

# コンサルテーションアシスタント —対話に基づく知的検索—

松生一郎 宮下敏昭

日本電気(株)関西C&C研究所

ユーザが情報検索時に用いる検索戦略を考慮に入れ、個々のユーザに適応した情報の絞り込みと再検索を実現する情報検索機能を提案する。特に、「ダイナミッククラスタリング」と呼ぶ検索結果のリアルタイム分類手法と、「延長検索」と呼ぶ"Retrieval by Examples"に類する検索手法について詳述する。「ダイナミッククラスタリング」はユーザによる情報の絞り込みを支援でき、ユーザ嗜好を反映したクラスタリングを行うことが期待される。「延長検索」は検索空間の拡大が実現でき、前回の検索実行時に明示していなかったが検索結果を見て好悪の反応を示した条件や緩めても良いと感じている条件を推定できると期待される。

## CONSULTATION ASSISTANT

### - AN INTERACTIVE AND INTELLIGENT RETRIEVAL SYSTEM -

Ichiro Matsuo Toshiaki Miyashita

Kansai C&C Reserch Laboratory, NEC Corporation

1-4-24 Shiromi, Chuo-Ku, Osaka 540, Japan

Information retrieval functions based on human retrieval tactics are proposed to adaptively filter and re-retrieve information. Especially a realtime clustering method called "Dynamic Clustering" and a re-retrieval method called "Extensible Retrieval" which is similar to "Retrieval by Examples" method are discussed in detail. "Dynamic Clustering" supports information filtering and it can reflect users' tastes. "Extensible Retrieval" expands retrieval space and it can find users' potential retrieval conditions.

## 1. はじめに

近年、中古車の無在庫販売サービスや、不動産物件販売サービス、住民票の自動発行サービスなど、サービス業におけるシステム化が急速に進展している。それらのシステムで利用されている端末には大きく分けて、端末操作者が顧客自身である無人機と、端末操作者が専任のオペレータである有人機とがある。無人機では提供できるサービスの品質に限界があり情報提示が中心となって物販には結びつきにくい等の問題がある。また有人機では専任のオペレータにしか操作が許されないため、ユーザが自ら希望を入力しながら自由にデータを散策する楽しみや手段が失われてしまう等の問題がある。

筆者らは、有人機と無人機の長所を併せ持つことで前述した問題点を解決する「コンサルテーション端末」の研究を行っており、その第一歩として「コンサルテーション端末」の核となる情報検索の方法に着目した。従来、情報検索の分野では検索対象を如何に指示し易くするかに注目し、形式言語（例えば、SQLなど）および自然言語によるインタフェース[1]が開発されてきた。一方で、柔軟な検索インタフェースとしてブラウジング[2][3]、ハイパーメディアに代表されるような言語インタフェースによらない直接操作による検索対象の指示方式が盛んに用いられるようになってきた。これらの技術により、記述のし易さ、操作のし易さは向上したが、人間が用いる検索の戦略までも考慮したインタフェースとしては不十分である。検索の戦略を考えるとという意味は、ある検索条件で検索した結果を眺めて、次にどのような条件変更を必要とするか、その意思決定のために検索結果をどのような形で提示することが望ましいかを考慮することである。

本稿では、検索のフィードフォワード過程とフィードバック過程をうまくミックスさせた情報検索機能を提案する。特に、フィードフォワード過程では「ダイナミッククラスタリング」と

呼ぶ検索結果のリアルタイム分類手法を、フィードバック過程では「Retrieval by Examples」に類する類似データ検索による検索空間の拡大手法をそれぞれ述べる。

## 2. 「コンサルテーション端末」の利用イメージ

筆者らが現在開発中の「コンサルテーション端末」は、有人機と無人機の間位置づけられる半有人機の利用イメージを想定している。すなわち、店舗を訪れた顧客は端末を自由に操作して、自分の希望条件を入力しつつ、見たい情報を検索する。そして具体的に商品購入の意思決定を行う、あるいは商品購入の契約や予約の手続きを行う段階になると、コンサルタントが顧客の入力情報を分析しつつ、専門的なアドバイスを交えて顧客と相談し、顧客の意思決定を支援する。図1に「コンサルテーション端末」のイメージを示す。

