

---

制約問合せ処理の最適化と  
制約データベースの枠組みに関する研究

---

(課題番号 17500097)

平成17年度～平成19年度科学研究費補助金  
( 基盤研究(C) )研究成果報告書

平成20年 5月

研究代表者 北上 始

広島市立大学大学院情報科学研究科教授

広島市立大学附属図書館



0002985824

## はしがき

本研究は、平成17年度から平成19年度までの3年間にわたって交付された科学研究費補助金（基盤研究（C））による研究課題「制約問合せ処理の最適化と制約データベースの枠組みに関する研究」（課題番号 17500097）の成果を取りまとめたものである。

近年、高性能PCの低価格化や情報ネットワークの浸透が牽引となり、インターネットに接続されたさまざまな情報システムが開発され、人と人、人とコミュニティ、人とモノ、コミュニティどうし、モノどうしなどでやりとりされるコミュニケーションの効率化が図られている。例えば、e-ビジネスの分野では、CRM、SCM、ERPなどを支援する大規模で複雑な情報システムが多数出現している。しかしながら、取り扱われるデータは巨大化の一途をたどっており、また、扱うデータが異種性をもつと同時に分散しているため、将来、従来の情報システムで対処することが処理コストや人間の作業効率などの面で困難になってくるものと考えられる。本研究では、制約データベースシステムを情報システムのエンジンとみなし、上記の問題を解決する。問題解決に当たっては、制約問合せ処理に含まれる制約問題として、(1)文字列制約、(2)ネットワーク制約、(3)計算幾何学的制約、(4)空間的制約、(5)数式表現制約の5つの枠組みを中心に研究を行った。

本研究は、研究分担者、その他の関係者との共同研究であり、ご協力頂いた関係各位に心からお礼を申し上げますと共に、本研究成果として得られた知見が今後の情報社会になんらかの形で役立つことを切望するものである。

平成20年5月

研究代表者 北上 始

(広島市立大学大学院情報科学研究科・教授)

## 目次

研究組織	-1-
研究発表	-2-
研究成果概要	-8-
研究成果	-11-
[1] 高木 允, 森 康真, 田村 慶一, 北上 始: ブログユーザ空間からの重複を許した頻出コミュニティ抽出法, 情報処理学会論文誌: 数理モデルと応用, Vol. 49, No. SIG4(TOM20), pp.93-104, 2008年3月.	-13-
[2] 田村 慶一, 森 康真, 北上 始: Extremal Optimization による調停グラフの交差数減少, 情報処理学会論文誌: 数理モデルと応用, Vol. 49, No. SIG4(TOM20), pp.105-116, 2008年3月.	-26-
[3] 浮田 健太, 黒木 進, 森 康真, 北上 始: 空間オブジェクト集合からの頻出近接属性パターン的高速抽出法, 電子情報通信学会 第19回データ工学ワークショップ(DEWS2008), Online Proceedings ( <a href="http://www.ieice.org/iss/de/DEWS/">http://www.ieice.org/iss/de/DEWS/</a> ), 7 pages, 2008年3月9日~3月11日.	-39-
[4] Makoto Takaki, Keiichi Tamura, Toshihide Sutou, and Hajime Kitakami: A New Dynamic Load Balancing Technique for Parallel Modified PrefixSpan with Distributed Worker Paradigm and Its Performance Evaluation, Proceedings of the Sixth International Symposium on High Performance Computing (ISHPC-VI), Lecture Notes in Computer Science (LNCS) Published in 2008, Springer-Verlag, Vol.4759, pp.227-237, Nara New Public Hall, Higashikasugano, Nara city, Japan, September 7-9, 2005.	-46-
[5] 荒木 康太郎, 田村 慶一, 加藤 智之, 北上 始: ミスマッチクラスタに対する最小汎化パターン抽出方式, 日本データベース学会論文誌 (DBSJ Letters), Vol.6, No.3, pp.5-8, 2007年12月.	-57-
[6] Kotaro Araki, Keiichi Tamura, Tomoyuki Kato, Yasuma Mori, and Hajime Kitakami: Extraction of Ambiguous Sequential Patterns with Least Minimum Generalization from Mismatch Clusters, Proceedings of The Third International Conference on Signal-Image Technology & Internet-Base system (SITIS' 2007): Track Information Management & Retrieval Technologies (IMRT), IEEE Computer Society Press, pp.32-39, Jiangong Jinjiang Hotel, Shanghai, China, December 16-19 in 2007.	-61-
[7] Makoto Takaki, Keiichi Tamura, Yasuma Mori, and Hajime Kitakami: A Extraction Method of Overlapping Cluster based on Network Structure Analysis, The 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, Workshops on Intelligent Web Interaction 2007 (IWI2007), IEEE Computer Society Press, pp.212-216, Silicon Valley in USA, November 2-5 in 2007.	-69-

