

第66回日本学生科学賞 最終審査進出研究作品概要

HIT016IT	高校	情報技術	東京都
学校名	東京都立大泉高等学校		
研究作品タイトル	JasperNet 顔動画像解析による日本語の発話予測		
研究者氏名 (共同の場合はグループ)	佐々木 俊輔		
指導教諭氏名	山口 貴史		

【動機】

読唇術は非常に難易度の高い行為であるため、機械学習による自動読唇の社会的な需要は非常に大きいものである。Assaelら(2016)によるLipNetは、95.2%の精度を達成したが、日本語では精度が低い。日本語の特性の考察をもとに、シーケンシャル画像による発話認識の精度向上のための手法を検討することには重要な社会的意義がある。

【方法】

本研究で提案したJasperNetは、喉の震えや顎の動きも訓練画像として使用した。また複雑な特徴を持つ画像を学習できるように、GoogLeNetを参考にしたInception Moduleを実装することで畳み込みニューラルネットワークの多層化を試みた。また日本語の発音の特徴を効率よく学習するために、Temporal Moduleを実装し時間方向の学習の効率化を図った。

【結果】

従来手法の訓練画像・ネットワーク構造におけるCTC損失は3.13、訓練画像のみに改良を加えたモデルの損失は2.97、ネットワーク構造にのみ改良を加えたモデルの損失は2.31、訓練画像・ネットワーク構造ともに改良を加えたJasperNetで学習したモデルの損失は2.10であった。

【まとめ】

訓練画像の改良・ネットワーク構造の改良ともに認識精度の向上に役立つことがわかり、特に日本語の特性を考慮して構築したネットワークの改良による影響が大きいと考えられる。訓練画像・ネットワーク構造ともに改良を加えたJasperNetモデルの損失が最も小さく、研究上の価値が示された。

【展望】

将来的に、より大きなデータセットにJasperNetを適用し、より安定した精度と汎化性能の獲得を目指し、最終的に実用化させることを目標とする。図書館内などの発声が適さない場所や、雑音が大きく声が届きにくい公共の場所等において、音声認識に代わるコミュニケーションツール

としての役割が期待できる。また言語学習者及び聴覚に障がいがある方の支援としての利用にも期待できる。さらに音声を入力することができない防犯カメラにおいて、会話内容の推定における活躍にも期待できる。