

# 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

石橋直樹<sup>†</sup> 細川宜秀<sup>††</sup> 清木 康<sup>†††</sup>

膨大な既存のデータベース群を連結する機構の実現により、それらの既存データベース群の価値は飛躍的に増大する。同一性、同義性、類似性、包含性などの多様な関連性を用いて異種データベース間を連結する場合、それらの関連性を明示的、かつ、静的に記述することにより、異種データベース間でのデータ結合を行う方法が用いられる。しかしながら、多くのデータベースが接続された広域ネットワーク環境では、この明示的記述のオーバーヘッドはきわめて大きいものとなる。また、ローカルデータベース内のデータ間の多様な関連性を計量する方式は多く提案されているが、これらはそれぞれが固有の適用範囲を持っており、それらを統合する環境は実現されていない。本論文では、データ間の多様な関連性を動的に計量し、異種のデータベース群を連結するメタレベルシステムのデータ統合方式を提案する。提案方式は、時空間を対象とした多様な計量機能をメタレベルに設定することにより、応用に応じた多様な時空間的計量を実現し、マルチデータベースシステムから適切な情報を検索・結合する。具体的には、時間的な計量空間、および、空間的な計量空間の設定、さらに、その機構における関連性の計量に要する文脈の定式化、および、実現を行った。また、提案方式はメタレベルシステムにおいて異種の計量システム群を統合することにより、それら計量システム固有の適用範囲を拡大する。提案方式の有効性を示すために時間的、および、空間的なデータ間の計量機構群を例示し、それらを用いた異種データベース間連結実験、および、その結果を示す。

## A Spatial and Temporal Context-dependent Data Integration Method for Heterogeneous Database Environments

NAOKI ISHIBASHI,<sup>†</sup> YOSHIHIDE HOSOKAWA<sup>††</sup> and YASUSHI KIYOKI<sup>†††</sup>

By creating an interconnection mechanism among a lot of legacy databases, the importance of those databases is enhanced significantly. To interconnect heterogeneous legacy databases, with conventional methods, as computing various inter-relationships like equality, synonymity, similarity or topological relationships, it is necessary to implement static and explicit pattern descriptions which represent various inter-relationships. However, the creation overhead of the static description is not ignorable specially in the global area network where many legacy databases are connected. There are many methods for computing various relationships among local data items. Those methods have their own applicative scopes, and an integration framework for those methods is not available. In this paper, we present a data integration method for the meta-level system which interconnects heterogeneous legacy databases by dynamically computing various inter-relationships. Our method has a capability to implement multiple computational mechanisms in the meta-level, so the method provides abilities to compute spatial and temporal relationships according to various applications. This method integrates these heterogeneous computational systems in the meta-level system. The applicative scopes of the computational methods are expanded through hybridization with others. The method provides data integration among heterogeneous databases by evaluating spatial and temporal relationships in context-dependent and independent ways.

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

<sup>††</sup> 筑波大学電子・情報工学系  
Institute of Information Sciences and Electronics, University of Tsukuba

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

### 1. はじめに

近年、広域ネットワークの高速化、および、データベース関連技術の向上により、多くのデータベース群がネットワークに接続されている。このような既存のデータベース群は、それぞれが独立に構築され、独立に管理・運用されている。これら異種データベース群

を相互に接続する機構の実現は、データベース群の価値を高めることができる<sup>2),3)</sup>。

時空間データベースの概念をマルチデータベース環境の抽象層(以下単にメタレベルと記す)へ導入することは、広域ネットワーク環境における異種データベース間の相互接続性(異種データベースの連結によってそれらの価値を高める性質)を向上させる。その理由は次のとおりである。互接続対象データベース群は、独立に作成されており、異なるドメインのデータを含むので、異種データベース間データ検索・結合のためのシステムは、そのデータのドメインを識別する機構とそのドメインに応じたデータ間の関連性計量機構を必要とする。そして、時間・空間データ検索・結合機能は、本質的に計量空間(計量を行うための空間)上に配置された異なるドメインのデータの識別機構、および、それらのデータ間の関連性計量機構を有している。

本論文では、広域ネットワーク環境に存在するデータベース群を対象とした、異なるドメインのデータ間の時間的関連性、および、空間的関連性の計量をとまなう異種データベース間データ検索・結合方式を提案する。この方式を融合型関連性計量機構を有する異種データベース間データ検索・結合方式、この方式を実現するマルチデータベースシステムを、以下ではメタレベルシステムと呼ぶ。

本論文で提案する異種データベース間結合方式におけるマルチデータベース環境において異種データベース間を時空間的に結合する方法論の本質は、メタレベルにおける時空間的関連性計量を行う計量機構群の実現にある。本論文では、時空間を対象とした多様な計量機能を、メタレベルに設定することにより、応用に応じた多様な時空間的計量を実現し、マルチデータベースシステムから適切な情報を検索・結合する方式を提案する。具体的には、時間的な計量空間、および、空間的な計量空間の設定、さらに、その機構における関連性の計量に要する文脈の定式化、および、実現方式を示す。

本方式は、次の2特徴を有する。

第1に、検索文脈とは独立かつ静的に定義される時間的・空間的関連性、ならびに検索文脈に応じて動的に定義される時間的・空間的関連性の計量をとまなう異種データベース間データ検索・結合機能の定式化を行っている点にある。ここで、前者の関連性を静的関連性、後者を動的関連性と定義する。文献8)において提案されている異種データベース間時間的・空間的データ検索・結合方式は、時間・空間に関する静的関

連性を対象としていた。本論文で対象とする時間・空間に関する動的関連性は、従来の異種データベース間での時間的・空間的データ検索・結合機能と比較して、計量対象関連性の数が多いので、それらの関連性に関する計量によって生成される新しい情報を獲得可能にする。

第2に、マルチデータベースシステムの構成を単純化する点にある。過去に提案された大局スキーマアーキテクチャ<sup>13)</sup>や連邦データベースアーキテクチャ<sup>2)</sup>、共通オントロジによる異種データベース間データ検索・結合方式<sup>7)</sup>などは、データのドメインが統一されている単一データベース環境におけるパターンマッチング技術に基づいたデータ検索・結合機能をマルチデータベース環境に適用するための方式として位置づけられる。これら過去において提案された方式は、データベース群に含まれる異なるドメインのデータを共通のドメインのデータに変換するための機構の実現に集中していた。具体的には、データ間の同一性、同義性、類似性、包含性などといった性質(以下単に関連性と記す)を、パターン記述の対集合として静的に記述する。これらの静的な関連性記述を対象に、同一性の計量を行い(すなわちパターンマッチングを適用し)、上記の関連性に応じたデータベース間統合を実現する。この過程を本論文では、“関連性変換”と呼ぶ。既存の方式において、同一性以外の関連性に応じた異種データベース間結合を実現する場合、この関連性変換が用いられる。多くのデータベース群が広域ネットワークに接続された状況では、多くのデータベース間に生じる多様な関連性に対応した関連性変換を実現する静的関連性記述の生成、および、更新のオーバーヘッドは非常に大きくなる。これを軽減する方法として、学習済みのニューラルネットワークを用いて関連性記述を動的に生成する方式<sup>12)</sup>、オントロジを用いて関連性記述を動的に生成する方式<sup>7)</sup>などが提案されている。以下では、異なるデータベース間データの関連性を明示的に記述し、パターンマッチングにより、データベース間の連結を行う方式(以下、単に“関連性記述方式”と記す)との比較において、本方式の有効性を検証する。

提案方式は、時間・空間データ検索・結合機能が異なるドメインのオブジェクトの識別機構、および、それらのオブジェクト間の関連性計量機構の実現により、データベース間データの関連性の記述を行うことなく、データベース間の検索・統合を可能にする。

本論文では、さらに、データ変換記述の数量、および、問合せ実行時間の観点から、提案方式と関連性記述方式を比較することにより、提案方式の有効性を

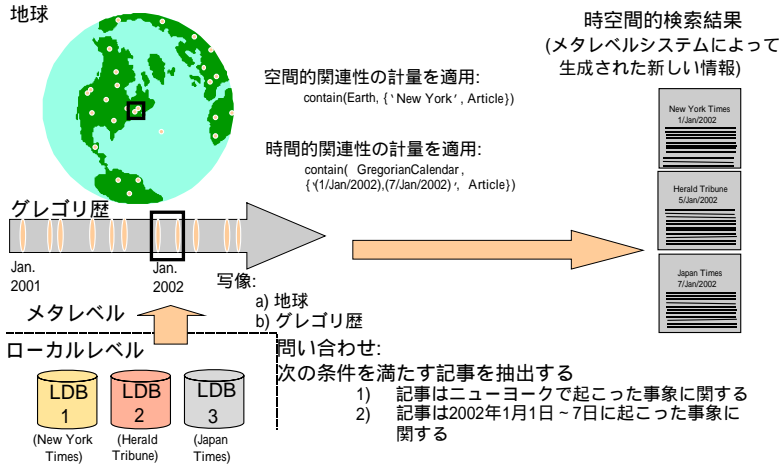


図 1 マルチデータベース環境における時空間的データ統合による新聞データベースからの知識獲得 1 (例 1)

Fig. 1 Knowledge extraction for newspaper databases by the spatial and temporal data integration of the multidatabase environment (Example 1).

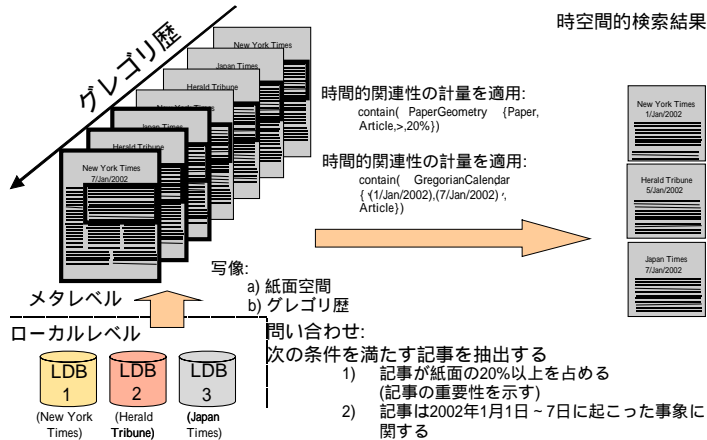


図 2 マルチデータベース環境における時空間的データ統合による新聞データベースからの知識獲得 2 (例 2)

Fig. 2 Knowledge extraction for newspaper databases by the spatial and temporal data integration of the multidatabase environment (Example 2).

