科学研究費助成事業　　研究成果報告書

平成　２６　年　　６　月　１７　日現在

機関番号：　３２６８２
研究種目：　基盤研究(B)
研究期間：　2010～2013
課題番号：　２２３０００２８

研究課題名（和文）組織間でのプライバシー保護疫学調査技術の研究

研究課題名（英文）Privacy-Preserving Epidemiology Study for Vertically Partitioned Datasets

研究代表者
菊池　浩明（KIKUCHI, HIROAKI）

明治大学・総合数理学部・教授

研究者番号：２０２６６３６５

交付決定額（研究期間全体）：（直接経費）　　10,900,000円、（間接経費）　　3,270,000円

研究成果の概要（和文）：本研究では，病歴と患者リストを管理する病院Xと喫煙や放射線従事者などの個人属性を有する組織Yの間で，喫煙とがん罹患率の相関を解析する疫学調査を目的とする．二つの集合のBloom Filter(BF)の内積を求めて，共通集合の大きさをもとめる事前確率をベイズ推定する近似手法を提案した．提案手法はBFにより通信コストを削減し，ベイズ推定により精度を改善する．
本方式をピロリ菌感染と胃がんの疫学調査に応用した．χ二乗推定により，ヘリコバクターピロリ菌の感染によるがんの危険度を，それぞれのデータを秘匿したままで計算することを示した．

研究成果の概要（英文）：This paper proposes a new privacy-preserving scheme for estimating the size of the intersection of two given secret subsets. Given the inner product of two Bloom filters (BFs) of the given sets, the proposed scheme applies Bayesian estimation under an assumption of beta distribution for an a priori probability of the size to be estimated. The BF retains the communication complexity and the Bayesian estimation improves the estimation accuracy.
A possible application of the proposed protocol is an epidemiological datasets regarding two attributes, Helicobacter pylori infection and stomach cancer. Assuming information related to Helicobacter Pylori infection and stomach cancer are separately collected, the protocol demonstrates that a Chi-squared test can be performed without disclosing the contents of the two confidential databases.

研究分野：　総合領域

科研費の分科・細目：　情報学・計算機システム・ネットワーク

キーワード：　ネットワークセキュリティ技術　プライバシー保護

１．研究開始当初の背景

個人と組織の間の情報の管理方法が大きく変化している．常時接続環境の普及とともに，個人からの情報発信は盛んになった．インターネットの持つ匿名性が情報発信のしきい意識を軽くしている．

一方，ひとたび発信された消費者や利用者からの個人情報を管理する組織の負担は重くなっている．個人情報の安全な管理と利用者に対する管理責任を明確にした個人情報保護法やガイドラインの浸透により，不用意な個人情報の交換は禁じられている．個人情報管理者としての組織は，それぞれが管理するユーザの属性や嗜好などの情報を持ち出せば，漏洩事件，他の組織との間で共有すれば個人情報の横流しとみなされる．それゆえ，ひとたび利用者から放たれた情報は組織に入ると取り出せない．

しかしながら，各組織が塩漬けにしている個人の属性や嗜好の情報には大きな潜在価値がある．例えば，膨大な商品の購買記録を元にその利用者の興味を予測して商品情報を提供する推薦システムは効果的な応用の代表例であろう．商品情報だけではなく，講義情報の推薦など，多くの分野に推薦の仕組みが導入されようとしている．ただし，これらの例は，単一の大規模な組織による閉じた情報推薦である．小規模な組織がそれぞれ管理する情報は，個人情報保護の観点から統合が難しい．書籍を購入した履歴からツアー旅行のパッケージを推薦することは困難なのが現状である．

組織間を越えた情報解析の最も難しくなっているものの一つに，疫学調査がある．道路・鉄道・航空会社の有する電車や航空券の搭乗履歴と病院が有するインフルエンザの患者リストを組み合わせれば，感染に至る経路を同定できる可能性があるが，前述の個人情報管理者の義務がそれを困難にしている．しかも，病歴は最も機微なプライバシー情報である．

