

不完全情報の広域問合せ

6X-5

宮崎収兄、大塚真吾
千葉工業大学情報工学科*

1 はじめに

オープンなネットワーク環境で不特定多数のデータベースを対象とするワールドワイドDBでは不完全DBの概念を基本とした処理を行う。不完全DBの代表としてnullを含んだDBがあるが、広域問合せの処理にはnullとは異なる枠組みが必要である。本稿では広域問合せのための不完全DBとnullとの比較を述べる。

2 広域問合せのための不完全DB

不特定多数のDBを対象とするため、従来のマルチDBとは異なって、スキーマ統合をできるだけ行わずにそのつど表現の差異を変換しながら問合せ処理を行う方法を検討している。その基礎として不完全DBの概念を導入する。以下では変換可能な属性名やデータ型の差異などは変換済みであると仮定して議論する。

不完全DB概念の導入は広域検索においてスキーマの不一致を吸収するためである。

1. 問合せ中のターゲット属性がDBに存在しない場合も処理を可能とする。
2. 問合せ中の条件属性がDBに存在しない場合も処理を可能とする。
3. 複数のDBからの解を統合し必要な解を得ることを可能とする。

まず、複数の類似したDBへの問合せを検討する。この機能は単純な選択(+射影)問合せで実現でき、必要な拡張は以下である。

1. 問合せ中のターゲット属性がDBに存在しない場合はその属性を含まない解を得る。
2. 条件付タプルの導入による拡張を行う。問合せ中の条件属性がDBに存在しない場合は条件付の解を得る。

3. 複数のDBからの解の統合は和を用いる。関係をタプル毎に属性が異なっても良いように拡張する。

次に複数のサイトにある関係の結合を考える。関係RとSにそれぞれ属性rとsがあるとする。

$$Answer = \sigma_{r=a \wedge s=b}(R \bowtie S)$$

問合せ元のサイトが対象サイトにあるスキーマを知らない場合、もっとも容易な方法は $R \bowtie S$ に対応する1つの関係Tを想定し、選択問合せ $\sigma_{r=a \wedge s=b}T$ を2つのサイトに送ることである。対象サイトでは上記の選択問合せで述べた拡張にしたがい、それぞれ条件付解を返す。問合せ元では2つの方法で処理が可能である。1つは2つの解が結合可能であることを判定し、解の結合を行うことである。もう1つはいったん2つの解の和を取り、その解を元に結合と同等な処理を行うことである。概念的に後者の採用を行うことにより、不完全DBの統一的な導入が可能になる。

不完全DBにおいて結合と同等な演算を以下のように導入する。 $T = R \cup S$ を考える。ただし、RとSは完全DBであるとする。不完全DBにおける自然結合はタプル対毎の共通な属性による結合を行うこととすると、

$$\begin{aligned} T \bowtie T &= R \bowtie S \cup R \bowtie R \cup S \bowtie S \\ &= R \bowtie S \cup R \cup S \end{aligned}$$

が成り立つ。従って、和の自然結合によって完全DBの自然結合に対応する結果を得ることができる。この演算では外結合と同様に結合により情報が失われることはない。これはオープン環境での問合せ処理には望ましい性質である。

上記の基本的な拡張を元に他の演算を定義すると以下のような拡張関係代数が得られる [3]。

- 選択は条件を真とするまたは真とする可能性のあるタプル集合を解とする。条件とする属性が存在しない場合は条件付タプルが生成される。
- 射影は通常の射影と同様。射影の対象属性が存在しないタプルの射影は属性なしのタプルとする。また、

*Global Query for Incomplete Information: Nobuyoshi Miyazaki and Shingo Otsuka, Department of Computer Science, Chiba Institute of Technology, miyazaki@cs.it-chiba.ac.jp

条件は対象属性にかかわらず結果タブルの条件となる。

- 直積は通常の直積と同様。
- 自然結合は通常自然結合と同様だが、タブル対毎に共通属性が異なる。 θ 結合は通常結合と同様に直積と選択の組み合わせで定義される。
- 和と差は通常和と同様で共通部分は差を用いて定義する。

