

2R-4

オブジェクトの再利用と類似検索

宇田川 佳久 ・ 辻 秀一 ・ 市川 照久
(三菱電機情報電子研究所)

1. はじめに

CAD, OAやエキスパート・システム(ES)用のデータベース管理システムでは, 従来の事務処理用のデータベースとは若干異なったデータ検索機能が重要になることが指摘されている[1]。CAD, OAやESにおけるデータ検索の重要な目的は, オブジェクト(設計オブジェクトや知識)を再利用することである。この場合, 検索条件を満たすオブジェクトを取り出すことに加え, 再利用の可能性を考慮した“より適切な”オブジェクトを選び出すことが必要になる。我々は, 部分一致に基づくオブジェクトの類似検索機能と類似度と相違度に基づくオブジェクトのランク付け手法を提案し, 実現してきた[2]。本文では, 従来の手法を述べるとともに, 部分一致に基づく手法との比較を行う。

2. 再利用のためのデータ検索機能

従来のデータベースでは, 検索条件を満たすデータはすべて等しく必要なものであり, ほとんどの場合, 検索されたデータ同士に“より適切である”といったランク付けをするようなことは無かった。

一方, CAD, OA, ESといった新しい応用におけるデータ検索の重要な目的は, オブジェクトを再利用することである。オブジェクト再利用のプロセスは, おおよそ次のようなものである。

- (1) 検索条件を満たすオブジェクトを検索する。
- (2) 適切なものを選ぶ。
- (3) オブジェクトを書き換える。

すなわち, 検索結果として最終的に必要となるオブジェクトは数個(ほとんどの場合1個)である。従って, 条件に合致するオブジェクトを提示するだけで良いのではなく, “ユーザの要求により近い”オブジェクトを順次に提示することが必要となる。

Reusing Objects and Similarity Retrieval Techniques
by Y. Udagawa, H. Tsuji, T. Ichikawa
MITSUBISHI ELECTRIC Corp.

3. ブーリアン検索によるアプローチ

3.1 重み付けブーリアン検索

検索式は複数の検索条件項から構成されていることが多い。ここで, 検索式に対する検索条件項の重要性が同じであるとは限らない。例えば, 自動車販売の対象者を選ぶには“年収額”が重要な検索条件項となり, “性別”はそれほど重要ではない。これに対し, ハンドバッグ販売の対象者を選ぶには, “性別”が重要で, “年収額”は比較的重要ではない。検索条件項に重みを付け, 条件項を満たすごとに重みを加えれば, 検索式を満たすデータに重み付けをすることができる。

図1は, 4個の検索条件項のANDによって生成される検索式のラティス構造を示したものである。実線で囲んだ部分は, 検索条件項A, B, C, Dに, それぞれ1, 2, 3, 1という重みを付けたとき, 5以上の重みになる検索式の集合を示している。

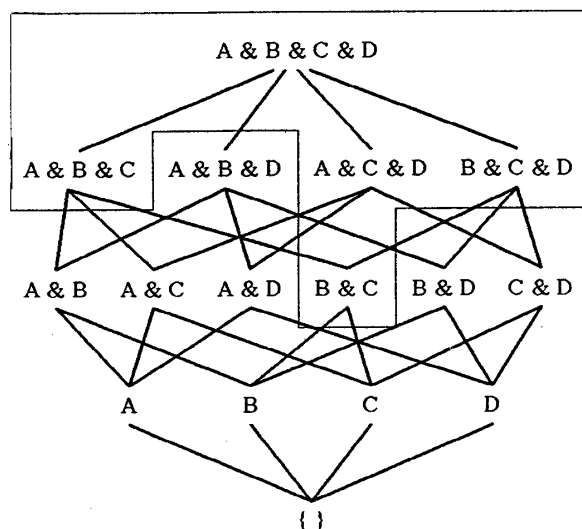


図1. 検索式のラティス構造

3.2 ファジィ検索

古典的集合論においては, 要素が集合に属するか否かを基準としていたのに対し, ファジィ集合論では, 要素の集合への帰属度を0から1の間の値を取るメンバシップ関数 μ_f によって定義している。ファジィ集