- [8] Makoto Takaki, Yasuma Mori, Keiichi Tamura, Susumu Kuroki and Hajime Kitakami: Method for Extracting Frequent Communities from Blog User Spaces, The 2007 International Conference on Parallel & Distributed Processing Techniques & Applications (PDPTA'07), Vol.II, pp.773-779, Monte Carlo Hotel, Las Vegas, Nevada in USA, July 2007. -74-
- [9] 高木 允, 森 康真, 田村 慶一, 黒木 進, 北上 始: ブログデータ集合から頻出なコミュニティ抽出手法, 日本データベース学会論文誌 (DBSJ Letters), Vol.6, No.1, pp.189-192, 2007年6月. -81-
- [10] 加藤 智之, 森 康真, 荒木 康太郎, 黒木 進, 北上 始: 可変長配列パターン抽出法におけるギブスサンプリングを用いた不要パターンの除去方式, 日本データベース学会論文誌 (DBSJ Letters), Vol.6, No.1, pp.65-68, 2007年6月. -85-
- [11] 北上 始: 日本のデータベース研究最前線, DNA配列の謎を解き明かす配列データマイニング, 月刊 DB マガジン, 翔泳社, pp.172-173, 2006年5月. -89-
- [12] 田村 慶一, 北上 始, 高木 允: Normalized Cutによるセグメント分割を用いた大規模無向グラフの並列クラスタリング手法, 先進的計算基盤システムシンポジウム SACSIS 2007, 情報処理学会シンポジウムシリーズ, ISSN 1344-0640, Vol.2007, No.5, pp.150-151, 2007年5月. -91-
- [13] 加藤 智之, 北上 始, 森 康真, 田村 慶一, 黒木 進: 極小かつ非冗長な可変長ワイルドカード領域を持つ頻出パターンの抽出, 電子情報通信学会論文誌 D「データ工学特集号」, Vol.J90, No.2, pp.281-291, 2007年2月. -94-
- [14] Makoto Takaki, Keiichi Tamura, and Hajime Kitakami: Dynamic Load Balancing Technique for Modified PrefixSpan on a Grid Environment with Distributed Worker Model, Proceedings of The 2006 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'06 & RTCOMP'06), Vol. II, Las Vegas, Nevada, USA, pp.895-901, June 26-30, 2006. -106-
- [15] 加藤 智之, 北上 始, 森 康真, 田村 慶一, 黒木 進: 極小な可変長ワイルドカード領域をもつ頻出配列パターンの抽出, 日本データベース学会論文誌(DBSJ Letters), Vol.5, No.1, pp.117-120, 2006年6月. -113-
- [16] Tomoyuki Kato, Hajime Kitakami, Makoto Takaki, Keiichi Tamura, Yasuma Mori, and Susumu Kuroki: Extraction for Frequent Sequential Patterns with Minimum Variable-Wildcard Regions, Proceedings of The 2006 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'06 & RTCOMP'06), Vol. II, Las Vegas, Nevada, USA, pp.825-831, Monte Carlo Hotel, Las Vegas, Nevada in USA, June 26-30, 2006. -117-

