

# 個人の興味, 関心, 嗜好を反映可能なデータベース検索手法の開発

三石 大

東北大学大学院教育情報学研究部

**要旨:** インターネット上で提供される講義資料や講義内容のアーカイブ化が進むと, 目的の資料や授業を探すための検索が必要となる。しかしながら, このような複数のデータからなる複合データは, その内容を単純な語句により説明することは極めて困難であり, 既存のリレーショナルデータベース等の技術をそのまま利用しては, 目的のデータを発見することは容易ではない。本稿では, このような複雑な特徴を有するデータからなるデータベースの効果的な検索のために, 我々が開発を進めている, 媒介変数を利用することにより個人の興味, 関心, 嗜好の分析をおこない, これをデータベースの検索結果に反映させることで, 利用者が目的とするデータの発見を支援する手法について述べ, 提案手法に基づくデータベースアプリケーションの構築と, これを利用した評価実験の結果を紹介する。

**キーワード:** マルチメディアデータベースアプリケーション, データマイニング, 感性検索, 嗜好分析, 媒介変数

## 1 はじめに

2002年度における東北大学インターネットスクール (ISTU) の発足や, マサチューセッツ工科大学の Open Courseware による講義資料の公開など, 近年のコンピュータ技術, ネットワーク技術の発展に伴い, 現在, 大学での講義内容を電子化し, これをネットワークを介して提供する試みが盛んになりつつある。またこのほかにも, コンピュータネットワークが一般の利用者にも普及するにつれ, このコンピュータネットワーク上では, 文書データやマルチメディアデータなどの様々な電子化情報を提供する, 多様なデータベースアプリケーションが多数構築されている。これらのデータベースは, web ブラウザ等を介してアクセス可能な形で提供されており, 一般の利用者が容易にデータを取得し, 利用可能となっている。

しかしながら, この様な web 上のデータベースアプリケーションにより, 個人が大量のデータを利用できるようになった反面, 一般の利用者が, 大量に提供されるデータの中から目的とするデータを取得することは容易ではない。特に, 講義データのように複数のデータからなる複合的なデータや映画や音楽などの複雑な内容を持つデータは, その特徴を複数の明示的な語句の単純な組み合わせにより表現す

ることが難しく, その結果, 既存のリレーショナルデータベースや全文検索システムを利用し, 目的とするデータを特定することは極めて困難である。

そこで著者らは, このような複雑な特徴を持つデータを効果的に検索し, 利用者が目的とするデータを効率的に発見できるように, 利用者の興味や関心, 嗜好などの感性を反映可能な検索手法の開発を行なっている。本稿では, 2002年12月に開催された情報処理学会データベースと Web 情報システム: DBWeb 2002 で発表を行なった内容を中心に, 我々が開発を進めている検索手法の1つである個人の嗜好を反映可能な検索手法を紹介する。これは, 媒介変数を利用することにより個人の嗜好分析をおこない, これをデータベースの検索結果に反映させることで, 利用者が目的とするデータの発見を支援するものである。ここでは, 我々の提案する手法の詳細と併せて, web 上のデータベースアプリケーションの1つとして料理のレシピ検索システムを例にとり, 提案手法に基づく個人の嗜好を反映可能なオンラインレシピ検索システムの設計, 実装, およびこれを利用した評価実験による有効性評価について述べる。

本稿は5章から構成される。2章ではまず, 大量のデータから効率的に目的のデータを検索するための

既存の検索手法の問題点を指摘し、我々の提案する媒介1変数を用いた嗜好分析による個人の嗜好を反映可能な検索手法について述べる。次に、3章では、2章で提案した手法の応用として、個人の嗜好を検索結果に反映可能なレシピ検索システムの設計、実装を行う。4章では、実装したシステムによる評価実験を行い、提案手法の有効性を評価する。最後に、5章で本稿のまとめを行う。

## 2 個人の嗜好を反映可能な検索

本章では、個人の嗜好を反映可能な既存の検索手法を挙げ、その問題点を指摘するとともに、我々の提案する媒介変数を用いた嗜好分析による検索手法について述べる。

### 2.1 個人の嗜好を反映可能な既存検索手法

大量のデータベースから目的のデータを効率的に取得するための手法として、データの利用目的や個人の嗜好、感性に基づき、これを検索結果に反映させる検索手法は多い。

一般に感性検索と呼ばれる検索手法では、予め個々のデータの特徴を何らかの感性語句により表現し、これを検索のためのインデックスとして提供する[1][2][8][9][12]。これにより、目的とするデータに対して利用者が抱く印象に基づきデータを検索することが可能となる。これらの感性検索では、予め特徴を示す語句となり得る感性語句を特定し、また、アンケート調査やデータの周波数解析など、何らかの方法でデータの特徴を分析し、想定された感性語句との対応付けを行う必要がある。

