

文脈認識をともなった時空間的関連性評価方式

細川 宜秀[†] 清木 康^{††}

本論文では文脈認識をともなう時空間的関連性評価方式を提案する。時空間表現は一般に次の2タイプによって表されるが、提案方式はタイプ1とタイプ2を対象とした時空間的関連性評価機能として実現する。(タイプ1)座標表現のみによって表現される時空間表現,(タイプ2)座標表現に文脈を組み合わせることで表現される時空間表現。ここで、本論文では座標表現と組み合わせることによって、時空間上の位置を特定するための要素を‘文脈’と定義し、時空間上の相対位置や方向を表す言語表現を指すものとする。さらに、時空間表現の集合から、与えられた文脈と同一の文脈を持つ時空間表現を選択することを‘文脈認識’と定義する。タイプ2の時空間表現にはタイプ1によって表現することが困難なものがあるので、提案方式はタイプ1を適用対象としている従来の時空間的関連性評価方式によって評価困難な時空間表現間の等価性・類似性・時空間的関連性の評価を可能にする。これによって、時空間データベースの適用範囲を拡大する。さらに、実験により検索精度についての本方式の有効性を検証する。

A Spatial and Temporal Evaluation Method with Context Recognition Functions

YOSHIHIDE HOSOKAWA[†] and YASUSHI KIYOKI^{††}

In this paper, we propose a spatial and temporal evaluation method with context recognition functions. There are following two types of spatial and temporal data items: (type-1) a coordinate, and (type-2) a pair of a coordinate and a context. We define an element for pointing a location on a spatio-temporal space as a ‘context’, which is given by a linguistic phrase for representing a position from the coordinate of a type-2 data items. We also define a process for extracting spatial and temporal data items with the same context from a given set of spatial and temporal data items as ‘context recognition’. The proposed method is realized to evaluate equality, similarity and spatial and temporal relationships between two type-1 or type-2 data items. By this, the proposed method makes it possible to expand the applicable scope of spatial and temporal database systems, because conventional methods are applied to only type-1 data items. We also present an implementation method for realizing the proposed spatial and temporal evaluation functions, and clarify feasibility and effectiveness of our method by showing several experimental results.

1. はじめに

近年の広域コンピュータ・ネットワーク技術の発展・普及にともない、地理情報や時刻などを表す時空間表現を含む情報源は増大している。さらに、モバイル・コンピューティング技術の実用化にともない、それらの情報源に含まれる時空間表現を介した新しい情報獲得のための技術、ならびに、それらの時空間表現の連結による新しい情報生成のための技術の確立が重要となっている。時空間表現間の時空間的関連性を評価す

るための機能(時空間的関連性評価機能)の実現は、そのような技術の確立に大きく貢献する。

時空間表現は一般に次の2タイプによって表現されるが、現在までに提案されている時空間的関連性評価機能^{(1),(2),(6),(8),(9),13)}は、そのうちの1つのタイプ(タイプ1)のみを対象としているので、次の2つのタイプを対象とした時空間的関連性評価機能の実現は、情報源に含まれる時空間表現を介した新しい情報獲得のための技術、ならびに、それらの時空間表現の連結による新しい情報生成のための技術の確立に貢献する。(タイプ1)座標表現のみによって構成される時空間

表現 $d_1 = (C)$

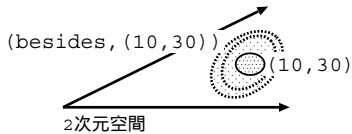
この時空間表現は、時空間上の位置を座標表現を用いて厳密に表現する。ここで C は (x, y) のような座標表現を表す。

[†] 筑波大学電子・情報工学系

Institute of Information Sciences and Electronics, University of Tsukuba

^{††} 慶應義塾大学環境情報学部

Faculty of Environmental Information, Keio University



時空間表現 (besides, (10, 30)) の文脈 besides は、座標表現 (10, 30) が指す時空間上の位置の周辺を表すのに貢献するが、その周辺を表す領域の境界の候補は不特定多数であるので、特定の境界が与えられることを前提とするタイプ 1 によって表現することは困難である。このようにタイプ 2 は、タイプ 1 では表現困難な時空間上の位置を表現することを可能にする。

図 1 タイプ 2 の時空間表現の主要な特徴

Fig. 1 The main feature of a type-2 spatial and temporal data item.

(タイプ 2) 座標表現に文脈を組み合わせることによって特定される時空間上の位置や時刻を表す時空間表現 $d_2 = (CX, C)$

ここで C は (x, y) のような座標表現を表し、 CX は文脈を表す。本論文では、座標表現 C と組み合わせることによって、時空間上の位置や時刻を特定するための要素を文脈と定義する。

この表現の主要な特徴は次のとおりである。

(特徴) タイプ 1 によって表現困難な時空間上の位置を表現可能にする。

図 1 に示すように、境界の定義が困難な時空間上の領域を表現可能にする。

新聞記事を対象としたドキュメント・データベースシステムはこのタイプの時空間表現を含む情報源として位置付けられる。

以後本論文では時空間上の位置、領域、時刻などを時空間的位置と総称する。時空間表現によって指される時空間的位置をその時空間表現の意味と定義する。

本論文ではタイプ 1 とタイプ 2 の時空間表現間の等価性、類似性、時空間的関連性を評価するための機能の実現方式を提案する。

現在までに提案されている時空間的関連性評価機能^{1),2),6),8),9),13)} はタイプ 1 間の時空間的関連性を評価することを目的としている。提案方式は、タイプ 1 では表現困難な時空間的位置を表現するタイプ 2 間の等価性、類似性、時空間的関連性評価を可能にするので、時空間データベースの適用範囲を拡大させる。ここで、Allen の 13 時区間モデル¹⁾ や Masunaga の時間モデル¹³⁾ は 1 次元の時間軸上の 2 つの区間の関連性を評価するための方式を示している。Egenhofer の Intersection Matrix²⁾ は 2 次元空間上の 2 つの座標表現の関連性を評価するための方式を示している。文献 6), 8), 9) において、我々は異種データベース環境を対象とした時空間的関連性評価機構の実現方式を示している。

また提案方式では、与えられた時空間表現の集合から、与えられた文脈と同一の文脈を持つ時空間表現を選択することを文脈認識と定義する。これより、提案方式を文脈認識をともなった時空間的関連性評価方式と定義する。

文脈認識機構を有する関連性評価モデルとして意味の数学モデル^{11),12)} が提案されている。意味の数学モデルは、複数の意味を持つ単語を文脈に応じてその意味を動的に確定するための意味的連想検索機構を実現している。この方式の特徴は約 2000 次元の正規直交空間上に単語を配置し、文脈認識機能を部分空間選択機能として実現している点にある。これにより約 2^{2000} の文脈の認識を可能にしている。さらに、単語間の意味的な等価性は選択された部分空間における距離計算によって評価している。一方、提案方式は、時空間的位置を特定するための文脈認識機能を実現している。これより座標表現に文脈を組み合わせることによって指される時空間的位置間の等価性、類似性、時空間的関連性の評価を可能にする。

さらに、提案方式における文脈とは、時空間上の相対位置や方向などを表す言語表現を指すものとする。これは次の 2 つの理由による。

第 1 の理由は、現在の計算技術を用いて現実的に実現可能な文脈認識機能が存在することを示すためである。一般に、文脈を表す言語表現と座標表現を組み合わせることによって表現される時空間表現が指す時空間的位置を特定するためには、その言語表現全体の認識を行う必要がある。特に、その時空間表現が文脈として文や文章などの一般的な言語表現を持つ場合には、文や文章を対象とした高度な文脈認識機能を実現する必要がある。しかし、自然言語処理技術に代表される現在の言語表現の認識技術を用いて、それらを認識し、その時空間表現が指す時空間的位置を特定することは困難である。そこで、提案方式では、文や文章などの一般的な言語表現を文脈とするのではなく、時空間上の相対位置や方向などを表す言語表現を文脈とすることにより、現在の計算技術を用いて現実的に実現可能な時空間表現の文脈認識機能を実現する。さらに、時空間上の相対位置や方向などを表す言語表現は、文などの一般的な言語表現に含まれる座標表現の近傍に出現するので、文などの一般的な言語表現を文脈とした時空間表現が表す時空間的位置の近似値を簡潔な計算によって導出することが可能となる。

第 2 の理由は、この言語表現がタイプ 1 によって表現困難な時空間的位置を表現可能にする典型的なタイプ 2 を構成する文脈として位置付けられるので、提案

方式が時空間データベースの適用範囲を拡大することを示すのに適しているからである。

空間的関連性評価と自然言語に関する研究として文献 3) がある。この研究では座標表現間の空間的関連性をそれと意味的に等価な自然言語によって表現し、データ操作時にその自然言語を介することにより、空間データベースや地理情報システムの利便性向上を達成している。この方式で使用される自然言語は空間的関連性とあらかじめ対応付けられている。提案方式はこの方式と異なり、座標表現のみでは表現困難な時空間的位置間の等価性、類似性、時空間的関連性を評価することを可能にする。これにより提案方式は時空間データベースの適用範囲を拡大する。

