

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 30 日現在

機関番号：33917

研究種目：基盤研究(C) (一般)

研究期間：2012～2015

課題番号：24500185

研究課題名(和文) プライバシー保護データマイニングを用いた交通データウェアハウス解析

研究課題名(英文) Data mining and privacy preservation for traffic data warehouse

研究代表者

河野 浩之 (KAWANO, HIROYUKI)

南山大学・理工学部・教授

研究者番号：70224813

交付決定額(研究期間全体)：(直接経費) 4,000,000円

研究成果の概要(和文)：本研究では、プライバシー保護の観点からデータマイニング技術の課題を整理し、交通情報を活用するロケーションサービスを応用例とするシステムアーキテクチャーを提案した。まず、空間データ処理におけるプライバシー保護に関わるデータマイニングの研究動向の調査を行った。次に、完全準同型暗号(Fully homomorphic encryption)のオープンソースHElib(Homomorphic Encryption library)を用いて、クラウド上のプロトタイプシステムを実装した。今後、計算機性能のさらなる向上により、クラウド上のデータセキュリティ向上に不可欠な技術であることを明らかにした。

研究成果の概要(英文)：In this research, from the point of view of data mining and privacy preservation, we proposed a cloud architecture of traffic information system for location service platform. First, we had a survey research of data mining related to privacy preservation and spatial data processing. Next, using the open source HElib(Homomorphic Encryption library) of fully homomorphic encryption, we tried to implement a prototype system in the cloud. In near future, a further improvement in computer performance revealed to be a key technology data security enhancement in the cloud architecture.

研究分野：情報システム

キーワード：データマイニング プライバシー保護

1. 研究開始当初の背景

(1) 2012 年当時、トランザクション高速化、高信頼性、クラウドアーキテクチャー、電力消費量低減などの研究が進むにつれ、大規模データを蓄積するデータベースが増加しており、データマイニング技術の応用も増えつつあった。特に、GPS 機能を備えるスマートフォン等の携帯デバイスを利用した位置情報サービスが浸透し、自動車に搭載する ETC 利用率も 85% 超となるなど、位置情報を含む行動履歴データベースが増加していた。

(2) 個人情報を含むデータベースに対するデータマイニングは、プライバシーに関する EU データ保護指令 (Data Protection Directive 95/46/EC) 等の観点から、データの利活用に問題があることが指摘されていた。特に、異なる運用組織において、複数のデータベースを安全に連携し、高度なデータ利活用を進めるには、プライバシー保護データマイニングアルゴリズム研究の重要性が増していた。

2. 研究の目的

(1) 交通データウェアハウスに対するデータマイニングの研究成果を発展させることを目的とし、位置情報を含む複数データベースを連携したデータ処理により、データの解析精度を高める研究を研究目的とした。

(2) 個人情報保護の観点から、位置情報を含む交通データには、プライバシー保護データマイニング (Privacy Preserving Data Mining) アルゴリズムが必要である。そこで、プライバシー保護データマイニングアルゴリズムを実装するシステムアーキテクチャを提案することを研究目的とした。

3. 研究の方法

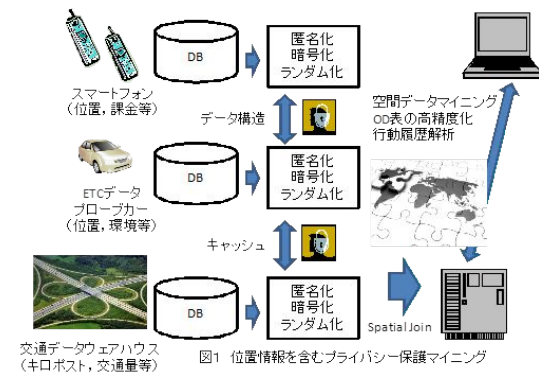
(1) 過去、阪神高速道路株式会社などから位置情報を含む交通データの提供を受けてきた。プライバシー保護の点で、一般車両の位置情報取得は困難であることから、パトロールカー等の特定車両に搭載された GPS による位置情報を含むプローブカーデータを用いて、データベース連携によりデータの精度を高めるための予備評価をすすめた。特に、位置情報を含むデータベース連携で有用な解析アルゴリズムを明確にするため、課題整理を行うと共に、プライバシー保護技術に関して広く研究調査を進めた。中でも、完全準同型暗号 (Fully homomorphic encryption) に着目し、応用の可能性について理解を深めた。