示す。

## 2. 提案方式の適用例

提案方式は、関連性計量機能を有する既存システム群を組み合わせる利用することによって、異種データベース間を結合する方式として位置づけられる。図 1、図 2 は、それぞれ、3 種類のローカルデータベースシステム (Japan Times, Herald Tribune, および、New York Times) を対象に、時間的関連性計量機構、および、空間的関連性計量機構を適用し、時空間的な条件に応じた情報獲得を表している。また、それぞれの例に対応する問合せを図 3、図 4 に示す。

例 1 は、記事が示す地図空間的な所在、および、書

かれた日時に応じ、異種データベース群から記事を獲得する問合せを表す。ここでは、まず、各データベースの有する記事データを、空間的関連性の計量機構が有する地球地図  $SC(1, 1)$  に写像する。同様に、各データベースの有する記事データを、時間的関連性の計量機構が有するグレゴリ歴  $SC(2, 1)$  に写像する。次に、各関連性計量機構のデータ形式に写像された値集合を、和演算機能  $unionV$  を用いて合成する ( $MV2$ )。さらに、空間的関連性計量機構  $SCR(1)$  において、New York に空間的に含まれる記事を選択する ( $R1$ )。同時に、時間的関連性計量機構  $SCR(2)$  において、時区間  $\langle(01/01/2002),(07/01/2002)\rangle$  に含まれる記事を選択する ( $R2$ )。最後に、 $R1, R2$  を合成することで、最

```

data_conversion("LDB1",
  {NameToName,"ArticleName",CS(3,1),Vn},
  {PointToPoint,"PublishedTime",CS(2,1),Vsp},
  {PointToPoint,"Location",CS(1,1),Vgp})->V1
data_conversion("LDB2",
  {NameToName,"ArticleName",CS(3,1),Vn},
  {PointToPoint,"PublishedTime",CS(2,1),Vsp},
  {PointToPoint,"Location",CS(1,1),Vgp})->V2
data_conversion("LDB2",
  {NameToName,"ArticleName",CS(3,1),Vn},
  {PointToPoint,"PublishedTime",CS(2,1),Vsp},
  {PointToPoint,"Location",CS(1,1),Vgp})->V3
unionV(V1,V2)->MV1
unionV(MV1,V3)->MV2
mselection('New York',MV2,
  {SCR(1),CS(1,1),contain,NULL})->R1
mselection('(01/01/2002),(07/01/2002)',MV2,
  {SCR(2),CS(2,1),contain,NULL})->R2
intersectionP(R1,R2)->RESULT

```

図3 問合せ1(例1)

Fig. 3 Query 1 (Example 1).

```

data_conversion("LDB1",
  {NameToName,"ArticleName",CS(3,1),Vn},
  {PointToPoint,"PublishedTime",CS(2,1),Vsp},
  {ArticleToRegion,"Article",CS(1,2),Vgp})->V1
data_conversion("LDB2",
  {NameToName,"ArticleName",CS(3,1),Vn},
  {PointToPoint,"PublishedTime",CS(2,1),Vsp},
  {ArticleToRegion,"Article",CS(1,2),Vgp})->V2
data_conversion("LDB2",
  {NameToName,"ArticleName",CS(3,1),Vn},
  {PointToPoint,"PublishedTime",CS(2,1),Vsp},
  {ArticleToRegion,"Article",CS(1,2),Vgp})->V3
unionV(V1,V2)->MV1
unionV(MV1,V3)->MV2
mselection("TopPage",MV2,
  {SCR(1),CS(1,1),contain,{},.20 % })->R1
mselection('(01/01/2002),(07/01/2002)',MV2,
  {SCR(2),CS(2,1),contain,NULL})->R2
intersectionP(R1,R2)->RESULT

```

図4 問合せ2(例2)

Fig. 4 Query 2 (Example 2).

最終的な出力を生成する (*RESULT*) .

例2は、紙面において記事が占める面積的割合、および、書かれた日時に応じ、異種データベース群から記事を獲得する問合せを表す。ここでは、まず、各データベースの有する記事データを、空間的関連性の計量機構が有する紙面空間  $SC(1, 2)$  に写像する。同様に、各データベースの有する記事データを、時間的関連性の計量機構が有するグレゴリ歴  $SC(2, 1)$  に写像する。次に、各関連性計量機構のデータ形式に写像された値集合を、和演算機能 *unionV* を用いて合成する (*MV2*)。さらに、空間的関連性計量機構 *SCR(1)* において、紙面における面積が20%以上の記事を選択する (*R1*)。同時に、時間的関連性計量機構 *SCR(2)* において、時区間 '(01/01/2002),(07/01/2002)' に含まれる記事を選択する (*R2*)。最後に、*R1*, *R2* を合成することで、最終的な出力を生成する (*RESULT*) .

提案方式は、このように、時間的、および、空間的関連性の計量に応じた異種データベース間のデータ結合・検索を実現する。

### 3. 関連性計量機構の実現方式

本論文で提案する異種データベース間結合方式にお

けるマルチデータベース環境において、異種データベース間の時空間的な結合は、メタレベルにおける時空間的関連性計量を行う計量機構群によって実現する。本方式は、メタレベルにおけるこれらの計量機構の実現により、データベース間データの関連性を明示的に記述することなく、データベース間の検索・統合を可能にする。

本論文では、時空間を対象とした多様な計量機能を、メタレベルシステムの基本機能として設定し、時空間的な結合・選択を必要とする応用に対して、それらの基本機能を組み合わせることによって、マルチデータベースシステムから適切な情報を検索・結合する方式を提案する。具体的には、時間的な計量空間、および、空間的な計量空間の設定、さらに、その機構における関連性の計量に要する文脈の定式化、および、実現を行った。本方式は、計量機構をそれぞれ独立に扱う枠組み、および、それらを組み合わせる問合せ環境を実現するため、新しい計量機構の追加が容易である特徴を有する。

本提案方式では、静的関連性、および、動的関連性を対象とした関連性計量機能群を組み合わせるための融合型関連性計量機構を実現するために、異種データベース間のデータ統合のための実行文脈を4種に分類する。さらに、この分類に従って、静的関連性、および、動的関連性を対象とした関連性計量機能群を定式化する。具体的には、文献8)では、(CX1)、および、(CX2)を提案しており、本論文では、それらに(CX3)、(CX4)を拡張している。

(CX1) 計量機構特定文脈  $k$

計量機構特定文脈  $k$  は、メタレベルシステムへの問合せにおいて用いられる関連性計量機構を特定する。 $k$  は、空間的な関連性計量機構、時間的関連性計量機構、同一性の関連性計量機構などから、問合せ処理に用いられる関連性計量機構を特定する際に用いられる。直感的には、 $k$  は時間的な関連性、あるいは、空間的な関連性を特定する文脈である。

(CX2) 計量空間特定文脈  $s$

メタレベルシステムは、計量空間特定文脈  $s$  を用いて、 $k$  で選択された特定の関連性計量機構の有する計量空間を識別する。直感的には、たとえば、 $k$  が空間的文脈として特定された場合において、この  $s$  により、メタレベルシステムは、図1で示した地図空間、CADなどで用いられる特定の閉空間、あるいは、図2で示した紙面空間などを特定する。

(CX3) 関連性識別文脈  $p$

関連性識別文脈  $p$  を用いることにより、メタレベルシステムは、関連性計量機構  $k$  が有する計量空間  $s$  において、計量対象となるデータ間の関連性を識別する。すなわち、 $p$  は、包含性、方位性、類似性、同一性などの関連性群から、計量対象となるデータ間の関連性の選択を行う。直感的には、 $k$  が空間的関連性の計量機構、 $s$  が地図を示している場合において、 $p$  は、包含性、方位性、あるいは、類似性などの関連性を識別する文脈である。たとえば、図1、図2では、時間的、または、空間的な包含性を示す“contain”が、 $p$  に対応する。また、各関連性計量機構は、計量を提供する関連性の構成が異なるため、 $p$  の候補は  $k$  に依存して異なる。

(CX4) 関連性確定文脈  $r$

関連性確定文脈  $r$  は、計量空間  $s$  において、識別された関連性  $p$  を計量する際に用いられる文脈を示す。すなわち、関連性識別文脈  $p$  を用いて選択された関連性は、 $r$  を用いることで、その役割が確定される。たとえば、距離関係の計量を行う関連性計量関数  $gnd1(s, gv1, gv2, \{r\})$  では、二値間の関連性を確定する距離値、および、オペレータ ( $op \in \{<, \leq, =, \geq, >\}$ ) が、 $r$  に相当する。図2で示した例では、“ $\{, 20\%$ ” が  $r$  に相当する。直感的には、 $k$  が空間的関連性、 $s$  が地図、 $p$  が距離性を示している場合において、この  $r$  により、“1 km 以下” など関連性の計量に用いるパラメータを確定する。この関連性確定文脈  $r$  に応じた関連性計量機構として、たとえば、動的な語意の評価を行うことで文脈に応じた単語間の類似性を計量する方式<sup>(10),(11),(17)</sup>、自然言語表現による文脈に応じた空間的な包含性の計量方式<sup>(6)</sup>などがあげられる。関連性確定文脈  $r$  は、特定の関連性識別文脈  $p$  の定める関連性に依存するものとして位置づけられる。本論文では、関連性確定文脈に応じて動的に確定する関連性を“動的関連性”、関連性確定文脈とは独立に固定的に決定される関連性を“静的関連性”と呼ぶ。

メタレベルにおける文脈に応じたデータ間関連性の計量を行い、異種データベース間結合を行う機能(以下単にデータベース間結合機能と記す)として、次の機能を提案する。

3.1 データベース間結合機能

我々は、異種データベース間結合機能として、次の機能を設計・実現した。

(MP1)  $mjoin\{V_1^{in}, V_2^{in}, \{k, s, p, r\}\} \rightarrow P^{out}$

この機能は、値集合  $V_1^{in}, V_2^{in}$  を対象とした結合機能に対応する。本機能は、 $k$  により識別される関連性計量機構、および、 $s$  により識別される計量空間において、 $p$  により選択される関連性、および、 $r$  により固定される関連性確定文脈を満たす場合、 $V_1^{in}, V_2^{in}$  を結合し、対集合  $V^{out}$  を生成する。ここで、値集合  $V^{in}$  は、次の値表現形式で表された値の集合をさす。この値表現は、計量空間上の値と対応する名前の対(名前付き値)によって記述される。 $sn$  は、値構造の識別子を表す。

$$[name, v_{[k,s,sn]}]$$

対集合  $P^{out}$  は、次の対表現の集合を表す。この対表現は、名前  $name_1$  を持つ値が、名前  $name_2$  を持つ値と関連があることを表す。

$$(name_1, name_2)$$

また、このデータベース間結合機能を、次の関連性計量関数を用いて実現する。

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

関連性計量関数  $f_p$  は、入力値  $v_1, v_2$  が、 $k$  により識別される関連性計量機構、および  $s$  により識別される計量空間において、関連性識別文脈  $p$ 、および、関連性確定文脈  $r$  により決定される関連性の判定を行う関数である。 $PF_{[k,p]}$  は、 $v_1$ 、および  $v_2$  が、 $p, r$  を満たす関係にある場合、真値を返す。

以上のデータベース間結合機能の実行モデルを、図5に示す。ユーザからの問合せは、4種の文脈  $k, s, p, r$  を用いて記述される。この問合せに応じて、メタレベルシステムの問合せ処理系は、ローカルデータベースを対象とした関数インタプリタを用いて、個々のローカルデータベース固有の問合せを発行し、以降の処理に必要なデータを獲得する。さらに、メタレベルシステム内において、以下で示す Step-1~Step-4の処理を行う。なお、Step-2~Step-4は、各関連性計量機構固

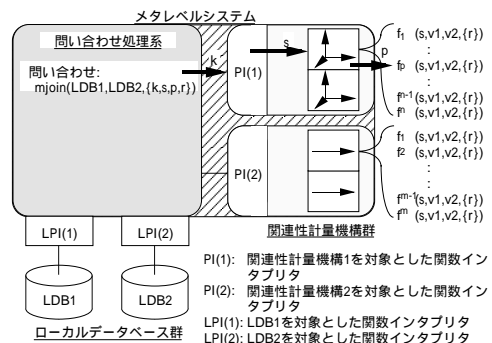


図5 メタレベル問合せの実行モデル  
Fig. 5 The execution model for the meta-level queries.

