

# 医療情報の匿名加工に関する取り組みと課題

国立保健医療科学院 木村映善

医療情報は要配慮個人情報があり、データの次元数も大きい故に、匿名化に関する課題が多数ある。そのため、丁寧なオプトアウトを前提として医療情報の収集を行い、研究開発に医療情報を提供できるように、「医療分野の研究開発に資するための匿名加工医療情報に関する法律」が制定された。一方、医療情報をオープンデータとして公開するための匿名化処理の検討は未だ途上である。本発表では、医療情報の匿名加工の議論に携わってきた経験から、医療情報の匿名化を困難する固有の課題、特に背景知識に関連があるものを紹介する。

## (1)再構成による特定個人の識別 - 精度の高い医療情報・画像の増大

例えば、CTやMRI等の断層撮影の結果は、複数のスライス画像となって出力される。各々のスライス画像を見ても特定個人を識別できないが、三次元再構成によって顔画像を生成できる。また、骨折やインプラント等は部位と形状から、レセプト情報の処置情報等と照合して特定するための有力な手がかりになる。ただちの定量的評価が困難であるけれども、機械学習によって抽出された特徴量を用いての照合可能性がこれから出てくるであろう。

## (2)身体特徴量 - IoT/ウェアラブルセンサー時代を見据えて

身体的をセンサーで測定して抽出された特徴量は特定個人の識別性が高いものについては、バイオメトリクス認証に使われることがあり、改正個人情報保護法2条2項1号では、個人識別符号として認定される身体的特徴がある。しかし、心電図等、単体では外見性、特定個人識別性がないものであっても、GPSによる位置情報、時間情報がメタデータとして付与されれば、経時的に計測情報が蓄積され、特定個人を識別されるリスクが高まる。今後、身体的特徴量に対する機械学習を利用して、治療効果を早期に予測、介入するような医療アプリケーションの開発が期待されるため、身体的特徴量を利用するデータの扱いについても整理が必要である。

## (3)背景知識の適用範囲の広さ

医療において時間や順番に関する情報も重要な手がかりを提供する情報である。例えば、入院時のレセプトでセンシティブな病名が伏せられていても、特定の疾患のみに利用される医薬品の存在で推測できる可能性がある。また、複数の疾病に適用できるような医薬品であっても、一般に公開されているクリニカルパスやプロトコル等と、処方の日付、パターンを突き合わせることで高い確度で類推することが可能である。また、医療分野はオントロジーの整備が精力的に取り組まれており、概念の階層関係の定義も豊富に公開されている。特定の語彙で照合できなくても、オントロジーを利用して上下概念も対象として照合のための探索範囲を広げることが可能である。

以上に述べたように、これまでの単純な照合のキーとなる定量的な情報のみならず、情報の再構成、時系列的蓄積、オントロジー等を活用した基礎的な背景知識を獲得する手法と機械学習を組み合わせることで、これまで以上に特定個人を識別できる可能性がある。匿名化手法の検討において機械学習への対抗手段や運用を検討していくことが重要である。