Webアプリケーションとして提供されるデータベースアプリケーションの場合、その利用の容易性により各利用者の利用履歴が得やすいという利点がある。そこで、利用者のデータベースの利用履歴から個人の嗜好を推測し、これをデータのクラスタリングや検索結果の提示に利用する手法がある[5][6][7]。これにより、検索結果の提示における各個人の嗜好を反映した絞り込みや順序付け、関連性の高いデータの提示が可能となり、利用者は効率的に目的とするデータを取得することができる。

また我々は、これまで、複雑で曖昧な特徴を持つマルチメディアデータベースに対し、利用履歴をもとに検索のためのインデックスを半自動生成する手

法を提案し、これによる検索システムの設計、実装を行ってきた[3][4][10][13]。これは、データベースの利用履歴から個人の嗜好を分析し、これをもとに、予め想定した感性語句等に関するデータの特徴を推測し、インデックスを生成するものである。これにより、データベース構築時には判っていなかったような、個々のデータの潜在的な特徴を推測し、これを検索に利用することが可能となり、複雑で曖昧な特徴を持つデータの中から目的とするデータの発見が容易となる。

しかし、これら利用履歴を利用する検索手法でも、データの特徴を示す語句となり得る語句、例えば印象を示す感性語句やデータに含まれている語句などを特定し、検索対象と関連付ける必要がある。

これに対し、明示的な語句により特徴を示すことが困難なデータも多い。例えば、料理に関するデータベースでは、検索対象となるデータである個々の料理の特徴として、その料理に使用する“食材”、“調理法”、さらに出来上がった料理の“味”などを考えることが出来る。このとき、食材や調理法といった特徴は、明示的な語句の組合せにより表現可能であるが、味を示す語句の意味は、個人や文化によってその使用方法に大きく差がある上、料理そのものや、用いる食材によっても異なることが多い。その結果、一般に味を示す語句を組み合わせても、料理の味に関する特徴を一意に示すことは困難である。

このような料理の味を表す語句のように検索対象の特徴を示す語句の意味が複雑であったり、またその意味が曖昧でデータにより相違が大きく、一意に定めることが困難なデータベースに対し、単純にこれらの語句を用いてデータの特徴を表現し、既存の検索手法を適用すると、その複雑性、曖昧性のために、利用者が意図しない検索結果を得る可能性が生じてしまい、そのままでは効率的な検索を行うことは困難である。

### 2.2 媒介変数を用いた嗜好分析による個人の嗜好を反映可能な検索

我々はこれまで、明示的な語句による特徴付けが困難なデータベースの効率的な検索のために、データベースの利用履歴から媒介変数による嗜好分析を行い、2個人の嗜好を検索結果に反映する手法を提案してきた[14][15]。これは、検索対象の特徴を明

示的な語句により表現することが困難な場合でも、検索対象の持つ特徴との関係が深く、かつその種別を何らかの特徴によりほぼ一意に特定可能な別のデータが存在する場合には、そのデータにより検索対象を分類し、これを媒介変数として個人の嗜好分析を行い、その分析結果を検索結果の評価関数として利用するものである。

例えば、料理を検索対象とするデータベース検索の場合、個人の味に関する嗜好を検索結果に反映できれば便利であるが、この場合、ワインの種別を媒介変数として利用することができる。ワインは、地域、気候、製法などによって渋味、ボリューム感、甘味、酸味等が異なり、様々な種別に分類することができる。また、食事とともに日常的に飲まれる酒であるため、料理の味との関係が深く、対応する料理の幅も広い上、各料理と相性の良いワインの種別が広く一般に知られている[11]。そのため、料理データベースの利用履歴から、ワインの種別を媒介変数として各利用者の嗜好を分析し、検索結果の提示に利用することができる。

簡単な例として、料理データベースの検索結果として得られた料理の中から、魚料理を頻繁に選択した利用者がいたとする。魚料理は白ワインとの相性が良いため、この利用者は白ワインと相性が良い料理が好みであると推測することができる。この時、他の豚肉料理や牛肉料理が赤ワインとの相性が良いのに対し、鶏肉料理は白ワインとの相性が良いことが判っている。そのため、魚料理が好みであると推測される利用者は、豚肉料理や牛肉料理と比較し、鶏肉料理の方が好みであろうと推測することができる。その結果、何らかの検索の結果として何種類かの肉料理が得られた際に、その利用者には鶏肉料理を優先的に提示することができる。

この様に、媒介変数を用いて嗜好分析を行い、これに基づき検索結果を提示することにより、明示的な語句による特徴付けが困難なデータベースの検索において個人の嗜好を検索結果に反映することが可能となり、その結果、大量に得られた検索結果の中から目的のデータを効率的に選択し、取得することができる。

