

機関番号：15401  
 研究種目：基盤研究（C）  
 研究期間：2007～2010  
 課題番号：19500123  
 研究課題名（和文） 個人情報保護する時空間データマイニングシステムとアルゴリズム  
 研究課題名（英文） Spatio-Temporal Data Mining Systems and Algorithms that Preserves Privacy  
 研究代表者  
 森本 康彦（MORIMOTO YASUHIKO）  
 広島大学・大学院工学研究院・准教授  
 研究者番号：00363010

## 研究成果の概要（和文）：

近年、個人情報保護に対する意識が高まっているが、旧来のマイニング技術の多くが個人情報保護の観点からは機能的に不十分であった。本研究では、データマイニングの過程で、個人情報を開示することのないような知識表現や計算方法の開発をすることで、この問題を解決した。具体的にはこの研究期間をとおしてグループ単位のパターン発掘技術や問合せ機能が開発できた。これらのマイニング技術で開示される情報はグループ単位の集計情報となっているため、デリケートな時空間情報が他者に知られることはない。

## 研究成果の概要（英文）：

Recently, we have to be aware of individual's privacy when we handle databases. However, conventional data mining functions cannot prevent individual's privacy well enough. The proposed research has been trying to solve this privacy problem of data mining. We have developed privacy preserving pattern mining functions and query functions. Those functions use groups instead of individual records and do not have to output individual records' values such as exact spatio-temporal information of individuals.

## 交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	900,000	270,000	1,170,000
2008年度	800,000	240,000	1,040,000
2009年度	800,000	240,000	1,040,000
2010年度	800,000	240,000	1,040,000
年度			
総計	3,300,000	990,000	4,290,000

研究分野： データベース工学

科研費の分科・細目： 情報学・知能情報学

キーワード： データマイニング、プライバシー、地理情報システム、時空間データベース

## 1. 研究開始当初の背景

大規模データベースからの知識発掘技術として「データマイニング」はこれまで大いなる発展を遂げ、様々な分野で応用されてき

た。その代表的な機能として「相関ルール発掘」があるが、これはある異なる2つ以上の事象が同時に発現するというパターンを発見するものである。このようなパターンは元々、併売の組み合わせ（例えば「ビールと

オムツが併売されやすい」など)を例に考案され、実際にこの技術を使って得られた知見は、商品の棚割りや仕入れ計画策定等に活用されてきた。しかし、競争の激しいマーケティングの現場では(特に電子商取引分野など、周りをとりまく環境が激しく変化している分野では)未来の仕入れ戦略や販売戦略を考えることが求められ、そのためには時間軸上の順序を考慮した知見が必要となる。「時系列パターン発掘」技術はこのような要求を満足させるために開発された。

データマイニングへのニーズが高まる一方で、これまで開発され利用されているマイニング機能の大部分は個別の取引記録、つまり個人単位のデータが入力となることを前提としている。特に、前述の時系列パターン発掘技術では、ある商品群を購入した個人が、後日の取引において何を購入しているかという情報、つまり個人識別子情報が不可欠である。

一方で、平成15年に個人情報保護法が成立するなど、近年では個人情報保護の立場から、個人識別子を含むデータ、個人単位での細かい取引記録等のデータ利用には細心の注意が必要となっている。

データマイニングの対象となるデータは、当初、当事者の意思とは関係なく収集され、解析されてきた。また、発見される知識の利用方法に対してもデータ所有者の裁量に任されていた。そのため、データマイニングアルゴリズム研究の多くは、その計算効率性や発見される知識の有用性にばかりに重点が置かれていた。その結果、旧来のマイニング技術のほとんどが個人情報保護の観点からは機能的に不十分となり、それゆえにデータマイニング技術を応用することができないというケースも増えてきていた。

## 2. 研究の目的

個人情報保護の観点から、データマイニングを目的として提供されるデータ内容が個人単位ではなく、集計済みのグループ単位のデータに変わってきている。例えば、2005年に、ある企業から我々に研究用として提供されたPOSデータは、元は個人単位の詳細な取引データであったが、分析目的で社外に持ち出すにあたり、店舗および日付毎の商品別販売数量、販売金額の集計データに変換されていた。このような状態であれば、このデータから個人単位の情報を特定することはできない。

本研究では、個人単位の情報を特定できない情報源から、できる限り有用なパターンを発見する技術の開発を目的としたものである。

とくに戦略的なデータマイニングを行うためには時間情報や位置情報を含むデータから知識を発掘してゆく必要があるが、特に時間情報や位置情報はデリケートな個人情報なので扱いに慎重にならなくてはならない。そのため、時空間情報を含むデータベースからのデータマイニングにおける個人情報保護手法について研究開発してゆくことを重要課題とした。

