

良い音データの集め方

牧野 浩二, 西崎 博光, 澤田 直輝

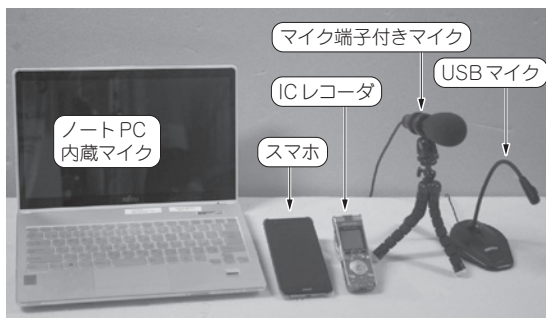


写真1 「音」認識人工知能を育てる実験セット

ディープ・ラーニングは、画像だけでなく音データも判別可能です。ここでは、どうすれば正答率の高いMy「音」認識人工知能を育てることができるのか、実験を交えて解説します。

● 育てるMy「音」認識人工知能

精度の高いMy「音」認識人工知能を育てるためには、良い音データを用意しておく必要があります。まず、必要なハードウェアやソフトウェアを紹介します。ディープ・ラーニング・アルゴリズムが、音を判別しやすくするための(正答率を上げるための)、学習用データを用意して、次のMy「音」認識人工知能をできるだけ正答率が高くなるように育ててみます(写真1)。

- (1) 生ずる音から物を当てるMy人工知能
- (2) 声から人間を当てるMy話者認識人工知能
- (3) ウェブから集めた学習用データを使って育てたMy生活音認識人工知能

人工知能で認識できそうな音の世界

● どのような音が判別できるか

▶人間が分かる音はディープ・ラーニングも分かる

人間が聞いて分かるものは、ディープ・ラーニングで大体判別できると考えてよいでしょう。人間でもお菓子の袋を振ったときの音は、何度も聞いているうちに分か

ります。誰がしゃべっているかも、慣れれば分かります。

また、声を聴くと「うれしそう」とか「疲れてそう」などと分かるときがあります。これは人生経験が積まれていくに従って敏感に感じ取れるようになります。これをディープ・ラーニングで行うという研究もあります。

その道のエキスパートになれば分かる音というものもあります。例えば列車や車の車軸をたたくと、緩んでいるのが分かるとか、てんぶらの揚げている音でおいしい温度を知るなどがあります。

たいた音で状態を知るのには「打音検査」という名前がよく知られており、この音を機械学習で判別しようという試みは、盛んに行われています⁽¹⁾。

● ざっくり人間と同じくらい学習時間もかかる

判別対象を学習するにはどの程度の量を学習すればよいのでしょうか。しっかりした基準はありません。感覚的には、人間が学習するのに必要な時間のオーダーに近い時間分の学習データが必要になるでしょう。

▶お菓子

例えばお菓子の音は、いろいろなお菓子を交互に振って、1つ当たり合計で3分程度も聞けば、大体聞き分けられるようになります。そこで、ディープ・ラーニングでは少し多めにとって、5分程度を学習用データとして用意すればよいと考えます。

▶声

誰の声を判別するときを考えましょう。似ている声の判別は難しいです。例えば「あ」を400回発音しても誰の声か覚えることは難しそうです。作文用紙1枚(400字)程度の文章を読んでもらえると、数名ならば誰の声か判別できそうです。

▶てんぶらを揚げる音

打音検査やてんぶらのおいしい温度はどうでしょう。これ分かるまでには何年も研修が必要となりますね。

今のところ音の聞き分けについて、人工知能が人間以上の性能を発揮したという発表はされていません。つまり、人間が行って習得までに時間がかかる音の判別は、やはり難しい問題となります。