- [17] 田村 慶一, 岩木 稔, 高木 允, 北上 始: PC クラスタにおける混合整数計画問題の並列処理とその性能評価, 情報処理学会論文誌: 数理モデル化と応用, Vol. 46, No.SIG 17(TOM 13), pp.56-69, 2005 年 12 月. -124-
- [18] Makoto Takaki, Keiichi Tamura, Toshihide Sutou and Hajime Kitakami: New Dynamic Load Balancing for Parallel Modified PrefixSpan with Distributed Worker Paradigm, Proceedings of International Special Workshop on Databases for Next Generation Researchers in Memoriam of Prof. Kambayashi (SWOD2005), pp.96-99, Tokyo in Japan, April 2005, also to be appeared in Proc. of ICDE Workshops, IEEE Computer Society Press, National Center of Science, Tokyo, Japan, p.1243, April 4, 2005. -137-
- [19] 高木 允, 田村 慶一, 周藤 俊秀, 北上 始: 並列 Modified PrefixSpan 法の並列化と動的負荷分散手法, 情報処理学会論文誌: 数理モデル化と応用, Vol. 46, No.SIG 10 (TOM 12), pp.138-152, 2005 年 6 月. -142-
- [20] 田村慶一, 高木允, 北上始: 遺伝的プログラミングによる調停グラフ交差数減少のための並列分散処理, FCS/MPS シンポジウム, 情報処理学会, pp.271-278, 2005 年 10 月. -156-
- [21] 劉 強, 北上 始, 黒木 進, 田村 慶一, 森 康真: 画像データベースにおける索引構造の効率化, 電子情報通信学会・第 16 回データ工学ワークショップ (DEWS2005), 電子情報通信学会データ工学研究専門委員会, Online Proceedings (<http://www.ieice.org/iss/de/DEWS/>), 2005 年 3 月. -164-

## 研究組織

研究代表者：北上 始（広島市立大学大学院情報科学研究科・教授）

研究分担者：黒木 進（広島市立大学大学院情報科学研究科・准教授）

研究分担者：田村 慶一（広島市立大学大学院情報科学研究科・助教）

研究分担者：森 康真（広島市立大学大学院情報科学研究科・助教）

## 交付決定額(配分額)

(金額単位：千円)

	直接経費	間接経費	合計
平成17年度	900	0	900
平成18年度	900	0	900
平成19年度	700	210	910
総計	2,500	210	2,710

## 研究発表

### (1) 雑誌論文(査読付き)

1. 高木 允, 田村 慶一, 周藤 俊秀, 北上 始: 並列 Modified PrefixSpan 法の並列化と動的負荷分散手法, 情報処理学会論文誌: 数理モデル化と応用, Vol. 46, No.SIG 10 (TOM 12), pp.138-152, 2005 年 6 月.
2. 田村 慶一, 岩木 稔, 高木 允, 北上 始: PC クラスタにおける混合整数計画問題の並列処理とその性能評価, 情報処理学会論文誌: 数理モデル化と応用, Vol. 46, No.SIG 17(TOM 13), pp.56-69, 2005 年 12 月.
3. 加藤 智之, 北上 始, 森 康真, 田村 慶一, 黒木 進: 極小な可変長ワイルドカード領域をもつ頻出配列パターンの抽出, 日本データベース学会論文誌 (DBSJ Letters), Vol.5, No.1, pp.117-120, 2006 年 6 月
4. 加藤 智之, 北上 始, 森 康真, 田村 慶一, 黒木 進: 極小かつ非冗長な可変長ワイルドカード領域を持つ頻出パターンの抽出, 電子情報通信学会論文誌, データ工学特集号, Vol.J90, No.2, pp.281-291, 2007 年 2 月.
5. 加藤 智之, 森 康真, 荒木 康太郎, 黒木 進, 北上 始: 可変長配列パターン抽出法におけるギブスサンプリングを用いた不要パターンの除去方式, 日本データベース学会論文誌 (DBSJ Letters) , Vol.6, No.1, pp.65-68, 2007 年 6 月.
6. 高木 允, 森 康真, 田村 慶一, 黒木 進, 北上 始: ブログデータ集合から頻出なコミュニティ抽出手法, 日本データベース学会論文誌 (DBSJ Letters) , Vol.6, No.1, pp.189-192, 2007 年 6 月.
7. 荒木 康太郎, 田村 慶一, 加藤 智之, 北上 始: ミスマッチクラスタに対する最小汎化パターン抽出方式, 日本データベース学会論文誌 (DBSJ Letters), Vol.6, No.3, pp.5-8, 2007 年 12 月.
8. 高木 允, 森 康真, 田村 慶一, 北上 始: ブログユーザ空間からの重複を許した頻出コミュニティ抽出法, 情報処理学会論文誌: 数理モデルと応用, Vol. 49, No. SIG4(TOM20), pp.93-104, 2008 年 3 月.
9. 田村 慶一, 森 康真, 北上 始: Extremal Optimization による調停グラフの交差数減少, 情報処理学会論文誌: 数理モデルと応用, Vol. 49, No. SIG4(TOM20), pp.105-116, 2008 年 3 月.

