(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2018-200529
(P2018-200529A)
(43)公開日　平成30年12月20日(2018.12.20)

(51)Int.Cl.		F I		テーマコード（参考）
G06F 21/62	(2013.01)	G06F	21/62	354

審査請求 未請求　請求項の数 4　OL　（全 11 頁）

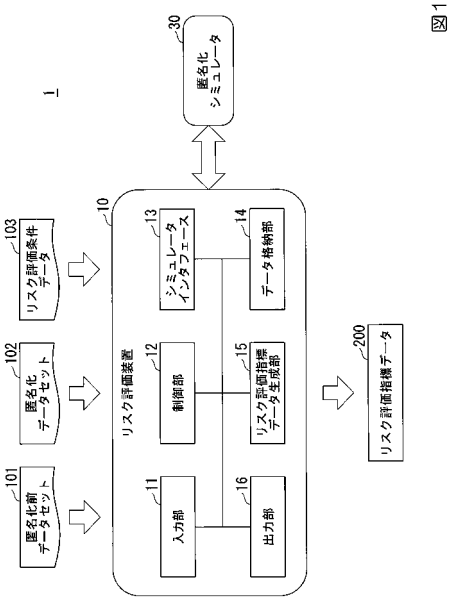
(21)出願番号　特願2017-104443 (P2017-104443)
(22)出願日　平成29年5月26日 (2017.5.26)

(71)出願人　000208891
ＫＤＤＩ株式会社
東京都新宿区西新宿二丁目３番２号
(74)代理人　100106909
弁理士　棚井　澄雄
(74)代理人　100064908
弁理士　志賀　正武
(74)代理人　100146835
弁理士　佐伯　義文
(72)発明者　三本　知明
埼玉県ふじみ野市大原二丁目１番１５号
株式会社ＫＤＤＩ総合研究所内
(72)発明者　清本　晋作
埼玉県ふじみ野市大原二丁目１番１５号
株式会社ＫＤＤＩ総合研究所内

(54)【発明の名称】リスク評価装置、リスク評価方法、及びコンピュータプログラム

(57)【要約】
【課題】攻撃者のモデルが弱いことを前提にすることによりリスク評価指標の有用性の向上を図る。
【解決手段】匿名化前データセットと匿名化データセットとリスク評価条件データを入力する入力部と、入力データセットを匿名化する匿名化シミュレータとの間でデータを送受するシミュレータインタフェースと、匿名化前データセットをシミュレータインタフェースにより匿名化シミュレータに入力して匿名化させる制御部と、リスク評価条件データの既知の個人情報データを使用して既知の個人情報データを有する個人情報レコードのみに匿名化データセットを絞り込み、絞り込み後の匿名化データセットと匿名化シミュレータのシミュレーション結果データセットとの比較を行い、比較の結果を使用してリスク評価指標データを生成するリスク評価指標データ生成部と、リスク評価指標データを出力する出力部と、を備える。
【選択図】図１

図1



【特許請求の範囲】
【請求項１】
匿名化される前の複数の個人情報レコードのデータセットである匿名化前データセットと、前記匿名化前データセットが特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により匿名化された匿名化データセットと、前記匿名化データセットのリスク評価条件を示すリスク評価条件データとを入力する入力部と、
前記特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により入力データセットを匿名化する匿名化シミュレータとの間でデータを送受するシミュレータインタフェースと、
前記匿名化前データセットを前記シミュレータインタフェースにより前記匿名化シミュレータに入力して匿名化させる制御部と、
前記リスク評価条件データが示すデータ絞り込み対象の既知の個人情報データを使用して当該既知の個人情報データを有する個人情報レコードのみに前記匿名化データセットを絞り込み、当該絞り込み後の匿名化データセットと前記匿名化シミュレータにより匿名化されたデータセットであるシミュレーション結果データセットとの比較を行い、当該比較の結果を使用して、個人情報の漏洩のリスク評価指標を示すリスク評価指標データを生成するリスク評価指標データ生成部と、
前記リスク評価指標データを出力する出力部と、
を備えるリスク評価装置。
【請求項２】
前記匿名化データセットは複数回の匿名化の各結果を含み、
前記リスク評価条件データは、リスク評価指標の閾値を含み、
前記リスク評価指標データ生成部は、あるデータ絞り込み対象の既知の個人情報データを使用した前記匿名化データセットのある回の匿名化の結果に対する前記リスク評価指標データのリスク評価指標が前記閾値以上である場合に、当該データ絞り込み対象の既知の個人情報データを使用して、前記匿名化データセットの次の回の匿名化の結果についての前記リスク評価指標データを生成する、
請求項１に記載のリスク評価装置。
【請求項３】
リスク評価装置が、匿名化される前の複数の個人情報レコードのデータセットである匿名化前データセットと、前記匿名化前データセットが特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により匿名化された匿名化データセットと、前記匿名化データセットのリスク評価条件を示すリスク評価条件データとを入力する入力ステップと、
前記リスク評価装置が、前記特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により入力データセットを匿名化する匿名化シミュレータとの間でデータを送受するシミュレータインタフェースステップと、
前記リスク評価装置が、前記匿名化前データセットを前記シミュレータインタフェースステップにより前記匿名化シミュレータに入力して匿名化させる制御ステップと、
前記リスク評価装置が、前記リスク評価条件データが示すデータ絞り込み対象の既知の個人情報データを使用して当該既知の個人情報データを有する個人情報レコードのみに前記匿名化データセットを絞り込み、当該絞り込み後の匿名化データセットと前記匿名化シミュレータにより匿名化されたデータセットであるシミュレーション結果データセットとの比較を行い、当該比較の結果を使用して、個人情報の漏洩のリスク評価指標を示すリスク評価指標データを生成するリスク評価指標データ生成ステップと、
前記リスク評価装置が、前記リスク評価指標データを出力する出力ステップと、
を含むリスク評価方法。
【請求項４】
コンピュータに、
匿名化される前の複数の個人情報レコードのデータセットである匿名化前データセットと、前記匿名化前データセットが特定の複数の匿名化方法のうちいずれか一つ又は複数の

