

[(本学科 2 年生向け) セミナール選択のための資料 (情報理論について)]

- ① 書店で例えばアルバイト情報誌などの本を立ち読みして、その内容を知っても、立ち読みされた情報誌は物理的には劣化しませんが、読んだ人にとっては、その情報誌の価値は下がります。
- ② 相手に、同じ内容の情報を伝えるだけならば、簡潔明瞭がよいですね。従量制のパケット代は安くて済みますから。(最近では定額制が世の主流のようですが。)
- ③ 待ち合わせの約束をした時に、2019 年 5 月 1 日 (水) と、わざわざ (水) を追記するのは何故でしょうか。

上記の①~③と似たような日常の話題は数多くあるでしょう。これらをもう少し、学問的に捉え直してみましよう。

- ① ある物理量を科学的に取り扱うためには、例えば時間を測るために時計があり、温度を測るのに温度計があるように簡単に入手可能な道具で、かつその量についていつでも同じ測定が可能な客観的基準が必要です。数学においても、例えば、ベクトル空間に内積と呼ばれる機能を導入することによって、ベクトルの長さや異なるベクトルの距離や角度が定義されユークリッド幾何学を代数解析的な形式で展開できました。では、情報という概念を数量的に扱うにはどうしたら良いのでしょうか。この疑問に答えるが如く、1948 年に Shannon はエントロピーという概念を用いて見事に情報を数量化してみせたのです。(そしてこの発見はその後の情報科学社会の発展に貢献してきました)しかも、情報量が満たすべき 3 つ (または 4 つ) の公理からエントロピーは定数倍を除いて一意に定まるのです。これは、一意性定理と呼ばれています。さらに、アルバイト情報誌を読む前よりも、読んだ後の方がより正確な情報を得たことになります。これは、相互情報量という概念です。つまり、通信路において入力情報が雑音や損失などの影響を受けてどれだけ正確に出力に側に受信されたかを表す量で、下記③の第二符号化定理に関係する量です。
- ② 圧縮ソフトはご存知でしょうか。LHA とか JPG や MP3 などです。情報量を圧縮して効率的に資源(記憶装置)を使いましょうという話です。圧縮にも種類があります。100% 誤り無く元に戻せるもの(可逆圧縮)とそうでないもの(非可逆圧縮)があります。文書等は、可逆圧縮でないと困りますね。しかし、人間の聴力で聞き分けられないほどの情報はバツサリ切って圧縮してしまって元に戻らなくても良さそうです。なので、MP3 などは非可逆圧縮です。実は、可逆圧縮の場合の圧縮限界は①で一意に定まったエントロピーなのです。これは、情報源符号化定理 (Shannon の第一符号化定理) と呼ばれるものです。
- ③ 日付だけでなく、曜日も併記することにより、その日にちに誤りがあった場合に発見できますよね。これを、誤り検出といいます。例えば、13 桁のバーコードや書籍に使われている ISBN の 10 桁 (最近では 13 桁) のうち最後の数字の 1 桁はチェックディジットと言われ、誤り検出のために付いています。②では、出来る限り圧縮して無駄を省きましたが、③では逆に無駄を追加して、誤りを検出、或いは、訂正できる符号を構成しようという話です。そして、そのような符号が存在することを保証しているのが、通信路符号化定理 (Shannon の第二符号化定理) と呼ばれるものです。