本書はInterface2021年1月号掲載の16ページに加筆修正を加えたものです.

AI自習ドリル

第1回　複数の評価項目を持つデータを任意のグループに分けてくれるk平均法

牧野 浩二，足立 悠

AI（Artificial Intelligence，人工知能）と言うと，ロボットが人間と話したり，治療薬を自動的に見つけたりなど，雲の上の技術に感じるかもしれません.

AIと言えば，現在はディープ・ラーニング（深層学習）が一番有望な技術として考えられているようですが，機械学習やデータ分析などと呼ばれる手法も，AI技術の一部です．この機械学習やデータ分析の方法はとてもたくさんあり，社会人ならば会社の業務，学生ならば実験データの解析に使えるものばかりです．そして，これらはデータ・サイエンスやビッグデータ，IoTの技術とも密接に関わっています.

AIドリルでは，AIを構成する機械学習やデータ分析の手法を毎回1つずつ取り上げます．その方法を簡単な例題を用いて紹介するだけでなく，理解を深めるために「実際に手を動かしてやってみよう!」という，ドリル形式で提供します．AI時代を生きる皆さんの理解の手助けになることを願っています.

1　k平均法でこんなことができる

k平均法（k-means）とは，多くのデータを指定の数に分類するための方法です．書籍によっては非階層型クラスタ分析として紹介されています.

1.1　こんなところで使われている

● 売れる商品の配置を見つける

最近はさまざまなポイント・カードがあります．これによって，どんな人がどんな商品を，いつどこで買ったのかをデータ化できます.

このデータを用いて，一緒に買われやすい商品を近くに配置すると，より買ってもらえることを期待できます（例：カレー粉の横に福神漬け）.

複数の業種で使えるカードの場合には，さまざまなデータが含まれていますので，従来は扱っていなかった商品も一緒に置くなどし，新たな商品の買い方を提案できそうです.

● 受験生をデータで分析する

大学では，どのような学生が集まる傾向があるのか，競合する大学はどこかなどを分類することで，より良い大学を目指しています.

受験生の傾向を分類することで，宣伝対象とする学生を絞れることが考えられます.

大学ごとの分類にも利用できるので，特色のある大学にするためにすべきことの検討にも使えそうです．例えば，大学ごとにグループ化して，「改革を行った場合に今のグループとは全く違う分類になるかどうか」を調べることなどに利用できそうです.

● 政府機関が調査に使うことも

実際にk平均法が使われている例として，内閣府の調査の事例を紹介します.

図1-1では，企業が抱える人材と，その企業が求める職種へのミスマッチがどの程度あるのかを，「日本的雇用慣行度」という独自の指標を作り，k平均法によって5つに分類しています.

図1-2は日中と夜間の労働者の増減を前年のデータと比較したものです．産業別の割合を4つのグループに分割することに利用しています.

1.2　分類に向くデータと向かないデータがある

k平均法では潜在的に分類できる要因がデータに含まれている必要があります．例を挙げて説明します.

● 工学部と文学部の学生を見分ける

工学部と文学部の学生を見分けることを考えたと