## (2) 国際会議発表(査読付き)

1. Makoto Takaki, Keiichi Tamura, Toshihide Sutou and Hajime Kitakami: New Dynamic Load Balancing for Parallel Modified PrefixSpan with Distributed Worker Paradigm, Proc. of International Special Workshop on Databases for Next Generation Researchers in Memoriam of Prof. Kambayashi (SWOD2005), pp.96-99, Tokyo in Japan, April 2005, also to be appeared in Proceedings of ICDE Workshops, IEEE Computer Society, National Center of Science, Tokyo, Japan, p.1243, April 4, 2005.
2. Shigetaka Tono, Hajime Kitakami, Keiichi Tamura, Yasuma MORI, Susumu Kuroki: Efficiently Mining Sequence Patterns With Variable-Length Wildcard Regions Using An Extended Modified PrefixSpan Method, The 13th Annual International Conference on Intelligent Systems for Molecular Biology(ISMB2005), Poster No.G22, p.147, Renaissance Center Marriott, Detroit, Michigan in USA, June 25-29, 2005.
3. Hajime Kitakami, Makoto Takaki, Keiichi Tamura, Tomoyuki Kato, and Susumu Kuroki: The Minimum Generalization of Flexible Sequence Patterns Extracted by Prefix-Projected Pattern-Growth Approach, The 3rd RECOMB Comparative Genomics Satellite Workshop, Poster, Trinity College, Dublin in Ireland, September 18-20, 2005.
4. Makoto Takaki, Keiichi Tamura, Toshihide Sutou, and Hajime Kitakami: A New Dynamic Load Balancing Technique for Parallel Modified PrefixSpan with Distributed Worker Paradigm and Its Performance Evaluation, Proceedings of the Sixth International Symposium on High Performance Computing (ISHPC-VI), Lecture Notes in Computer Science (LNCS) Published in 2008, Springer-Verlag, Vol.4759, pp.227-237, Nara New Public Hall, Higashikasugano, Nara city, Japan, September 7-9, 2005.
5. Tomoyuki Kato, Hajime Kitakami, Makoto Takaki, Keiichi Tamura, Yasuma Mori, and Susumu Kuroki: Extraction for Frequent Sequential Patterns with Minimum Variable-Wildcard Regions, Proceedings of The 2006 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'06 & RTCOMP'06), Vol. II, Monte Carlo Hotel, Las Vegas, Nevada in USA, pp.825-831, June 26-30, 2006.



6. Makoto Takaki, Keiichi Tamura, and Hajime Kitakami: Dynamic Load Balancing Technique for Modified PrefixSpan on a Grid Environment with Distributed Worker Model, Proceedings of The 2006 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'06 & RTCOMP'06), Vol. II, Las Vegas, Nevada, USA, pp.895-901, Monte Carlo Hotel, Las Vegas, Nevada in USA, June 26-30, 2006.
7. Makoto Takaki, Yasuma Mori, Keiichi Tamura, Susumu Kuroki and Hajime Kitakami: Method for Extracting Frequent Communities from Blog User Spaces, The 2007 International Conference on Parallel & Distributed Processing Techniques & Applications (PDPTA'07), Vol.II, pp.773-779, Monte Carlo Hotel, Las Vegas, Nevada in USA, July 2007.
8. Makoto Takaki, Keiichi Tamura, Yasuma Mori, and Hajime Kitakami: A Extraction Method of Overlapping Cluster based on Network Structure Analysis, The 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, Workshops on Intelligent Web Interaction 2007 (IWI2007), IEEE Computer Society Press, pp.212-216, The Fremont Marriott, Silicon Valley in USA, November 2-5 in 2007.
9. Kotaro Araki, Keiichi Tamura, Tomoyuki Kato, Yasuma Mori, and Hajime Kitakami: Extraction of Ambiguous Sequential Patterns with Least Minimum Generalization from Mismatch Clusters, Proceedings of The Third International Conference on Signal-Image Technology & Internet-Base System (SITIS' 2007): Track Information Management & Retrieval Technologies (IMRT), IEEE Computer Society Press, pp.32-39, Jiangong Jinjiang Hotel, Shanghai, China, December 16-19 in 2007.

