

# 事象間の因果関係を扱う動的な文脈解釈機能を有する 意味的連想検索方式の実現

鷹野 孝典<sup>†</sup> 関子 泰三<sup>†</sup> 清木 康<sup>††</sup>

本稿では、事象に関する文脈解釈機能を有する事象間の因果関係計量による意味的連想検索方式を提案する。本方式は、同時あるいは一連に起こっている事象群を文脈（検索語列）として指定することにより、その事象群の原因または結果の事象群を動的に検索することを可能とする。本方式は、事象間の因果関係計量を行うためのベクトル空間の生成を、因果関係を表現するベクトル空間の自己相関行列を直交化して行うことにより、直接的な因果関係のみでなく、間接的な因果関係の計量を実現する。本方式により、ある事象について、直接的および間接的な因果関係にある事象群が記述された文書データを検索結果の上位に獲得することが可能となる。本稿では、擬似事象データおよび宇宙工学分野を対象とした事象データを用いた実験により、本方式の有効性、および実現可能性を検証する。

## A Semantic Associative Search Method with Dynamic Context-Awareness Functions for Computing Causal Relationships of Event Data Sets

KOSUKE TAKANO,<sup>†</sup> TAIZO ZUSHI<sup>†</sup> and YASUSHI KIYOKI<sup>††</sup>

In this paper we present a semantic associative search method with dynamic context-awareness functions for computing causal relationship between events. This method dynamically computes causal relationships between a “context” (query words) and event data sets in a search domain. A “context” represents the event data set to be simultaneously observed. In this method, an orthogonal space for search is generated from the self-correlation matrix of a causal relationship matrix, and this orthogonal search space enables to measure not only direct causal relationships, but also indirect ones. This method makes it possible to retrieve documents ranked higher which are described about an event data set, which has both direct and indirect causal relationships to a given event data set. We clarify effectiveness and feasibility of the method by experiments using a synthetic data set and a data set for an aerospace engineering.

### 1. はじめに

#### 1.1 背景

近年、高性能コンピュータの普及やデータベース技術の発達にともない、様々な組織において、大量の文書データ群が電子ファイルとして蓄積され、組織内の共有知識資源として活用するためにデータベースシステム上で管理されている。データベースや情報検索システムの研究分野においては、これらの文書データ群から目的に合った文書データを獲得するための情報検

索方式として、ベクトル空間モデルを用いた検索方式の有効性が示されている<sup>1),3)-5),8),9)</sup>。

一方、これまでのベクトル空間モデルを用いた検索方式は、データ間の意味的な等価性や類似性についての関連性計量を行う方式が提案されてきたが、データ間の因果関係についての関連性計量を行う方式は実現されていなかった。

SMART システム<sup>8),9)</sup> や LSI<sup>1)</sup> においては、文書データ群を文書データ中に出現する単語の出現頻度に基づいてベクトルデータ化する。SMART システムや LSI では、これらのベクトル表現されたデータに対して、ベクトル間の内積計算やコサイン距離計算を行うことにより、検索語と文書データ間の意味的な等価性や類似性についての静的な関係を計量し、検索者の与えた検索語について類似的関連性のある文書データの

<sup>†</sup> 慶應義塾大学政策・メディア研究科  
Graduate School of Media and Governance, Keio University

<sup>††</sup> 慶應義塾大学環境情報学部  
Faculty of Environmental Information, Keio University

検索を実現する．また、意味の数学モデルによる意味的連想検索方式<sup>3)-5)</sup>では、分野別の専門知識を利用して、その分野の「意味」を形式的に計量することのできるベクトル空間を生成し、そのベクトル空間において文脈解釈機能をともなったベクトル演算を行うことにより、利用者が指定した文脈に応じて、意味的に近い情報の動的な検索を実現している．しかし、ある事象の原因や結果について記述された文書データ群を検索するといった検索要求がある場合、単語間の類似性を算出するための計量系のみでは十分ではなく、事象間の因果関係を扱う計量系が必要となる．

このような背景から、筆者らは、これまで事象間の因果関係が計量可能なベクトル空間生成方式を提案した<sup>15)</sup>．文献 15) に示されている方式では、事象間の因果関係を表現する因果関係マトリクスの組に対して、ベクトル間の内積計算やコサイン距離計算といった相関量計算を行うことにより、事象間の因果関係計量を実現しており、検索者の検索目的に応じて、「ある事象の原因についての検索（原因検索）」、「ある事象の結果についての検索（結果検索）」、および「ある事象の原因である事象群/結果である事象群の両方についての検索（原因・結果検索）」といった目的別の因果関係検索を実現している．

本稿では、文献 15) で示される方式により生成した因果関係を計量するベクトル空間を意味の数学モデルによる意味的連想検索方式に適用することにより、事象に関する文脈解釈機能をともなった因果関係計量を行う意味的連想検索を実現する方式を提案する．

文献 15) では、発生した事象を示すためのキーワード群を問合せとして与え、そのキーワード群の表す事象の原因として相関の強い事象を表す文書群、および結果として相関量の強い事象を表す文書群を、それぞれ独立に検索結果の上位にランキングすることを可能とする相関量計算を実現する方式が提案されている．

本稿において示す方式は、このような原因となる事象、および結果となる事象についての独立な計量機能を有している．本方式では、さらに、問合せにおいて、発生した事象に加えて、その原因および結果を得るための状況を規定する文脈を与えることにより、文脈に応じた事象の原因や結果を究明することを可能とする．

文献 15) に対して、本方式は、問合せとして、原因を究明する場合においては、発生した事象を示すキーワードに加えて、その事象の原因となる事象群を特定するための状況を文脈として与え、その文脈において、キーワードが表す事象を引き起こす原因となる事象に関する文書群を検索結果の上位にランキングすること

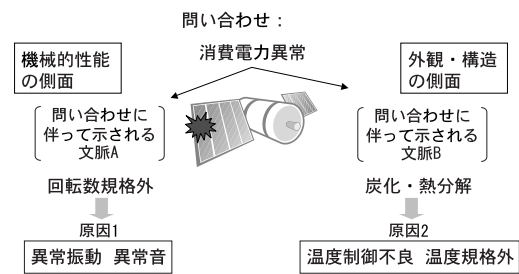


図 1 文脈解釈機能による動的な因果関係検索

Fig. 1 A search with context-awareness functions for dynamically computing causal relationships.

を可能とする方式である．この方式により、図 1 に示すように、様々な状況（文脈）における、その事象の原因となる事象群を各状況（文脈）ごとに独立に究明することが可能となる．発生した事象の結果を究明する場合においても、その結果となる事象群を特定するための状況を文脈として与えることにより、多様な状況に応じて、各々独立に、発生した事象の結果として引き起こされる事象に関する文書群を獲得し、発生する結果を状況ごとに究明することが可能となる．

本稿では、因果関係を扱うベクトル空間生成方式を、意味の数学モデルによる意味的連想検索方式に適用するための実現方式について示す．さらに、擬似事象データや宇宙工学分野を対象とした事象データを用いた実験により、提案方式の有効性、および実現可能性を検証する．

## 1.2 因果関係計量による検索例

ここでは、本方式における因果関係計量による検索例について述べる．本方式では、文書の内容を示す事象単語をメタデータを付与した文書データ群を対象として、因果関係計量による検索を行う．事象単語のメタデータ設定が適切な文書データの例として、システムの不具合情報等の事例データベースに格納されている文書データ群等がある．

以下に、本方式により実現される因果関係計量による検索例について具体的に述べる．

### 1.2.1 文脈解釈機能による動的な因果関係検索

本方式では、文脈解釈機能による動的な因果関係計量による情報検索を実現する．

図 1 は、消費電力異常というキーワードに加えて、異なる文脈を表す事象単語を指定した場合に、より状況に即した原因を検索している例を示している．この例では、文脈語として、機能的性能の側面を表す回転数規格外を指定した場合には、消費電力異常の原因として振動系の異常に関する「異常振動」や「異常音」といった事象を検索している．一方、文脈語として、

外観・構造の側面を表す炭化・熱分解を指定した場合には、原因として温度系の異常に関する「温度規格外」や「温度制御不良」といった事象を検索している。

このように、提案方式により実現した意味的検索空間を用いた検索では、ある事象について、原因・結果である事象群が、機械的性能、電気的性能、光学的性能等複数の側面を持つという観点で多義性を持っている場合において、発生した事象を示すキーワードに加えて、同時あるいは一連に起こっている事象群を文脈（検索語列）として指定することにより、文脈として与えられる事象群の原因または結果として相関の強い事象群を動的に検索することが可能となる。