提案方式の特徴は次のとおりである。

(1) 時空間データベースの適用範囲の拡大

提案する実現方式の利点を明確にするために、タイプ 2 を対象とし、提案方式と実現方式の異なる時空間的関連性評価方式を比較する。

提案方式とは、時空間表現を座標表現と文脈を表す言語表現の対によって表現し、時空間表現間の時空間的関連性評価機能を、座標表現間の時空間的関連性評価機能と文脈を表す言語表現間の表現形式の同一性を評価する機能を組み合わせて実現する方式である。

一方、タイプ 2 を対象とし、提案方式と実現方式の異なる時空間的関連性評価方式とは、データベース構築時に定義されるタイプ 2 やデータ検索時に検索条件として与えられるタイプ 2 を意味的に等価なタイプ 1 に変換し、変換されたタイプ 1 に対してこれまでに提案されてきた時空間的関連性評価機能を適用する方式である。本論文ではこの方式を「表現形式変換をともなう時空間的関連性評価方式」と呼ぶことにする。

ここで、タイプ 2 の特徴でも述べたようにタイプ 2 をタイプ 1 によって表現することは一般に困難であるが、データベース設計者のデータ定義の視点やデータ検索者のデータ検索の視点を認識することによって、タイプ 2 をそれと意味的に等価なタイプ 1 に変換することが可能な場合がある。ただし、この比較は表現形式変換をともなう時空間的関連性評価方式が実現可能な場合にのみ成立するものである。提案方式は、表現形式変換をともなう時空間的関連性評価方式との比較において次の 2 つの利点を有する。

提案方式の利点 1 提案方式は、表現形式変換をともなう時空間的関連性評価方式との比較において、

ドキュメント・データへの適用性が高い。

ドキュメント・データは一般に、時空間上の座標表現に対応する言語表現と文脈を表す言語表現を含む。そのため、ドキュメント・データから提案する時空間表現（座標表現と文脈を表す言語表現の対）を抽出することは可能である。

一方、表現形式変換をともなう時空間的関連性評価方式は、ドキュメント・データに含まれるタイプ 2 からそれと意味的に等価なタイプ 1 への変換表に従って、従来の時空間的関連性評価機能に適用可能な時空間表現を生成するが、ドキュメント・データは一般にその変換表を構成するのに必要なデータを含んでいないので、表現形式変換をともなう時空間的関連性評価方式はドキュメント・データ作成者にその変換表の作成を強要する。

このように、提案方式は、表現形式変換をともなう時空間的関連性評価方式と異なりタイプ 2 からタイプ 1 への変換表を作成するオーバーヘッドがないので、ドキュメント・データへの適用性が高いといえる。本論文では提案方式のドキュメント・データへの適用性に関する検証実験を行い、その適用性を明らかにする。さらに、提案方式が時空間データベースの適用範囲を拡大することを明らかにする。

提案方式の利点 2 提案方式は、表現形式変換をともなう時空間的関連性評価方式との比較において、実現に要するオーバーヘッドを軽減して、時空間データベースの適用範囲を拡大する。

表現形式変換をともなう時空間的関連性評価方式では、データ定義の視点やデータ検索の視点を認識する機構の実現が必要である。この機構の実現方法として、タイプ 2 からそれと意味的に等価なタイプ 1 への変換表をあらかじめ用意し、その表から、与えられたタイプ 2 と意味的に等価なタイプ 1 を検索する方法が考えられる。このとき m 個のタイプ 2 を n 個の視点に応じて意味的に等価なタイプ 1 を検索するためには、 $m \times n$ 個のタイプ 2 とタイプ 1 の対からなる変換表を作成する必要がある。データベース設計者のデータ定義の視点やデータ検索者のデータ検索の視点は無数にあるのが一般的なもので、表現形式変換をともなう時空間的関連性評価方式の実現オーバーヘッドは大きい。

一方、提案方式はタイプ 2 からタイプ 1 への変換を要しないので、表現形式変換をともなう時空間的関連性評価方式との比較において、提案方式は

その変換を行うための機構の実現に要するオーバーヘッド，ならびに，その実行に要するオーバーヘッドなく文脈認識を行うことを可能にする．

これより提案方式は，表現形式変換をともなう時空間的関連性評価方式との比較において，実現に要するオーバーヘッドを軽減して，時空間データベースの適用範囲を拡大する．

(2) 文脈認識機能の実現方式の単純化

提案方式において表現形式が異なる言語表現を異なる意味を持つものと見なし，文脈を表す言語表現の選択機能をその表現形式のパターンマッチングによって実現する．これより文脈認識機能を簡潔に実現する．

ただし，パターンマッチングは表現形式に関する同一性を評価するので，異なる表現を持つが同じ意味を表す文脈を同じものとして認識することは困難である．たとえば，(at the end of, {(20000101),(20010101)}) と (late in, {(20000101),(20010101)}) はともに‘2000年終盤’を表す2つの時空間表現を表しているが，パターンマッチングによる文脈認識機能によってこれらを等価なものとして認識することは困難である．

そこで提案方式では，この問題を解決するためにソーラスやオントロジ，意味的連想検索機能^{(11),(12)}などの言語表現の同義性を扱うための方式を利用するための枠組みを実現する．具体的には，言語表現の同義性を扱うための機構の出力結果を同義の文脈を表す言語表現の集合として受け取り，その集合に含まれる言語表現のどれかと表現形式が一致する言語表現を持つ時空間表現を抽出する機能を実現する．また，この機構はアプリケーションにその利用目的に応じた同義の言語表現集合を設定させることによって，多様な文脈の言語表現の同義性を扱うことを可能にする．これは提案方式のアプリケーション数を増大させる．

本論文では，同じ意味を表す2つの時空間表現が同一の座標表現と異なる言語表現によって表されるとき，それらの言語表現は同義（意味的に等価）であるとする．この文脈を表す言語表現間の同義に関する関連性を文脈を表す言語表現の同義性と定義する．

(3) 時空間表現の言語独立性

提案方式は，特定の言語に依存せずに実現するために，特定の言語に依存するいかなる制約も設定しない．これより提案方式は，言語を用いて表現可能なすべての相対位置や方向を表す文脈を認識対象とする．さらに，新しい言語表現が時空間表現を構成す

る言語表現として生み出された場合に提案方式の変更を要しないという利点を持つ．

2. 文脈認識をともなう時空間的関連性評価機構の実現方式

本章では提案する文脈認識をともなった時空間的関連性評価機構の実現方式について述べる．提案方式の主要な特徴は，タイプ2間の等価性，類似性，時空間的関連性を評価するための機能を実現する点にある．これより，従来の座標表現のみによって構成される時空間表現に加えて，座標表現と言語表現を組み合わせることで表現される時空間表現を対象とした時空間的関連性評価を可能にするので，時空間データベースの適用範囲を拡大する．

さらに，文脈認識機能を文脈を表す言語表現を対象としたパターンマッチングによって実現する．これより文脈認識機能を簡潔に実現する．

提案する文脈認識をともなう時空間的関連評価は次の手順によって実行される．図2は，直線{(10,10),(50,50)}を対角線に持つ長方形内にある建物の前を表す空間表現を選択する際の提案方式の実行手続きの実行例を表す．

(Step-1) 時空間表現の文脈認識

利用者が欲する時空間表現と同じ文脈を有する時空間表現を選択する．この手続きは文脈を表す言語表現を対象としたパターンマッチングによって実現する．図2におけるこの手続きでは，‘in the front of’ という文脈を有する空間表現を選択する．

(Step-2) 座標表現を対象とした時空間的関連性評価

(Step-1)の結果として抽出された時空間表現の集合から利用者によって与えられた時空間的関連性を満たす座標表現を有する時空間表現を選択する．図2におけるこの手続きでは，‘in the front of’ という文脈を有する空間表現の集合から，直線{(10,10),(50,50)}を対角線に持つ長方形に含まれる座標表現を有する空間表現を選択する．その結果，‘in the front of’ という文脈と座標表現(20,30)を有する1つの空間表現が抽出される．

一方，従来の時空間的関連性評価方式^{(1),(2),(6),(8),(9),(13)}を図2の時空間表現群に適用した場合，座標表現(20,30)を有する4つの空間表現が抽出される．このうち3つは利用者が欲する空間表現ではない．このように提案する文脈認識機能は利用者の欲する時空間表現のみを時空間上から抽出することを可能にする．

2.1 データ構造

本節では時空間表現の文脈を認識し，その意味を特

利用者の時空間的位置特定の条件:
直線 $\{(10, 10), (50, 50)\}$ を対角線にもつ長方形内にある
建物の前を表す空間表現を選択する。

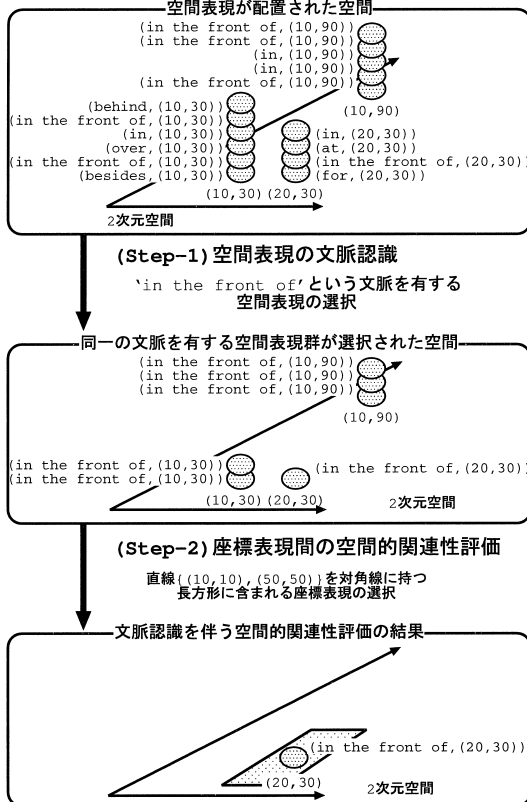


図2 文脈認識をともなう空間的関連性評価方式(提案方式)の実行例

Fig. 2 An example of spatial evaluation with context recognition functions (our method).

