Space Adaptation: Privacy-preserving Multiparty Collaborative Mining with Geometric Perturbation

Keke Chen  Ling Liu
Outline

- Introduction
  - Service-based collaborative mining
  - Privacy issues
- Geometric perturbation
  - Concept
  - Challenges in service based collaborative mining
- Space Adaptation Protocol
Introduction

- Service-based multiparty collaborative mining
Privacy issues in this paradigm

- The shared data may contain sensitive information that is important to the owner.

The goal

- Find the model without leaking the sensitive information to any of the involved parties.

Assumption:

- Data are encrypted in transmission.
- Semi-honest parties, without collusion.
Geometric Data Perturbation

- \( G(X) = RX + T + D \)
  - \( X \): dataset
  - \( R \): “random rotation matrix”
  - \( T \): “random translation matrix”
  - \( D \): random noise, for perturbing distances

\( Gi \) represents the parameters \((R,T,D)\),
\( Gi(X) \) is the perturbed data

- Particularly good for many classification models
  - Models trained with perturbed data keep similar prediction accuracy
  - kNN, kernel methods, SVMs, linear classifiers, and more
When it goes to multiparty...

- Each party has its own secret $G_i$
  - Different $G_i \rightarrow$ different data space
  - But mining can be only done on a unified space
  - So we need to **securely unify these $G_i$ to $G_t$**

- Potential attacks
  - The revealed information is valuable, only when
    - the data owner is identified, given $G_i(X)$ or $G_t(X)$
    - when $G_i(X)$ or $G_t(X)$ is known, $X$ can be estimated precisely

- Privacy threats from
  - Other curious data providers
  - Curious service provider
Space Adaptation (SA)

- SA is one of the approaches unifying Gi
  - Utilize the fact that geometric transformations are transformable to each other

- The “space adaptor” : transform Gi to Gt
  - $G_t(X) = S_{i \rightarrow t}(G_i(X))$, $S$ is the space adaptor, $G_t$ is the unified perturbation
  - Each party knows $G_t$, and thus holds $S_{i \rightarrow t}$, $G_i$, and $G_i(X)$
Space Adaptation

- **Protocol**
  - Prevent service provider identifying source
    - Shuffle perturbed data between data providers
  - Prevent data provider breaching privacy from the received perturbed data
    - Locally optimized Gi [Chen&Liu SDM07]
  - Ignore the details ...
Evaluation

- **Source identifiability**
  \[ \pi = \Pr(\text{source is identified}) \]

- **Normalized privacy guarantee**
  - Maximum privacy guarantee \( b_i \)
  - Locally optimized privacy guarantee \( \rho_i \)
  - Normalized privacy guarantee: \( \rho_i/b_i \)

- **Overall risk of privacy breach for DPi**
  For DP j receiving perturbed data from DP i
  risk: \( 1* (1 - \rho_i/b_i) \)

  For service provider
  risk: \( 1/n * (1 - \rho_i'/bi) \)

  \( \rho_i' \) is the privacy guarantee of Gt to Xi
Conclusion

- Properties of geometric perturbation
- SA protocol considers
  - Source identifiability
  - attacks to the perturbed data
- Future work
  - Remove the semi-honest assumption
- Contact the authors
  - Keke Chen (kekechen@cc.gatech.edu)
  - Ling Liu (lingliu@cc.gatech.edu)