### 1.2.2 間接的な因果関係検索

システムの不具合情報等の事例データベースにおいては、1つの文書中に、ある事象  $\varepsilon$  と直接的な原因や結果である事象群のみではなく、間接的な因果関係にある事象について述べてある文書データ群も多数存在すると考えられる。このような文書データを検索結果の上位に獲得するためには、事象間の直接的な因果関係のみでなく、間接的な因果関係を計量する方式が必要である。

本方式により実現される意味的検索空間は、事象間の因果関係を表現する因果関係マトリクス<sup>2)</sup>の自己相関行列を直交化することにより生成する。この際、ある事象と共通の原因や結果を持つ事象は、意味的検索空間上のある同一の直交軸に集約される傾向にある。この結果、本方式では、ある事象  $\varepsilon$  の原因検索や結果検索を行う場合に、事象  $\varepsilon$  と直接的な因果関係にある事象群だけでなく、直接的な因果関係がある事象群に対して、これらの事象群と共通の原因や結果を持つといった間接的な因果関係にある事象群についての因果関連性計量が可能となる。

## 2. 関連研究

因果関係に関する文書データ検索に応用可能な方式として、ベイジアン・ネットワーク<sup>2),7)</sup>がある。ベイジアン・ネットワークは、 $n$ 個の確率変数  $X_1, X_2, \dots, X_n$  間に成立する因果関係や相関関係といった依存関係を条件付き確率として設定し、その同時確率分布を有向グラフを用いて表現したものである。

ベイジアン・ネットワークを適用した文書データ検索方式は、検索問合せとして入力した事象単語について、同時確率が大きいと推定される事象について記述されている文書データを、適合文書データとして検索結果の上位にランキングする方式として実現される。したがって、ベイジアン・ネットワークによる検索方

式では、検索問合せとしてある事象単語を入力した場合に、その事象の原因または結果として、発生確率は低い、専門知識により重要であると判断されるような事象群（たとえば、爆発事故を連鎖的に起こす可能性のある等危険性の高い因果的関連事象）について記述された文書データを検索ランキングの上位に獲得することは考慮されない。

これに対して、提案方式により実現される検索方式では、ベクトル成分の数値における重みにより、索引語の重要性や関連の強さを表現することが可能であるため、発生確率は低い、専門知識により重要であると判断されるような因果的関連事象群について記述された文書データを検索ランキングの上位に獲得することが可能となる。

### 3. 因果関係を計量するベクトル空間の意味的連想検索への適用と実現方式

本章では、事象間の因果関係を計量するためのベクトル空間（以下、因果関係マトリクス）を、意味的数学モデルによる意味的連想検索方式に適用するための実現方式について示す。

因果関係マトリクスとは、対象とする事象単語群について、事象単語が表す事象間の因果関係を設定した以下の3つのマトリクス  $M$ ,  $M_c$ , および  $M_e$  (図2) である。

$M$  : 原因・結果両方について関係する事象単語、および自分自身を表す事象単語に1を設定し、それ以外の事象単語に0を設定したベクトル（原因・結果ベクトル）群から構成されるマトリクス

$M_c$  : 原因のみに関係する事象単語、および自分自身を表す事象単語に1を設定し、それ以外の事象単語に0を設定したベクトル（原因ベクトル）群から構成されるマトリクス

$M_e$  : 結果のみに関係する事象単語、および自分自身を表す事象単語に1を設定し、それ以外の事象単語に0を設定したベクトル（結果ベクトル）群から構成されるマトリクス

検索目的に応じて、検索対象データおよび問合せ語のベクトルについて原因ベクトル、結果ベクトル、原因・結果ベクトルの組合せを決定し、ベクトル間の相関量計算を行うことで、「原因となる事象について記述されたデータ」、「結果となる事象について記述されたデータ」、および「原因・結果両方の事象について記述されたデータ」をそれぞれ独立に検索可能となる。因果関係マトリクスの生成方式については、4章で述べる。

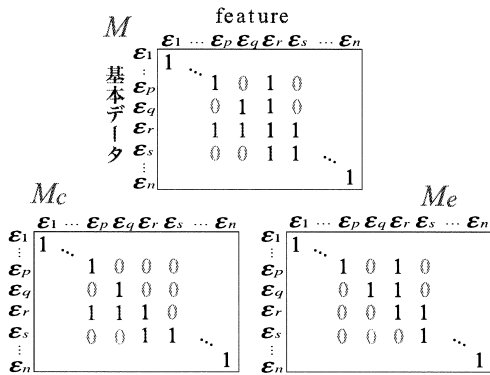


図 2 因果関係マトリクス  $M$ ,  $M_c$ ,  $M_e$ .  
Fig.2 Vector space matrices of causality  $M$ ,  $M_c$ ,  $M_e$ .

一方、一般的な情報検索手法であるパターンマッチング検索では、静的かつ明示的に与えられた記述に対する単純なパターン照合でのみ検索を行うが、実際は、データを持つ意味や、データ間の関係性は静的に決めうるものではなく、文脈や状況、あるいはユーザの視点に応じて動的に変化するものである。意味的連想検索では、分野別の専門知識を利用して、その分野の「意味」を形式的に計量することのできるベクトル空間である「メタデータ空間」を生成する。メタデータ空間における文脈解釈をともなったベクトル計算により、指定した文脈に対して、意味的に近い情報を動的に検索することを可能にしている。

因果関係マトリクスを意味的連想検索に適用することにより、検索者の検索目的に応じて、ある事象の原因検索、結果検索、および原因・結果検索といった事象間の因果関係を扱う情報検索が、文脈解釈機能による動的な情報検索を可能とする意味空間上で実現される。

以下に、意味の数学モデルによる意味的連想検索方式の概要について述べる。次に、因果関係マトリクスを、意味の数学モデルによる意味的連想検索方式へ適用する際の実現方式について具体的に述べる。

### 3.1 意味の数学モデルによる意味的連想検索方式

#### 3.1.1 概要

意味の数学モデルによる意味的連想検索方式の概要について述べる。詳細は、文献 3)~5) に示されている。

意味的連想検索方式は、メタデータ空間における文脈解釈、ベクトル計算により、利用者が指定した文脈に対して、意味的に近い情報を動的に検索することを可能とする検索方式である。ここで、メタデータ空間とは、分野別の専門知識を利用して、その分野の「意味」を形式的に計量することのできるベクトル空間である。意味的連想検索方式において、検索者は与える

文脈を複数の単語を用いて表現する。検索者が与える単語の集合をコンテキストと呼ぶ。

#### 3.1.2 メタデータ空間生成方式

以下に、メタデータ空間の生成プロセスを示す。

- (a) 対象とする分野を表現するために必要な特徴語（以下、feature）群を準備する。対象分野の専門辞書等を用いて、各見出し語を説明している説明文中の単語を抽出し、この集合を feature 群とする。これにより、その分野の意味を表現するのに必要な単語群が定義される。
- (b) 対象とする分野の基本的な用語である、基本データ群を準備する。(a)と同様に、専門辞書を用いて、見出し用語群を抽出し、この集合を基本データ群と定義する。
- (c) feature 群を用いて、各基本データの特徴付けを行う。同様の専門辞書を用いて基本データの説明文を調べ、説明文をもとに、関係のある feature には 1 を、逆の意味で用いられている feature には -1 を、関係のない feature には 0 を、それぞれ設定する。この方法で、すべての基本データに対して、feature による特徴付けを行う。
- (d) 以上の feature による基本データの特徴付けマトリクスから、意味的連想検索のためのメタデータ空間を生成する。

以上のプロセスにより、対象分野における意味の形式的な計算を可能とするメタデータ空間を生成する。

### 3.2 因果関係マトリクスの意味的連想検索方式への適用

因果関係マトリクスを、意味の数学モデルによる意味的連想検索方式へ適用する際の実現方式について述べる。提案方式は、以下のステップ (Step-I ~ Step-IV) で実現される。

#### Step-I ベクトル空間マトリクスの生成

4 章で示す生成方式により、3 つの因果関係マトリクス  $M$ ,  $M_c$ ,  $M_e$  を生成する (図 2)。これらの因果関係マトリクスは、意味の数学モデルにおけるメタデータ行列として設定するために用いられる。

#### Step-II MDS の生成

意味的検索空間 MDS 生成のためのメタデータ行列として、因果関係マトリクス  $M$  (もしくは  $M_c$ ,  $M_e$ ) を設定する。 $M$  の相関行列  $M^T M$  (もしくは  $M_c^T M_c$ ,  $M_e^T M_e$ ) を固有値分解し、固有ベクトルから構成される正規直交空間 MDS

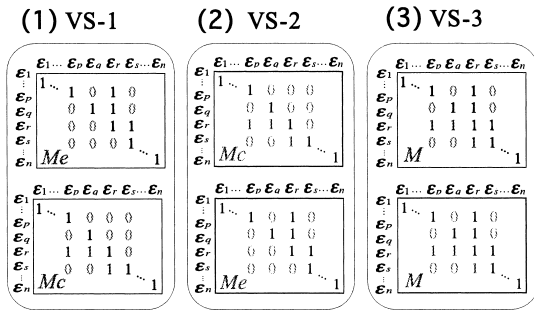


図3 ベクトル空間マトリクス組の組(上段: 検索対象データベクトル形成のためのマトリクス, 下段: 問合せ語ベクトル形成のためのマトリクス)

Fig.3 Combinations of the vector space matrices.