### (3) 学会発表(査読付き)

1. 劉 強, 北上 始, 黒木 進, 田村 慶一, 森 康真: 画像データベースにおける索引構造の効率化, 電子情報通信学会・第 16 回データ工学ワークショップ (DEWS2005), 電子情報通信学会データ工学研究専門委員会, Online Proceedings (<http://www.ieice.org/iss/de/DEWS/>), 九十九島観光ホテル (佐世保市), 2005 年 2 月 28 日~3 月 2 日.
2. 周藤 俊秀, 高木 允, 田村 慶一, 北上 始: グリッド環境下における Modified PrefixSpan 法の並列処理とその動的負荷分散方式, 先進的計算基盤システムシンポジウム(SACSYS2005), 情報処理学会, pp.161-168, つくば国際会議場 (つくば市), 2005 年 5 月 18 日~20 日.

3. 田村 慶一, 高木 允, 北上 始, 札埜 有香: 遺伝的プログラミングによる調停グラフの交差数減少の並列分散処理とそのグリッド環境下における性能評価, 先進的計算基盤システムシンポジウム (SAC SIS2005), 情報処理学会, pp.258-259, つくば国際会議場(つくば市), 2005年5月18日~20日.
4. 高木 允, 田村 慶一, 北上 始: 並列 Modified PrefixSpan 法のグリッド化とその動的負荷分散手法, FCS/MPS シンポジウム, 名古屋, pp.241-248, 2005年.
5. 田村 慶一, 高木 允, 北上 始: 遺伝的プログラミングによる調停グラフ交差数減少のための並列分散処理, FCS/MPS シンポジウム, 名古屋大学(名古屋市), pp.271-278, 2005年10月11日~13日.
6. 高木 允, 田村慶一, 北上 始: グリッド環境下での分散型ワーカモデルを用いた Modified PrefixSpan 法の動的負荷分散方式, 電子情報通信学会・第17回データ工学ワークショップ(DEWS2006), 電子情報通信学会データ工学研究専門委員会, Online Proceedings (<http://www.ieice.org/iss/de/DEWS/>), 沖縄コンベンションセンター(宜野湾市), 2006年3月1日~3日.
7. 加藤 智之, 北上 始, 森 康真, 田村 慶一, 黒木 進: 極小な可変長ワイルドカード領域を持つ頻出配列パターンの抽出, 電子情報通信学会・第17回データ工学ワークショップ (DEWS2006), 電子情報通信学会データ工学研究専門委員会, Online Proceedings (<http://www.ieice.org/iss/de/DEWS/>), 沖縄コンベンションセンター(宜野湾市), 2006年3月1日~3日.
8. 高木 允, 田村 慶一, 北上 始: Modified PrefixSpan 法のための PC クラスタ間の効率的な動的負荷分散方式にむけて, 先進的計算基盤システムシンポジウム(SAC SIS2006), 情報処理学会シンポジウムシリーズ, Vol.2006, No.5, pp.232-233, 大阪国際会議場(大阪市), 2006年5月.
9. 高木 允, 森 康真, 黒木 進, 北上 始: 時系列ログデータからの頻出な密部分グラフの抽出, データベースと Web 情報システムに関するシンポジウム(DBWeb2006), 情報処理学会, ポスター, 京都大学百周年時計台記念館(京都市), 2006年11月30日~12月1日.
10. 高木 允, 森 康真, 田村 慶一, 黒木 進, 北上 始: プログユーザ空間からの頻出なコミュニティ抽出法, 電子情報通信学会 第18回データ工学ワークショップ(DEWS2007), Online Proceedings, 8 pages, 広島プリンスホテル(広島市), 2007年2月28日~3月2日.
11. 加藤 智之, 森 康真, 荒木 康太郎, 黒木 進, 北上 始: 可変長配列パターン抽出法におけるギブスサンプリングを用いた不要パターンの除去方式, 電子情報通信学会・第18回データ工学ワークショップ(DEWS2007), 電子情報通信学会データ工学研究専門委員会, Online Proceedings (<http://www.ieice.org/iss/de/DEWS/>), 広島プリンスホテル(広島市), 2007年2月28日~3月2日.