２．研究の目的

(1) プライバシー保護の課題
本研究では，公開鍵暗号技術の安全性に基づく，プライバシー保護データマイニング技術 (Privacy-Preserving Data Mining, 以後 PPDM)をこの問題に適用する．PPDM と呼ばれる技術には，確率的にノイズを加える摂動アプローチと暗号学的アプローチの二つの大きな流れがある．暗号学アプローチは秘匿レベルが高く，精度も保証されるが，大きな計算コストがかかる． PPDM を用いて疫学調査を行うプロトコルを開発する．

(2) スケーラビリティの課題
大規模な医療データに PPDM を適用するために，各種のデータ構造の適用を検討する．

(3) 実証実験の課題
理論的な考察だけではなく，現実の医療疫学問題に提案手法を適用して，その実現可能性や実現に必須の課題を明らかにする．

３．研究の方法

(1) 垂直分割型プライバシー保護データマイニング技術の調査
これまでに提案されている多くの PPDM 研究のサーベイを行い，本研究目的への適用可能性を調査する．実施可能性は，問題の整合性に加えて処理時間などのパフォーマンスや精度などの品質，実現容易性などについて多方面から評価する．

(2) がん疫学調査の実施に関わる調査
全国の地域がんセンターで管理されている患者データの形式と属性，その規模や種類などを調査し，プライバシー保護疫学調査システムの入力するための前処理を行う．放射線業務従事者協会で管理されている従事者データベースについても同様に調査を行う．

(3) プライバシーを保護した疫学調査プロトコルの設計
暗号学アプローチによる，ID の整合を保証しない新しいプライバシー保護疫学調査プロトコルを設計する．プロトコルの計算量を明らかにし，その性能を予測する．選択平文攻撃やプロトコルから漏れる情報量などの暗号学的な安全性評価を行う．プロトコルを守らない組織に対しては，ゼロ知識証明などの技術を適用した悪意の不正者(malicious)モデルへの拡張を行う．

(4) プライバシーを保護した疫学調査システムの実装
提案プロトコルを Java により実装する．ハッシュや準同型性を満たす公開鍵アルゴリズムなどの要素技術を実現し，単一のホスト上で複数のデータベースの間のプライバシー保護疫学調査の計算を行うシステムを構築する．次に，遠隔地にある組織間で疫学調査を可能とするように，SSL/TLS ベースで相互に計算するシステムへ拡張する．

(5) 提案方式の実データへの適用と評価
実装されたシステムを実際の患者データと従事者データに適用し，年齢別などの各属性についての罹患率を計算する．暗号化する際に生じるデータ量や処理時間の増大を計測し，適用可能な患者数規模を評価する．同姓同名の患者や住所変更や氏名改名などの特殊データの存在が精度に及ぼす影響も調査する．

(6) 実証実験に基づくプロトコルの改良
実験結果に基づき問題点を検討し，必要に応じてプロトコルとシステムを改良する．罹患率以外の疫学調査や他の情報推薦システムなどへ適用可能性を検討する．

４．研究成果

(1) プライバシー保護疫学プロトコルを提案した．基本となる秘匿内積プロトコルを応用した放射線疫学調査プロトコルを提案し，国際会議にて発表した．（学会発表 2, 11）

(2) Bloom Filter を用いた効率の良い秘匿内積プロトコルを提案した.
(雑誌論文 1, 学会発表 1, 8, 10)
図 2 に Bloom Filter (BF)の原理を表す. 集合 S の要素について複数のハッシュ関数 H1, H2 を用いて BF のビット位置を決めて登録 (join)し, 与えられた要素がその BF に登録済みかどうかを, 同様の手順でハッシュ関数にかけて BF の位置を決定し, 全ての位置が 1 にセットしていたらし S に所属すると判断する. 集合の要素数 n に BF のサイズが依存することなく, 大きな圧縮効果がある反面, 偽陽性, 例えば, 図の a4 の例の様に, 本来 S の要素でないのに要素であると誤って判定されることがある.
提案方式では, 小さなサイズの BF を幾つも用意し, 秘匿内積プロトコルを用いて二つの集合の交わりを計算する. 図 3 に示す様に, 検査の繰り返し回数 s が増加するに従って, 計算の制度が向上することを示した.
この提案方式は, 従来の秘匿内積プロトコル[FNP, CT, KNV]などと比較して, 計算コストが小さく O(n), 通信コストも繰り返し回数 s と BF のビット長 m の積である. これらの関係を図 4 に整理する.