## 3. 研究の方法

データをある程度のグループ単位で集計し、集計情報として保存管理することで、個人レベルの情報を特定できない形にできる。集計情報からは個人レベルの情報を特定できないため、このような状態のデータをマイニング対象とすることができれば個人情報を保護することができる。

従来のデータマイニング技術の多くは、個人レベルの情報を必要とするが、研究代表者らは、集計情報の時系列の変化などに着目したパターン発掘技術を開発した。

その一方で、集計情報からのスカイライン集合問い合わせ機能の開発も行った。

データベース中のデータの中で、いずれかの属性値で他のデータに比べ良いものを含むデータを「スカイライン」と呼び、それを列挙する機能を「スカイライン問合せ」と呼ぶ。データ収集技術の飛躍的な発展や、記憶媒体の大容量化に伴い、データベースに蓄積されるデータ数が急速に増えているが、スカイライン問合せは、大規模データを効果的にフィルタリングするための技術として、近年、注目され、盛んに研究されている。しかし、スカイライン上のオブジェクトは個人レベルのデータなので、この機能や、この機能により得られる知識は、個人情報を含むデータベースに対して適用することができない。

研究代表者らはスカイライン問い合わせの機能をグループ単位で行う新技術の開発を行った。グループ単位とすることにより個人情報を保護することが可能となるため、スカイライン問い合わせ機能の適用可能範囲を格段に広げることが期待できる。

## 4. 研究成果

従来のスカイライン問合せはレコード単位での選別を行うのに対して、候補者らは集合単位でレコードを選別するまったく新しい枠組みを構築した。スカイライン問合せの重要性が高まっている一方で、近年、個人情報保護に対する意識が高まり、それがスカイライン問合せ技術を応用するにあたっての

障害となってきた。候補者らの行ったこの技術を利用すれば、個々のデータ内容を開示することなく、スカイライン問合せの利点を享受することが可能となるため、この研究成果の意義は大きい。

これらの業績に関連して、2010年度末までに IPSJ TOD, IEICE 英文論文誌などを含む4編の査読付学術雑誌論文を公表しているほか、6編の査読付国際会議論文を公表している。公表された論文のリストは後述のとおりである。国際会議論文の中には、採択率が29%であった DBKDA 2010, 同33%であった PAKDD 2009 など厳しい競争率の中で選ばれているものも多く、論文数のみならず、個々の論文の評価も高いとあってよい。また、特筆すべき点として、フランスで行われた DBKDA 2010 国際会議で候補者らの論文は「Best Paper Award」を受賞している。

時間情報、位置情報に特化したグループ単位のデータマイニング技術としては、時間情報スカイライン集合問い合わせ、空間情報スカイライン問い合わせ機能を開発した。これらに関しては、現時点でまだ未公表のものもあるが、既発表分に関しては評判もよく、2010年には研究代表者が DNIS2010 という国際会議で招待講演を行っている。

また、本研究成果の応用実験をとおして、開発した機能が多くがクラウドコンピューティング環境からのデータマイニングに有効であることが予見された。そのため、研究期間終盤では、分散データベース上のマイニングにおける個人情報保護手法の開発も行った。この機能に関しては引き続き研究を継続し、クラウドコンピューティングの発展に貢献できる技術に育てあげていく予定である。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 7 件)

1. Md. Anisuzzaman Siddique, Yasuhiko Morimoto, “Efficient Maintenance of all k-Dominant Skyline Query Results for Frequently Updated Database”, IARIA International Journal on Advances in Software, Vol. 3, No. 3/4, 査読有, 2010, pp. 424-433.

2. Yasuhiko Morimoto, “Co-location Pattern Mining for Unevenly Distributed Data: Algorithm, Experiments, and Applications,” International Journal of Computational Science and Engineering, Vol. 5, No. 3/4, 査読有, 2010, pp.

185-196.

3. Md. Anisuzzaman Siddique, Yasuhiko Morimoto, “k-dominant and Extended k-dominant Skyline Computation by Using Statistics”, International Journal on Computer Science and Engineering (IJCSE), Vol. 2, Issue 5, 査読有, 2010, pp. 1934-1943.

4. Md. Anisuzzaman Siddique, Yasuhiko Morimoto, “Algorithm for Computing Convex Skyline Objectsets on Numerical Databases”, The Institute of Electronics, Information and Communication Engineers (IEICE) Transactions on Information and Systems, Vol. E93-D, No. 10, 査読有, 2010, pp. 2709-2716.

5. Md. Anisuzzaman Siddique, Yasuhiko Morimoto, “Continuous k-Dominant Skyline Computation by Using Divide and Conquer Strategy”, The Information Processing Society of Japan (IPSJ) Transactions on Databases, Vol. 3, No. 2, 査読有, 2010, pp. 48-60.