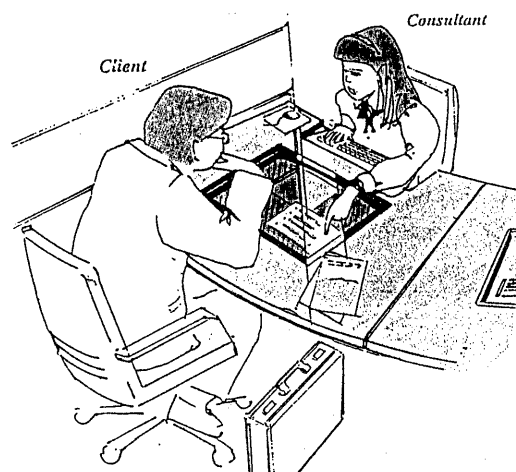


図1 「コンサルテーション端末」のイメージ

## 3. 対話に基づく知的検索

### 3.1 ユーザタイプと検索戦略

顧客が情報検索をする場面を想定した場合、以下のようなユーザタイプが考えられる。

- ユーザタイプ(1) 欲しい情報が持つべき属性およびその値が割り合い明確なユーザ。  
 ユーザタイプ(2) (1)のタイプに近いが、さらに良い物はないかと探求するユーザ。  
 ユーザタイプ(3) サンプルから欲しい情報を発想しようとするユーザ。

上記の3つのタイプのユーザが情報検索時に取ると想定される検索戦略上の特徴を表1にまとめる。

表1 ユーザタイプと検索戦略

タイプ	検索戦略上の特徴
(1)	検索条件をきっちり与える。 ・検索結果のレビュー時にも条件として与えた属性の値の比較が主となる。 ・検索結果が各属性毎にソートされていることを望む。
(2)	必要最小限の検索条件を与える。 ・検索結果に応じて条件式の変更を伴う。 ・条件式変更のトリガーとして、検索結果の特徴を示す情報を望む。
(3)	思い着くままの検索条件を与える。 ・検索結果の中から比較的良いものをサンプルとして、サンプルに類似した情報を追いかけることで検索条件を固める。

### 3.2 「コンサルテーション端末」に搭載する対話検索機能

3.1節で述べたユーザタイプと検索戦略を考慮して、筆者らはユーザとの対話に基づき、個々のユーザ（顧客）に適応した検索情報の絞り込みと情報提示を実現し、ユーザの嗜好を反映し

た再検索手段を提供する、対話に基づく知的検索（以下、対話検索と呼ぶ）技術を開発して、「コンサルテーション端末」に搭載する事を目指している。現在開発を進めている対話検索機能には以下のものがある。

#### (1) オーバービュー表示機能

（→ユーザタイプ(1)の支援）

検索されたデータの情報をユーザの重視している項目を残して情報圧縮し、画面に一覧表示することによって複数のデータを比較検討させ、希望するデータを選択しやすくする。希望するデータが選択されたらそのデータの詳細表示を行う。

#### (2) プライオリティ表示機能

（→ユーザタイプ(1)の支援）

ユーザの希望条件と検索されたデータとの間に距離(非適合度)を導入し、距離の昇順にデータをソートして、ユーザ希望との適合度が高いデータから順番に表示する。

#### (3) 典型例表示機能

（→ユーザタイプ(2)の支援）

検索されたデータを複数クラスに分類し、各クラスから代表データを1件ずつパイロットデータとして表示する。表示されたパイロットデータをもとにユーザが希望するクラスを選択する。選択されたクラス中のデータが設定された表示件数よりも多いならば、そのクラスをさらに複数クラスに分類して各クラスより1件ずつパイロットデータを表示する。こうして、「クラス分け→クラス選択」を繰り返して表示データを絞り込む。

#### (4) 延長検索機能

（→ユーザタイプ(3)の支援）

オーバービュー表示を行った後、ユーザに希望データを決定させるが、希望にぴったりと合致するデータがなかった場合に、表示したデータの良し悪しを○×で入力させる。この○×評価をもとにしてシステムが検索条件を修正して再検索を行う。

