Fuzzy 集合を用いた情報検索モデル

橋　本　　　寛

1．はじめに

収集した一定の情報を後の検索に便利な形で蓄積しておき，必要に応じ検索することを通常情報検索と呼んでいる。ここでは，情報として科学技術関係の文献，または論文のようなものを考える。したがって，ここでの情報検索とは文献検索の意味である。また検索方式としては，文献の内容を示すキーワードというものを各文献に与えて蓄積しておき，文献をさがす場合は，そのキーワードを手がかりとして検索する方式を考える。このキーワード方式の文献検索に関して情報検索の数学的モデルを構成し，その若干の性質を述べる。モデルの構成にあたってはFuzzy集合の理論を採用する。Fuzzy集合はL.A.Zadeh〔1〕によって提出されたものである。この集合は境界の明確でない集合として考えることができ，その理論はあいまいさのあるもの，または2値的に決めがたいものをとりあつかうのに都合がよい。これまでFuzzy集合の理論を種々の分野へ適用することが試みられている。しかし，最もよく適用できる分野とおもわれる情報検索に関してはほとんど適用されておらず，また以下で述べるような形で，その理論を積極的に活用したモデルはみられない。

従来の文献検索では，その文献の内容を表わすキーワード間の関連というものが十分考慮されていない。そのためあるキーワードを手がかりにして文献をさがしている場合，それと同義のキーワードまたは関連のあるキーワードがあっても，機械的検索においては無視されてしまい，その関連のあるキーワードの付けられている文献はとり出せないことになる。したがって，そ

のような文献をもとり出すためにはキーワード間の関連を考慮して検索することが必要である。しかしながら，キーワード間の関連をとり入れた検索手法には従来満足すべきものがなかった。これに対して，ここではFuzzy集合の理論を採用して，キーワード間の関連を考慮する検索手法を構成したものである。

そのキーワード間の関連を考慮するために，関連度行列なるものを導入し，これによってキーワード間の関連を表示する。一方，文献の検索要求は論理関数の形の質問式で表現し検索を実行するが，検索の実行にあたってはFuzzy集合の演算を採用する。Fuzzy集合の演算は人間の感覚と大体一致しているとおもわれ，またその性質も容易に知ることができるので都合がよい。

以下で述べるようなモデルは，従来提案されておらず，今後の情報検索の研究に役立つものと考えられる。しかし，意味の問題のような情報検索の本質的困難を解決するにはこのモデルだけでは不十分であり，幅広い接近法がのぞまれる。

2．演算の定義

x，y を0と1の間の数値をとる変数とするとき，$x \vee y$，$x \wedge y$ および $\bar{x}$ をつぎのように定義する。

$$x \vee y \equiv \max\{x, y\}, \quad x \wedge y \equiv \min\{x, y\}, \quad \bar{x} \equiv 1 - x$$

たとえば

$$0.8 \vee 0.5 = \max\{0.8, 0.5\} = 0.8, \quad 0 \wedge 1 = \min\{0, 1\} = 0$$

$$\overline{0.7} = 1 - 0.7 = 0.3$$

また各要素が0と1の間の数値であるベクトル $\mathbf{x}$，$\mathbf{y}$ に関しても同様に定義する。すなわち

$$\mathbf{x} = [x_1, x_2, \cdots, x_n]', \quad \mathbf{y} = [y_1, y_2, \cdots, y_n]'$$

に対して

$$\mathbf{x} \vee \mathbf{y} = [x_1 \vee y_1, x_2 \vee y_2, \cdots, x_n \vee y_n]'$$

$$\mathbf{x} \wedge \mathbf{y} = [x_1 \wedge y_1, x_2 \wedge y_2, \cdots, x_n \wedge y_n]'$$

$$\bar{\mathbf{x}} = [\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n]'$$

ただし，“′”は転置を示す。

たとえば　$\mathbf{x} = [0, 0.4, 1]'$，$\mathbf{y} = [0.2, 0.5, 0]'$　のとき

$$\mathbf{x} \vee \mathbf{y} = [0 \vee 0.2,\ 0.4 \vee 0.5,\ 1 \vee 0]' = [0.2,\ 0.5,\ 1]'$$

$$\mathbf{x} \wedge \mathbf{y} = [0 \wedge 0.2,\ 0.4 \wedge 0.5,\ 1 \wedge 0]' = [0,\ 0.4,\ 0]'$$

$$\bar{\mathbf{x}} = [\bar{0},\ \overline{0.4},\ \bar{1}]' = [1,\ 0.6,\ 0]'$$

さらに，行列間の演算に関しても同様に定める。各要素が0と1の間の数値をとる行列 $A = [a_{ij}]$，$B = [b_{ij}]$，$C = [c_{ij}]$ に対して

$$A \vee B = [a_{ij} \vee b_{ij}],\quad A \wedge B = [a_{ij} \wedge b_{ij}],\quad \bar{A} = [\bar{a}_{ij}]$$

$$A \circ C = [(a_{i1} \wedge c_{1j}) \vee (a_{i2} \wedge c_{2j}) \vee \cdots \vee (a_{im} \wedge c_{mj})]$$