定するため機構のデータ構造について述べる。

時空間的位置は一般に時空間上の位置, 領域, 時刻などを表す座標表現だけでなく, 座標表現に文脈を組み合わせることで特定される。提案する文脈認識をともなう時空間的関連性評価機構のデータ構造は次のように時空間上の座標表現 C と文脈を表す言語表現 L の対によって構成する。

$$\text{時空間表現} = (L, C)$$

さらに, 提案方式では L を文字列表現とする。これより提案する文脈認識をともなう時空間的関連性評価機構のデータ構造は特定の言語に依存しないので, 1章で述べた時空間表現の言語独立性を達成する。

2.2 基本機能群

提案する文脈認識をともなう時空間的関連性評価機構は本章冒頭で述べた2手続きに対応する次の2つの関数を組み合わせて実現する。

$$(F-1) \text{ recognize_context}(D^{in}, u_l) \rightarrow D^{out}$$

この関数は, 提案方式の (Step-1) を実行し, 時空間表現の集合 D^{in} から与えられた文脈を表す言語表現 u_l と表現形式が一致する言語表現を文脈を持つ時空間表現を選択する。ここで u_l を文脈認識条件と定義する。

D^{out} は選択された時空間表現の集合を表す。

$$D^{out} = \{d | d \in D^{in} \wedge d.L = u_l\}$$

ここで d は D^{in} に含まれる時空間表現, $d.L$ は d を構成する言語表現を表す。また, $d.L$ と u_l の同一性を表現形式のパターンマッチングによって評価する。

$$(F-2) \text{ select_data}(D^{in}, \{f_c, u_c\}) \rightarrow D^{out}$$

この関数は, 提案方式の (Step-2) を実行し, 与えられた時空間的関連性を満たす座標表現を持つ時空間表現を, 時空間表現の集合 D^{in} から選択する。時空間表現の選択条件は, 時空間的関連性評価を行うためのブール関数 f_c と比較対象となる座標表現 u_c によって構成され, f_c が真を返したとき満たされるものとする。また, ブール関数 f_c として座標表現を対象とした従来の時空間的関連性評価機能^{(1),(2),(6),(8),(9),(13)}を利用する。これは, ここでの時空間表現の選択が座標表現間の時空間的関連性に基づいて行われるからである。この時空間表現の選択条件 $\{f_c, u_c\}$ を座標表現選択条件と定義する。

D^{out} は選択された時空間表現の集合を表す。

$$D^{out} = \{d | d \in D^{in} \wedge f_c(d.C, u_c) = \text{true}\}$$

ここで, d は D^{in} に含まれる時空間表現, $d.C$ は d を構成する座標表現を表す。

さらに時空間表現の集合に対する基本機能として次の3集合演算を設定する。

$$(F-3) \text{ union}(D_1^{in}, D_2^{in}) \rightarrow D^{out}$$

$$(F-4) \text{ difference}(D_1^{in}, D_2^{in}) \rightarrow D^{out}$$

$$(F-5) \text{ intersection}(D_1^{in}, D_2^{in}) \rightarrow D^{out}$$

2.3 実行手続き

提案する文脈認識をともなう時空間的関連性評価方式の実行手続きを第2.2節で述べた2つの関数を組み合わせる。

$$\text{select_data}(\text{recognize_context}(D^{in}, u_l), \{f_c, u_c\})$$

2.4 文脈を表す言語表現の同義性に関する提案方式のアプローチ

本節では文脈を表す言語表現の同義性を扱った文脈認識をともなう時空間的関連性評価方式の実現方式について述べる。

提案する文脈認識をともなう時空間的関連性評価方式において表現形式が異なる言語表現を異なる意味を持つものと見なし、言語表現の選択機能をその表現形式のパターンマッチングによって実現しているが、パターンマッチングは表現形式に関する同一性を評価するので、異なる表現を持つが同じ意味を表す文脈(を表す言語表現)を同じものとして認識することは困難である。たとえば、(at the end of, {(20000101),(20010101)})と(late in, {(20000101),(20010101)})はともに‘2000年終盤’を表す2つの時空間表現を表しているが、パターンマッチングによる文脈認識機能によってこれらを等価なものとして扱うことは困難である。

そこで提案方式では、シソーラスなどの言語表現の同義性を扱うための技術を文脈を表す言語表現の同義性を扱うための機能として利用するための枠組みを実現する。具体的には関数(F-1)を、言語表現の同義性を扱うための機構の出力結果を同義の文脈を表す言語表現集合として受け取り、その集合に含まれる言語表現のどれかと表現形式が一致する言語表現を持つ時空間表現を抽出するように拡張する。

(F-1') $recognize_context(D^{in}, U_l) \rightarrow D^{out}$

この関数は同義の文脈を表す言語表現集合 U_l を文脈認識条件として受け取り、 U_l の要素のうち少なくとも1つの言語表現を持つ時空間表現を D^{in} から選択する。

D^{out} は選択された時空間表現の集合を表す。

$$D^{out} = \{d | (\exists u_l)(u_l \in U_l \wedge d \in D^{in} \wedge d.L = u_l)\}$$

ここで u_l は U_l に含まれる言語表現、 d は D^{in} に含まれる時空間表現、 $d.L$ は d を構成する言語表現を表す。また、 $d.L$ と u_l の同一性を表現形式のパターンマッチングによって評価する。

3. 実験

本実験では提案方式(文脈認識をともなった時空間的関連性評価方式)の有効性を明らかにするために次の3検証項目について統計的検証を行った。

検証項目1: 従来方式に対する提案方式の優位性

本検証では従来の時空間的関連性評価方式^{1),2),6),8),9),13)}(従来方式(文脈認識を行わない時空間的関連性評価方式))に対する提案方式の優位性を示す。

検証項目2: 文脈を表す言語表現の同義性に関する提案方式のアプローチの妥当性

本検証では次の2つの問合せ群を文脈認識をとも

なう時空間的関連性評価機能を用いて実行させ、それらの検索結果の精度を比較することによって、文脈を表す言語表現の同義性に関する提案方式のアプローチの妥当性を示す。

問合せA群(提案方式) 同義の文脈を表す言語表現の集合を文脈認識条件として含む問合せ群

問合せB群(比較対象方式) 問合せA群に与えられる同義の文脈を表す言語表現の集合の1要素を文脈認識条件として含む問合せ群

ここで文脈を表す言語表現の同義性に関する提案方式のアプローチとは、外部から同義の文脈を表す言語表現集合を受け取り、その集合に含まれる言語表現のどれかと表現形式が一致する言語表現を持つ時空間表現を抽出するというものである。

検証項目3: 提案方式の適用性

提案方式の応用例を示すことによって提案方式の適用性を検証する。応用例として時空間的関連性評価をともなうドキュメント・データ検索を取り上げ、提案方式を組み込んだドキュメント・データ検索方式と従来方式(文脈認識を行わない時空間的関連性評価方式)を組み込んだドキュメント・データ検索方式を実現し、それらの検索結果の精度を比較することによって、従来方式が適用できなかった範囲に提案方式が適用できることを明らかにする。

実験データとして英字新聞 The Japan Times¹⁰⁾ の2000年1月から2001年5月までの野茂英雄投手に関する新聞記事44件を使用した。

本実験では実験システムをオブジェクト・リレーショナル・データベースシステムである Informix Internet Foundation.2000⁷⁾ 上に構築し、Sun Enterprise 450 (CPU: UltraSPARC-II (296 MHz) × 2, Memory: 1GB, OS: Solaris 2.6) 上で実行させた。また、時空間表現を構成する座標表現を対象とした時空間的関連性評価機能(2.2節の関数(F-2)で使用する時空間的関連性評価関数)として文献9)の時空間的関連性評価機能を使用した。これはドキュメント・データに含まれる座標表現群が地理空間上の地点や時間軸上の期間など異種の座標表現によって構成されており、文献9)の時空間的関連性評価機能はそのような異種の座標表現の集合に適用することが可能であるからである。これより座標表現の表現形式は文献9)の表現形式に従い、{計量機構識別子, 計量空間識別子, 型名, 値}の4つ組構造とした。たとえば、西暦という時間軸上の時刻(20000101)は文献9)では

表 1 実験で使った座標表現を対象とした時空間的関連性評価機能
Table 1 Boolean functions for evaluating spatial and temporal relationships used in the experiments.

