様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　６　月　２７　日現在

機関番号：　８２６３６
研究種目：　若手研究(B)
研究期間：　2015～2017
課題番号：　１５Ｋ１６０７４

研究課題名（和文）クラウドロボティクス基盤を用いた大規模データからの動作と対話の学習

研究課題名（英文）Learning Motions and Responses from Large-Scale Data on a Cloud-Robotics Platform

研究代表者

杉浦　孔明（Sugiura, Komei）

国立研究開発法人情報通信研究機構・先進的音声翻訳研究開発推進センター先進的音声技術研究室・主任研究員

研究者番号：６０４７０４７３

交付決定額（研究期間全体）：（直接経費）　　3,100,000円

研究成果の概要（和文）：本研究は，動作等の非言語知識の学習手法を開発するとともに，クラウドロボティクス基盤を用いたロボットとのマルチモーダル対話を実現することを目的とする．Deep Neural Network (DNN)に基づく動作予測手法を構築するとともに，クラウドロボティクス基盤Rospeexを高度化し、５万ユニークユーザを達成した．生活支援ロボット上に概念実証システムを構築し，１万種類以上の消耗品の知識について音声対話を可能とした．また，状況に応じてユーザの命令を理解可能なマルチモーダル言語理解手法Latent Classifier GANを開発した．

研究成果の概要（英文）：In this study, we aim to develop a learning method of non-verbal knowledge such as motion and multimodal dialogue with a robot using a cloud robotics platform. Time series prediction method based on Deep Neural Network which introduced Dynamic Pre-training was proposed and its effectiveness was validated by using a standard motion data. In addition to improving the cloud robotics platform Rospeex, we have built a domestic service robot that has over 10,000 multimodal concepts. We also developed a multimodal language understanding method Latent Classifier GAN (LAC-GAN) that can understand user commands according to the situation.

研究分野：　知能ロボティクス

キーワード：　知能ロボティクス　マルチモーダル言語理解　クラウドロボティクス　模倣学習　機械学習

１．研究開始当初の背景

半世紀以上も前から，人との音声コミュニケーションが可能なロボットは，音声翻訳機と並ぶ夢のシステムであった．今日，音声言語処理技術は大きく進歩し，スマートフォンや PC などのコンシューマデバイスにおいて，音声エージェントが広く利用可能になっている．また，接客などを目的とした人型のロボットが市販されるなど，コミュニケーションを行うロボットに対して活発に研究開発が行われている．

ただし，あらかじめ決められたシナリオでは流暢に会話するように見えても，シナリオ外の内容については支離滅裂な会話を行うロボットは多い．一方，ハードウェア上の制約を持つロボットの中には，指示された物を運ぶなどの物理的な利便性を提供できないものもある．これらの事実が示すことは，コミュニケーションと物理的な利便性の双方において，実応用レベルのロボットを構築することは実際には簡単ではない，ということである．移動・物体認識・把持などの基本機能から，多様な話者に対する音声認識や指示の意味理解にいたるまで，要素技術の精緻化と高度な機能統合が必要とされる．

時々刻々変化する実世界における感覚運動系にグラウンドした言語処理は，多くの関連課題を有する挑戦的な分野である．例えば，「ペットボトルを片付けて」という発話は頻出するユーザ指示であるが，ロボットが行動を開始するために十分な情報を含んでいない．一方，十分な情報を含む命令文は，不自然であることが多い．例えば，「現在把持中のペットボトルをキッチンの棚の３段目の右側に片付けておいて」のような発話をユーザが行うことは考えにくい．

上記の問題への単純なアプローチとして，スロット値がすべて確定するまで聞き返すアプローチが現状では多いものの，このアプローチでは「シリアルはどこですか？」「キッチンのどの棚ですか？」「棚の何番目の段ですか？」など多くの確認発話が生成されるため，動作実行するに必要な時間が長く，不便である．

