

## 相互インデックス法とベクトル演算によるファジィシソーラス生成法

木 瀬 若 桜<sup>†</sup> 田 中 秀 俊<sup>†</sup>  
佐 藤 裕 幸<sup>†</sup> 野 本 弘 平<sup>††</sup>

事例ベース推論などにおける類似事例検索では、データベースに蓄積された情報について、検索条件に対する一致の度合いを評価し、それをを用いて検索を行っている。先に提案した相互インデックス法では、各事例に複数のインデックスが設定されており、インデックス間の相互関係のファジィシソーラスを生成することで検索を実現している。しかし、現在利用しているファジィシソーラスでは検索結果に差がつきにくいという課題が存在している。この課題を解決するため、本稿ではベクトル演算によりファジィシソーラスを生成する手法を提案する。実データを用いて検証を行った結果、提案手法では類似度のばらつきが大きくなり、所期の目的が達成できることが確認された。

## A Study on Fuzzy Thesaurus Using Vector Operation for Cross Index Method

WAKASA KISE<sup>†</sup>, HIDETOSHI TANAKA<sup>†</sup>, HIROYUKI SATO<sup>†</sup>  
and KOHEI NOMOTO<sup>††</sup>

We presented a fuzzy thesaurus generation algorithm based on cross-occurrence frequency for similar case retrieval. The algorithm was applied to a decision support system to plan a rescue operation against a disaster and was evaluated using 200 real cases. Then useful cases were retrieved by the resultant fuzzy thesaurus, but a strange phenomenon was found. Degree of a relation between a pair of index attribute value will be unnaturally weak, if their occurrence frequencies are very different each other. In order to this problem, a modified algorithm using vector operation is shown.

### 1. ま え が き

人間が意思決定を行う場合には、現在の状況に類似した過去の事例を参考にする場合が多い。このような場合に利用されるアプローチの一つとして、現在の状況に類似した過去の事例を検索する類似事例検索がある。このような類似事例検索のためにファジィシソーラスによる方法が有効な手段として存在する。しかし、キーワードの同時出現にもとづく文献検索の分野で標準的なファジィシソーラスは、ここで検討の対象としている複数のインデックスにそれぞれ1個のインデックス値が付与されている状況には対応できない。そこで、我々はあるインデックスに属するインデックス値間の相互関係を評価し、ファジィシソーラスを生成するため相互インデックス法を提案している<sup>1)2)</sup>。

試作システムを作成し、この方法により実際の事例群に対する評価を行ったところ、検索結果に差がつきにくいという課題が存在することが判明した。これに加え、インデックス値対の出現頻度の差が大きい場合、関連度が不自然に低く算出されるという現象が見られた。

上記の課題を解決するため、本稿ではベクトル演算によりファジィシソーラスを生成する手法を提案する。

### 2. 相互出現頻度に基づくファジィシソーラス生成法

過去の事例には、複数のインデックスが設定されており、インデックス  $k$  にはそれぞれに属する  $m_k$  個のインデックス値の集合の中から1個のインデックス値  $i_k$  が付与されているものとする(図1)。

類似事例検索は、各インデックスのインデックス値を指定することにより行われるため、シソーラスは各インデックス毎に必要である。

しかし、あるインデックス値(これを対象インデックスと記す)に属するインデックス値のシソーラスを生成しようとしても、同一のインデックスに属するインデッ

<sup>†</sup> 三菱電機株式会社 情報技術総合研究所  
Information Technology R&D Center, Mitsubishi Electric Corporation  
<sup>††</sup> 三菱電機株式会社 デザイン研究所  
Industrial Design Center, Mitsubishi Electric Corporation

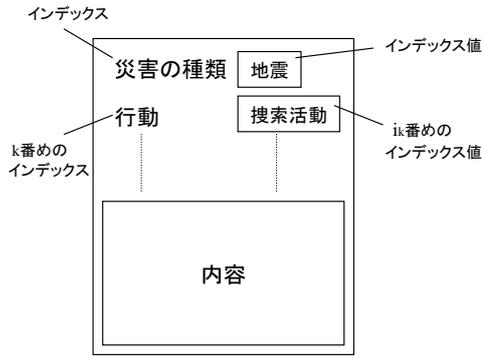


図1 事例のインデックス構造

クス値は各事例にそれぞれ1個排他的に付与されている。そこで、我々は他のインデックス(これを参照インデックスと記す)を利用して、同一の事例における相互出現頻度を利用してファジィシソーラスを生成する相互インデックス法を提案している<sup>1)</sup>。

まず、対象インデックス  $k$  と参照インデックス  $l$  の事例への出現状況を表す行列(これを出現行列と記す)  $A(k)$ ,  $A(l)$  をそれぞれ生成する。