6. Taufik Djatna, Yasuhiko Morimoto, “Reviving Dominated Points in Skyline Query,” International Journal of Computer Science and Network Security, Vol. 1, No. 9, 査読有, 2009, pp. 347-354.

7. Taufik Djatna, Yasuhiko Morimoto, “Attribute Selection for Numerical Databases that Contain Correlations,” International Journal of Software Informatics, Vol. 2, No. 2, 査読有, 2008, pp. 125-140.

[学会発表] (計 14 件)

1. Md. Anisuzzaman Siddique, Yasuhiko Morimoto, “Efficient Computation for k-dominant Skyline Queries with Domination Power Index,” Proceedings of The 13th IEEE International Conference on Computer and Information Technology (ICCIT 2010), pp. 382-387, 23-25 Dec 2010. (バンガラディッシュ・ダッカ)

2. Md. Anisuzzaman Siddique, Yasuhiko Morimoto, “Extended k-dominant Skyline in High Dimensional Space,” Proceedings of The International Conference on Information Science and Applications (ICISA 2010), Vol. 2, pp. 520-527, 21-23

Apr 2010. (韓国・ソウル)

3. Md. Anisuzzaman Siddique, Yasuhiko Morimoto, “Efficient Maintenance of k-dominant Skyline for Frequently Updated Database,” Proceedings of The 2nd International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2010), pp. 107-110, 11-16 Apr 2010. (フランス・レメヌイエール)

4. Yasuhiko Morimoto, Md. Anisuzzaman Siddique, ” Skyline Sets Query and Its Extension to Spatio-temporal Databases,” Proceedings of The 6th International Workshop on Databases in Networked Information System, Springer, LNCS 5999, pp. 317-329, 29-31 Mar 2010. (福島)

5. Md. Anisuzzaman Siddique, Yasuhiko Morimoto, ” k-Dominant Skyline Computation by Using Sort-Filtering Method,” The 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2009), Lecture Notes in Artificial Intelligence, vol. 5476 (LNAI 5476), Springer, pp. 839-848, 27-30 Apr 2009. (タイ・バンコク)

6. Md. Anisuzzaman Siddique, Taufik Djatna, Yasuhiko Morimoto, ” Algorithm for Computing Skyline ObjectSet on Numerical Databases,” The 3rd International Workshop on Data-Mining and Statistical Science (DMSS2008), Japanese Society of Artificial Intelligence (JSAI), pp. 40-43, 25 Sep 2008. (東京)

7. Taufik Djatna, Yasuhiko Morimoto, ” Feature Selection for Databases that Contain Correlated Attributes by Using Two Dimensional Rules,” The 3rd International Workshop on Data-Mining and Statistical Science (DMSS2008), Japanese Society of Artificial Intelligence (JSAI), pp. 48-51, 25 Sep 2008. (東京)

8. 森本康彦, “時空間データマイニングとその応用,” 2008年度日本建築学会大会, 情報システム技術部門 研究協議会資料, pp. 45- 54, (パネルディスカッション「地域情報化ー社会的課題の解決手段としての意義と展望」), 2008年9月18-20日. (広島)

9. Yasuhiko Morimoto, ” Geometrical

Information Fusion from WWW and Its Related Information,” Proceedings of The 5nd International Workshop on Databases in Networked Information System, Springer LNCS 4777, pp. 16-32, 17-19 Oct 2007. (福島)

(他 5件)

[図書] (計 2件)

1. 森本康彦他, 電子情報通信学会, オンライン知識ベース “知識の森 (http://www.ieice-hbkb.org/portal/ より 順次公開予定)”, (7群6編6章 “頻出パターンの抽出” を分担執筆)

2. 青木孝夫, 坂田省吾, 林光緒, 平手友彦, 森本康彦編, 培風館, “知の根源を問う”, 2008, 260頁 (共編, 分担執筆, pp. 95-105)

[その他]

○商業誌への寄稿  
森本康彦, “個人情報保護法施行下のデータマイニング,” DB Magazine, Vol. 16, No. 12, 翔泳社, pp. 142-143, 2007.

○受賞  
フランスで開催された国際会議 International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2010)において発表した(「学会発表」項目の3の)論文が「Best Paper Award」を受賞した。

○ホームページ情報  
以下のホームページにおいてこれまでの研究成果を公開している。  
<http://www.morimo.com/morimo-ken/>

6. 研究組織

(1) 研究代表者  
森本 康彦 (MORIMOTO YASUHIKO)  
広島大学・大学院工学研究院・准教授  
研究者番号：00363010

(2) 研究分担者  
なし

(3) 連携研究者  
なし