解説

データベース

Database Technologies

石川 佳治^{*1} 喜連川 優^{*2} ^{*1}名古屋大学 ^{*2}東京大学Yoshiharu Ishikawa^{*1} and Masaru Kitsuregawa^{*2} ^{*1}Nagoya University ^{*2}The University of Tokyo

1. はじめに

データベースの研究は、Codd による 1970 年代の関係データモデルの提案以降、綿々と発展を続けてきた。すでに、関係データベースによるデータ管理は、情報社会を支える基盤技術として極めて重要な役割を果たしている [1]。一方で、データベース分野における研究対象も拡大が進んでいる。1990 年代からのインターネットおよびウェブの急速な発展を受けて、従来のデータベース技術が主として定型的なビジネスデータを対象としていたのに対し、半構造データ、ウェブデータ、ストリームデータ等、多様かつ大規模なデータをどのように活用するかがデータベース分野における重要な課題の一つとなっている。

このような流れを受け、筆者の喜連川を代表とする科学研究費補助金特定領域研究「情報爆発 IT 基盤」が 2006 年度から 2010 年度までの 5 年間をかけて進行中である [2] [3]。この特定領域研究は、爆発的に増大を続ける情報にどのように立ち向かうかという大きな目標に対し、データベース分野の研究者のみならず、自然言語処理、メディア情報処理、ネットワーク、分散コンピューティングなど、様々な分野の研究者が結集した情報分野を横断するプロジェクトである。新しいサーチエンジン Tsubaki やポストクラウド基盤 InTrigger など、多様な分野に対する研究開発を進めている。

以降では、近年のデータベース分野のホットな研究分野の中から、ロボティクスに関連の深い話題をいくつか選んで紹介する。まず、センサ環境における情報の効果的なモニタリングのためにデータベース技術を活用するセンサデータベースの技術を紹介する。次いで、これらの分野に関係が深く、最近着目を浴びている確率的データベースについて述べる。そして、自動車、ロボットなどの移動する対象

物のデータベースにおける表現や問い合わせ処理を目的とする移動オブジェクトデータベースについて説明する。誌面の都合もあり、それぞれの話題について簡単に触れることしかできないが、データベース研究者の問題意識を共有いただけると幸いである。

2. センサデータベース

センサ機器の普及に伴い、センサ環境において大量のデータが生み出されるようになっている。スマートグリッドではすでに数千万個のセンサノードが利用されるなど研究が活発化しつつある。この状況を受け、センサ情報処理のためのデータベース技術の開発が進められている [4]。以下では、様々な取り組みの中から、無線センサネットワーク [5] におけるデータベース技術を例にとり解説する。その典型的な利用形態の一つとして、屋内外における環境モニタリングが挙げられる。図 1 に、このような状況におけるセンサデータベースの概念図を示す。センサノードは通常バッテリーで駆動されることから、消費電力の削減が重要となる。

2.1 TinyDB: センサデータベースシステムの実例

TinyDB は、センサデータベース技術に関する初期の研究としてよく知られている [6] [7]。TinyDB では、収集・集約対象となっているセンサデータを、仮想的な関係データベースとして抽象的に表現する。センサネットワークにお

図 1 センサデータベースの概念

原稿受付 2010 年 1 月 12 日

キーワード: Database Technologies, Sensor Databases, Probabilistic Databases, Moving Object Databases

^{*1}〒464-8601 名古屋市千種区不老町^{*2}〒153-8505 目黒区駒場 4-6-1^{*1}Chikusa-ku, Nagoya-shi, Aichi^{*2}Meguro-ku, Tokyo

けるデータ集約の要求を、SQL [1] 風の問い合わせ言語を用いて容易に記述できる。以下は文献 [7] からの例であり、音量のセンシングにより、6 階の会議中の部屋を探す問い合わせである。

```
SELECT AVG(volume), room
FROM sensors
WHERE floor = 6
GROUP BY room
HAVING AVG(volume) > threshold
SAMPLE PERIOD 30s
```