時空間的関連性評価機能	真を返す条件
$t1d_contain(s, v_1, v_2, null)$	時間軸 s 上の座標表現 v_1 は座標表現 v_2 に含まれる .
$t1d_before(s, v_1, v_2, null)$	時間軸 s 上の座標表現 v_1 は座標表現 v_2 より過去にある .
$t1d_disjoint(s, v_1, v_2, null)$	時間軸 s 上の座標表現 v_1 は座標表現 v_2 と重複期間がない .
$s2d_inside(s, v_1, v_2, null)$	2次元空間 s 上の座標表現 v_1 は座標表現 v_2 の内部にある .
$s2d_equal(s, v_1, v_2, null)$	2次元空間 s 上の座標表現 v_1 は座標表現 v_2 と一致する .

$\{1,10,T1dVp,(20000101)\}$ と表現される . 地球上の 2 点 (10,10) と (100,100) を対角線とする長方形領域は文献 9) では $\{2,20,S2dVr,(10,10),(100,100)\}$ と表現される . ここで , 計量機構識別子 '1' は座標表現に適用する関連性評価機構が時間的関連性評価機構であることを , '2' は座標表現に適用する関連性評価機構が空間的関連性評価機構であることを表す . 計量空間識別子 '10' は座標空間が西暦であることを , '20' は座標空間が地球であることを表す . 型名 'T1dVp' は西暦の時刻 , 'S2dVr' は地球上の長方形領域を表す . その他の型として本実験では西暦の時区間を表す 'T1dVi' , 地球上の地点を表す 'S2dVs' を使用した .

次に本実験で使用した文献 9) の時空間的関連性評価機能について述べる . この機能は計量空間上に配置された 2 つの座標表現 v_1, v_2 間に時空間的関連性 p があるとき真 , さもなければ偽を返す関数として定義される .

$$(PF_{[k,p]}) f_p(s, v_1, v_2, \{r\}) \rightarrow \text{boolean}$$

ここで , k は計量機構識別子 , s は計量空間識別子 , p は時空間的関連性の識別子 , r は関連性確定文脈を表す . 関連性確定文脈は 2 つの座標表現間の関連性を確定するための要素である . たとえば , 距離に関する関連性 (距離性) を評価する空間的関連性評価関数 $f_{gnd1}(s, v_1, v_2, \{d, op\})$ は関連性確定文脈を引数に持つ空間的関連性評価関数である . この関数は , 2 座標表現間の関連性を確定する距離値 d , および , オペレータ $op(\in \{<, \leq, =, \geq, >\})$ からなる関連性確定文脈を引数にとる . 例として $f_{gnd1}(20, v_1, v_2, \{\{2,20,S2dSl,1000M\}, \leq\})$ は 2 つの座標表現 v_1, v_2 間が '1000メートル以下の距離' にある場合真を返し , さもなければ偽を返す . 同様に $f_{gnd1}(20, v_1, v_2, \{\{2,20,S2dSl,2000M\}, \leq\})$ は 2 つの座標表現 v_1, v_2 間が '2000メートル以下の距離' にある場合真を返し , さもなければ偽を返す . ここで 'S2dSl' は地球上の距離を表す . このように関連性確定文脈はそれを与えることによって 2 つの座標表現間の関連性を確定する要素として機能する . 表 1 は本実験で使用した座標表現を対象とした時空間的関連

性評価機能群を表す . ここで関連性確定文脈 $null$ は関連性確定文脈をとらない時空間的関連性評価関数であることを表す . また , 関数名には p を使用した . また実験に使用した問合せ中に関数名の前に '!' があるものが存在するが , この '!' は否定を表し , f_p の出力を反転させる .

3.1 実験 1

本実験では , 従来方式 (文脈認識を行わない時空間的関連性評価方式) に対する提案方式 (文脈認識をとらぬ時空間的関連性評価方式) の優位性 (検証項目 1) を示す .

3.1.1 実験方法

提案方式と従来方式の検索結果の精度を比較することによって , 従来方式に対する提案方式の優位性を示す .

3.1.2 実験データ

新聞記事に含まれる時空間表現から提案方式と従来方式に適用する 2 つのデータ群 (表 2 , 表 3) を作成した .

表 2 は提案方式に適用する実験データのデータ構造と , データ項目の一部を表す . ここで時空間表現の座標表現として新聞記事に現れる地名や曜日・日付・年などの言語表現に対応する時空間上の座標表現を用いた . 文脈を表す言語表現は新聞記事中に現れ , 地名や日付・年などの言語表現が含まれる前置詞句から抽出した . また時間表現に関しては 'last year' や 'last week' などの副詞句として表現されているものが新聞記事中に数多く出現したので , そのような副詞句からも抽出した . たとえば 2000 年の記事中に現れる 'last year' を表す時空間表現を期間を表す時空間表現 (during, $\{1,10,T1dVi,(19990101),(19991231)\}$) として抽出した .

表 3 は従来方式に適用する実験データのデータ構造と , データ項目の一部を表す . 従来方式では文脈認識を行わないので , 提案方式に適用する実験データから文脈を表す言語表現を取り除いて作成した .

表 4 と表 5 は本実験で使用した新聞記事中に現れた文脈を表す言語表現とその数を表す .

表 2 実験 1 の方式 1 (提案する文脈認識をともなう時空間的関連性評価方式) に適用した実験データの例
 Table 2 An example of the experimental data for the Method-1 (the proposed method) in Experiment-1.

記事の識別子	座標表現の意味を確定する言語表現	座標表現
20000602	on	{1,10,T1dVp,(20000602)}
20000602	in	{2,20,S2dVs,(256,162)}
20000602	since	{1,10,T1dVi,(19451001),(19451101)}
20000602	during	{1,10,T1dVi,(19451001),(19451101)}

表 3 実験 1 の方式 2 (従来の文脈認識を行わない時空間的関連性評価方式) に適用した実験データの例

Table 3 An example of the experimental data for the Method-2 (the conventional method) in Experiment-1.

記事の識別子	座標表現
20000602	{1,10,T1dVp,(20000602)}
20000602	{2,20,S2dVs,(256,162)}
20000602	{1,10,T1dVi,(19451001),(19451101)}

表 4 実験に使用した新聞記事に現れる時間表現を構成する言語表現
 Table 4 Linguistic phrases of temporal representations in the articles used in the experiments.

言語表現	この言語表現を含む時間表現の数
on	61
in	29
during	18
since	10
of	4
after	2
from	2
at the end of	1
late in	1
later	1
before	1
through	1
to	1
for about	1
earlier	1
until	1

表 5 実験に使用した新聞記事に現れる空間表現を構成する言語表現
 Table 5 Linguistic phrases of spatial representations in the articles used in the experiments.

言語表現	この言語表現を含む時間表現の数
in	58
at	7
from	2

3.1.3 検証方法と実験に使用した問合せ

本実験では情報検索システムの評価指標である再現率と適合率を用いて提案方式の有効性検証を行う。これは再現率と適合率が情報検索分野における主要な検索結果の精度に関する情報検索システムの評価指標として位置付けられるからである。ここで再現率と適合率を次の式によって求めるものとする。

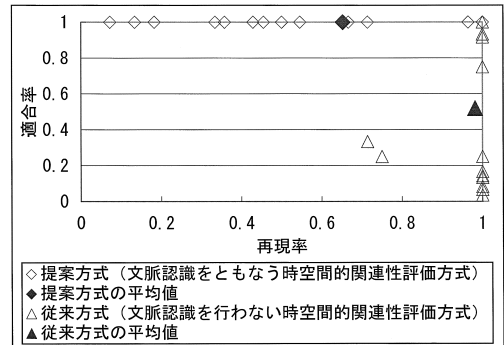


図 3 実験 1 の結果

Fig. 3 The results of Experiment-1.

$$\text{再現率} = \frac{\text{提案方式が検索した正解記事数}}{\text{正解記事数}}$$

$$\text{適合率} = \frac{\text{提案方式が検索した正解記事数}}{\text{提案方式が検索した結果数}}$$

なお正解記事は著者が決定した。

表 4 と表 5 のすべて言語表現を文脈認識条件として使用した。これは 1 章で述べた提案方式の言語独立性を保持する状況の中で提案方式が機能することを示すためである。

さらに、提案方式が座標表現選択条件と独立に文脈認識が可能であることを示すために複数の座標表現選択条件(具体的には 3 つの座標表現選択条件)を設定し、23 個の文脈認識条件と組み合わせで表現される合計 69 個の問合せを実行させた。ただしこのうち 1 件以上の検索結果を返した問合せ 29 個を検証の対象とした。これは検索されなかったことが正しい場合に、再現率と適合率が検索結果数 0 を検索されなかったことが正しいものとして扱うことができないためである。

実験に使用した問合せの例を表 6 に示す。これは提案方式に対する文脈認識条件 ‘on’ を含む問合せと従来方式に対する問合せを表す。座標表現選択条件として 3 つの条件を設定した。

3.1.4 実験結果と考察

図 3 は 29 個の問合せに対する提案方式と従来方式の再現率と適合率を表す。さらに各指標の平均値をグ

表 6 実験 1 で使用した文脈認識条件 'on' を持つ問合せ

Table 6 Queries for extracting temporal representations with the context 'on' in Experiment-1.

