

制約を用いた異種分散情報源の探索手法

— 位置情報に応じた情報提供のためのフレームワーク —

柳 奇亨* 石川 佳治† 北川 博之†

* 筑波大学 理工学研究科 † 筑波大学 電子・情報工学系

1. はじめに

GPSや携帯端末の発達により、携帯端末を保持したユーザや自動車などの移動オブジェクトの位置情報を得ることが容易となってきている。一方、オンライン、特に Web 上で利用可能な情報源の数も日々増えているが、これらの中にはある地域内の情報を提供する情報源や、提供されるそれぞれの情報に位置情報が付随しているような情報源が存在する。GPSにより取得した位置情報をもとに各時点のユーザの位置に応じた情報提供を行うような情報提供サービスはすでに存在するが、このようなサービスは異種の複数の情報源の存在を考慮しているわけではない。よって本稿では、多種多様な情報源を統合し、位置情報に応じた情報を提供するための枠組みを提案する。

2. 情報源記述を用いる情報統合方式

2.1 概要

まず、ここでは本研究の基礎となる、情報源記述を用いる情報統合方式について簡単に述べる。リレーショナルデータベース、Webサーバなど、ネットワーク上には多種多様な情報源が存在する。異種の情報源を統合し、ユーザの要求に応じて適切な情報を提供するためには、情報源の異種性を吸収し、与えられた問合せに対して情報源を選択し、問合せの発行と結果の統合を行う機構が必要となる。このために広く用いられているアプローチは、与えられた問合せに対して情報源を選択し問合せ処理を行う**メディエータ (mediator)**と、各情報源をメディエータの処理機能に応じて抽象化する**ラッパー (wrapper)**からなるアプローチである。本研究でもこのアプローチを採用する。

メディエータにおける情報統合のための問合せ処理を実現するためには、各情報源が有する情報を適切に表現するための枠組みが必要となる。このような枠組みには、多様な情報源を表現するための柔軟性や、効率的な問合せ処理プランを作成できるための簡潔かつ十分な表現能力が求められる。このような情報源に関するメタ情報を表現するためのアプローチのひとつとして、*Information Manifold* システムで用いられた**情報源記述 (source description)**のアプローチがある¹⁾。本研究ではこのアプローチを拡張して、位置情報に基づく情報提供のフレームワークを提案する。

2.2 データモデル

本研究における情報源記述では、*Information Manifold* のアプローチ¹⁾と同様、リレーションとクラスの両者の概念を用いるデータモデルを採用する。各情報源が有する情報を仮想的なリレーションやクラスにマッピングするのは各情報源に対して構築されたラッパーの役割であり、メディエータではラッパーにより抽象化されたリレーションやクラスの集まりとして情報源をとらえることになる。ただし、実際に情報源記述を行う際には、クラスとオブジェクトの情報をリレーションにマッピングすることで記述の統一化と処理の簡略化を図っている。

クラスのリレーションへのマッピングについて例を示す。たとえば、クラス *Restaurant* に対しては、同名の単項リレーション *Restaurant* を定義する。*Restaurant* リレーションの各タプルには *Restaurant* オブジェクトのオブジェクト識別子 (OID) が格納されると考える。また、クラス *Restaurant* の各属性ごとに、(OID, 属性値) という2項のタプルの集合を保持する同名のリレーションを構築する。たとえば、*Restaurant* クラスの *Name* 属性については、*Name* という名前の2項リレーションを構築する。このようにクラスとリレーションを統一的に表現することで、問合せ処理におけるプラン構築の処理が軽減される。

2.3 問合せと情報源記述の例

ユーザからメディエータに与えられる問合せの例を示す。この問合せは、現在ユーザが存在する位置から距離1000メートル以内のレストランで、レストランの評価値が2.5以上であるようなものの名前、位置、評価値をすべて列挙する問合せである。

$$q(n, p, s) \leftarrow \text{Restaurant}(r), \text{Name}(r, n), \\ \text{Position}(r, p), \text{Rating}(n, s, w), \\ \text{dist}(\text{my_pos}, p) \leq 1000, s \geq 2.5$$

ここでは4つのリレーションを用いた問合せを行っている。最初の3つのリレーションは、クラス *Restaurant* をリレーションにマッピングして得られたものであり、*Name* はレストラン名を、*Position* はレストランの位置を保持している。なお、ここでは位置情報の表現のために、点情報を表現する *point* データ型が存在することを想定している。*Rating* は3項のリレーションであり、属性は順に、レストラン名、レストランの評価、評価した人物名の情報からなる。*my_pos* は、ユーザの現在位置の情報を与える、現在時刻に応じて定まる定数である。*dist* は *point* 型上に定義された距離関数である。

A Heterogeneous Information Integration Framework for Mobile Users Based on Positional Information
Gihyong Ryu*, Yoshiharu Ishikawa†, Hiroyuki Kitagawa†
*Master's Program in Sci. and Eng., Univ. of Tsukuba
†Inst. of Inf. Sci. and Elec., Univ. of Tsukuba
{non,ishikawa,kitagawa}@dmlab.is.tsukuba.ac.jp

