背景知識の違いによる匿名加工データへの攻撃者モデルの評価

伊藤 聡志（菊池研究室）

概要：匿名加工は，購買履歴データのような元データから個人が識別されることを防ぐために，個人識別情報を加工する技術である．データを匿名加工する際には，データを悪用しようとする攻撃者を想定し，リスクを評価する必要がある．しかしながら，データに対する攻撃者をどう想定したらよいかはいまだ不明である．本研究では，履歴データのある属性から背景知識を得る攻撃者を想定し，攻撃者の持つ背景知識に当てはまるレコード数とユーザ数を用い，データリスク評価の理論的なモデルを提案する．また，提案したモデルを用いて実際の履歴データのリスク評価実験を行う．

1. はじめに

匿名加工は，元データから個人が識別されることを防ぐために，個人識別情報を加工する技術である．企業や組織は収集したビッグデータを活用する際，そのデータ内の個人が再識別されるリスクを評価し，匿名加工することを求められる．一方，そのようなデータから個人を識別しようとする攻撃者は，公開されている匿名加工データだけでなく，追加の背景知識を用いることが予想される．しかしながら，データについてのどのような背景知識が危険であるのか，どういった攻撃者が危険であるのかは不明であった．加えて，購買履歴データのようなデータを匿名加工する際に，どの属性（列）を加工したらよいのかを判断するための指標も不明であった．

Domingo-Ferrer らはデータセットに対する攻撃者想定として，最大知識攻撃者モデルを提案した [1]．最大知識攻撃者モデルでは，攻撃者は元のデータセットと匿名加工されたデータの両方のすべてを背景知識として持っていることを想定されている．しかしながら，この攻撃者想定はあまりにも協力であり，現実的ではない．また El Eman らは，データセットに対して現実世界で実行されうる 4 種類の攻撃（故意の攻撃，故意でない攻撃，データ侵害，公開データ）を想定し，それらのリスクを測定した [2]．しかしながら，これらのリスク評価には「攻撃が行われる確率」といった主観的な値や，「データ侵害が発生する確率」といった時間や場合によって変化する値が用いられており，求めるのが困難であった．

本研究の目的は，現実的な攻撃者の想定とデータセットのリスク評価である．我々はレコードと属性によって構成される履歴データに注目し，ある属性から背景知識を得る攻撃者を想定する．攻撃者の持つ背景知識に当てはまるレコード数と顧客数を用い，データの危険度の理論的なモデルを提案する．また，提案したモデルを評価するために，公開データセットを用いた実験を行う．

2. 基礎定義

2.1 データモデル

本研究では，レコード（行）と属性（列）によって構成され，個人を表す識別子を持つ履歴データを研究する．記号等を以下のように定義する．

定義 2.1 履歴データを T とし，T のレコード数を m，ユーザ数を n とする．履歴 T の属性 X の取りうる値の集合を D_X とし，T における X のユニークな値の数を ω_X とする．すなわち，$\omega_X = |D_X|$ である．D_X の要素 x について，履歴 T で x を満たすレコード（行）インデックスの集合を R_x とし，x を満たすユーザの集合を U_x とする．T を匿名化してユーザ ID を仮名化した匿名化データを T' とする．

例 2.1 T の例として，3 人のユーザ（ユーザ 1,2,3）の 3 日間（2010/12/1〜2010/12/3）の購買履歴データ T_{example} を表 1 に示す．例えば，仮名 2 は 2010/12/1 にパンを購入していることがわかる．T_{example} は $m = 10, n = 3$ の履歴データであり，X=Date のとき，$D_X = \{2010/12/1, 2010/12/2, 2010/12/3\}$, $\omega_X = 3$ である．また，$x = 2010/12/1$ のとき，$R_x = \{1, 2, 3, 4\}, U_x = \{1, 2\}$ である．