近年，画像処理・音声処理・自然言語処理などの分野で深層学習と大規模データを組み合わせたデータ指向アプローチが成功し，ロボティクスにおいても言語・非言語知識を利用したコミュニケーションに関する試みが実応用と結びつく事例も報告されるようになった．本研究では，動作等の非言語知識の学習手法を開発するとともに，クラウドロボティクス基盤を用いた音声対話，および状況に依存したマルチモーダル言語理解手法を実現する．

２．研究の目的

以下では，本研究で取り組んだ（１）Deep Neural Network (DNN)による動作予測，（２）クラウドロボティクス基盤 Rospeex，（３）状況に依存したマルチモーダル言語理解，の３点について目的を述べる．

（１）DNN による動作予測

模倣学習分野においては，Kinect 等の RGB-D カメラが広く用いられている．これらの安価なデバイスで得られた動画像においては，全ての骨格情報が観測可能であるとは限らない．すなわち，隠れた関節角を欠損値として扱うか，関節角の推定値を求める必要がある．

一般的な時系列予測問題を扱ったものは非常に多く存在する．予測問題における DNN の構造を検討したものに[1]がある．[2]では，2 つの restricted Boltzmann machine からなる Deep Belief Network を用いた時系列予測手法が提案されている．一方，DNN において学習データの提示法を検討した研究としては，Curriculum Learning と呼ばれるアプローチがある．Curriculum Learning では画像認識や言語モデルが議論の中心であるが，予測など他のタスクについても有効であることが示唆される．

このような背景から，Kinect 等の安価なデバイスで得られた関節角時系列の予測問題に取り組んだ．提案手法では，時系列に特化した Pre-Training 手法を用い，動作予測に DNN を適用する．

（２）クラウドロボティクス基盤 Rospeex

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．この問題を解決するためにクラウドロボティクス基盤を構築することはコミュニティへの貢献は大きいと考えられる．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．

このような背景から，音声対話向けクラウドロボティクス基盤 Rospeex を我々は構築し，2013 年 9 月から運用してきた．多言語の音声認識・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．主な適用先は生活支援ロボット等，人にサービスを提供するロボットである．このようなロボットにおいては，高騒音環境における認識精度向上や多言語対応などの問題から，音声対話機能の構築が高コストであり，Rospeex を利用することでコストの低減が可能である．Rospeex の長期実証実験において得られた結果を解析し，Rospeex の高度化に取り組む．

（3）状況に依存したマルチモーダル言語理解

ロボットとの音声対話において，不完全情報および記号接地に対応した言語処理は，多くの関連課題を有する挑戦的な分野である．例として，日常環境で「新聞片付けておいて」という音声指示をロボットが実行するタスクを考える．環境中には「新聞（に分類され得るオブジェクト）」が複数存在する可能性があるうえ，「どれを」「どこに」「どうやって」片付けるか，など様々なレベルで曖昧性が存在するため，望ましい動作を実行することは簡単ではない．そこで，物体操作タスクにおける不完全情報を含む言語理解の一例として，動詞のない命令文を入力としたマルチモーダル言語理解手法の構築に取り組む．

３．研究の方法

（1）DNN による動作予測

２－（1）の目的のもと，関節角時系列の予測問題に取り組んだ．提案手法では，Dynamic Pre-Training (DPT)手法を用いて DNN を訓練する．DPT の独自性は，DPT では誤差関数にペナルティ項を必要としないことと，DPT は時系列に特化した変換を用いる点である．

DPT は，Pre-Traning におけるオートエンコーダの学習を対象とする．DPT では，入力時系列を順序を保ったまま部分時系列に分割する．各部分時系列は，反復回数に応じて変化する重要度が割り当てられる．各部分時系列を重み付けして結合し，実際の学習に用いるサンプルを作成する．手法の詳細は[雑誌論文⑦]を参照されたい．