仮想的な関係表 `sensors` はセンサノードの情報を表現しており、センシングされる音量を属性 `volume` として表す。上記問い合わせは、30 秒ごとに各部屋の音量の平均をとり、それが閾値以上の場合に結果を出力する。問い合わせを受けると、TinyDB システムは、センサデータ収集のため木構造ネットワークを動的に構築する。問い合わせの形式でセンサネットワーク上のデータ集約を記述できることは、ユーザの労力を省くだけでなく、処理コストの面でも利点がある。TinyDB は、指定された要求を満たすために消費電力をできるだけ小さくするような問い合わせ処理プランを導出する。

2.2 モデルベースの問い合わせ

このような考え方を発展させ、データ集約処理を確率モデルでモデル化し、センサネットワークにおける問い合わせに確率的な推定処理を融合するアプローチが提案されている。このような問い合わせはしばしばモデルベースの問い合わせと呼ばれる。例えば BBQ システム [8] では、

```
SELECT nodeId, temp ± .1°C, conf(.95)
WHERE nodeId in {1..8}
```

という問い合わせが記述できる。これは、センサノード 1～8 における温度を返す問い合わせであるが、精度や信頼度が指定できることが特徴である。例えば問い合わせ結果に (1, 16.3) というタプルが含まれる場合、これはセンサノード 1 における温度測定値が 16.2°C～16.4°C の範囲にある信頼度（確率）が 95%以上であることを意味する。

モデルベースの問い合わせの処理では確率モデルが積極的に活用される。上記 BBQ においては、センシング回数削減のため、センサノードの履歴情報や複数のセンサ入力の相関関係が活用される。例えば、前回までの問い合わせ結果をもとに構築した確率的モデルにより、あるセンサノードの現在の気温が指定された精度と信頼度を満たして推定できるなら、そのノードへのセンシング処理が省略される。また、地理的に近いセンサノードのペアについて、一方だけのセンシングを行い、もう一方の測定値は履歴・相関情報から確率的に推定するといったアプローチもとられる。

ここでは、センサデータベースに関する研究のほんの一例を紹介した。確率的情報の活用はロボティクス研究でも

Researchers			Memberships			
Name	Affil	Pr	Name	Society	Membership	Pr
Mike	MIT	0.3	Mike	IEEE	Member	0.9
	CMU	0.2	Mike	IEEE	Fellow	0.2
	Google	0.5	Mary	ACM	Student Member	0.5
Mary	MIT	1.0	John	ACM	Member	0.7
John	CMU	0.7	John	IEEE	Member	0.8
	MIT	0.3				

図 2 確率的データベースの例

盛んであるが [9]、データベース分野においても重要な課題である。次章で、さらに関連する話題に触れる。

3. 確率的データベース

3.1 確率的データベースとは

センサ情報に限らず、インターネット上のデータなど、今日では十分に信頼がおけないデータが存在しており、しばしば確率的モデリングが用いられる。また、データマイニングの結果も確率的に表現されることが多い。このような状況を受け、近年確率的データベースに関する研究が盛んになっている [10]。例を図 2 に示す。ここでは、研究者と彼らの学会への所属情報を表現している。各関係表において、属性 `Pr` は各タプルの存在確率を示す。ここでは、関係 `Researchers` における研究者の所属は排他的である（研究者は同時に複数機関に所属しない）とモデル化している。

各大学および企業が、IEEE にどの程度関与しているかの確信度を求める問い合わせを以下に示す。

```
SELECT x.Affil, confidence()
FROM Researchers x, Memberships y
WHERE x.Name = y.Name and y.Membership = 'IEEE'
GROUP BY x.Affil
```

確信度の計算法の詳細については省略するが、複雑なデータベース問い合わせに対する確率的推論を含む問い合わせの表現・処理技術は、この数年で急速に進展している [10]。