(3) 千葉がんセンターと協力して, 提案手法を現実のデータに適用し, ピロリ菌感染による胃がんの相対危険度を算出した.
(雑誌論文 1, 学会発表 9)
図 2 に, この実験で行ったピロリ菌と胃がんの分割表を表す. ピロリ菌に感染していた 2629 人の内, 80 名が千葉がんセンターに登録していたことを表している. この罹患率 80/2629 は一般的な罹患率に比べて 9.7 倍高く, すなわち, ピロリ菌に感染しているとがんになるリスクがあることを示している. この結果の確率検定の結果は, 統計量 χ =17.81 であり, 98%を超える P 値で統計的に有意である.

(4) 研究成果のプライバシー保護データマイニング技術を応用して, 情報推薦システムや信頼交渉システム, ソーシャルネットワークでのラベル推定などへのプライバシー保護を実現した.
(雑誌論文 2, 3, 学会発表 3,4,5,6,7,12, 13)

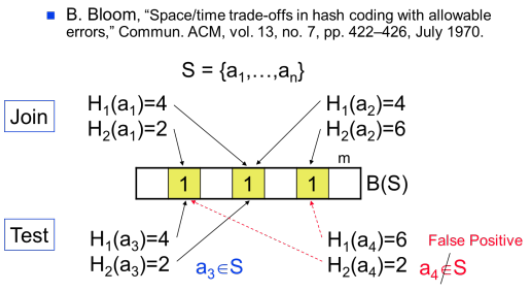

図 1 Bloom Filter の原理

	がん患者数	非登録者	計
ピロリ菌	80	2,549	2,629
非保有者	346	106,988	107,334
計	426	2,999,574	3,000,000

相対危険度 RR = 9.70 (ピロリ菌はがんに9.7倍なりやすい)
有意性 χ = 17.81 (98%以上の確からしさで有意)

図 2 ピロリ菌の相対危険度

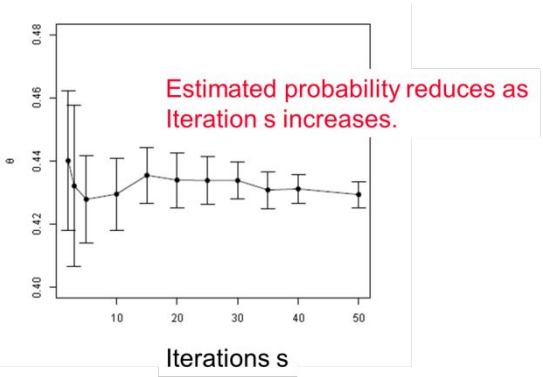

図 3 Bloom Filter による精度向上

		FNP[3]	CT	KNV[5]	proposed
primitive		OPE	Blind RSA	SSP, BF	SSP, BF
Comp.	A	$n_A \log\log n_B$	$2n_A+1$	m	ms
	B	$n_B+n_A \log\log n_B$	n_A+n_B+1	0	0
		$O(n_A \log\log n_B)$	O(n)	$O(n^2)$	O(n)
Comm.		n_A+n_B	$2n_A+n_B$	m+1	ms+1

$n^2 > m > kn$ $m = n/\ln 2$

図 4 提案方式の性能評価

5. 主な発表論文等

〔雑誌論文〕(計 3 件)

1. Hiroaki Kikuchi, Jun Sakuma, Bloom Filter Bootstrap: Privacy-Preserving Estimation of the Size of an Intersection, Journal of Information Processing,Vol. 22 (2014) No. 2 pp. 388-400. (**推薦論文**, 査読あり)

http://dx.doi.org/10.2197/ipsjjip.22.388

2. Hiroaki Kikuchi , Anna Mochizuki, Privacy-preserving Collaborative Filtering Using Randomized Response, Journal of Information Processing, Vol. 21 (2013) No. 4 pp. 617-623, **（査読あり）**
https://www.jstage.jst.go.jp/article/ipsjjip/21/4/21_617/_article
3. Tangtisanon Pikulkaew and Hiroaki Kikuchi, "Perfect Privacy-preserving Automated Trust Negotiation", Journal of Information Processing, IPSJ, Vol. 19, pp.451-462, 2011.
（査読あり）

