H-048

好ましい会話を検出するためのラベルを維持するデータ拡張
Data Augmentation to Maintain Labels for Detecting Concentration on Conversation

上野晃英† 島川博光†
Koyo Ueno Hiromitsu Shimakawa

1. はじめに

本研究では，会話の盛り上がり分類のためのラベルを維持するデータ拡張手法を提案する．会話の盛り上がり分類のためのデータ拡張に人間が会話に対して持っている感覚を取り入れる．人間が感覚的に似ていると判別したデータ間の類似度とデータ拡張によって生成されたデータセット内の類似度を比較する．そして類似度別にデータセットを作成し，複数のモデルを訓練する．

実験の結果，人間の感覚をデータ拡張に取り入れることは有効であることが分かった．これにより．良質の訓練データを作れるので，良好な人間関係の構築を機械的に支援するために会話の盛り上がりの判別器が構築可能となる．

2. データ拡張

データ拡張はモデルの汎化性能を向上させる有効な手法である．データ拡張とは既存のラベルの付いているデータを変形して新しいデータを作成し，モデルがクラス内の共通の特徴を学習できるようにすることを目的としている．

大規模なデータセットの収集が困難な場合にはデータ拡張により機械学習モデルを適切に学習することができるようにする必要がある．データ拡張はCNNを含む深層学習に、優れた性能を発揮させるために必要不可欠な前処理である[1]．しかし効果的なデータ拡張の手法はデータの種類や特性により異なっている．

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判別の精度向上のためのデータ拡張手法の策定を目指している．そのために、まず実験にて，マイクを用い会話データを収集する．収集した会話データには、盛り上がりが歩かないかのラベルが付与されている者とする．そしてここで収集した会話データに対して異なる手法や異なる条件でデータ拡張を実施する．最終的に各手法，各条件別に拡張したデータで判別モデルの学習を行い，精度を評価する．上記の手法の概要を図 1 に示す．

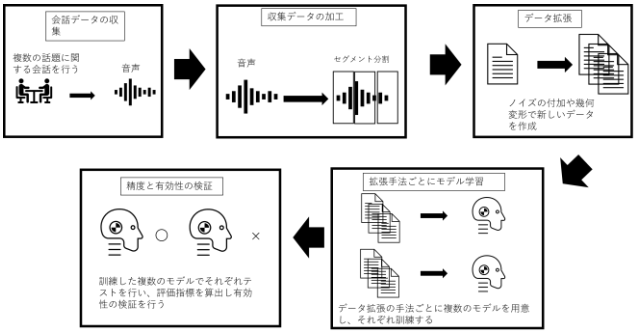

図 1　手法概要図

3.1 データ拡張

有効なデータ拡張の手法はデータの種類や特性により大きく異なっている．そのため有効なデータ拡張手法が確立されていない種類のデータも多くある．

そのため本研究は会話データの盛り上がり分類に特化したデータ拡張手法の提案と指針策定を目的としている．本研究では会話データの盛り上がり分類におけるデータ拡張で人間の感覚を取り入れることを提案する．

会話の中で人間は意識的にあるいは無意識のうちに会話が盛り上がっているかどうかをある程度認識している．例えば盛り上がっている場合は声のトーンが高くなったり，会話のレスポンスが早くなったりする傾向があることを感覚的に理解している．

そもそも会話が盛り上がっているかどうかは人間の感覚に依存している．本研究の実験で取得したデータでも人間が判断することにより手動でラベル付けしており，この例にもれない．ラベル付けが人間の感覚に依存しているのでデータ拡張にも人間の感覚を取り入れることは有効であると考える．

本研究では，加算ノイズ，乗算ノイズ，滑らかな曲線の畳み込み，時間軸方向の伸縮操作の 4 種のデータ拡張手法について検証する．これらは Terry [2]らによりウェアラブルセンサーから取得した加速度の時系列データを用いた分類タスクのためのデータ拡張手法として提案されたものである．

3.2 人間の感覚の統合

本研究では各データ拡張手法の有効性の検証のために正解率（accuracy）を利用する．異なるデータ拡張手法やパラ

† 立命館大学情報理工学部

メータの条件ごとに前節で述べたモデルを複数学習させてそれぞれ正解率（accuracy）を算出し比較する．

本研究では会話の興奮度の分類がうまく機能するように，データ拡張を成功させるための基準として人間の直感を導入する．深層学習ではモデルだけでなく，データ拡張手法についても人間が任意に設定できるハイパーパラメータがある．通常これらのパラメータは複数のパターンの中から精度などの数値の指標から適したものが選択される．