(2) 交通データウェアハウスの一部データをレプリケートし、プライバシー保護データマイニングの実験環境プロトタイプの実装を行った。初めに、プライバシー保護データマイニングが可能なクラウドアーキテクチャを提案し、CMS ソフトウェアと、GPL ソフト

ウェアである “HElib” を用いたプロトタイプシステムの構築を進めた。プロトタイプ実装環境下において、データマイニングアルゴリズムを実行するために、データ容量やデータ転送性能などを考察し、提案したクラウドアーキテクチャの性能評価を進めた。

4. 研究成果

2012 年度から 2015 年度の研究期間において、図 1 に示すような、複数データベース統合を実現し、データマイニングによる空間データに対する解析精度の向上を目指した。研究着手当初、プライバシー保護データマイニングを実現するための基礎技術となる「匿名化、暗号化、ランダム化」のいずれに可能性があるか明らかではなかった。



このような状況を踏まえて、研究計画期間内に、次の 5 項目について研究の進展を得た。

- (1) GPS 利用が進むにつれ、既存の交通データの精度高め、より精密な交通モデル評価が可能であること。
- (2) 位置情報の統合利用は有用であるが、個人情報であるため、プライバシー保護データマイニング技術が重要であること。
- (3) プライバシー保護と関連するデータマイニングやデータ同定の研究動向について調査し、完全準同型暗号 (Fully homomorphic encryption) の可能性に着目したこと。
- (4) 研究代表者が指導する学生らにより、提案アーキテクチャの実装を試み、性能測定結果を求めたこと。
- (5) 複数のデータベース連携を可能とするクラウドアーキテクチャにおいて、プライバシー保護データマイニングアルゴリズムの基礎となる演算を実現するために、“HElib”を用いたプロトタイプシステムを構築したこと。

以下、上記の 5 項目に関わる研究成果について詳細に述べる。

- (1) 2012 年 7 月に開催された EURO2012 において、“Analysis of optimal vehicle speed: from supersonic wave devices to GPS transceivers” を発表した。

本研究で扱うデータを蓄積している阪神高速道路株式会社の運用する交通データウェアハウスには、路上に500メートル間隔で設置された超音波センサーにより収集された、交通情報が蓄積されている。しかしながら、多くは点もしくは線によるデータ収集であることから測定精度が必ずしも高くない。パトロールカーなどに搭載されているGPSの活用により、速度、緯度経度データの精度の高い収集を行い、さらに、天候データ等をもちいることで、連続的なデータを利用した最適モデルについて検討を行った。

2012年の研究着手当時、プライバシー保護データマイニング技術は、垂直分割モデルと水平分割モデル、さらに、匿名化と暗号化とランダム化などの手法が提案され研究されていた。大きな課題は、プライバシー保護性能を高めても、データ同定技術により、データセキュリティを担保することが困難であった。また、暗号化技術についても、計算機性能面の課題があった。しかし、研究期間中の2013年に、HElibとして公開された暗号化ライブラリは、2009年にCraig Gentryが発表した完全準同型暗号技術に基づいており、その有用性を認識した。

- (2) 2014年に開催された第20回IFORS国際会議において、“Advanced traffic monitoring system by probe vehicles under privacy preservation”を公表した。

先行研究において論じたように、GPSを搭載した車両により、高い精度で「速度、加速度、位置等」の交通データ収集と蓄積が可能であることに基づき、既存のセンサーとGPSデータを融合し、交通密度や交通量など、渋滞予測などの基礎となる交通データの精度向上を試みた。また、この種のデータ収集を一般車両に広げるに際して、プライバシー保護の観点に着目したデータマイニングを必要とすることから、新たなシステムアーキテクチャが必要となることを論じた。