提案方式に対する文脈認識条件 'on' を持つ問合せ	従来方式に対する問合せ
select_data(recognize_context(D ⁱⁿ , on), {t1d_contain, 10, {1,10,T1dVi,(20000401),(20001101)}, null})	select_data(D ⁱⁿ , {t1d_contain, 10, {1,10,T1dVi,(20000401),(20001101)}, null})
select_data(recognize_context(D ⁱⁿ , on), {t1d_contain, 10, {1,10,T1dVi,(19900401),(20000101)}, null})	select_data(D ⁱⁿ , {t1d_contain, 10, {1,10,T1dVi,(19900401),(20000101)}, null})
select_data(recognize_context(D ⁱⁿ , on), {t1d_before, 10, {1,10,T1dVp,(20000101)}, null})	select_data(D ⁱⁿ , {t1d_before, 10, {1,10,T1dVp,(20000101)}, null})

ラフにプロットした。

この結果より提案方式は正解となる時空間表現のみを検索することが確認された。すなわち、提案方式は言語に依存することなく、さらに座標表現選択条件に依存することなく、文脈認識が機能していることが明らかとなった。

また実験結果より提案方式の再現率は分散したが、これは同義の文脈を表す言語表現を文脈認識条件として別々に使用したためである。提案方式における言語表現の同義性の扱い方法については 2.4 節で述べたが、この有効性に関する検証を次節の実験 2 で述べる。

一方、従来方式に関して適合率が総じて低くなった。さらにこの結果を分析すると、適合率が低い問合せの共通点として出現回数が少なかった言語表現を文脈認識条件として持つことが確認された。これは従来方式が文脈を認識する機能を持たないので、同一の座標表現を有するが異なる意味を表す時空間表現の識別ができなかったためである。特に、同一の座標表現を有するが異なる意味を表す時空間表現がまれに出現した場合に従来方式の適合率が顕著に低下することが明らかとなった。

以上により同一の座標表現を有するが異なる意味を表す時空間表現の識別に関して、従来方式との比較において提案方式の優位性が明らかとなった。さらに文脈を表す言語表現に関して言語に依存した制限を設けないことによって 1 章で述べた提案方式の言語独立性を保持した状況の中で提案方式が機能することを明らかにした。特に表 4 に示すように一般に文脈の種類は多いが、提案方式は特定の文脈に依存することなく文脈の認識を可能にしている点に大きな価値を持っているといえる。

3.2 実験 2

本実験では文脈を表す言語表現の同義性に関する提案方式のアプローチの妥当性 (検証項目 2) を検証する。

3.2.1 実験方法

本検証では次の 2 つの問合せ群を提案方式 (文脈認識をともなう時空間的関連性評価方式) を用いて実行

させ、それらの検索結果の精度を比較することによって、文脈を表す言語表現の同義性に関する提案方式のアプローチの妥当性を示す。

問合せ A 群 (提案方式) 同義の文脈を表す言語表現の集合を文脈認識条件として含む問合せ群

問合せ B 群 (比較対象方式) 問合せ A 群に与えられる同義の文脈を表す言語表現の集合の 1 要素を文脈認識条件として含む問合せ群

ここで文脈を表す言語表現の同義性に関する提案方式のアプローチとは、外部から同義の文脈を表す言語表現集合を受け取り、その集合に含まれる言語表現のどれかと表現形式が一致する言語表現を持つ時空間表現を抽出するというものである。

3.2.2 実験データ

実験 1 において提案方式 (文脈認識をともなう時空間的関連性評価方式) に適用した実験データを本実験でも使用した。これはその実験データが同義の文脈を表す言語表現の集合を複数含んでいるからである。

3.2.3 検証方法と実験に使用した問合せ

本実験でも実験 1 と同様に、情報検索システムの評価指標として位置付けられる再現率と適合率を使用した。これは再現率と適合率が情報検索分野における主要な検索結果の精度に関する情報検索システムの評価指標として位置付けられるからである。再現率と適合率は 3.1.3 項で述べた式に従う。

次に本実験で使用した問合せについて述べる。問合せ A 群として、表 4 と表 5 に含まれる次の 3 つの同義の言語表現の集合を文脈認識条件として含む問合せ群を設定した。

(同義の言語表現集合 1) 時間表現の {on, in, during}

これらの言語表現は、それらを文脈として含む時空間表現に「出来事が時空間上で発生した時点や期間」という意味付けを与える。新聞記事の多くは時空間上で発生した出来事について記述しているので、時点での出来事の発生ならび期間中の出来事の発生を同義とし、その時点ならびに期間の意味付けを与える言語表現を同義とした。

(同義の言語表現集合 2) 時間表現の {late in, at the

表 7 実験 2 で使用した問合せ
Table 7 Queries used in Experiment-2.

同義の言語表現集合	問合せ A 群 (文脈を表す言語表現 の同義性を考慮した問合せ群)	問合せ B 群 (文脈を表す言語表現 の同義性を考慮しない問合せ群)
同義の言語表現集合 1 {on, in, during }	select_data(recognize_context(D ⁱⁿ , {on, in, during}), {t1d_before, 10, {1,10,T1dVp,(20010501)}, null})	select_data(recognize_context(D ⁱⁿ , on), {t1d_before, 10, {1,10,T1dVp,(20010501)}, null}) select_data(recognize_context(D ⁱⁿ , in), {t1d_before, 10, {1,10,T1dVp,(20010501)}, null}) select_data(recognize_context(D ⁱⁿ , during), {t1d_before, 10, {1,10,T1dVp,(20010501)}, null})
同義の言語表現集合 2 {late in, at the end of}	union(union(union(select_data(recognize_context(D ⁱⁿ , {late in, at the end of}), {!t1d_disjoint, 10, {1,10,T1dVi,(19990901,20000101)}, null}), select_data(recognize_context(D ⁱⁿ , {late in, at the end of}), {!t1d_disjoint, 10, {1,10,T1dVi,(20000901,20010101)}, null})), select_data(D ⁱⁿ , {t1d_contain, 10, {1,10,T1dVi,(19990901,20000101)}, null})), select_data(D ⁱⁿ , {t1d_contain, 10, {1,10,T1dVi,(20000901,20010101)}, null}))	union(select_data(recognize_context(D ⁱⁿ ,late in), {!t1d_disjoint, 10, {1,10,T1dVi,(19990901,20000101)}, null}), select_data(recognize_context(D ⁱⁿ ,late in), {!t1d_disjoint, 10, {1,10,T1dVi,(20000901,20010101)}, null})) union(select_data(recognize_context(D ⁱⁿ ,at the end of), {!t1d_disjoint, 10, {1,10,T1dVi,(19990901,20000101)}, null}), select_data(recognize_context(D ⁱⁿ ,at the end of), {!t1d_disjoint, 10, {1,10,T1dVi,(20000901,20010101)}, null})) union(select_data(D ⁱⁿ , {t1d_contain, 10, {1,10,T1dVi,(19990901,20000101)}, null}), select_data(D ⁱⁿ , {t1d_contain, 10, {1,10,T1dVi,(20000901,20010101)}, null}))
同義の言語表現集合 3 {in, at}	select_data(recognize_context(D ⁱⁿ ,{in, at}), {s2d_inside, 20, {2,20,S2dVr,(100,100),(350,200)}, null})	select_data(recognize_context(D ⁱⁿ ,in), {s2d_inside, 20, {2,20,S2dVr,(100,100),(350,200)}, null}) select_data(recognize_context(D ⁱⁿ ,at), {s2d_inside, 20, {2,20,S2dVr,(100,100),(350,200)}, null})

end of}

これらの言語表現は「ある期間の終わり」という同義の文脈を表す。

(同義の言語表現集合 3) 空間表現の {in, at}

これらの言語表現は「出来事が時空間上で発生する地点(あるいは領域)」という同義の文脈を表す。

問合せ B 群として上記 3 つの同義の言語表現集合の 1 要素を文脈認識条件として含む問合せ群を設定した。表 7 は本実験で使用した問合せ A 群と問合せ B 群を構成する問合せを表す。

さらに、提案方式が座標表現選択条件と独立に文脈認識が可能であることを示すために複数の座標表現選択条件を設定し、8 個の問合せからなる問合せ A 群と

21 個の問合せからなる問合せ B 群を設定し実行させた。表 8 は座標表現選択条件ならびに各問合せに対する正解数を表す。

3.2.4 実験結果と考察

図 4 は 8 個の問合せを含む問合せ A 群(同義の文脈を表す言語表現の集合を文脈認識条件として含む問合せ群)と 21 個の問合せを含む問合せ B 群(問合せ A 群に与えられる同義の文脈を表す言語表現の集合の 1 要素を文脈認識条件として含む問合せ群)の検索結果の再現率と適合率を表す。さらに各指標の平均値をグラフにプロットした。

この結果より提案方式が問合せ A 群に対して座標表現選択条件に依存することなく、すべての正解デー

表 8 実験 2 で使用した問合せの座標表現選択条件と正解数

Table 8 Conditions for selecting coordinates used in the queries and the number of correct data items.

