

## OS-06 「情報の保護と中立性に配慮したデータ分析」

オーガナイザ：中川 裕志（東京大学）  
 神寫 敏弘（産業技術総合研究所）  
 佐久間 淳（筑波大学）  
 荒井 ひろみ（理化学研究所）

データマイニング、情報推薦、情報検索など種々の IT がサービス化し、これらを多くの利用者が使っている現在、IT サービスの社会性への配慮が重視されつつある。例えば、データマイニングにおける個人情報ないしプライバシー情報一般の保護、情報推薦システムにおける推薦の公正性、中立性の確保、情報検索における質問者のプライバシーである検索質問の秘匿、地図や動画サイトにおける映像情報から個人を特定できないようなぼかし処理など、IT サービスが社会的に受け入れられるための配慮すべき点は枚挙にいとまがない。この社会性配慮のためには原データないし結果データに何からの摂動処理、例えば差分プライバシーに見られる結果への雑音加算が有力だが、その一方で得られた結果の正確さも維持しなければならない。このような問題意識のもと、このオーガナイズドセッションでは、情報の保護と中立性に配慮した IT サービスを実現するためのデータ分析、制御、マイニング、ないしデータ公開における重要なテーマごとに招待講演 5 件と一般講演 4 件、および総合討論を 3 セッションに分けて行い、将来の方向性を議論した。参加者は 1 セッション当たり平均 25 名であった。以下では招待講演の概要と総合討論について報告する。

「中立性・公正性に配慮したデータ分析」(神寫 敏弘：産総研)では、データベース内のデータの分布や相関ルールによって、試験など個人へ影響のある事柄で不公正な判断がされたことを検出すること、および不公正な判断の原因となったデータベース中の特徴量を変換して不公正な判断が起きないようにする技術に関する現状について報告された。データを故意に用いて不公正な判断をしない場合も問題が生じ得る。総合討論において神寫氏は、さらに難しい問題として、ラタニャ・スウィーニーという黒人系の名前(著名なプライバシー保護研究者)で検索すると「逮捕」のようなネガティブなキーワードが示唆されるが、典型的なアングロサクソン系の名前で検索すると中立的なキーワードしか表示されない例を示した。実際は、検索システムは故意にキーワードを細工したわけではなく、収益を最大化するために、利用者のクリック率を最適化する過程での副作用であった。したがって、公正性の確保はデータを偏見なく利用するだけでは実現できない例もあり、難しい問題であることが報告された。

「官庁統計の情報保護基準」(星野伸明：金沢大学)では、官公庁の提供する公的統計における開示リスクすなわち、公的統計から個人識別が可能である確率を定式化している。この確率は、(a) 公開データにおける個人識別のために照合する擬似識別子(住所、年齢、性別など)が同じ基準で記録されている、(b) 公開ファイルに個人が含まれている、(c) 個人が母集団一意である、という事象の結合確率をもとに計算されることを示した。公的には、第三者委員会における個人識別可能性の基準の明確化が重要である。

「差分プライバシーについての暗号理論的な考察」(松田隆宏：産総研)では、暗号論的立場からの差分プライバシーを分析している。すなわち差分プライバシーは、計算能力無限の攻撃者に対し、識別不可能性を保証している点で強力である。一方で、条件に一致するデータ件数を質問する場合、同じ質問の繰返し回数が増えると統計量としての精度が上がるが、プライバシーは危険になるトレードオフについて説明している。

「プライバシーを考慮した防犯カメラ映像処理」(船富卓哉：京都大学)では、道路や店舗などの監視カメラから得たビデオ画像を公開するにあたって、個人情報を保護する方法を紹介した。すなわち、人間の映っている部分を予測、認識し抽象的な図形(円筒)で機械的に置き換える手法の研究成果を報告している。

「位置情報サービスにおけるプライバシー保護技術」(川本淳平：筑波大学)では、個人の滞在した位置情報を含むデータベースを利用しないし公開するにあたってのプライバシー保護手法について紹介している。位置情報を存在した点ではなく  $k$  人を含む領域に拡大して  $k$ -匿名性を確保する方法があるが、領域が大きくなりすぎる問題がある。また、移動履歴に関しては、長い距離の移動履歴を対象にするとデータがスパースになり  $k$ -匿名性の実現が難しくなるので、少ない地点間での移動履歴に限定する方向が研究されていることが報告された。

総合討論で佐久間は個人のゲノム情報が数十ドル払えば簡単に入手できる状況を指摘し、「ネット上であなたのゲノム情報、公開しますよ」という恐喝の可能性に触れた。健康保険料、就職などに不利益を生じかねず、重大な問題である。一方、生体認証はパスワード認証より安全と思われているが、指紋のような生体情報は一度盗まれたら、パスワードのように取り替えがきくというものではなく、かえって危険性があることも指摘された。

近未来の社会と個人と情報の関係についての深い議論ができた有意義なセッションであった。

[中川 裕志(東京大学)]