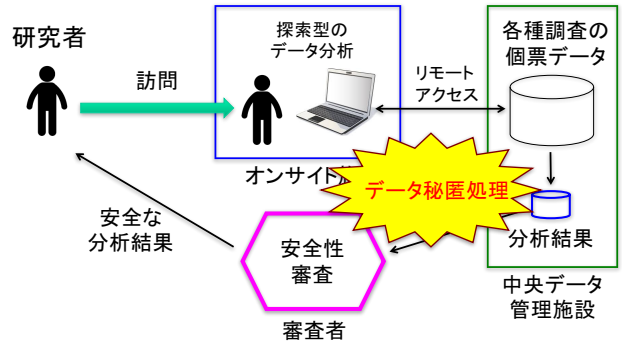
表データのセル秘匿処理の最適化

南　和宏　　モデリング研究系　准教授

オンサイト利用の安全性審査



表データからの情報漏えい

外部者攻撃　　内部者攻撃

地域、職種から個人を特定

自分の属性は知っている

度数分布表

居住	職種	P_2	P_3	P_4		
M_1	20	15	30	20		
M_2	72	20	1	30	10	133
M_3	38	38	15	40	2	133
合計	130	73	46	90	22	361

集計表

	P_1	P_2	P_3	P_4	P_5	合計
M_1	360	450	720	400	360	2290
M_2	1440	540	22	570	320	2892
M_3	722	1178	375	800	363	3438
合計	2522	2168	1117	1770	1043	8620

収入の合計

収入を特定

自分の収入を引けばもう一人の収入が分かる

表セルの１次秘匿

度数分布表
- 最小度数ルール

集計表
- 占有性ルール
  - (n, k)-ルール
  - p%ルール

表セルの２次秘匿

- 秘匿セル変数の可能範囲（*秘匿インターバル*）の幅wの長さがしきい値 t（*度数分布表では10*）以上であること

	P_1	P_2	P_3	合計
M_1	x_{11}	24	x_{13}	72
M_2	x_{21}	38	x_{23}	116
M_3	40	39	42	121
合計	98	101	110	309

1. 最小値問題

$a_{23} = \min x_{23}$

拘束条件:
$x_{11}+x_{13} = 72-24$
$x_{21}+x_{23} = 116-38$
$x_{11}+x_{21} = 98-40$
$x_{13}+x_{23} = 110-42$
$(x_{11},x_{13},x_{21},x_{23}) \geq 0$

and

2. 最大値問題

$\overline{a_{23}} = \max x_{23}$

拘束条件:
$x_{11}+x_{13} = 72-24$
$x_{21}+x_{23} = 116-38$
$x_{11}+x_{21} = 98-40$
$x_{13}+x_{23} = 110-42$
$(x_{11},x_{13},x_{21},x_{23}) \geq 0.$

⇨ *秘匿インターバル* $w = max\, x_{23} - min\, x_{23} = 68 - 20 = 48 > 10$

秘匿セル数の最小化問題

- 秘匿パターン　$y_i \in \{0,1\} \quad i = 1,\ldots,n$

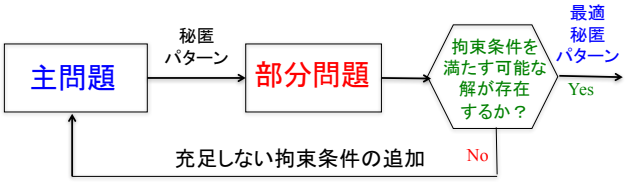
10	NA
5	80

$\longleftrightarrow$ (0,1,0,0)

- 目的関数: 秘匿セル数　$\sum_{i=1}^{n} y_i$
- 拘束条件: 秘匿パターンに対応するテーブルの安全性

Benders分割による効率化

- 主問題と部分問題に分割
  - 主問題: 秘匿パターンの最適化
  - 部分問題: 各秘匿セルの拘束条件のチェック
- 大部分の問題で効率的に実行
- アルゴリズムが終了した場合は、*最適解を保証*



● 主問題

Minimize　$\sum_{i=1}^{n} y_i$

subject to:　$y_p = 1 \quad \forall p \in \mathcal{P}$

$y_i \in \{0,1\} \quad i = 1,\ldots,n$

$v^{jT} \geq \beta^j \quad j \in \Phi$　　y_iに関する拘束条件（最初は空集合）

● 部分問題

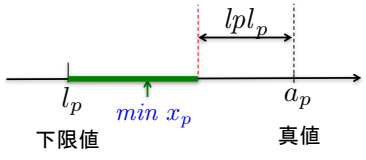
$-lpl_p \geq min \quad x_p$

subject to:　$Ax^{l,p} = 0 \quad \lambda$

$\begin{pmatrix} Ax^{l,p} \geq 0 \\ (-A)x^{l,p} \geq 0 \end{pmatrix}$

$x_i^{l,p} \geq (l_i - a_i)y_i \quad i = 1,\ldots,n \quad \mu_l$

$x_i^{l,p} \leq (u_i - a_i)y_i \quad i = 1,\ldots,n \quad \mu_u$



今後の課題

- 多次元テーブルへの対抗
- 差分攻撃への対策として、複数テーブルの同時処理