（2）クラウドロボティクス基盤 Rospeex

2018 年 3 月 31 日現在，Rospeex は 5 万ユニークユーザ以上に利用されている．以下では，Rospeex を用いた長期実証実験によって得られた結果を解析する．具体的には，2014/1/1 から 2014/11/28 までのアクセス記録をもとに，実際の利用における音声認識ログを解析した[雑誌論文⑥]. ログに含まれる発話の音声認識結果の総数は, 44,960 であった．ただし，無音など明らかに発話が含まれないものは取り除いた．音声認識結果を以下のカテゴリに分類する．

(a) 挨拶・雑談: 日常会話（例 : こんにちは）
(b) 一問一答型質問: 対話履歴を必要としない情報源への問い合わせ（例 : 今何時）
(c) 移動・把持: 移動や把持に関連する動作指示発話（例 : 止まれ）
(d) 家電操作: 音声リモコンのように家電を操作する発話（例 : テレビを消して）
(e) 認識・学習: センサ入力の学習または認識を指示する発話（例 : ここはどこ）
(f) 一般的な指示: (c)-(e)以外でロボットの行動を指示する発話（例 : 手を上げろ）
(g) その他（検索・回答，判別不能）: (a)-(f)以外の発話．主に，質問への応答，音声認識誤りまたは判別不能な発話を含む．

（3）状況に依存したマルチモーダル言語理解

２－（3）の目的のもと，動詞のない命令文からの物体操作可能性の推定に取り組んだ．一般に，動作スロットが埋まっていない場合は，確認発話により聞き返しを行うこともできるが，実世界情報に基づいて動作スロットを補完できれば利便性の向上につながる．具

表 1 CATS ベンチマークにおける性能評価

Method	Score (E_1)
DPT-DRNN (proposed)	1451
RBM (baseline)	1622

体的には，LAtent Classifier Generative Adversarial Nets (LAC-GAN)を提案した[学会発表①]．LAC-GAN は Generative Adversarial Nets[3]を拡張し，分類器として利用するものである．GAN は，画像や文の生成などに適用され，品質の良い疑似サンプルの生成が報告されている．LAC-GAN はこれらと関連するが，GAN における Generator/Discriminator に加え，特徴抽出を行う Extractor を有することが異なる．手法の詳細は[学会発表①]を参照されたい．

４．研究成果

（１）DNN による動作予測

提案手法の有効性を検討するため，時系列予測の性能評価を行った．本研究では，時系列予測のベンチマークとして標準的に用いられている CATS を用いた．CATS ベンチマークは，5000 フレームの人工データから 100 フレームの欠損値を予測するタスクである．

表 1 に提案手法である DPT を導入した Deep Recurrent Neural Network による結果を示す．評価尺度として，CATS ベンチマークにおいて使用されている誤差の指標である E1 を用いた．表より，ベースラインと比較し，本手法の誤差が小さいことがわかる．

次に動作予測に対する提案手法の評価を行った．評価において標準的なデータセットを用いることは重要であり，本研究では MSR Action3D Dataset を用いる．本データセットは 10 人の被験者に 20 種類の動作を行なわせ，Kinect により収録したものである．各動作は平均 120 フレームほどであり，少なくとも 3 回動作が繰り返される．評価では，モデルの予測にはその動作を行った被験者の情報は使われていない．

図 1 は，データセット中の「手を左右に振る動作」に対して，提案手法を適用した結果を示す．図において，上図・中図・下図に手首の特徴量に対する x, y, z 軸の軌道を示す．結果の詳細については，[学会発表⑦]を参照されたい．