次に、対象インデックスと参照インデックスの出現行列から相互出現頻度行列  $C(k, l)$  を作成する。

$$C(k, l) = A(k) \oplus A(l)^T \quad (1)$$

$A \oplus B$  は行列  $A$  と行列  $B$  の sum-min 合成と呼ばれる演算であり、式(2)のように定義される。

$$[A \oplus B]_{i,k} = \sum_j ([A]_{i,j} \wedge [B]_{j,k}) \quad (2)$$

但し、 $[A]_{i,j}$  は行列  $A$  の  $(i, j)$  要素である。

上記の相互出現頻度行列を利用して、対象インデックスに属するインデックス値間の関係、すなわち間接的同時出現行列  $G(k|l)$  を式(3)のように定義する。

$$G(k|l) = C(k, l) \oplus C(k, l)^T \quad (3)$$

さらに、この間接的同時出現行列を各インデックス値の出現頻度により正規化し、ファジィ関連語行列  $R(k|l)$  を式(4)のように算出する。

$$[R(k|l)]_{i_k, j_k} = \min\{[\Psi(k|l)]_{i_k, j_k}, [\Psi(k|l)]_{j_k, i_k}\} \quad (4)$$

但し、 $[\Psi(k|l)]_{j_k, i_k}$  は以下のように与えられる。

$$[\Psi(k|l)]_{i_k, j_k} = \begin{cases} \frac{[G(k|l)]_{i_k, j_k}}{[G(k|l)]_{i_k, i_k}} & ; [G(k|l)]_{i_k, i_k} \neq 0 \\ 1 & ; [G(k|l)]_{i_k, i_k} = 0 \end{cases}$$

このような従来の相互インデックス法においては、関連度を集合論的定義にもとづいて求めている。まず、事例は各キーワードやインデックス値の概念を具体的に示す指標であると考え、従来手法では、概念  $A$  と概念  $B$  との共通部分  $A \cap B$  が概念  $A$  に含まれる割合と、概念  $B$  に含まれる割合のうちの小さい方を関連度とす

る(図2)。相互インデックス法では、まず2つのインデックスにおいてそれぞれのインデックス値の出現行列を作成し、これらを基に相互出現行列を求める。さらにこの相互出現行列を利用してインデックス値間の関係の強さを示す間接的同時出現行列を算出する。これらを正規化することによりファジィ関連語行列を算出し、ファジィシソーラスとして利用する。

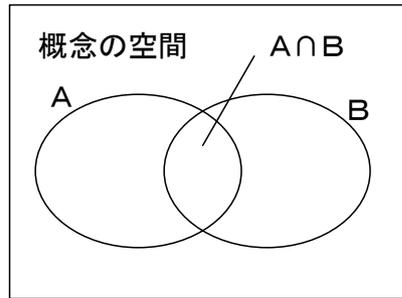


図2 インデックス値の概念の包含関係

これら従来手法についてファジィシソーラスを生成したところ、以下のような課題があることが判明した。

- 生成される類似度の数値にあまり差がつかず、検索結果に差がつきにくい。
  - 出現頻度の差が大きいインデックス対の関連度が不自然に低く算出される。
- この原因としては、次のようなことが考えられる。
- 集合論的定義にもとづきファジィシソーラスを求めている。
  - sum-min 合成を連続して行っていることにより、どのシソーラスも類似した値となってしまう。

### 3. ベクトル演算によるファジィシソーラス生成法

上記課題を解決するため、本稿では、相互インデックス行列からファジィ関連語行列を算出する際に、従来利用していた集合論的定義にもとづく sum-min 合成に代えて、ベクトル演算による手法によりファジィシソーラスを生成する方法を提案する。

文献検索等の分野において関連度を算出する際に、上記の集合論的定義と同様に用いられる手法として、ベクトル演算、その中でもベクトル間の差を利用した方法が知られている。特に cos 関数を利用した方法は、文献におけるキーワード間の類似測度を算出する際に頻繁に利用される<sup>3)</sup>。これはベクトル全体の方向の差異により関連度を定義する手法であり、以降、この考え方をファジィシソーラス生成に利用することを検討する。

相互出現頻度行列の各行には対象インデックス、各列

には参照インデックスのインデックス値が対応している．これにより行列の値はインデックス間の同時出現頻度を示している．そこで、この相互出現頻度行列により、次のような方針によりファジィ関連語行列の算出を行う（図3）．

- 各行を一つのベクトルとみなし、相互出現頻度をベクトル間の内積により評価する．
- インデックス値が現れる事例数を利用して各ベクトルのノルムにより正規化し、関連度を算出する．