を生成する. 行列  $X$  の転置行列を  $X^T$ , 行列  $X$  の自己相関行列を得る関数を  $S_1$ , 自己相関行列  $X^T X$  を固有値分解し, 正規直交空間  $MDS$  を得る関数を  $S_2$  とすると,

$$\begin{aligned} S_1: X &\mapsto X^T X \\ S_2: S_1(X) &\mapsto MDS \\ (X = M, M_c, M_e) \end{aligned}$$

$MDS$  は, ある発生した事象について, その原因や結果となる事象群を得るための状況を規定する文脈を与えることにより, 文脈に応じた事象間の因果関係の計量を可能とする意味的検索空間である.

### Step-III 検索語, および検索対象データベクトルの生成

検索語列  $Q$  は, 因果関係マトリクス  $M (M_c, M_e)$  に基本データとして設定されている事象単語  $\mathcal{E}$  の列からなり, そのうち発生した事象として指定する事象単語の集合を  $\mathcal{K}$ , 発生した状況の文脈として指定する事象単語の集合を  $\mathcal{CT}$  とする. 以下, 発生した事象を示す事象単語の集合  $\mathcal{K}$  をキーワードと呼び, 発生した状況の文脈を表す事象単語の集合  $\mathcal{CT}$  を文脈語と呼ぶ. また, 検索対象データ (事象単語  $\mathcal{E}$  がメタデータとして設定されている) を  $\mathcal{P}$  とする.

検索語列  $Q$ , および検索対象データ  $\mathcal{P}$  をベクトル化するために, 4章で示す3つの因果関係マトリクス組 ( $VS-1, VS-2, VS-3$ ) (図3) を, 検索語ベクトル, および検索対象データベクトルを形成するためのメタデータ行列として設定する.

因果関係マトリクス  $M (M_c, M_e)$  に基本データとして設定されている  $n$  個の事象単語  $\mathcal{E}_i (i =$

$1 \cdots n$ ) は, 3つの因果関係マトリクス  $M_c, M_e, M$  により, それぞれのマトリクスに対応する行ベクトル  $e_c, e_e, e$  に変換される.

因果関係マトリクス  $M_c, M_e, M$  による事象ベクトル変換関数を  $\mathcal{V}_c, \mathcal{V}_e, \mathcal{V}$  とすると,

$$\begin{aligned} \mathcal{V}_c: \mathcal{E}_i &\mapsto e_{ci} \\ \mathcal{V}_e: \mathcal{E}_i &\mapsto e_{ei} \quad (i = 1 \cdots n) \\ \mathcal{V}: \mathcal{E}_i &\mapsto e_i \end{aligned}$$

$VS-1 \sim VS-3$  の因果関係マトリクス組から, 検索者の検索目的に応じて1つの因果関係マトリクス組が選択される.

#### (1) ある事象の原因である事象群を検索する目的の場合

因果関係マトリクス組  $VS-1 (M_c, M_e)$  を, 検索語ベクトル形成のためのメタデータ行列, および検索対象データベクトル形成のためのメタデータ行列として設定する. それぞれのメタデータ行列を用いて, 検索語 (キーワードおよび文脈語) ベクトル  $q$ , 検索対象データベクトル  $p$  は和 (OR) 演算により, 以下のように合成される.

$$\begin{aligned} q &= \sum_{\mathcal{E}_i \in \mathcal{K}} \mathcal{V}_c(\mathcal{E}_i) + \sum_{\mathcal{E}_j \in \mathcal{CT}} \mathcal{V}_e(\mathcal{E}_j) \\ p &= \sum_{\mathcal{E}_k \in \mathcal{P}} \mathcal{V}_e(\mathcal{E}_k) \end{aligned}$$

$VS-1$  では, 検索語ベクトルとして事象の「原因ベクトル」, 検索対象データベクトルとして事象の「結果ベクトル」を用いて設定するため, 問合せとして与えられた事象単語に対して「原因となる事象について記述されたデータ」の検索を実現する. このベクトルセットを用いた場合の検索を, 以下「原因検索」と呼ぶ.

#### (2) ある事象の結果である事象群を検索する目的の場合

(1) と同様に, 因果関係マトリクス組  $VS-2 (M_e, M_c)$  を, 検索語ベクトル, および検索対象データベクトルを形成するためのメタデータ行列として設定する. 検索語ベクトル  $q$ , 検索対象データベクトル  $p$  は, 以下のように合成される.

$$\begin{aligned} q &= \sum_{\mathcal{E}_i \in \mathcal{K}} \mathcal{V}_e(\mathcal{E}_i) + \sum_{\mathcal{E}_j \in \mathcal{CT}} \mathcal{V}_c(\mathcal{E}_j) \\ p &= \sum_{\mathcal{E}_k \in \mathcal{P}} \mathcal{V}_c(\mathcal{E}_k) \end{aligned}$$

$VS-2$  では, 検索語ベクトルとして事象の「結果ベクトル」, 検索対象データベク

表 1 各検索で設定されているメタデータ行列  
Table 1 Metadata matrices for each search.

	設定するメタデータ空間		
	MDS の生成	検索語(キーワードおよび 文脈語)ベクトルの形成	検索対象データ ベクトルの形成
原因検索	$M$ (もしくは $M_c, M_e$ )	$VS-1 (M_c, M_e)$	
結果検索	$M$ (もしくは $M_c, M_e$ )	$VS-2 (M_e, M_c)$	
原因・結果検索	$M$ (もしくは $M_c, M_e$ )	$VS-3 (M, M)$	

トルとして事象の「原因ベクトル」を用いて設定するため、問合せとして与えられた事象単語に対して「結果となる事象について記述されたデータ」の検索を実現する。このベクトルセットを用いた場合の検索を、以下「結果検索」と呼ぶ。

- (3) ある事象の原因である事象群, および結果である事象群の両方について検索する目的の場合

(1) と同様に, 因果関係マトリクスの組  $VS-3 (M, M)$  を, 検索語ベクトル, および検索対象データベクトルを形成するためのメタデータ行列として設定する。検索語ベクトル  $q$ , 検索対象データベクトル  $p$  は, 以下のように合成される。

$$q = \sum_{\mathcal{E}_i \in \mathcal{K}} \mathcal{V}(\mathcal{E}_i) + \sum_{\mathcal{E}_j \in \mathcal{CT}} \mathcal{V}(\mathcal{E}_j)$$

$$p = \sum_{\mathcal{E}_k \in \mathcal{P}} \mathcal{V}(\mathcal{E}_k)$$

$VS-3$  では, 検索語ベクトル, 検索対象データベクトルとして, 事象の「原因・原因ベクトル」を用いて設定するため, 問合せとして与えられた事象単語に対して「原因・結果両方の事象について記述されたデータ」の検索を実現する。このベクトルセットを用いた場合の検索を, 以下「原因・結果検索」と呼ぶ。