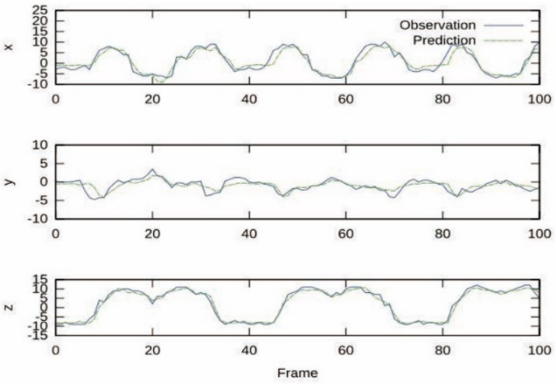

図 1　DPT による手首位置の予測

（２）クラウドロボティクス基盤 Rospeex

Rospeex 上で収集した発話ログのうち，頻度が 3 以上のものを分析対象として発話を分類した結果を表 2 に示す．表より，約半数の発話は挨拶・雑談や一問一答型の質問であったことがわかる．これらの発話に対しては，一般的に提供されている質問応答や雑談対話のクラウドサービスを用いることが有効であると考えられる．一方，(c)-(f)の指示関連発話はロボットごとに機能を実装する必要がある．「その他」カテゴリに分類された発話も多いため，音声認識精度の向上や対話履歴の解析は今後の課題である．結果の詳細については，[雑誌論文⑥]を参照されたい．

（３）状況に依存したマルチモーダル言語理解

物体操作マルチモーダルデータセットを用いて，提案手法とベースライン手法（AC-GAN）を比較評価した．標準的な手順に従い，検証セットの精度が最大値を示したモデルを各手法の最良モデルとした．最良モデルを用いて，テストセットの精度を検証した結果を表 2 に示す．

表において，「PA」の有無は，入力に対して Pre-Activation を行うかどうかを示す．表に

表 2　Rospeex 発話ログの分類

カテゴリ	発話数	割合 [%]
挨拶・雑談	1894	31.59
一問一答型質問	1153	19.23
移動・把持	258	4.30
家電操作	229	3.82
認識・学習	215	3.59
一般的な指示	41	0.68
その他（検索・回答，判別不能）	2205	36.78
合計	5995	100

表 3　LAC-GAN のテストセット精度

	テストセット精度
Baseline (AC-GAN, PA 無)	50.7%
Baseline (AC-GAN, PA 有)	58.2%
Extractor のみ	61.1%
提案手法 (LAC-GAN)	67.1%

おいて,「Extractor のみ」は，Extractor の出力の精度を示す. すなわち, 6 層の単純なフィードフォワードネットワークにおける精度を示す.

表より, AC-GAN と比較して, LAC-GAN は高い精度を示した．この結果は，特徴量をそのまま用いる AC-GAN より, 特徴抽出を行い, 分類に関係が深い特徴のみを用いた方が有利であることを示唆している. これは, Generator の機能であるサンプル生成により，Discriminator に入力されるサンプル数が擬似的に拡張され，汎化性能に寄与したことが示唆される.

（4）社会展開・水平展開

上述した手法の構築および解析と並行し，研究成果の水平展開を行った．太陽フレア予測に LAC-GAN を応用し，専門家の予測精度を圧倒的に上回る世界最高性能を達成した．その結果，天体物理学の最高峰ジャーナルである The Astrophysical Journal に 2 年連続で採択された．また，コミュニティ先導活動としては，ロボカップ 2017 世界大会の運営を行うとともに，生活支援ロボットを標準化した．多言語音声対話クラウドロボティクス基盤 Rospeex の社会展開を進め，5 万ユニークユーザを達成するとともに，Rospeex 用に開発した合成音声を複数の企業・研究機関にライセンシングした.

＜引用文献＞

[1] S.F. Crone, M. Hibon, and K. Nikolopoulos, “Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction,” International Journal of Forecasting, vol.27, no.3, pp.635–660, 2011.

[2] T. Kuremoto, S. Kimura, K. Kobayashi, and M. Obayashi, “Time series forecasting using a deep belief network with restricted boltzmann machines,” Neurocomputing, vol.137, pp.47–56, 2014.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” Advances in Neural Information Processing Systems, pp.2672–2680, 2014.

