

人工知能と人間の共存する生態系における学習

吉田 洋

現代の人工知能は大規模データから機械学習によって自動的に知識を獲得している。そして、特に自然言語を扱う人工知能の場合 web 上に存在する大規模テキスト集合から学習を行う事が多い。その大規模テキスト集合の一つである Twitter にはすでに多数の人工知能や会話 bot、自動出力が存在し、人工知能による出力は最終的に広い範囲に渡ってみられることが予想される。そのため、これらの出力は取り除くことが困難になるであろうと予測される。その結果、機械の出力がコーパスの中に含まれ、人工知能は人工知能から学習する環境になりうると予想される。

そこで、本研究では人工知能同士の出力が互いに影響しあう環境における学習を想定し、自然言語系の人工無能を対象に、学習済みの人工無能からの出力をもとに学習した場合の出力への影響を明らかにすることを目指した。

そのために Twitter 上に存在する実際の人間の投稿から学習した人工無能の出力(以降、第一世代人工無能)を用意し、第一世代人工無能の出力のみから学習した第二世代人工無能と人と第一世代人工無能の出力を 1:1 で組み合わせた教師データから学習した人工無能の出力と第一世代人工無能の出力の比較や、人工無能の出力から学習し、その出力から学習を行うことを繰り返した第 n 世代人工無能の出力の世代ごとの比較を行い、人間からの出力の比率や自己学習の回数による変化を調べた。

その結果、自己学習を重ねるたびに特徴が増幅されることが示されたが、人間からの出力の比率と人工無能からの出力の特徴の変化の観察が今後の課題となった

(指導教員 手塚太郎)