ここで、情報源として以下の4つの情報源があったとしよう。

Source 1: 地域限定のレストラン情報 Contents: $V_1(r) \subseteq Restaurant(r), Position(r, p), contain(rect_1, p)$ Capabilities: $(\{Name(r, n), Category(r, c)\}, 1)$
Source 2: イタリアンレストランの情報 (地域限定) Contents: $V_2(i) \subseteq Italian_Restaurant(i), Position(i, p), contain(rect_2, p)$ Capabilities: $(\{Area(i, a)\}, 1)$
Source 3: 高級レストランの情報 Contents: $V_3(r) \subseteq Restaurant(r), Position(r, p), DinnerPrice(r, d), d \geq 10000$ Capabilities: $(\{Name(r, n), contain(-, p)\}, 1)$
Source 4: レストランの評価の情報 Contents: $V_4(r) \subseteq Rating(n, s, w)$ Capabilities: $(\{n, \geq(s, -), \dots\}, 1)$

各々の情報源に対する情報源記述は、その情報源がどのような内容を保持しているかを表す**内容記述** (contents description) と、その情報源の問合せ処理能力を表す**問合せ能力記述** (capability description) からなる。内容記述には、その情報源が提供する情報が、メディアータが想定している仮想的なリレーションのうちのどの部分に対応する情報を含んでいるかを表現している。問合せ能力記述には、その情報源に与えることができる問合せ条件が記述されている。

上の例で、情報源1は地域限定のレストラン情報を示している。たとえば、東京都内のレストランの情報を提供しているような情報源がこれに相当する。 $rect_1$ は、情報源1が情報提供する地域を包含する最小矩形領域 (minimum bounding box, MBB) の定数を表している。 $contain(rect_1, p)$ は、 p が $rect_1$ 中に含まれる場合に真となる述語である。問合せ能力記述における $(\{Name(r, n), Category(r, c)\}, 1)$ は、この情報源に対する問合せ条件として、レストラン名 $Name(r, n)$ に対する値の指定、あるいはレストランのカテゴリ $Category(r, c)$ への値の指定が可能であることを示している。1という数値は、この情報源に対する問合せは、少なくとも $Name(r, n), Category(r, c)$ のうちの一つの値を指定しなければならないことを示している。なお、複数の条件が指定された場合は、論理積をとった問合せを情報源に指定可能であるものとする。

情報源2では情報源1と同様に地域限定の情報を提供しているが、イタリアンレストランの情報のみを提供する点異なる。なお、*Italian_Restaurant* クラスは *Restaurant* クラスのサブクラスである。この情報源に対しては渋谷などの地区名を指定した問合せが可能である。情報源3は、ディナーの標準価格 $DinnerPrice(r, d)$ が10,000以上の高級レストランに関する情報を保持している。この情報源は、レストラン名による検索機能と領域指定による検索機能を提供している。 $contain(-, p)$ は、

述語 $contain$ の第一引数に任意の定数の矩形領域を指定して検索が行えることを示している。情報源4はレストランの評価の情報を提供している情報源であり、レストラン名や評価値に対する条件を用いた問合せが可能である。

3. 問合せの処理

2.3節で示した問合せの処理は以下のように進められる。

1) 情報源4に対し、

$$Tmp(n, s) \leftarrow Rating(n, s, w), s \geq 2.5$$

という条件で問合せを行い、評価の高いレストラン名を得る。

2) 情報源1に対する問合せを行う。まず、ユーザの現在位置 my_pos と $rect_1$ を比較し、 my_pos から $rect_1$ までの最短距離が1,000メートル以下の場合にのみ、以下の問合せを実行する。

$$Res_1(n, p, s) \leftarrow Tmp(n, s), Restaurant(r), Name(r, n), Position(r, p), dist(my_pos, p) \leq 1000$$

$Name(r, n)$ を条件として情報源1に問合せを与え、結果として得られたレストランのうち my_pos から1,000メートル以内のものを選択する。

3) 情報源2については、条件として要求されている地区名を与えることができないため、問合せを行うことはできない。よって考慮の対象から外す。

4) 情報源3に対する処理を行うため、まず my_pos を中心とする半径1,000メートルの円に外接する MBB $rect_3$ を求める。そして以下の問合せを実行する。

$$Res_3(n, p, s) \leftarrow Tmp(n, s), Restaurant(r), Name(r, n), Position(r, p), contain(rect_3, p), dist(my_pos, p) \leq 1000$$

$Name(r, n) \wedge contain(rect_3, p)$ という条件の論理積を情報源3に対する問合せ条件として与えることにより、より絞り込まれた結果を取得することができる。

5) 最終結果として $Res_1 \cup Res_3$ をユーザに返す。

4. まとめと今後の課題

本稿では、位置情報に基づく情報の提供のための、異種分散情報統合のための枠組みについて提案した。今後は提案手法の詳細化を行い、プロトタイプ的设计と実装について検討する。

参考文献

- [1] A.Y. Levy, A. Rajaraman, J.J. Ordille, "Querying Heterogeneous Information Sources Using Source Descriptions", in *Proc. VLDB*, pp. 251-262, 1996.