

■ 目 次

■ シリーズの刊行にあたって	iii
■ まえがき	v
Chapter 1 第 1 章 データ解析におけるプライバシー保護技術の概要	1
1.1 仮名化・匿名化によるデータの外部提供	1
1.2 差分プライバシーを適用した統計量公開	2
1.3 秘密計算によるデータ解析	4
Chapter 2 第 2 章 パーソナルデータ提供におけるプライバシーの問題	7
2.1 パーソナルデータ提供にかかわるプライバシーの侵害	8
2.1.1 マサチューセッツ州の事例	10
2.1.2 AOL の事例	11
2.1.3 Netflix の事例	13
2.1.4 NY 市 Taxi Ride の事例	15
2.2 パーソナルデータの利用とプライバシー保護技術	17
Chapter 3 第 3 章 パーソナルデータ提供におけるデータの構成要素	19
3.1 パーソナルデータの実例とデータの種別	19
3.2 識別情報	20
3.2.1 直接識別情報と間接識別情報	20
3.2.2 識別と特定の違い	23
3.3 履歴情報	24
3.3.1 識別情報として働く履歴情報	25
3.3.2 特異性	25
3.3.3 習慣性	25

3.3.4	一意性	26
3.4	要配慮情報	26
3.5	識別情報／要配慮情報と履歴情報の境界	27
3.6	連絡情報	27
3.7	個人に被害を与える情報	28
3.8	データベースに関する情報	28
3.9	個人情報保護法との関係	29
3.10	技術と法制度	31

Chapter 4

第 4 章	パーソナルデータ提供のリスクと有用性	35
4.1	データ提供のプロセス	35
4.1.1	データ提供における自明なリスクと非自明なリスク	37
4.1.2	データ提供において想定する攻撃者	37
4.2	個人属性データ提供に伴う特定と連結	38
4.2.1	個人属性データにおける特定と連結	39
4.2.2	個人属性データにおける特定を経ない連結	41
4.3	履歴データ提供に伴う特定と連結	42
4.4	データ提供に伴う特定のリスク評価: k 匿名性	43
4.5	データ提供に伴う特定のリスク評価: 標本一意性と母集団一意性	45
4.5.1	母集団一意であるレコード数の推定	47
4.5.2	母集団一意かつ標本一意であるレコード数の推定	49
4.6	個人属性データ提供に伴う属性推定	50

Chapter 5

第 5 章	パーソナルデータの匿名化	53
5.1	パーソナルデータの匿名化のプロセス	53
5.2	仮名化における直接識別情報の扱い	54
5.2.1	仮名 ID の構成	55
5.2.2	対応表による仮名化	55
5.2.3	鍵付きハッシュ関数による仮名化	56
5.3	匿名化における間接識別情報の扱い	57

5.3.1	再符号化	58
5.3.2	トップコーディングとボトムコーディング	59
5.3.3	抑制	59
5.3.4	マイクロアグリゲーション	60
5.3.5	加工方法の適用例	61
5.4	一般化階層構造に基づく k 匿名化	62
5.4.1	一般化階層構造	62
5.4.2	有用性と匿名性のトレードオフ	64
5.4.3	最適な k 匿名化は NP 困難	66
5.4.4	Incognito	66
5.5	仮名化／匿名化データの提供における注意点	69
5.5.1	仮名化／匿名化データの並び順	69
5.5.2	履歴データの仮名化／匿名化	69
第 6 章 識別不可能性と攻撃者モデル 73		
6.1	計算と秘匿性	73
6.2	記法	75
6.2.1	多項式時間アルゴリズムと多項式領域アルゴリズム	75
6.2.2	決定的アルゴリズムと確率的アルゴリズム	76
6.2.3	無視できる関数	76
6.2.4	確率	77
6.3	識別不可能性	77
6.4	情報理論的識別不可能性	78
6.4.1	情報理論的識別不可能性の定義	78
6.4.2	情報理論的識別不可能性に基づく秘匿性	79
6.5	計算量的識別不可能性	80
6.5.1	計算量的識別不可能性の定義	80
6.5.2	計算量的識別不可能性に基づく秘匿性	81
6.6	識別不可能性に基づく秘匿性と攻撃者モデル	81
6.6.1	情報理論的識別不可能性における攻撃者モデル	82
6.6.2	計算量的識別不可能性における攻撃者モデル	82
6.7	データ匿名化が想定する攻撃者モデル	83

