審 査 の 結 果 の 要 旨

論 文 提 出 者 氏 名　　楊　　斌

データマイニングとプライバシー保護の接点に位置する研究として、2000年代に入ってから対象データに記載されたプライベート情報を保護しつつデータマイニングを行うプライバシー保護データマイニングの研究が盛んになってきている。プライベート情報とはデータベースにおける個人名などの個人情報が想定されている。一方、複数のパーティと呼ばれる組織が各々の保持するデータベースを統合してデータマイニングをする場合は、各パーティのデータベースそのものがプライベート情報と見なされ、他の機関に漏れないように保護することが必要になる。例えば、病院における患者情報のデータベースは、複数の病院が統合してデータマイニングすれば感染症の流行状況の把握、流行予測などの有益な疫学的知見が得られる可能性がある。しかし、各病院のデータベースは外部に流出しないようにしなければならない。このような目的を持つプライバシー保護データマイニングにおいて、複数のパーティの保持するデータベースを暗号化したうえで、データマイニングを行う手法が重要な課題となってきている。

本論文は「Encryption-based Privacy Preserving Data Mining － Collusion Resistant Protocol and Clustering on Secure Networks」
（暗号化に基づくプライバシー保護データマイニング―結託耐性プロトコルとセキュアネットワーク上のクラスタリング）と題し、６章からなる。

第１章「Introduction」（序論）では、プライバシー保護データマイニングの背景となる近年のデータ資源状況を述べ、その必要性、有用性を論じている。

第２章「Privacy-Preserving Data Mining」（プライバシー保護データマイニング）では、まず、プライバシー保護データマイニング分野を概観している。すなわち、プライバシー保護データマイニングは、(1)暗号技術を用いたセキュア計算による手法、(2)データマイニングの対象となるデータベースに摂動を加える入力摂動による手法、(3)データベースへの質問に雑音を加算する出力摂動による手法に分類されることが述べられている。これら３種の手法のうち、本論文で対象としている(1)のセキュア計算による手法において使用される準同型性公開鍵暗号、セキュア計算プロトコル、ランダムシェアなどの概念を導入し、次に既存研究について説明している。

第３章「Collusion-Resistant Privacy-Preserving Data Mining」（結託耐性プライバシー保護データマイニング）では、暗号技術を用いたセキュア計算プロトコルによるデータマイニングにおける従来の手法では実現できていなかった結託耐性を扱っている。複数パーティの各々が保持するデータを統合的に利用するプライバシー保護データマイニングでは、あるパーティのデータを複数パーティが結託して暴こうとする攻撃が問題となる。提案手法では攻撃目標とされたパーティ以外の全パーティが結託しても、攻撃目標のパーティが保持するデータは暴かれないセキュア計算プロトコルを提案して

いる。ここでは、各パーティが保持するデータを分割して暗号化したものを他の全パーティに送り、最終結果をランダムに分解したものを得ることが目的である。各パーティは受け取ったデータは暗号化したまま計算を行う。これによって自分以外の全パーティが結託しても、自分のデータは保護できる。この方法によって複数のデータの和に対する微分可能な関数が実行できるので、種々のデータマイニングアルゴリズムを対象にしたプライバシー保護データマイニングが実現できることが示されている。

第４章「Private Clustering on Secure Networks: Gibbs Sampling」(セキュアネットワークにおけるプライベートなクラスタリング：ギッブスサンプリング)では、パーティはネットワークでつながれた個人であり、直接つながっている接続相手との接続情報しか分からないという状況におけるプライバシー保護データマイニングを提案している。すなわち、全パーティは自分のネットワークの接続情報は他者へは流出しないという条件の下でクラスタリングを行う暗号を利用したプライバシー保護データマイニングの方法について検討し、クラスタリング手法としてプライバシー保護ギッブスサンプリングを提案している。提案手法の性能評価を行い、高い精度が得られることを実証している。

第５章「Private Clustering on Secure Networks: EM Algorithm」(セキュアネットワークにおけるプライベートなクラスタリング：ＥＭアルゴリズム)では、第４章と同じネットワークを対象にしたクラスタリングを実現するプライバシー保護できるＥＭアルゴリズムを提案している。第４章のギッブスサンプリングに比べて同じ精度を保持しつつ計算速度は高速であることを実証している。

第６章「Conclusion」（結論）は本論文のまとめである。

以上を要するに、本論文は複数パーティのデータベースを対象にし、準同型性公開鍵暗号を利用したプライバシー保護能力の高いデータマイニング手法として、従来提案されていなかった結託耐性のあるアルゴリズムを提案している。さらに、ネットワークにおける接続情報などの個人情報のプライバシー保護という制約を満たしたクラスタリング手法を提案し、実装したうえで性能評価している。これらの研究成果によって数理情報学分野の技術発展に寄与した。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。