**Step-IV** 検索語, および検索対象データベクトルの  $MDS$  への写像

Step-III により, 得られた検索語ベクトル  $q$ , および検索対象データベクトル  $p$  を,  $MDS$  上へ写像する。ベクトル  $x$  が  $MDS$  へのベクトル写像関数  $\mathcal{D}$  により,  $x'$  に写像されるとすると,

$$\mathcal{D}: x \mapsto x'$$

検索語ベクトル  $q$ , および検索対象データベクトル  $p$  は  $\mathcal{D}$  により, それぞれ  $MDS$  上で  $q', p'$  に写像される。

$$q' = \mathcal{D}(q)$$

$$p' = \mathcal{D}(p)$$

ベクトル  $q'$  を無限大ノルムで正規化したベクトルを意味重心  $G_+$  と呼び, 意味的検索空間  $MDS$  における部分空間選択に用いられる。 $G_+$  により選択された部分空間上に  $p'$  を写像し, そのノルムを計量することにより, 検索語列(キーワードおよび文脈語)と検索対象データ  $\mathcal{P}$  の相関量計算が行われる。

以上のステップにより, 検索者の検索目的に応じた事象間の因果関係を扱う情報検索が, 文脈語  $\mathcal{CT}$  によって与えられる文脈に応じて, 文脈解釈機能をともなった動的な情報検索を可能とする意味的検索空間上で実現される。各検索で設定されるメタデータ行列を表 1 に示す。

#### 4. 事象間の因果関係を扱うベクトル空間生成方式

事象間の因果関係を扱うベクトル空間生成方式の概要について述べる。本方式では, 検索対象領域における事象間の因果関係が計量可能なベクトル空間を生成する。詳細は, 文献 15) に示されている。

##### 4.1 因果関係マトリクスの作成

因果関係マトリクスとして, 3 つの特徴行列  $M_c, M_e, M$  を生成する(図 2)。これらのマトリクスの生成ステップを, 以下に示す。

**Step-I** *feature* (特徴語) 群, 基本データ群の設定  
3 つのマトリクス  $M, M_c, M_e$  に対し,  $n$  個の事象語群  $\mathcal{E}_1 \sim \mathcal{E}_n$  を *feature*, 基本データとして, それぞれマトリクスの横軸, 縦軸に設定し,  $n \times n$  の正方行列を生成する。

##### Step-II 特徴付けの設定

Step-I で *feature* 群, および基本データ群を設定した 3 つのマトリクス  $M, M_c, M_e$  について, それぞれ特徴付けの設定を行う。それぞれのマトリクスに対する特徴付けの設定方法を以下に示す。

$M$ : 基本データ中の事象単語について, *feature* の中から, 自分自身を表す事象単語, 原因である事象単語, および結果である事象単語に 1 を設定し, それ以外の事象単語に 0 を設定

する．この操作を基本データ中の全事象単語に適用し， $M$  を生成する．

$M_c$ ：基本データ中の事象単語について，*feature*の中から，自分自身を表す事象単語，および原因である事象単語に 1 を設定し，それ以外の事象単語に 0 を設定する．この操作を基本データ中の全事象単語に適用し， $M_c$  を生成する．

$M_e$ ：基本データ中の事象単語について，*feature*の中から，自分自身を表す事象単語，および結果である事象単語に 1 を設定し，それ以外の事象単語に 0 を設定する．この操作を本データ中の全事象単語に適用し， $M_e$  を生成する．

## 4.2 因果関係マトリクスの組の設定

4.1 節で生成した 3 つの因果関係マトリクス  $M$ ， $M_c$ ， $M_e$  を用いて，検索目的に応じた因果関係マトリクスの組  $VS-1 \sim VS-3$  (図 3) の設定方法について述べる．

(a) ある事象の原因である事象群を検索する目的の場合

検索対象データベクトル形成のための因果関係マトリクスとして  $M_e$  を，問合せ語ベクトル形成のための因果関係マトリクスとして  $M_c$  を設定する．この因果関係マトリクスの組を  $VS-1$  とする．

(b) ある事象の結果である事象群を検索する目的の場合

検索対象データベクトル形成のための因果関係マトリクスとして  $M_c$  を，問合せ語ベクトル形成のための因果関係マトリクスとして  $M_e$  を設定する．この因果関係マトリクスの組を  $VS-2$  とする．

(c) ある事象の原因である事象群，および結果である事象群の両方について検索する目的の場合

検索対象データベクトル形成のための因果関係マトリクス，問合せ語ベクトル形成のための因果関係マトリクスともに  $M$  を設定する．この因果関係マトリクスの組を  $VS-3$  とする．

以上設定した因果関係マトリクスの組 ( $VS-1 \sim VS-3$ ) を，内積計算やコサイン距離計算等のベクトル演算やベクトル空間モデルによる検索方式に適用することで，検索者の検索目的に応じた因果関係による情報検索が実現可能となる．

## 4.3 検索語事象ベクトルの拡張

ここでは，3 つ以上離れた因果関係にまで及び検索

を可能とするために，検索者が与える検索語事象ベクトルを拡張する方式を示す．事象ベクトルを拡張するアルゴリズムを以下に示す．

- (1) 与えられた事象ベクトルの要素のうち“1”が付与されている事象を列挙する．
- (2) 列挙したそれぞれの事象について，事象ベクトルの“1”が付与されている要素に拡張関数  $f_{expand}$  を適用する．
- (3) (2)における各ベクトルを元の検索語事象ベクトルに足し合わせる．

(1)~(3)の操作を再帰的に適用し，再帰のフェーズが拡張定数を超えるまで続ける．拡張定数は，検索語事象ベクトルの拡張をどこまで行うかを決定する定数である． $n$  個 ( $n \geq 2$ ) 離れた事象の因果関連性を計量する場合は，拡張定数を  $n-2$  に設定する．すべての因果関連性を有する事象を検索する場合は，拡張定数を設定する必要はない．

拡張関数  $f_{expand}$  としては，アプリケーションに応じて様々な関数を設定することが可能であるが，一般的には次のような関数を適用する．

$$f_{expand}(x, phase) = x * k^{phase}$$

$x$  は対象としているベクトルの要素の値， $k$  は任意の定数， $phase$  は再帰のフェーズを表す．より離れた事象ほど因果関連性を低くするには， $k < 1$  とし，逆に，より離れた事象ほど因果関連性を高くするには， $k > 1$  と設定すればよい．また， $phase$  は，初期値が“0”で，再帰するごとにインクリメントする．

## 5. 実 験

### 5.1 実 験 A

#### 5.1.1 実 験 目 的

実験 A は，文献 15) に対して，本方式の特徴である文脈解釈機能による動的な因果関連性計量についての実験 C および D を行う前提として，問合せとして与える事象データと，提案方式により検索結果の上位に獲得可能な事象データ間の因果関係に関する性質を確認する．本実験では，ある事象  $\mathcal{E}$  の原因検索や結果検索を行う場合に，事象  $\mathcal{E}$  と直接的な因果関係にある事象群だけでなく，直接的な因果関係にある事象群に対して，これらの事象と共通の原因や結果を持つといった間接的な因果関係にある事象群との因果関連性を計量し，検索可能なことを検証する．

#### 5.1.2 実 験 環 境

3 章および 4 章に示した方式により，実験システムを構築した．実験 A では，事象 00~37 の 38 個の擬

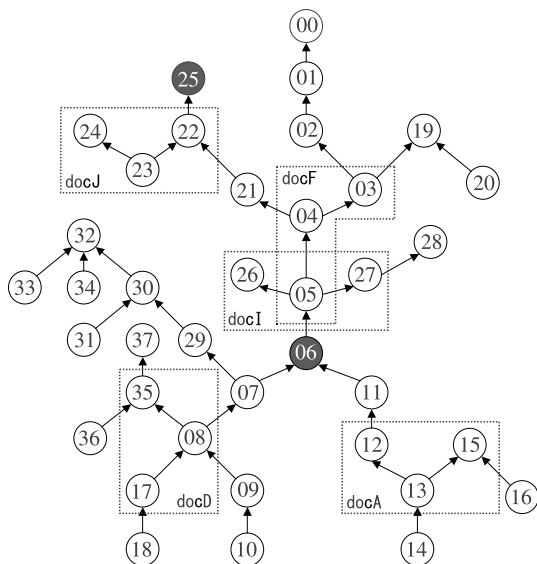


図 4 擬似事象データセット (実験 A)

Fig. 4 A synthetic event data set (exp. A).

表 2 意味的検索空間 MDS の構成 (実験 A)

Table 2 A description of MDS (exp. A).

	単語, 次元数
feature	38 語
基本データ	38 語
空間 MDS の次元	36 次元

表 3 文書データのメタデータ (実験 A)

Table 3 Metadata of documents (exp. A).