第 7 章 統計量の公開における差分プライバシーの理論 85

- 7.1 統計量の公開 85
- 7.2 統計量公開におけるプライバシー 86
 - 7.2.1 独立性検定 88
 - 7.2.2 事例 1: 統計量公開がプライバシーの侵害を起こしていない例 89
 - 7.2.3 事例 2: 統計量公開がプライバシーの侵害を起こしている例 91
- 7.3 完全秘匿性に基づく安全性の議論 93
- 7.4 完全秘匿の不可能性 95
- 7.5 「弱い秘匿性」の実現 96
 - 7.5.1 まったく秘匿性がないケース 97
 - 7.5.2 「弱い秘匿性」があるケース 97
 - 7.5.3 「弱い秘匿性」は確率アルゴリズムによって実現される 98
 - 7.5.4 「弱い秘匿性」と有用性 98
- 7.6 ϵ -差分プライバシー 99
 - 7.6.1 差分プライバシーは「弱い秘匿性」を保証する 100
- 7.7 (ϵ, δ) -差分プライバシー 101
- 7.8 ϵ の解釈と隣接性の定義 102
- 7.9 δ の解釈 104
- 7.10 差分プライバシーにおける攻撃者モデル 105
 - 7.10.1 差分プライバシーにおける攻撃者の背景知識 106
 - 7.10.2 差分プライバシーにおける攻撃者の攻撃アルゴリズム 106
 - 7.10.3 事後分布の差による攻撃の評価 107
 - 7.10.4 semantic privacy 108
 - 7.10.5 semantic privacy と差分プライバシーは等価である 109

第 8 章 差分プライバシーのメカニズム 113

- 8.1 確率アルゴリズムとしてのメカニズム 113
 - 8.1.1 randomized response 113
- 8.2 メカニズムの評価基準 115
- 8.3 ラプラスメカニズム 117

8.3.1	ℓ_1 敏感度	117
8.3.2	ラプラス分布によるランダム化	119
8.3.3	ラプラスメカニズムのプライバシー	119
8.3.4	ラプラスメカニズムの正確性	120
8.3.5	ラプラスメカニズムに基づくクエリ	121
8.3.6	ラプラスメカニズムの事例	122
8.4	ガウシアンメカニズム	123
8.4.1	ガウシアンメカニズムのプライバシー	124
8.5	指数メカニズム	124
8.5.1	指数メカニズムのプライバシー	125
8.5.2	指数メカニズムの正確性	126
8.5.3	指数メカニズムの事例	127
8.6	レコードの独立性	128
8.7	複数回のクエリに対する差分プライバシーの保証	129
8.7.1	差分プライバシーの合成定理	129
8.7.2	最適な合成定理	130
8.7.3	同じクエリの複数回の問い合わせは得か	131
8.8	合成定理の応用	132
8.9	疎な出力	134
8.9.1	閾値メカニズム	135

第 9 章 差分プライバシーと機械学習 139

9.1	経験損失最小化	139
9.1.1	経験損失最小化による教師あり学習	139
9.1.2	汎化損失	141
9.1.3	正則化	141
9.2	経験損失最小化における差分プライバシー	142
9.2.1	差分プライバシーを保証した経験損失最小化の有用性	143
9.3	出力摂動法による差分プライバシーの保証	144
9.3.1	強凸性	144
9.3.2	正則化経験損失の目的関数の敏感度	145
9.3.3	正則化経験損失における出力摂動法の差分プライバシー	147

9.3.4	正則化経験損失における出力摂動法の有用性解析	148
9.4	目的関数摂動法による差分プライバシーの保証	149
9.4.1	正則化経験損失における目的関数摂動法	149
9.4.2	正則化経験損失における目的摂動法の差分プライバシー	150
9.4.3	正則化経験損失における目的摂動法の有用性解析	151