ただし，A，Bは（l，m）次行列，Cは（m，n）次行列とする。このとき，A∘Cは（l，n）次行列となる。すなわち通常の行列積の積和を∧，∨でおきかえて定義したものである。

たとえば

$$A = \begin{bmatrix} 1 & 0.8 \\ 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 0.3 & 0.5 \\ 0 & 1 \end{bmatrix}$$

に対しては

$$A \vee B = \begin{bmatrix} 1 & 0.8 \\ 0 & 1 \end{bmatrix} \quad A \wedge B = \begin{bmatrix} 0.3 & 0.5 \\ 0 & 0 \end{bmatrix} \quad \bar{A} = \begin{bmatrix} 0 & 0.2 \\ 1 & 1 \end{bmatrix}$$

$$A \circ B = \begin{bmatrix} 1 & 0.8 \\ 0 & 0 \end{bmatrix} \circ \begin{bmatrix} 0.3 & 0.5 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} (1 \wedge 0.3) \vee (0.8 \wedge 0), & (1 \wedge 0.5) \vee (0.8 \wedge 1) \\ (0 \wedge 0.3) \vee (0 \wedge 0), & (0 \wedge 0.5) \vee (0 \wedge 1) \end{bmatrix} = \begin{bmatrix} 0.3 & 0.8 \\ 0 & 0 \end{bmatrix}$$

以下においては，この行列算を用いてモデルを構成する。それによってモデルがきわめて簡明に表現できる。

3．各種行列と質問式

蓄積されている文献の集合を $\{D_1, D_2, \cdots, D_m\}$ とし，またそれらに付けられているキーワードの集合を $\{K_1, K_2, \cdots, K_n\}$ とする。以下，モデルの

構成において必要ないくつかの定義をおこなう。

(1)　文献キーワード行列

文献とキーワードの間の関係を示す文献キーワード行列Ａの要素a_{ij}はつぎのようにして定義される。

$$a_{ij} = \begin{cases} 0 \cdots\cdots \text{文献}D_i\text{にキーワード}K_j\text{が付けられていないとき} \\ \\ 1 \cdots\cdots \text{文献}D_i\text{にキーワード}K_j\text{が付けられているとき} \end{cases}$$

たとえば，今５個の文献と４個のキーワードがあり，それらの間の関係がつぎの文献キーワード行列Ａで与えられたとする。

$$A = \begin{array}{c} \\ D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{array} \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & K_3 & K_4 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

このとき，この行列から文献D_1には２つのキーワードK_1とK_2が付けられており，またキーワードK_1は文献D_1，D_3，D_4に付けられていることがわかる。さらに，キーワードK_1とK_2が同時に付けられている文献としてはD_1とD_3があることなどがわかる。

(2)　キーワード関連度行列と文献関連度行列

キーワードK_iとK_jとの間の関連の度合，すなわち関連度を r_{ij} で示す。この r_{ij} は０と１の間の数値をとるものとし，関連の大きい場合ほどその値は１に近くなるものとする。またK_jとK_iとの間の関連度はK_iとK_jとの間の関連度に等しいものと考え，自分自身に対する関連度は１とする。すなわち

$$0 \leqq r_{ij} \leqq 1, \quad r_{ij} = r_{ji}, \quad r_{ii} = 1$$

この関連度r_{ij}を（ｉ，ｊ）要素とする行列をキーワード関連度行列とよび，以下行列Ｒで示す。

同様にして文献間の関連度行列を考え，それを行列Ｓで示す。

なお，この行列ＲおよびＳをどのようにして求めるかについては，ここでは議論しない。何か適当な方法で，たとえば統計的な手法ですでに求められているものと仮定する。

(3)　質問式

文献の要求者から与えられた要求を分析して，それをキーワードに関する論理式で表現したものを検索の質問式とよぶ。この質問式Qはつぎのようにしてつくる。

あるキーワードK_iに関連の大きい文献をさがすときは

$$Q=K_i$$

2つのキーワードK_iとK_jのすくなくとも一方に関連の大きい文献をさがすときは

$$Q=K_i\vee K_j$$

2つのキーワードK_iとK_jの両方に関連の大きい文献をさがすときは

$$Q=K_i\wedge K_j$$

あるキーワードK_iに関連の小さい文献をさがすときは

$$Q=\bar{K}_i$$

とする。またこれらを組み合わせてさらに複雑な質問式をつくることができる。たとえば

$$Q=K_1\vee(K_2\wedge\bar{K}_3)$$

これはK_1に関連の大きいか，またはK_2に関連が大きくかつK_3に関連の小さい文献をさがす場合の質問式である。

一般に，キーワード$K_1, K_2, \cdots, K_n$から構成される質問式Qを

$$Q=Q(K_1, K_2, \cdots, K_n)$$

で表わす。なお，質問式$Q(K_1, K_2, \cdots, K_n)$の中には必ずしもすべてのキーワード$K_1, K_2, \cdots, K_n$が含まれているわけではない。その中のいくつかのキーワードを用いて構成されているものと考える。これは添字を簡略化するためである。