有の操作を用いて実行する必要があるため、問合せ処理系は、関数インタプリタを介して、メタレベルシステムの機能を各関連性計量機構固有の操作に変換し、処理を行う。

(Step-1) 計量機構特定文脈  $k$  に対応する関連性計量機構を特定し、ローカルデータベースより得られたデータを渡す。

(Step-2) 計量空間特定文脈  $s$  に対応する計量空間を識別し、各関連性計量機構内の計量空間にデータを写像する。

(Step-3) 関連性識別文脈  $p$  に対応する関連性を識別し、対応する関連性計量関数にデータを渡す。

(Step-4) Step-3 により識別された関連性計量関数に対し、関連性確定文脈  $r$  を渡し、関連性の計量を実行する。

また、MP1と同様に、異種データベース群を対象とした選択機能として、次の機能を設計・実現した。

(MP2)  $mselection\{V^{in}, A^{in}, \{k, s, p, r\}\} \rightarrow V^{out}$

この機能は、値集合  $V^{in}$  を対象とした選択機能に対応し、ユーザ定義の値集合  $A^{in}$  と4種の文脈  $k, s, p, r$  を満たす関係にある場合、 $V^{in}$  から部分的な値集合  $V^{out}$  を出力する。 $A^{in}$  は、 $V^{in}$  と同様の値表現形式で記述される。

### 3.2 ローカルデータベースを対象としたフィルタリング機能

本方式では、ローカルデータベースを対象としたフィルタリング機能として、次の機能を実現する。

(MP3)  $filterS\{V^{in}, \{k, s\}\} \rightarrow V^{out}$

この機能は、ローカルデータベースシステムの有する値集合  $V^{in}$  から、 $k$  により識別される関連性計量機構、および、 $s$  により識別される計量空間に写像可能である値集合  $V^{out}$  を選択する。また、この機能を用いて実現する。

(MQPF<sub>1</sub>)  $f_{[MQP, s]}(v, \{k, s\}) \rightarrow boolean$

この関数は、入力値  $v$  が、 $k$  により識別される関連性計量機構、および、 $s$  により識別される計量空間に写像可能である場合、真値を返す。

同様に、本方式では、次のフィルタリング機能を実現する。

(MP4)  $filterP1\{V^{in}, \{k, p\}\} \rightarrow V^{out}$

(MP5)  $filterP2\{V^{in}, \{k, p\}\} \rightarrow V^{out}$

これらのフィルタリング機能は、ローカルデータベースシステムの有する値集合  $V^{in}$  から、 $k$  により識別される関連性計量機構に写像可能であり、かつ、 $p$  により識別される関連性計量関数に適用可能である値集合  $V^{out}$  を選択する。MP4は関連性計量関数の

第1引数、MP5は第2引数への適用可能性を評価する。また、この機能を用いて実現する。

(MQPF<sub>2</sub>)  $f_{[MQP, p]}(v, \{k, p\}) \rightarrow boolean$

この関数は、入力値  $v$  が、 $k$  により識別される関連性計量機構に写像可能であり、かつ、 $p$  により識別される関連性計量関数に適用可能である場合、真値を返す。

## 4. システム構成

本章では、すでに提案したメタレベルシステムアーキテクチャ<sup>(8),(9)</sup>を拡張し、文脈をともなった時空間的な関連性の計量を行うデータベース間結合機能の実現方式を示す。

実際に、異種データベース間の関連性計量機構を実現するために、図6で示すマルチデータベースシステムを構築した。メタレベルシステムは、メタレベル問合せ処理系、関連性計量機構 ( $SCR(k)$ ) 群、および、その関数インタプリタより構成される。メタレベル問合せ処理系は、ユーザからメタレベルシステムに与えられた問合せを解釈・実行するシステムである。関連性計量機構  $SCR(k)$  は、データ間の関連性を計量する機構で、前述した計量空間 ( $CS(k, s)$ ) 群、関連性計量関数 ( $PF_{[k, p]}$ ) 群を有する。メタレベル関数インタプリタは、メタレベルに設定された基本機能群を、対応する計量モデルの操作に翻訳するシステムである。ローカルデータベースシステム(以下  $L_h$  と記す)を対象とした関数インタプリタは、メタレベルの操作に対応するローカルデータベースシステムの操作に翻訳し、ローカルデータベースシステムから得られる結果をメタレベルのデータ形式に翻訳するシステムである。したがって、ローカルデータベースを対象とした関数インタプリタは、データ変換機能  $C_j^{[k, h]}$  を含む。

図7は、 $L_1$ 、および  $L_2$  を対象に提案方式を適用し、 $SCR(k)$  の有する関連性計量関数群を組み合わせ

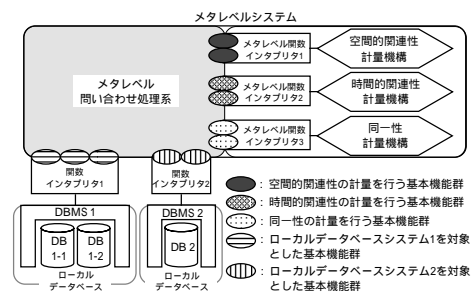


図6 マルチデータベースシステムの構成

Fig. 6 The multidatabase system architecture.

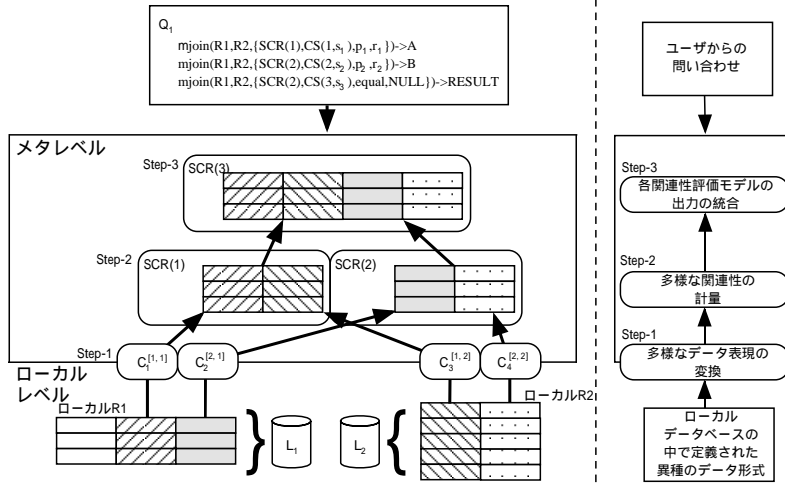


図 7 提案方式による異種データベース間結合の過程

Fig. 7 A process of data integration among heterogeneous databases by the proposed method.

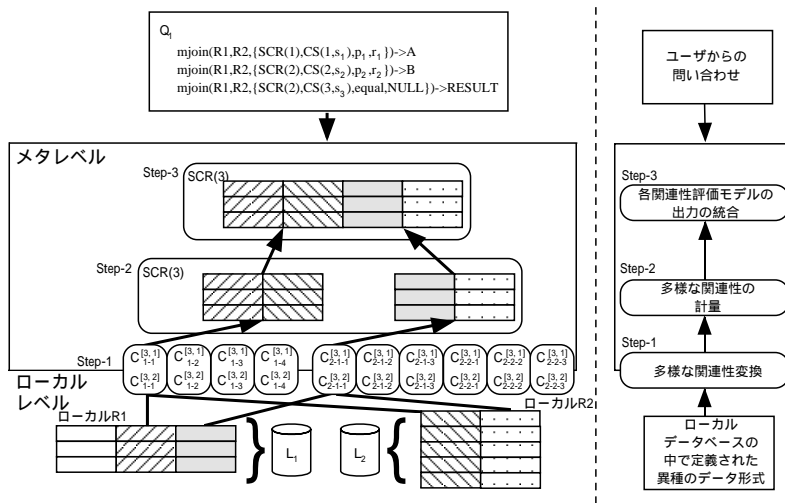


図 8 関係性記述方式による異種データベース間結合の過程

Fig. 8 A process of data integration among heterogeneous databases by the methods using the relationship conversions.

せ、データベース間結合を行う過程を示す。このデータ統合は以下の手順により実行される。

- (Step-1) ローカルデータベース群が有するデータ集合を対象に、対応する関連性計量機構の標準的なデータ形式へと変換、および、写像する。
- (Step-2) Step-1 で得られる中間結果を対象に、各々の関連性計量機構が有する関連性計量関数を適用し、メタレベルのデータ集合を結合する。
- (Step-3) Step-2 で得られる各関連性計量機構の結合結果を統合する。

提案方式によりデータベース間結合を行うためには、

単一の関連性計量機構  $SCR(k)$  と、単一のローカルデータベース  $L_h$  の間に、特定量  $D_{[k,h]}$  の変換記述を設定する必要があるため、提案方式では、式 (1) のようにそれらの総和  $D$  に相当する変換記述が必要となる。

$$D = \sum_{k=1}^l \sum_{h=1}^m D_{[k,h]} \tag{1}$$

一方、比較対象とする“関係性記述方式”によるマルチデータベースシステムは、単一の計量機構、および、多様な関連性に対応する静的記述、すなわち、関

連性変換を行う変換機能の集合を実現するものとする(図8)。ここでは、特定の関連性計量機構  $SCR(k)$ 、特定のローカルデータベース  $L_h$  の対、特定の関連性識別文脈  $p$ 、および、特定の関連性確定文脈  $r$  に応じ、記述量  $D_{[k,h,p,r]}$  の変換記述を設定する必要がある。したがって、この変換記述の総量  $D$  は式(2)で表される。ここで、 $q(k)$  は、関連性計量機構  $SCR(k)$  が計量対象とする関連性識別文脈の数を示し、 $y(p)$  は、関連性識別文脈  $p$  の計量において用いられる関連性確定文脈  $r$  の総数を示す。

$$D = \sum_{h=1}^{mC_2} \sum_{k=1}^l \sum_{p=1}^{q(k)} \sum_{r=1}^{y(p)} D_{[h,k,p,r]} \quad (2)$$

多くのデータベースが接続された広域ネットワークを対象とした場合、式(2)で示したように、要求される変換の記述量が組合せ数で増加するため、変換に要求される記述のオーバーヘッドが大きくなる。提案方式は、関連性変換を用いず、多様なデータ間の関連性を直接的に計量する機能を実現するため、その記述量を著しく減少させる。

### 5. メタレベルシステムにおける時空間的関連性計量機構の実現方式

本論文での主要な提案として、前述のメタレベルシステムアーキテクチャを対象とした、関連性確定文脈の評価による時空間的データベース間結合関数の実現方式を示す。この方式は次の3機構からなる。

#### (1) 空間的関連性計量機構 $SCR(1)$

ローカルデータベース群を空間的に連結するために、Egenhofer の提案した空間的関連性モデル<sup>5),6)</sup>を拡張し、メタレベルシステムの有する関連性計量機構  $SCR(1)$  として実現した。

$SCR(1)$  の空間構造として次のものを設定した。

#### (SSC(1)) 2次元ユークリッド空間

本方式では異種データベース間データ統合のために2次元空間を対象とした。

$SCR(1)$  のデータ構造は以下のように設定した。

- (VSC(1,1)) 2次元空間上の点,  $(x_1, y_1)$
- (VSC(1,2)) 2次元空間上の線,  $((x_1, y_1), \dots, (x_n, y_n))$
- (VSC(1,3)) 2次元空間上の領域,  $\{(x_1, y_1), \dots, (x_n, y_n)\}$
- (VSC(1,4)) 2次元空間上の長さ,  $(length)$
- (VSC(1,5)) 2次元空間上の面積,  $\{area\}$

$VSC(1,1)$ ,  $VSC(1,2)$ , および,  $VSC(1,3)$  は2次元空間上のベクター型のデータ構造を示す。

$VSC(1,4)$ , および,  $VSC(1,5)$  はスカラー型のデータ構造を示す。

また,  $SCR(1)$  において実現した関連性計量関数群の一部を次に例示する。

#### (R1) 包含関係

$$(PF_{[1,2]}) \text{ gt1\_contain}(s, gv_1, gv_2, \{op, r\}) \rightarrow \text{boolean}$$