	座標表現選択条件	説明	正解数
同義の言語 表現集合 1 に 関する問合せ の座標表現 選択条件	{t1d_before, 10, {1,10,T1dVp,(20010501)}, null}	2001/5/1 以前の 時間表現選択	43
	{t1d_before, 10, {1,10,T1dVp,(20000501)}, null}	2000/5/1 以前の 時間表現選択	17
	{t1d_before, 10, {1,10,T1dVp,(19990501)}, null}	1999/5/1 以前の 時間表現選択	8
同義の言語 表現集合 2 に 関する問合せ の座標表現 選択条件	{!t1d_disjoint, 10, {1,10,T1dVi,(19990901),(20000101)}, null}	1999 年と 2000 年 終盤を表す 時間表現選択	7
	{!t1d_disjoint, 10, {1,10,T1dVi,(20000901),(20010101)}, null}		
	{t1d_contain, 10, {1,10,T1dVi,(19990901),(20000101)}, null}		
	{t1d_contain, 10, {1,10,T1dVi,(20000901),(20010101)}, null}		
	{t1d_disjoint, 10, {1,10,T1dVi,(20000901),(20010101)}, null}	2000 年 終盤を表す 時間表現選択	5
{t1d_contain, 10, {1,10,T1dVi,(20000901),(20010101)}, null}			
同義の言語 表現集合 3 に 関する問合せ の座標表現 選択条件	{s2d_inside, 20, {2,20,S2dVr,(100,100),(350,200)}, null}	U.S.A 内の 座標表現選択	36
	{s2d_inside, 20, {2,20,S2dVr,(300,100),(350,200)}, null}	U.S.A 東部の 座標表現選択	13
	{s2d_equal, 20, {2,20,S2dVs,(310,156)}, null}	ボルティモア地域の 座標表現選択	3

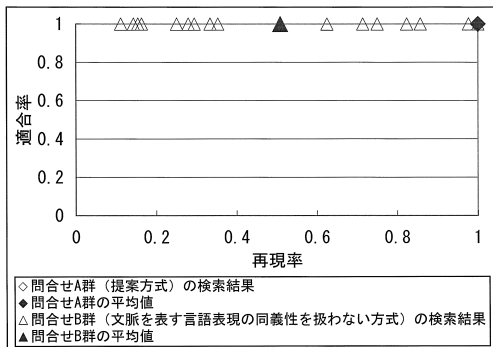


図 4 実験 2 の結果

Fig. 4 The results of Experiment-2.

タのみを検索することを確認した。

一方、問合せ B 群では再現率は分散した。これは同義の文脈を表す言語表現を文脈認識条件として別々に使用したためである。

以上より、文脈を表す言語表現の同義性に関する提案方式のアプローチの妥当性を明らかにした。

3.3 実験 3

本実験では提案方式(文脈認識をともなった時空間的関連性評価方式)の適用性(検証項目 3)を検証する。

3.3.1 実験方法

提案方式の応用例を示すことによって提案方式の適用性を検証する。応用例として時空間的関連性評価をともなうドキュメント・データ検索を取り上げ、提案方式を組み込んだドキュメント・データ検索方式と従

来方式(文脈認識を行わない時空間的関連性評価方式)を組み込んだドキュメント・データ検索方式を実現し、それらの検索結果の精度を比較することによって、従来方式が適用できなかった範囲に提案方式が適用できることを明らかにする。ここでドキュメント・データ検索方式として、現在の主要なドキュメント・データ検索方式の 1 つとして位置付けられ、ドキュメント・データに対する検索キーワードの包含性を評価する方式を使用した。

方式 1 文脈認識をともなう時空間的関連性評価をともなうドキュメント・データ検索方式

この方式は提案方式をドキュメント・データ検索方式に適用した方式である。

この方式のデータ構造を新聞記事の識別子、新聞記事データ、新聞記事に含まれる座標表現、および、文脈を表す言語表現の組として実現する。実行手続きは次のとおりである。(手続き 1)新聞記事の組集合から文脈認識条件と一致する新聞記事の組を抽出する。(手続き 2)(手続き 1)の結果得られた組集合の座標表現に時空間的関連性評価機能を適用し、利用者によって与えられた時空間的関連性を満足する新聞記事の組を抽出する(手続き 3)(手続き 2)の結果得られた組集合の新聞記事データに対する検索キーワードの包含性を評価し、検索キーワードを含む組を抽出する。ここで、(手続き 1)と(手続き 2)それぞれは、本論文 2 章で述べた提案する文脈認識をともなう時

表 9 実験 3 に使用した実験データの例
Table 9 A sample of the experimental data for Experimentn-3.

記事の 識別子	記事記事 データ	時空間表現	
		言語表現	座標表現
20000602	In Chicago, .. (省略).. walked two.	in	{2,20,S2dVs,(256,162)}
20000602	In Chicago, .. (省略).. walked two.	on	{1,10,T1dVp,(20000602)}
20000602	In Chicago, .. (省略).. walked two.	since	{1,10,T1dVi,(19451001),(19451101)}
20000602	In Chicago, .. (省略).. walked two.	during	{1,10,T1dVi,(19451001),(19451101)}

空間的関連性評価手続き (Step-1) と (Step-2) に対応する。

方式 2-1 文脈認識を行わない時空間的関連性評価をともなうドキュメント・データ検索方式 1

この方式は従来の文脈認識を行わない時空間的関連性評価機能をドキュメント・データ検索方式に適用した方式として位置付けられる。

この方式のデータ構造を新聞記事の識別子, 新聞記事データ, および, 新聞記事に含まれる座標表現の組として実現する。実行手続きは次のとおりである。(手続き 1) 新聞記事の組集合の座標表現に時空間的関連性評価機能を適用し, 利用者によって与えられた時空間的関連性を満足する新聞記事の組を抽出する。(手続き 2) (手続き 1) の結果得られた組集合の新聞記事データに対する検索キーワードの包含性を評価し, 検索キーワードを含む組を抽出する。

方式 2-2 文脈認識を行わない時空間的関連性評価をともなうドキュメント・データ検索方式 2

この方式は従来の文脈認識を行わない時空間的関連性評価機能をドキュメント・データ検索方式に適用した方式として位置付けられる。さらに方式 2-1 と異なり, この方式において新聞記事データに対する文脈認識条件の包含性評価により文脈の認識を試みる。これはこの方式がドキュメント・データ検索における文脈認識機能の実現形態の 1 つであるからである。

この方式のデータ構造を新聞記事の識別子, 新聞記事データ, および, 新聞記事に含まれる座標表現の組として実現する。実行手順は次のとおりである。(手続き 1) 新聞記事の組集合の座標表現に時空間的関連性評価機能を適用し, 利用者によって与えられた時空間的関連性を満足する新聞記事の組を抽出する。(手続き 2) (手続き 1) の結果得られた組集合の新聞記事データに対する検索キーワードと文脈認識条件の包含性を評価し, 検索キーワードと文脈認識条件を含む組を抽出する。

方式 3 時空間的関連性評価をともなわないドキュメント・データ検索方式

この方式は現在の主要なドキュメント・データ検索方式の 1 つとして位置付けられる。新聞記事データの集合に対する検索キーワードの包含性を評価し, 検索キーワードを含む新聞記事データを抽出する。

3.3.2 実験データ

実験データのデータ構造として新聞記事の識別子, 新聞記事の全文データ, ならびに, 新聞記事に含まれる時空間表現の組構造を設定した。表 9 は実験データの一部を表す。

3.3.3 検証方法と実験に使用した問合せ

本実験でも実験 1, 2 と同様に, 情報検索分野における主要なシステムの評価指標として位置付けられる再現率と適合率を使用した。これは再現率と適合率が情報検索分野における主要な検索結果の精度に関する情報検索システムの評価指標として位置付けられるからである。再現率と適合率は第 3.1.3 節で述べた式に従う。

本実験では提案方式のドキュメント・データ検索方式への適用性を示すために, 新聞記事データに対して複数の検索キーワード群を設定した。さらに, 提案方式が座標表現選択条件と独立に文脈認識が可能であることを示すために複数の座標表現選択条件を設定し, 6 つの問合せ群合計 26 個の問合せを実行させた。表 10 は検索条件ならびに正解数を表す。

3.3.4 実験結果と考察

図 5, 図 6, 図 7, 図 8, 図 9, 図 10 はそれぞれ問合せ第一群, 問合せ第二群, 問合せ第三群, 問合せ第四群, 問合せ第五群, 問合せ第六群に対する各方式の検索結果の再現率と適合率の平均値を表す。

検索結果の適合率に関して, 図 5, 図 7, 図 8, 図 9, 図 10 より, 問合せ第二群以外の問合せ群について提案方式はつねに高い適合率を示した。一方, 図 6 ではどの方式の適合率も著しく低かったが, それでも他方式と比較して提案方式の適合率は高くなった。図 6 でどの方式の適合率も著しく低かった理由は, 新聞記

表 10 実験 3 で使用した問合せ群ごとの検索条件
Table 10 The retrieval conditions of the Query Groups used in Experiment-3.

