

もっと紹介！ Standard シリーズ「MATH CONNECT」

章扉と章末コラムで、大学入学共通テストでも重視されている日常とのつながりを学ぶことができます。

本資料では、昨今注目されている **データサイエンス** の題材をご紹介します。

数学 I



▲ 数学 I Standard p.169

具体的な数値や表で説明することで
読み進めやすくなっています。

実際の運用についても触れ、
日常と数学のつながりを、よりリアルに
感じられるようにしています。

題材に関する章扉イラストによって
親しみやすくしています。

身近な疑問を取り上げることで、
学習への意欲を引き出します。

この章の内容の概要や意義を
示しています。

▼ 数学 I Standard p.196

196

MATH CONNECT
マスコネクト

動画共有サイトのおすすめ機能と相関係数

インターネットの動画共有サイトで動画を見ていくと、利用者の好みに合わせておすすめの動画が次々と表示される。動画共有サイトには、膨大な数の動画があるが、その中からどのようにして一人一人の利用者に合わせたおすすめの動画を選び出しているのだろうか。

利用者の行動履歴から、傾向が近い利用者の評価を参考に動画を推薦するという方法がある。このような推薦アイテムの選び方を協調フィルタリングといい、その仕組みには相関係数が使われている。その仕組みを単純化してみよう。

まず、利用者のサイト内の行動履歴から、利用者の見た動画に対する評価を数値化する。仮にこれが0から5までの整数で数値化されていたとする。このとき、①から⑩までの10本の動画に対する利用者P、A、B、C、Dの評価が次の表のようになっているとして、利用者Pへのおすすめ動画を考える。ここで、「－」と書かれているのは、未視聴を意味する。

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	Pとの相関係数
P	2	－	5	0	3	－	5	1	－	4	
A	－	5	1	4	4	4	1	0	3	1	-0.388
B	－	1	3	1	5	4	5	2	5	3	0.714
C	3	4	4	2	2	3	4	3	4	2	0.556
D	4	1	5	－	4	3	4	2	－	3	0.672
推薦度		2				3.3			4.5		

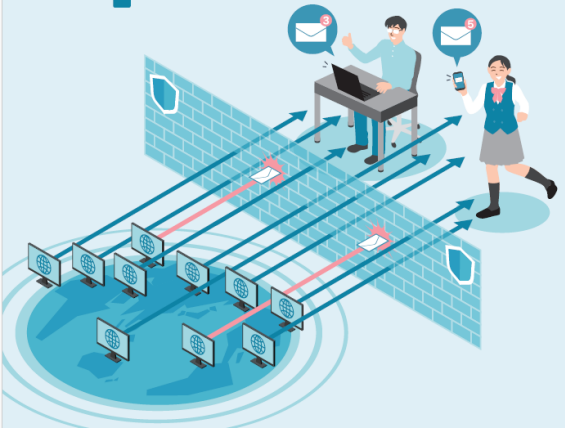
次に、利用者Pと他の利用者の評価傾向の類似性を調べる。例えば、PとAに対して、2人とも見たことのある動画に対する評価の相関係数が計算できる。この値が大きいほど、2人の評価傾向が似ていることになる。

このように調べると、B、C、Dとの相関係数が0.5以上と大きくなっている。そこで、Pが未視聴の動画②、⑥、⑨に対する、B、C、Dの評価の平均値（ただし、未視聴の動画は除く）を調べ、これを「推薦度」とすると、動画⑨の推薦度が最も高い。こうして、Pへのおすすめ動画として⑨が選出できる。

実際には、これを基本としてさまざまな手法が組み合わせられて運用されている。

5

1 章 場合の数と確率



迷惑メールの疑いのある電子メールを、自動で検知する機能がある。

➡ p.64 迷惑メールのフィルタリング機能と条件付き確率

この章では、日常生活のさまざまな場面で活用されている
5 「場合の数と確率」を学んでいこう。

▲ 数学 A Standard p.5

64

MATH CONNECT
マスコネクト

迷惑メールのフィルタリング機能と条件付き確率

電子メールの中には、勧誘や宣伝を目的として、受信者の意向を無視して無差別に送られてくるものがある。このような迷惑メール（スパムメール）を、正当なメールと区別するには、どのような判定法が用いられているのだろうか。

ここでは、これまでの電子メールの情報から、届いたメールが
5 スパムメールである確率は 40 %
スパムメールであるとき、本文に特定の単語 W が含まれる確率は 80 %
正当なメールであるとき、本文に特定の単語 W が含まれる確率は 5 %
であると分かっているとしよう。