## 2.3 媒介変数を用いた検索手順

我々の提案する、媒介変数を用いた嗜好分析による検索手法の手順を以下に示す。

(1) 準備: 先ずはじめに、予め検索対象と媒介変数との関係を定義する。すなわち、検索対象の集合を  $T$ 、媒介変数の集合を  $M$  としたとき、検索対象と媒介変数との関係に基づき、以下のように、任意の検索対象  $\forall t \in T$  が与えられると、この検索対象に対応した媒介変数  $\exists m \in M$  を出力する写像関数  $R(t)$  を定義する。

$$R: \{t \in T\} \rightarrow \{m \in M\}$$

また、各利用者ごとに、その嗜好の初期値を定義する。ここでは、各利用者の嗜好を媒介変数に対する得点からなるベクトル空間により表現することとする。すなわち、 $n$  個の媒介変数  $m_1, m_2, \dots, m_n$  があつた場合、利用者  $u$  の媒介変数  $m_i$  に対する得点を  $s_{miu}$  とし、ある利用者  $u$  の嗜好  $\vec{p}_u$  を以下のように定義し、その初期値として、各値を 0 とする。

$$\vec{P}_u = \langle s_{m1u}, s_{m2u}, \dots, s_{mnu} \rangle$$

(2) 嗜好分析: 利用者のデータベース利用履歴に基づき、各利用者の嗜好分析を行う。すなわち、ある利用者  $u$  がなんらかの方法でデータベースを検索し、得られた検索結果の中からあるデータ  $t$  を選択した際に、これに基づき、利用者  $u$  の嗜好  $\vec{p}_u$  の値を更新する。嗜好  $\vec{p}_u$  の更新方法としては多数考えることができるが、例えば、最も簡単な方法として、利用者  $u$  のデータ  $t$  に対応する媒介変数  $m = R(t)$  に対する得点  $s_{R(t)u}$  への加算とし、以下のように定義することができる。

$$s_{R(t)u} + 1 \rightarrow s_{R(t)u}$$

(3) 検索結果提示への嗜好反映: データベースの検索結果に対し、上記 (2) で得られた嗜好を反映させ、利用者  $u$  に提示する。ある利用者  $u$  が、なんらかの方法でデータベースを検索し、その結果として  $k$  個のデータ  $\{t_i \mid t_i \in T, i = 1, 2, \dots, k\}$  が得られたとする。このとき、利用者  $u$  の嗜好  $\vec{p}_u$  に基づき、各データ  $t$

$1, t_2, \dots, t_k$  に対する得点はそれぞれ以下のように求めることができる。

$$t_1:SR(1)u, t_2:SR(2)u, \dots, t_k:SR(t_k)u$$

この得点に基づき検索結果の並べ替えを行い、利用者に提示する。その結果、推測された利用者の嗜好に基づき、その利用者が欲するであろうと思われるデータを優先的に提示することができる。

### 3 個人の嗜好を反映可能なオンラインレシピ検索システム

本章では、提案手法の一応用である、個人の嗜好を反映可能なオンラインレシピ検索システムの web アプリケーションとしての設計、実装について述べる。

#### 3.1 オンラインレシピ検索システムにおける個人の嗜好反映可能な検索手法

本システムは、食材を検索キーとして料理の検索を行い、検索結果として料理名のリストを出力する。このとき、本稿で提案する手法に基づき、検索結果として得られる料理名リストを利用者の好みに応じて並べ替えを行い、提示する。利用者が得られた料理名リストの中から料理を選択すると、そのレシピと調理例の写真を参照することができる。

本システムでは、利用者の好みに応じた順序で検索結果を提示するために、検索対象  $T$  である料理に対する媒介変数  $M$  としてワインの種別を利用する。そのためにまず、予め、赤、白、ロゼ、スパークリングの各ワインの種別を渋みやボリューム感等の特徴に基づき分類する。例えば赤ワインは、渋味およびボリューム感に応じて 16 種類に大別することができる (図 1)。今回は、文献 [11] に基づき、表 1 に示すように、ワインの種別を計 44 種別に分類した。

また、同様に文献 [11] に基づき、各料理からワインの種別への写像関数  $R(t)$  を定義する。例えば、図 1 中の赤ワイン  $A$  と相性の良い料理として、“牛肉のたたき”、“こんにゃくの芥子炒め”、“麻婆豆腐”、“まぐろの造り”、“焼きなす”などを挙げることができる。そこで、これらの料理を引数とした場合に、“赤ワイン  $A$ ”を出力するよう、写像関数  $R(t)$  を定義する。これにより、ワインの種別を料理の味に関す

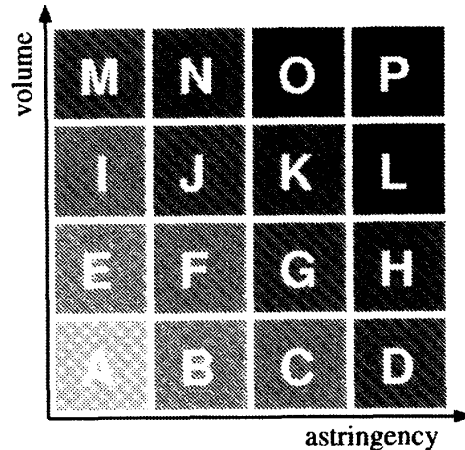


図 1: 赤ワインの種別の分類

表 1: ワインの種別数

赤ワイン	16 種別
白ワイン	18 種別
ロゼワイン	6 種別
スパークリングワイン	4 種別
計	44 種別

る嗜好分析のための媒介変数として利用可能となる。

次に、ある利用者  $u$  が、検索結果として得られた料理名のリストの中からレシピを参照するために特定の料理  $t$  を選択した際に、その利用者毎に、選択された料理が対応するワインの種別  $m = R(t)$  に対する得点  $SR(t)u$  にポイントを加算する。これを繰り返すことにより、個々の利用者の嗜好の違いにより、ワインの種別毎の得点に偏りが生じる事が予想される。その結果、得点が高いワインの種別に適合する料理が、その利用者が好む料理であると推測することができる。従って、この得点に応じて、検索結果の料理名リストを並べ替えることにより、利用者の好みの順に料理名リストを提示することが可能となる。