さらに複数DBからの解の和を元に情報の統合・抽出を行うために自然結合を元にしたマージや無条件タブル選択などの演算を導入する。なお、[3]ではマージを1回の自然結合により行う形で定義したが、複数回の自然結合を考慮した修正が必要である。

3 nullによる不完全DBとの比較

不完全DBの代表は関係へのnullの導入である。nullの導入には各種の問題があるが、意味的にはunknownを表すと考えることが多い。また、その処理には大別し2つの方法がある。

null値表 nullを含んだ表で演算を行う。nullは条件の評価などでは $\text{null} \neq \text{null}$ として扱い、重複の判定においては $\text{null} = \text{null}$ として扱う[2]。

条件付表 nullを変数に置き換え、さらに条件の表現を導入した表に拡張して処理する[1]。

null値表による方法はCoddの提案の延長で、実現が容易なためSQLはこの方法にもとづいている。これに対し、条件付表による方法は不完全DBの可能世界意味論と一致する。可能世界意味論では不完全DBを可能世界の集合として考える。1つの可能世界は通常関係と同様にnullでない値を持つ関係で表現される。

2つの方法による問合せ処理の結果は異なる。例えば、nullを含む2つの関係RとSを考える。RとSのタブルがすべて同じ場合、この2つの関係の可能世界集合は等しい。このとき $R=S$ と書くとすれば、完全DBでは以下が成り立つ。

$$R \bowtie R = R \bowtie S = R, R \cup R = R \cup S = R$$

nullがある場合はこれらの等式が成り立たない場合がある。まず、null値表による演算では以下ようになる。

$$R \bowtie R = R \bowtie S \neq R, R \cup R = R \cup S = R$$

すなわち、関係の自分自身との自然結合を行った場合、元の関係が得られない。これに対し、条件付表による演算では以下ようになる。

$$R \bowtie R = R, R \bowtie S \neq R, R \cup R = R, R \cup S \neq R$$

すなわち、自分自身との自然結合などでは元の関係が得られるが、同一の別の関係との自然結合や和では元の関係とは異なった結果が得られる。このように2つの方法は2項演算の結果が大きく異なる。なお、選択と射影の単項演算については条件表現による可能解の精密化以外は類似の結果が得られる。理論的には条件付表の方が整合性があるが、nullはunknown以外の意味に使う場合もあり、どちらの方法にも問題が残っている。

広域問合せのための不完全DBは条件表現の導入を除くとnull値による演算と類似している。すなわち、以下ようになる。

$$R \bowtie R = R \bowtie S \neq R, R \cup R = R \cup S = R$$

但し、自然結合については属性の有無についての考え方が異なるので、null値による自然結合の結果と広域問合せでの自然結合の結果は異なる。なお、等結合など属性を指定した結合の場合は条件表現の導入による差異以外は類似している。従って、広域問合せのための不完全DBは意味論的にはnull値表による不完全DBに条件表現を導入し拡張した形になっている。また、条件のない関係では以下が成り立ち、ローカルDBに(null値表方式による)nullがある場合も広域問合せ処理で容易に扱うことができる。

$$R \bowtie_{\text{null値表}} R \subset R \bowtie_{\text{条件付表}} R = R \subset R \bowtie_{\text{広域}} R \\ R \cup_{\text{null値表}} R = R \cup_{\text{広域}} R = R \subset R \cup_{\text{条件付表}} R$$

4 まとめ

オープン環境での広域検索のための不完全DB概念とそこでの関係代数演算を検討した。さらに広域検索のための不完全DBとnullによる不完全DBの比較を行い、それらの関連を示した。本研究は文部省科学研究費重点領域研究の補助によるものである。

参考文献

- [1] Abiteboul, S., Hull, R., and Vianu, V., Foundations of Databases, Chapter 19, Addison-Wesley, 1995
- [2] Date, C.J., An Introduction to Database Systems, Sixth Edition, 1995
- [3] 大塚真吾、宮崎収兄, “広域検索のための関係代数の拡張”, 情報処理学会DBシステム研究会(風林火山ワークショップ), 109 pp.25-30 (1996)