$PF_{[1,2]}$  は、特定のユークリッド空間  $s$  において、ベクター値  $gv_1$  がベクター値  $gv_2$  を含み、かつ  $gv_1$ ,  $gv_2$  が、下式で示される割合  $r$  とオペレータ  $op$  の関係を満たす場合、真値を返す動的関連性の計量関数である。割合  $r$  は関連性確定文脈に対応する。式中  $gv_1^\circ$  は、ベクター値  $gv_1$  の内部領域を示す。

$$r = \frac{\text{Area}(gv_1^\circ \cap gv_2^\circ)}{\text{Area}(gv_1^\circ)}$$

$$op \in \{<, \leq, =, \geq, >\}$$

$PF_{[1,2]}$  は、 $VSC(1,2)$ 、または、 $VSC(1,3)$  の構造を有する1組のベクター型データ値の包含性を計量する。 $SCR(1)$  には、包含関係を対象とした33個の動的関連性の計量関数を構築した。

$$(PF_{[1,3]}) \text{ gt1\_contain}(s, gv_1, gv_2, NULL) \rightarrow \text{boolean}$$

$PF_{[1,3]}$  は、特定のユークリッド空間  $s$  において、領域  $A$  が領域  $B$  を含む場合、真値を返す静的関連性の計量関数で、 $VSC(1,1)$ 、 $VSC(1,2)$ 、または、 $VSC(1,3)$  の構造を有する1組のベクター型データ値の包含性を計量する。同様に、 $SCR(1)$  では、包含関係を計量する50個の包含関係を対象とした静的関連性の計量関数を設定・実現した。

#### (R2) 方位関係

$$(PF_{[1,84]}) \text{ gcd1\_north}(s, gv_1, gv_2, NULL) \rightarrow \text{boolean}$$

$PF_{[1,84]}$  は、空間的な方位性を計量する静的関連性計量関数の一例で、 $VSC(1,1)$ 、 $VSC(1,2)$ 、または、 $VSC(1,3)$  の構造を有する1組のベクター型データ値の方位性を計量する。同計量関数は、特定のユークリッド空間  $s$  においてベクター値  $gv_1$  がベクター値  $gv_2$  の北に位置する場合、真値を返す。同様に、 $SCR(1)$  では、方位関係を計量する16個の静的関連性計量関数を実現した。

#### (R3) 大小関係

$$(PF_{[1,103]}) \text{ gnl1\_eq}(s, gv_1, gv_2, NULL) \rightarrow \text{boolean}$$



$PF_{[1,103]}$  は、空間的な大小関係を計量する静的関連性計量関数の一例である。同計量関数は特定のユークリッド空間  $s$  上においてデータ値  $gv_1$  とデータ値  $gv_2$  の長さが等しい場合、真値を返す。この  $gv_1$ 、および  $gv_2$  には、 $VSC(1,2)$ 、 $VSC(1,3)$ 、および、 $VSC(1,4)$  のいずれかの構造を持つ。入力に  $VSC(1,2)$  を受ける場合、同計量関数は対応する線の長さを計量する。入力に  $VSC(1,3)$  を受ける場合、同計量関数は対応する領域の外周を計量する。同様に  $SCR(1)$  では、大小関係を計量する 6 個の静的関連性計量関数を実現した。

#### (R4) 距離関係

$$(PF_{[1,106]}) \quad gnd1(s, gv_1, gv_2, \{op, r\}) \rightarrow \text{boolean}$$

$PF_{[1,106]}$  は、空間的な距離性を計量する動的関連性計量関数の一例で、 $VSC(1,1)$ 、 $VSC(1,2)$ 、または、 $VSC(1,3)$  の構造を有する 1 組のベクター型データ値の距離性を計量する。同計量関数は、特定のユークリッド空間  $s$  上においてベクター値  $gv_1$  とベクター値  $gv_2$  間の距離が、距離  $r$  と下式で示す  $op$  の関係にある場合、真値を返す。同様に  $SCR(1)$  では、距離関係を計量する 6 個の動的関連性計量関数を実現した。

$$op \in \{<, \leq, =, \geq, >\}$$

#### (2) 時間的関連性計量機構 $SCR(2)$

ローカルデータベース群を時間的に連結するために、Allen の時区間モデル<sup>1)</sup>を  $SCR(2)$  として位置づけ、時間的な関連性計量機構  $SCR(2)$  を実現した。

$SCR(2)$  の空間構造は以下のように設定した。

#### (SSC(2)) 1次元ベクトル空間

$SCR(2)$  のデータ構造は以下のように設定した。

( $VSC(2,1)$ ) 1次元空間上の点、 $\{(tc)\}$

( $VSC(2,2)$ ) 1次元空間上の区間、 $\{(tc_1), (tc_2)\}$

( $VSC(2,3)$ ) 1次元空間上の大きさ、 $\{duration\}$

$VSC(2,1)$ 、および、 $VSC(2,2)$  はベクター型として位置づけられ、 $VSC(2,3)$  はベクター値の大きさをしめすスカラー型として位置づけられる。

$SCR(2)$  に実現した関連性計量関数を以下に例示する。

#### (R5) 包含関係

$$(PF_{[2,12]}) \quad tl1\_overlap(s, tv_1, tv_2, \{op_1, r_1, op_2, r_2\}) \rightarrow \text{boolean}$$

$PF_{[2,12]}$  は、包含関係を扱う動的関連性計量関数の 1 つであり、 $VSC(2,2)$  で記述された

ベクター型のデータ対を計量する。同計量関数は、特定の 1 次元ベクトル空間  $s$  上において、 $tv_1$  と  $tv_2$  が部分的に重複し、かつ、 $tv_1$ 、 $tv_2$  と、下式で示される割合  $r_1$ 、 $r_2$  が、それぞれオペレータ  $op_1$ 、 $op_2$  の関係を満たす場合、真値を返す。同様に  $SCR(2)$  には、包含関係を対象とした 5 個の動的関連性計量関数を実現した。

$$r_1 = \frac{Duration(tv_1^o \cap tv_2^o)}{Duration(tv_1^o)}$$

$$r_2 = \frac{Duration(tv_1^o \cap tv_2^o)}{Duration(tv_2^o)}$$

$$op_1, op_2 \in \{<, \leq, =, \geq, >\}$$

$$(PF_{[2,13]}) \quad tl1\_overlap(s, tv_1, tv_2, NULL) \rightarrow \text{boolean}$$

$PF_{[2,13]}$  は、包含関係を扱う静的関連性計量関数の 1 つであり、 $VSC(2,2)$  で記述されたベクター型のデータ対を計量する。同計量関数は、特定の 1 次元ベクトル空間  $s$  上において  $tv_1$  と  $tv_2$  が部分重複する場合、真値を返す。同様に  $SCR(2)$  には、包含関係を対象とした 13 個の静的関連性計量関数を実現した。

#### (R6) 前後関係

$$(PF_{[2,19]}) \quad tdc1\_before(s, tv_1, tv_2, NULL) \rightarrow \text{boolean}$$

$PF_{[2,19]}$  は、前後関係を扱う静的関連性計量関数の 1 つであり、 $VSC(2,2)$  で記述されたベクター型のデータ対を計量する。同計量関数は、特定の 1 次元ベクトル空間  $s$  上において、 $tv_1$  が  $tv_2$  の前方に位置する場合、真値を返す。同様に  $SCR(2)$  には 4 個の前後関係を対象とした静的関連性計量関数を実現した。

#### (R7) 大小関係

$$(PF_{[2,23]}) \quad tl1\_lt(s, tv_1, tv_2, NULL) \rightarrow \text{boolean}$$

$PF_{[2,23]}$  は、時間長を扱う静的関連性計量関数の 1 つであり、 $VSC(2,2)$ 、または、 $VSC(2,3)$  で記述されたデータ対を計量する。同計量関数は、特定の 1 次元ベクトル空間  $s$  上において  $tv_1$  が時間的に  $tv_2$  より短い場合、真値を返す。同様に  $SCR(2)$  には、大小関係を対象とした 3 個の静的関連性計量関数を実現した。

#### (R8) 距離関係

$$(PF_{[2,16]}) \quad td1(s, tv_1, tv_2, \{op, r\}) \rightarrow \text{boolean}$$

$PF_{[2,16]}$  は、時間的な距離を扱う動的関連

性計量関数の1つで、 $VSC(2, 1)$ 、または、 $VSC(2, 2)$ で記述されたベクター型のデータ対を計量する。同計量関数は、特定の1次元ベクトル空間  $s$  上において、 $tv_1$  と  $tv_2$  の間の時間的な間隔が、距離  $r$  とオペレータ  $op$  の関係を満たす場合、真値を返す。同様に  $SCR(2)$  には、距離関係を対象とした3個の動的関連性計量関数を実現した。

$op \in \{<, \leq, =, \geq, >\}$

### (3) 同一性計量機構 $SCR(3)$

同一性の計量機構を  $SCR(3)$  として定義し、次の空間構造を対象とする関連性計量機構  $SCR(3)$  として実現した。

( $SSC(3)$ ) 文字列集合空間:  $\{< string_1 >, < string_2 >, \dots, < string_n >\}$

$SSC(3)$  に写像される文字列は以下のデータ構造を持つ。

( $VSC(3, 1)$ ) 文字列,  $< string\_value >$

$SCR(3)$  は以下の関連性計量関数を有する。

#### (R9) 同一性

( $PF_{[3,1]}$ )  $sp\_equal(s, s1, s2, NULL) \rightarrow$   
*boolean*

$PF_{[3,1]}$  は、 $s1$ 、および  $s2$  が特定の0次元の文字列集合空間  $s$  において一致した場合、真値を返す静的関連性計量関数で、パターンマッチングを用いて実現した。 $PF_{[3,1]}$  は、 $SCR(3)$  の有する唯一の計量関数である。

## 6. 実験

提案方式は、ローカルデータベース内のデータを、対応する特定の関連性計量機構の形式に変換し、データ間の関連性を直接的に計量することにより、多様な時空間的関連性に応じたデータベース間結合を実現する。したがって、提案方式は、データ変換の観点においては、ローカルデータベース内データを対象に、関連性計量機構の形式への変換を必要とする。提案方式において要求される変換記述量は、式(1)において定式化した。

一方、従来のマルチデータベースシステムの実現方式<sup>2),7),13)</sup>は、データベース群に含まれる異なるドメインのデータを、共通のドメインのデータに変換するための機構の実現に集中していた。これらの方式では、共通のドメインに変換されたデータを対象にパターンマッチングを適用することで、異種データベース間の関連性を計量し、異種データベース間結合を実現する。提案方式との比較対照として、パターンマッチング技

術に基づいた異種データベース間データ検索・結合方式を、“関連性記述方式”と記す。この関連性記述方式を用いて、5章において述べたすべての時空間的関連性に応じた異種データベース間結合を実現するためには、すべてのデータベース間に内在するすべての関連性を、静的に記述する必要が生じる。この関連性記述の必要性より、従来のマルチデータベースシステムの実現方式は、ローカルデータベースの数量、計量対象の関連性の数量などに応じて、要求される関連性記述が著しく増加するため、実現可能性が低下する性質を有する。本論文では、関連性記述方式が有する性質として、データベースの数量、関連性の数量などに応じた関連性記述の増加傾向を、式(2)において定式化した。