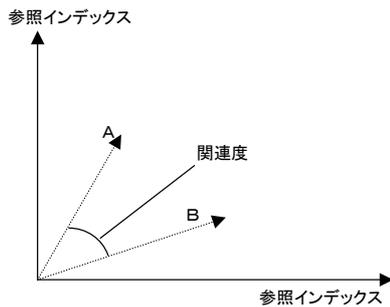


図3 ベクトルによるインデックス値間の評価

この方法に基づけば、ファジィ関連語行列  $R'(k|l)$  は以下のように相互出現頻度行列から直接求められる．

$$[R'(k|l)]_{i_k, j_k} = \begin{cases} \frac{\sum_{i_l=1}^{m_l} ([C(k|l)]_{i_k, i_l} \cdot [C(k|l)]_{j_k, i_l})}{N} & ; N \neq 0 \\ 1; N = 0 \text{ and } i_k = j_k & \\ 0; N = 0 \text{ and } i_k \neq j_k & \end{cases}$$

$$N = \left( \sum_{i_l=1}^{m_l} ([C(k|l)]_{i_k, i_l})^2 \right)^{\frac{1}{2}} \left( \sum_{j_l=1}^{m_l} ([C(k|l)]_{j_k, i_l})^2 \right)^{\frac{1}{2}}$$

以上のベクトル演算にもとづくファジィ関連語行列生成法のアルゴリズムを、図4に示す．

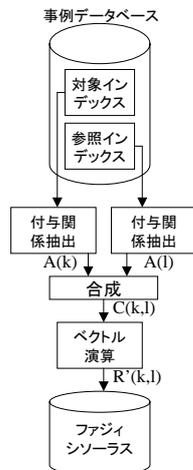


図4 ベクトル演算にもとづくファジィ関連語生成法

#### 4. 実データによる評価

災害派遣活動について類似事例検索システムを試作し<sup>1)</sup>、以上に述べた相互インデックス法による検証を行った．事例として、過去の記事を元に200件あまりを登録した．各事例には、災害の種類、発生時刻、発生場所、被害物、行動の5種類について、それぞれインデックス値が付与されている．その中で、参照インデックスを災害の種類ならびに被害物、対象インデックスを行動としたファジィシソーラスを従来手法と提案手法により求めた．さらに、算出したファジィシソーラスについて、対角成分を除く各行の関連度の平均と分散、同じく対角行を除くすべてのデータにおける平均と分散を算出した．なお、紙面の都合で各表の列のインデックス値は一部表示を省略した．

表1 災害の種類による行動のファジィ関連語行列（従来法）

	搜索	救助	監視	人員	復旧
搜索活動	1.00	0.21	0.14	0.37	0.39
救助活動	0.21	1.00	0.00	0.39	0.24
監視活動	0.14	0.00	1.00	0.39	0.15
人員輸送	0.37	0.39	0.39	1.00	0.37
復旧作業	0.39	0.24	0.15	0.37	1.00

表2 災害の種類による行動のファジィ関連語行列（提案法）

	搜索	救助	監視	人員	復旧
搜索活動	1.00	0.59	0.70	0.91	0.65
救助活動	0.59	1.00	0.00	0.36	0.62
監視活動	0.70	0.00	1.00	0.91	0.34
人員輸送	0.91	0.36	0.91	1.00	0.64
復旧作業	0.65	0.62	0.34	0.64	1.00

表3 被害物による行動のファジィ関連語行列（従来法）

	搜索	救助	監視	人員	復旧
搜索活動	1.00	0.21	0.14	0.37	0.41
救助活動	0.21	1.00	0.21	0.52	0.15
監視活動	0.14	0.21	1.00	0.30	0.24
人員輸送	0.37	0.52	0.30	1.00	0.44
復旧作業	0.41	0.15	0.24	0.44	1.00

表4 被害物による行動のファジィ関連語行列（提案法）

	搜索	救助	監視	人員	復旧
搜索活動	1.00	0.96	0.55	0.94	0.52
救助活動	0.96	1.00	0.35	0.85	0.29
監視活動	0.55	0.35	1.00	0.58	0.97
人員輸送	0.94	0.85	0.58	1.00	0.59
復旧作業	0.52	0.29	0.97	0.59	1.00

今回提案した、ベクトル演算によるファジィシソーラスならびに従来手法である集合論的定義に基づく手法による比較結果について考察する．

はじめに、平均値について考察する．提案手法では、従来手法によるファジィシソーラスと比較すると値は全体的に大きくなった．これは、従来手法では論理積により min をとっていたため、比較的值が小さく算出されがちであったということが背景にある．例として「災害

