論文審査の要旨および学識確認結果

報告番号	甲／乙第　　号	氏　名	Jia　Jing

論文審査担当者：				
	主査	慶應義塾大学教授	博士(工学)	西　宏章
	副査	慶應義塾大学教授	博士(工学)	重野　寛
		慶應義塾大学准教授	博士(情報理工学)	金子　晋丈
		慶應義塾大学教授	博士(工学)	久保　亮吾
		アアルト大学教授	Ph.D	Valeriy Vyatkin

(論文審査の要旨)

修士（工学)、Jia Jing 君の学位請求論文は、「Advances in Formal Anonymization for Practical Applications ― Efficient and Secure Techniques for Structured and Unstructured Data ― (実用的応用に向けた形式的匿名化の新展開 ― 構造化データおよび非構造化データに対する効率的かつ安全な手法 ―)と題し、6 章で構成されている。

現在社会が直面する、医療データなど機微情報を含むデータの利活用におけるプライバシー保護問題、特に情報漏洩のリスクや柔軟性の欠如、処理のスケーラビリティといった諸問題を解決する手法を提案している。また、差分プライバシー・k-匿名性・階層的匿名化・適応的匿名化といった異なるアプローチにも対応している。

本論文の構成は以下のとおりである。第 1 章において、機微情報を含む医療・臨床データの活用において直面するプライバシー保護の課題を概観し、匿名情報流通インフラの重要性と、そのインフラがもたらす新たな情報社会の展望について述べている。

2 章では、背景および関連研究について述べている。プライバシー保護技術に関する理論的基盤と応用事例を網羅的に整理し、本研究で利用する差分プライバシー、k-匿名性などについて説明し、それぞれの長所と限界を論じている。次に、準同型暗号と匿名化技術の統合事例について触れ、既存手法が直面する性能面や適応性の問題を明らかにしている。また、構造化データと非構造化データの双方における医療データ匿名化の実例を紹介し、本研究の立ち位置を明確化している。

3 章では、現状の匿名化インフラにおいて、情報集約のためクラウド上で匿名化することから、クラウド業者が不要な個人情報を集めざるを得ない状況に対する解法を与えている。たとえ信頼できないサーバ上であったとしても、準同型暗号を用いた安全な差分プライバシーによる匿名化手法を提案している。理論的解析とともに、実際の医療データセットを用いた評価により、匿名化手法としての正しさと実用的な計算性能の両立を達成することが示されている。

4 章では、3 章の拡張として、k-匿名化などクラスタリングを行う匿名化への応用を図っている。階層的な k-匿名化処理を秘密計算および秘密分散を用いて実現する手法を提案し、複数段階にわたる汎化処理を用いて、複雑な k-匿名化処理を秘密裏に実行することができる。特に、秘密計算の計算コストは大きいため、内包する階層的構造を利用し、近隣のデータとの一般化処理においては秘密計算を、遠くのデータでは計算コストが少ない秘密分散を用いることで全体の計算コストの低減を図っている。この場合も同様に医療データセットを用いて有効性を示している。

5 章では、さらに、数値データではなく、電子カルテの文章による自由記述といった非構造データに対して k-匿名性を適用する新たな方式を提案している。自然言語処理が必要であるため BERT (Bidirectional Encoder Representations from Transformers)を用いた識別子・準識別子の抽出と、抽出されたワードに対して個別に匿名化レベルを調整することを可能としている。従来の画一的な匿名化では失われていた文脈的有用性を保持するため実運用での利便性が高い。評価実験により、従来手法に比べて情報損失が小さく、臨床研究用途における利活用の可能性を確認している。

以上、本研究は情報の匿名化インフラを構築するうえで障害となっていた複数の問題点を解決しており、情報の安全な公開を達成するために必須となる基本技術の構築を図っていることから、その貢献は工学上少なくない。

よって，本論文の著者は博士(工学)の学位を受ける資格があるものと認める。

学識確認結果	学位請求論文を中心にして関連学術について上記審査会委員で試問を行い、当該学術に関し広く深い学識を有することを確認した。 また、語学（英語）についても十分な学力を有することを確認した。