以上の性質に応じて、本実験では、提案方式の有効性を示すために、次の2つの検証実験を行った。

(実験1) 変換記述オーバーヘッドに関して、提案方式と関連性記述方式を比較することにより、提案方式によるマルチデータベースシステム実現の容易性を検証する。

(実験2) 異種データベース間データ検索・結合に要する処理時間に関して、提案方式と関連性記述方式を比較することにより、提案方式の実用性を検証する。

### 6.1 実験 1

#### 6.1.1 実験手法

本実験では、すべての関連性識別文脈、および、いくつかの関連性確定文脈を問合せとして発行し、提案方式、および、提案方式の比較対象として設定した関連性記述方式が、これらすべての文脈の計量に要するデータ変換の記述量を検証する。

空間データを含む任意のデータベースが連結された一般的なマルチデータベース環境における提案方式の実現容易性を検証するために、本実験では、文献5)で示されるすべての空間的関連性を満たす空間データの最小セットを2つのデータベース  $L_1, L_2$  に格納した(図9)。さらに、文献6)で示される関連性確定文脈によって決定されるすべての空間的関連性を満たす空間データの代表セットをデータベース  $L_1, L_2$  に格納した。ここで、本実験で扱う関連性確定文脈が、いくつかの代表値をとる理由は次のとおりである。第1の理由は、本実験で扱う関連性確定文脈は、連続値ドメイン上に定義され、そのすべてを扱うことは不可能であることによる。第2の理由は、関連性変換の記述(以下、単に“関連性記述”と記す)の個数は、関連性確定文脈の値に依存しないので、 $y$  個の関連性確定文

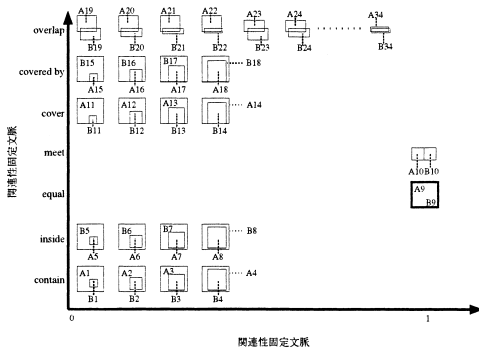


図 9 実験に用いる空間的的属性値群

Fig. 9 Spatial data values for the experiments.

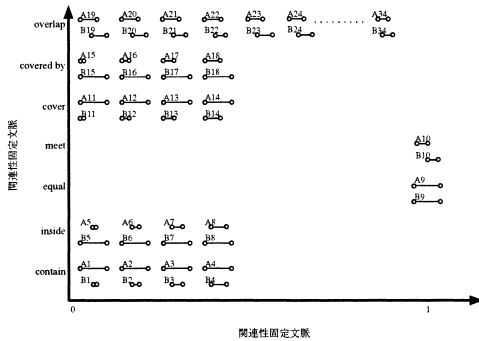


図 10 実験に用いる時間的的属性値群

Fig. 10 Temporal data values for the experiments.

脈における空間的関係に関する関連性記述個数は、 $y \times 1$  個の関連性確定文脈をとる空間的関係に対する関連性記述個数により計測されることによる。

同様に、時間データを含む任意のデータベースが連結されたマルチデータベース環境における提案方式の実現容易性を検証するために、本実験では、文献 1) で示されるすべての時間的関連性を満たす時間データの最小セット、および、時間的関連性確定文脈によって決定されるすべての時間的関連性を満たす時間データの代表セットを  $L_1, L_2$  それぞれに格納した(図 10)。

本実験では、実験システムとして、次の 2 種類のマルチデータベースシステムを実現した。

(MLS1) 関連性記述方式によるメタレベルシステム  
 関連性記述方式による MLS1 は、時空間的な関連性を同一性計量機構  $SCR(3)$  を用いて計量する。したがって、すべての時空間的な関連性を計量するために、多量な関連性記述を要求する。このため、MLS1 では、次の 80 種の記述を構成した。MLS1 は、これらの関連性記述を選択的に使用することで、時空間的な関連性の計量を行う。時間的な関連性の変換として、 $L_1, L_2$  の有するデータ値は、RC1~RC40 の共通のパターン記述

表 3 提案方式が有する時間的データ表現への変換記述  
 Table 3 Conversion descriptions of the proposed method for the temporal data representations.

| 名称     | 定義              | 記述内容                             |
|--------|-----------------|----------------------------------|
| (RC81) | $(C_1^{[2,1]})$ | $L_1$ のデータ値を、時間的データ表現(時区間)へと変換する |
| (RC82) | $(C_1^{[2,2]})$ | $L_2$ のデータ値を、時間的データ表現(時区間)へと変換する |

表 4 提案方式が有する空間的データ表現への変換記述  
 Table 4 Conversion descriptions of the proposed method for the spatial data representations.

| 名称     | 定義              | 記述内容                            |
|--------|-----------------|---------------------------------|
| (RC83) | $(C_1^{[1,1]})$ | $L_1$ のデータ値を、空間的データ表現(領域)へと変換する |
| (RC84) | $(C_1^{[1,2]})$ | $L_2$ のデータ値を、空間的データ表現(領域)へと変換する |

に変換される(表 1)ここでは、 $L_1, L_2$  のデータ値が、特定の時間的な関連性を満たす場合、静的に記述される。同様に、空間的な関連性記述として、RC41~RC80 のパターン記述を設定した(表 2)。

(MLS2) 提案メタレベルシステム

提案方式に対応する MLS2 は、空間的関連性計量機構  $SCR(1)$ 、時間的関連性計量機構  $SCR(2)$ 、および、同一性計量機構  $SCR(3)$  を有する。MLS2 は、本実験での問合せに要求される関連性の計量機構を有するため、これらの関連性は、対応する計量機構において、直接的に計量される。このため、データ変換機能は  $L_1, L_2$  のデータ表現を対応する計量機構のデータ表現へ変換するシンプルなデータ表現の変換系として実現される。時間的なデータ表現の変換として、 $L_1, L_2$  の有するデータ値は、2 種のデータ変換記述 RC81, RC82 を用いて時間的計量機構にマッピングされる(表 3)。同様に、空間的なデータ表現の変換として、2 種のデータ変換記述 RC83, RC84 を設定した(表 4)。

6.1.2 実験結果

図 11, 図 12 は、空間データを対象とした異種データベース間データ検索・結合機能の記述量(実現オーバーヘッド)を表す。図 13, 図 14 は、時間データ、および、空間データを対象とした異種データベース間データ検索・結合機能の記述量(実現オーバーヘッド)を表す。これらより、MLS2 の実現に要したデータ変換記述の総量(図 11, 図 12, 図 13, 図 14 の縦軸は、記述量の累積を示している)は、MLS1 の実現に要した関連性記述の総量より小さいことが確認された。これは、提案方式は、式 (1) で表されるように、広域ネットワーク環境に連結されたデータベースに含まれる総データ数に一致するデータ変換記述数を要するのに対

表 1 関連性記述方式が有する時間的関連性の変換記述  
Table 1 Conversion descriptions of the method using the relationship conversions for computing the temporal relationships.

| 名称     | 定義   | 関連性   |
|--------|--|---|
| (RC1)  | $(C_{t_1}^{[3,1]}), (C_{t_1}^{[3,2]})$       | $L_1$ の値と $L_2$ の値が等しい (equal)  |
| (RC2)  | $(C_{t_2}^{[3,1]}), (C_{t_2}^{[3,2]})$       | $L_1$ の値と $L_2$ の値が外接する (meet)  |
| (RC3)  | $(C_{t_3}^{[3,1]}), (C_{t_3}^{[3,2]})$       | $L_1$ の値が $L_2$ の値を含む (contain)   |
| (RC4)  | $(C_{t_4}^{[3,1]}), (C_{t_4}^{[3,2]})$       | $L_1$ の値が $L_2$ の値を含み (contain), かつ, $L_1$ の値の 10% 以下の領域と $L_2$ の値が重なる                              |
| (RC5)  | $(C_{t_5}^{[3,1]}), (C_{t_5}^{[3,2]})$       | $L_1$ の値が $L_2$ の値を含み (contain), かつ, $L_1$ の値の 30% 以下の領域と $L_2$ の値が重なる                              |
| (RC6)  | $(C_{t_6}^{[3,1]}), (C_{t_6}^{[3,2]})$       | $L_1$ の値が $L_2$ の値を含み (contain), かつ, $L_1$ の値の 50% 以下の領域と $L_2$ の値が重なる                              |
| (RC7)  | $(C_{t_7}^{[3,1]}), (C_{t_7}^{[3,2]})$       | $L_1$ の値が $L_2$ の値を含み (contain), かつ, $L_1$ の値の 70% 以下の領域と $L_2$ の値が重なる                              |
| (RC8)  | $(C_{t_8}^{[3,1]}), (C_{t_8}^{[3,2]})$       | $L_1$ の値が $L_2$ の値に含まれる (inside)  |
| (RC9)  | $(C_{t_9}^{[3,1]}), (C_{t_9}^{[3,2]})$       | $L_1$ の値が $L_2$ の値に含まれ (inside), かつ, $L_2$ の値の 10% 以下の領域と $L_1$ の値が重なる                              |
| (RC10) | $(C_{t_{10}}^{[3,1]}), (C_{t_{10}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に含まれ (inside), かつ, $L_2$ の値の 30% 以下の領域と $L_1$ の値が重なる                              |
| (RC11) | $(C_{t_{11}}^{[3,1]}), (C_{t_{11}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に含まれ (inside), かつ, $L_2$ の値の 50% 以下の領域と $L_1$ の値が重なる                              |
| (RC12) | $(C_{t_{12}}^{[3,1]}), (C_{t_{12}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に含まれ (inside), かつ, $L_2$ の値の 70% 以下の領域と $L_1$ の値が重なる                              |
| (RC13) | $(C_{t_{13}}^{[3,1]}), (C_{t_{13}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に内接する (cover)   |
| (RC14) | $(C_{t_{14}}^{[3,1]}), (C_{t_{14}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に内接し (cover), かつ, $L_1$ の値の 10% 以下の領域と $L_2$ の値が重なる                               |
| (RC15) | $(C_{t_{15}}^{[3,1]}), (C_{t_{15}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に内接し (cover), かつ, $L_1$ の値の 30% 以下の領域と $L_2$ の値が重なる                               |
| (RC16) | $(C_{t_{16}}^{[3,1]}), (C_{t_{16}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に内接し (cover), かつ, $L_1$ の値の 50% 以下の領域と $L_2$ の値が重なる                               |
| (RC17) | $(C_{t_{17}}^{[3,1]}), (C_{t_{17}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に内接し (cover), かつ, $L_1$ の値の 70% 以下の領域と $L_2$ の値が重なる                               |
| (RC18) | $(C_{t_{18}}^{[3,1]}), (C_{t_{18}}^{[3,2]})$ | $L_2$ の値が $L_1$ の値に内接する (covered by)  |
| (RC19) | $(C_{t_{19}}^{[3,1]}), (C_{t_{19}}^{[3,2]})$ | $L_2$ の値が $L_1$ の値に内接し (covered by), かつ, $L_2$ の値の 10% 以下の領域と $L_1$ の値が重なる                          |
| (RC20) | $(C_{t_{20}}^{[3,1]}), (C_{t_{20}}^{[3,2]})$ | $L_2$ の値が $L_1$ の値に内接し (covered by), かつ, $L_2$ の値の 30% 以下の領域と $L_1$ の値が重なる                          |
| (RC21) | $(C_{t_{21}}^{[3,1]}), (C_{t_{21}}^{[3,2]})$ | $L_2$ の値が $L_1$ の値に内接し (covered by), かつ, $L_2$ の値の 50% 以下の領域と $L_1$ の値が重なる                          |
| (RC22) | $(C_{t_{22}}^{[3,1]}), (C_{t_{22}}^{[3,2]})$ | $L_2$ の値が $L_1$ の値に内接し (covered by), かつ, $L_2$ の値の 70% 以下の領域と $L_1$ の値が重なる                          |
| (RC23) | $(C_{t_{23}}^{[3,1]}), (C_{t_{23}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複する (overlap)   |
| (RC24) | $(C_{t_{24}}^{[3,1]}), (C_{t_{24}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 10% 以下が $L_2$ の値に, $L_2$ の値の 10% 以下が $L_1$ の値に重なる |
| (RC25) | $(C_{t_{25}}^{[3,1]}), (C_{t_{25}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 10% 以下が $L_2$ の値に, $L_2$ の値の 30% 以下が $L_1$ の値に重なる |
| (RC26) | $(C_{t_{26}}^{[3,1]}), (C_{t_{26}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 10% 以下が $L_2$ の値に, $L_2$ の値の 50% 以下が $L_1$ の値に重なる |
| (RC27) | $(C_{t_{27}}^{[3,1]}), (C_{t_{27}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 10% 以下が $L_2$ の値に, $L_2$ の値の 70% 以下が $L_1$ の値に重なる |
| (RC28) | $(C_{t_{28}}^{[3,1]}), (C_{t_{28}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 30% 以下が $L_2$ の値に, $L_2$ の値の 10% 以下が $L_1$ の値に重なる |
| (RC29) | $(C_{t_{29}}^{[3,1]}), (C_{t_{29}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 30% 以下が $L_2$ の値に, $L_2$ の値の 30% 以下が $L_1$ の値に重なる |
| (RC30) | $(C_{t_{30}}^{[3,1]}), (C_{t_{30}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 30% 以下が $L_2$ の値に, $L_2$ の値の 50% 以下が $L_1$ の値に重なる |
| (RC31) | $(C_{t_{31}}^{[3,1]}), (C_{t_{31}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 30% 以下が $L_2$ の値に, $L_2$ の値の 70% 以下が $L_1$ の値に重なる |
| (RC32) | $(C_{t_{32}}^{[3,1]}), (C_{t_{32}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 50% 以下が $L_2$ の値に, $L_2$ の値の 10% 以下が $L_1$ の値に重なる |
| (RC33) | $(C_{t_{33}}^{[3,1]}), (C_{t_{33}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 50% 以下が $L_2$ の値に, $L_2$ の値の 30% 以下が $L_1$ の値に重なる |
| (RC34) | $(C_{t_{34}}^{[3,1]}), (C_{t_{34}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 50% 以下が $L_2$ の値に, $L_2$ の値の 50% 以下が $L_1$ の値に重なる |
| (RC35) | $(C_{t_{35}}^{[3,1]}), (C_{t_{35}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 50% 以下が $L_2$ の値に, $L_2$ の値の 70% 以下が $L_1$ の値に重なる |
| (RC36) | $(C_{t_{36}}^{[3,1]}), (C_{t_{36}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 70% 以下が $L_2$ の値に, $L_2$ の値の 10% 以下が $L_1$ の値に重なる |
| (RC37) | $(C_{t_{37}}^{[3,1]}), (C_{t_{37}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 70% 以下が $L_2$ の値に, $L_2$ の値の 30% 以下が $L_1$ の値に重なる |
| (RC38) | $(C_{t_{38}}^{[3,1]}), (C_{t_{38}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 70% 以下が $L_2$ の値に, $L_2$ の値の 50% 以下が $L_1$ の値に重なる |
| (RC39) | $(C_{t_{39}}^{[3,1]}), (C_{t_{39}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 70% 以下が $L_2$ の値に, $L_2$ の値の 70% 以下が $L_1$ の値に重なる |
| (RC40) | $(C_{t_{40}}^{[3,1]}), (C_{t_{40}}^{[3,2]})$ | $L_1$ の値と $L_2$ の値が離散している (disjoint)  |