深層学習が取り扱う大部分のタスクと異なり，本研究で扱うデータのラベルは完全に人間の感覚のみに依存している．

その他にこのデータの大きな特徴としてラベルの不安定さが挙げられる．まず，同じデータのラベリングでもラベルを付ける人が異なればラベルが変わる可能性が多分にある．さらに同じ人物がラベリングしていたとしても別の時間や日にラベリングすればラベルが変わってしまう可能性は十分にあると考えられる．

こういった不安定さをデータ拡張の時点で人間の感覚を利用して取り入れることは有効ではないかと予測した．そのためハイパーパラメータを人間の感覚で決めることを考える．各データ拡張手法を用いて実験で取得した生のデータを大きく変形させた場合や少量のみ変形を加えた場合が考えられる．どの程度のデータの変形まで人間が聞いて許容できるか，元のデータがもつラベルと同一であると判断できるかという観点で評価することで人間の感覚をデータ拡張に取り入れる．

ただし人間の感覚という主観的な指標のままでは評価が困難であるために客観的な数値指標を導入し，人間の感覚を数値化する．ここではデータ拡張の変形度合を評価する数値指標として DTW を利用する．DTW から類似度を算出するためにまず DTW を算出する． これは比較した二つのデータに対して各サンプルを重複ありかつ順番が入れ替わらないようにしてマッチングしつつ最小化したものである．

DTW を比較したふたつのデータに対して各サンプルを重複ありかつ順番が入れ替わらないようにしてマッチングしつつ最大化した値で除算する．これによりデータ自体の振幅の大きさによる DTW への影響を抑えられ，異なるデータ間で比較できるようになる．

最後にこの数値を 1 から引くことで類似度を算出する．類似度は最大値が 1，最小値が 0 となる．つまりどの程度低い類似度まで大きな変形を加えたデータ拡張まで許容できるかといった観点で人間の感覚を数値化する．

データ拡張手法，類似度を異なる条件で設定しデータ拡張手法を行う．このようにして拡張したデータでモデルを学習させる．そして精度を算出し，比較して評価することが本研究における一連の流れである．

4. 人間の感覚によるデータ拡張への影響

本研究ではデータ拡張に人間の感覚を導入する．任意のデータ拡張手法でデータを変形させた場合における，変動の大きさの人間の感覚による許容の度合と精度への影響を調査する．ここでの変動の度合は DTW から算出した類似度で数値化する． 本研究ではデータの前処理としてセグメントに分割しメルスペクトルグラムへ変換する．また正解率を算出し，精度評価するために VGG16 を転移学習で利用する．

表 1　人間の感覚の導入による時間軸方向の伸縮の正解率

Similarity	0.2177	0.3787	0.5224	0.668	0.7843
Accuracy	0.7379	0.7634	0.787	0.7857	0.7857

表 1 は時間軸方向の伸縮での人間の感覚による精度への影響を調べたものである．時間軸方向の伸縮による変形で類似度を低下させていくと類似度 0．7843 から類似度 0．5224 までは精度に大きな変化は見られない．類似度が 0．3787 の時点でやや精度が低下し始め，類似度が 0．2177 まで低下すると精度がさらに低下していることが確認できる．時間軸方向の伸縮では類似度を約 0．7843 まで低下させても人間の感覚でラベルの判別が可能である．つまり人間の感覚で判別可能な境界より大きくデータを変形させるほど精度が低下することが確認できる．このことから時間軸方向の伸縮については人間の感覚でデータ拡張のハイパーパラメータを調整することが精度を向上させることが分かる．

乗算ノイズと滑らかな曲線の畳み込みについては類似度を変化させても精度の変動はほとんど確認できなかった．

加算ノイズでは人間の感覚でラベルが判別できる限界点と判別モデルの精度が低下し始める境界が一致しなかった．

5. おわりに

本研究では会話の盛り上がり分類で人間の感覚をデータ拡張に取り入れるデータ拡張手法を提案した．本研究で人間の感覚をデータ拡張に取り入れるデータ拡張手法の有効性を検証した結果，加算ノイズ以外，データ拡張において，人間の感覚を取り入れることが有効であることが分かった．

今後は半教師あり学習を用いて，人間が聞いて判別する工程を省いて，自動化することで，人間の感覚によるデータ拡張の低コスト化を進める．

参考文献

[1] Park, Daniel S. et al.，“SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition，” Interspeech, 2019.

[2] Terry T. Um et al.，“Data augmentation of wearable sensor data for parkinson’s disease monitoring using convolutional neural networks，” Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI '17), 2017, pp. 216–220.