4．検索手順とモデル

(1) 演算子としてのキーワード

文献キーワード行列AおよびキーワードK_iに対して　AK_i

で行列Aのｉ番目の列ベクトルを示すことにする。またRK_iや（A∘R）K_iについても同様で，それらはそれぞれ行列RおよびA∘Rのｉ番目の列ベクトルを示すものとする。

たとえば

$$A=\begin{array}{c}\\ D_1\\ D_2\\ D_3\\ D_4\\ D_5\end{array}\begin{array}{c}\begin{array}{cccc}K_1 & K_2 & K_3 & K_4\end{array}\\ \begin{bmatrix}1 & 1 & 0 & 0\\ 0 & 1 & 1 & 0\\ 1 & 1 & 0 & 1\\ 1 & 0 & 1 & 1\\ 0 & 0 & 0 & 1\end{bmatrix}\end{array}
\qquad
R=\begin{array}{c}\\ K_1\\ K_2\\ K_3\\ K_4\end{array}\begin{array}{c}\begin{array}{cccc}K_1 & K_2 & K_3 & K_4\end{array}\\ \begin{bmatrix}1 & 0 & 0 & 0.5\\ 0 & 1 & 0.1 & 0.4\\ 0 & 0.1 & 1 & 0.8\\ 0.5 & 0.4 & 0.8 & 1\end{bmatrix}\end{array}$$

とするとき

$$AK_2=\begin{bmatrix}1\\1\\1\\0\\0\end{bmatrix}\qquad RK_3=\begin{bmatrix}0\\0.1\\1\\0.8\end{bmatrix}$$

また

$$A\circ R=\begin{array}{c}\\ D_1\\ D_2\\ D_3\\ D_4\\ D_5\end{array}\begin{array}{c}\begin{array}{cccc}K_1 & K_2 & K_3 & K_4\end{array}\\ \begin{bmatrix}1 & 1 & 0 & 0\\ 0 & 1 & 1 & 0\\ 1 & 1 & 0 & 1\\ 1 & 0 & 1 & 1\\ 0 & 0 & 0 & 1\end{bmatrix}\end{array}\circ\begin{array}{c}\begin{array}{cccc}K_1 & K_2 & K_3 & K_4\end{array}\\ \begin{bmatrix}1 & 0 & 0 & 0.5\\ 0 & 1 & 0.1 & 0.4\\ 0 & 0.1 & 1 & 0.8\\ 0.5 & 0.4 & 0.8 & 1\end{bmatrix}\end{array}=\begin{array}{c}\begin{array}{cccc}K_1 & K_2 & K_3 & K_4\end{array}\\ \begin{bmatrix}1 & 1 & 0.1 & 0.5\\ 0 & 1 & 1 & 0.8\\ 1 & 1 & 0.8 & 1\\ 1 & 0.4 & 1 & 1\\ 0.5 & 0.4 & 0.8 & 1\end{bmatrix}\end{array}$$

$$(A\circ R)K_4=\begin{bmatrix}0.5\\0.8\\1\\1\\1\end{bmatrix}$$

(2)　文献とキーワード間の関連

以下において，キーワードK_iとK_jとの間の関連度 r_{ij} を $r(K_i,\ K_j)$で示す。

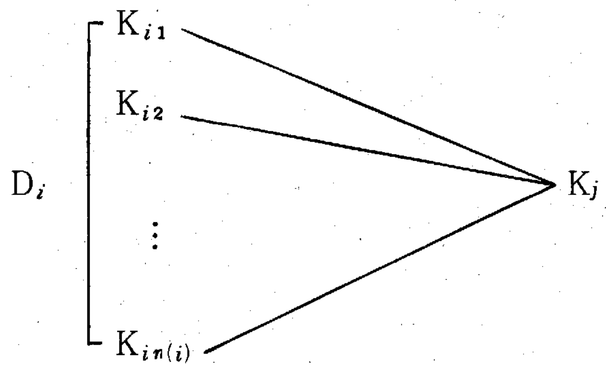
文献D_iに付けられているキーワードを K_{i1}, K_{i2}, …, $K_{in(i)}$ とするとき，このキーワードの集合 $\{K_{i1},\ K_{i2},\ \cdots,\ K_{in(i)}\}$ とキーワードK_jとの間の関連の度合を

$$r(\{K_{i1},\ K_{i2},\ \cdots,\ K_{in(i)}\},\ K_j)\quad または\quad r(D_i,\ K_j)$$

で示し，これをつぎのように定義する。

$$\max\{r(K_{i1}, K_j), r(K_{i2}, K_j), \cdots, r(K_{in(i)}, K_j)\}$$

すなわち，その文献D_iに付けられているキーワードの中で，K_jに対し最も大きな関連度を示すもののその値をもって，文献D_iとキーワードK_jとの間の



関連の度合とするものである。

たとえば，文献とキーワードの関係およびキーワード間の関連度がつぎの行列AおよびRで与えられているとする。

$$A = \begin{array}{c} \\ D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{array} \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & K_3 & K_4 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \qquad R = \begin{array}{c} \\ K_1 \\ K_2 \\ K_3 \\ K_4 \end{array} \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & K_3 & K_4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0.5 \\ 0 & 1 & 0.1 & 0.4 \\ 0 & 0.1 & 1 & 0.8 \\ 0.5 & 0.4 & 0.8 & 1 \end{bmatrix} \end{array}$$

