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

令和　４ 年　 ５ 月 ３１ 日現在

機関番号：１２６０１
研究種目：若手研究
研究期間：2019～2021
課題番号：１９Ｋ２０３４９

研究課題名（和文）仮想現実環境を利用した家庭内行動の生成によるデータセットの効率的な大規模化

研究課題名（英文）Efficient Data Augmentation of Household Behavior with Simulation in Virtual Environments

研究代表者
郷津　優介（GOUTSU, Yusuke）

東京大学・生産技術研究所・特任研究員

研究者番号：８０８１６８２７

交付決定額（研究期間全体）：（直接経費）　　2,700,000円

研究成果の概要（和文）：本研究では，人間の全身動作とテキストはそれぞれ姿勢と単語からなる同じ系列データとして，系列変換モデルを用いた行動と言語の融合研究，その中でも主に動作の言語記述に関する研究を行った．系列の正当性の評価を従来のように逐次的に要素ごとに行うのではなく，要素の探索により終端状態にまで達した系列全体を評価する枠組みを取り入れ，その結果を系列変換モデルの学習に利用する手法を提案した．これにより，長い系列の生成に対して予測誤差が少なくなり，単に「歩く」だけでなく「数歩だけ前に歩く」や「弧を描くように歩く」などのように，観測した動作を詳細に記述するテキストまで生成できることを実現した．

研究成果の学術的意義や社会的意義
人間の動作を詳細に記述できるということは，動作の細かな差異とそれに対応するテキストの関係を捉えられる，即ち行動と言語の高度な表現関係まで取り扱えるようになったことを意味する．これにより，例えばスポーツ解析などにおいて，上級プレイヤーの熟練された動作を予め学習しておくことで初級プレイヤーとの差異を指摘し，更にはどのように動作を修正すれば上達できるかを助言するなどの動作支援への応用に繋がる．このことは，周りに熟練者がいなくてもシステムとのインタラクションを通してコーチングを受けることができるという点で社会的に非常に重要である．

研究成果の概要（英文）：We have tackled our research related to a fusion of human behavior and language with a sequence-to-sequence translation model, in which human whole-body movement and text are considered as the same sequential data consisting of postures and words respectively, and the main research topic is linguistic description of human motion. Specifically, our approach incorporates a framework for evaluating the validity of entire sequence that has reached the final state through search of sequence elements, and uses the results to train the translation model. This is quite different from previous approaches that evaluate the element-wise prediction sequentially. By reducing the prediction error for the generation of long sequences, not only "walk" but also "walk forwards a few steps" or "walk a quarter circle clockwise" describing observed human motion in detail can be appropriately generated.

研究分野：コンピュータビジョン

キーワード：動作認識　行動認識　言語生成　身体動作　系列変換　敵対的学習　ニューラルネットワーク

１．研究開始当初の背景

膨大な情報が溢れているインターネットなどから目的に合ったデータセットを発見できる画像・映像と異なり，家庭内の人間の行動，とりわけ関節の 3 次元位置などの身体情報を研究対象とする場合には，特定のタスクのために作成された比較的に小規模なデータセットを利用していることが多い．これは，家庭内の身体動作を含むデータセットが被験者のプライバシーや環境セッティングのコストの問題などから計測・収集が困難であることに起因する．また，人間の行動を取り扱ったタスクの精度はデータセットのドメインに依存する傾向があるため，既存のデータセットを利用したからといって目的のタスクまで遂行できるとは限らない．そのため，データセット作成のための動作の計測・収集によるコストは不可避の問題である．そこで，既存の行動データセットを効率的に大規模化できるように，どのようにデータ拡張を行っていくのかということが重要な問いになってくる．

２．研究の目的

家庭内の人間の動作とその内容を記述したテキストはそれぞれ姿勢と単語で構成される系列データである．動作とテキストをペアとして，両系列の意味レベルでの対応関係を利用して系列を生成するために，行動と言語を双方向に結ぶ系列変換モデルを構築する．さらに，系列変換モデルの介在変数の操作により基準に対してバリエーションのある動作・テキストを生成することで，汎用的にデータ拡張できる枠組みを構築する．これらにより，既存の小規模なデータセットに対してデータ拡張機能を適用することで，様々な表現を持った動作・テキストのペアを生成し，擬似的な大規模化により動作認識の性能を向上させることができる．ただし，系列変換モデルを学習するためのデータセットは，仮想現実空間上に構築した家庭内環境を利用して効率的に収集する必要がある．

