［報　告］計算科学・計算機科学人材育成のためのスーパーコンピュータ無償提供利用報告

工学部電気情報物理工学科「アドバンス創造工学」プログラム「深層学習による歌声音声変換」

伊藤　彰則
工学研究科　通信工学専攻

１．概要

工学部電気情報物理工学科では、「Step-QI スクール」という学部生用の教育プログラムを提供しています。「アドバンス創造工学」はその一環であり、学部 2 年生・3 年生を対象として、研究室での研修を通して研究の体験をするプログラムです。

同学科の伊藤・能勢研究室では、「歌唱音声の操作と評価」という題目で研修を行いました。このテーマで 2 名の学生が研修を行いましたが、うち 1 名は音声操作に深層学習（ディープラーニング）の手法を用いるため、大量の計算を必要とします。そのため、サイバーサイエンスセンター提供のスーパーコンピュータ無償提供制度を利用させていただきました。

２．研修内容

歌唱音声には、基本的な要素である音高やテンポ、歌唱技術であるビブラートやこぶしなどのほか、知覚要素である「熱唱度」があります[1]。本研修では、深層学習を用いた写像関数を利用することにより、熱唱でない音声を熱唱音声に変換する手法を検討しました。

手法は以下の通りです。最初に音声分析合成系 World[2]を利用して歌唱音声を基本周波数(F0)、スペクトルおよび非周期性指標に分解します。次に、F0 およびスペクトルを熱唱音声に変換するためのネットワークを深層学習により推定します。非熱唱音声が与えられたとき、F0 とスペクトルを学習済みネットワークによって変換し、元の非周期性指標と合わせて再合成し、熱唱音声を得ます。

実験には R を利用し、深層学習フレームワークとして RSNNS[3]を利用しました。学習・評価データは 11 名が日本語のポピュラーソングを「熱唱」「非熱唱」の 2 通りで歌唱した音声で、それぞれ 3〜5 秒程度の音声を切り出して使っています。スペクトルの変換には通常の多層ニューラルネットワーク、F0 の変換にはエルマン型のリカレントニューラルネットワーク(RNN)を利用しました。

評価実験を行い、スペクトルの変換はあまり効果がなく、F0 の変換はやや効果あり、という結果が得られました。本研修の成果は、2018 年 3 月の電子情報通信学会総合大会の学生ポスターセッションで発表しました[4]。

３．所感など

工学部電気情報物理工学科には教育用計算機がありますが、主に演習用であり、今回のように一つのジョブに数時間を要する計算には向いていません。研究室所有の計算サーバにも空きがなく、今回のスーパーコンピュータ無償提供利用制度は大変助かりました。関係各位に感謝申し上げます。

参考文献

1. R. Daido, M. Ito, S. Makino and A. Ito, “Automatic evaluation of singing enthusiasm for karaoke”, Computer Speech and Language, Vol. 28, No. 2, pp. 501-517, 2014
2. M. Morise, F. Yokomori, and K. Ozawa. “WORLD: a vocoder-based high-quality speech synthesis system for real-time applications.” IEICE TRANSACTIONS on Information and Systems 99, no. 7 (2016): 1877-1884.
3. C. N. Bergmeir and S. J. M. Benítez, “Neural networks in R using the Stuttgart neural network simulator: RSNNS,” Journal of Statistical Software, 2012, Vol. 46, No. i07, 2012.
4. 早坂、伊藤:「RNN を用いた FO 操作による歌唱音声の熱唱化の検討」,電子情報通信学会 2018 年総合大会、2018-3.