

個人特定性低減のための 加工について

2014年5月13日

中央大学経済学部

伊藤 伸介

1. 「(仮称)準個人情報」の定義等について

①個人または個人の情報通信端末等に付番され、継続して共用される情報

例えば、「雇用保険被保険者番号」は該当するのではないか。雇用保険被保険者番号は、個人に紐づけされているだけでなく、転職した場合でも個人に対して固定された番号になっていると考えられる(ビジネス目的での使用は困難と思われる)。

②個人の生体的・身体的特性に関する情報

「顔認識データ」のように「それだけで」一意に特定することが可能な情報と、「人種」のように(地域を狭める等の)条件付きで一意の特定が可能な情報については区別したほうがよいのではないだろうか。

③特徴的な行動の履歴

(例えば)移動履歴に限らず、一意に該当しかつ特異なレコードについては、「特徴的」なレコードと考えられる。

・「①個人または個人の情報通信端末等にかかわる」**個体識別情報**(名前, 住所等の直接的な識別子等)を任意の一連番号に置き換えたデータとしても, 変数の組み合わせによって一意となる場合が生じる。

・「②個人の身体的特性にかかわる」**情報**についても, 詳細な地域情報が含まれる場合等を勘案すると, 列削除だけでは十分でない場合があると思われる。

* 属性によっては, 列削除のような手法を用いた場合, 情報量損失はむしろ大きくなる。

①や②に該当する「(仮称)準個人情報」を「(仮称)個人特定性低減データ」に加工するためには, 列削除(直接的な識別子の削除等)だけでなく, さらなる加工が必要かと思われる。

・「③特徴的な行動の履歴」については、「特殊な一意(special uniques)」に該当するレコードがそれに該当する可能性がある。「疫学的に特異であるために、本質的に(intrinsically)まれな属性群の組み合わせを有する」レコードは「特殊な一意」だと考えられる (Elliot(2001))。

←政府統計の分野で用いられている用語

⇒少数のキー変数の組み合わせでも一意になるレコードは、特殊な一意に該当するとみなされる。

→特殊な一意に該当するレコードの探索および特殊な一意に対する加工方法の検討

・「③特徴的な行動の履歴」のなかで、位置データと日時を含む移動履歴のような情報に関しては、例えば、時点(時間)情報や地点(地域)情報を粗くする等の措置を行うことによって、特定性低減の程度を高める必要がある。

2. 「(仮称)個人特定性低減データ」を作成するための加工方法について

政府統計マイクロデータにおける技術的な匿名化措置

1) 非攪乱的な手法

レコードの一部抽出(サンプリング, リサンプリング), 再符号化(リコーディング), トップ(ボトム)・コーディング, レコード削除等

2) 攪乱的な手法

スワッピング, ノイズの付加, ミクロアグリゲーション等

* わが国の政府統計の匿名データについては, 直接的な識別子の削除だけでなく(技術的な手法としては, 最低限必須の加工方法という位置づけ), レコードの一部抽出(サンプリング, リサンプリング), 再符号化, レコード削除といった非攪乱的な手法が基本的な匿名化技法として用いられてきた(必須ではないと思われるが優先度は高い)。

* 国勢調査の匿名データにおいては, 上記の非攪乱的な手法だけでなく, 攪乱的手法としてスワッピングが適用された。

・わが国の政府統計の匿名データの作成においてレコードの一部抽出(サンプリング, リサンプリング)は, 有力な技法の1つとしてこれまで用いられてきた。

→「個人データ」及び「準個人データ」から「個人特定性低減データ」へ加工するための方法の1つとして, レコードの一部抽出(サンプリング)を行うことによって, 個人特定性の低減が可能になると思われる。

⇒レコードの一部抽出を行う場合, 抽出率や抽出方法を公開しないということも考えられる。

・「(仮称)個人特定性低減データ」の作成において「特徴的な行動履歴」を考慮するために、少数のキー変数の組み合わせに基づいて、特殊な一意に該当すると思われるレコードを探索することが考えられる。



「特徴的な行動履歴」については、該当する変数をすべて削除するよりも、特殊な一意に該当するレコードを削除するほうが情報の損失を小さくする意味では望ましい方法だと考える。

提供先(受領者)に課せられる制約について

- 「(仮称)個人特定性低減データ」の提供元は、その安全性について第三者機関による承認が与えられたとしても、それを提供することによって生じる責任を負っていると思われる。
- 他方、「(仮称)個人特定性低減データ」の提供先(受領者)においても、その安全な利用や節度ある分析について技術的な自主的な考慮や、情報の保管や取扱いなど運用面に関する配慮が必要であろう(インターネット回線に接続された状態での使用に対する規制は必要か?)。

⇒運用面では、「適正管理」や「データアクセス・提供のシステム」(オンサイト施設等)、「守秘義務」、個人情報情報が漏えいした場合の「罰則・ペナルティ」に関する法整備が求められるのではないか。

参考

統計法

調査票情報等の管理及び情報漏えい等の対策に関するガイドライン(総務省政策統括官(統計基準担当)決定)

第三者機関に届出る項目について

- ・提供される個人特定性低減データの概要

名称, 種類, データが収集された時期(期間), データに含まれる属性, データに含まれるレコード数 等

- ・個人特定性低減データの加工方法

- ・個人特定性低減データ作成のための"最低条件"の適用状況
- ・"最低条件"以外の加工の方法

⇒ここで、加工方法には、第三者機関へ届け出ることが可能な情報と、それが困難な情報があると思われる。(届け出するための資料の作成が容易でない等の理由で)

参考文献

Elliot, M. (2001) “Disclosure Risk Assessment”, Doyle *et al.*(eds.)*Confidentiality, Disclosure, and Data Access: Theory and Practical Application for Statistical Agencies*, Elsevier Science, Amsterdam, pp.75-90.