2.2 攻撃者モデル

本研究では，攻撃者が履歴 T に属するユーザ u の属性 X についての背景知識 x を偶然得ることを想定する．

表 1　履歴データ T の例 T_{example}

User ID	Date	Time	Goods	Price	Number
1	2010/12/1	8:45	Bread	1.45	2
1	2010/12/1	8:45	Book	3.75	1
1	2010/12/1	20:10	Tea	0.85	2
2	2010/12/1	10:03	Bread	1.45	3
1	2010/12/2	15:07	Tea	0.85	3
3	2010/12/2	11:57	Bread	1.45	4
3	2010/12/2	11:57	Juice	1.25	4
3	2010/12/3	15:54	Book	3.75	1
3	2010/12/3	15:54	Tea	0.85	10
3	2010/12/3	15:54	Juice	1.45	10

定義 2.2　攻撃者が背景知識 x を得る確率 $Pr(x)$ は，x の T における頻度に比例する，すなわち，$Pr(x) = |R_x|/m$ である．また，T のレコード数 m と属性 X の種類数 ω_X は与えられているものとする．

匿名化データ T' を与えられた攻撃者は，背景知識として x を含む T のレコードにアクセスできるとき，対応する T' の仮名の真のユーザの候補として U_x を得る．従って，再識別を表す事象 idf が生起するリスクを，x の条件付確率として次のように定める．

定義 2.3　攻撃者が背景知識 x から個人を識別 (idf) する条件付き確率 $Pr(\text{idf}|x)$ を $Pr(\text{idf}|x) = 1/|U_x|$ とする．

定義 2.2,2.3 より，攻撃者が背景知識 x を得ることと，攻撃者が背景知識 x から個人を識別することの同時確率 $Pr(\text{idf}, x)$ は，

$$Pr(\text{idf}, x) = Pr(x)Pr(\text{idf}|x) = \frac{|R_x|}{m}\frac{1}{|U_x|}$$

である．また，ここで $|R_x|/|U_x| = \alpha_x$ とおくと，

$$Pr(\text{idf}, x) = \frac{\alpha_x}{m}$$

とも表せる．α_x は x についてのユーザ当たりの平均レコード数 [レコード/人] を意味しており，本論文の解析に重要な役割を果たす．そこで，これを次のように定義する．

定義 2.4　背景知識 x による平均レコード数を α_x とする．属性 X における α_x の平均を α_X と表し，$\alpha_X = \frac{1}{\omega_X}\sum_{x\in D_X}\alpha_x$ とする．

例 2.2　T_{example} の Date 属性についての x, $|R_x|$, $Pr(x)$, $|U_x|$, $Pr(\text{idf}|x)$, $Pr(\text{idf}, x)$ を表 2 に示す．T_{example} の Date 属性の場合，$D_X = \{2010/12/1, 2010/12/2, 2010/12/3\}$ である．攻撃者が背景知識 x =2010/12/3 を得る確率は，$R_x = \{8, 9, 10\}$ であるため $Pr(x) = 3/10$ であり，その背景知識からユーザ u を識別できる確率は，$U_x = \{3\}$ なので，$Pr(\text{idf}|x) = 1/1$ となる．この場合，攻撃者が背景知識 x によって u を識別できる確率は $Pr(\text{idf}, x) = Pr(x)Pr(\text{idf}|x) = 0.3 \cdot 1 = 0.3$ である．または，$\alpha_x = 3/1 = 3$ であるので，

$$Pr(\text{idf}, x) = \frac{\alpha_x}{m} = \frac{3}{10} = 0.3$$

表 2　T_{example} の Date 属性に対する攻撃者の識別確率

x	$\lvert R_x\rvert$	$Pr(x)$	$\lvert U_x\rvert$	$Pr(\text{idf}\lvert x)$	$Pr(\text{idf}, x)$
2010/12/1	4	0.4	2	0.5	0.2
2010/12/2	3	0.3	2	0.5	0.15
2010/12/3	3	0.3	1	1	0.3
合計	10	1.0			0.65

である．

2.3　リスクモデル

本研究では以下に定義する平均識別確率 $Pr(\text{idf}, X)$ を，履歴 T の属性 X に関する危険度とする．

定義 2.5　(平均識別確率) 履歴 T の属性 X のある値を背景知識 x として与えられた攻撃者により，あるユーザ u が識別される確率 $Pr(\text{idf}|x)$ の期待値を，属性 X の平均識別確率 $Pr(\text{idf}, X)$ とする．

