

AI 自習ドリル

第7回 クロス集計やポジ・ネガ解析でアンケートを読み解く

牧野 浩二, 足立 悠

● 人工知能とデータ・サイエンスとの関係

AI自習ドリルでは、しばしばデータ・サイエンスに関わる手法も取り上げています。データ・サイエンスはAIを活用するために必要なものであり、今後の社会人スキルとして望まれる技術です。データ・サイエンスとは、ウェブ上にあるあらゆる情報、自分で取得した多数のセンサ・データなどから有用な情報を引き出し、意思決定に役立てるものです。

AIに入力するデータを扱う方法や、AIから得られたデータを扱うときにも、直接的ではないにしろ、データ・サイエンスの技術が役に立ちます。AIを活用するためにもさまざまなデータ・サイエンスの手法を知っておくことが大切です。

● トライすること

雑誌やテレビといったメディア、YouTubeやLINEといったSNS、そして駅前やショッピング・モールといった外出先など、さまざまな場面でアンケートが行われています。アンケートは集めた後、それを活用す

ることが重要です。今回はアンケート結果を集計する方法として、単純集計、クロス集計、およびポジ・ネガ解析を紹介します。この集計と分析を行うことで、人間がアンケート結果を解析しやすくなります。アンケートはさまざまな場面で利用されていますので、データの分析方法を身に着けることで、仕事や研究に役に立つと思います。

今回はGoogleスプレッドシートを使います。Googleスプレッドシートでは、Excelとほぼ同じことが無料でできます。ExcelがインストールされていないPCもあるでしょうから、本稿ではGoogleスプレッドシートを使うことにしました。

● AIを使って自由記述欄の解析も

アンケートにはたいてい、自由記述欄があります。サンプルとなる文章を10例用意し、それぞれポジティブな単語とネガティブな単語、それ以外の単語がどの程度の割合で含まれているかを調べます。

1 アンケート分析の方法あれこれ

● 単純集計…最も基本的な集計法

単純集計とはアンケートの項目ごとに集計した結果のことです。例えばレストランで、「これまでに来店した回数」や「来店のきっかけ」などの、幾つかアンケートを取った項目について、それぞれ円グラフ(図1-1)や棒グラフ(図1-2)で表したものです。図1-1と図1-2から、単純集計を行うことで項目ごとの回答の傾向を把握できます。他方で、単純集計の方法では異なる項目間の関係性は分かりません

▶例1…レストランで「これまでに来店した回数」(図1-1)

▶例2…レストランへの「来店のきっかけ」(図1-2)

● クロス集計…項目間の関係性を調べる集計法

クロス集計は「睡眠時間と年代の関係」(表1-1)や、「会話の頻度と都市規模の関係」(表1-2)のように、異なる2つの項目間の関係性を調べるための集計方法です。さらに表にまとめるだけでなく、表中に網掛けをして強調することも行われています(表1-1)。クロス集計は2つの項目の関係性を、人間が見やすい形にまとめることができるため、さまざまな場面で使われています。

▶例1…睡眠時間と年代の関係(表1-1)

表1-1からは多くの年代で6~7時間の睡眠が多いことが分かります。他方で40代では6~7時間の睡眠が少ないことが分かります(5~6時間の睡眠が多い)。50代以降は、年齢が進むにつれて睡眠時間が延びる