し, 関連性記述方式は, 異種データベース間データ検索・結合機能を実現するために, 式 (2) で表される関連性記述数を要するからである. 本実験では, 2 つのデータベース間における各空間的関連性, および, 時間的関連性を満たすデータの組が 1 となるように設定したが, 広域ネットワーク環境には多数のデータベースが連結されているので, 提案方式と関連性記述方式との

変換記述量の差はさらに大きくなる. さらに, ローカルデータベース内のすべてのデータ間に, 対応する時間空間的関連性をすべて設定すること (すなわちすべての関連性記述を設定すること) は困難な作業であるといえる. これらの結果により, 提案方式は, 関連性記述方式と比較してシステム実現容易性が高く, 広域ネットワーク環境に対する適性が高いことが分かる.

表 2 関連性記述方式が有する空間的関連性の変換記述

Table 2 Conversion descriptions of the method using the relationship conversions for computing the spatial relationships.

| 名称     | 定義   | 関連性   |
|--------|--|---|
| (RC41) | $(C_{s_1}^{[3,1]}), (C_{s_2}^{[3,2]})$       | $L_1$ の値と $L_2$ の値が等しい (equal)  |
| (RC42) | $(C_{s_2}^{[3,1]}), (C_{s_3}^{[3,2]})$       | $L_1$ の値と $L_2$ の値が外接する (meet)  |
| (RC43) | $(C_{s_3}^{[3,1]}), (C_{s_3}^{[3,2]})$       | $L_1$ の値が $L_2$ の値を含む (contain)   |
| (RC44) | $(C_{s_4}^{[3,1]}), (C_{s_4}^{[3,2]})$       | $L_1$ の値が $L_2$ の値を含む (contain), かつ, $L_1$ の値の 10% 以下の領域と $L_2$ の値が重なる                              |
| (RC45) | $(C_{s_5}^{[3,1]}), (C_{s_5}^{[3,2]})$       | $L_1$ の値が $L_2$ の値を含む (contain), かつ, $L_1$ の値の 30% 以下の領域と $L_2$ の値が重なる                              |
| (RC46) | $(C_{s_6}^{[3,1]}), (C_{s_6}^{[3,2]})$       | $L_1$ の値が $L_2$ の値を含む (contain), かつ, $L_1$ の値の 50% 以下の領域と $L_2$ の値が重なる                              |
| (RC47) | $(C_{s_7}^{[3,1]}), (C_{s_7}^{[3,2]})$       | $L_1$ の値が $L_2$ の値を含む (contain), かつ, $L_1$ の値の 70% 以下の領域と $L_2$ の値が重なる                              |
| (RC48) | $(C_{s_8}^{[3,1]}), (C_{s_8}^{[3,2]})$       | $L_1$ の値が $L_2$ の値に含まれる (inside)  |
| (RC49) | $(C_{s_9}^{[3,1]}), (C_{s_9}^{[3,2]})$       | $L_1$ の値が $L_2$ の値に含まれ (inside), かつ, $L_2$ の値の 10% 以下の領域と $L_1$ の値が重なる                              |
| (RC50) | $(C_{s_{10}}^{[3,1]}), (C_{s_{10}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に含まれ (inside), かつ, $L_2$ の値の 30% 以下の領域と $L_1$ の値が重なる                              |
| (RC51) | $(C_{s_{11}}^{[3,1]}), (C_{s_{11}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に含まれ (inside), かつ, $L_2$ の値の 50% 以下の領域と $L_1$ の値が重なる                              |
| (RC52) | $(C_{s_{12}}^{[3,1]}), (C_{s_{12}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に含まれ (inside), かつ, $L_2$ の値の 70% 以下の領域と $L_1$ の値が重なる                              |
| (RC53) | $(C_{s_{13}}^{[3,1]}), (C_{s_{13}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に内接する (cover)   |
| (RC54) | $(C_{s_{14}}^{[3,1]}), (C_{s_{14}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に内接し (cover), かつ, $L_1$ の値の 10% 以下の領域と $L_2$ の値が重なる                               |
| (RC55) | $(C_{s_{15}}^{[3,1]}), (C_{s_{15}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に内接し (cover), かつ, $L_1$ の値の 30% 以下の領域と $L_2$ の値が重なる                               |
| (RC56) | $(C_{s_{16}}^{[3,1]}), (C_{s_{16}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に内接し (cover), かつ, $L_1$ の値の 50% 以下の領域と $L_2$ の値が重なる                               |
| (RC57) | $(C_{s_{17}}^{[3,1]}), (C_{s_{17}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に内接し (cover), かつ, $L_1$ の値の 70% 以下の領域と $L_2$ の値が重なる                               |
| (RC58) | $(C_{s_{18}}^{[3,1]}), (C_{s_{18}}^{[3,2]})$ | $L_2$ の値が $L_1$ の値に内接する (covered by)  |
| (RC59) | $(C_{s_{19}}^{[3,1]}), (C_{s_{19}}^{[3,2]})$ | $L_2$ の値が $L_1$ の値に内接し (covered by), かつ, $L_2$ の値の 10% 以下の領域と $L_1$ の値が重なる                          |
| (RC60) | $(C_{s_{20}}^{[3,1]}), (C_{s_{20}}^{[3,2]})$ | $L_2$ の値が $L_1$ の値に内接し (covered by), かつ, $L_2$ の値の 30% 以下の領域と $L_1$ の値が重なる                          |
| (RC61) | $(C_{s_{21}}^{[3,1]}), (C_{s_{21}}^{[3,2]})$ | $L_2$ の値が $L_1$ の値に内接し (covered by), かつ, $L_2$ の値の 50% 以下の領域と $L_1$ の値が重なる                          |
| (RC62) | $(C_{s_{22}}^{[3,1]}), (C_{s_{22}}^{[3,2]})$ | $L_2$ の値が $L_1$ の値に内接し (covered by), かつ, $L_2$ の値の 70% 以下の領域と $L_1$ の値が重なる                          |
| (RC63) | $(C_{s_{23}}^{[3,1]}), (C_{s_{23}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複する (overlap)   |
| (RC64) | $(C_{s_{24}}^{[3,1]}), (C_{s_{24}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 10% 以下が $L_2$ の値に, $L_2$ の値の 10% 以下が $L_1$ の値に重なる |
| (RC65) | $(C_{s_{25}}^{[3,1]}), (C_{s_{25}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 10% 以下が $L_2$ の値に, $L_2$ の値の 30% 以下が $L_1$ の値に重なる |
| (RC66) | $(C_{s_{26}}^{[3,1]}), (C_{s_{26}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 10% 以下が $L_2$ の値に, $L_2$ の値の 50% 以下が $L_1$ の値に重なる |
| (RC67) | $(C_{s_{27}}^{[3,1]}), (C_{s_{27}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 10% 以下が $L_2$ の値に, $L_2$ の値の 70% 以下が $L_1$ の値に重なる |
| (RC68) | $(C_{s_{28}}^{[3,1]}), (C_{s_{28}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 30% 以下が $L_2$ の値に, $L_2$ の値の 10% 以下が $L_1$ の値に重なる |
| (RC69) | $(C_{s_{29}}^{[3,1]}), (C_{s_{29}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 30% 以下が $L_2$ の値に, $L_2$ の値の 30% 以下が $L_1$ の値に重なる |
| (RC70) | $(C_{s_{30}}^{[3,1]}), (C_{s_{30}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 30% 以下が $L_2$ の値に, $L_2$ の値の 50% 以下が $L_1$ の値に重なる |
| (RC71) | $(C_{s_{31}}^{[3,1]}), (C_{s_{31}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 30% 以下が $L_2$ の値に, $L_2$ の値の 70% 以下が $L_1$ の値に重なる |
| (RC72) | $(C_{s_{32}}^{[3,1]}), (C_{s_{32}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 50% 以下が $L_2$ の値に, $L_2$ の値の 10% 以下が $L_1$ の値に重なる |
| (RC73) | $(C_{s_{33}}^{[3,1]}), (C_{s_{33}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 50% 以下が $L_2$ の値に, $L_2$ の値の 30% 以下が $L_1$ の値に重なる |
| (RC74) | $(C_{s_{34}}^{[3,1]}), (C_{s_{34}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 50% 以下が $L_2$ の値に, $L_2$ の値の 50% 以下が $L_1$ の値に重なる |
| (RC75) | $(C_{s_{35}}^{[3,1]}), (C_{s_{35}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 50% 以下が $L_2$ の値に, $L_2$ の値の 70% 以下が $L_1$ の値に重なる |
| (RC76) | $(C_{s_{36}}^{[3,1]}), (C_{s_{36}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 70% 以下が $L_2$ の値に, $L_2$ の値の 10% 以下が $L_1$ の値に重なる |
| (RC77) | $(C_{s_{37}}^{[3,1]}), (C_{s_{37}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 70% 以下が $L_2$ の値に, $L_2$ の値の 30% 以下が $L_1$ の値に重なる |
| (RC78) | $(C_{s_{38}}^{[3,1]}), (C_{s_{38}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 70% 以下が $L_2$ の値に, $L_2$ の値の 50% 以下が $L_1$ の値に重なる |
| (RC79) | $(C_{s_{39}}^{[3,1]}), (C_{s_{39}}^{[3,2]})$ | $L_1$ の値が $L_2$ の値に部分重複し (overlap), かつ, $L_1$ の値の 70% 以下が $L_2$ の値に, $L_2$ の値の 70% 以下が $L_1$ の値に重なる |
| (RC80) | $(C_{s_{40}}^{[3,1]}), (C_{s_{40}}^{[3,2]})$ | $L_1$ の値と $L_2$ の値が離散している (disjoint)  |