12. 荒木 康太郎, 田村 慶一, 加藤 智之, 黒木 進, 北上 始: 曖昧検索に基づく最小汎化パターンの抽出法, 電子情報通信学会・第 18 回データ工学ワークショップ(DEWS2007), 電子情報通信学会データ工学研究専門委員会, Online Proceedings (<http://www.ieice.org/iss/de/DEWS/>), 広島プリンスホテル (広島市), 2007 年 2 月 28 日~3 月 2 日.
13. 田村 慶一, 北上 始, 高木 允: Normalized Cut によるセグメント分割を用いた大規模無向グラフの並列クラスタリング手法, 先進的計算基盤システムシンポジウム SACSIS 2007, 情報処理学会シンポジウムシリーズ, ISSN 1344-0640, Vol.2007, No.5, pp.150-151, 学術総合センター講堂・会議室 (東京都), 2007 年 5 月 23 日~25 日.
14. 澤田 祐介, 田村 慶一, 荒木 康太郎, 北上 始: PC クラスタ上でのディスクベースサフィックス木の並列構築方式, 第 9 回 IEEE 広島支部学生シンポジウム, CD-ROM, d-22, 4 pages, 鳥取大学 (鳥取市), 2007 年 11 月 24 日~25 日.
15. 荒木 康太郎, 田村 慶一, 加藤 智之, 北上 始: ミスマッチクラスタから最小汎化された曖昧配列パターン集合の抽出方式, 第 9 回 IEEE 広島支部学生シンポジウム, CD-ROM, d-18, 4 pages, 鳥取大学 (鳥取市), 2007 年 11 月 24 日~25 日.
16. 澤田 祐介, 田村 慶一, 荒木 康太郎, 高木 允, 北上 始: PC クラスタ上でのディスクベースサフィックス木の並列構築方式, 電子情報通信学会 第 19 回データ工学ワークショップ (DEWS2008), Online Proceedings (<http://www.ieice.org/iss/de/DEWS/>), 8 pages, 宮崎国際会議場 (宮崎市), 2008 年 3 月 9 日~3 月 11 日.
17. 浮田 健太, 黒木 進, 森 康真, 北上 始: 空間オブジェクト集合からの頻出近接属性パターンの高速抽出法, 電子情報通信学会 第 19 回データ工学ワークショップ (DEWS2008), Online Proceedings (<http://www.ieice.org/iss/de/DEWS/>), 7 pages, 宮崎国際会議場 (宮崎市), 2008 年 3 月 9 日~3 月 11 日.
18. 河野 修久, 加藤 智之, 田村 慶一, 北上 始: 配列データベースから類似部分配列を抽出するための GS 最適化手法に関する考察, 電子情報通信学会 第 19 回データ工学ワークショップ (DEWS2008), Online Proceedings (<http://www.ieice.org/iss/de/DEWS/>), 8 pages, 宮崎国際会議場 (宮崎市), 2008 年 3 月 9 日~3 月 11 日.
19. 木村 浩明, 荒木 康太郎, 田村 慶一, 澤田 祐介, 北上 始: 段階的一般化法によるミスマッチクラスタからの最小汎化パターンの抽出, 電子情報通信学会 第 19 回データ工学ワークショップ (DEWS2008), Online Proceedings (<http://www.ieice.org/iss/de/DEWS/>), 8 pages, 宮崎国際会議場 (宮崎市), 2008 年 3 月 9 日~3 月 11 日.

