

H-034

# 好ましい会話を検出するための時系列のデータ拡張 Data Augmentation of Time Series Data for Detecting Concentration on Conversation

上野晃英<sup>†</sup> 島川博光<sup>†</sup>  
Koyo Ueno Hiromitsu Shimakawa

## 1. はじめに

近年は機械学習の発展に伴い多くのデータ拡張の手法がデータの種類の問わず研究、提案されてきた。しかし Um[1] には有効なデータ拡張の手法はデータの種類や特性ごとに異なっていると説いている。さらに効果的にデータ拡張するためにはデータ拡張の対象となるデータに関する専門知識が必要な場合もあり、効果的なデータ拡張の手法が明らかでない分野も多い。

コロナ禍のため意思疎通能力が衰えた人が増えているため、著者らは会話を円滑に進める研究を進めている。人間の会話における会話の盛り上がりや興ざめを検知したいが実験で収集できるデータの量には時間や費用の制約のために限りがあり、機械学習を行うのに十分なデータが集まらない。よって、会話の盛り上がりや興ざめの特徴をニューラルネットワークで抽出するための構造を表現して取り入れることは困難である。本研究は、人間の会話を記録した体動や音声の時系列データを、機械学習、特にディープラーニングなどを用いてデータ拡張することに着目する。

提案手法により、収集に手間がかかる会話データの特徴を、ディープラーニングを含む機械学習を用いて学習することができるようになり、分析モデルの構築が可能となる。会話の盛り上がりや興ざめを機械的に検知できるようになることで、これまで個人のスキルに依存した会話の円滑化のためのスキル、例えば会話を活性化する話題の選択の仕方や間の取り方、合図地の取り方や頻度など、いわゆる会話スキルの有無を判別することにつながる。さらに会話の盛り上がりや興ざめを機械学習モデルで表現できるので、人間同士の会話のみではなく、人間と機械の会話の円滑化にも応用が期待できる。

## 2. 会話を示す時系列データの拡張

### 2.1 時系列データのデータ拡張

Um[1] には時系列データのデータ拡張手法として Permutation, Time-warping, Magnitude-warping, Jittering, Scaling を用いている。Permutation はデータをセグメント化し部分的に入れ替える。Time-warping はデータを部分的に時間軸方向に長くしたり短くしたりする手法である。

Magnitude-warping はデータを滑らかな曲線で畳み込む。Jittering はランダムなノイズを付加する手法である。Scaling はランダムなスカラー値を乗算することでデータの大きさを変化させる。

### 2.2 姿勢推定

本研究では体動データの取得のために学習済み深層学習モデルを用いたトップダウン型の姿勢推定を用いる。今回利用するトップダウン型の姿勢推定はまず物体検出で人を検出し、その後に画像データから体の各部位の位置とそれらの接続関係を推論するものである。実験で取得した映像データから体の各部位の位置情報の変動を示す時系列データを取得してデータ拡張と学習に用いる。Maji[3] には姿勢推定に YOLOv5 を用いているが本研究では最新の YOLOv7 を用いる。

## 3. 会話状態の検知とデータ拡張

本研究では会話の盛り上がりや興ざめの検知のための会話データに特化したデータ拡張の手法の提案をする。

本研究の手法概要図を図 1 に示す。

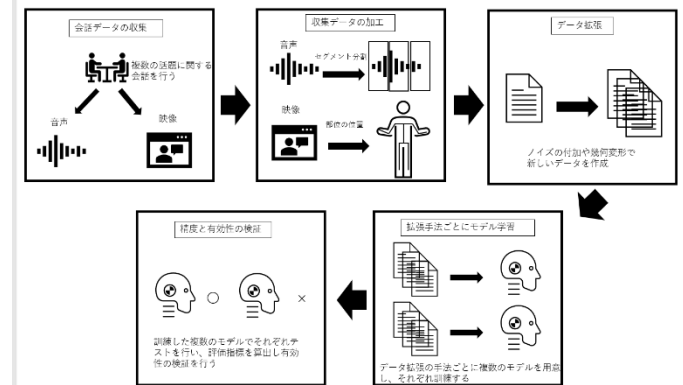


図 1 手法概要図

本研究では会話時に取得した音声データと映像データをもとに音声と体動の時系列データを取得する。取得した少数の実験データに対してデータ拡張を行う。データ拡張手法ごとにデータセットを分けてそれぞれで CNN を学習させ、評価を行う。

### 3.1 音声データと姿勢推定による体動データの取得

会話が盛り上がっている時には姿勢が前のめりになったり

<sup>†</sup> 立命館大学情報理工学部

身振り手振りが大きくなったり体動に関する特徴が現れる。そのため本研究では YOLOv7 を用いた姿勢推定により体動データの取得の取得を行う。体の各部位の位置の変動の時系列データを、会話の実験から取得した映像データに YOLOv7 を適用することにより、取得する。YOLOv7 での姿勢推定では最終的に身体の 17 か所のうちそれぞれ最も確信度の高い位置の x 座標、y 座標、確信度が得られる。これらの座標の時間的推移を体動データの変動とする。実際に姿勢推定を用いて取得した体動データの一部の例を図 2 に示す。