このとき

$$r(D_1, K_1) = r(\{K_1, K_2\}, K_1) = \max\{r(K_1, K_1), r(K_2, K_1)\}$$
$$= \max\{1, 0\} = 1$$

$$r(D_3, K_3) = r(\{K_1, K_2, K_4\}, K_3) = \max\{r(K_1, K_3), r(K_2, K_3),$$
$$r(K_4, K_3)\} = \max\{0, 0.1, 0.8\} = 0.8$$

上の例からわかるように，文献D_3は$\{K_1, K_2, K_4\}$なるキーワードの集合で表現されているが，これとキーワードK_3との間の関連の度合は，集合$\{K_1, K_2, K_4\}$の中にK_3と関連の深いキーワードK_4が含まれているために，その中にK_3というキーワードは含まれていないにもかかわらず，0.8という値になる。このようなことは検索においてきわめて都合のよいことである。

ここで，先に定めた文献キーワード行列Aの要素a_{ij}を$a(D_i, K_j)$と書く

ことにすると，文献D_iに付けられているキーワードがK_{i1}，K_{i2}，…，$K_{in(i)}$であるからp＝1，2，…，n(i)に対しては

$$a(D_i,\ K_{ip})=1$$

である。

一方，集合$\{K_{i1},\ K_{i2},\ \cdots,\ K_{in(i)}\}$に含まれていないキーワードすなわち文献$D_i$に付けられていないキーワード$K_l$に対しては

$$a_{il}=a(D_i,\ K_l)=0$$

である。

したがって，文献D_iとキーワードK_jとの間の関連の度合 $r(D_i,\ K_j)$ はつぎのようにも表わされる。

$$\begin{aligned}
r(D_i,K_j)&=\max\{r(K_{i1},K_j),\ r(K_{i2},K_j),\cdots,\ r(K_{in(i)},K_j)\}\\
&=\max\{a_{i1}\wedge r(K_1,K_j),a_{i2}\wedge r(K_2,K_j),\cdots,\ a_{in}\wedge r(K_n,K_j)\}\\
&=(a_{i1}\wedge r_{1j})\vee(a_{i2}\wedge r_{2j})\vee\cdots\vee(a_{in}\wedge r_{nj})\\
&=[a_{i1},a_{i2},\cdots,a_{in}]\circ[r_{1j},r_{2j},\cdots,r_{nj}]
\end{aligned}$$

ここでベクトル$[a_{i1},\ a_{i2},\ \cdots,\ a_{in}]$および$[r_{1j},\ r_{2j},\ \cdots,\ r_{nj}]'$について考えると，前者は文献キーワード行列Aの第i行目のベクトル，また後者はキーワード関連度行列Rの第j列目のベクトルである。したがって，$r(D_i,\ K_j)$をi番目の要素とする列ベクトル$\mathbf{b}_j$を考えれば，$\mathbf{b}_j$は

$$\mathbf{b}_j=\begin{bmatrix} r(D_1,K_j)\\ r(D_2,K_j)\\ \cdots\cdots\cdots\\ r(D_m,K_j)\end{bmatrix}=\begin{bmatrix}(a_{11}\wedge r_{1j})\vee(a_{12}\wedge r_{2j})\vee\cdots\vee(a_{1n}\wedge r_{nj})\\ (a_{21}\wedge r_{1j})\vee(a_{22}\wedge r_{2j})\vee\cdots\vee(a_{2n}\wedge r_{nj})\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ (a_{m1}\wedge r_{1j})\vee(a_{m2}\wedge r_{2j})\vee\cdots\vee(a_{mn}\wedge r_{nj})\end{bmatrix}$$

$$=\begin{bmatrix}a_{11} & a_{12}\cdots a_{1n}\\ a_{21} & a_{22}\cdots a_{2n}\\ \cdots & \cdots\cdots\\ a_{m1} & a_{m2}\cdots a_{mn}\end{bmatrix}\circ\begin{bmatrix}r_{1j}\\ r_{2j}\\ \cdots\\ r_{nj}\end{bmatrix}$$

$$=A\circ(RK_j)=(A\circ R)K_j$$

で表わされる。さらに$\mathbf{b}_j$を j 番目の列ベクトルとする行列Bを考えれば，行列Bは

$$\begin{aligned} B &= [\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_n] = [A \circ (RK_1), A \circ (RK_2), \cdots, A \circ (RK_n)] \\ &= A \circ [RK_1, RK_2, \cdots, RK_n] \\ &= A \circ R \end{aligned}$$