図2に「コンサルテーション端末」のモジュール構成を示す。

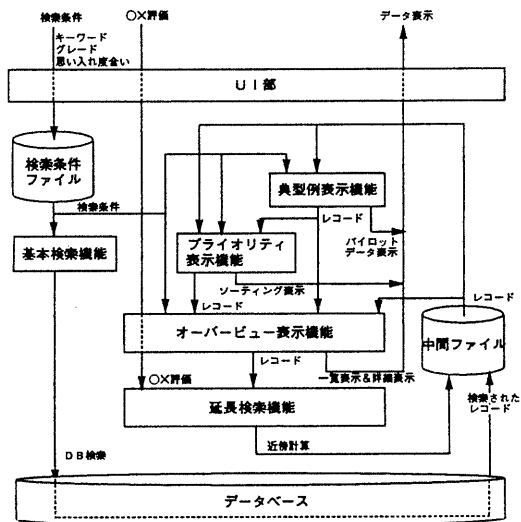


図2 「コンサルテーション端末」の構成

本稿では以下に、対話検索機能の中で代表的な機能である(3)の典型例表示機能と(4)の延長検索機能について詳細に述べる。

#### 4. 典型例表示機能と延長検索機能

##### 4.1 旅行相談システムを対象とした

##### データ構成の例

表2に宿泊施設選択を行う旅行相談システムをアプリケーションとした場合の登録データ例の一部を示す。

表2 旅行相談システムにおける登録データの例(一部)

条件 宿	地区	立地	グルメ グレード*	眺望 グレード*
ホテルA	琵琶湖	湖畔	5	7
ホテルB	鳥羽	海岸	7	9
ホテルC	洛中	市街	8	2

ここで、「地区」「立地」のように定性的な値をとる条件を定性条件と呼び、「グルメグレード」「眺望グレード」のように数値を取る条件を定量条件と呼ぶことにする。グレード値は1から10までの整数値をとることにする。

検索開始時にユーザは表1に示されるような複数条件の中から希望する条件を選んで入力する。定性条件に関しては具体的な定性値をキーワードのメニュー選択により入力する。定量条件に関しては1~10までの値の中から希望するグレード値をスクロールバーを用いて入力し、更にその定量条件に関する「思い入れ度合い」をメニュー選択により入力する。「思い入れ度合い」は表3のように定義し、以下で説明する典型例表示機能および延長検索機能でデータ間の距離を計算する際に重み $w$ として用いている。

表3 思い入れ度合いと対応する重み $w$ の値の例

思い入れ度合い		重み $w$ の値
強	この定量条件は絶対に譲れない	5
中	この定量条件は充たして欲しい	3
弱	できたらこの定量条件は充たされるのが望ましい	1

検索の最初の段階では、入力された条件のうち定性条件を用いてデータベース検索をする「基本検索処理」を行い、検索されたデータを中間ファイルにセーブする。以上の検索の初期過程を図3に図示する。

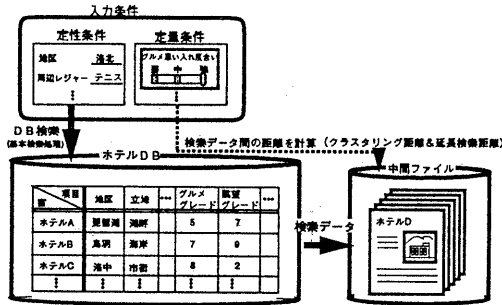


図3 検索の初期過程の例

## 4.2 典型例表示機能

### 4.2.1 典型例表示機能のねらい

「基本検索処理」を行った結果、得られた検索データ数が設定された表示件数より多かった場合に典型例表示機能を用いてデータの絞り込みを行う。

典型例表示機能では、ユーザによるデータの絞り込みを支援するために、検索されたデータのクラスタリングを行う。クラスタリングを行う際には、分類の観点（どういう風にカテゴリー分けすればユーザが喜ぶか）が重要になると考えられる。そこで筆者らはクラスタリングを行う時点で、ユーザが入力した希望条件を分類観点とし、検索実行時に動的にクラスタリングを行う「ダイナミッククラスタリング」と呼ぶ方式を提案した。予め用意された分類観点だけでクラスタリングを行う通常の静的なクラスタリング方式ではデータ登録時にしかクラスタリングが行われない。このようなデータが登録された時点でクラスタが決定してしまう方式[4]と比べて、「ダイナミッククラスタリング」はユーザ嗜好を反映した柔軟性を持つクラスタリングが実現できると考える。

