

サービスロボットのための実世界知識を扱う クラウド型音声基盤の構築

A Cloud Robotics Platform for Human-Robot Real-World Interactions

2151018



研究代表者

国立研究開発法人
情報通信研究機構

主任研究員

杉 浦 孔 明

[研究の目的]

スマートフォンへの音声インタフェース機能の導入は、音声言語処理機能を広く一般に認知させた。2013年の音声認識分野の市場規模は世界全体で約880億円であり、2018年には1700億円程度へと増加すると予測されている[1]。この背景には、機械学習手法の進展とその大規模データへの適用とともに、音声認識機能がクラウドサービス化され、非力な端末でも高品質なサービスを利用できるようになったことがある[3]。

一方、クラウド基盤のログとして収集される大規模なデータを用いてサービスの品質を向上することは、音声認識分野に限らず一般的に行われてきた。例えば、ユーザの購買情報から推薦精度を向上させたり、クリック数から検索精度を向上させるなどの応用は幅広く行われている。これらは、検索・推薦・認識等について高性能なサービスを行うと同時に、多く利用されることでさらに性能改善を行うものである。つまり、クラウドサービスを提供することでスパイラル的改善が可能であるため、クラウドサービス提供者にもメリットがある。音声認識分野においても、クラウドサービスの提供と大規模ログコーパスを用いた精度改善が行われている。

今後、ネットワークに接続されるロボットやIoTデバイスが増加するに従い、それらの機器を対象としたクラウドサービスも増加すると考

えられる。実際に、クラウドロボティクス分野においては、物体認識や軌道計画などの応用について研究が始まりつつある。一方、人と共存するロボットにおいては音声対話機能の構築が高コストであり、現状では高品質なサービスが難しい[4]。ロボット向けのクラウドロボティクス基盤を構築することで、開発コミュニティに貢献できると考えられる。ただし、このような基盤構築は、音声認識・合成についての基礎技術からクラウド基盤の構築・運用やロボットへの適用に至るまでの包括的な技術開発が必要であり、簡単な課題ではない。

上記の背景から、音声対話向けクラウドロボティクス基盤 rospeek を我々は構築し、ロボットのコミュニケーション機能の開発者に提供してきた。本研究では、rospeek の高速化を行うとともに、ヒューマノイド型ロボットにおける画像処理との統合を行う。

[研究の内容, 成果]

Rospeek は、4か国語（日英中韓）の音声認識・合成に対応しており、学術研究目的に限り無償・登録不要で公開している。2016年1月までに3万ユニークユーザに利用されている。図1に rospeek のユニークユーザの分布を示す。

本研究では、rospeek の高速化を行うとともに、高精度化に向けた検討のため、クラウド基盤上に得られたログを解析した。また、概念実

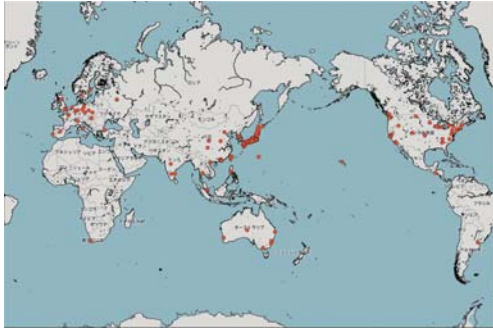


図1 rospeek ユーザの IP 分布

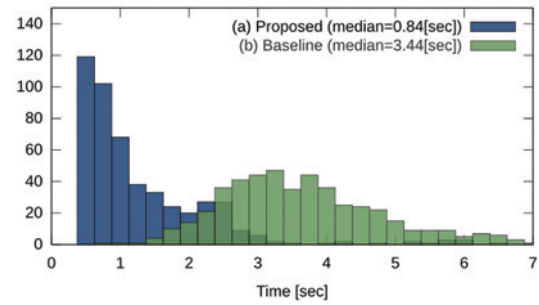


図3 音声認識レスポンス時間のヒストグラム。(a) 分割送信, (b) 一括送信。



図2 rospeek を用いたロボット

証として、ヒューマノイド型ロボット（図2）に rospeek を導入し、画像処理との統合を行った。図2に示すロボットは、カメラ画像を入力として[2]の手法を用いて説明文を生成し、rospeekにより音声で説明を行う。本報告では、これらの成果について述べる。

1. 分割送信による高速化

音声認識を高速化する手段のひとつに音声の分割送信がある。これは、入力中の音声をセグメントに分割し、サーバへ逐次送信するしくみである。分割送信によるレスポンス時間の改善について評価を行った[5]。

本実験の目的は、代表的な使用環境における分割送信と一括送信の時間差の調査である。代表的な使用環境を模擬するため、無線 LAN 環境を用い、rospeek サーバとは異なるネットワーク上からサービスを使用した。本実験ではあらゆる帯域および使用状況におけるクラウドサーバとの通信速度を計測することを目的としていない。各国・各ユーザに依存して異なる通

信環境をすべて模擬することは合理的ではないためである。実験に用いたサーバの CPU は Intel 製 X5690 (12 コア, 3.47 GHz), メモリは 200 GB であった。