例えば、ある利用者が本システムの利用を繰り返した結果、図 2 に示すような形でポイントが加算されたとする。ここでは、G の得点が最も高く、次いで F, H... の順に高い。また、A, I, M, N, P には得点がない。その結果、この利用者は、ボリューム感が控え目で渋みがほどほどある赤ワインの種別である G と相性の良い料理、例えば“牛肉の醤油焼き”や“ハンバーグ”などの料理が好みであると推測することができる。これら G に対応する料理を最も上位

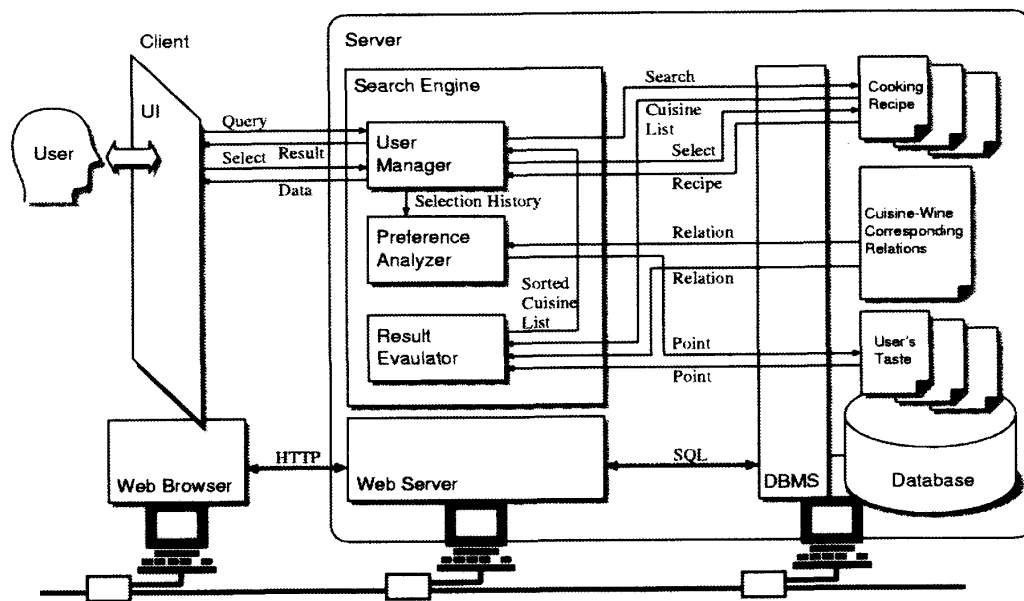


図 3: システムアーキテクチャ

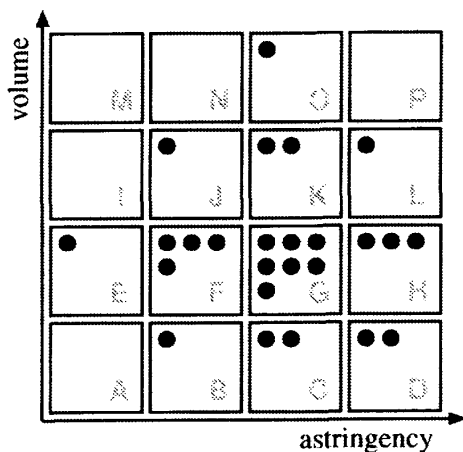


図 2: 赤ワインの種別に対するポイントの加算例

に提示し、次いで F, H…に対応する料理を提示し、A や I 等の種別に対応する料理は最下位に提示することができる。

### 3.2 システムの設計と実装

個人の嗜好を反映可能なオンラインレシピ検索システムを実装するために、今回、そのシステムアーキテクチャを図 3 に示す形の web アプリケーションとして設計した。