〔学会発表〕（計　13 件）

1. Hiroaki Kikuchi, Jun Sakuma, "Bloom Filter Bootstrap: Privacy-Preserving Estimation of the Size of an Intersection", Data and Applications Security and Privacy XXVII (DBSec 2013), Springer, LNCS Volume 7964, pp 145-163, 2013.
2. Hiroaki Kikuchi , Tomoki Sato , Jun Sakuma, "Privacy-Preserving Protocol for Epidemiology in Effect of Radiation", Proceedings of the 2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS '13), pp. 831-836, IEEE, 2013.
3. Hiroaki Kikuchi, Privacy-Preserving Data Mining for Medical Applications, 2013 AI Forum, Tamkang, Taiwan, 2013. **(招待講演)**
4. Hiroaki Kikuchi, "Application for Privacy-Preserving Epidemic analysis and Bays Estimation of Size of Intersection using Bloom Filter", The 8th International Workshop on Security (IWSEC2013), Okinawa, 2013. **(招待講演)**
5. **菊池浩明，ライフログに関するプライバシーの課題とプライバシー保護データマイニングの展望，信学技報**, vol. 113, no. 326, ISEC2013-64, pp. 35-38, 2013**.**
（招待講演）
6. H. Kikuchi, Yoshiki Aoki, Msayuki Terada, Kazuhiko Ishii, Kimihiko Sekino, "Accuracy of Privacy-preserving Collaborative Filtering Based on Quasi-homomorphic Similarity", 2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing (ATC 2012), pp. 555-562, IEEE, 2012.
7. Hiromi Arai and Jun Sakuma, "Privacy Preserving Semi-Supervised Learning for Labeled Graphs", Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2011),
8. **菊池浩明，佐久間淳**, Bloom Filter **を用いた積集合サイズのベイズ推定とそのプライバシー保護疫学調査への応用**,**コンピュータセキュリティシンポジウム** (CSS 2012), pp 216-223, IPSJ, 2012.
（優秀論文賞受賞）
9. **菊池浩明，佐久間淳，三上春夫，プライバシーを保護したピロリ菌疫学調査**, 2012 **年度人工知能学会全国大会（第** 26 **回）**, 3I2-OS-20-9, pp. 1-4, 2012.
10. **菊池浩明，佐久間淳**, "Bloom **フィルタを用いたマッチング数の秘匿比較**", **コンピュータセキュリティシンポジウム** 2011, **情報処理学会** , 2C4-3, pp. 516-521, 2011**.**
11. **菊池 浩明**, **橋本 英樹**, **康永 秀生**, **渋谷 健司**, DPC **データセットによるプライバシーを保護した治療戦略の比較**, **コンピュータセキュリティシンポジウム** (CSS 2013), **情報処理学会，** 1D1-4, pp. 110-117, 2013.
12. Shuang Wu, Tadanori Teruya, Junpei Kawamoto, Jun Sakuma, Hiroaki Kikuchi, "Privacy-preserving Logistic Regression", **第** 27 **回人工知能学会全国大会**, 3L1-OS-06a-3**，** 2013.
13. Shuang Wu, Junpei Kawamoto, Hiroaki Kikuchi, Jun Sakuma, "Privacy-preserving Online Logistic Regression Based on Homomorphic Encryption", **信学技報**, vol. 113, no. 139, IBISML2013-10, pp. 67-74, 2013.

６．研究組織

(1)研究代表者
菊池　浩明（KIKUCHI, Hiroaki）
明治大学・総合数理学部・教授
研究者番号　20266365

(2)研究分担者
佐久間　淳（SAKUMA, Jun）
筑波大学・大学院システム情報工学研究科（系）・准教授
研究者番号 90376963

(3)連携研究者
なし.