

意外と難しい繊細な日本語表現の判定に挑戦してみる

人工知能で「自分好み」の日本語記事をセレクト

ご購入はこちら

鈴木 直人

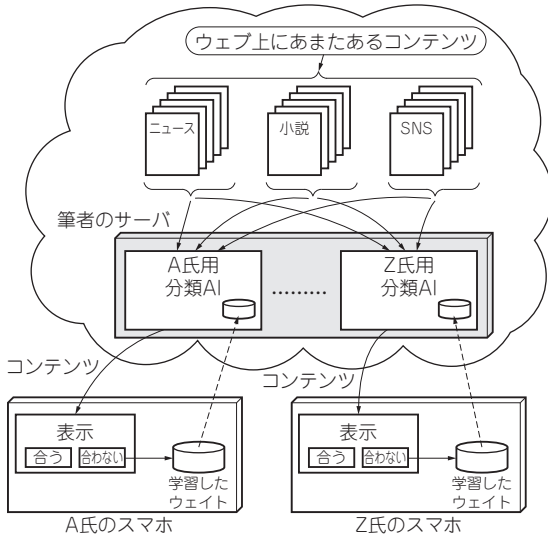


図1 筆者が作った「分類AI」ができること…数多あるウェブ上のコンテンツの中から自分好みの文章を選んでくれる

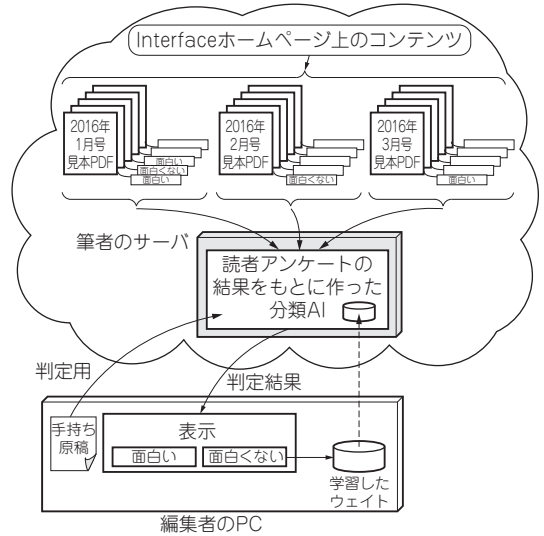


図2 本稿の実験例…図1の「分類AI」でInterface誌のある月号の記事(ウェブで公開している見本PDF)が読者の好みに合うか/合わないのかを判定してみる

● ここでやること…日本語文章の合う/合わないを人工知能で選ぶ

あるまとまった文章(日本語)を自分の「好み/好みでない」と仕分けてくれるプログラムを作り、これをTensorFlow上のGPUで動かします。

目の前にある何らかのまとまった文章、例えばニュースや偶然訪れたウェブ・ページ、たまたま手に取った小説があるとします。数ページめくってみて、自分に「合う/合わない」は何となく感じると思います。この「合う/合わない」の判断を人工知能に学習させるのが狙いです。検索エンジンを一歩進めた「好み分類AI」です(図1)。

筆者が元々作ったのは、個人の好みを学習し、それぞれの好みに合致したコンテンツを提示してくれる分類AIです(図1)。

今回はInterface誌の読者の好みを学習し、編集者が取ってきた記事が読者に喜ばれるのかを判定してくれる分類AIを作りました(図2)。

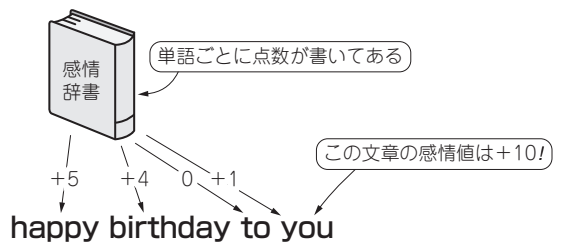


図3 従来の判定方法…感情判定辞書を利用し文章に点数をつける

● ディープ・ラーニング& GPUを使う

好みの文章を判定するとなると、普通、以下の手法が考えられます。

- 1, 用意した単語辞書とどれだけ合致するか
- 2, ウェブ上から入手した感情判定辞書を利用し喜びの点数が一定値を超えているか(図3)

さて、上記のような判定方法は、天邪鬼を自認する筆者にとって「ちょっと違うんだよなあ?」というパターンがほとんどです。自分と同じ感覚で選択をしてくれ