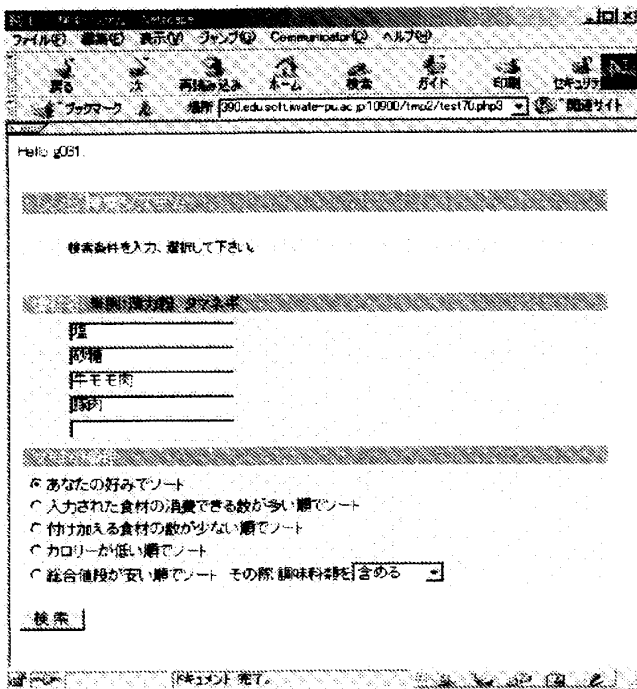
本システムは、レシピ検索機能を提供するサーバと、これを利用するためのユーザインタフェースを提供するクライアントから構成される。サーバは、web サーバ上に構築されレシピ検索機能を提供する

アプリケーションサーバと、これに接続された DBMS 上に構築されるデータベースからなる。

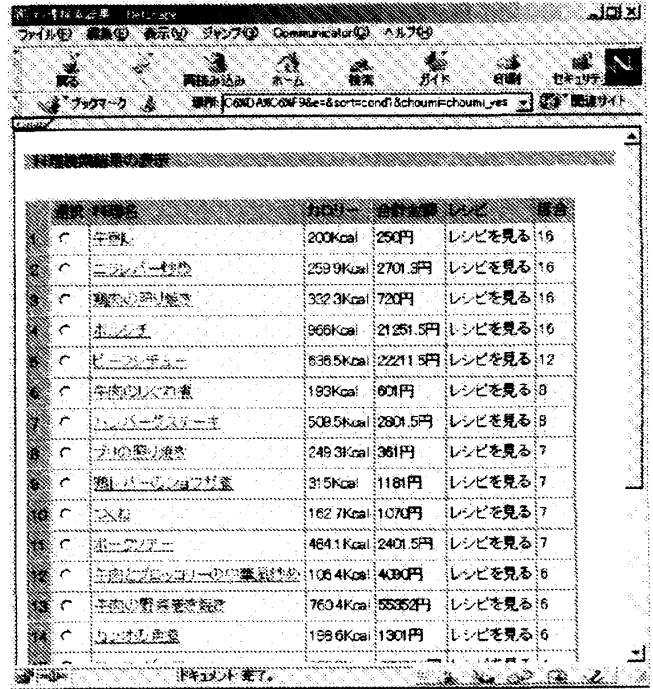
サーチエンジンは、ユーザ管理機構、ユーザ嗜好分析機構、検索結果評価機構の 3 つの機構からなる。ユーザ管理機構は、クライアントを介して利用者とのインタラクションを行い、各利用者を特定するためのユーザ認証や、検索要求に応じたレシピデータベースの検索、およびその結果の提示を行う。ユーザ嗜好分析機構は、利用者の利用履歴から、各利用者毎の嗜好を分析、記録する。検索結果評価機構は、分析された利用者の嗜好に基づき、検索結果として得られた料理名リストの並べ替えを行う。

またデータベースには、具体的な料理名と必要な食材など、各料理のレシピに関するデータに加え、検索対象から媒介変数への写像関数としての料理とワインの種別との対応関係、ワインの種別に対する得点により表現された各利用者毎の嗜好を格納し、これを DBMS を介して提供する。

まず、利用者が食材を検索キーとしたレシピの検索要求を行うと、ユーザ管理機構により該当する料理名の検索を行う。この結果得られた料理名リストは、利用者の嗜好に基づき検索結果評価機構により並べ替えられ、利用者に提示される。次に、利用者が提示された料理名から任意の料理を選択すると、そのレシピが検索、提示されるとともに、選択履歴がユーザ嗜好分析機構に入力され、利用者毎の嗜好



(a): 食材をキーとする検索



(b): 検索結果の提示

図 4: レシピ検索システムの実行例

表 2: レシピ検索システムの実装環境

Operating System:	Solaris 2.7
DBMS:	PostgreSQL 7.0.2
Web サーバ:	Apache 1.3.12
記述言語:	PHP 3.0.15-i18n-ja

を更新する。

以上の設計に基づき、今回、オンラインレシピ検索システムを UNIX システム上に実装し (表 2)、レシピデータとして 230 件の料理を登録した。実装されたシステムの実行例を図 4 に示す。利用者が本システムを繰り返し利用した後、web ブラウザから食材名を入力し、検索要求を行った結果、その利用者が過去に選択した料理に対応するワインの種別に加算された得点に応じた順序で、料理の一覧が提示されていることが確認された。