3.2 BayesStore：確率的推論機構の統合

確率的推論機能をデータベースシステムの内部に組み込むという意欲的な試みも進められている。BayesStore [11] では、データベースシステムの問い合わせ処理機構にベイジアンネットワークによる推論機構を組み込んでいる。

図 3 に BayesStore のデータ表現の例を示す。図 3 左では、各部屋の各センサごとに気温と明るさのセンシング結果を保持している。BayesStore では、これらのセンサデータに加え、センサ間の確率的な依存関係を表すベイジアンネットワークも合わせて管理する。図 3 右は、部屋 1 のセンサ 1 の明るさがセンサ 1 の気温に確率的に依存することを前提に記述される条件付確率表（CPT）である。大規模なデータに対しては CPT が非常に大きくなるため、BayesStore ではそのサイズを削減するための工夫がなされている。

Sensor				CPT (Light Temp)		
Room	Sid	Temp	Light	(Sensor 1 at Room 1)		
1	1		Drk	Light	Temp	Prob
1	2	Hot		Brk	Hot	0.8
1	3	Cold	Drk	Drk	Hot	0.2
2	1		Brk	Drk	Cold	0.8
2	2			Brk	Cold	0.2
2	3	Hot				

図3 BayesStore のデータ表現

Location				Movement					
tag	t	l	p(t)	tag	t	t'	l	l'	p(l' l)
Bob	7	H1	0.8	Bob	7	8	H1	H1	0.625
Bob	7	H2	0.15	Bob	7	8	H1	H2	0.25
Bob	7	O1	0.05	Bob	7	8	H1	O1	0.125
Bob	8	H1	0.50	Bob	7	8	H1	O2	0.0
		⋮		Bob	7	8	H2	H1	0.0
		⋮		Bob	7	8	H2	H2	0.666
		⋮					⋮		

図4 Markovian Streams のデータ表現

BayesStore では、関係代数 [1] の拡張による確率的な推論を伴う問い合わせの機能が実現されている。例えば、上記 Sensor 表から気温が低い部屋を探す問い合わせ $\sigma_{\text{Temp} = \text{'Cold'}}(\text{Sensor})$ を実行すると、通常の関係代数では 3 番目のタプルのみが選択される。一方 BayesStore の場合には、明かりが暗い部屋は気温が低い可能性があるという依存関係をもとに確率的推論が行われ、例えば 1 行目のタプルも選択される。このようなデータ問い合わせ機能以外に、確率モデル自体を問い合わせする機能も提供されている。

3.3 Markovian Streams: 確率的データストリーム

確率的データベースシステムではないが、センサデータから得られる確率的情報の活用という観点から、Markovian Streams [12] を紹介する。このプロジェクトでは、RFID タグにより取得したセンサデータを確率モデルに基づいて確率的ストリームの形式に洗練・加工する。図4にこの例を示す。この例は、ユーザ Bob に RFID タグを付与し、複数のセンサでその行動をモニタリングした結果から、粒子フィルタ (particle filter) [9] を用いて得られた移動軌跡を示している。Location 表は、例えば時刻 7 に Bob が部屋 (l) H1 にいた確率が 0.8 であるといった、存在確率を表す。一方、Movement 表は、時刻 7 に Bob が部屋 H1 にいたとき時刻 8 に部屋 H2 に移動した確率が 0.25 であることを示している。このような確率的データは、ユーザの移動に応じて継続的に得られるため、一種のデータストリームとなる。

Markovian Streams では、このような確率的データストリームに対する問い合わせを支援している。図5は、「Bob が部屋 H2 から部屋 O2 に移動した」というイベントを検

図5 Markovian Streams における問い合わせ

出する問い合わせを表しており、非決定性オートマトンで記述されている。このプロジェクトでは、このような問い合わせを効率的に支援するため、確率的データストリームに対する索引機構の開発などを行っている。