- (3) 2014年度末にむけて、プライバシー保護データマイニング関係の研究を調査し、「プローブカーによる交通データ測定の課題について 行動履歴データに対するプライバシー保護の観点から」(アカデミア南山大学紀要数理情報編)としてまとめた。

先行研究を中心に、交通データウェアハウスに格納されている超音波感知器などと、プローブカーとしてのパトロールカーにより収集された交通データを融合するデータ処理について紹介した。また、位置情報を含む行動履歴データに関わるプライバシー保護を

扱うため、プライバシー保護と関連するデータマイニングやデータ同定の研究の調査結果を含んでいる。

- (4) 2014年度に、研究代表者の所属する南山大学情報理工学部で卒業研究の指導を行った「完全準同型暗号を用いたクラウド上の演算機能の実装」では、提案アーキテクチャの実装を試み、測定結果を求めた。但し、データ提供者との関係で、本研究では実データを用いない実験とした(<http://www.seto.nanzan-u.ac.jp/msie/gr-thesis/2014/11se111.pdf>)

暗号化データ容量は、データセキュリティを高めるためのパディング部分を含むものの、元データの約1000倍を超すことから、クラウドにデータを高速にアップロードするために、一層のネットワークの高速化が必要と評価した。もっとも、セキュリティレベルにより処理時間は異なるものの、HElibを用いたクラウド設計は十分可能であると結論づけた。

- (5) 2015年7月に開催されたEURO2015において、“Cloud Architecture of Traffic Mining Systems under Privacy Preservation”を公表した。

日本における交通データ収集システムは、各組織によって個別に開発されてきた。また、各組織間で、一部の交通データの相互交換がなされているが十分ではない。先行研究で論じたとおり、車両に搭載された各種センサーにより精度の高い交通データの収集が可能となっていることから、あらたな交通データ利活用のチャンスが生じている。しかし、データセキュリティの観点から、重要な個人情報データをデータクラウドに直接蓄積することは困難である。そこで、プライバシー保護データマイニングが可能なクラウドアーキテクチャを提案し、“HElib”を用いることで、交通データマイニングを実行するためのプロトタイプシステムを構築し、データ転送性能などを考察した。

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5件)

河野 浩之、プローブカーによる交通データ測定の課題について 行動履歴データに対するプライバシー保護の観点から、アカデミア南山大学紀要数理情報編、査読無、第14巻、2014、pp.1-10
http://www.seto.nanzan-u.ac.jp/st/nas/academia/vol_014pdf/14-028-037.pdf

梅村 信夫、河野 浩之、沢田 篤史、学生

主導型授業評価」支援システムの構築：
LAMPとR言語によるクラウド型アプリケーションの試作、JeLA 会誌、査読有、日本 e-Learning 学会、Vol.13、2013/07、pp.68-78
<http://ci.nii.ac.jp/naid/40019767306>

Kawano, H., Fröschl, K., Nemecek, R., Dunkl, R., Weber, S.、Modelling of Pathodynamics for Process Epidemiology (Extended Abstracts)、査読有、Computer Aided Systems Theory (Eurocast 2013)、2013/02、pp.289-290

[学会発表](計 6 件)

Kawano, H.、 “ Cloud Architecture of Traffic Mining Systems under Privacy Preservation ” 、 27th European Conference on Operational Research (EURO2015) 、 2015/07/15、 Glasgow, (England)

Kawano, H.、 “ Advanced traffic monitoring system by probe vehicles under privacy preservation ” 、 20th Conference of the International Federation of Operational Research Societies 、 2014/07/17、 Barcelona (Spain)

Kawano, H.、 “ Analysis of optimal vehicle speed: from supersonic wave devices to GPS transceivers ” 、 Proc. of 25th European Conference on Operational Research (EURO 2012)、2012/07/10 , Vilnius (Lithuania)

6 . 研究組織

(1)研究代表者

河野 浩之 (KAWANO, Hiroyuki)

南山大学・理工学部・教授

研究者番号：7 0 2 2 4 8 1 3