### 4.2.1 「ダイナミッククラスタリング」のアルゴリズム

以下にダイナミッククラスタリングのアルゴリズムを記述する。「基本検索処理」で残ったデータに対し、定量条件を用いて式1で定義した距離 $Dist_a$ をクラスタリング距離として、近接するデータを同一のクラスタに所属するデータとしてまとめる。

式1 データ $d_x$ とデータ $d_y$ とのクラスタリング距離

$$Dist_a(d_x, d_y) = \sqrt{\sum_{j=1}^M \{(g_{xj} - g_{yj})^2 w_j\}}$$

$g_{xj}$ :  $d_x$ の定量条件jに関するグレード

$g_{yj}$ :  $d_y$ の定量条件jに関するグレード

$w_j$ : jに関するユーザの思い入れ度合い

M: ユーザが入力した定量条件の数

近接するデータをまとめてクラスタを生成する方式としては、これまでにいろいろなもの提案されている。筆者らはクラスタ生成方式として最短距離法（クラスタとクラスタの距離を各クラスタに属するデータの対の中で最も近い2つのデータの距離と考える）[5]を用いている。例えば、「基本検索処理」で検索されたデータが表3に示すデータ間距離を持つ場合、図4のように階層的にクラスタが生成される。

表3 データ間距離の例

	ホテルB	ホテルC	ホテルD
ホテルA	5	14	10
ホテルB		15	7
ホテルC			9

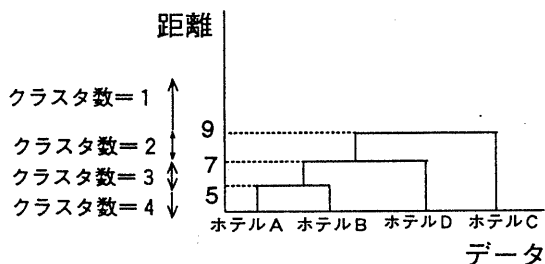


図4 表3のデータ例のクラスタリング結果

さらに、このデータをユーザが指定した定量条件軸から成る空間上に配置する（但しユーザがグルメと眺望の2つの定量条件を指定した場合を仮定している）。まず軸間に強弱が付けられなかった場合、検索結果である4つのホテルは図5-aのように配置され、近接するデータを同一のクラスターにまとめると、図5-aのようなクラスタリング結果が得られる。次に、ユーザの思い入れ度合いを加味して分類軸間に強弱を付ける(図5-b)。この場合はグルメに対する思い入れ度合いを強、眺望に対する思い入れ度合いを弱と設定した場合を示している。データ間の距離尺度として

式1に示した尺度を用いるため、データ間の位置関係が図5-bのように変化し、その結果、クラスタリングの形状も図5-bのように変化する（この図5-bは、表3, 図4のデータ例に対応している）。このように、ユーザの希望条件を反映して分類クラスが動的に変化する。

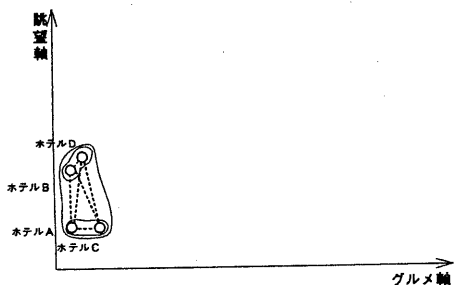


図5-a 軸間に強弱が付けられなかった場合のクラスタリング結果

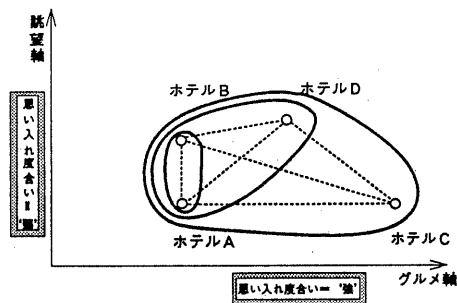


図5-b 軸間に強弱が付けられた場合のクラスタリング結果

このクラス分けが済んだら各クラスターから代表的なデータを選びパイロットデータ（典型例）として表示する。パイロットデータは各クラスターの重心に最も近いデータを選択することにする。生成されたあるクラスター $C_n$ の重心は式2によって求める。また求めた $C_n$ の重心と $C_n$ に属するデータとの距離は式3により求める。

