様式３

論　文　内　容　の　要　旨

氏　名　（　三本　知明　）	
論文題名	A study on Utility-aware Privacy-preserving Techniques （有用性を考慮したプライバシ保護技術に関する研究）

論文内容の要旨

Privacy is a fundamental human right and the constitutions of many countries mention the right to privacy. However, the development of computer science is threatening that right. Security and privacy measures are essential for the safe and secure utilization of data. The main difference between privacy and security is who is the attacker. As for security, the data needs to be protected from outsiders, and the correct data needs to be delivered to authorized recipients. Privacy, on the other hand, needs to take into account that even authorized recipients are attackers because data subjects have rights associated with their data. Therefore, it is important to strike a balance between privacy and utility, and technologies are needed to achieve both requirements.

This dissertation presents privacy-preserving techniques that focus mainly on the utility of data and the understandability of privacy strength. Privacy data can be divided into structured data and unstructured data. Furthermore, there are static and dynamic data for each type of data. De-identification techniques also need to address the issues of privacy and usefulness in terms of data format. We propose privacy-preserving techniques for these data formats. The experimental results show that our methods can improve the utility while maintaining privacy.

Privacy-preserving techniques include not only de-identification techniques but also secure computation. This dissertation deals with privacy-preserving techniques in general that may improve utility, and secure computation is also covered. In the secure computation, we focus on one of the most basic components, the two-party comparison protocol, and taxonomize the protocol from the perspective of each user's input and output. Furthermore, we show that it is possible to construct arbitrary types of protocols by organizing the conversion methods between protocols.

In addition, we discuss differential privacy for further research. We review the procedure for constructing the differential privacy mechanism and confirm that there are situations in which it is difficult to calculate sensitivity. We discuss the definition of differential privacy to derive sensitivity even in such a situation. We focus on the t-test as a case study and construct a differential privacy mechanism that satisfies the new definition.

様式 7

論文審査の結果の要旨及び担当者

氏　名	（　三本　知明　）		
論文審査担当者		（職）	氏　名
	主 査	教授	宮地　充子
	副 査	教授	滝根　哲哉
	副 査	教授	丸田　章博
	副 査	教授	馬場口　登
	副 査	教授	三瓶　政一
	副 査	教授	井上　恭
	副 査	教授	鷲尾　隆
	副 査	教授	駒谷　和範
	副 査	教授	佐久間　淳（筑波大学　システム情報系）

論文審査の結果の要旨

IoT 機器の普及に伴い、様々なデータが収集されるようになった。今後、5G の普及に伴い、各種データが様々な機器から収集されるようになるだろう。これらビッグデータの利活用により、医療、教育、行政など様々な分野の課題を解決するといわれている。しかしながら、データにはプライバシが含まれており、プライバシの確保とデータ利活用の有用性は両刃で、その両立は非常に難しい。現在、過度なプライバシ強化の方向に進む傾向にあり、プライバシを保護しつつデータ有用性を劣化させない匿名化技術の確立は重要な課題である。プライバシ保護技術の重要な技術の一つである入力プライバシ保護技術とは、データを直接処理（非識別化）してプライバシを保護する技術を意味する。本論文では主に入力プライバシ保護技術を扱っている。本技術は、主にデータ主体から信頼できる機関に収集されたデータを対象とし、企業や団体で利用されることを想定する。データセットをデータ構造とデータタイプの観点から分類している。具体的にはデータ構造は構造化データと非構造化データに、データタイプは静的データと動的データに分類している。本論文ではプライバシと有用性の 2 つの指標で提案方式を評価するが、プライバシ指標は個人が再識別されうる確率を用いる。これは k-匿名性を一般化した指標である。本論文ではプライバシと有用性の両立を目指し様々なデータにおいてプライバシを保護しつつ有用性を確保した匿名化技術を提案している。

第 1 章では、プライバシの概念とプライバシ保護の必要性について論じている。第 2 章では、これまで研究されてきたプライバシ保護技術について、k-匿名化ベース、差分プライバシベース、秘密計算ベースに大別して紹介しており、本論文で必要となる関連研究についても紹介している。第 3 章では、主なプライバシ保護技術に関する関連研究をとその問題点を考察している。