## 6.2 実験 2

### 6.2.1 実験手法

提案方式, および, 関連性記述方式が異種データベース間を連結する処理の実行時間を比較し, 提案方式の実用性を検証する.

関連性記述方式, および, 提案方式における問合せ処理の実行時間は, 次のように定式化される.

(1) 関連性記述方式における問合せ処理の実行時間

式 (3) は, 関連性記述方式における 2 つの異種データベース間の連結手続きを表す. ここでは, まず, す

べての変換記述  $SRT$  より, 計量機構特定文脈  $k$ , 関連性識別文脈  $p$ , 関連性確定文脈  $r$  に応じた変換記述を選択する. 次に, ローカルデータベース  $L_1$  のデータ値と, 選択された変換記述を結合し, さらに, ローカルデータベース  $L_2$  のデータ値と結合する.

$$join(L_2, join(L_1, selection(SRT, \\ ==, [k, p, r], ==), ==)) \quad (3)$$

ここで,  $L_1$  のデータ数を  $d_1$ ,  $L_2$  のデータ数を  $d_2$ ,  $SRT$  のデータ数を  $D$  とした場合, データ間の同一性の比較回数の総数  $E$  は, 1)  $SRT$  を対象とした選択

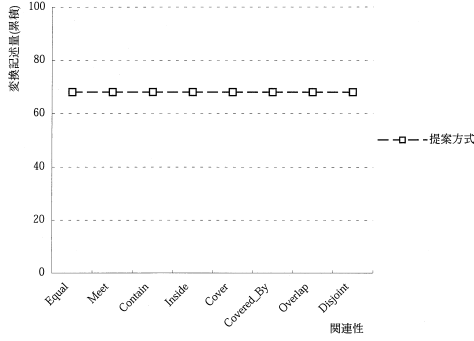


図 11 提案方式において空間的な関連性の計量に要求される変換記述量の推移

Fig. 11 Amount change of required conversion descriptions for computing the spatial relationships by the proposed method.

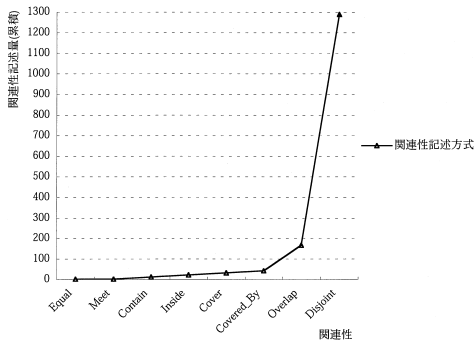


図 12 関連性記述方式において空間的な関連性の計量に要求される関連性記述量の推移

Fig. 12 Amount change of required conversion descriptions for computing the spatial relationships by the method using relationship conversions.

演算における比較回数  $E_{selection}$ 、2) 1の結果と  $L_1$  のデータ値を対象とした結合演算における比較回数  $E_{join_1}$ 、および、3) 2の結果と  $L_2$  のデータ値を対象とした結合演算における比較回数  $E_{join_2}$  の和より得られる(式(4))  $ssf$  は、選択演算  $selection$  における選択率を表し、 $jsf1$  は結合演算  $join_1$  における結合率を表す。

$$\begin{aligned}
 E &= E_{selection} + E_{join_1} + E_{join_2} \\
 &= D + ((ssf \times D) \times d1) + \\
 &\quad ((jsf1 \times (ssf \times D) \times d1) \times d2) \quad (4)
 \end{aligned}$$

これより、関連性記述方式における2つのデータベース間の連結に要する実行時間  $T$  を、式(5)により定式化する。ここで、 $Ppm$  は、1組のデータ記述を対象としたパターンマッチングの単位処理時間を表す。

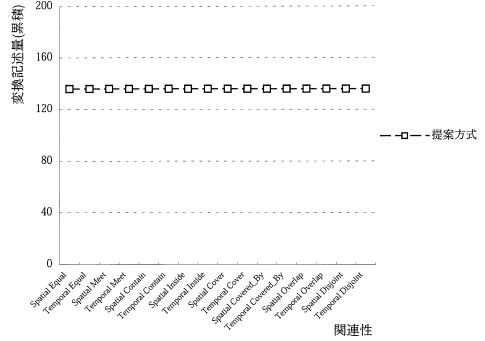


図 13 提案方式において時空間的な関連性の計量に要求される変換記述量の推移

Fig. 13 Amount change of required conversion descriptions for computing the spatial and temporal relationships by the proposed method.

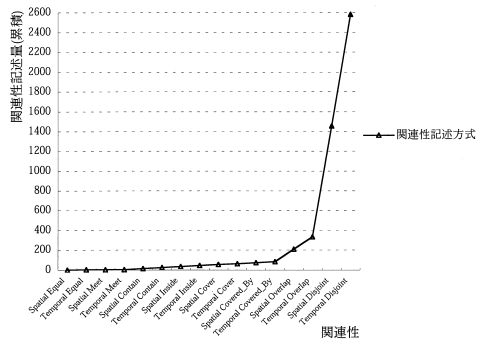


図 14 関連性記述方式において時空間的な関連性の計量に要求される関連性記述量の推移

Fig. 14 Amount change of required conversion descriptions for computing the spatial and temporal relationships by the method using relationship conversions.

$$\begin{aligned}
 T &= Ppm \times (E_{selection} + E_{join_1} + E_{join_2}) \\
 &= Ppm \times (D + ((ssf \times D) \times d1) + \\
 &\quad ((jsf1 \times (ssf \times D) \times d1) \times d2)) \quad (5)
 \end{aligned}$$

(2) 提案方式における問合せ処理の実行時間

式(6)は、提案方式における2つの異種データベース間の連結手続きを表す。ここで、 $f_{[k,p,r]}$  は、計量機構特定文脈  $k$ 、関連性識別文脈  $p$ 、関連性確定文脈  $r$  に応じた関連性計量機能を表す。

$$join(L_1, L_2, f_{[k,p,r]}) \quad (6)$$

提案方式におけるデータ間の関連性の総計量回数  $E$  は、式(7)より求められる。

$$E = d1 \times d2 \quad (7)$$

これより、提案方式における2つのデータベース間の連結に要する実行時間  $T$  を、式(8)によって定式化する。ここで、 $P(f_{[k,p,r]})$  は、文脈  $k, p, r$  に応じた関連性計量機能  $f_{[k,p,r]}$  が、1組のデータ対の計量

に要する単位処理時間を表す．

$$T = P(f_{[k,p,r]}) \times (d1 \times d2) \quad (8)$$

### 6.2.2 実験結果

図 15, 図 16, 図 17 は提案方式, 図 18, 図 19, 図 20 は関連性記述方式における, データベース数の増加にともなう問合せ処理時間の推移を示している．

図 15, 図 18 は, 長方形データを対象とした静的な空間的関連性 *Overlap* の計量をともなう異種データベース間データ結合処理時間を表す．図 16, 図 19 は, 長方形データを対象とした動的な空間的関連性 *Overlap (>, 50%)* の計量をともなう異種データベース間データ結合処理時間を表す．図 17, 図 20 は, 多角形データを対象とした静的な空間的関連性 *Overlap* の計量をともなう異種データベース間データ結合処理時間を表す．表 5 は, 我々が実装したパターンマッチング・プログラム, および, 空間的関連性 *Overlap* を計量するプログラムにおける 1 回の比較に要する実行時間 (実測値) を表す．これらの値を,  $P_{pm}$ ,  $P(overlap, rectangle)$ ,  $P(overlap (>, 50%), rectangle)$ ,  $P(overlap, polygon)$  に使用した．これらの実測値は, 次の実験環境により計測した: Sun Enterprise 450 (CPU: UltraSPARC-II (296 MHz) × 2, Memory: 1 GB, OS: Solaris 2.6). *ssf*, *jsf1* は, それぞれ, 0.143 とした．これは, データベースに格納されている情報が, 空間的な関連性を均等に満たしているという前提を設定し, 関連性識別文脈  $p$  の総数  $q$  を用いて,  $1/q$  より求めた．また, データベース数の変化にともない, *SFR* のサイズ ( $D$ ) を, それぞれ, 2,000,000, 6,000,000, 12,000,000, 20,000,000, 30,000,000, 42,000,000, 56,000,000, 72,000,000, 90,000,000 に変化させた．

実験 1 の結果より, 提案方式が, データベースに含まれるデータ数  $n$  に一致してメタレベルにおける関連性計量機能の探索範囲が拡大する．関連性記述方式では, 探索範囲が  $O(n^2)$  で拡大するので, 図 15, 図 18, 図 16, 図 19 より, 多数のデータベースが連結された広域ネットワーク環境において, 提案方式における 1 回の比較に要する処理時間が小さい関連性計量をともなう異種データベース間データ結合の処理時間は, 関連性記述方式のそれと比較して短くなる．そして, メタレベルシステムに連結されるデータベース数の増加にともない, 関連性記述方式の *SFR* のサイズが劇的に大きくなるので, 提案方式と関連性記述方式との処理時間の差が大きくなる．

一方, 図 17, 図 20 より, 提案方式における 1 回の比較に要する処理時間が大きい関連性計量をともなう

表 5 関連性計量に要求される単位計量時間

Table 5 Estimation time for computing a couple of data values.

| 名称                    | 対象データ | 単位計量時間 (秒) |
|-----------------------|-------|------------|
| $P_{pm}$              | 文字列   | 0.00000788 |
| $P(overlap)$          | 四辺形   | 0.0000309  |
| $P(overlap (>, 50%))$ | 四辺形   | 0.0000491  |
| $P(overlap)$          | 多角形   | 0.0044836  |

