

技術士一次試験基礎科目 この問題が難しい R01 空間距離

令和1年度

I-2-3 表1は、文書A～文書F中に含まれる単語とその単語の発生回数を示す。ここでは問題を簡単にするため、各文書には単語1、単語2、単語3の3種類の単語のみが出現するものとする。各文書の特徴を、出現する単語の発生回数を要素とするベクトルで表現する。文書Aの特徴を表すベクトルは  $\vec{A} = (7, 3, 2)$  となる。また、ベクトル  $\vec{A}$  のノルムは、 $\|\vec{A}\|_2 = \sqrt{7^2 + 3^2 + 2^2} = \sqrt{62}$  と計算できる。

2つの文書Xと文書Y間の距離を(式1)により算出すると定義する。2つの文書の類似度が高ければ、距離の値は0に近づく。文書Aに最も類似する文書はどれか。

表1 文書と単語の発生回数

	文書A	文書B	文書C	文書D	文書E	文書F
単語1	7	2	70	21	1	7
単語2	3	3	3	9	2	30
単語3	2	0	2	6	3	20

$$\text{文書Xと文書Yの距離} = 1 - \frac{\vec{X} \cdot \vec{Y}}{\|\vec{X}\|_2 \|\vec{Y}\|_2} \quad (\text{式1})$$

(式1)において、 $\vec{X} = (x_1, x_2, x_3)$ 、 $\vec{Y} = (y_1, y_2, y_3)$ であれば、

$$\vec{X} \cdot \vec{Y} = x_1 \cdot y_1 + x_2 \cdot y_2 + x_3 \cdot y_3, \quad \|\vec{X}\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2}, \quad \|\vec{Y}\|_2 = \sqrt{y_1^2 + y_2^2 + y_3^2}$$

- ① 文書B    ② 文書C    ③ 文書D    ④ 文書E    ⑤ 文書F

R01-1-2-3 正答 ③

空間距離に関する問題です。

この問題を解くにはひらめきが必要です。ひらめけば一瞬にして答に至ります。

文書D/文書Aで、単語1 =  $21 / 7 = 3$ 、同じく単語2 =  $9 / 3 = 3$ 、単語3 =  $6 / 2 =$

3 となっていますので、文書 A と文書 D の空間距離はゼロとなります。

表 1 文書と単語の発生回数

	文書A	文書B	文書C	文書D	文書E	文書F
単語1	7	2	70	21	1	7
単語2	3	3	3	9	2	30
単語3	2	0	2	6	3	20

$$\text{文書 X と文書 Y の距離} = 1 - \frac{\vec{X} \cdot \vec{Y}}{\|\vec{X}\|_2 \|\vec{Y}\|_2} \quad (\text{式 1})$$

(式 1) において、 $\vec{X} = (x_1, x_2, x_3)$ 、 $\vec{Y} = (y_1, y_2, y_3)$  であれば、

$$\vec{X} \cdot \vec{Y} = x_1 \cdot y_1 + x_2 \cdot y_2 + x_3 \cdot y_3, \quad \|\vec{X}\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2}, \quad \|\vec{Y}\|_2 = \sqrt{y_1^2 + y_2^2 + y_3^2}$$

確認のために実際に計算してみました。文書 A と文書 D 間の空間距離がゼロであることがわかります。

空間距離 (計算結果)

	文書A	文書B	文書C	文書D	文書E	文書F
文書A	0.000	0.190	0.089	0.000	0.355	0.381
文書B	0.190	0.000	0.410	0.190	0.407	0.215
文書C	0.089	0.410	0.000	0.089	0.687	0.759
文書D	0.000	0.190	0.089	0.000	0.355	0.381
文書E	0.355	0.407	0.687	0.355	0.000	0.076
文書F	0.381	0.215	0.759	0.381	0.076	0.000

(補足)

a を定数として、 $y_1 = a x_1$ 、 $y_2 = a x_2$ 、 $y_3 = a x_3$

$$\vec{X} \cdot \vec{Y} = a (x_1^2 + x_2^2 + x_3^2)$$

$$\|\vec{X}\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2}, \quad \|\vec{Y}\|_2 = a \sqrt{x_1^2 + x_2^2 + x_3^2}$$

文書 X と文書 Y の距離

$$= 1 - a (x_1^2 + x_2^2 + x_3^2) / a (x_1^2 + x_2^2 + x_3^2)$$

$$= 0$$