#### (4) 学会発表(査読なし)

1. 高木 允【学生奨励賞】，田村 慶一，北上 始：分散型ワーカモデルを用いたグリッド環境における Modified PrefixSpan 法の動的負荷分散方式，第 68 回全国大会（学会創立 45 周年記念大会）講演論文集(3)，pp.149-150，工学院大学 新宿キャンパス（東京都），2006 年 3 月 7 日～9 日。
2. 高木 允，田村 慶一，北上 始：グリッド環境下での分散型ワーカモデルを用いた Modified PrefixSpan 法の動的負荷分散方式，情報処理学会・数理モデル化と問題解決研究会，pp.51-57，2006 年 6 月 26 日～29 日。
3. 加藤 智之，北上 始，高木 允，田村 慶一，森 康真，黒木 進：可変長ワイルドカード領域を持つ極小な頻出配列パターンの抽出，情報処理学会・数理モデル化と問題解決研究会，pp.9-14，2006 年 6 月 26 日～29 日。
4. 高木 允，森 康真，田村 慶一，黒木 進，北上 始：プログラマー空間からの頻出な部分グラフの抽出，情報処理学会第 69 回全国大会講演論文集（第 1 分冊），pp.373-374，早稲田大学 大久保キャンパス（東京都），2007 年 3 月 6 日～8 日。
5. 荒木 康太郎，田村 慶一，加藤 智之，黒木 進，北上 始：最小汎化された曖昧な頻出配列パターンの抽出，情報処理学会第 69 回全国大会講演論文集（第 4 分冊），pp.671-672，早稲田大学 大久保キャンパス（東京都），2007 年 3 月 6 日～8 日。
6. 高木 允【学生奨励賞】，田村 慶一，森 康真，北上 始：ネットワーク構造解析に基づく重複クラスタ抽出法，情報処理学会 第 142 回 データベースシステム研究会・第 87 回 情報学基礎研究会，pp.39-45，東京工業大学 大岡山キャンパス（東京都），2007 年 5 月 31 日（木）～ 6 月 1 日（金）。
7. 高木 允，森 康真，田村 慶一，黒木 進，北上 始：頻出コミュニティのプログラマー空間からの抽出手法，情報処理学会・数理モデル化と問題解決研究会，pp.33-36，2007 年 6 月 25 日。
8. 高木 允【プレゼンテーション賞】，田村 慶一，森 康真，北上 始：プログラマー空間からの重複を許した頻出コミュニティ抽出法，情報処理学会 第 66 回数理モデル化と問題解決研究会，pp.49-52，名古屋大学 ベンチャービジネスラボラトリ（名古屋市），2007 年 9 月 3 日～4 日。

#### (5) 図書

1. 北上 始：日本のデータベース研究最前線，DNA 配列の謎を解き明かす配列データマイニング，月刊 DB マガジン，翔泳社，pp.172-173，2006 年 5 月。

## 研究成果概要

制約データベースに対する制約問合せ処理では、制約問題を解く方法の研究が中心的な課題である。研究開始時点では、制約データベース処理に関係するさまざまな枠組みの制約問題に着目し、調査を行ったが、研究を具体的にすすめるに当たり、(1)文字列制約、(2)ネットワーク制約、(3)計算幾何学的制約、(4)空間的制約、(5)数式表現制約の5つの枠組みに着目した。そして、5つの枠組みのそれぞれに対して、代表的な応用例を挙げ、制約問題を解く方法について研究を行った。また、扱うデータの大規模化に対処するために、上記の制約問題のいくつかに対して、並列分散環境下で処理する方法の研究も行った。以下、本研究で得られた成果について述べる。ただし、文中に表記する引用論文の番号は、本報告書の目次に掲載されている既発表論文の番号とする。

### ■文字列制約問題

文字列（あるいは単語列）が蓄積されているデータベースから頻出部分文字列を抽出するという文字列制約問題を解くために、生命情報科学分野におけるアミノ酸配列や遺伝子配列などの分子配列データベースに着目し、可変長ワイルドカード領域や曖昧文字表現などを含む正規表現の頻出パターンを抽出する方法の研究を行った。その結果、可変長ワイルドカード領域については、従来の可変長頻出パターン抽出の手法では見落されていた頻出パターンを抽出することができるようになった<sup>[11], [13], [15], [16]</sup>。また、抽出されるパターンにおいて、明らかに不要と見なされるパターンを除去するために、ギブスサンプリングを応用する方法を明らかにした<sup>[10]</sup>。曖昧文字表現については、曖昧検索（類似性検索）の結果として得られる部分文字列の集合（ミスマッチクラスタと呼ばれる）に対して汎化処理を施し、ミスマッチクラスタの全体像を把握する方法を明らかにした<sup>[5], [6]</sup>。

