

表データの最適セル秘匿処理に対するマッチング攻撃の実証的評価

統計数理研究所 南 和宏

1. はじめに

度数分布表、集計表等の表データは公的マイクロデータの分析における基本となる記述統計である。しかし、表データのセル値は機密情報を含む場合があり、適切な表セル秘匿処理が必要となる。この問題は、行計、列計の線形式を内包する表データに対し、機密情報を含む1次秘匿セルの集合の値の保護を拘束条件とし、情報損失を最小化する2次秘匿セルの集合を決定する問題として定式化される。一次秘匿セルの安全性の要件は取りうる値の不確定性で評価し、具体的には可能なセル値のインターバル幅を指標とする。

2. 現状の秘匿処理ツールの課題

現在、幾つかの表データのための秘匿処理ツール [1,2]が開発されているが、提供される最適なセル秘匿処理は決定論的アルゴリズムとして実装されている。したがって、秘匿された表データの秘匿セルを候補値で補完した表データに対して、秘匿処理を実行して、もとの秘匿パターンが再現されるかどうかのマッチングを行うことで候補値の組の集合を絞り込みが可能となり、その結果、秘匿セルのインターバル幅が想定よりも絞り込まれるリスクが生じる。

3. マッチング攻撃の実証的評価

表データに含まれる複数の秘匿セルの候補値の全ての組み合わせは、行計、列計に関する不定方程式を解くことで、零空間を構成する列ベクトルの係数の組み合わせとして表現できる。我々は候補値の組で補完した候補の表データに前述のマッチング検査を実施し、各機密セルの「実際の」インターバル幅を算出する攻撃プログラムを開発した。人工的に生成した異なるセル数の表データに対してこのマッチング攻撃を実施した結果、約40～60%の一次秘匿セルについて安全性要件が破られてしまうことが判明し、特に不定方程式の零空間の次元が小さいと秘匿セル値の絞り込みが顕著になることを確認した。

参考文献

1. Minami, K.; Abe, Y. Statistical Disclosure Control for Tabular Data in R. Romanian Statistical Review 2017, pp. 67–76.
2. tau-ARGUS homepage. <http://research.cbs.nl/casc/tau.html>.