで, より正確で柔軟な対応が可能となることを期待している.

参考文献

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, <https://arxiv.org/pdf/1810.04805.pdf>, (2020 年 12 月アクセス)

[2] 駒谷 和範, 一問一答, 進化するヒトと機械の音声コミュニケーション, pp. 180-191