

受領者投稿

感情コミュニケーション研究における資源不足の解消に向けて

千葉工業大学 情報科学部 情報工学科 准教授 有本 泰子
(2014年度受領者)

深層学習の普及により音声から感情を認識する技術や表現豊かな合成音声を実現する技術は飛躍的に進歩した。感情認識では高精度に話者の感情を認識することが可能であるし、音声合成では人間の声と区別がつかないほどの精巧な音声を作り出すことができる。深層学習がこれまでのこうした技術のブレークスルーとなったことは疑いようがない。しかしながら、深層学習もまた大きな問題を抱えている。それは、これまで利用されてきた機械学習の手法以上に多くの学習データを必要とすることである。学習に使用可能なデータが少なければ、学習が収束しなかったり、学習に使用したデータに依存したモデルを構築することになり、汎用性の低いモデルとなる。つまり、一人の話者の感情を認識することはできるものの、他の話者の感情は認識できないシステムになる。音声合成も同様である。例えばニュースを読み上げる場面の音声を合成することはできても、友人とのカジュアルな対話を実現する音声を合成することはできない。これらは全て、音声資源の不足に起因する問題である。

2014年度～2015年度に立石財団からご支援いただいた「言語情報と音響情報の統合的利用による感情音声コーパスの大規模化」のテーマのもと、感情コミュニケーション場面の音声資源不足の問題解決にチャレンジした。当時、感情ラベルが付与された日本語対話コーパスがふたつ存在していたものの、それらに付与されていた感情ラベルはひとつが10種類の感情カテゴリでありもうひとつは6種類の感情次元であったため、これらを併用した学習は不可能であった。そこで、コーパスの共通化を行うことを目的に、互いのコーパスに付与されている感情カテゴリあるいは感情次元をもう一方のコー

パスにも自動で付与する感情推定モデルをサポートベクターマシンにより構築した。感情推定モデルの説明変数には、学習に使用するコーパスに含まれる音声の音響的特徴量のみならずその音声に付与されている感情ラベルを利用した。その結果、感情ラベルを利用したモデルを使用したほうが推定精度が向上することを示した。これらのモデルを他のコーパスに適用することで、感情ラベルが付与されていないコーパスにも感情ラベルが付与でき、感情コミュニケーション研究における資源不足の解消の一翼を担うことが可能である。

現在は、Affect Burstと呼ばれる無意識な感情の噴出である笑い声や叫び声を対象に、深層学習を利用した音響イベント検出や音声合成の研究を行っている。笑い声や叫び声は対話をしていれば必ず出現する現象ではないため、Affect Burstの音声資源不足は感情を含んだ音声よりもさらに厳しい状況にある。全く関係のないデータで事前に学習を行う転移学習や信号処理などを駆使してデータを拡張するデータオーグメンテーションの技術により、資源不足を解消することが検討されているが、その利用可能範囲は限定的であり、感情コミュニケーション研究における資源不足の抜本的な解決には至っていない。支援を受けた研究手法を拡張し、人的コストをかけずに大規模にデータを収集し、コーパスを共通で利用可能とする技術がさらに求められるだろう。