５．主な発表論文等

〔雑誌論文〕（計９件）

① A. Magassouba, K. Sugiura, H. Kawai, "A Multimodal Classifier Generative Adversarial Network for Carry and Place Tasks from Ambiguous Language Instructions", IEEE Robotics and Automation Letters, 査読有, 印刷中.
DOI: 10.1109/LRA.2018.2849607

② K. Sugiura, "SuMo-SS: Submodular Optimization Sensor Scattering for Deploying Sensor Networks by Drones", IEEE Robotics and Automation Letters, 査読有, 印刷中.
DOI: 10.1109/LRA.2018.2849604

③ N. Nishizuka, K. Sugiura, Y. Kubo, M. Den, and M. Ishii, "Deep Flare Net (DeFN) Model for Solar Flare Prediction", The Astrophysical Journal, 査読有, Vol. 858, Issue 2, 113 (8pages), 2018.
http://iopscience.iop.org/article/10.3847/1538-4357/aab9a7/pdf

④ 奥川雅之, 伊藤暢浩, 岡田浩之, 植村渉, 高橋友一, 杉浦孔明, "ロボカップ西暦 2050 年を目指して", 知能情報ファジィ学会誌, 査読無, Vol. 29, No.2, pp. 42-54, 2017.

⑤ 杉浦孔明, "模倣学習における確率ロボティクスの新展開", システム制御情報学会誌, 査読無, Vol. 60, No. 12, pp. 521-527, 2016.

⑥ 杉浦孔明, "ロボットによる大規模言語学習に向けて -実世界知識の利活用とクラウドロボティクス基盤の構築-", 計測と制御, 査読無, Vol. 55, No. 10, pp. 884-889, 2016.

⑦ 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, 査読無, Vol. 99, No. 6, pp. 500-504, 2016.

⑧ 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, 査読無, Vol. 31, No. 2, pp. 230-236, 2016.

⑨ L. Iocchi, D. Holz, J. Ruiz-del-Solar, K. Sugiura, and T. van der Zant, "RoboCup@Home: Analysis and Results of Evolving Competitions for Domestic and Service Robots," Artificial Intelligence, 査読有, Vol. 229, pp. 258-281, 2015.

〔学会発表〕（計１２件）

① K. Sugiura and H. Kawai, "Grounded Language Understanding for Manipulation Instructions Using GAN-Based Classification", IEEE ASRU, Okinawa, Japan, 2017.

② 杉浦孔明, "ロボットの音声コミュニケーション技術～言葉や能力の壁を越えるデータ指向知能に向けて", 音学シンポジウム 2017, お茶の水女子大学, 2017 年 6 月 18 日.

③ K. Sugiura, "Cloud Robotics for Building Conversational Robots", IROS 2016 Workshop on Machine Learning Methods for High-Level Cognitive Capabilities in Robotics, 大田, 韓国, Oct. 14, 2016.

④ K. Sugiura and K. Zettsu, "Analysis of Long-Term and Large-Scale Experiments on Robot Dialogues Using a Cloud Robotics Platform", ACM/IEEE HRI, Christchurch, New Zealand, 2016.

⑤ K. Sugiura, "Statistic Imitation Learning and Human-Robot Communication", The 2nd International Workshop on Cognitive Neuroscience Robotics, Sankei Conference Osaka, Feb. 21, 2016.

⑥ K. Sugiura and K. Zettsu, "Rospeex: A Cloud Robotics Platform for Human-Robot Spoken Dialogues", IEEE/RSJ IROS, Hamburg, Germany, Oct 1, 2015.

⑦ 杉浦孔明, 是津耕司: "Dynamic Pre-training を導入した Deep Neural Network による関節角時系列の予測", 第 33 回日本ロボット学会学術講演会, 山形大学, 2015.

〔図書〕（計０件）

〔産業財産権〕

○出願状況（計０件）

○取得状況（計０件）

〔その他〕
ホームページ等
Rospeex website: http://rospeex.org/

６．研究組織
(1)研究代表者
杉浦　孔明（SUGIURA, Komei）
国立研究開発法人情報通信研究機構・先進的音声翻訳研究開発推進センター先進的音声技術研究室・主任研究員
研究者番号：60470473