

機械学習にチャレンジ ～機械学習とは?と開発環境～

企画連携課 小山 洋太

近年、ディープラーニング(深層学習)の登場によりAI(人工知能)に注目が集まっていますが、実現する手法として機械学習があります。ここでは機械学習について説明し、機械学習を実装するための手段としてPythonによる開発環境の導入についてご紹介いたします。

AIと機械学習 ～AIと機械学習の関係～

2012年、画像認識の競技会であるILSVRCにおいて、ディープラーニングによる手法を用いたチームが、画像認識率を大きく改善したことを皮切りに、ディープラーニングを中心としたAIブームが巻き起こり、現在も活発にAIの研究・開発・活用が進められています。ディープラーニングは機械学習の手法の一つです。機械学習とはアルゴリズムや統計モデルなど、AIを実現するための手段であり、AIは、明確な定義はなされていませんが、人間が行う判断や分析などいわゆる人間の「知能」を電子計算機上で実現したものです。すなわち、AIを実現するためには機械学習といわれるアルゴリズムや統計モデルを電子計算機に実装する必要があります。

機械学習あれこれ

～学習の種類・手法・アルゴリズム～

機械学習には、主に以下に示す2つのグループがあります。(その他には、強化学習などもあります。)

①教師あり学習

例えば、良品・不良品のデータをあらかじめ用意し、それを学習させることで新規データが良品であるか、不良品であるかを判断するなど、判断の基準を与え、それに対する回答を調整することで学習させる手法です。代表的な手法として回帰分析やクラス分類があります。

②教師なし学習

①とは異なり、判断の基準を事前に用意することはせず、例えば、与えられたデータの分布などの特徴や構造などから判断・分析を行うものです。代表的な手法として、クラスタリングがあります。

◎回帰分析

例えば、住宅の価格は築年数や駅などの施設からの距離など、様々な要因から決まります。このような住宅価格(目的変数)を築年数や距離(説明変数)から求める式を見出す分析です。

◎クラス分類

例えば、与えられた写真がイヌかネコかを判別する場合、事前にたくさんのイヌとネコの写真から各々の特徴を見出しておくことで、新たに与えられた写真がイヌかネコかを判別することができます。見出した特徴を基準にして、与えられたものがどこに所属するかを判断する手法がクラス分類です。

◎クラスタリング

例えば、赤いリングと青いリングのデータについて色の赤味

に注目すると赤いリングでは強く、青いリングでは弱くなります。そうすると、たくさんの赤いリングと青いリングのデータは色の赤味の強弱により2つの群を分けることができます。このように分布・構造に着目してデータを複数の群に分けるものがクラスタリングです。

これらの手法を実現するアルゴリズムの1つがディープラーニングであり、他にも決定木・回帰木やサポートベクターマシンなどたくさんあります。これらは各々に特徴があるため、与えられたデータの質や状況(例えばサンプル数)からどの手法を選ぶかを判断する必要があります。

このように機械学習ではまず「どのようなことをしたいか?」、「教師データはあるのか?」から手法を決定した上で、適当なアルゴリズムを選択することになります。

機械学習を始めるには

～Python with Anaconda～

機械学習の分野ではプログラミング言語のPythonが多く開発に用いられます。Pythonでの開発環境としては、例えばAnacondaといわれるディストリビューション(必要なソフトウェアを集めたもの)があり、Windows機でも比較的簡単に実現することが可能です。

表 回帰分析結果

順位	機器	回帰係数
1	A189	8.06
2	A504	6.00
3	A514	5.18
4	A40	4.85
5	H00	4.31

表はAnacondaでPython開発環境を導入したのち、scikit-learnという機械学習ライブラリを使って、当センターの電力使用量と機器の稼働状況を回帰分析し、その結果から回帰係数の上位5位の機器とその回帰係数を示したものです。アルゴリズムにはSVM(サポートベクターマシン)というものを使用し、決定係数は0.6程度でした。

機器名は仮名化していますが、電力消費量の大きい試験機が上位にくる(回帰係数が大きい)結果になり、簡単ではありませんが電力消費量と機器稼働の相関が読み取れる結果となりました。

ここまで述べたように、機械学習は目的やデータの質や状況によりどのような手法を使うか?また、どのようなアルゴリズムを使うか?(さらにはパラメータ調整なども必要)を考えて行う必要があります。比較的気軽に始めることが可能ですので、ぜひこの機会にチャレンジしてみてください。