第 4 章では、最も多く研究されている構造化静的データに着目し、実際に組織間のデータ提供を前提とした際の、非識別化されたデータセットのプライバシ評価手法および有用性の評価指標が提案されている。提案するプライバシ評価手法は、複数の非識別化手法の組み合わせにより非識別化されたデータセットのプライバシを評価することが可能である。k-匿名性はその指標のわかりやすさから、広く採用されているプライバシ指標であるが、確率的な非識別化処理に対するプライバシ評価はできない。そこでシステム設計や法制度を考慮した組織間のデータ提供を考慮した攻撃者モデルを構築することで、確率的な非識別化処理を適用した場合でも統一的な k-匿名性に準じるプライバシ評価を行うことを可能としている。実験では同程度のプライバシ強度を持つデータセットであっても、複数の非識別化手法を組み合わせることで、より高い有用性を維持できることを確認している。

第 5 章では、構造化動的データを扱っている。特にデータ形式として行列を想定し、行列分解を非識別化手法の一つとして扱っている。提案手法は容易に k-匿名化や撹乱などの非識別化手法を組み合わせることが可能であり、k-匿名

化と行列分解を組み合わせることで得られる非識別化データセットは k-匿名性を有する。実験では非識別化前後のデータセットから個人を再識別する攻撃と、異なる時間帯のデータセットから、同一個人を特定する二種類の攻撃者を想定し、前者に関しては行列分解と撹乱により、柔軟にデータセットのプライバシ強度をコントロールできることが確認され、後者に関しては行列分解のみであってもプライバシ強度を飛躍的に向上することが確認された。また、4 章と同様、同程度のプライバシ強度を持つデータセットであっても、非識別化手法を組み合わせることで、より高い有用性を維持できることを確認している。

第 6 章では、静的非構造化データに対する非識別化手法を扱っている。ここでは特に文章データを扱い、Web 検索を行う攻撃者を想定した対策アルゴリズムが提案されている。アルゴリズムは大きく事前準備部、攻撃部、プライバシ保護部の 3 つのサブルーチンから構成され、事前準備部では文書データを構造化データに変換し、各単語の情報量を計算する。攻撃部では実際に情報量の高い単語を用いた Web 検索によって文書データを収集し、各文書に対して追加情報を取得可能であるかを判定している。プライバシ保護部では攻撃が成功した単語を削除、あるいは一般化することで Web 検索による追加情報の取得を阻害している。実験では実際の文書データとして学校で発生した事故レポートと裁判文書を利用し、人手で非識別化した文書にはプライバシリスクが残っていることが確認されている。提案アルゴリズムは、プライバシリスクのある単語のみを削除・一般化するため、必要十分な非識別化処理で済み、有用性の維持が可能である。なお、本提案は動的データにも適用可能である。

第 7 章では、データが信頼できる機関に保管されている場合に、そのデータに含まれるプライバシ情報を保護することを目的とする差分プライバシ技術に着目されている。差分プライバシを満たす撹乱メカニズムは任意のデータセットを想定し、1 レコードがクエリの出力に与える結果センシティビティの導出が必要となる。しかし、t 検定や相関係数のように任意のデータセットを考慮するとセンシティビティの導出が困難とされる。さらに従来の差分プライバシを変更し、ダミーデータの追加を考慮した差分プライバシの定義が提案されている。さらに従来の差分プライバシを満たすメカニズムはこの新しい定義も満たすことが示されており、ケーススタディとして、t 検定の t 値をクエリとしたダミーデータの追加を考慮した差分プライバシメカニズムを構築し、評価を実施している。本提案により、センシティビティの導出が困難な場合であっても直接的にセンシティビティの導出が可能で、提案する定義を満たすメカニズムの構築が可能となっている。

第 8 章では、プライバシを保護したデータの利用を促進するセキュアな計算の基盤技術である二値の大小比較プロトコルの改良を目的とし、入出力の違いから大小比較プロトコルのタイプを分類し、任意のタイプに変換するプロトコルが提案されている。さらに、木構造を用いた効率的な二値の大小比較プロトコルも提案されており、タイプ間の変換プロトコルを組み合わせることで任意のタイプで効率的な二値の大小比較プロトコルを実現可能としている。

第 9 章では、分類されたデータセットに応じたデータ利活用の有用性を保持しつつプライバシを確保する本論文の成果である匿名化技術について述べられている。

以上のように、本論文では有用性を考慮したプライバシ保護技術に関する研究が扱われ、様々なデータにおいてプライバシを保護しつつ、有用性を確保した匿名化技術の研究が提案されている。また、構造化データと非構造化データ、静的データと動的データのそれぞれの観点においてプライバシ保護と有用性を確立する方式が提案されており、さらに提案方式は理論的に評価されている。本論文で提案されたプライバシを保護しつつデータ有用性を劣化させない匿名化技術は、今後ますます重要となるビッグデータの利活用における貢献は計り知れない。

よって本論文は博士論文として価値あるものと認める。