式2 クラスター $C_n$ の重心

$$\text{Grav}(C_n) = \left[ -\sum_{A^i=1} g_{i1}, -\sum_{A^i=1} g_{i2}, \dots, -\sum_{A^i=1} g_{iM} \right]$$

但し、 $g_i \in C_n$  ( $1 \leq i \leq A$ ,  $A$ はクラスター $C_n$ に属するデータの個数)

$g_{ij}$  ( $1 \leq j \leq M$ ): ユーザが入力した  $M$ 個の定量条件のうちの  $j$ 番目の定量条件のグレード

式3 クラスター $C_n$ に属するデータ $d_x$ とクラスター $C_n$ の重心 $\text{Grav}(C_n)$ との距離

$$\text{Dist}(d_x, \text{Grav}(C_n)) = \sqrt{\sum_{i=1}^M \{(g_{xi} - \text{Grav}(C_n)_i)^2 w_i\}}$$

$g_{xi}$ :  $d_x$ の定量条件 $i$ に関するグレード  
 $\text{Grav}(C_n)_i$ : グループ $C_n$ の重心 $\text{Grav}(C_n)$ の定量条件 $i$ に関するグレード

$w_i$ : 定量条件 $i$ に関するユーザの思い入れ度合い

$M$ : ユーザが条件設定した定量条件の個数

次に、ユーザは表示されたパイロットデータを見て、気に入ったクラスタを選択する。選択されたクラスタ内のメンバー（所属データ）数が予め設定された表示件数以下ならば、そのメンバーを全て表示する。選択されたクラスタ（選択クラスタと呼ぶ）のメンバー数が表示件数を越える場合は、図4のように得られたクラスタリング結果を参照して、選択クラスタを複数のクラスタに分割する。そして分割されてきた各クラスタからパイロットデータを選んで表示し、ユーザに気に入ったパイロットデータを選択させる。このように、「クラス分け→クラス選択」を繰り返してユーザの希望するデータを段階的に絞り込む。

クラスタリング距離は、ユーザが実行時に指定した定量条件のみに対して、ユーザが付した思い入れ度合いの重み付けを加味したノルムをとっている。こうして、ユーザが検索実行時に入力した希望条件を反映して分類クラスが動的に変化するため、ユーザによる希望データの絞り込みが支援できる。またこうして得られたクラスタリング結果は、ユーザ嗜好（ユーザが興味を引かれる傾向）を反映していると考えられる。

### 4.3 延長検索機能

#### 4.3.1 延長検索機能のねらい

典型例表示機能では、ユーザの指定した定量条件だけで距離計算を行うが、検索されたデータをユーザが実際に見て、希望条件として明示していなかった定量条件に対して好悪の反応を示すということが考えられる。延長検索機能はこのような要因を加味して、明示しなかったが魅力に感じている条件や緩めても良いと感じている条件を加味して、検索条件を修正し、ユーザの希望するデータの類似データを再検索し表示する。これは、“Retrieval by Examples”と呼ばれる検索手法[6]の一実現方式である。

#### 4.3.2 延長検索のアルゴリズム

以下に延長検索のアルゴリズムを記述する。まずユーザに表示データの良し悪しを評価させ、希望に近いと感じたデータに対して○、希望とかなり違うと感じたデータに対して×をつけさせ、そのどちらでもないデータに対しては何もつけさせない。ここで、○をつけられたデータをポジティブサンプル、×をつけられたデータをネガティブサンプルと呼ぶ。

ユーザによって○×評価がなされたら、ポジティブサンプルの近傍データの集合(PNS: Positive Neighbors Set と呼ぶ)と、ネガティブサンプルの近傍データの集合(NNS: Negative Neighbors Set と呼ぶ)を計算する。ここで近傍距離 $Dist_b$ (延長検索距離と呼ぶ)は式4のように定義し、近傍のしきい値(どの距離以内に含まれるデータを近傍データとするか)は式5のように、ポジティブサンプルとネガティブサンプルとのペア間の距離の半分の値としている。