#### 4 評価実験

我々の提案する、媒介変数を用いた嗜好分析による個人の嗜好を反映可能な検索手法の有効性確認のために、実装したシステムを用いて実施した評価実験とその評価を述べる。

#### 4.1 実験内容

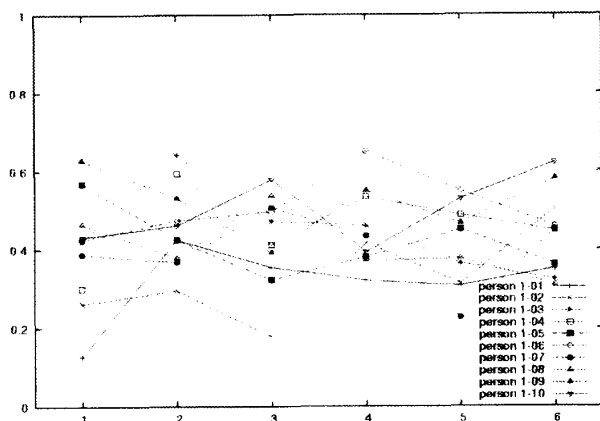
実験内容は、提案手法に基づく検索機能を組み込まない、すなわち個人の嗜好を反映させない検索手法によるシステムと、提案手法に基づく検索機能を組み込み、個々の利用者の嗜好を反映させるシステムの 2 種類を用意し、それぞれ数名の利用者により繰り返しシステムを利用してもらうことにより、その利用者の好み検索結果に反映されたかどうかを比較するものである。

具体的には、被験者を無作為に 2 つのグループに分け、それぞれ、既存手法によるシステム、もしくは提案手法によるシステムを利用し、1 日 2 食、1 食あたり 1 品目の料理を 6 週間分、計 84 食分の献立を作成してもらう。このとき、献立として選択された料理がその利用者の好みに基づき選択された料理であるとし、対応するワインの種別への得点を加算するとともに、検索結果として得られた料理リスト中のどの位置に提示された料理が選択されたかを記録する。

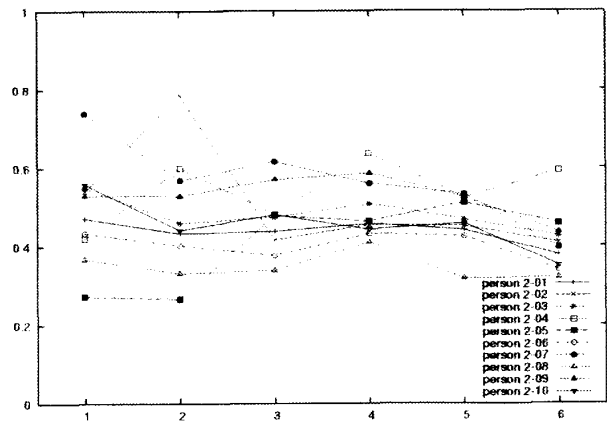
(a): グループ A (提案手法無し)

(b): グループ B (提案手法有り)

図 5: 分散値の変化



(a): グループ A(提案手法無し)



(b): グループ B(提案手法有り)

図 5: 分散値の変化

次に、利用者  $u$  の第  $i$  回目の検索結果として得られた料理リストの大きさを  $n_{ui}$ 、この時の選択位置を  $p_{ui}$  としたとき、各利用者ごとに、1 週間、すなわち 14 食分を 1 セットとして、以下のように、第  $w$  週の 0 を基準とする選択位置の相対値 ( $p_{uw} = n_{ui}$ ) の分散を

$$v_{uw} = \frac{1}{14} \sum_{i=14(w-1)+1}^{14w} \left\{ \frac{p_{ui}}{n_{ui}} \right\}^2$$

を求める。

提案手法に基づき個人の嗜好を反映できていた場合、データベースの利用を繰り返すうちに、利用者の好みと思われる料理が検索結果のリストの上位に提示されることとなるため、利用者はリストの上位から料理を選択することが多くなることが予想され、その結果、上記で求める分散値が減少していくことが予想される。逆に、好みを反映していない場合、選択位置は不特定となるため、分散値は利用回数によらず不確定となる。すなわち、提案手法に基づかないシステムを利用したグループにおいて分散値が不確定となり、また、提案手法に基づくシステムを利用したグループにおいて分散値の減少が確認できれば、提案手法による効果があると判断することができる。

#### 4.2 実験結果と評価

今回、学部学生から 20 名の被験者を募り、これを無作為に 10 名ずつの 2 グループに分け、実験を行った。実験結果として、提案手法に基づく検索機能を

表 3: 分散値の増減人数

グループ	増加	減少
A (提案手法なし)	6 名	4 名
B (提案手法有り)	2 名	8 名

組み込んでいないシステムを利用したグループ (グループ A) の各利用者毎の分散値の変化を図 5 (a) に、また、提案手法に基づく検索機能を組み込んだシステムを利用したグループ (グループ B) の各利用者毎の分散値の変化を図 5 (b) に示す。

実験結果をみると、各個人によって分散値に大きく差異があるが、グループ A では、各週で分散値が上下し、利用回数によらずに不確定な値となっている。これに対し、グループ B では、利用回数を重ねるごとに分散値が減少する傾向にあることが読み取れる。また、第 1 週と第 6 週とでの分散値の増減を比較しても、グループ A に比べ、グループ B では分散値が下がった人数が多い (表 3)。