で表わされる。この行列Bはキーワード間の関連を考慮した場合の文献キーワード行列と考えることができる。

文献とキーワード間の関連の度合を上のように定義することが実際の場合に妥当かどうかは検討を要する。この定義を変えれば，ここで述べるモデルとは別のモデルが構成され，それに応じた結果が得られるであろう。

(3) 文献と質問式との間の関連

これはつぎのように定める。ある文献D_iとキーワードK_1, K_2, …, K_nから構成される質問式$Q = Q(K_1, K_2, \cdots, K_n)$との間の関連の度合$r(D_i, Q)$は

$$\begin{aligned} r(D_i, Q) &\equiv r(D_i, Q(K_1, K_2, \cdots, K_n)) \\ &\equiv Q(r(D_i, K_1), r(D_i, K_2), \cdots, r(D_i, K_n)) \end{aligned}$$

で与えられるとする。

たとえば

$$r(D_i, K_1 \vee K_2) = r(D_i, K_1) \vee r(D_i, K_2)$$

$$r(D_i, K_1 \wedge \bar{K}_2) = r(D_i, K_1) \wedge \overline{r(D_i, K_2)}$$

具体的に，文献キーワード行列Aおよびキーワード関連度行列Rがつぎのように与えられているものとする。

$$A = \begin{array}{c} \\ D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{array} \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & K_3 & K_4 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \qquad R = \begin{array}{c} \\ K_1 \\ K_2 \\ K_3 \\ K_4 \end{array} \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & K_3 & K_4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0.5 \\ 0 & 1 & 0.1 & 0.4 \\ 0 & 0.1 & 1 & 0.8 \\ 0.5 & 0.4 & 0.8 & 1 \end{bmatrix} \end{array}$$

このとき，文献と質問式との間の関連の度合の例として$r(D_1, K_1 \vee K_2)$と$r(D_5, K_1 \wedge \bar{K}_2)$を求めてみる。

$$r(D_1, K_1 \vee K_2) = r(D_1, K_1) \vee r(D_1, K_2)$$
$$= r(\{K_1, K_2\}, K_1) \vee r(\{K_1, K_2\}, K_2)$$
$$= (r(K_1, K_1) \vee r(K_2, K_1)) \vee (r(K_1, K_2) \vee r(K_2, K_2))$$
$$= (1 \vee 0) \vee (0 \vee 1) = 1 \vee 1 = 1$$

$$r(D_5, K_1 \wedge \bar{K}_2) = r(D_5, K_1) \wedge \overline{r(D_5, K_2)}$$
$$= r(\{K_4\}, K_1) \wedge \overline{r(\{K_4\}, K_2)}$$
$$= r(K_4, K_1) \wedge \overline{r(K_4, K_2)}$$
$$= 0.5 \wedge \overline{0.4} = 0.5 \wedge 0.6 = 0.5$$

さらに，質問$Q = K_1 \wedge K_3$について，すべての文献との間の関連の度合を計算してみるとつぎの表のようになる。

	$r(D_i,\ K_1)$	$r(D_i,\ K_3)$	$r(D_i,\ K_1 \wedge K_3)$
D_1	1	0.1	0.1
D_2	0	1	0
D_3	1	0.8	0.8
D_4	1	1	1
D_5	0.5	0.8	0.5

この　$r(D_i,\ K_1 \wedge K_3)$の大きい文献ほど質問$Q = K_1 \wedge K_3$によく適合した文献であると考え，文献をとり出す場合はこの値の大きいものから順に必要な個数だけとり出すものとする。一般の質問Qに対しても同様に考える

また，　$r(D_i,\ Q)$ をｉ番目の要素とする列ベクトル$\mathbf{r}$ を考えると，$\mathbf{r}$ は

$$\mathbf{r} = \begin{bmatrix} r(D_1, Q) \\ r(D_2, Q) \\ \cdots\cdots \\ r(D_m, Q) \end{bmatrix} = \begin{bmatrix} r(D_1, Q(K_1, K_2, \cdots, K_n)) \\ r(D_2, Q(K_1, K_2, \cdots, K_n)) \\ \cdots\cdots \\ r(D_m, Q(K_1, K_2, \cdots, K_n)) \end{bmatrix}$$

$$= \begin{bmatrix} Q(r(D_1, K_1), r(D_1, K_2), \cdots, r(D_1, K_n)) \\ Q(r(D_2, K_1), r(D_2, K_2), \cdots, r(D_2, K_n)) \\ \cdots\cdots \\ Q(r(D_m, K_1), r(D_m, K_2), \cdots, r(D_m, K_n)) \end{bmatrix}$$

