秘匿性と実用性を兼ねるプライバシ保護技術とAIの社会適用

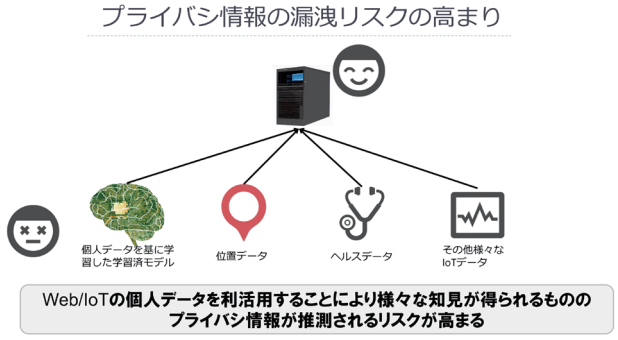
清 研究室

清 雄一
Yuichi SEI

日本では昨今の「個人情報の保護に関する法律（個人情報保護法）」の改正を受けて、特定の個人を識別できないように〝匿名化〟すれば、本人の同意を得ずに一定の条件下で個人データの自由な利活用が可能になりました。例えば、医療機関などで蓄積された医療ビッグデータが新薬の開発や治療効果の分析などに役立てられ始めています。

一方で、それは個人情報の漏えいリスクと常に隣り合わせであるといえます。かつて米国では、名前を伏せた医療データを作成したものの、「誕生日」や「性別」、「自宅の郵便番号」といった情報から個人が特定可能なことが指摘され、公開を取りやめた経緯があります。現にこの三つの情報だけで、米国の全人口の87％が特定できるとの研究報告もあります。

さらに、IoT（モノのインターネット）の進展により、今後、個人に関する多様かつ大量のデータが生成されるようになると、例え匿名化していたとしても、データの結びつきから個人が特定されてしまうリスクもでてきます。欧州連合（EU）では、個人データを集めたり利用したりするEU域内の企業などを対象に、個人情報に関する厳密な保護ルールである一般データ保護規則（GDPR）を課しています。



ノイズを与えて保護する

このような背景において、清雄一教授は十分な秘匿性を確保しつつ、ビッグデータの解析にも適した実用性の高い「プライバシー保護データマイニング（マイニング：採掘）」技術を提案しています。その中でも、ノイズを与えることでプライバシを保護する「ローカル差分プライバシ」と呼ばれる手法を研究しています。

この手法は、米アップルや米グーグルなども導入している最新のプライバシ保護手法です。与えるノイズが大きいほどプライバシを強く保護できますが、ノイズを大きくしてプライバシを過剰に保護してしまうと、データの有用性が下がって利活用しにくいというトレードオフの関係があります。

複雑なIoTデータを対象に

ローカル差分プライバシはこれまで、年齢や性別など誤差がなく種類も少ない従来型のデータを対象にした研究が主流でしたが、清教授は環境中に埋め込んだセンサなどによって観測されるIoTデータを対象にしています。例えば、画像認識により推測された年齢や性別、また体温や発汗量などから推定される新型コロナウイルスへの感染の有無など、IoT環境から得られたデータは多種類なうえ、観測誤差や欠損も含みます。さらに、IoT環境ではデータを収集するサーバも一つではなく、一人の人のデータが複数の

キーワード

プライバシ、データマイニング、人工知能、ソフトウェア工学

所　属	大学院情報理工学研究科 情報学専攻
メンバー	清 雄一 教授
所属学会	情報処理学会、電子情報通信学会、日本ソフトウェア科学会、米国電気電子学会（IEEE）、米国電気電子学会コンピュータ学会（IEEE Computer Society）
E-mail	seiuny@uec.ac.jp

サーバに送られるという状況にあります。

こうした複雑かつ多様なＩｏＴデータについて、清教授はノイズを乗せることでプライバシを守りつつ、データを大量に集めた場合でも個人を特定されずに、統計解析や機械学習を高精度に行えることを確認しました。清教授は「観測誤差を考慮してノイズを加え、さらに統計的に処理することで分析精度が向上した」とそのポイントを語ります。

IoT環境で収集されるデータの特徴

AIで河川の水位を予測

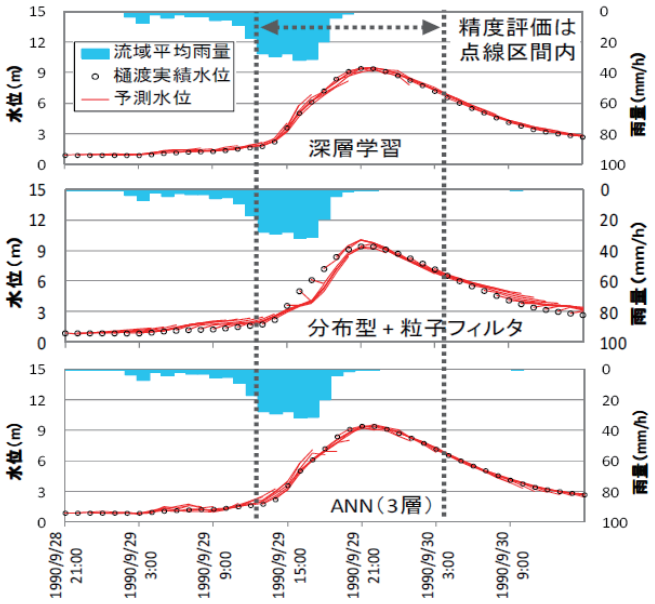
また、全く異なるテーマとして、人工知能（AI）技術を使って物理現象を推測する研究にも取り組んでいます。一つの例が、AIを使った河川の水位予測です。河川の水位予測は、従来、物理モデルや一般的な機械学習を使ったモデルの導入にとどまっていました。これに対して、清教授は日本工営（株）と共同研究を進め、昨今注目されているAI技術の一つであるディープラーニング（深層学習）を初めてこの領域に適用し、予測の精度を向上させました。

ディープラーニングの活用により、従来手法よりも予測誤差を数十％減らすことができました。大雨や洪水時に河川の水位をリアルタイムに測定し、例えば、現在から6時間後までの1時間ごとの水位を随時更新しながら予測できれば、河川が危険水域を越えた際などに、素早く的確な警報を発することができるかもしれません。

「炎上」や隠語の予測なども

そのほかツイッターの投稿データをAIで分析し、投稿内容にデマが含まれていないかどうかを調べる真偽の予測や、〝炎上〟しそうな投稿の予測などのほか、麻薬などの隠語を検出するといった興味深い研究も行っています。サッカーのPK戦において、選手やボールの位置や速度から、攻撃側と守備側がどのようなポジショニングを取るべきかを推測するゲーム理論の手法を用いた研究なども手がけているそうです。

【取材・文＝藤木信穂】



河川水位の予測における実データとの一致度
（一番上が提案手法、それ以外は従来手法）