文書 ID	事象メタデータ	
docA	12 13 15	事象 06 の原因についての擬似文書データ
docB	07 08 09	
docC	11 12 13	
docD	08 17 35	
docE	08 17 18	
docF	03 04 05	事象 25 の原因についての擬似文書データ
docG	04 05 06	
docH	04 21 22	
docI	05 26 27	
docJ	22 23 24	

似事象データ (図 4) を用意し, 意味の数学モデルへ適用するための因果関係マトリクス  $M$ ,  $M_c$ ,  $M_e$  を生成した. 因果関係マトリクス  $M$  を用いて生成した意味的検索空間の構成を表 2 に示す. また, 検索対象データとして, 事象を示すメタデータのみが設定された擬似的な文書データを 10 件設定した. 各文書データに設定したメタデータの内容を表 3 に示す. 本実験では, 4.3 節に示した方式において,  $k = 0.5$  と設定し, 検索語事象ベクトルを拡張した.

表 4 実験 A の検索結果 (1)

Table 4 A search result (1) of exp. A.

contextA-1: 事象 06

順位	提案方式		コサイン距離	
	文書 ID	相関値	文書 ID	相関値
1	docC	0.744677	docC	0.663107
2	docA†	0.708786	docB	0.634876
3	docB	0.471318	docE	0.409566
4	docD‡	0.361522	docA†	0.368191
5	docE	0.356743	docD‡	0.351056
6	docI	0.141862	docG	0.143522
7	docG	0.099407	docF	0.000000
8	docJ	0.096294	docH	0.000000
9	docH	0.075818	docI	0.000000
10	docF	0.015604	docJ	0.000000

### 5.1.3 実験方法

事象 06 と事象 25 をそれぞれ検索語として, 原因検索を行い, 提案方式による検索とコサイン距離による検索を行った場合との検索結果の比較を行う. 事象 06 については, 文書データ docA ~ docE が正解セットである. 事象 06 を検索語として原因検索を行った場合の, 間接的な因果関係があると考えられる事象 15 および事象 35 が, それぞれ文書データ docA および docD のメタデータの 1 つとして設定されている. 事象 25 については, 文書データ docF ~ docJ が正解セットである. 事象 25 を検索語として原因検索を行った場合の, 間接的な因果関係があると考えられる事象 03, 事象 26 と 27, および事象 24 が, それぞれ文書データ docF, docI および docJ のメタデータの 1 つとして設定されている. 事象 06 および事象 25 を検索語として, 原因検索を行った場合について, 提案方式による検索では, コサイン距離による検索よりも, それぞれの間接的な因果関係にある事象について述べている文書データ docA や docD, および docF, docI, および docJ を検索結果の上位に獲得可能なことを確認する.

### 5.1.4 実験結果

実験結果を表 4, 表 5 に示す.

### 5.1.5 実験考察

表 4 の実験結果は, 事象 06 を検索語として, 原因検索を行った場合の検索結果を示している. この実験結果より, 事象 06 と間接的な因果関係があると判断される事象 15 および事象 35 を, それぞれメタデータに含んでいる文書データ docA および docD が, 提案方式による検索では, それぞれ 2 位および 4 位とコサイン距離による検索 (4 位および 5 位) よりも上位に検索されていることが確認できる.

また, 表 5 の実験結果は, 事象 25 を検索語として,



表 5 実験 A の検索結果 (2)

Table 5 A search result (2) of exp. A.

contextA-2: 事象 25

順位	提案方式		コサイン距離	
	文書 ID	相関値	文書 ID	相関値
1	docJ†	0.819879	docH	0.797053
2	docH	0.536157	docJ†	0.688434
3	docF‡	0.438741	docG	0.226423
4	docG	0.386171	docF‡	0.187017
5	docI§	0.266089	docI§	0.070328
6	docA	0.138155	docC	0.030768
7	docC	0.096285	docB	0.029458
8	docD	0.087498	docE	0.015384
9	docE	0.083948	docA	0.013830
10	docB	0.080706	docD	0.013186

原因検索を行った場合の検索結果を示している。この実験結果より、事象 25 と間接的な因果関係があると判断される事象 24 および事象 03 を、それぞれメタデータとして含んでいる文書データ *docJ*, *docF* がそれぞれ、1 位、3 位とコサイン距離による検索 (2 位および 4 位) よりも上位に検索されていることが確認できる。事象 25 と間接的な因果関係があると判断される事象 26 および事象 27 をともにメタデータとして含んでいる文書データ *docI* については、提案方式による検索とコサイン距離による検索では、検索順位はどちらも 5 位と変化はないが、コサイン距離による検索では、*docI* の相関値は 0.070328 であり、1 位の *docH* の相関値 0.797053 と比較して関連性が少ないと計量されていることが確認できる。

これらの実験結果は、間接的な因果関係にある事象を記述した文書データを対象とした場合に、提案方式による検索では、コサイン距離による検索よりも順位や相関値が上がる傾向にあることを示している。これは、提案方式による検索を行った場合では、間接的な因果関係にある事象についても計量を行っているのに対し、コサイン距離を適用して検索を行った場合には、間接的な因果関係にある事象の計量を行っていないためである。

以上の考察により、ある事象についての因果関係検索を行う場合に、事象と直接的な因果関係にある事象群だけでなく、直接的な因果関係にある事象群と共通の原因や結果を持つといった間接的な因果関係にある事象群との因果関連性を計量し、検索可能なことが確認できた。

## 5.2 実験 B

### 5.2.1 実験目的

実験 B では、文献 15) に示されている、原因系の事象と結果系の事象についての独立な計量機能を、提

表 6 意味的検索空間 *MDS* の構成 (実験 B, C)Table 6 A description of *MDS* (exp. B and C).

	単語、次元数
feature	365 語
基本データ	365 語
空間 <i>MDS</i> の次元	354 次元

案方式においても有していることを確認する。本実験では、提案方式により生成した意味的検索空間において実現した目的別検索、原因検索 (*VS-1*)、結果検索 (*VS-2*) および原因・結果検索 (*VS-3*) について、それぞれが検索目的に応じた文書検索が可能であることを検証する。

### 5.2.2 実験環境

実験 B では、意味の数学モデルへ適用するためのメタデータ行列作成のための単語群として、宇宙開発事業団 (現、宇宙航空研究開発機構) で提供されるドキュメントの不具合原因および不具合現象についての分類表の単語群 365 語を用いた。これらの単語群を基本データ、および *feature* としてそれぞれ設定し、3 つのマトリクス *M*, *M<sub>c</sub>*, *M<sub>e</sub>* を生成した。それぞれのマトリクスにおいて、各基本データの *feature* による特徴付けは、宇宙工学の専門家が行った。*M* を用いて生成した意味的検索空間の構成を表 6 に示す。検索対象データとして、メタデータのみが設定された擬似的な文書データを 30 件設定した。各文書データに設定したメタデータの内容を表 13 (後出) に示す。なお、本実験では検索語事象ベクトルの拡張は行っていない。

### 5.2.3 実験方法

下記に示す、事象単語を表す *contextB-1*, *contextB-2* を、5.2.2 項で示した意味的連想検索システムへコンテキスト (検索語) として入力し、それぞれ原因検索、結果検索および原因・結果検索を行う。表 13 中の文書データの中から、*contextB-1*, *contextB-2* について、それぞれ原因について述べた文書データ 2 件、結果について述べた文書データ 2 件をあらかじめ設定しておく。これらの文書データについて、指定したコンテキストに対する原因検索、結果検索および原因・結果検索の検索結果を比較する。

(*contextB-1*) 色むら・光沢むら・色違い・しみ・汚れ・変色・斑点

(*contextB-2*) 異常雑音

### 5.2.4 実験結果

5.2.3 実験方法で示した *contextB-1*, *contextB-2* をコンテキスト (検索語) として入力し、原因検索、結果検索、原因・結果検索を行った実験結果を表 7, 表 8

表 7 実験 B の検索結果 (1)

Table 7 A search result (1) of exp. B.

contextB-1: 色むら・光沢むら・色違い・しみ・汚れ・変色・斑点

文書 ID	原因検索		結果検索		原因・結果検索	
	順位	相関値	順位	相関値	順位	相関値
doc09	1	0.559260	25	0.181398	4	0.586840
doc08	5	0.422268	20	0.243327	8	0.527639
doc12	24	0.153958	1	0.600280	5	0.565532
doc11	14	0.224462	2	0.572128	3	0.664330