$$= Q\left(\begin{bmatrix} r(D_1, K_1) \\ r(D_2, K_1) \\ \vdots \\ r(D_m, K_1) \end{bmatrix}, \begin{bmatrix} r(D_1, K_2) \\ r(D_2, K_2) \\ \vdots \\ r(D_m, K_2) \end{bmatrix}, \cdots, \begin{bmatrix} r(D_1, K_n) \\ r(D_2, K_n) \\ \vdots \\ r(D_m, K_n) \end{bmatrix}\right)$$

$$= Q\ (\mathbf{b}_1,\ \mathbf{b}_2,\ \cdots,\ \mathbf{b}_n)$$

$$= Q((A \circ R)K_1, (A \circ R)K_2, \cdots, (A \circ R)K_n)$$

で与えられる。

（例）

$$A = \begin{array}{c} \\ D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{array}\begin{array}{c} \begin{array}{cccc} K_1 & K_2 & K_3 & K_4 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \qquad R = \begin{array}{c} \\ K_1 \\ K_2 \\ K_3 \\ K_4 \end{array}\begin{array}{c} \begin{array}{cccc} K_1 & K_2 & K_3 & K_4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0.5 \\ 0 & 1 & 0.1 & 0.4 \\ 0 & 0.1 & 1 & 0.8 \\ 0.5 & 0.4 & 0.8 & 1 \end{bmatrix} \end{array}$$

$$Q = K_1 \wedge K_2$$

$$\mathbf{r} = \begin{bmatrix} r(D_1, K_1\wedge K_2) \\ r(D_2, K_1\wedge K_2) \\ r(D_3, K_1\wedge K_2) \\ r(D_4, K_1\wedge K_2) \\ r(D_5, K_1\wedge K_2) \end{bmatrix} = \begin{bmatrix} r(D_1, K_1)\wedge r(D_1, K_2) \\ r(D_2, K_1)\wedge r(D_2, K_2) \\ r(D_3, K_1)\wedge r(D_3, K_2) \\ r(D_4, K_1)\wedge r(D_4, K_2) \\ r(D_5, K_1)\wedge r(D_5, K_2) \end{bmatrix}$$

$$= \begin{bmatrix} r(D_1, K_1) \\ r(D_2, K_1) \\ r(D_3, K_1) \\ r(D_4, K_1) \\ r(D_5, K_1) \end{bmatrix} \wedge \begin{bmatrix} r(D_1, K_2) \\ r(D_2, K_2) \\ r(D_3, K_2) \\ r(D_4, K_2) \\ r(D_5, K_2) \end{bmatrix} = (A \circ R)K_1 \wedge (A \circ R)K_2$$

ここで

$$\mathbf{r} = Q((A \circ R)K_1, (A \circ R)K_2, \cdots, (A \circ R)K_n)$$

の右辺を

$$(A \circ R)Q(K_1, K_2, \cdots, K_n) \text{ または } (A \circ R)Q$$

と書くことにする。

たとえば

$$(A \circ R)K_1 \vee (A \circ R)K_2 = (A \circ R)(K_1 \vee K_2)$$

$$\overline{(A \circ R)K_1} = (A \circ R)\overline{K_1}$$

さらに，A ∘ Rに相当する部分を行列Wで示すと，結局検索の結果$\mathbf{r}$ は

$$\mathbf{r} = WQ$$

で与えられる。これがここで目的としたFuzzy集合を用いた情報検索のモデルである。

また文献間の関連度行列Sを考慮する場合の検索の結果$\mathbf{r}$ は

$$\mathbf{r} = (S \circ A \circ R)\ Q$$

で与えられる。この場合もS ∘ A ∘ Rに相当する部分を行列Wで示せば，結果$\mathbf{r}$ は

$$\mathbf{r} = WQ$$

となる。文献間の関連を考慮しないとき，すなわち行列Sが単位行列であれば

$$W = S \circ A \circ R = A \circ R$$

となる。

次節で，このモデルのいくつかの性質を列挙する。なお，$\mathbf{r} = (A \circ R)Q$ においてRが単位行列であれば，$\mathbf{r} = AQ$となり，これは従来の同義語をみとめない場合の2値的検索モデルに相当し，Rが対称行列であって0，1の要素をもち対角要素がすべて1であれば，同義語をみとめる場合の2値的検索モデルに相当する。前者後者ともにこのとき，$\mathbf{r}$ は0，1の2値のみからなるベクトルとなる。

（例）

$$A = \begin{array}{c} \\ D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{array} \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & K_3 & K_4 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

(i)　$Q = K_1 \wedge K_2$のとき

$$\mathbf{r}=AQ=A(K_1\wedge K_2)=AK_1\wedge AK_2=\begin{bmatrix}1\\0\\1\\1\\0\end{bmatrix}\wedge\begin{bmatrix}1\\1\\1\\0\\0\end{bmatrix}=\begin{array}{c}D_1\\D_2\\D_3\\D_4\\D_5\end{array}\begin{bmatrix}1\\0\\1\\0\\0\end{bmatrix}$$

