

②データが十分にあるか

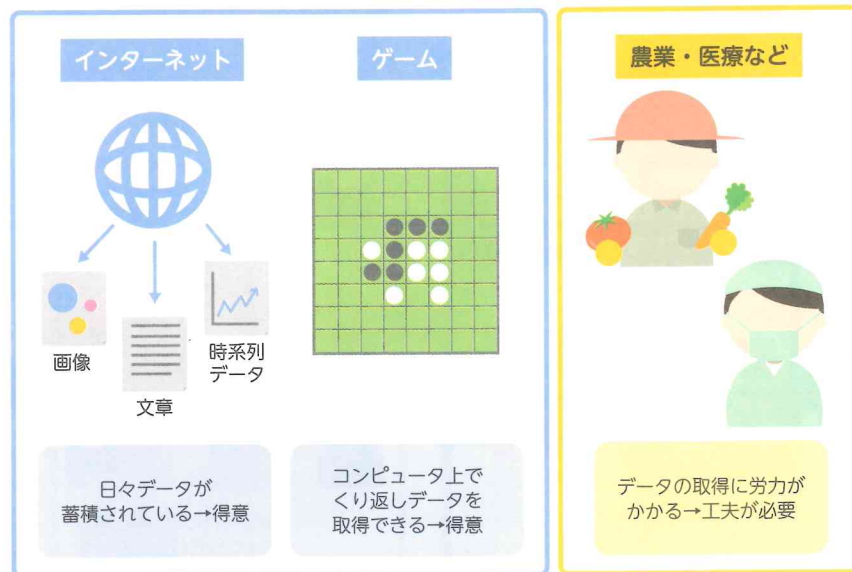
機械学習において、単にデータが「ある」というだけでは不十分なケースもあります。つまりこの場合は「**十分にあるか**」という点が重要です。

データ数が十分といえるかどうかは、適用する問題の難易度やデータセットの質により大きく異なります。とりわけ、画像データ分類など入力するデータが大きいケースでは、それぞれのクラス（分類する対象）のデータが数千から数万単位が必要であると言われています。

近年では、インターネット上の情報であれば比較的に簡単に大量のデータを確保できます。またゲームなど、何度もくり返し試行することが可能な問題もデータ数の確保がしやすいため、機械学習の得意分野と言ってよいでしょう。

一方で、データの入手がオフラインとなってしまう分野や、そもそもあまりひんぱんに起こらない現象を扱う分野では、データ数の少なさが学習のボトルネックとなる場合があります。

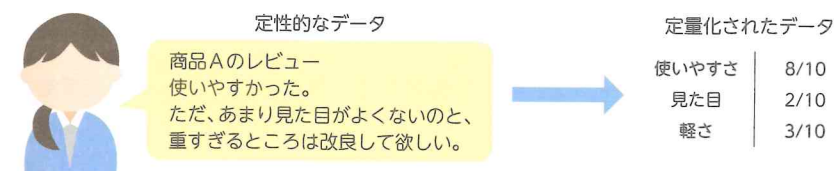
■ 十分なデータが簡単に手に入るかどうか



③データが定量的であるか

一般的に、機械学習の入出力データは数値で表されている必要があります。そのため、数値で表せない定性的なデータ（性質にまつわるデータ）に対して機械学習を適用する場合、これらを**定量的なデータに変換**しなければなりません。たとえば「あるサービスの顧客満足度を向上させる」という課題に機械学習を適用しようとした場合、出力を「顧客満足度の向上」という定性的な表現から「顧客満足度アンケートの数値が〇〇以上」といった定量的な表現に変える必要があるのです。そのため、「顧客データから今後の事業の方向性を決めたい」などといった定性的かつ定量的なデータに変換することが難しい課題の解決は、あまり得意とはいえないでしょう。

■ レビューの定量化



④推論の過程がわからなくてもよい

こちらは、Section9とも重なるポイントです。機械学習は、学習データを入力したときの出力が正答に近い値になるよう、モデルを自動で最適化（学習）するアルゴリズムです（教師あり学習）。つまり、必ずしも人間の思考のように推論が進むとは限らず、その過程を見ても根拠がわからないことが多いのです。そのため機械学習で病気を診断したとして、「あなたは〇〇という病気である可能性が高いです。しかし根拠はわかりません」という結論が出るかもしれません。これでは当然、患者を納得させることもできないでしょう。このように、根拠が重要となる推論が必要な分野において、機械学習のみで結論を出すことは難しいのです。ただし近年は、この問題に対処するために機械学習の推論根拠を可視化する研究が行われており、今後は有効に活用できる可能性があります。

まとめ

■ データ量や推論の用途に注目して判断する