表 8 実験 B の検索結果 (2)

Table 8 A search result (2) of exp. B.

contextB-2: 異常雑音

文書 ID	原因検索		結果検索		原因・結果検索	
	順位	相関値	順位	相関値	順位	相関値
doc26	1	0.484424	20	0.178828	8	0.530468
doc25	3	0.421999	19	0.200069	13	0.455055
doc30	27	0.230328	1	0.514344	7	0.536498
doc29	12	0.293432	3	0.493342	4	0.574856

に示す。表中の上段に示す 2 件の文書データは、入力したコンテキストについて、それぞれ原因について述べている文書データである。また、表中の下段に示す 2 件の文書データは、入力したコンテキストについて、それぞれ結果について述べている文書データである。

### 5.2.5 実験考察

表 7 に示す実験結果は、contextB-1「色むら・光沢むら・色違い・しみ・汚れ・変色・斑点」に対する検索結果を示している。この検索結果において、「原因検索」では、「色むら・光沢むら・色違い・しみ・汚れ・変色・斑点」の原因について述べた文書データ doc09 (1 位), doc08 (5 位) を上位に検索し、結果について述べた文書データ doc11 (14 位), doc12 (24 位) を下位に検索している。また、「結果検索」では、結果について述べた文書データ doc12 (1 位), doc11 (2 位) を上位に検索し、原因について述べた文書データ doc08 (25 位), doc09 (20 位) を下位に検索している。さらに、「原因・結果検索」では、原因について述べた文書データ doc09 (4 位), doc08 (8 位), および結果について述べた文書データ doc11 (3 位), doc12 (5 位) それぞれを上位に検索している。これらの検索結果は、提案方式による原因検索、結果検索、および原因・結果検索が、それぞれの検索目的に応じた検索が実現されていることを示している。

また、表 8 に示す実験結果は、contextB-2「異常雑音」に対する検索結果を示している。この実験結果についても、contextB-1 の示す実験結果についてと同様の考察ができる。

以上の実験結果は、提案方式により生成した意味的検索空間が、原因検索、結果検索および原因・結果検索について、それぞれの検索目的に応じた検索が可能であることを示している。

## 5.3 実験 C

### 5.3.1 実験目的

実験 C は、文献 15) に対して、本方式の特徴である文脈解釈機能をともなった因果関連性計量を示すものである。本実験では、提案方式により生成した意味的検索空間が、複数の異なる側面の原因、結果を持つ事象単語に加えて、同時あるいは一連に起こっていると考えられる事象単語を文脈として追加したコンテキスト (検索語) を与えることで、文脈に応じた動的な事象間の因果関係の計量による情報検索が可能であることを検証する。

### 5.3.2 実験環境

実験 B と同じ実験環境を用いた。

### 5.3.3 実験方法

ある事象単語に加えて、同時あるいは一連に起こっていると考えられる 2 種類の事象単語を文脈として与えたコンテキスト (検索語) を、下記に示す contextC-1a, contextC-1b, および contextC-2a, contextC-2b のように設定する。これらのコンテキストを、5.2.2 項で示した意味的連想検索システムへ入力して検索を行う。contextC-1a, contextC-1b についての検索は、「原因検索」を行う。contextC-2a, contextC-2b についての検索は、「結果検索」を行う。表 13 中の文書データの中から、contextC-1a, contextC-1b につい

表 9 実験 C の検索結果 (1)  
Table 9 A search result (1) of exp. C.

原因検索 (VS-1)		
contextC-1a		
消費電力異常 回転数規格外		
文書 ID	順位	相関値
doc01	1	0.408987
doc02	4	0.325340
contextC-1b		
消費電力異常 炭化・熱分解		
文書 ID	順位	相関値
doc02	1	0.439171
doc01	2	0.368898

表 10 実験 C の検索結果 (2)  
Table 10 A search result (2) of exp. C.

結果検索 (VS-2)		
contextC-2a		
亀裂 (割れ/クラック) 回転ムラ/トレンド変化		
文書 ID	順位	相関値
doc17	1	0.376697
doc18	8	0.277768
contextC-2b		
亀裂 (割れ/クラック) 視野視線異常		
文書 ID	順位	相関値
doc18	1	0.475623
doc17	5	0.390772

ては、原因について述べた文書データをそれぞれ 1 件ずつ設定し (doc01, doc02), contextC-2a, contextC-2b については、結果について述べた文書データをそれぞれ 1 件ずつ設定しておく (doc17, doc18)。これらの文書データについて、contextC-1a と contextC-1b の検索結果、および contextC-2a と contextC-2b の検索結果を比較する。

(contextC-1a) 消費電力異常 回転数規格外

(contextC-1b) 消費電力異常 炭化・熱分解

(contextC-2a) 亀裂 (割れ/クラック) 回転ムラ/  
トレンド変化

(contextC-2b) 亀裂 (割れ/クラック) 視野視線異常

### 5.3.4 実験結果

実験結果を表 9, 表 10 に示す。

### 5.3.5 実験考察

表 9 に示す結果は、contextC-1a 「消費電力異常 回転数規格外」、および contextC-1b 「消費電力異常 炭化・熱分解」をコンテキストとして、「原因検索」したときの検索結果を示している。doc01, doc02 はいずれも「消費電力異常」の原因について述べた文書データである。doc01 は、表 13 に示される事象単語「ガ

タ・緩み 異常振動 異常発振」がメタデータとして設定されており、「消費電力異常」の振動異常に関する原因について述べた文書データである。同様に、doc02 は、事象単語「制御不十分 温度制御不良 温度規格外」がメタデータとして設定されており、「消費電力異常」の温度異常に関する原因について述べた文書データである。表 9 に示す実験結果では、contextC-1a, および contextC-1b をコンテキストとして「原因検索」を行った場合に、doc01, および doc02 がそれぞれ 1 位に検索されている。

この検索結果より、回転に関する原因および温度に関する原因等の複数の異なる側面の原因を持つ事象単語「消費電力異常」に加えて、同時あるいは一連に起こっていると考えられる事象単語「回転数規格外」または「炭化・熱分解」を文脈としてコンテキストに追加することで、それぞれの文脈に応じた動的な事象間の因果関係の計量による検索が実現されていることが確認できる。

また、表 10 に示す結果は、contextC-2a 「亀裂 (割れ/クラック) 回転ムラ/トレンド変化」、および contextC-2b 「亀裂 (割れ/クラック) 視野視線異常」をコンテキストとして、「結果検索」をしたときの検索結果を示している。doc17, doc18 はいずれも「亀裂 (割れ/クラック)」の結果について述べた文書データである。doc17 は、表 13 に示される事象単語「異常音 異常振動 画素間感度偏差異常」がメタデータとして設定されており、「亀裂 (割れ/クラック)」の振動異常に関する結果について述べた文書データである。同様に、doc18 は、事象単語「レベル特性異常 光源 (測定系, 校正系) のノイズ 画素欠損」がメタデータとして設定されており、「亀裂 (割れ/クラック)」の光学異常に関する結果について述べた文書データである。

表 10 に示す実験結果では、contextC-1a, および contextC-1b をコンテキストとして「結果検索」を行った場合に、doc17, および doc18 がそれぞれ 1 位に検索されている。

この検索結果より、振動異常に関する結果および光学異常に関する結果等の複数の異なる側面の結果を持つ事象単語「亀裂 (割れ/クラック)」に加えて、同時あるいは一連に起こっていると考えられる事象単語「回転ムラ/トレンド変化」、「視野視線異常」を文脈としてコンテキストに追加することで、それぞれの文脈に応じた動的な事象間の因果関係の計量による検索が実現されていることが確認できる。

以上の実験結果は、提案方式により生成した意味的検索空間が、複数の異なる側面の原因、結果を持つ事

象単語に加えて、同時あるいは一連に起こっていると考えられる事象単語を文脈としてコンテキストに追加することで、文脈に応じた動的な事象間の因果関係の計量による情報検索が可能であること示している。

## 5.4 実験 D

### 5.4.1 実験目的

実験 D は、実験 C に引き続き、文脈解釈機能による因果関連性計量についての実験を行う。本実験では、提案方式の特徴である、ある事象単語に加えて状況を示す事象単語を文脈としてコンテキストに追加した場合の検索では、コサイン距離を用いた場合の検索では容易に得られないデータについての情報獲得が可能であることを検証する。

