

答えを学習しなくても特徴を予測できる「自己組織化マップ」

牧野 浩二, 寺田 英嗣

● 実際の需要予測等に使われていて進化中

自己組織化マップは英語でSelf-Organizing Mapと言われ、その頭文字をとってSOMと略されています。これは1980年代前半にT.コホネン (Teuvo Kohonen) が提案した手法で、自己組織的に(教師データがなく与えられたデータだけで自動的にという意味)いろいろなデータが分類されることで話題になりました。

しかし、当時のコンピュータが非力であったため、分類できるデータの数に限界がありました。そのため人間が想像できる程度の分類にしか役立ちませんでした。

2000年代に入りコンピュータの性能が飛躍的に向上したのをきっかけに、多数のデータが扱えるようになったため、第2次ブームが起きました。これにより分類するだけでなく、学習しておいた分類結果に新たなデータを入力すると、そのデータの性質や特徴を予測できるようになりました。

例えば、脈拍解析やゲノム解析、電力需要の予測や気象予測、土砂生産量など、工学の分野を超えて利用されています。そしてさらに、今回紹介する2次元の平面マップだけでなく、マップを球面にしたりトーラス(ドーナツ型)にしたりなど、今でも改良が進んでいます注1。

● マップを作り新たなデータをマップに当てはめることが可能

図1に人工知能のアルゴリズムを整理しました。自己組織化マップ(以降、SOM)は、ニューラル・ネットワークの教師なし学習に位置付けられます。ここでは、これまで紹介したアルゴリズムとの違いを説明します。

SOMを使う場合は、似ているデータを集めて「2次元の表(マップ)」を作ることが目的です。2次元で表現する点は主成分分析に似ていますが、縦軸や横軸に意味はありません。似ているデータが集まるので、クラスタ分析の結果に関連性があります。

SOMによって作られたマップ上に新たなデータを当てはめることで、データの分類や今後の予測などに

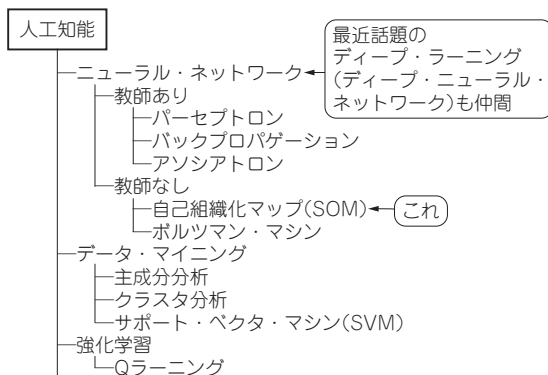


図1 人工知能のアルゴリズムあれこれ

利用できる点が主成分分析やクラスタ分析とは異なります。

そのため、アンケート結果の分析にも用いることができますが、電力需要や気象などの「予測」にも使うことができます。例えば気象予測では、同じ天気の日には気温や気圧などが似ていることを利用します。対象とする地域(日本のある都市)の毎日の(または毎月や毎年の平均の)気温や気圧など測定できるデータを入力としてSOMによって分類することでマップを作ります。そして、今日の気温や気圧などをそのマップへの入力とすると、どのデータに一番近いのかが分かります。そして、そのデータの天気と同じ天気になると予測できます。

体験準備…Rのインストール

論より証拠、まずはSOMを体験してみましょう。主成分分析やクラスタ分析でも使った「R」注2を使って、SOMを扱ってみます。Rを使うといろいろな角

注1: これを応用した研究として、Interface2016年7月号 特集第8章「研究! 生体センシング×機械学習」で触り方を判別する方法を紹介しています。T.コホネンのグループが公開している本家本元のプログラムを使ったものです。

注2: Rはプログラミング言語および開発環境のことです。