また評価実験終了後、被験者に、本システムを利用して求める料理が得られたかどうかという質問に対し、1. 「ほとんど得られなかった」から、4. 「十分得られた」までの 4 段階評価で回答してもらう簡単なアンケートを行なった。アンケート結果を表 4 に示す。このアンケート結果だけでは定量的な評価を行なうには十分ではないが、この結果においても、提案手法に基づかないシステムを利用したグループ

表 4: システムの利用に関するアンケート結果

グループ	回答 1	回答 2	回答 3	回答 4	無回答
A	0名	4名	4名	1名	1名
B	0名	2名	5名	1名	2名

回答 1: 求める料理をほとんど得られなかった

回答 2: 求める料理をあまり得られなかった

回答 3: 求める料理を概ね得られた

回答 4: 求める料理を十分得られた

に比較して提案手法に基づくシステムを利用したグループからの方が、若干ではあるが良い評価を得ている。

以上の結果から、料理を検索対象とし、提案手法に基づきワインを媒介変数として利用したオンラインレシピ検索システムにおいて、利用回数を重ねることにより、利用者の嗜好を反映した検索結果を得ることが出来ていることが判る。すなわち、本稿で提案する、媒介変数を用いた嗜好分析による個人の嗜好を反映可能な検索手法により、効率的なデータベース検索を行えるといえる。

## 5 おわりに

本稿では、ネットワーク上のデータベースアプリケーション、とりわけ、明示的な語句による特定が困難な特徴を持つデータからなるデータベースの効果的な利用のために、我々の提案する媒介変数を用いた嗜好分析による、個人の嗜好を反映可能なデータベース検索手法を紹介した。これは、検索対象の特徴と関連の深い他のデータを媒介変数として各利用者のデータベース利用履歴に基づき個人の嗜好を分析し、検索結果の提示に利用するものである。これにより、データベースの利用を繰り返すことで個人の嗜好を検索結果に反映することが可能となり、目的とするデータを効率的に発見することが可能となる。

また今回、料理のレシピデータベースを例にとり、提案手法に基づくオンラインレシピ検索システムを設計、実装した。本システムでは検索対象を料理のレシピとし、料理の味とワインの種別との関係に基づき媒介変数としてワインの種別を利用することで、個人の味に関する嗜好を検索結果に反映させるものである。

実装されたシステムにより評価実験を行った結果、本システムの利用を繰り返すことにより、検索結果の提示に個人の嗜好を反映できていることを確認した。すなわち、本稿で提案する検索手法により、検索結果の提示に利用者の感性や嗜好を反映し、得られた大量のデータの中から目的とするデータを効率的に選択、取得することが可能となり、特徴を示す語句の意味が複雑で曖昧であるなど、明示的な語句による特徴付けが難しいデータからなるデータベースの効果的な利用が可能となるといえる。

しかしながら、我々の提案する検索手法では、検索対象から媒介変数への写像関数を一意に定義する必要はあるが、検索対象と媒介変数の間に、必ずしもこのような関係を定義できるとは限らない。例えば、ある料理は、2種類以上のワインとの相性が良いことが知られている。また逆に、ワインは、和食との相性があまり良くないことも知られている。そのため、より広範囲な検索対象に対応するためには、他の媒介変数、例えば和食に対する日本酒などを加えるなど、複数の媒介変数を組み合わせた写像関数等を検討する必要がある。

また、本提案手法では、データベースの利用履歴を利用して、検索対象となるデータに対する個人の嗜好を分析するため、データベースの利用開始時には個人の嗜好を検索結果に反映することができない。この問題を解決するためには、例えば、データベースの利用開始前に、媒介変数の種別に基づき検索対象に関する簡単な質問を行なうなど、予め利用者の大まかな嗜好を推定できる方法などが必要となる。

今後、これらの問題を解決し、また、より効果的な検索手法を検討するとともに、ISTUにおける講義の検索等への応用を検討していく予定である。

## 参考文献

- [1] Fukuda, M., Sugita, K. and Shibata, Y.: Per-ceptional Retrieving Method for Distributed De-sign Image Database System, *Trans. IPSJapan*, Vol.39, No.2, pp.158-169(1998).
- [2] Ishihara, S., Ishihara, K. and Nagamachi, M.: Analysis of Individual Differences in Kan-sei Evaluation Data Based on Cluster Analysis, *KANSEI Engineering International*, Vol.1, No.1, pp. 49-58 (1999).