この節では、確率的データベースに対する様々なアプローチについて述べた。今後の社会では、センシングやデータマイニングなどによるデータがさらに膨大に得られると考えられ、確率的な情報をスケラブルに扱える確率的データベース技術が重要になるものと思われる。

4. 移動オブジェクトデータベース

データベースの研究分野でロボティクスに関連が深い領域として、移動オブジェクトデータベース [13] について紹介する。移動オブジェクトとは、自動車、歩行者など、それ自身が移動するデータを指し、移動ロボットなども含む広い概念である。移動オブジェクトデータベースは、移動オブジェクトそれ自身、また、移動オブジェクトを支える周辺環境に関する情報を表現し、蓄積・活用しようとするデータベースである。地図データなど、複数次元の空間上のデータを扱う空間データベース [14] の一分野とも考えられ、また、モバイルデータベース (モバイル環境におけるデータベース) とも関連が深い。

移動オブジェクトデータベースに関する代表的な研究テーマとしては、移動オブジェクト表現のためのデータモデル、様々な条件で移動オブジェクトを検索する問い合わせ処理、移動予測・移動分析、移動軌跡データからのデータマイニングなどがある。移動オブジェクトデータベースに対する問い合わせの例としては、例えば「10 分後に、指定する領域内に存在する自動車をすべて求めよ」などがある。この場合、データベース中には多数の自動車の現在位置と移動予測に必要な情報が管理されており、それらを用いた問い合わせを行うことになる。なお、先に述べた Markovian Streams のフレームワークも、移動オブジェクトに対する問い合わせ処理の一種ととらえることができる。

ここでは、ロボティクスにも関連し、前章で述べた確率的データベースとも関連する話題として、筆者のうち石川の研究を紹介する [15]。移動ロボットが逐次的に、自身のセンサや移動履歴をもとに自己の位置を確率的に推定する状況を考える。ロボットの移動を、ガウス過程に基づくカルマンフィルタ [9] により表現・追跡する場合、各時点のロボットの推定位置はガウス分布を用いて曖昧に表現される

図 6 曖昧な位置に基づく範囲問い合わせ

ことになる。ここで、ロボットから距離 10 [m] 以内にいる目標物（障害物や獲得すべきアイテムなど）の情報を検索したいとする。ロボットの内部では目標物のデータが、その座標値（ここでは簡単のため点データとする）とともに管理されているとする。

この問い合わせは、ロボットの位置が正確に知られている場合には容易に処理することができる。ロボットの位置と各目標物の距離を計算し、10 [m] 以下であるかを判断すればよい。しかし、ロボットの位置がガウス分布で表現される場合、ロボットの位置と目標物の距離は近いこともある（図 6）。

このような問題に対して文献 [15] では、範囲問い合わせの概念を拡張し、「10 [m] 以内に位置する確率が閾値以上であるような目標物をすべて検索せよ」と、確率的な範囲問い合わせを考えている。この場合、各目標物が条件を満たすかどうかの検証のため、コストの高い数値積分が必要となるが、候補削減のための工夫や索引技術を活用することにより、必要最小限の目標物にのみ数値積分を適用する、効率的なアルゴリズムを提案している。曖昧な位置情報の利用は、ロボティクスに限らずモバイル環境においても重要な課題となっており、興味深い研究課題となっている。

5. おわりに

本稿では、データベースの研究分野の中から、ロボティクスに関連の深い領域について、最近の話題をいくつか取り上げて紹介した。大規模センサ情報の活用のためのセンサデータベースについて述べ、関連する領域として確率的データベースを述べた。また、移動ロボットと関連する移動オブジェクトデータベースについても紹介した。今後、セ