	新聞記事データ に対する検索 キーワード群	時空間表現の選択条件		正解数
		文脈認識条件	座標表現選択条件	
問合せ 第一群	Nomo, win, won	{since}	{t1d_contain, 10, {1,10,T1dVi,(20000101),(20010601)}, null} {t1d_contain, 10, {1,10,T1dVi,(20000101),(20010101)}, null} {t1d_contain, 10, {1,10,T1dVi,(20000401),(20000701)}, null} {t1d_before, 10, {1,10,T1dVp,(19990101)}, null} {t1d_before, 10, {1,10,T1dVp,(19900101)}, null}	3 3 3 3 2 2
問合せ 第二群	Nomo, win, won	{on,in,during}	{t1d_contain, 10, {1,10,T1dVi,(20000401),(20000430)}, null} {t1d_contain, 10, {1,10,T1dVi,(20000401),(20000531)}, null} {t1d_contain, 10, {1,10,T1dVi,(20000401),(20000630)}, null} {t1d_contain, 10, {1,10,T1dVi,(20000401),(20001101)}, null} {t1d_contain, 10, {1,10,T1dVi,(20000101),(20010101)}, null} {t1d_contain, 10, {1,10,T1dVi,(20000101),(20010601)}, null}	1 2 2 8 8 13
問合せ 第三群	no-hitter, complete game	{since}	{t1d_before, 10, {1,10,T1dVp,(20000101)}, null} {t1d_before, 10, {1,10,T1dVp,(19980101)}, null} {t1d_before, 10, {1,10,T1dVp,(19960101)}, null}	3 3 3
問合せ 第四群	Orioles	{in, at}	{s2d_inside, 20, {2, 20, S2dVr, (100,100),(350,200)}, null} {s2d_inside, 20, {2, 20, S2dVr, (100,100),(350,200)}, null} {s2d_equal, 20, {2, 20, S2dVs, (310,156)}, null}	3 3 2
問合せ 第五群	Tigers	{on,in,during}	{t1d_contain, 10, {1,10,T1dVi,(20000401),(20000701)}, null} {t1d_contain, 10, {1,10,T1dVi,(20000401),(20001101)}, null} {t1d_contain, 10, {1,10,T1dVi,(20000401),(20010601)}, null}	14 28 28
問合せ 第六群	Tigers	{since}	{t1d_before, 10, {1,10,T1dVp,(20020101)}, null} {t1d_before, 10, {1,10,T1dVp,(20010101)}, null} {t1d_before, 10, {1,10,T1dVp,(20000101)}, null} {t1d_before, 10, {1,10,T1dVp,(19990101)}, null} {t1d_before, 10, {1,10,T1dVp,(19900101)}, null}	3 3 3 3 2

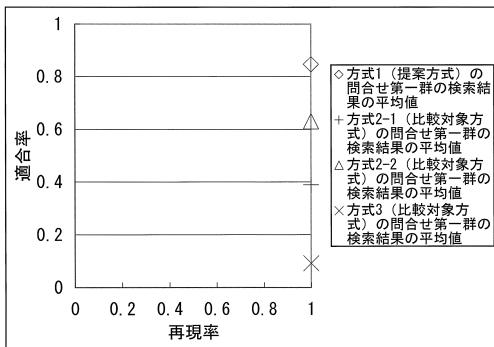


図 5 実験 3 における問合せ第一群の検索結果

Fig. 5 The results of the Query Group-1 in Experiment-3.

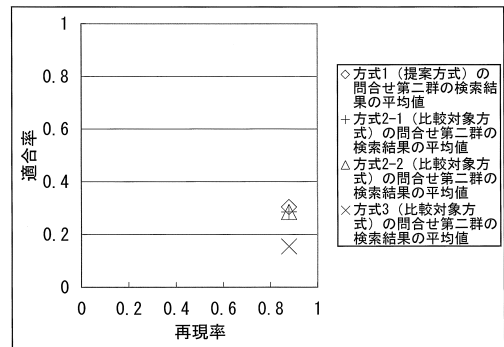


図 6 実験 3 における問合せ第二群の検索結果

Fig. 6 The results of the Query Group-2 in Experiment-3.

事に対する検索キーワードが利用者の欲する新聞記事を検索するための条件を十分に表現できなかったためである。具体的には、野茂投手勝利に関する記事の検索するために 'Nomo', 'win', 'won' の 3 単語を検索キーワードとして使用したが、本実験で使用した記事の主要な内容が勝敗に関するものであり、特に、野茂投手敗北に関する記事では、相手チームの勝利に関する内容が記述されているので、野茂投手敗北の記事においても 'Nomo', 'win', 'won' が登場し、結果とし

て野茂投手敗北記事が不正解データとして検索結果に含まれたからである。これより、提案方式が検索結果中に含まれる不正解データを削減するように機能していることが明らかとなった。

検索結果の再現率に関して、図 5、図 6、図 7、図 8、図 9 より、問合せ第一群から問合せ第五群の検索結果の再現率は同じであった。問合せ第六群については提案方式の再現率が従来方式より低下した(図 10)。この理由は、問合せ第六群が等価な意味を持つ時空間表

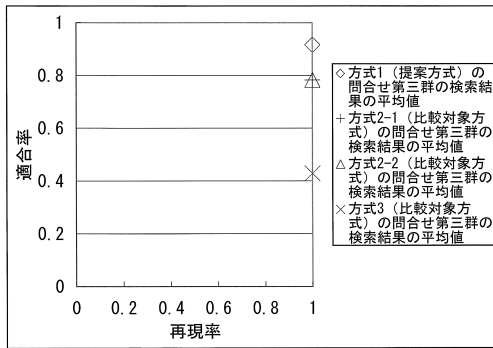


図 7 実験 3 における問合せ第三群の検索結果

Fig. 7 The results of the Query Group-3 in Experiment-3.

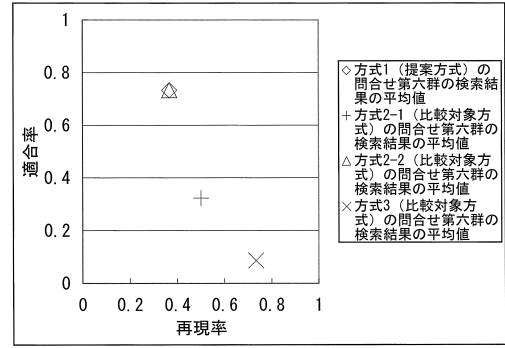


図 10 実験 3 における問合せ第六群の検索結果

Fig. 10 The results of the Query Group-6 in Experiment-3.

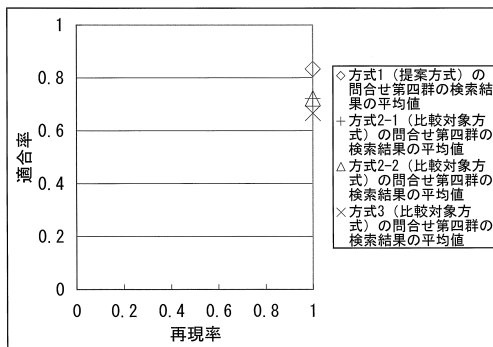


図 8 実験 3 における問合せ第四群の検索結果

Fig. 8 The results of the Query Group-4 in Experiment-3.

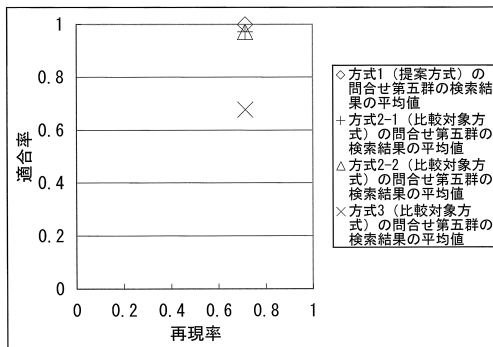


図 9 実験 3 における問合せ第五群の検索結果

Fig. 9 The results of the Query Group-5 in Experiment-3.

