

6G-4

文脈による意味決定機能をもつデータベース検索支援システム

人見 洋一, 清木 康, 北川 高嗣  
筑波大学

1 はじめに

データベース・システムは、情報の抽出、共有、および、統合を柔軟かつ効率的に実現するためのシステムとして位置づけることができる。現在までに提案されてきたデータベース・システムにおける情報抽出の基本操作は、パターン・マッチングによる検索であり、異なる表現形態であるが同一の意味をもつデータや近い意味をもつデータの検索を動的に行うことはできない。これには、シソーラスを用いて同義語を照会する方法があるが、その同義語は、シソーラスの設計時に静的に決定され、また、同義であることの定義が明確でないため、ある特定の状況(文脈)において、同義あるいは近いという判定を、検索時に動的に行うことはできない。

我々は、状況(文脈)に応じた動的な言葉の意味解釈を行う機能を実現することにより、高水準なデータベース検索[3]を行うことが可能になるものと考えている。この機能を実現する方式として、意味の数学モデル[1][2]が提案されている。このモデルは、状況(文脈)に応じた動的な意味の解釈を計算によって実現するモデルである。

本稿では、意味の数学モデルの実現方式の検証について述べる。本モデルは、言葉のもつ意味を扱うために、言葉をベクトルによって表現したデータ(以下で述べるデータ行列)を用いる。我々は、本モデルの実現を行い、生成したデータを用いて実験を行った。ここでは、その実現方式および実験の概要を示す。

2 意味の数学モデルの概要

意味の数学モデルの概要を実現するモジュール群の構成を、図1に示す。具体的な定式化については、[1][2]において、述べられている。

**Step-1: 前提** まず、言葉に関するいくつかの特徴づけが、データ行列の形で与えられているものとする。このデータ行列をAとする。

**Step-2: イメージ空間の設定** 行列Aを2ノルムで正規化し、相関行列を作成する。その相関行列を固有値分解し、固有ベクトルを求める。その固有ベクトルの張る空間をイメージ空間と定義する。

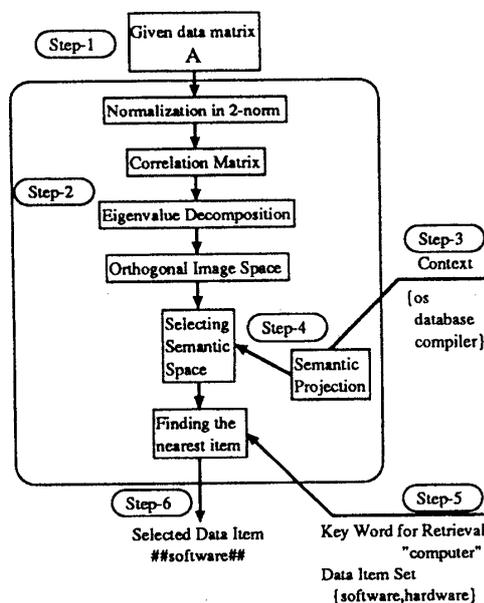


図1: モジュール群の構成

**Step-3: 文脈** 検索語の言葉の意味を決定するための文脈を与える。

**Step-4: 意味空間の設定** 文脈の各言葉をイメージ空間へフーリエ展開し、フーリエ係数を求める。各固有ベクトル毎に、そのフーリエ係数の総和を求める。そして、しきい値εを越えた、その総和に対応する固有ベクトルの張る空間を意味空間と定義する。

**Step-5: 検索語・比較対象語** 検索語および比較対象語を与える。比較対象語とは、検索語が持っている複数の意味を、個々の単語で表したものである。

**Step-6: 距離比較** 意味空間において、検索語に最も近い比較対象語を選択する。その選択された言葉を、その状況(文脈)に応じた検索語の持つ意味とする。

3 実験

本実験では、本モデル実現の検証を行った。本モデルは、言葉の持つ意味を扱うため、言葉をベクトルによって表現したデータ(データ行列)を必要とする。我々は、対象のデータとして、2000の言葉(keywords: 英単語)の各々を、850の特徴(features: 英単語)に関する値によってベクトルで表現したデータ行列を作成した。データ行列の構成を図2に示す。この図では、