の種類」により評価された「行動」のファジィシソーラスで検証する。

まず、「行動」と「災害の種類」の間の相互出現頻度行列を表5に示す。

表5 「行動」と「災害の種類」の間の相互出現頻度行列

	暴風雨	豪雪	洪水	津波	地震
搜索活動	5	0	0	17	21
救助活動	1	0	0	1	6
監視活動	0	0	0	0	0
人員輸送	3	0	0	2	6
物資輸送	0	0	0	0	0
復旧作業	15	4	2	0	13

	噴火	火災	海難	飛行機事故	その他
搜索活動	29	5	3	8	2
救助活動	0	3	3	4	1
監視活動	13	0	0	0	0
人員輸送	17	2	1	2	0
物資輸送	0	0	0	0	0
復旧作業	8	7	2	0	3

表5より、インデックス値「搜索活動」と「救助活動」について、インデックス「災害の種類」に対する相互出現行列の値ならびに sum-min 合成による論理積、ならびに通常の乗算による積の結果得られるベクトルは、それぞれ次のとおりである。

表6 ベクトルの論理積ならびに通常の積

	暴風雨	豪雪	洪水	津波	地震
搜索	5	0	0	17	21
救助	1	0	0	1	6
論理積	1	0	0	1	6
通常積	5	0	0	17	126

	噴火	火災	海難	飛行機事故	その他
搜索	29	5	3	8	2
救助	0	3	3	4	1
論理積	0	3	3	4	1
通常積	0	15	9	32	2

表6より、搜索活動と救助活動の sum-min 合成の値は19になる。また、搜索活動の総和 (= 出現回数) は90、救助活動の総和は19となり、その結果として従来手法による関連度は0.21となる。

次に、今回利用しているベクトル演算の場合、内積ならびにノルムにより cos 値を算出することにより関連度を算出している。表6より、インデックス「災害の種類」に対する内積値は、積のベクトルの要素を足しあわせればよいため206となる。一方、搜索活動ならびに救助活動の要素の値から、それぞれのベクトルのノルムを算出すると、ほぼ41、8.5となる。従ってこの場合の関連度は0.59である。また、上記の内積を算出する際のそれぞれの要素毎の積をみると、地震活動が126と値が際立っており、これが全体の類似度を押し上げる結果となっている。一方、従来手法の場合にはこれが min 値

により決まるため、地震活動の場合でも6となり、これも関連度が低く算出される要因である。

このように、従来手法では、事例数が多いインデックス値と少ないインデックス値の組合せの場合には、結果的に事例数が少ないインデックス値の数に引きずられ、全体の値が小さくなってしまふ。これに対して、ベクトル演算の場合、ベクトル自体の方向を重視しており、相互インデックス行列の各要素が小さい値、すなわち事例の出現頻度が少ないインデックス値についても、互いのインデックス値の一方が0でなければ、出現頻度を考慮した関連度を算出することがわかる。

次に、算出された関連度行列の分散について考察する。今回算出したファジィシソーラスの各要素に対応する分散を表7に示す。

表7 ファジィシソーラスの分散

	搜索	救助	監視	人員	復旧	全体
表1	0.278	0.210	0.170	0.380	0.288	0.265
表2	0.713	0.393	0.488	0.705	0.563	0.572
表3	0.283	0.273	0.223	0.408	0.310	0.299
表4	0.743	0.613	0.613	0.740	0.593	0.660

今回生成したすべてのシソーラスについて、提案手法による分散が他の従来手法による分散を上回った。このことから、検索結果に差がつかないという状況はある程度改善されると推測される。この理由も、先ほどの平均値に対する考察と同様に、提案手法においては出現頻度の多寡を考慮してシソーラスを生成しており、sum-min 合成の低い要素の値に影響されずに関連度が算出されるためである。

## 5. む す び

ファジィシソーラスにおける関連度に差がつきにくいという課題を解決するため、集合論的定義にもとづく論理演算による手法に代わって基本的なベクトル演算により直接関連語行列を算出する方法を提案した。実データを用いた検証の結果、提案手法ではファジィ関連語行列のばらつきが大きくなり、所期の目的が達成できることがわかった。また、上記課題のみならず、各インデックス値の事例数の多寡を考慮したシソーラスを算出することが確認された。

## 参 考 文 献

- 1) 野本, 木瀬, 小菅, "類似事例検索のための相互出現頻度に基づくファジィシソーラスの生成法," 日本ファジィ学会誌, Vol.8, No.6, pp.1134-1143, 1996.
- 2) 野本, "正規化相互出現頻度に基づくファジィ関連語行列生成法," 第13回ファジィシステムシンポジウム, pp.211-214, 1997.
- 3) 伊藤, "情報検索", ソフトウェア講座 19, 昭晃堂, 1986.