定義 2.4 より，

$$Pr(\text{idf}, X) = \sum_{x\in D_X} Pr(\text{idf}, x) = \sum_{x\in D_X}\frac{\alpha_x}{m}$$

である．

例 2.3　X =Date の場合，T_{example} の属性 X から背景知識 x を得た攻撃者の平均識別確率は

$$Pr(\text{idf}, X) = \sum_{x\in D_X}\frac{\alpha_x}{m} = \frac{2 + 1.5 + 3}{10} = 0.65$$

である．これは，攻撃者が T_{example} の Date 属性からあるユーザ u の背景知識を得たとき，u を平均 65% の確率で識別できることを意味する．

また，リスクの計算コストを以下のように定義する．

定義 2.6　リスク計算のコストは，計算に用いるレコード数に比例する．

例 2.4　履歴 T_{example} の全レコードの Price 属性の平均値を求める場合, 計算コストは 10 である．また, 2010/12/1 の Price 属性の平均値を求める場合, 計算コストは 4 である．

3.　リスク近似モデルの提案

平均識別確率を求めるためには，定義 2.5 より，履歴 T の属性 X に出現するすべての x について，α_x を求める必要がある．しかしながら，ビッグデータに対してすべての α_x を計算するのは困難であるため，これを近似する方法を検討する．平均識別確率を計算するモデルとして，本章では以下の 3 つを提案する．

(1) 平均モデル
(2) 最小コストモデル
(3) サンプリングモデル

3.1　厳密解

匿名化データ T' の再識別のリスクは，攻撃者に与えられる背景知識の属性 X に依存して決まる．そこで，X を

与えられた時の再識別リスク $R(X)$ を，属性 X の平均識別確率と定める．すなわち，$R(X) = Pr(\mathrm{idf}, X)$ とする．$R(X)$ の厳密解を求めるためには，履歴 T の属性 X に出現するすべての x について α_x を求める必要があるため，この場合の計算コストは m である．

3.2 平均モデル

平均モデルは，属性 X のリスクを α_x の平均 α_X を用いて求めるモデルである．以下のように定義を行う．

定義 3.1 平均モデルによって求められる属性 X のリスクを $R_{mean}(X)$ で示し，

$$R_{mean}(X) = \frac{\alpha_X \omega_X}{m}$$

とする．

平均レコード数の平均 α_X で定めた平均モデルのリスクは次のように厳密解を与えている．

定理 3.1 $R_{mean}(X)$ を定義 3.1 による，平均モデルによって求められるリスクとする．このとき，$R_{mean}(X) = Pr(\mathrm{idf}, X)$ である．

(Proof) 定義 3.1,2.4 より，

$$\begin{aligned} R_{mean}(X) &= \frac{\alpha_X \omega_X}{m} \\ &= \frac{\omega_X}{m} \sum_{x \in D_X} \frac{\alpha_x}{\omega_X} \\ &= \sum_{x \in D_X} \frac{\alpha_x}{m} \\ &= \sum_{x \in D_X} Pr(x) Pr(\mathrm{idf}|x) \\ &= Pr(\mathrm{idf}, X) \end{aligned}$$

であり，定理 3.1 を得る． (Q.E.D)

例 3.1 T_{example} の Date 属性の場合，$\alpha_X = (2 + 1.5 + 3)/3 = 13/6$ であるため，

$$R_{mean}(X) = \frac{\alpha_X \omega_X}{m} = \frac{13/6 \cdot 3}{10} = 0.65$$

である．

このモデルでは α_X を求める際に，履歴 T の属性 X に出現するすべての x について α_x を計算する必要があるため，この場合の計算コストは m である．

3.3 最小コストモデル

最小コストモデルは，全ての x について $\alpha_x = 1$ と近似して，属性 X のリスクを最小の計算コストで求めるモデルである．以下のように定義を行う．

定義 3.2 最小コストモデルによる属性 X のリスクを，

$$R_{cost}(X) = \frac{\omega_X}{m}$$

とする．

例 3.2 T_{example} の Date 属性の場合，

$$R_{cost}(X) = \frac{\omega_X}{m} = \frac{3}{10} = 0.3$$

である．

定義 2.2 より，T のレコード数 m と属性 X の種類数 ω_X は与えられている情報であり，このモデルでは履歴 T のレコードを用いて α_X 等を計算する必要が無いため，計算コストは 0 である．