上記の可変長頻出パターンを抽出する処理を高速化するために、並列分散処理の方法を明らかにした。具体的には、可変長頻出パターン抽出処理で生成される列挙木は平衡木ではなく、どの節点についてもそれを頂点とする部分木の深さを予め見積もることができないという問題に着目し、PC クラスタ環境下でマスタ・タスク・ステイル法（マスタワーカモデル）<sup>[19]</sup>やキャッシュベースドランダム・ステイル法（分散型ワーカモデル）<sup>[4], [18]</sup>を開発した。さらに、グリッド環境下における負荷分散方法についても検討を行い、キャッシュベースドマルチキャスト・ステイル法と呼ばれるネットワークの遅延に左右されにくい動的負荷分散方法を開発した<sup>[11], [14]</sup>。

### ■ネットワーク制約

ネットワーク状に接続されている物事の関係が蓄積されているデータベースを理解するというネットワーク制約問題を解くために、プログユーザ空間に

着目し、クラスタリングの研究を行った。その結果、ブログユーザ空間において、ブログどうしのつながりを表すネットワークを抽出し、そのネットワークデータに対して、*Newman* のアルゴリズムを適用することにより、ブログユーザのコミュニティを抽出する方法を明らかにした<sup>[8],[9]</sup>。また、この方法では、重複を許すコミュニティの抽出には直接つながらないことから、*Newman* のアルゴリズムを改良する方法についても明らかにした<sup>[11],[7]</sup>。

その他、大規模ネットワークからのコミュニティ抽出に応用可能と考えられる *Normalized Cut* によるクラスタリングアルゴリズムの並列処理方式についても明らかにした<sup>[12]</sup>。

### ■空間的制約

空間情報を蓄積するデータベースの空間的制約問題を解くために、携帯電話による Web アクセス履歴などに着目し、2次元空間に広がる空間オブジェクトの点座標データを蓄積するデータベースから頻出な近接属性パターンを高速に抽出する方法の研究を行った。その結果、点座標データの集合を空間分割する方法を明らかにした<sup>[3]</sup>。また、画像データベースを対象に類似画像を効率的に検索するために、従来の索引構造を改良する方法の検討もおこなった<sup>[21]</sup>。

### ■計算幾何学的制約

二種類の異なる接続関係をもつ木構造データどうしをつなぎ合わせた調停グラフは、一般に多くの交差が含まれる。それぞれの木の接続関係を維持しつつ、同じ親をもつ辺を入れ替えることにより、この交差を減少させる処理は NP 完全であり、計算幾何学的な制約問題である。ここでは、この問題を解くために、分子進化系統樹と生物分類樹とを同じ種どうしでつなぎ合わせた調停グラフに着目した。その結果、この問題を並列分散型の遺伝的プログラミングにより解決する方法を明らかにした<sup>[20]</sup>。また、*Extremal Optimization* と呼ばれる最適化手法を用いる方が単体性能の向上に有効であることがわかった<sup>[2]</sup>。

### ■数式表現制約

数式表現制約問題を解くために、何種類かの仕事を何台かの機械で処理する工場における機械スケジューリング問題に着目した。処理する機械ごとにかかる仕事の処理時間などはデータベースに蓄積されている。機械スケジューリング問題は、混合整数計画法により解くことが可能だが、整数変数の数により大きく処理時間が左右されることが知られている。共有メモリー型並列計算機により計算時間を短縮する研究は、すでに行われているが、複数台の PC をネットワークにより結合した PC クラスタによる並列計算の方法は十分には研究されていない。ここでは、PC クラスタ環境でマスタ・タスク・ステイル法 (マスタワーカモデル) を応用することにより、高速に計算する方法を開発した<sup>[17]</sup>。