したがって，この場合質問$Q=K_1\wedge K_2$に適合する文献として，D_1とD_3がとり出される。

(ii) $Q=K_1\vee\bar{K}_1$のとき

$$\mathbf{r}=AQ=A(K_1\vee\bar{K}_1)=AK_1\vee\overline{AK_1}=\begin{bmatrix}1\\0\\1\\1\\0\end{bmatrix}\vee\begin{bmatrix}0\\1\\0\\0\\1\end{bmatrix}=\begin{array}{c}D_1\\D_2\\D_3\\D_4\\D_5\end{array}\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}$$

この質問$Q=K_1\vee\bar{K}_1$に対してはすべての文献が該当することになる。それは質問の意味を考えても当然である。このように，2値的な場合では補元律が成立する。

(ii) $Q=K_1\wedge\bar{K}_1$のとき

$$\mathbf{r}=AQ=A(K_1\wedge\bar{K}_1)=AK_1\wedge\overline{AK_1}=\begin{bmatrix}1\\0\\1\\1\\0\end{bmatrix}\wedge\begin{bmatrix}0\\1\\0\\0\\1\end{bmatrix}=\begin{array}{c}D_1\\D_2\\D_3\\D_4\\D_5\end{array}\begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix}$$

この場合にはいかなる文献もこの質問に該当しない。

一方，Fuzzy集合の演算で注意しなければならないことは通常の集合で成立する補元律が一般には成立しないことである。たとえば質問式を$K_j\wedge\bar{K}_j$として検索したとき，上の例で示すように従来の検索では何も検索されないが，Fuzzy集合の演算を採用した場合は必ずしもそうではない。それは$K_j\wedge\bar{K}_j$の解釈がことなるからでなる。すなわち，2値的な場合であれば質問式$K_j\wedge\bar{K}_j$はキーワードK_jに関連が有りかつK_jに関連の無い文献，いいかえるとキーワードK_jが付けられている文献であってかつK_jの付けられていない文献をさがせという質問式であって，内容的に矛盾しているものである。これに対し，

Fuzzy集合の演算を採用したときは，$K_j \wedge \bar{K}_j$はK_jに関連の大きい文献であって，またかつK_jに関連の小さい文献をさがせということになり，任意の文献とその質問式との間の関連の度合は一般に次の関係で与えられる。

$$0.5 \geqq r(D_i,\ K_j \wedge \bar{K}_j) \geqq 0$$

すなわち，必ずしも0にはならないわけである。また質問式が$K_j \vee \bar{K}_j$の場合も同様であって，そのときは

$$1 \geqq r(D_i,\ K_j \vee \bar{K}_j) \geqq 0.5$$

となる。上の関係は基本的にはつぎの性質によるものである。

$1 \geqq x \geqq 0$のとき

$$0.5 \geqq x \wedge \bar{x} \geqq 0,\quad 1 \geqq x \vee \bar{x} \geqq 0.5$$

たとえば，$x = 0.6$のとき

$$x \wedge \bar{x} = 0.6 \wedge \overline{0.6} = 0.6 \wedge 0.4 = 0.4$$

$$x \vee \bar{x} = 0.6 \vee \overline{0.6} = 0.6 \vee 0.4 = 0.6$$

各文献に対する上の関係を例で示すとつぎのようになる。

（例）

$$A = \begin{array}{c} \\ D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{array} \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & K_3 & K_4 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \qquad R = \begin{array}{c} \\ K_1 \\ K_2 \\ K_3 \\ K_4 \end{array} \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & K_3 & K_4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0.5 \\ 0 & 1 & 0.1 & 0.4 \\ 0 & 0.1 & 1 & 0.8 \\ 0.5 & 0.4 & 0.8 & 1 \end{bmatrix} \end{array}$$

$$A \circ R = \begin{array}{c} \\ D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{array} \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & K_3 & K_4 \end{array} \\ \begin{bmatrix} 1 & 1 & 0.1 & 0.5 \\ 0 & 1 & 1 & 0.8 \\ 1 & 1 & 0.8 & 1 \\ 1 & 0.4 & 1 & 1 \\ 0.5 & 0.4 & 0.8 & 1 \end{bmatrix} \end{array}$$

(i)　$Q = K_1 \vee \bar{K}_1$のとき

$$\mathbf{r} = (A \circ R)Q = (A \circ R)(K_1 \vee \bar{K}_1) = (A \circ R)K_1 \vee \overline{(A \circ R)K_1}$$

$$= \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0.5 \end{bmatrix} \vee \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0.5 \end{bmatrix} = \begin{array}{c} D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{array} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0.5 \end{bmatrix}$$

(ii)　$Q=K_1 \wedge \bar{K}_1$のとき

$$\mathbf{r}=(A \circ R)Q=(A \circ R)(K_1 \wedge \bar{K}_1)=(A \circ R)K_1 \wedge \overline{(A \circ R)K_1}$$

$$= \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0.5 \end{bmatrix} \wedge \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0.5 \end{bmatrix} = \begin{array}{c} D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{array} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0.5 \end{bmatrix}$$