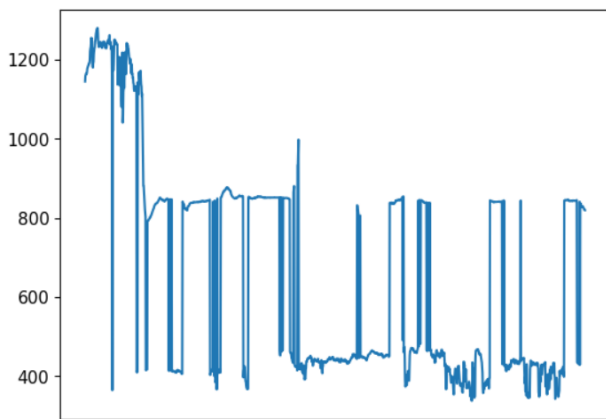


図 2 取得した体動データの例

さらに取得したデータに関して、機械学習を行う際の入力データサイズを統一することと単一のデータ内に盛り上がりや興覚めの複数の特徴の混在を防ぎラベルを一意に決定するためにセグメント分割を行う。

### 3.2 データ拡張の手法

データ拡張はデータセットのサイズを増加させるとともに訓練データの偏りや少なさから生じる過適合を防ぎ、より汎化能力の高いモデルを作成することにつながる

同一の会話から取得したデータに変動が生じる要因として被験者とマイクやカメラの距離など位置関係の違いが考えられる。音声データであれば実験の最中に被験者の姿勢が変化しマイクとの距離が変化することで音声データの大きさが変動すると考えられる。そのため音声データに対して大きさを変化させるためにランダムに乗算ノイズを付加することで音声データの被験者とマイクの位置関係の変化によるデータの変動をシミュレートすることができる。

その他に変動をもたらすものとして盛り上がりや興覚めの特徴の時間的位置の変化があげられる。入力データとして最終的に固定サイズのセグメントとして任意の位置から抽出するため、どのタイミングで盛り上がりや興覚めの特徴が現れるかは重要ではない。そのため入力データ内で盛り上がりや興覚めの特徴の位置を変動させることでデータ拡張することは有効であると予測される。

一つ目のデータ拡張手法として入力データをいくつかの少数のセグメントに分割して入れ替える手法がある。入力データ自体が元の実験で取得したデータをセグメント化し

たものであるため、二段階に分けてセグメント化していることになる。定義域を定めたセグメント数をランダムに決定し、それをランダムに並べなおし、新しいデータを作成する。二つ目のデータ拡張手法は取得したデータを時間軸方向に拡大、縮小するものである。これを入力データの一部、あるいは全体にわたって適応することで新しいデータを作成する。三つ目のデータ拡張手法はランダムな数値をデータに対して乗算するものである。四つ目のデータ拡張手法はランダムな滑らかな曲線を畳み込むことで新しいデータを作成する手法である。五つ目のデータ拡張手法はランダムなノイズを加算して新しいデータを作成するものである。六つ目のデータ拡張手法は入力データをセグメント分割しこれらをランダムに重複を許して入力データと同一のサイズになるまで選択し新しいデータを作成するものである。ここで新しいデータを作成する際に同一の入力データをセグメント分割したもののみを用いる方法と同一被験者の同一の実験で取得した別の入力データからそれぞれ生成されたセグメントを組み合わせる二種類のパターンでデータ拡張を行う。さらにこれらの手法を複数組み合わせることも考えられる。

### 3.3 精度と有効性の検証

それぞれのデータ拡張手法ごとに分類モデルを訓練し複数のモデルを用意する。ベースラインとしてデータ拡張なしの SVMなどを想定する。単一のデータ拡張手法で用意したデータセットで訓練したモデル、複数のデータ拡張手法を組み合わせで用意したデータセットで訓練したモデルの精度評価を行い、会話データにおいてどのようなデータ拡張の手法が有効であるか検証する。

## 4. おわりに

本研究では会話の盛り上がりや興覚めの検知をするために音声データと体動データからなる会話データに特化したデータ拡張手法を提案した。今後は実際に実験を行い、実験で取得したデータに対してこれらのデータ拡張手法を適用しその有効性を検証し、その効果について議論する。

### 参考文献

- [1] Terry T. Um, et.al, "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks" ICMI '17: Proceedings of the 19th ACM International Conference on Multimodal Interaction Pages 216–220(November 2017).
- [2] Wu Liu, Qian Bao, Yu Sun, Tao Mei, Recent Advances of Monocular 2D and 3D Human Pose Estimation: A Deep Learning Perspective ", ACM Computing Surveys, Volume 55Issue 4Article No.: 80pp 1–41.
- [3] Debapriya Maji, et.al, "YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss", IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (2022)