### 5.4.2 実験環境

実験 B と同じ実験環境を用いた。ただし、検索対象データには、実験 B でのマトリクス  $M$ ,  $M_c$ ,  $M_e$  の生成に用いた事象単語群 365 語を用いた。これらの事象単語群は、表 13 に示すように文書データのメタデータとして設定可能である。

### 5.4.3 実験方法

ある事象単語に加えて、事象単語を文脈として与えたコンテキスト（検索語）を、下記に示す contextD-1a, contextD-1b, および contextD-2a, contextD-2b のように設定する。contextD-1a, contextD-1b についての検索は、「原因検索」を行う。contextD-2a, contextD-2b についての検索は、「結果検索」を行う。  
 (contextD-1a) 絶縁不良 導通不良 ハンダ付不良  
 (contextD-1b) 絶縁不良 温度制御不良  
 (contextD-2a) 導通不良 接触不良 温度不良  
 (contextD-2b) 導通不良 誤配線 使用材料・原料 欠陥

また、これらのコンテキストを用いて検索を行った場合の正解として、宇宙工学の専門家が、それぞれのコンテキストのキーワードを表す事象単語の原因、または結果である事象単語を、以下に示すように設定した。

- contextD-1a, D-1b のキーワード「絶縁不良」の原因を示す事象単語：「組立て上の不良」、「温度異常」
- contextD-2a, D-2b のキーワード「導通不良」の結果を示す事象単語：「コマンド受付不良」、「不着火」

contextD-1a と contextD-1b, および contextD-2a と contextD-2b をコンテキストとして指定した場合において、提案方式による検索とコサイン距離による検索を行った場合とで検索結果の比較を行う。

### 5.4.4 実験結果

実験結果を表 11, 表 12 示す。

### 5.4.5 実験考察

表 11 は、コンテキストとして contextD-1a および contextD-1b を指定して、原因検索を行った結果を示している。提案方式による検索では、「絶縁不良」に加えて、状況を示す事象単語「導通不良」、「ハンダ付不良」を文脈としてコンテキストに追加することにより、「絶縁不良」の原因である事象「組立て上の不良」を検索結果の上位（3 位）に獲得している。また、同様に「温度制御不良」を文脈としてコンテキストに追加することにより、「絶縁不良」の原因である事象「温度異常」を検索結果の上位（7 位）に獲得している。これに対して、コサイン距離による検索では、「組立て上の不良」や「温度異常」は検索結果の上位に獲得されていないことが確認できる。

また、表 12 は、コンテキストとして contextD-2a および contextD-2b を指定して、結果検索を行った結果を示している。この場合も同様に、提案方式による検索では、「導通不良」に加えて、状況を示す事象単語「接触不良」、「温度不良」、または「誤配線」、「使用材料・原料欠陥」を文脈として指定することで、「導通不良」の結果である事象「コマンド受付不良」、「不着火」を、それぞれ指定した文脈に応じて検索結果の上位（4 位, 9 位）に獲得している。これに対して、コサイン距離による検索では、「コマンド受付不良」や「不着火」は検索結果の上位に獲得されていないことが確認できる。

これらは、提案方式による検索では、コサイン距離を用いた場合の検索のような直接的な因果関係の連鎖についての絞り込みによる検索を行っているのではなく、コンテキストに状況を示す事象単語を文脈として追加することにより、その文脈に応じた動的な因果関係計量による検索を行っているためである。

また、本実験では、365 語の事象単語群を検索対象データとして実験を行ったが、これらの単語群を、表 13 に示すように文書データのメタデータとして付与することにより、実験 C と同様に文書データを検索対象データとして検索を行うことが可能である。

以上の実験結果は、提案方式による検索では、問合せとして、ある事象単語に加えてその状況を示す事象単語を文脈として指定することにより、コサイン距離を用いた場合の検索では容易に得られないデータについての情報獲得が可能であることを示している。

表 11 実験 D の検索結果 (1)  
Table 11 A search result (1) of exp. D.

原因検索 (VS-1)

	提案方式			
	contextD-1a		contextD-1b	
	順位	相関値	順位	相関値
事象単語				
組立て上の不良	3	0.291437	26	0.208319
温度異常	19	0.233291	7	0.241409

	コサイン距離			
	contextD-1a		contextD-1b	
	順位	相関値	順位	相関値
事象単語				
組立て上の不良	38	0.143499	58	0.000000
温度異常	23	0.186410	23	0.151185

表 12 実験 D の検索結果 (2)  
Table 12 A search result (2) of exp. D.

結果検索 (VS-2)

	提案方式			
	contextD-2a		contextD-2b	
	順位	相関値	順位	相関値
事象単語				
コマンド受付不良	4	0.272604	45	0.170714
不着火	89	0.233291	9	0.372155

	コサイン距離			
	contextD-2a		contextD-2b	
	順位	相関値	順位	相関値
事象単語				
コマンド受付不良	63	0.070754	31	0.081992
不着火	47	0.103142	23	0.119522

## 6. 実験全体の考察

ここでは、文書メタデータの作成、および問合せ処理の計算量について考察する。

### 6.1 文書メタデータの作成

提案方式では、検索対象となる文書データをベクトル化するために、これらの文書データに事象単語群がメタデータとして付与されていることが前提となる。今回の実験では、擬似的な文書データとして、事象単語のメタデータを数個設定して作成した。1つの擬似文書データに設定した事象単語は、それぞれが因果関連性があるものを選択した。これは、提案方式の応用例の1つであるシステムの不具合情報等の事例データベースにおいて、1つの文書データ中に、ある発生事象のみでなく、それについての原因や結果である事象群についての記述を行うことにより、利用者間の共有知識資源としての文書データの有効性が高まると判断したためである。

実際の文書データにメタデータの付与を行うには、

文書データからパターンマッチング等により自動で事象単語を抜き出して抽出する方法、および文書の作成者や専門家が文書データの内容に照らし合わせて、事象単語を付与する方法がある。パターンマッチング等を用いた自動抽出方法では、必ずしも文書データの内容に忠実なメタデータを抜き出せるとは限らないため、最終的には、人間がメタデータの妥当性を判断する必要がある。このため、文書作成者あるいは専門家が、文書データの内容を示す事象単語群をメタデータとして付与していくことが重要であると考えられる。

### 6.2 問合せ処理の計算量

問合せを行う前提として、意味的検索空間（以下、本節において意味空間とする）の生成を行う。意味空間の生成は、行列の固有値計算を行うため、ベクトル次元数に応じた大きな計算量を必要とするが、因果関係マトリクスを定義した後の意味空間の生成は初期的に1回だけ行えばよい。その意味空間上への任意の検索対象データをマッピングすることにより、意味的連想検索を適用することができる。このとき、意味空間

表 13 文書データのメタデータ (実験 B, C)  
Table 13 Metadata of documents (exp. B and C).

文書 ID	事象メタデータ
doc01	ガタ・緩み 異常振動 異常発振
doc02	制御不十分 温度制御不良 温度規格外
doc03	異常雑音 回路設計不十分 ハンダ付不良
doc04	リニアリティ異常 過渡応答異常 光源 (測定系, 校正系) のノイズ
doc05	発光スペクトル異常 帯域外特性異常 異常メッセージ出力
doc06	画像上のノイズ 画素欠陥 (白キズ等) 画素欠損
doc07	腐食・酸化・硫化・窒化・侵食・酸化膜の形成 炭化・熱分解 作業環境不良 (温湿度, 清浄度, 汚染ガス)
doc08	治工具不良 図面指示誤り, 漏れまたは不明確 温度規格外
doc09	技量不足 不注意・ボカミス 発熱・オーバヒート
doc10	物性 (屈折率, 分散, 泡, 脈理, 吸収率・輻射率) 異常 画像上のノイズ 透過率 (含反射率) 異常
doc11	コンタミネーション 素子感度 (含ブルーミング) 異常 コーティング劣化
doc12	材料特性異常 耐環境 (放射線, 紫外線) 特性異常 周辺減光異常
doc13	強度 機構設計不十分 材料適用ミス
doc14	強度 剥がれ・剥離・浮き上り・ふくれ 孔内/外径不良
doc15	剥がれ・剥離・浮き上り・ふくれ 寸法不良 異部品装着
doc16	画素欠損 画素欠陥 (白キズ等) 画質異常
doc17	異常音 異常振動 画素間感度偏差異常
doc18	レベル特性異常 光源 (測定系, 校正系) のノイズ 画素欠損
doc19	図面指示誤り, 漏れまたは不明確 手順書指示誤り, 漏れまたは不明確 作業指示不遵
doc20	図面等の誤読 判断の誤り・勘違い 技量不足
doc21	腐食・酸化・硫化・窒化・侵食・酸化膜の形成 機械・設備不良 作業環境不良 (温湿度, 清浄度, 汚染ガス)
doc22	利得低下・利得増大 出力規格外れ 動作不安定
doc23	出力変動 波形不良・位相不良 画像上のノイズ
doc24	早期起動, 早期停止 チャタリング システムダウン
doc25	回路設計不十分 実装不良 ハンダ付不良
doc26	接触不良 実装設計不十分 ガタ・緩み
doc27	実装不良 異常振動 図面等の誤読
doc28	出力異常 動作不安定 画像上のノイズ
doc29	動作せず, 始動不能, 停止不能 電圧の規格外れ・過渡電圧 画素間暗電流偏差異常
doc30	画質異常 利得低下・利得増大 光源 (測定系, 校正系) のノイズ