式4 データ $d_x$ とデータ $d_y$ との延長検索距離

$$Dist_b(d_x, d_y) = \sqrt{\left[ \sum_{j=1}^C \{(g_{xj} - g_{yj})^2\} \right]}$$

$g_{xj}$ :  $d_x$ の定量条件 $j$ に関するグレード  
 $g_{yj}$ :  $d_y$ の定量条件 $j$ に関するグレード  
 $C$ : データベースに登録されている定量条件の総数

式5 ポジティブサンプル $d_p$ と  
ネガティブサンプル $d_n$ との間の  
近傍距離しきい値 $Th(d_p, d_n)$

$$Th(d_p, d_n) = \frac{1}{2} \text{Dist}(d_p, d_n)$$

$$= \frac{1}{2} \sqrt{[\sum_{i=1}^c \{(g_{p1} - g_{n1})^2\}]}$$

$g_{p1}$ :  $d_p$ の定量条件 $i$ に関するグレード  
 $g_{n1}$ :  $d_n$ の定量条件 $i$ に関するグレード  
 $C$ : データベースに登録されている  
 定量条件の総数

PNSとNNSを求めたら、PNSに含まれるデータからNNSに含まれるデータを除き、残ったデータをユーザが希望したデータと類似するデータとして再検索し表示する。

延長検索の実行例を図6に図示する。

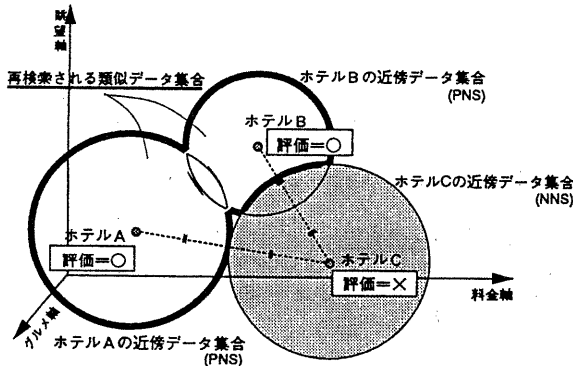


図6 延長検索の実行例

延長検索距離は、定量条件に対してユーザが付した思い入れ度合いの重み付けを外し、ユーザが明示しなかった条件も含めた全ての定量条件に対してノルムをとっている点で、典型例表示機能におけるクラスタリング距離と異なっている。このように近傍距離計算の過程においてユーザが入力した定量条件以外の条件も加味したノルムをとるため、検索空間の拡大が実現で

きる。またこれにより、明示していなかったが実際のデータを見て好悪の反応を示した条件や緩めても良いと感じている条件を推定できると考える。

## 5. おわりに

本稿では、個々のユーザに適応した情報の絞り込み及び情報提示を実現するとともに、ユーザとの対話に基づいて再検索手段を提供する対話検索機能について述べた。今後は、現在開発している「コンサルテーション端末」のプロトタイプを完成し、その有効性を評価する。また、検索要求の分析を通して、対話検索技術を深め、コンサルテーション機能の拡張を目指して研究を進めていく。

## 謝辞

本研究を行う機会を与えて下さった、日本電気(株)関西C&C研究所の真名垣所長、宮井課長に感謝致します。

## 参考文献

- [1]谷, 久保, 市山, 会森「情報検索におけるサーチャーの知識を用いた自然言語からの検索式生成」(情処第44回全国大会3G-9, 1992)
- [2]笠原, 岸本「画像データベースナビゲーション」(信学技報IE88-46, 1988)
- [3]宮沢, 木下, 横山, 松下「APTBook: パラパラめくりによる情報提供環境」(情処第41回全国大会6G-3, 1990)
- [4]Fisher, D.H. 「AutoClass: A Bayesian Classification System」(Proc. of 5th Conf. on MACHINE LEARNING, 1988)
- [5]Anderberg, M.R. 「Cluster Analysis for Applications」(Academic Press, 1973)
- [6]大野「確率論理による学習型文献検索」(人工知能学会研究会資料, 1988)