以上のことからわかるように $(A \circ R)(K_j \vee \bar{K}_j)$ の要素が常に1であるとはかぎらないし，また $(A \circ R)(K_j \wedge \bar{K}_j)$ が常に零ベクトルであるとはいえない。

5．主要な性質

以下に列挙する性質はいずれもほぼ自明であるから，いくつかのものを除いて，証明は省略する。Aは文献キーワード行列，R，R_1，R_2はキーワード関連度行列，Sは文献関連度行列，K_i，K_jはキーワードを示す。

(1)　$(A \circ R)K_i = A \circ (RK_i)$

$(A \circ R^{(2)})K_i = A \circ R \circ (RK_i)$　（ここに$R^{(2)}$は$R \circ R$を示す）

$(S \circ A \circ R)K_i = S \circ A \circ (RK_i)$

（証明）

$$\begin{aligned}(S \circ A \circ R)K_i &= ((S \circ A) \circ R)K_i \\ &= (S \circ A) \circ (RK_i) \\ &= S \circ A \circ (RK_i)\end{aligned}$$

(2)　$(A \circ R)(K_i \vee K_j) = (A \circ R)(K_j \vee K_i)$

$(A \circ R)(K_i \wedge K_j) = (A \circ R)(K_j \wedge K_i)$

(3)　$(A\circ R)\bar{K}_i=\overline{(A\circ R)}K_i$

(4)　$(A\circ(R_1\vee R_2))K_i=(A\circ R_1)K_i\vee(A\circ R_2)K_i$

$(A\circ(R_1\vee R_2))\bar{K}_i=\overline{(A\circ R_1}\wedge\overline{A\circ R_2)}K_i$

（証明）

$$
\begin{aligned}
(A\circ(R_1\vee R_2))\bar{K}_i&=\overline{(A\circ(R_1\vee R_2))K_i}\\
&=\overline{(A\circ R_1\vee A\circ R_2)K_i}\\
&=\overline{A\circ R_1\vee A\circ R_2}\,K_i\\
&=\overline{(A\circ R_1}\wedge\overline{A\circ R_2)}K_i
\end{aligned}
$$

(5)　$(A\circ R)\ (\bar{K}_i\wedge\bar{K}_j)=(\overline{A\circ R})\ (K_i\wedge K_j)=(A\circ R)\ \overline{(K_i\vee K_j)}$

（証明）

$$
\begin{aligned}
(A\circ R)\ (\bar{K}_i\wedge\bar{K}_j)&=\overline{(A\circ R)K_i}\wedge\overline{(A\circ R)K_j}\\
&=\overline{(A\circ R)}K_i\wedge\overline{(A\circ R)}K_j\\
&=\overline{(A\circ R)}\ (K_i\wedge K_j)
\end{aligned}
$$

$$
\begin{aligned}
\overline{(A\circ R)K_i}\wedge\overline{(A\circ R)K_j}&=\overline{(A\circ R)K_i\vee(A\circ R)K_j}\\
&=(A\circ R)\ \overline{(K_i\vee K_j)}
\end{aligned}
$$

(6)　$(A\circ(R_1\vee R_2))\ (K_i\vee K_j)=(A\circ R_1)\ (K_i\vee K_j)\vee(A\circ R_2)\ (K_i\vee K_j)$

（証明）

$$
\begin{aligned}
(A\circ(R_1\vee R_2))\ (K_i\vee K_j)&=(A\circ R_1\vee A\circ R_2)\ (K_i\vee K_j)\\
&=(A\circ R_1\vee A\circ R_2)K_i\\
&\qquad\vee(A\circ R_1\vee A\circ R_2)K_j\\
&=(A\circ R_1)K_i\vee(A\circ R_2)K_i\\
&\qquad\vee(A\circ R_1)K_j\vee(A\circ R_2)K_j\\
&=(A\circ R_1)\ (K_i\vee K_j)\\
&\qquad\vee(A\circ R_2)\ (K_i\vee K_j)
\end{aligned}
$$

6．ま　と　め

これまでの議論から明らかなように，Fuzzy 集合の理論は，その定義およ

び性質が検索モデルと直接対応しており，情報検索においてきわめて有用である。また，そのモデルをもとにして，Fuzzy集合の理論で議論されている以上の興味深い性質を得ることができる。

しかしながら，情報検索における主要な問題は文献に対しいかにしてキーワードを与えるか，またキーワード間の関連度をいかにして測るかということなどである。すなわちここでのモデルでいえば，文献キーワード行列のAとか，キーワード関連度行列のRをいかにして求めるかということである。ここでは，これらについては何も述べていない。これらの問題はきわめて困難な問題であるが，今後さらに議論を展開するためには，これらの問題に対しても有効であるようなモデルの構成が必要であろう。

なお本論文は東亜経済学会・山口経済学会合同定例研究会（昭和48年10月）における資料に加筆修正したものである。熱心にご討論いただいた諸氏に深謝する。

文　献

【1】 L. A. Zadeh : "Fuzzy Sets", *Information and Control* 8, 338−353 (1965)