コーパスとして ATR503 文を用い、声優 1 名による収録を行った。そのうち発話区間検出に問題がない 495 文についてレスポンス時間を求めた。レスポンス時間とは、クラウドサービスからのレスポンスが得られた時刻から発話検出 (VAD) 終了時刻を引いたものである。

実験の結果を図3に示す。一括送信でのレスポンス時間の中央値は 3.44 秒であったが、分割送信では 0.84 秒に短縮されている。我々の経験では、ロボットとのインタラクションにおいて 1 秒以内にレスポンスを返せない場合、話者が再度発話を行ってしまい、それが誤認識されるなど、望ましくない挙動に陥りやすい。本実験より、分割送信によって待ち時間を減らせることが示唆された。

2. 音声合成に関する個人依存性の解析

これまでロボティクスにおいてはスタンドアロンアプローチが採られてきたため、ロボット開発者が求める音声処理機能の性質について、多様かつ多人数のユーザを扱った定量的な解析が困難であった。クラウドロボティクス基盤を構築することで、ロボット対話開発における個人依存性を調査することが初めて可能になる[6][7]。

いま、あるユーザ u_i の音声合成リクエスト履

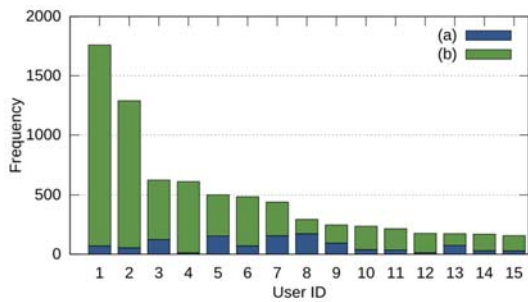


図4 上位15ユーザの音声合成リクエスト数のヒストグラム。(a) N_{uniq} および (b) N_c を表す。

歴全体を文集合 S_i と定義する。 S_i に含まれる文の種類を N_{uniq} とすると、 S_i 中でキャッシュを利用できるリクエスト数 N_c は、 $N_c = \|S_i\| - N_{uniq}$ と表すことができる。ここで、擬似キャッシュヒット率 r_p を $r_p = N_c / \|S_i\|$ と定義する。 r_p が高いユーザは、同じ言い回しを繰り返し使っていると考えられる。2014/1/1 から2015/5/31 までの rospeek のユーザのうち、10回以上合成リクエストを行ったユーザを抽出した。そのうち、上位15ユーザについて解析した結果を図4に示す。

図において、(a)は文の種類 N_{uniq} 、(b)は N_c を表す。 N_{uniq} は N_c に比べて小さく、上位ユーザは100~200種類の文を繰り返し使っていることがわかる。実験の詳細については、[6][7]を参照されたい。上位88ユーザについて平均 $\bar{r}_p = 0.504$ であった。

rospeek をヒューマノイド上に実装し、画像処理と統合した音声対話を行わせた。

[今後の研究の方向, 課題]

クラウドロボティクス技術は屋内レベルから屋外、都市に至るまで多岐に亘る応用を持つ。代表的な応用対象としては、スマートホームデバイス、ドローン、プロブカーなどが挙げられる。

社会展開の手段としてのクラウドサービス提供には、知財であるデータセットを公開する必

要がないというメリットがある。また、ソフトウェアの影響を計測したい場合、ダウンロード数では実際に使用したユーザ数を反映していない可能性がある。一方、クラウドサービスではサーバ上へのアクセス数からユーザが使用した回数を計測することができる。

一方、クラウドサービス提供にはデメリットも存在する。クラウドサービスの知名度が高くなると、サービスそのものやウェブサイトが攻撃や不正利用の対象になりやすい。また、ロボットの機能が全てクラウド化されるべきなのではなく、リアルタイムの衝突回避など、クラウド化が容認されにくい機能も存在する。

ユーザ数の増加が見込めないようであれば、クラウドサービス化以外の方法が費用対効果の面で有利になる場合もある。

[参考文献]

- [1] NEDO 技術戦略研究センター, “TSC Foresight vol.8”, 2015.
- [2] Karpathy, Andrej, and Li Fei-Fei. “Deep visual-semantic alignments for generating image descriptions,” In Proc. of IEEE CVPR, pp. 3128-3137, 2015.

[成果の発表, 論文等]

- [3] 杉浦孔明: “ビッグデータの利活用によるロボットの音声コミュニケーション基盤構築”, 電子情報通信学会誌, to appear.
- [4] 杉浦孔明: “ロボカップ@ホーム: 人と共存するロボットのベンチマークテスト”, 人工知能, Vol. 31, No. 2, pp. 230-236, 2016.
- [5] K. Sugiura and K. Zettsu: “Analysis of Long-Term and Large-Scale Experiments on Robot Dialogues Using a Cloud Robotics Platform”, In Proc. ACM/IEEE HRI, pp. 525-526, 2016.
- [6] K. Sugiura and K. Zettsu: “Rospeek: A Cloud Robotics Platform for Human-Robot Spoken Dialogues”, In Proc. IEEE/RSJ IROS, pp. 6155-6160, 2015.
- [7] 杉浦孔明, 是津耕司: “クラウドロボティクス基盤 rospeek の長期実証実験と大規模ロボット対話データの解析”, 第33回日本ロボット学会学術講演会, RSJ2015AC3D1-03, 2015.