異種データベース間データ結合処理時間は, 関連性記述方式と比較してはるかに大きい．しかし, メタレベルシステムに連結されるデータベース数の増加にともない, 関連性記述方式と提案方式の実行時間の差は小さくなり, データベース数が多い領域では, 図 15 ~ 図 19 と同様の傾向を示す．

これより, 異種データベース間データ結合処理時間に関して, 関連性記述方式との比較において, 提案方式を多数のデータベースが連結された広域ネットワーク環境に適用することが有効であることが確認された．

## 7. 関連研究

マルチデータベースシステムを実現する方式は, 異種データベース間のデータ統合の観点において, 次の 2 種に分類される．提案方式は方式 B に位置づけられる．

方式 A (関連性記述方式) メタレベルの共通データ記述を対象としたパターンマッチングによる異種データベース間のデータ統合方式

大局スキーマによる異種データベース間データ統合の実現方式は, データベース群に共通のスキーマを提供し, 共通のスキーマ上に変換されたデータ値を対象にパターンマッチングを適用することで, データベース間統合を実現する<sup>4),13)</sup>．これら大局スキーマのアプローチによるマルチデータベースシステムの実現方式は, RDBMS が有するような単一データベースを対象とした計量機構を異種データベース環境に適用したデータベース統合方式として位置づけられる．

方式 B (提案方式) メタレベルの有する複数の関連性計量機構の統合による異種データベース間データ統合方式

複数の計量機構を用いたデータ統合の実現方式として, ND/SR+<sup>15)</sup>, パターン記述, および, その意味の計量を行う連想検索方式<sup>17)</sup>が提案されている．ND/SR+は, パターンマッチング機構とリージョン・アルジェブラを統合し, RDB のような構造化データ, および, HTML データのよ

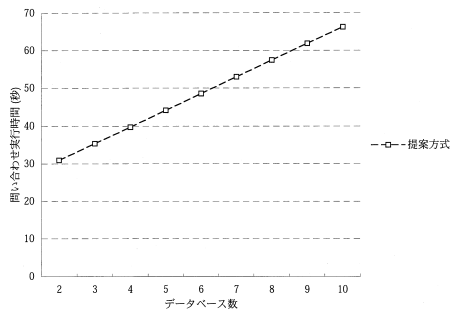


図 15 提案方式において四辺形を対象とした関連性計量機能 overlap の処理に要する実行時間

Fig. 15 Query execution time for computing “overlap” between rectangles by the proposed method.

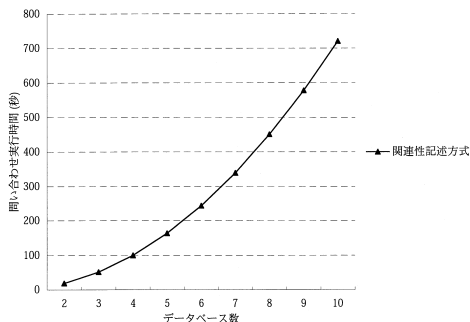


図 18 関連性記述方式において四辺形を対象とした関連性計量機能 overlap の処理に要する実行時間

Fig. 18 Query execution time for computing “overlap” between rectangles by the method using the relationship conversions.

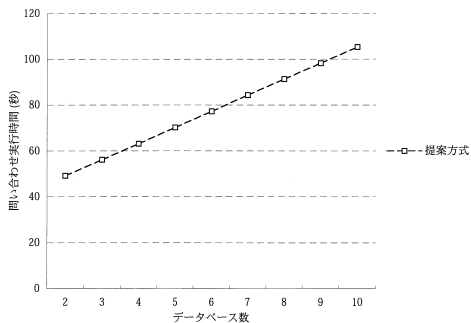


図 16 提案方式において四辺形を対象とした関連性確定文脈をともなう関連性計量機能 overlap の処理に要する実行時間

Fig. 16 Query execution time for computing “overlap” with a ratio context between rectangles by the proposed method.

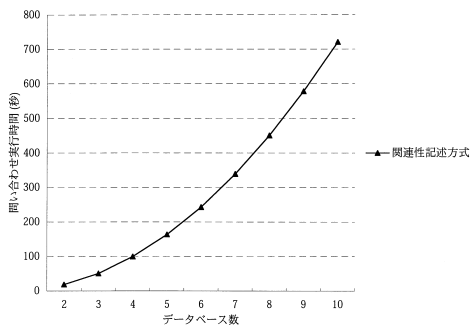


図 19 関連性記述方式において四辺形を対象とした関連性確定文脈をともなう関連性計量機能 overlap の処理に要する実行時間

Fig. 19 Query execution time for computing “overlap” with a ratio context between rectangles by the method with the relationship conversions.

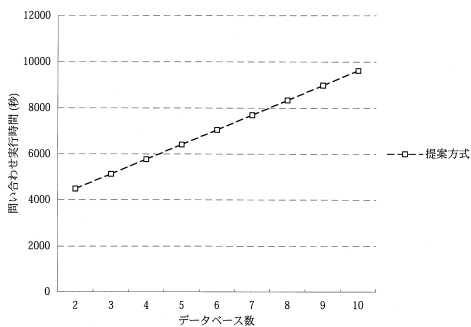


図 17 提案方式において多角形を対象とした関連性計量機能 overlap の処理に要する実行時間

Fig. 17 Query execution time for computing “overlap” between polygons by the proposed method.

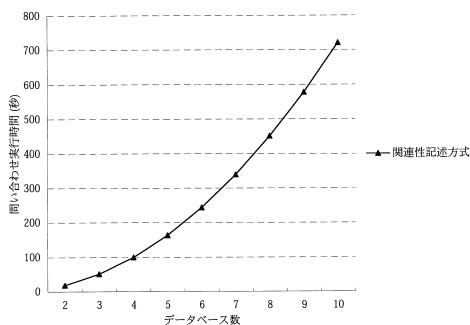


図 20 関連性記述方式において多角形を対象とした関連性計量機能 overlap の処理に要する実行時間

Fig. 20 Query execution time for computing “overlap” between polygons by the method using the relationship conversions.



うな半構造化データの統合に適用する。文献17)の連想検索方式は、パターンマッチング機構と意味的連想機構を統合し、主観的、および、客観的な視点に応じたマルチメディアデータの検索を実現する。空間データベース間統合の実現方式としては、異種GIS群を対象とした共通のメタデータ構造と標準APIの実現方式<sup>16)</sup>が、メタレベルで空間処理を適用する方式として提案されている。提案方式は、静的関連性、および、動的関連性を対象とした関連性計量機能群を、文献1), 5), 10), 11), 14)などのこれまでに提案されてきた多様な計量機能群の一段抽象度の高いメタレベルにおいて定式化することにより、これらの多様な計量機能群の実現方式に依存することなく、それらの多様な計量機能群を統合するための枠組みを実現している。

## 8. 結 論

本論文では、データ検索・結合文脈に応じて確定されるデータ間の時間的・空間的関連性の計量をとまなう異種データベース間時間的・空間的データ検索・結合方式を提案した。本方式の特徴は、データ検索・結合文脈と独立かつ静的に定義される時間的・空間的関連性に加えて、データ検索・結合文脈に応じて動的に定義される時間的・空間的関連性を動的に計量する機能群を、データベースの実現レベルより一段抽象度の高いメタレベルに実現することにより、異種データベース群の検索・結合問合せ数を増加させ、生成される新しい情報の総量を増加させる。さらに、実験により、関連性記述方式との比較における提案方式の実現容易性、ならびに、実用性を示した。

今後の課題としては、時空間的関連性計量システムの最適化、高速化方式、メタレベルシステムにおけるデータマイニング機構、および、時空間的関連性を含む多様な関連性の計量をとまなう異種データベース間のアクティブ化機構などを実現する予定である。

謝辞 本研究にあたって貴重なご助言をいただいた筑波大学電子・情報工学系北川高嗣先生に感謝の意を表します。

## 参 考 文 献

- 1) Allen, J.F.: Maintaining Knowledge about Temporal Intervals, *Comm. ACM*, Vol.26, No.11, pp.832-843 (1983).
- 2) Bouguettaya, A., Benatallah, B. and Elgarmid, A.: An Overview of Multidatabase Sys-

- tems: Past and Present, *Management of Heterogeneous and Autonomous Database Systems*, Elgermid, A., Rukinkiewicz, M. and Sheth, A. (Eds.), pp.1-27, Morgan Kaufmann (1998).
- 3) Bright, M.W., Hurson, A.R. and Pakzad, S.: A Taxonomy and Current Issues in Multidatabase Systems, *Computer*, Vol.25, No.3, pp.50-60 (1992).
- 4) Batini, C., Lenzerini, M. and Navathe, S.B.: A comparative analysis of methodologies for database schema integration, *ACM Computing Surveys*, Vol.18, No.4, pp.324-364 (1986)
- 5) Egenhofer, M.J.: Spatial Relations: Models, Inferences, and their Future Application, *Proc. Advanced Database Symposium*, Tokyo, Japan, December 2-4, separate volume (1996).
- 6) Egenhofer, M.J., Rashid, A. and Shariff, B.M.: Metric Details for Natural-Language Spatial Relations, *ACM Trans. Information Systems*, Vol.16, No.4, pp.295-321 (1998).
- 7) Goñi, A., Mena, E. and Illarramendi, A.: Querying Heterogeneous and Distributed Data Repositories using Ontologies, *Information Modelling and Knowledge Bases IX*, pp.19-34, IOS Press (1998).
- 8) 細川直秀, 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.SIG 8 (TOD4), pp.95-111 (1999).
- 9) Hosokawa, Y. and Kiyoki, Y.: Functional and parallel query processing and query optimization for multidatabase systems, *Proc. 17th IASTED International Conference on Applied Informatics*, Austria, pp.101-106 (1999).
- 10) 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).
- 11) 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 (1995).
- 12) Li, W.S. and Clifton, C.: Semantic Integration in Heterogeneous Database Using Neural Networks, *Proc. 20th VLDB Conference*, pp.1-12 (1994).
- 13) Litwin, W., Mark, L. and Roussopoulos, N.: Interoperability of Multiple Autonomous Databases, *ACM Comp. Surveys*, Vol.22, No.3,

pp.267-293 (1990).

- 14) Masunaga, Y.: A temporal expansion to the multimedia object model in OMEGA, *Proc. DASFAA '95*, pp.430-440 (1995).
- 15) Morishima, A. and Kitagawa, H.: A Data Modeling and Query Processing Schema for Integration of Structured Document Repositories and Relational Databases, *Proc. 5th International Conference on Database Systems for Advanced Applications (DASFAA '97)*, pp.145-154 (1997).
- 16) Open GIS Consortium: The OpenGIS<sup>TM</sup> Guide, Introduction to Interoperable Geoprocessing, Part 1, Available via WWW from <http://www.opengis.org/public/> (1996).
- 17) Yoshida, N., Kiyoki, Y. and Kitagawa, T.: An Associative Search Method Based on Symbolic Filtering and Semantic Ordering for Database Systems, *Proc. 7th IFIP 2.6 Working Conference on Database Semantics (DS-7)*, Switzerland, pp.215-237 (1997).

(平成 13 年 6 月 27 日受付)

(平成 13 年 9 月 14 日採録)

(担当編集委員 有澤 博)



石橋 直樹

1973 年生。1996 年慶應義塾大学総合政策学部卒業。1998 年同大学大学院政策・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開発に従事。ACM 会員。



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・情報工学系に勤務。現在、同学系助手。データベースシステムに関する研究に従事。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 年同大学教授。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, 電子情報通信学会各会員。