

# 軽くてシンプルでLチカ相当「K-means」を動かす

ご購入はこちら

佐藤 聖

ここからは、筆者おすすめのAIアルゴリズムを「習うより慣れろ」で動かしながら理解していきます。1人でもくもくと始められるように体験サンプルを用意しています(写真1)。

## 軽くてシンプルK-meansの特徴

K-meansは機械学習アルゴリズムとしては、かなり目立たない存在です。しかし処理負荷が軽くてシンプルという良さがあります。Lチカに相当するような基本的なアルゴリズムなので紹介しておきます。処理性能に制約がある組み込み系コンピュータにも最適です。

一般的にはあえてK-meansを選択する意味はないように感じられるかもしれませんが、数ある教師データなし分類アルゴリズムの中でも、特別に精度が高いわけではなく、初期値に依存するので使い方が難しいのも事実です。

クラスタリング・アルゴリズムでは、scikit-learnでも採用されている密度準拠クラスタリング(DBSCAN)の方が一般的に分類精度が良いとされています。また、無数の派生アルゴリズム(K-means++など)が改良版として開発されています。

K-meansは欠点はあるもののアルゴリズムが誰にでも理解しやすく、なぜそのような分類結果になったのかを説明しやすいです。マーケティング・リサーチなどで分類する目的であらかじめ分類数を決めてから利用するなら、応用範囲が広いアルゴリズムだと思います。

教師なしデータ(ラベル付けされていないデータ)でも分類できたり、分類数を指定してデータを分類できたりなど、他にはない特徴があります。例えばラベル付きのデータ・セットを作るときやデータの特徴がはっきりしていてディープ・ラーニングなどの複雑なアルゴリズムを用いなくても分類できるときに役立ちます。

### ● 用途

09 K-meansで分類実験.ipynbを開くとノートブックにサンプル・プログラムがあるので動かしてみましよう。実際に動かすと、人的作業なしで学習モ

デルを作成し、そのモデルに従って分類してくれます。残念ながらその様子は見えません。例えば人間が $(x_1, y_1)$ と $(x_2, y_2)$ の2つを通る直線を引いて分類するなどのルールをプログラムしなくても計算によって編み出してデータが分類されます。

一見するとこれだけかと思われるかもしれませんが、人間が分類すればいいとも簡単なことです。数値データを見ると分かりませんが、グラフ化すれば大体幾つのクラスター数に分けられそうかは一目で分かるはずです。

このように人間が行えば簡単なことをプログラムで毎日実行するような場合、手作業で処理してもよいのですが手間がかかる上に作業者が分類ルールを理解している必要があります。K-meansを応用すると分類のためのモデル構築が行えるので本格的な分析の前処理やデータのラベル付け処理に利用できると思います。

例えばお店の売り上げ分析での応用としては顧客の来店頻度を推定するようリアルタイム分析につながると思います。いつもレジ係をしている人ならばよく見かける顧客の来店時間帯や購入商品の種類を経験的に覚えているかもしれません。しかしセルフレジが普及してきており単純な買い物では人間の経験則が働かず、顧客がポイント・カードを忘れて精算時に顧客情報とのひも付けができないとリピート層の顧客をお店

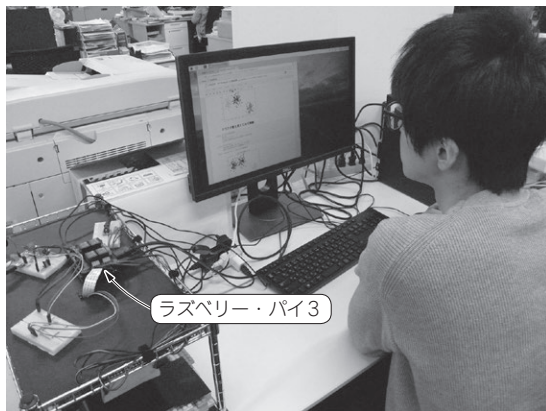


写真1 職場でも自宅でも1人でもくもくと始められるのが今回用意した体験サンプル