３．研究の方法

動作とテキストを系列データとみなして双方向の系列変換モデルによりお互いを生成するための枠組みを構築する．提案手法は，エンコーダとデコーダと呼ばれるニューラルネットワークから成る系列変換モデルに対して，近年様々な分野で活用されている敵対的学習の枠組みを取り入れたものである．具体的には，系列変換モデルである生成器と入力された系列データが実在データか生成データかを判別する識別器で構成される．また，学習は生成器と識別器の事前学習と両器を用いた敵対的学習の 2 段階となっている（図１）．

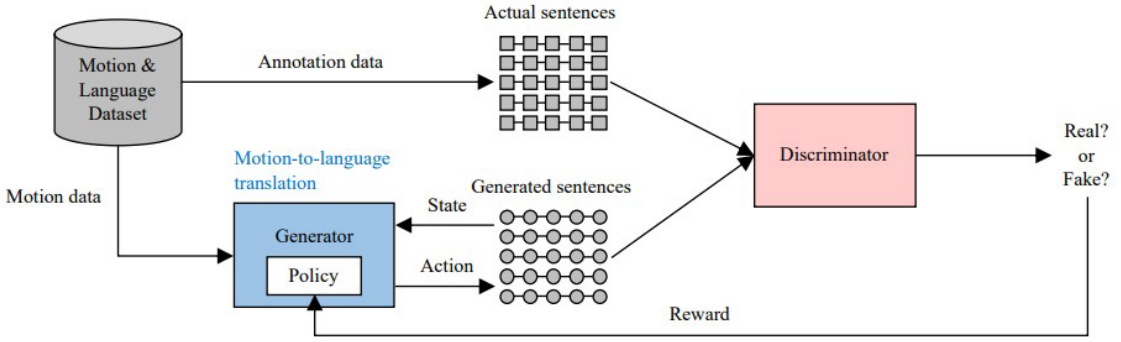


図１．系列変換モデルに敵対的学習と強化学習の枠組みを取り入れた提案手法の概観

従来の系列変換モデルのエンコーダとデコーダの学習では,正解系列を用いた逐次学習により尤度最大化を行うが，予測時はモデルの出力系列に置き換わるために予測誤差が徐々に蓄積する傾向にあった．また，学習時には過去の系列の履歴から次の要素を予測できるようにモデリングするが，テスト時には系列単位での正解類似度を評価尺度としており，その結果を最適化に用いている訳ではない．そこで，本研究では，これまで探索してきた系列を状態，要素の探索行動により終端状態にまで到達した系列の評価を報酬とする強化学習の枠組みを取り入れて，その報酬を敵対的学習における識別器の出力として生成器である系列変換モデルの重みを更新する手法を提案した．

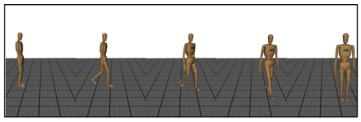
当初の計画では，系列変換モデルの学習に必要なデータセットを効率的に作成するために，簡易的なモーションキャプチャを装着し，実世界の人間の動作と家庭内環境を模した仮想現実世界にいるアバターの動作を同期させることで計測・収集のコストを軽減する予定であったが，研究期間中に所属が変わってしまったことでその方法での実現が難しくなり，既存の行動・言語データセットを利用するだけに留めた．

４．研究成果

行動・言語データセットとして既存の The KIT Motion-Language Dataset を利用する．このデータセットは，光学式モーションキャプチャシステムにより人間の全身動作を取得し，独自に開発したアノテーションツールにより動作を記述した複数のテキストを付与したものである．なお，データセット内の動作データ数は約 3900 個，付与したテキスト数は約 6300 個であった．提案する系列変換モデルのエンコーダとデコーダに関して，系列データの深層学習手法として代表的な Recurrent Neural Network (RNN)を利用している．近年では，RNN ユニットを複数重ねた stacked RNN や入力の系列データを前向きと後ろ向きの双方向に重ねた bidirectional RNN も報告されており，本研究でもエンコーダは 2 層の stacked bidirectional RNNs を，デコーダには 2 層の stacked unidirectional RNNs を適用した．ここで，隠れ状態の数はそれぞれ 64 と 128 とした．

図２では，動作データに対して生成された複数のテキストの中で確率の高かった上位 3 個を順に表示している.また,比較対象として先行研究による動作認識の結果も並べている.例えば,(c)の「バイオリンを弾く」動作に対して，“a person plays the violin”と“a person is playing the violin”という 2 種類の表現を持ったテキストが生成されている．(f)の「フロアから立ち上がる」動作では，“floor”と“ground”が同義語として使い分けられている.また,提案手法により生成されたテキストは人間の動作を詳細に記述できていることが分かる．例えば，(b)-(d)の動作ではそれぞれ“both hands”，“left hand”，“right hand”のように使用している手の判別も正しく行われている.(a)と(e)の「歩く」動作では,先行研究と比較して,単に「歩く」だけでなく「弧を描きながら歩く」「数歩だけ前に歩く」などのように，観測した動作を詳細に記述したテキストが正しく生成されている.実際の定量評価でも,人間の動作を入力としてその内容を記述したテキストを出力するタスクにおいて先行研究を上回る精度を達成できた.