	W が含まれる確率	W が含まれない確率
スパムメール（メール全体の 40 %）	80 %	20 %
正当なメール（メール全体の 60 %）	5 %	95 %

ここで、新たに届いた電子メールの本文に単語 W が含まれていたとする。このとき、この電子メールがスパムメールである条件付き確率を求めてみよう。

電子メールがスパムメールであるという事象を A、電子メールの本文に単語 W が含まれるという事象を B とおくと

$$P(A \cap B) = \frac{40}{100} \times \frac{80}{100} = \frac{3200}{10000}$$

同様に考えて

$$P(\bar{A} \cap B) = \frac{60}{100} \times \frac{5}{100} = \frac{300}{10000}$$

よって

$$P(B) = P(A \cap B) + P(\bar{A} \cap B) = \frac{3200}{10000} + \frac{300}{10000} = \frac{3500}{10000}$$

したがって、求める条件付き確率は

$$P_B(A) = \frac{P(A \cap B)}{P(B)} = \frac{3200}{10000} \div \frac{3500}{10000} = \frac{32}{35} \approx 0.914 \dots$$

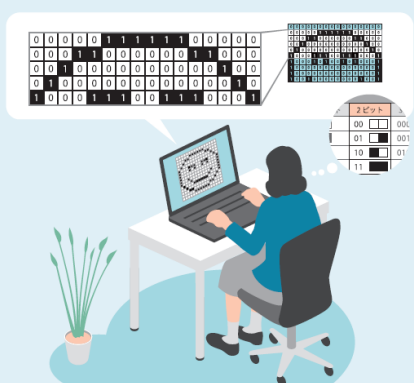
ゆえに、この電子メールがスパムメールである確率は約 90 % である。

このような仕組みをベイジアンフィルタという。実際には、「特定の単語 W」のように 1 つだけでなく、複数の単語についての確率を用いるなどして運用されている。

▲ 数学 A Standard p.64

113

3 章 数学と人間の活動



コンピュータの内部の計算では、整数の性質が活用されている。

➡ p.160 コンピュータの計算と 2 進法

5 この章では、整数の性質とその日常への活用など、さまざまな「数学と人間の活動」について、理解を深めていこう。

▲ 数学 A Standard p.113

160

MATH CONNECT
マスコネクト

コンピュータの計算と 2 進法

この章で学んだ 2 進法を用いると、数を 0 と 1 の並びで表すことができる。コンピュータでは、電気信号や電圧の有無で 0 と 1 を表し、2 進法を利用して数の記録・計算・処理を行っている。

コンピュータの内部では、負の数をどのように扱っているのだろうか。

2 進法で「4 桁以下の正負の整数」を扱うとしよう。そのとき、コンピュータは「5 桁の整数」として記録することで、正と負を区別して扱う。

例えば、正の数 $a = 9 = 1001_{(2)}$ について、コンピュータは、この数の左に「0」をつけた 5 桁の数「01001₍₂₎」として記録する。

そして、負の数 $-a$ を、コンピュータは

さらに 1 つ桁が多い 6 桁の数 100000₍₂₎ から正の数 a を引いた数として記録する。この数を a の補数という。

正の数 $a = 9 = 1001_{(2)}$ について

$$100000_{(2)} - 1001_{(2)} = 10111_{(2)}$$

であるから、「10111₍₂₎」が $-a$ として記録される。

なぜ、 $-a$ として a の補数が用いられるのだろうか。

$b = 13 = 1101_{(2)}$ について、「 $b - a$ 」と「 $b + (a \text{ の補数})$ 」の計算を比較して考えてみよう。

$\begin{array}{r} 01101 \leftarrow b \\ - 01001 \leftarrow a \\ \hline 00100 \end{array}$	$\begin{array}{r} 01101 \leftarrow b \\ + 10111 \leftarrow a \text{ の補数} \\ \hline 100100 \end{array}$
---	--

下から 5 桁目までは、ともに「00100₍₂₎」となり、一致している。

一般に、「 $b - a$ 」と「 $b + (a \text{ の補数})$ 」の計算について、上から 1 桁目を消去すると、これらの結果が一致することが知られている。

このように、補数を用いることで、コンピュータの内部で引き算を足し算として計算でき、演算回路が簡単になったり、演算速度が速くなったりするメリットがある。

▲ 数学 A Standard p.160



本社
支社・出張所

〒114-8524 東京都北区堀船 2-17-1 Tel : 03-5390-7320（高校教育部）
札幌 011-562-5721 仙台 022-297-2666 東京 03-5390-7467 金沢 076-222-7581 名古屋 052-950-2260
大阪 06-4967-1356 広島 082-568-2577 福岡 092-771-1536 鹿児島 099-213-1770 那覇 098-834-8084