

注目ディープ・ラーニング AIメカニズム実験室

▶ 第1章 メカニズム実験①…対話する人工知能

第2章 メカニズム実験②…勝手に成長する人工知能

第3章 メカニズム実験③…それらしい写真を自動生成する人工知能

山本 大輝, 速川 徹



図1 人工知能の一大分野に「会話」がある

会話はインターフェースとしてこれからいろいろポテンシャルがある。学習させると、自分の好きなキャラクタといつでも会話することもできる

コンピュータに人間のように言葉を読み書きさせた…過去、そして今なお研究者や技術者が挑んでいる大きな課題です。そのような分野は「自然言語処理」と呼ばれますが、現在、ディープ・ラーニングによってその課題が解決されつつあります。一般に有名な成果としては、

- LINEやマイクロソフトの「りんな」など、問いかげに人間のように応えてくれる「Bot^{注1}」
- Google 翻訳などの異なる言語間の「自動翻訳」

などが挙げられるでしょう。手法やライブラリも洗練された今、個人のマシンでもディープ・ラーニングを用いた自然言語処理を試すことが可能になりました。ここでは有名な小説の口調をまねて問いかげに答えてくれる「夏目漱石Bot」を作成してみましょう(図1)。

● 広がる世界

今回は学習用データ手配の都合から、著作権が切れ

た夏目漱石の小説を利用しましたが、手元にテキスト形式で用意できる会話データがあれば、Twitter上の会話でも、お好きなキャラクタが出てくる本のセリフでも何でもBotにできます。いろいろと試してみてください。例えば、

- 好きな小説の主人公との会話を楽しむ
- 亡くなった人と会話する
- 尊敬する人から励ましの言葉をもらう
- 面接の想定問答
- コールセンターの応答
- スーパーマーケットの接客

などの装置が作れるようになります。

自然言語処理でよく使われるモデル 「リカレント・ニューラル・ネットワーク」

● 時系列データのモデル化に効果

ディープ・ラーニングで扱えるデータには、画像や音声、数値の並びから文章まで多岐に渡りますが、その中でも時間通りに並べられるデータを時系列データといいます。例えば次のようなものは時系列データです。

- 日ごとの日本の平均気温の推移
- 心電図の電位
- 毎月の家庭の電気代

自然言語処理の対象となる文章や会話も時系列データの一種です。「吾輩は/猫/で/ある」というように、単語が読む順番に並んでいるわけです。

時系列データの扱いにはリカレント・ニューラル・ネットワーク(以下RNNと略す)と呼ばれるモデルが有効です。RNNのアイデアはごく単純で、隠れ層の出力を隠れ層の入力に戻してやるだけです。

図2(b)のように、隠れ層の入力には、ある時間の

注1: Botはコンピュータによる自動発音システム。