3.4 サンプリングモデル

サンプリングモデルは，D_X からランダムに選んだ複数個の要素についての α_x を求め，これの平均を属性 X の平均レコード数 α_X の近似値であるとして属性 X のリスクを求めるモデルである．このとき，サンプリングするのは 1 つのレコードではなく，D_X からランダムに選んだ複数個の要素を満たすすべてのレコードであることを注意せよ．例えば，T_{example} の Date 属性のうち “2010/12/1” がランダムに選ばれた場合，T_{example} からこれを満たすレコード（この場合 4 レコード）をすべてサンプリングする．以下のように定義を行う．

定義 3.3 s をサンプリング数とし，$D'_X = \{x_1, \ldots, x_s\}$ を D_X からランダムにサンプリングされた，要素が s 個の部分集合とする．このとき，$\alpha_{x'} = \frac{1}{s} \sum_{i=1}^{s} \alpha_{x_i}$ とする．最小コストモデルによる属性 X のリスク $R_{sample}(X)$ を，

$$R_{sample}(X) = \frac{\alpha_{x'} \omega_X}{m}$$

とする．

例 3.3 T_{example} の X =Date 属性の場合，$s = 2$, $D'_X = \{2010/12/1, 2010/12/3\}$ とすると，$\alpha_{x_1} = 2$, $\alpha_{x_2} = 3$ であるため，

$$R_{sample}(X) = \frac{\alpha_{x'} \omega_X}{m} = \frac{2.5 \cdot 3}{10} = 0.75$$

である．

このモデルでの $\alpha_{x'}$ の計算コストは，D'_X の要素が $1/\omega_X$ で一様に選ばれるならば，これは $p = \frac{1}{\omega_X}$, 期待値 $\mu = \frac{m}{\omega_X}$ の二項分布であるため，sm/ω_X である．

4. 評価実験

4.1 実験目的

前節で提案したモデルを用いて，実際のデータに対するリスク評価実験を行う．実験のために，UCI Machine Learning Repository[3] より公開されている以下の 3 つのデータセットを用いる．

(1) T_1: Online Retail Data Set [4]
(2) T_2: Diabetes Data Set [5]
(3) T_3: Adult Data Set [6]

T_1, T_2, T_3 はそれぞれ，英国の 1 年間の購買履歴データ，10

表 3　公開データセット T_1, T_2, T_3 の統計量

	m	n	属性数
T_1	38,087	400	7
T_2	101,766	71,518	50
T_3	32,561	32,561	16

表 4　各モデルによって近似された平均識別確率

T	X	$R_{mean}(X)$	$R_{cost}(X)$	$R_{sample}(X)(s=10)$
T_1	購買時刻	*0.3217	0.0145	*[0.1411, 0.5998]
	購買日	0.1860	0.0076	[0.1267, 0.2786]
	購買商品	0.0965	*0.0730	[0.0718, 0.0982]
	単価	0.0121	0.0048	[0.0036, 0.0132]
	個数	0.0080	0.0025	[0.0017, 0.0152]
T_2	入院日数	*1.45E-04	*1.38E-04	*[1.46E-04, 1.52E-04]
	年齢	1.33E-04	9.83E-05	[1.21E-04 , 1.42E-04]
	人種	7.73E-05	5.90E-05	[6.92E-05, 8.31E-05]
	性別	3.78E-05	2.95E-05	[3.08E-05 , 4.30E-05]
T_3	年齢	*2.24E-03	*2.24E-03	*[2.24E-03, 2.24E-03]
	職業	4.61E-04	4.61E-04	[4.61E-04, 4.61E-04]
	婚姻状況	2.15E-04	2.15E-04	[2.15E-04, 2.15E-04]
	人種	1.54E-04	1.54E-04	[1.54E-04, 1.54E-04]

