



実際のコンピュータでは

有名なアメリカの数学者であるクロード・シャノンは、ジャグラーや一輪車乗りとしても有名です。そして、ここで学んだゲームのような実験を数多く手掛けました。彼は情報量をビット（「はい・いいえ」による決定の回数）で測りました。メッセージの情報量は、既に知っていることが何であるかによって違ってくることを発見しました。ある質問をすることによって、他の質問をしないでよくなることは、よくあることです。この場合、しないでよくなったメッセージに含まれる情報量は少ないのです。たとえば、コイン投げの結果は、裏か表かですから、普通1ビットです。でも、コインが「いかさま」で、10回中9回も表が出るなら、その情報は1ビットよりも少なくなります。どうやって、1回のコイン投げの結果を1回未満の質問で知ることができるのでしょうか。単純です。次のような質問をすればよいのです。「2回のコイン投げの結果はどちらも表だったか?」。いかさまコインの結果を書き並べると、こういった「表表」になることは、だいたい80%の回数で起こります。20%は「いいえ」なので、この場合ももっと質問をする必要があります。でも平均してみれば、コイン投げの回数よりも少ない回数で、いかさまコイン投げの結果を当てることができるでしょう。



シャノンは、メッセージの情報量のことを「エントロピー」と呼びました。「エントロピー」は、事象の数、つまりコイン投げの場合は、表・裏なので2だけではなく、それが起こる「確率」にも影響を受けます。ありえない出来事、つまりびっくりするような情報は、そのメッセージを当てるのに多くの回数の質問が必要です。ちょうど学校に来るのにヘリコプターを使ったときのように、それまで知らなかった情報をつけ加えているのですから。

メッセージのエントロピーは、コンピュータ科学者にとって大変重要です。エントロピーよりも小さな大きさにメッセージを圧縮することはできません。また、最良の圧縮法は、数を当てるゲームと同じです。答えを何ビットかの情報として記憶しておくことで、決定木を使ってコンピュータのプログラムは推測を行なうことができます。だから、情報を元のように再構築することができるのです。圧縮システムの中には、テキストファイルのファイルサイズを4分の1にまで小さくできるものもあります。これは、記憶容量の節約に大いに役立っています。

数を当てるゲームの方法は、コンピュータインタフェースにも使われています。利用者が次にどのキーを押すかを予想して候補を表示することにより、身体に障害があってキー操作が不自由な人でも平均で2回の「はい・いいえ」の答えだけでローマ字1字を打つことができます。この仕組みは携帯電話の文字入力にも使われています



解答とヒント

1回の「はい・いいえ」で答える回答は1ビットの情報に対応しています。その質問が「50よりも大きいですか」といった単純なものか、「20と60の間ですか」といったもう少し複雑なものであるかには関係しません。

数を当てるゲームでは、特定の方法で質問を選んできると、答えの並びを2進法で表したものと一致します。3は2進法では011ですが、決定木での質問を並べると「いいえ、はい、はい」になります。「いいえ」の代わりに0、「はい」の代わりに1を書いたときの結果は、3を2進法で表したものと同じです。

年齢を当てる決定木は、小さな数の方へ片寄るでしょう。

文の中の伏せ字を当てる決定木は、前の文字が何であるかに関係します。これは英語の場合、どの文字の次にどの文字が来やすいかという統計が研究として確立しているからです。日本語の場合にも、前後にある既知の字と辞書を使って候補を絞ることができます。