

先端情報技術 10月3日授業レポート

201821636 村松直哉

2018 年 10 月 3 日

1 設問1

問 学習サンプルさえ増やせば、音声認識の課題は全てDNNで解決すると考えられるか？それでも解決できない課題や問題があるとすれば、それは何か？

私はDNNでは音声認識に関する全ての問題は、解決できないと考える。ここで「全ての問題の解決」とは、人間並みが高確率で成功させる課題に関して、同程度以上の確率で成功させることができること意味する。主張の主な理由を2つ述べる。

1つはいわゆるone-shot learningの課題である。DNNといわれるアルゴリズムは多くあるが、その中でも、対象の学習データが1つのような状態を認識するのは非常に難しい。

新しい言葉や、対話者の造語などは、事前に大量の学習サンプルを確保することが難しい。画像処理におけるone-shot learningの論文は、多く発表されている。しかし、音声認識分野ではまだそれほどない。また、one-shot learningの精度は、人間の認識精度よりも大きく劣っている。

この課題は、学習サンプル数の問題ではなく、特徴量抽出の問題である。そのため、サンプルを増やしただけでは、解決できない課題の1つであるとする。

2つ目は、状況に応じた受け答えに関する課題である。例えば「あれ」「これ」など、こそあど言葉を正確に認識することは難しい。なぜなら、対話者同士の事前知識が必要である。通常のDNNでは、記憶機能が明確になく、ニューロン同士の重みとその機能を果たしていると考えられる。しかし、現在のニューラルネットワークでは、具体的な情報を引き出すことができない。記憶を参照するような対話を認識するためには、DNNが具体的なデータを扱えるようにする必要がある。

以上のような理由から、現在のDNN技術だけで、学習サンプルさえ増やせば、音声認識課題を全て解決することができないと考える。