

図1

A=01010101010101010101010101010101
01010101010101010101010101010101...

B=110010010000111111011010101010
00100010000101101000110000...

C=1110000011000000111011101000
11011100111001100001011001...

末にコルモゴロフ複雑性と呼ばれる概念にたどり着きます。ランダムな列とはそのコルモゴロフ複雑性が高い列です。

もう1人はアメリカの数学者ソロモノフ（1926-2009）です。ソロモノフは初期の人工知能研究（1950年代〜60年代）の中で、良い予測ができることと

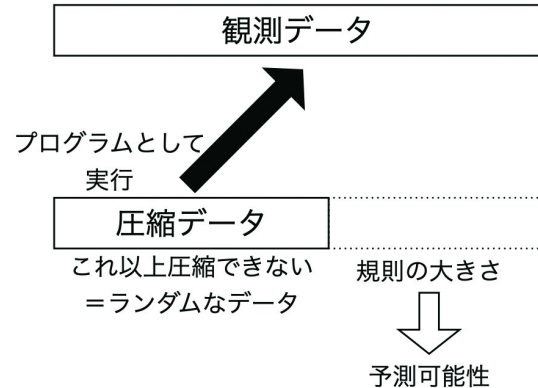
大きく圧縮できることが同じであることを見抜き、コルモゴロフ複雑性の概念にコルモゴロフより若干早く独立にたどり着いています。

0と1の有限列に対して、そのコルモゴロフ複雑性とは、その文字列を出力するプログラムのうち文字列としての最短の長さ（ビット）を言います。最初の例のAの10億桁（100メガバイトくらい）を出力するプログラムは非常に短く書けます。しかし規則が分らない10億桁の列を出力するためには、その列そのものを情報として持つ必要があります。プログラムも10億桁くらいの長さが必要になります。コルモゴロフ複雑性はデータの複雑さや情報量を表現しているのです。

私が最近興味を持って研究しているのは、最善な予測が持つ性質です。計算に必要な時間やメモリ容量を忘れてデータが大量にある場合の予測の極限の振る舞いを考えると、適当な条件のもとでは最

善な予測が存在することを証明できます。その最善な予測は、より単純な数理モデルに、より高い確率を付与したものです。「計算により規則がどのように発見されるのか」に関する理解が深まることで、より汎用的な機械学習のアルゴリズムの発見につながると信じています。

最後に私のゼミや研究の進め方を紹介します。数学のゼミは輪講形式が一般的で、事前に決めた教科書を読んできて、学生が順番にその内容を説明します。教員や他の学生から不十分な点を指摘されたり、関連項目について質問されたりして、理解を深めます。オンラインでも事前に準備したノートやタブレットやペンタブで書いたりして、同じことを行っています。他の研究者と議論する時も同様で



す。最近はオンラインでの研究会が多く参加しやすくなった反面、時差のために夜中になるなど、時間と気持ちの管理に苦慮しています。また長時間画面に向かうことになり、大きな画面などの情報機器環境や、部屋の中で体を動かせる環境の整備が重要になっていると思います。

研究最前線

THE FRONT LINE OF RESEARCH

理工学部

ランダムの使い方

PROFILE

宮部 賢志

MIYABE Kenshi

理工学部准教授
専門：計算論、確率論

1982年 岐阜県生まれ
2005年 京都大学理学部卒業
2010年 京都大学大学院理学研究科博士後期課程修了 博士(理学)
2014年 明治大学理工学部専任講師
2017年から現職

主な著書・論文

『新しい微積分<上><下>』(共著・講談社・2017年)
『確率と哲学』(監訳・九夏社・2020年)
『確率統計入門』(共著・森北出版・2020年)

所属学会

日本数学会、日本応用数理学会、Association for Symbolic Logic、Association Computability in Europe

私はランダム性について研究しています。ランダムとは、秩序や規則が見つけられず、予測が難しい状態を言います。データを理解したり、予測したりするのに必要な概念の一つです。

何かがランダムであるとはどういうことでしょうか。例えば、図1のような0と1からなる無限の

列A、B、Cがあったとします。Aは0と1が交互に並んでいるのでランダムではないでしょう。BとCには一見して規則がなさそうです。実はBは円周率 π （円周の直径に対する比率で3.14159...）を通常の10進法ではなく2進法で書いたものです。Cは計算機でつくった乱数（メルセンヌ・

ツイスタ）です。最初はランダムに見えても規則が見つかるランダムに見えなくなるように、ランダム性は知識や計算能力などに相対的な概念です。

ランダム性に自然な数学的定義を与えるのに貢献した2人の数学者を紹介します。1人目はロシアの数学者コルモゴロフ（1903

ー1987）です。コルモゴロフは若い頃に『確率論の基礎概念』（1933）を著し、確率概念を公理化する（数学的に議論できるようにする）ことに成功します。また、自身の確率の公理を現実の問題に応用する時に、与えられたデータの複雑さを数値で表現したという動機から、1960年代