の生成は必要なく, 検索対象データの意味的検索空間へのマッピングは, 意味空間の生成時間に比べて大変短い。

問合せ処理は, 大きく部分空間選択処理と, 検索対象データのノルム計算処理からなる。部分空間選択処理の計算量は, 意味空間へマッピングされた検索対象データ数に依存しておらず, ノルム計算時間と比較して大変短い。ノルム計算の計算量は, 検索対象データ数 ( $N$ ) のオーダ, すなわち  $O(N)$  であり, ノルム計算結果をランキングする場合は, そのランキング (ソーティング) にかかる計算量  $O(N \log_2 N)$  が必要になる。

このノルム計算について, 提案方式では, 文献 6) に示す高速化アルゴリズムにより,  $O(N)$  以下の計算によりノルムの大きい (高い相関を有する) 検索対象データを獲得することが可能となる。

## 7. 結 論

本稿では, 事象に関する文脈解釈機能を有する事象間の因果関係計量による意味的連想検索方式を示した。

まず, 因果関係を扱うベクトル空間生成方式を, 意味の数学モデルによる意味的連想検索方式に適用するための実現方式について示した。次に, 擬似事象データおよび宇宙工学分野を対象とした事象データを用いた検索実験を行い, 提案方式の有効性, および実現可能性の確認を行った。実験 A および B により, 提案方式により実現される基本的な因果関係検索機能について確認した。さらに, 文献 15) に対する提案方式の特徴として次の項目について確認した。

- 提案方式により, 文脈解釈機能による動的な因果関係検索を実現される。本方式により, 問合せとして, 発生した事象を示すキーワードに加えて, その原因や結果となる事象群を特定するための状況を文脈として与えることにより, 多様な状況に応じて, 発生した事象の原因や結果である事象に関して記述された文書群を獲得し, 発生した事象の原因や結果を状況ごとに究明することが可能となる (実験 C)。
- また, 提案方式では, 問合せとして, ある事象単語に加えて状況を示す事象単語を文脈として指定

して検索を行うことにより、コサイン距離を用いた場合の検索では容易に得られないデータについての情報獲得が可能となる（実験 D）。

実験 A では、因果関係マトリクスをコサイン距離に適用した場合に、間接的な因果関係の計量は行えないが、直接的な因果関係を計量する場合においては良い結果を示していることが確認できる。したがって、検索者の検索目的に応じて、提案方式による検索、および内積計算やコサイン距離計算による検索のいずれか、または両者を組み合わせて適切に利用することにより、より柔軟な因果関係検索環境を実現できると考えられる。

さらに今後は、大量の実事象データ・文書データを対象とした実験システムを構築し、本方式の有効性評価実験等を行っていく予定である。

謝辞 本研究に関して、多くの貴重なご助言をいただいた、宇宙航空研究開発機構但田育直氏、および波内みさ氏に感謝申し上げます。

#### 参考文献

- 1) Deerwester, S.C., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R.A.: Indexing by latent semantic analysis, *J. Am. Soc. Inf. Sci.*, Vol.41, No.6, pp.391-407 (1991).
- 2) Fung, R. and Del Favero, B.: Applying bayesian networks to Information Retrieval, *Comm. ACM*, Vol.38, No.3, pp.42-48 (1995).
- 3) Kiyoki, Y., Kitagawa, T. and Hayama, T.: A metadatabase system for semantic image search by a mathematical model of meaning, *ACM SIGMOD Record*, Vol.23, No.4, pp.34-41 (Dec. 1994).
- 4) Kiyoki, Y., Kitagawa, T. and Hayama, T.: A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning, *Multimedia Data Management — using metadata to integrate and apply digital media*, Sheth, A. and Klas, W. (Ed.), Chapter 7, McGrawHill (1998).
- 5) 清木 康, 金子昌史, 北川高嗣: 意味の数学モデルによる画像データベース探索方式とその学習機構, 電子情報通信学会論文誌 D-II, Vol.J79-D-II, No.4, pp.509-519 (1996).
- 6) 宮川明子, 清木 康, 宮原隆行, 北川高嗣: 画像データベースを対象とした意味的連想検索の高速化アルゴリズム, 情報処理学会論文誌: データベース, Vol.41, No.SIG 1(TOD5), pp.1-10 (2000).
- 7) Pearl, J.: *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Francisco, CA (1988).
- 8) Salton, G.: *The SMART Retrieval System — Experiments in Automatic Document Processing*, Prentice Hall Inc., Englewood Cliffs, NJ (1971).
- 9) Salton, G., Wong, A. and Yang, C.S.: A vector space model for automatic indexing, *Comm. ACM*, Vol.18, No.11, pp.613-620 (1975).
- 10) Tada, I., Namiuchi, M., Nakagawa, K., Takano, K., Zushi, T. and Kiyoki, Y.: An Application of the Semantic Associative Search Method to Nonconformance Information Retrieval, *International Symposium on Space Technology and Science* (June 2004).
- 11) 鷹野孝典, 清木 康: 異分野データベース群を対象とした意味的検索空間統合プロセスの実現, *DBSJ Letters*, Vol.1, No.1, pp.55-58 (2002).
- 12) 鷹野孝典, 図子泰三, 清木 康, 但田育直, 波内みさ: 時間的因果関係を扱う動的な文脈解釈を伴った意味的連想検索方式の実現, 情報処理学会研究報告 DBS-131, pp.483-489 (2003).
- 13) 図子泰三, 吉田尚史, 清木 康: ドキュメントデータ群を対象とした文脈依存動的クラスタリングの再帰的適用による意味的知識発見方式, 情報処理学会論文誌: データベース, Vol.43, No.SIG 2(TOD13), pp.216-230 (2002).
- 14) 図子泰三, 鷹野孝典, 清木 康: 事象データ群の時間的因果関係を扱う意味的連想検索方式, 情報処理学会研究報告 DBS-130, pp.71-77 (2003).
- 15) 図子泰三, 清木 康, 鷹野孝典, 但田育直, 波内みさ: 事象データ間の因果関連性計量機能をともなったベクトル空間検索方式, 情報処理学会論文誌: データベース, Vol.45, No.SIG 7(TOD22), pp.124-136 (2004).

(平成 16 年 9 月 20 日受付)

(平成 17 年 1 月 1 日採録)

(担当編集委員 吉岡 真治)



鷹野 孝典 (学生会員)

1998 年慶應義塾大学環境情報学部卒業。同年 (株) 図研入社。2003 年慶應義塾大学大学院政策・メディア研究科修士課程修了。同年より、同大学院同研究科博士課程に在学。データベースシステムの研究に従事。ACM, 日本データベース学会各会員。



図子 泰三（学生会員）

1999年慶應義塾大学環境情報学部卒業。2002年同大学大学院政策・メディア研究科修士課程修了。現在、同大学院同研究科後期博士課程に在学中。データベースシステム、デー

タマイニング、情報検索に関する研究に従事。日本データベース学会会員。



清木 康（正会員）

1978年慶應義塾大学工学部電気工学科卒業。1983年同大学大学院工学研究科博士課程修了。工学博士。同年日本電信電話公社武蔵野電気通信研究所入所。1984～1995年筑波大

学電子・情報工学系講師，助教授を経て，1996年慶應義塾大学環境情報学部助教授，1998年同学部教授。データベースシステム，知識ベースシステム，マルチメディアシステムの研究に従事。ACM，IEEE，電子情報通信学会，日本データベース学会各会員。

---