当初の計画では，テキストからそれに対応する動作を生成するタスクも同様に系列変換モデルを用いた双方向生成という形で実現していく予定であった．しかし，動作はテキストよりも多次元且つ連続的な系列データであり，提案モデルでは高精度な生成結果を得られなかった．本研究により，行動と言語の関係性を抽出し，人間の主観と合致する両者の意味空間が構築できていることを確認できた．

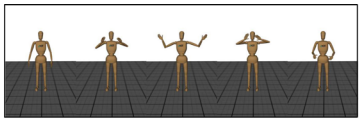


References:
subject walks a quarter circle clockwise
a person walks in a 9 0 degrees curve to his right

Ours:
a person walks a quarter circle to the right
a person walks a quarter circle clockwise
a person walks a quarter circle clockwise starting with right

M. Plappert et al. [8]:
a person walks in a circle clockwise
a person walks in a circle to the right
a person walks in a circle

(a) Walking (quarter circle curve)

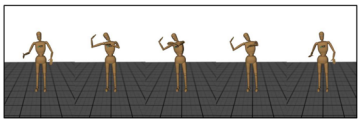


References:
a person waves with both hands
person waving with both hands
a human waves with both hands

Ours:
a person waves with both hands
a person waves with its right hand
a person waves with his right hand

M. Plappert et al. [8]:
a person waves with both hands
a person waves with his right hand
someone waves with both hands

(b) Waving (both hands)

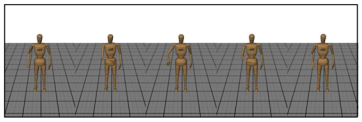


References:
a person is playing violin
a person plays the violin with the left hand

Ours:
a person plays the violin with its left hand
a person is playing the violin
a person plays the violin with its right hand

M. Plappert et al. [8]:
a person plays the violin
a person plays violin
a person plays the violin with its left hand

(c) Playing violin (left hand)

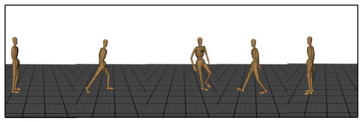


References:
person washing something with the right hand
someone moves the right hand in a circular path
a human wipes a table with his right hand

Ours:
a person wipes something with his right hand
a person wipes something with its right hand
a person wipes something with his left hand

M. Plappert et al. [8]:
a person wipes something
a person wipes something with its left hand
a person wipes something with its right hand

(d) Wiping (right hand)

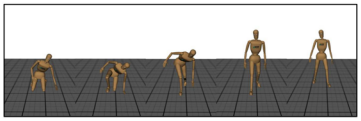


References:
a person walks forwards and turns around
a person walks forwards a few steps then turns 1 8 0 degrees to the right and keeps walking

Ours:
a person walks 2 steps forward turns around and walks back
a person walks 2 steps forward turns 1 8 0 degrees to the right and walks back
a person walks forward turns 1 8 0 degrees to the right and walks back

M. Plappert et al. [8]:
a person walks forward turns around and walks back
a person walks 2 steps forward turns around and walks back
a person walks 2 steps forward turns 1 8 0 degrees on the left foot and walks back

(e) Walking (turn around)



References:
a kneeling person stands up from the floor

Ours:
a person stands up from the floor
a person stands up from the ground
a person stands up on the ground

M. Plappert et al. [8]:
a person stands up
a person stands up from the floor
a person stands up from the ground

(f) Standing up

図２．人間の全身動作からテキストを生成する実験の比較結果

5．主な発表論文等

〔雑誌論文〕　計0件

〔学会発表〕　計3件（うち招待講演　0件／うち国際学会　2件）

1．発表者名 Yusuke Goutsu, Tetsunari Inamura
2．発表標題 Linguistic Descriptions of Human Motion with Generative Adversarial Seq2Seq Learning
3．学会等名 Proceedings of 2021 IEEE International Conference on Robotics and Automation (ICRA2021)（国際学会）
4．発表年 2021年

1．発表者名 Yusuke Goutsu, Tetsunari Inamura
2．発表標題 How Can a Human Motion Dataset Be Collected Effectively? - Roadmap for Human Motion Data Augmentation -
3．学会等名 Proceedings of the 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE2019)（国際学会）
4．発表年 2019年

1．発表者名 郷津優介，稲邑哲也
2．発表標題 系列変換と方策勾配法を用いた敵対的学習による動作-説明文間の双方向生成
3．学会等名 第37回日本ロボット学会学術講演会
4．発表年 2019年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関