現を選択するための条件を十分に含んでいなかったためである。このとき等価な意味を持つ時空間表現とは 'since 1950' と 'in 50 years' であった。これらの意味的な同一性の扱い方法は、'since 時刻' によって表される時空間表現を選択するための条件と 'in 期間' によって表される時空間表現を選択するための条件を含む問合せを提案方式に実行させることによって達成される（本実験で使用した問合せは 'since 時刻' によって表さ

れる時空間表現を選択するための条件のみを含んでいた）。このように提案方式は利用者が意味的に等価であると判断した時空間表現に対する選択条件によって構成される問合せを実行することによって、時空間表現間の意味的等価性を評価することを可能にする。

方式 1 は方式 2-2 との比較において、すべての問合せ群の検索結果の再現率を低下させずにつねに高い適合率を導いた。これは 1 つの新聞記事に複数の時空間表現が含まれているのが一般的であるが、方式 2-2 は新聞記事に含まれる文脈を表す言語表現と座標表現間の対応関係を扱えないので、検索結果において不正解となる新聞記事データの数が多くなったからである。これより方式 1 は、方式 2-2 のような新聞記事に対する文脈認識条件の包含性評価によって文脈認識を試みた方式より有効であることを確認した。さらに、ドキュメント・データ検索という実際のアプリケーションにおいて、提案する文脈認識機構の妥当性を確認した。

最後に方式 1 は方式 3 との比較において適合率を大幅に向上させることが可能となった。Web ページ群に代表される大規模なドキュメント・データ群を対象としたドキュメント・データ検索において、その検索結果に含まれる不正解データ数の多さが問題となっているが、提案方式を不正解データを削減するためのフィルタリング機能として適用することは有効であると考えられる。

以上より提案方式は従来の時空間的関連性評価方式との比較において、ドキュメント・データへの適用性が高いことが明らかとなり、時空間データベースの応用範囲を拡大したことが明らかとなった。

3.4 実行時間の高速化, 問合せ処理の最適化, データベースサイズに関する考察

提案方式の実行時間の高速化に関しては、現在までに提案されている文字列照合高速化アルゴリズム、

ハッシュや B-tree などの文字列データ群からの高速文字列検索技術⁵⁾，ならびに，R-tree⁴⁾などのインデックスによる座標表現の高速検索技術を適用することによって実現可能である．これは提案方式が文脈認識機構を単純な文字列データ選択を行う関数として実現しており，座標表現の選択に関しては従来の時空間的関連性評価方式を利用しているからである．これより今後，文字列データ検索方式ならびに座標表現を対象とした時空間的関連性評価方式のさらなる高速化が達成されることによって提案方式の高速化が達成される．

提案方式を組み込んだデータベースシステムの間合せ最適化に関して，提案方式は関数として実現されているので，リレーショナル・データベースにおいて実現されている関数適用の変更による間合せ最適化方式¹⁴⁾を利用することによって，提案方式を組み込んだデータベースシステムの間合せ最適化処理を実現することが可能である．

提案方式のデータサイズは従来方式との比較において大きくなる．これは提案方式の時空間表現が座標表現と文脈を表す文字列表現の対によって表現されるからである．さらに提案方式が適用される時空間データベースのサイズは表 2 のように同一の座標表現に複数の文脈が関連付けられるので，従来方式のデータベースサイズと比較して大きくなる．たとえば表 2 のデータ項目数は表 3 のデータ項目数と比較して多い．ただし，現在計算機のメモリならびにハードディスクが安価に入手可能であり，提案方式を用いた時空間データベースの構築の妨げにはならないと考えられる．

また，データサイズの違いにともなう実行時間の比較において提案方式は従来方式より劣る．しかしながら提案方式は検索結果に含まれる不正解となるデータを従来方式より少なく抑えることができるので，利用者が要求する情報を獲得するまでに要する時間は提案方式が有利であると考えられる．

以上の考察により，提案方式の実行時間の高速化・間合せ処理の最適化については現在までに提案されている高速化・最適化手法を利用することによって達成可能であり，データベースサイズに関しては提案方式は従来方式との比較において必ずしも不利ではないことを明らかにした．

4. おわりに

本論文では文脈認識をともなう時空間的関連性評価機構の実現方式を提案した．

時空間的位置は次の 2 タイプによって表現され，提案方式は，タイプ 1 とタイプ 2 を対象とした時空間的

関連性評価方式として位置付けられる．

(タイプ 1) 時空間上の位置，領域，時刻などを表す座標表現

(タイプ 2) 座標表現に文脈を組み合わせることによって特定される時空間上の位置や時刻を表す時空間表現

さらに，提案方式では，与えられた時空間表現の集合から，与えられた文脈と同一の文脈を持つ時空間表現を選択することを文脈認識と定義した．これより，提案方式を文脈認識をともなう時空間的関連性評価方式と定義した．

提案方式の特徴は次のとおりである．

(1) 時空間データベースの適用範囲の拡大

提案方式は，タイプ 1 のみを対象としている従来の時空間的関連性評価方式と異なり，タイプ 1 とタイプ 2 を対象とした時空間的関連性評価方式として位置付けられる．さらに，タイプ 2 からそれと意味的に等価なタイプ 1 に変換することは困難なので，提案方式の適用範囲は従来の時空間的関連性評価方式のそれより広いと考えられる．そこで，本論文では，時空間的関連性評価方式の新しい適用範囲としてドキュメント・データ検索を取り上げ，検証実験によって，提案方式がドキュメント・データに対して高い適用性があることを明らかにした．これにより，提案方式が時空間データベースの適用範囲を拡大することを明らかにした．

(2) 文脈認識機能の実現の単純化

提案方式において，表現形式が異なる言語表現を異なる意味を持つものと見なし，文脈を表す言語表現の選択機能をその表現形式のパターンマッチングによって実現した．これより文脈認識機能を簡潔に実現した．

さらに，文脈を表す言語表現の同義性を扱うために，シソーラスやオントロジ，意味的連想検索機能^{11),12)}などの言語表現の同義性を扱うための方式を利用するための枠組みを実現した．文脈を表す言語表現の同義性を扱うための提案アプローチの有効性を検証実験によって明らかにした．

(3) 時空間表現の言語独立性

提案方式を特定の言語に依存せずを実現するために，特定の言語に依存するいかなる制約も設定しない．これより提案方式は，言語を用いて表現可能なすべての相対位置や方向を表す文脈を認識対象とする．本論文では，実験に使用した新聞記事データに出現するすべての文脈を表す言語表現について提案する文脈認識機能を適用することによって，提案方式の

言語独立性を確認した。

今後の課題として、(1)異なる表現形式だが同一の意味を持つ時空間表現の同義性を扱うための枠組みの実現、(2)ドキュメント・データに含まれる時空間表現の自動抽出機構の実現、(3)提案方式のWWW環境への適用などがあげられる。

謝辞 筑波大学芸術学系原田昭先生、五十嵐浩也先生、ならびに、筑波大学電子・情報工学系北川高嗣先生には本研究の遂行にあたりご助言をいただいた。ここに感謝の意を表す。

参 考 文 献

- 1) Allen, J.F.: Maintaining Knowledge about Temporal Intervals, *Comm. ACM*, No.26, pp.832-843 (1983).
- 2) Egenhofer, M.J.: Spatial Relations: Models, Inferences, and their Future Application, *Proc. Advanced Database Symposium '96*, separate volume, Japan (1996).
- 3) Egenhofer, M.J., Rashid, A. and Shariff, B.M.: Metric Details for Natural-Language Spatial-Relations, *ACM Trans. Infor. Syst.*, Vol.16, No.4, pp.295-321 (1998).
- 4) Guttman, A.: R-Trees: A Dynamic Index Structure for Spatial Searching, *Proc. 1984 ACM SIGMOD Int'l. Conf. on Mgmt. of Data*, pp.45-57 (1984).
- 5) 萩原 宏, 西原清一: 現代データ構造とプログラム技法, オーム社 (1987).
- 6) 細川宜秀, 石橋直樹, 八代夕紀子, 清木 康: マルチデータベース環境における時間的・空間的関連性評価によるデータ結合方式, 情報処理学会論文誌: データベース, Vol.40, No.SIG8(TOD 4), pp.99-111 (1999).
- 7) Internet Foundation2000, Informix. <http://www.informix.com/>
- 8) Ishibashi, N., Hosokawa, Y. and Kiyoki, Y.: A Spatial and Temporal Data Integration Method for Heterogeneous Database Environments, *Proc. IASTED International Conference APPLIED INFORMATICS Symposium 2. Networks, Parallel and Distributed Processing and Applications*, pp.323-330 (2001).
- 9) 石橋直樹, 細川宜秀, 清木 康: 時空間的文脈に応じた動的関連性計量機構を有する異種データベース間結合方式, 情報処理学会論文誌: データベース, Vol.43, No.SIG2(TOD13), pp.128-145 (2002).
- 10) *The Japan Times Monthly Bound Volume*, The Japan Times, Ltd.
- 11) Kitagawa, T. and Kiyoki, Y.: The mathematical model of meaning and its application to multidatabase systems, *Proc. 3rd IEEE Int. Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems*, pp.130-135 (1993).
- 12) Kiyoki, Y., Kitagawa, T. and Hitomi, Y.: A fundamental framework for realizing semantic interoperability in a multidatabase environment, *Journal of Integrated Computer-Aided Engineering*, Vol.2, No.1, pp.3-20, John Wiley & Sons (1995).
- 13) Masunaga, Y.: A temporal expansion to the multimedia object model in OMEGA, *Proc. 4th International Conference on Database Systems for Advanced Applications (DASFAA'95)*, pp.430-440 (1995).
- 14) Smith, J.M. and Chang, P.Y.-T.: Optimizing the Performance of a Relational Algebra Database Interface, *Comm. ACM*, Vol.18, No.10, pp.568-579 (1975).

(平成 13 年 12 月 21 日受付)

(平成 14 年 4 月 4 日採録)

(担当編集委員 加藤 俊一)



細川 宜秀(正会員)

1972年生。1995年筑波大学第三学群情報学類卒業。1999年同大学院工学研究科博士課程中退。1999年筑波大学電子・情報工学系助手を経て、2002年名古屋工業大学電気情報工学科助手、現在に至る。マルチデータベースシステム、時空間処理をともなう情報検索システムの研究に従事。日本感性工学会会員。



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, 日本ソフトウェア科学会各会員。