表 5　各モデルのコストと誤差

	計算コスト	誤差
平均モデル	38087	0
サンプリングモデル	131.3	0.073
最小コストモデル	0	0.178

年間の糖尿病患者・入院データ，国税調査による所得データである．各データの m, n, 属性数を表 3 に示す．

4.2　提案モデルの精度と計算コスト

T_1, T_2, T_3 の各属性の危険度を平均モデル，最小コストモデル，サンプリングモデルによって効率よく求める．表 4 に各モデルの評価値を示す．定理 3.1 により，平均モデルによる評価値 $R_{mean}(X)$ は平均識別確率の厳密解と一致する．サンプリングモデルによる評価値 $R_{sample}(X)$ は，$s = 10$ のときの 90%($\mu \pm \delta$) の信頼区間を示している．表中の*印がついている値は，そのデータで最も危険であると評価された属性のリスクである．例えば T_1 について，平均モデル（=厳密解）では購買時刻属性が最も危険であると評価されているのに対し，最小コストモデルでは購買商品属性が最も危険であると評価されている．サンプリングモデルにおいては，信頼区間の半順序関係における極大値となる属性は購買時刻であった．

表 5 に各モデルのコストと誤差の値を示し，図 1 に T_1 の購買日属性についての，各モデルの計算コストと誤差の散布図を示す．X 軸は計算コスト（レコード数）の対数であり，Y 軸は厳密解 $Pr(\mathrm{idf}, X)$ との絶対誤差である．図中の赤い点がこれらのモデルの結果を表している．灰色の点は D_X の ω_X 個の要素のリスク評価結果を示しており，それらの重心をサンプリングモデルの代表の点としている．サンプリングモデルはこれらの ω_X 個の点から s 個をランダムに選んでリスク評価をすることに注意せよ．

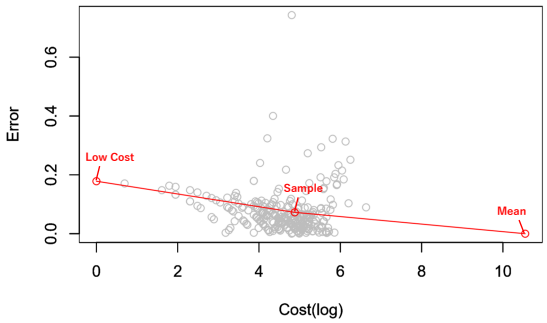

図 1　各モデルのコストと誤差の散布図

5.　まとめ

本稿では，履歴データのある属性から背景知識を得る攻撃者を想定し，その平均識別確率を用いてデータのリスク評価を行うモデルを提案した．また，平均識別確率を近似する 3 つのモデルを提案し，それらを用いて購買履歴データ，入院記録データ，世帯収入データの 3 つの実際のデータのリスク評価を行い，どの属性が危険であるのかを評価した．匿名加工をする際にこのリスク評価モデルを用いることによって，どの属性を加工・削除するか？等の加工指針を立てることができる．

参考文献

[1] Josep Domingo-Ferrer, Sara Ricci and Jordi Soria-Comas, “Disclosure Risk Assessment via Record Linkage by a Maximum-Knowledge Attacker”, 2015 Thirteenth Annual Conference on Privacy, Security and Trust (PST), *IEEE*, 2015.

[2] Khaled El Emam, Luk Arbuckle, “Anonymizing Health Data Case Studies and Methods to Get You Started”, *O’Reilly*, 2013.

[3] UCI Machine Learning Repository, `http://archive.ics.uci.edu/ml/index.php`, December 17, 2018.

[4] Online Retail Data Set, `https://archive.ics.uci.edu/ml/datasets/online+retail`, December 17, 2018.

[5] Diabetes 130-US hospitals for years 1999-2008 Data Set , `https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008`, December 17, 2018.

[6] Adult Data Set , `https://archive.ics.uci.edu/ml/datasets/adult`, December 17, 2018.

業績

(1) 伊藤 聡志，菊池 浩明，中川 裕志，“背景知識の違いによる匿名加工データの攻撃者モデルの分類と評価”，コンピュータセキュリティシンポジウム 2017 (CSS-2017)，pp. 1–8，2017

(2) Satoshi Ito, Hiroaki Kikuchi, Hiroshi Nakagawa, “Attacker Models with a Variety of Background Knowledge of Payment History”, MDAI-2018, USB proceedings, Spain, pp. 178–189, 2018.

他，国際学会 2 件，国内学会 3 件