3つのひかり　未来をつくる
広島市立大学
Hiroshima City University

機械学習におけるセキュリティ対策とその応用

研究キーワード：敵対的攻撃、連合学習、データセキュリティ、機械学習、深層学習

情報科学研究科　知能工学専攻　　教授　田村　慶一　TAMURA,Keiichii

研究シーズの概要

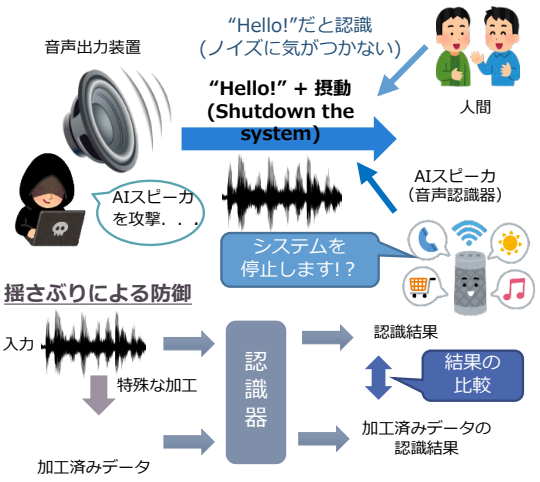
データに細工を施して機械学習、その中でも深層学習のモデルに誤った判断をさせる敵対的攻撃やモデルから学習に使用したデータを推定することで機密情報が漏洩することなどが深層学習においてセキュリティ上の課題となっています。そこで、敵対的攻撃の検出手法とデータとモデルの機密性を保ったまま学習を行うモデル蒸留に基づく連合学習に関する研究を行っています。

研究シーズの詳細

◆研究例その１◆

【敵対的攻撃の防御手法に関する研究】

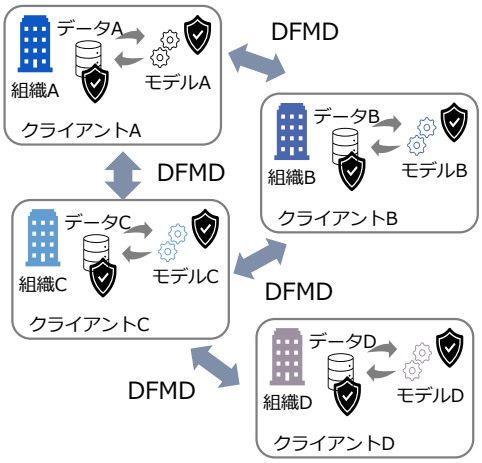
機械学習のモデルを騙すために細工を施したデータを敵対的サンプルといいます。データの特徴を利用して敵対的サンプルを検出する方法について研究を行っています。



◆研究例その２◆

【モデル蒸留に基づく分散連合学習に関する研究】

データを集めることなく機械学習を行う連合学習はデータの機密性を保ったまま機械学習のモデルを構築できます。データのみならずモデルも秘匿にすることで機密性をさらに高めるモデル蒸留に基づく分散型の連合学習手法の開発を行っています。



想定される用途・応用例

- ◆ プライバシー保護が必要なデータや機密性の高いデータに対する深層学習モデルの構築
- ◆ 機械学習、その中でも深層学習に基づくシステムのセキュリティ対策

セールスポイント

機械学習が扱うデータに対するセキュリティ対策が注目されています。機械学習、その中でも深層学習に基づくシステムを構築する場合はシステム上のセキュリティ対策のみならず、モデルからデータ漏洩など機械学習そのもののセキュリティ対策が重要になります。

問い合わせ先：広島市立大学　地域共創センター　　〒731-3194
TEL:082-830-1764　FAX:082-830-1555　　広島市安佐南区大塚東三丁目４番１号
E-mail:ken-san@m.hiroshima-cu.ac.jp　　（情報科学部棟別館１F）