- [ 3 ] Mitsuishi, T., Sasaki, J. and Funyu, Y.: A Pro-posal of Semi-automatic Indexing Algorithm for Multi-media Database with Users' Sensibility, *Proc. of the 2000 Spring Conference of KOSES& International Sensibility Ergonomics Symposium*, pp. 120-125 (2000).
- [ 4 ] Mitsuishi, T., Sasaki, J. and Funyu, Y.: A Design of A Kansei Retrieval System for Distributed Multi-media Databases, *Proc. of 15th ICOIN*, IEEE Computer Society Press, pp. 285-290 (2001).
- [ 5 ] Okada, R., Lee, E.-S., Kinoshita, T. and Shi-ratori, N.: A Method for Personalized Web Searching with Hierarchical Document Cluster-ing, *Trans. IPS Japan*, Vol. 39, No. 4, pp. 867-877 (1998).
- [ 6 ] 風間一洋, 佐藤進也, 清水奨, 神林隆: WWWのユーザ操作履歴による HTML 文書の相関関係の解析, *情報処理学会論文誌*, Vol. 40, No. 5, pp. 2450 - 2459 (1999).
- [ 7 ] 桑田喜隆, 谷津正志, 小泉宣夫: ユーザモデルに基づく技術支援情報の自動配信サービス, *情報処理学会論文誌*, Vol. 40, No. 11, pp. 3896 - 3905 (1999).
- [ 8 ] 原田将治, 伊藤幸宏, 中谷広正: 感性語句を含む自然言語文による画像検索のための形状特徴空間の構築, *情報処理学会論文誌*, Vol. 40, No. 5, pp. 2356 - 2366 (1999).
- [ 9 ] 吉田尚史, 清木康, 北川高嗣: 意味的連想検索機能を持つメディア情報検索システムの実現方式, *情報処理学会論文誌*, Vol. 39, No. 4, pp. 911 - 922 (1998).
- [ 10 ] 三浦真奈美, 三石大, 佐々木淳, 船生豊: 感性語句による音楽データベース検索システムの構築, 第 62 回情報処理学会全国大会講演論文集 (3), pp. 69 - 70 (2001).
- [ 11 ] 田崎真也: 田崎真也が選ぶ毎日飲むワイン, 新星出版社 (1998).
- [ 12 ] 佐藤聡, 菊地幸平, 北上始: 音楽データを対象としたイメージ検索のための感情価の自動生成, *情報処理学会研究会報告 99-DBS-118*, pp. 57 - 64 (1999).
- [ 13 ] 三石大: 媒介変数を用いた嗜好分析による個人の嗜好を反映可能なデータベース検索, データベースと Web 情報システムに関するシンポジウム (DBWeb2002) 予稿集 (情報処理学会シンポジウムシリーズ), Vol. 2002, No. 19, pp. 291 - 298 (2002).
- [ 14 ] 三石大, 多田和彦, 佐々木淳, 船生豊: 媒介変数を用いた嗜好分析による感性情報検索の提案, *情報処理学会研究報告 2001-DBS-124*, Vol. 2001, No. 44, pp. 1 - 8 (2001).
- [ 15 ] 多田和彦, 三石大, 佐々木淳, 船生豊: 媒介変数により個人の嗜好を反映可能なレシピ検索システムの構築, *情報処理学会研究報告 2002-DBS-126*, pp. 137 - 144 (2002).

## **Development of A Database Retrieval Method which Reflects Individual Interests, Concerns, and Preferences**

**Takashi MITSUISHI**

*Graduate School of Educational Informatics, Tohoku University*

In this paper, we describe a new retrieval method we are developing in order to find target data efficiently from databases, which are consisting of some indistinguishable data by specific words because of their complex characteristics. It could reflect individual preferences of users on retrieval results using mediation variables, and facilitate users to find target or preferable data. Based on our methods, we design and implement an on-line recipe retrieval system as an example of database applications, perform an experiment with the implemented system, and show the effectiveness of our methods with the experimental result.

Now, we are archiving lecture data for Internet School of Tohoku University (ISTU) in order to provide them via the network. Some multimedia databases are constructed on the web, and are providing their data (e.g. movies, music and others). The archives grow larger, some effective retrieval methods become to be necessary to find target data. However, the characteristics of the contents that are consisting of complex data such as the lecture data, movies data, music, and other multimedia data are hard to be expressed by some specific words. It makes difficult to distinguish different data and find target data from archives when we are using some existing relational database or full-text search methods as it was.

In order to find target data efficiently from such complex databases, we are proposing a new retrieval method. This method uses the other data as mediation variables, which have some relation to target data. It defines mapping function from target data to mediation variables in advance based on the relation between them. It estimates individual interests, concerns or preferences of different users through the mediation variables, and uses estimated preferences as evaluation functions for retrieving. Thus, it reflects them on retrieval result, and facilitate users to find target or preferable data, although there is not some specific words to distinguish target data.

Based on the proposing method, we designed and implemented an on-line recipe retrieval system as an example of web based database application. It estimates users' individual preferences for dishes through wine types as mediation variables based on our method, and recommends preferable dishes to a user when he / she queries by materials of dishes. And, we performed experiment using implemented system. We prepared two systems; one is the system without our method, and the other hand is implemented with our methods, and compared effects between them. As a result, we found that subjects could find target data efficiently by the system with our method compared with the system without our method, and confirmed the effectiveness that our method could reflect individual preferences on retrieval results.

**Keywords:** multimedia database application, data mining, preference analysis, kansei retrieval, preference analysis, mediation variable