第 10 章	秘密計算の定式化と安全性	153
10.1	秘密計算	153
10.1.1	マルチパーティ秘密計算	154
10.1.2	アウトソーシング型秘密計算	156
10.1.3	秘密計算の実現	157
10.2	秘密計算プロトコル	158
10.2.1	イデアルモデルとリアルモデル	159
10.3	攻撃者モデル	160
10.4	秘密計算の正当性と秘匿性	161
10.5	秘密計算の秘匿性の定義	162
10.6	差分プライバシーと秘密計算における攻撃者の違い	164
10.7	秘密計算の攻撃者モデル	165
10.8	秘密計算の構成法	166

第 11 章	秘密鍵暗号と公開鍵暗号	167
11.1	秘密鍵暗号	167
11.1.1	秘密鍵暗号の定式化	167
11.1.2	秘密鍵暗号による通信	168
11.1.3	ワンタイムパッド	169
11.1.4	ワンタイムパッドの完全秘匿性	170
11.2	公開鍵暗号	170
11.2.1	公開鍵暗号の定式化	171
11.2.2	公開鍵暗号による通信	171
11.2.3	ElGamal 暗号	172
11.2.4	ElGamal 暗号の秘匿性	176

第 12 章	準同型暗号による秘密計算	181
12.1	準同型暗号	181
12.1.1	加法準同型暗号	181
12.1.2	乗法準同型暗号	182
12.1.3	完全準同型暗号	183
12.2	準同型暗号による秘密計算の安全性	183
12.3	準同型暗号による秘密計算: 独立性検定への応用	183
12.3.1	分割表計算の 2-party 秘密計算	184
12.3.2	分割表計算の 2-party プロトコルの秘匿性	187
12.3.3	分割表計算のアウトソーシング型秘密計算	188

第 13 章	秘匿回路による秘密計算	189
13.1	秘匿回路	189
13.1.1	秘匿回路の定式化	190
13.2	紛失送信	191
13.3	秘匿回路生成	194
13.4	秘匿回路評価	196
13.5	秘匿回路評価の秘匿性	196
13.6	秘匿回路評価の実行例	197

第 14 章	秘密分散による秘密計算	199
14.1	秘密分散	199
14.1.1	加法的シェアによる秘密分散	200
14.1.2	多項式による秘密分散	201
14.2	秘密分散による秘密計算	201
14.2.1	秘密分散による加算と公開された数の乗算	201
14.2.2	秘密分散による乗算: 非公開な数の乗算	204
14.3	秘密分散による汎用的な秘密計算と実装	205
14.4	秘密分散による秘密計算の安全性	206
14.5	秘密分散による秘密計算の実装	206

- 参考文献 207
- 索引 211

表 記法

記法	定義	記法	定義
\mathbb{N}	自然数集合	x	データ, 入力, 平文
\mathbb{Z}	整数集合	X	データの定義域
\mathbb{R}	実数値集合	n	データ数
\emptyset	空集合	d	データの属性数/次元数
$\mathbb{Z}/q\mathbb{Z}$	q を法とする剰余環	D	データ集合/データベース
$\{0, 1\}^*$	任意長のビット列	\mathcal{D}	データ集合/データベースの定義域
poly	(任意の) 多項式	f	データ解析関数
negl	無視できる関数	q	クエリ
\mathcal{A}	確率的多項式アルゴリズム	y	出力
κ	セキュリティパラメータ	Y	出力の定義域
H	暗号学的ハッシュ関数	Ω	確率変数の定義域
k	秘密鍵暗号の鍵	S	確率変数の定義域の部分集合
pk	公開鍵暗号の公開鍵	Δ	大域敏感度
sk	公開鍵暗号の秘密鍵	m	メカニズム
Gen	鍵生成関数	$ \cdot $	絶対値, 集合/ビット列のサイズ
Enc	暗号化関数	$\ \cdot\ $	ノルム
Dec	復号関数	$a\ b$	bit 例 a, b の連結
\oplus	加法準同型演算	$a \boxplus b$	bit 例 a, b の排他的論理和
\otimes	乗法準同型演算	$a \in_R A$	集合 A からのランダム要素選択