A retrieval support system for database  
with function for selecting appropriate meanings  
according to contexts  
Youichi Hitomi, Yasushi Kiyoki, Takashi Kitagawa  
University of Tsukuba

表1 実験結果

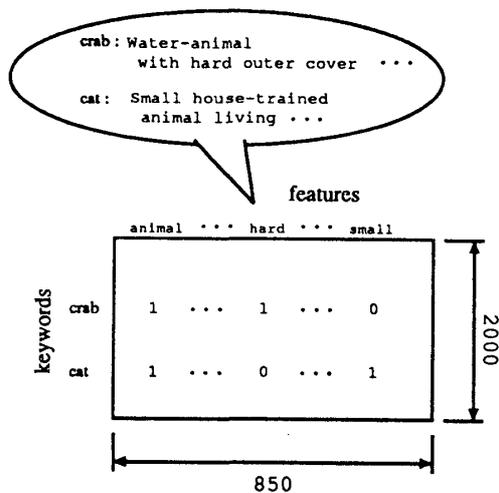


図2: データ行列の作成

'crab' と 'cat' を例として示している。各英単語は、図の上部に示しているように定義されているとする。そのとき、データ行列は、図の下部に示すように設定される。例えば、'crab' は説明に 'water', 'hard' の単語が使われているので、その列の 'crab' の行は、1 が設定される。

このデータ行列を対象とした、実験を行った結果を表1に示す。No.1 から No.4 の実験結果は、意図した回答が得られた結果であり、文脈を変えることにより、選択された意味空間が変化した。そのため、比較対象語群の中から選択された言葉が、文脈に応じて変化した。No.5 の実験では、意図した回答が得られなかったが、選択された意味空間は、変化した。これは、No.5 の実験で使用したデータ項目 (data items) のイメージ空間における相対的な位置が正しくないためと考えられる。

この実験結果から、本方式により状況 (文脈) に応じた言葉の意味の動的な解釈を行う計算が可能となり、高水準なデータベース検索の実現に向けての指針を得ることができた。

#### 4 まとめ

本稿では、意味の数学モデルの実現方式の検証を行うための実験を行い、その実験により本モデルの実現方式の正当性を示した。現在の段階では、データ行列の生成において、データ行列を1と0の値によって定義している。今後は、各言葉の定義文における各特徴 (features) の重要度を考慮したデータ行列の要素の値づけが必要であると考えられる。また、データ行列の生成

No.	Key Word for Retrieval	Context	Data Item Set		Selected Data Item
			Distance between Key-Word and Data-Items		
1	band		belt	party	
		band put round middle of body to keep clothing in place	1.664320	1.742350	belt
		group of persons journeying, working, together	2.055697	1.819425	party
2	case		instance	box	
		special example of some general fact, law	0.796605	1.026701	instance
		for solids, four sided and with cover	1.027877	1.022699	box
3	race		people	competition	
		person	0.556714	0.570695	people
		make attempt to do or be better than other	1.200716	1.041146	competition
4	link		bond	torch	
		force uniting or keeping fixed	1.141210	1.149725	bond
		bit of burning wood for giving light	1.893141	1.845212	torch
5	rent		slit	borrow	
		long narrow opening	0.981427	0.848490	borrow
		get the use of (money, thing) for a time	0.772784	0.596107	borrow

の段階において、言葉をより正確に表現するために、データ検索の結果を評価し、それをフィードバックすることにより、データ行列を自動修正する方法が有効であると考えられる。

#### 参考文献

- [1] 北川 高嗣, 清木 康, “意味の数学モデルとデータベースへの応用可能性について,” 電子情報通信学会データ工学研究会 DE92-6, pp.43-50, May 1992.
- [2] T. Kitagawa and Y. Kiyoki, “The mathematical model of meaning and its application to multi-database systems,” to appear in Proceedings of 3rd IEEE Int. Workshop on Interoperability in Multidatabase Systems, Apr. 1993.
- [3] H. Shimizu, Y. Kiyoki, A. Sekijima and N. Kamibayashi, “A decision making support system for selecting appropriate online databases,” to appear in Proceedings of 3rd IEEE Int. Workshop on Interoperability in Multidatabase Systems, pp.322-329, Apr. 1991