ンサ情報のさらなる増加や、ロボット自体が蓄積・活用するデータの量が増大することが考えられ、ロボティクスにおいてもデータベース技術の活用が求められてくると考えられる。将来的には、ロボットの行動認識や移動計画のモジュールにデータベースシステムが統合されていくこともあり得るであろう。そのような状況においては、本稿で述べたセンサデータベース、確率的データベース、移動オブジェクトデータベースなどの技術の蓄積は大いに役立つに違いない。近い将来には両分野の連携による研究・開発も重要となるであろう。

謝 辞 本研究の一部は科学研究費（21013023）による。

参 考 文 献

- [1] H. Garcia-Molina, J.D. Ullman and J. Widom: Database Systems: The Complete Book, 2nd ed. Prentice Hall, 2008.
- [2] 科学研究費特定領域研究「情報 IT 爆発基盤」ホームページ: <http://www.infoplosion.nii.ac.jp/info-plosion/>
- [3] 柴山, 鳥澤, 田浦, 河野 (編): 特集「情報爆発時代におけるわくわくする IT の創出を目指して」, 情報処理, vol.49, no.8, pp.1-77, 2008.
- [4] V. Kumar (ed.): "Special Section on Sensor Network Technology and Sensor Data Management," ACM SIGMOD Record, vol.32, no.4, pp.16-82, 2003.
- [5] 安藤, 田村, 戸辺, 南: センサネットワーク技術: ユビキタス情報環境の構築に向けて, 東京電機大学出版局, 2005.
- [6] TinyDB ホームページ: <http://telegraph.cs.berkeley.edu/tinydb/>
- [7] S.R. Madden, M.J. Franklin, J.M. Hellerstein and W. Hong: "TinyDB: An aquisitional query processing system for sensor networks," ACM TODS, vol.30, no.1, pp.122-173, 2005.
- [8] A. Deshpande, C. Guestrin, S.R. Madden, J.M. Hellerstein and W. Hong: "Model-based approximate querying in sensor networks," VLDB Journal, vol.14, no.4, pp.417-443, 2005.
- [9] S. Thrun, W. Burgard and D. Fox: 確率ロボティクス, 毎日コミュニケーションズ, 2007.
- [10] N. Dalvi, C. Ré and D. Suciu: "Probabilistic databases: diamonds in the dirt," CACM, vol.52, no.7, pp.86-94, 2009.
- [11] D.Z. Wang, E. Michelakis, M. Garofalakis and J.M. Hellerstein: BayesStore: "Managing large, uncertain data repositories with probabilistic graphical models," pp.340-351, Proc. of Very Large Data Bases (VLDB), 2008.
- [12] J. Letchner, C. Ré, M. Balazinska and M. Philipose: "Challenges for event queries over Markovian streams," IEEE Internet Computing, vol.12, no.6, pp.30-36, 2008.
- [13] R.H. Güting and M. Schneider: Moving Objects Databases. Morgan Kaufmann, 2005.
- [14] P. Rigaux, M. Scholl and A. Voisard: Spatial Databases with Application to GIS. Morgan Kaufmann, 2001.
- [15] Y. Ishikawa, Y. Iijima and J.X. Xu: "Spatial range querying for Gaussian-based imprecise query objects," Proc. of International Conference on Data Engineering (ICDE), pp.676-687, 2009.

石川佳治 (Yoshiharu Ishikawa)

筑波大学大学院工学研究科単位取得退学, 奈良先端科学技術大学院大学助手, 筑波大学講師, 助教授を経て, 現在名古屋大学情報基盤センター教授, 博士 (工学), 日本データベース学会, 情報処理学会, 電子情報通信学会, 人工知能学会, ACM, IEEE 各会員,

(日本ロボット学会正会員)

喜連川優 (Masaru Kitsuregawa)

東京大学大学院工学系研究科情報工学専攻修了, 同大学生産技術研究所教授, 戦略情報融合国際研究センター長, 専門はデータベース工学, 文部科学官, 情報処理学会副会長, 経済産業省情報大航海プロジェクト戦略会議委員長, 文部科学省特定領域研究「情報爆発」代

表等を務める。