合論をデータ検索に適用したものがファジィ検索である。

ファジィ検索では、検索式（が定めるデータ集合）に対するデータの帰属度を、次のルールに従って計算する。

$$mf(Q_1 \text{ AND } Q_2) = \min(mf(Q_1), mf(Q_2))$$

$$mf(Q_1 \text{ OR } Q_2) = \max(mf(Q_1), mf(Q_2))$$

$$mf(\sim Q_1) = 1 - mf(Q_1)$$

ここで、 Q_1 、 Q_2 は任意のブーリアン検索式である。

4. 部分一致によるオブジェクトの検索

我々は、ADAMと称するオブジェクト指向リレーショナル・データベースを開発し、回路図の管理に適用してきた。ADAMでは、前節で述べたオブジェクトの再利用を前提としたデータ検索を行うために、部分一致によるオブジェクトの検索機能を実現した。

4.1 検索条件の指定

ADAMでは、これまでに、属性値に基づく回路図の検索機能を実現し、試使用してきたが、この方法は不便なことが多いことが判明した。そこで、今回は回路図を検索条件として回路図を検索するアプローチを取った。この方法によれば、ユーザは欲しいと思っている回路図の概略図（ルアー回路図）を通常の回路図の編集と同様な方法で編集し、その後、このルアー回路図に類似する回路図の検索をシステムに依頼するだけで所望の回路図を検索できるようにした。

4.2 “より適切な”回路図の検索

ADAMのデータ・ディレクトリは、回路図とそれを構成する部品回路の引用情報を管理している[2]。このデータ・ディレクトリを使えば、ユーザが検索条件として指定した概略回路図を構成する部分回路を含む回路図を検索することができる。こうして得られた回路図の集合が、候補回路図の集合である。この中から最適な回路図を選択するわけであるが、回路図名を見ただけでは、どの回路図が適しているのか判別することは難しい。ADAMではユーザから与えられたルアー回路図と候補回路図との類似度と相違度を以下のルールによって計算し、回路図の選択の基礎情報としている。

$$\begin{aligned} \text{類似度} & \quad \text{Cardinal}(X \cap Y) \\ \text{相違度} & \quad \text{Cardinal}(X - Y) \\ & \quad + \text{Cardinal}(Y - X) \end{aligned}$$

ただし、 X 、 Y は、それぞれ、ルアー回路図の部分回路の集合、候補回路図の部分回路の集合である。従来の類似検索が、オブジェクトの類似度のみに着目していたのに対し、ADAMでは類似度に加え相違度も扱っていることが特色である。図2は、ルアー回路図と類似回路図との概念的な対応を示している。

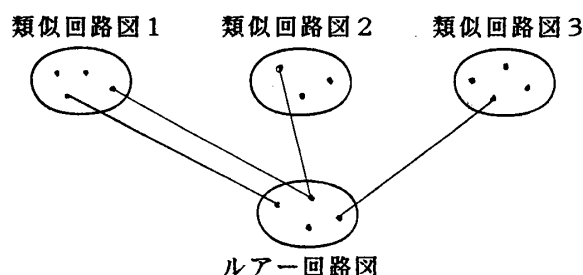


図2. ルアー回路図と類似回路図

5. おわりに

表1は、重み付けブーリアン検索、ファジィ検索、部分一致によるオブジェクトの検索手法の特徴をまとめたものである。オブジェクトは、一般に、階層構造を成す。部分一致による検索手法は、この階層構造を利用するものであり、オブジェクトによるオブジェクトの検索ができること、類似度のみならず相違度も扱えることが特徴である。オブジェクト指向データベースでは、部分一致による類似検索機能が必須であると考えている。

表1. 検索手法の比較

手法	重み付け	ファジィ	部分一致
オブジェクトの階層	無し	無し	有り
検索条件	ブール式	ブール式	オブジェクト
類似度	モデル	帰属度	モデル
相違度	無し	無し	モデル

参考文献

- [1] Held, H. J. and Carlis, V. J. : MATCH--A new high-level relational operator for pattern matching, CACM. 30. 1. (1987).
- [2] 宇田川, 辻, 市川: オブジェクト指向リレーショナル・データベースADAM ..., 情報処理学会アドバンスト・データベースシステム・シンポジウム予稿集 (1988).