

# 学習データ作りの基本 ご購入はこちら

佐藤 聖

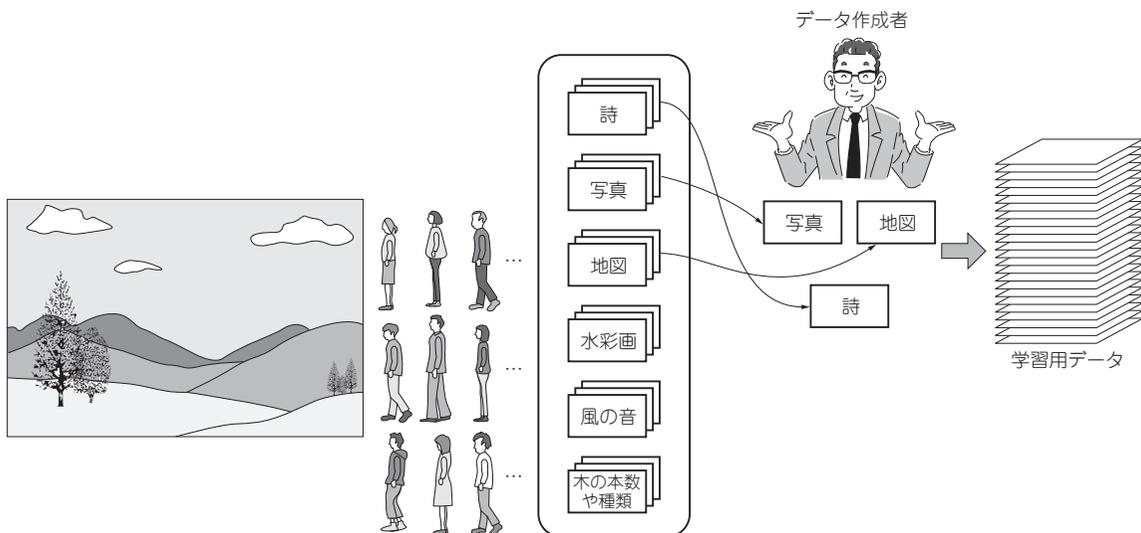


図1 いかに工夫して価値や意味を表現するかが人工知能開発成功のカギ

## いい人工知能作り成功のカギ

### ● データの作り方を工夫して意味を表現する

人工知能には、よく知られた基本形のアルゴリズムだけでなく、目的や用途に応じてさまざまな改良版アルゴリズムが開発されています。毎年1万種類くらいは増加していると言われています。

そもそもアルゴリズムの種類が増えるのは、有用なデータの組み合わせが膨大にあるためです。学習用データとして作られるときに、人間の意味づけや価値が織り込まれて表現されます。学習用データ構築は創作的な要素があり、感覚/出来事/物質などの自然から得られるデータ、株価や広さなど人工的なルールから得られるデータなどの組み合わせからなります。

例えば人間は山の風景を表現しようとしたとき、詩、写真や衛星写真、地図、水彩画、風の音、地響きなどの形で表現しようとするでしょう。しかしコンピュータはデジタル計算機なので、2進数で計算し

なければならないという制約があります。そこで学習用データの作り手としては、「いかに作り方を工夫して価値や意味を表現するか」が、人工知能開発の成功のカギだと思います(図1)。

### ● データの量と質のトレードオフ

本特集では教師あり学習を前提とした学習用データ(コラム)の収集だけにとどまらず、加工、データセット作成、有用性検証などを説明していきます。人工知能の性能は一般的に「データの量」×「データの質」×「学習モデルの精度」のバランスと、各要素のレベルで決まります。データの量と質を今回の特集で取り上げます。

人工知能では一般的にデータ量が多いほど学習に利用できるデータが増えるため、パターン認識や予測精度が向上します。しかし、同時に学習量が増えるため、データを学習してモデルを構築するのにより時間がかかります。個人的には人工知能システム開発の7~8割の時間は、データ収集から学習用データ(デー