匿名化方法により匿名化された匿名化データセットと、前記匿名化データセットのリスク評価条件を示すリスク評価条件データとを入力する入力機能と、
前記特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により入力データセットを匿名化する匿名化シミュレータとの間でデータを送受するシミュレータインタフェース機能と、
前記匿名化前データセットを前記シミュレータインタフェース機能により前記匿名化シミュレータに入力して匿名化させる制御機能と、
前記リスク評価条件データが示すデータ絞り込み対象の既知の個人情報データを使用して当該既知の個人情報データを有する個人情報レコードのみに前記匿名化データセットを絞り込み、当該絞り込み後の匿名化データセットと前記匿名化シミュレータにより匿名化されたデータセットであるシミュレーション結果データセットとの比較を行い、当該比較の結果を使用して、個人情報の漏洩のリスク評価指標を示すリスク評価指標データを生成するリスク評価指標データ生成機能と、
前記リスク評価指標データを出力する出力機能と、
を実現させるためのコンピュータプログラム。

【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、リスク評価装置、リスク評価方法、及びコンピュータプログラムに関する。
【背景技術】
【０００２】
従来、複数の個人情報が匿名化された匿名化データセットのリスク評価指標として、ｋ－匿名性、ｌ－多様性、差分プライバシーなどが知られている。
【０００３】
ｋ－匿名性については例えば非特許文献１に開示されている。ｌ－多様性については例えば非特許文献２に開示されている。差分プライバシーについては例えば非特許文献３に開示されている。
また、差分プライバシーとｋ－匿名性を組み合わせた場合の指標については例えば非特許文献４に開示されている。さらにサンプリングを加えた場合の指標については例えば非特許文献５に開示されている。
【先行技術文献】
【非特許文献】
【０００４】
【非特許文献１】P. Samarati and L. Sweeney, “Generalizing data to provide anonymity when disclosing information”, In Proc. of PODS 1998, p. 188, 1998.
【非特許文献２】A. Machanavajjhala, J. Gehrke, and D. Kifer, “l-diversity: Privacy beyond k-anonymity”, In Proc. of ICDE’06, pp. 24-35, 2006.
【非特許文献３】C. Dwork, “Differential privacy”, In Proc. of ICALP 2006, Vol. 4052, pp.1-12, 2006.
【非特許文献４】Kamalika Chaudhuri and Nina Mishra, “When random sampling preserves privacy”, In Annual International Cryptology Conference, pp.198-213. Springer, 2006.
【非特許文献５】Ninghui Li, Wahbeh Qardaji, and Dong Su, “On sampling, anonymization, and differential privacy or, k-anonymization meets differential privacy”, In Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, pp. 32-33. ACM, 2012.
【発明の概要】
【発明が解決しようとする課題】
【０００５】
しかし、上述した従来のリスク評価指標では、攻撃者のモデルが強力であって匿名化デ

ータセットが十分な安全性を担保するように強固に匿名化されている場合に、有用性が極端に低くなる場合があった。

【０００６】

本発明は、このような事情を考慮してなされたものであり、その目的は、攻撃者のモデルが弱いことを前提にすることによりリスク評価指標の有用性の向上を図ることにある。

【課題を解決するための手段】

【０００７】

（１）本発明の一態様は、匿名化される前の複数の個人情報レコードのデータセットである匿名化前データセットと、前記匿名化前データセットが特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により匿名化された匿名化データセットと、前記匿名化データセットのリスク評価条件を示すリスク評価条件データとを入力する入力部と、前記特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により入力データセットを匿名化する匿名化シミュレータとの間でデータを送受するシミュレータインタフェースと、前記匿名化前データセットを前記シミュレータインタフェースにより前記匿名化シミュレータに入力して匿名化させる制御部と、前記リスク評価条件データが示すデータ絞り込み対象の既知の個人情報データを使用して当該既知の個人情報データを有する個人情報レコードのみに前記匿名化データセットを絞り込み、当該絞り込み後の匿名化データセットと前記匿名化シミュレータにより匿名化されたデータセットであるシミュレーション結果データセットとの比較を行い、当該比較の結果を使用して、個人情報の漏洩のリスク評価指標を示すリスク評価指標データを生成するリスク評価指標データ生成部と、前記リスク評価指標データを出力する出力部と、を備えるリスク評価装置である。

【０００８】

（２）本発明の一態様は、上記（１）のリスク評価装置において、前記匿名化データセットは複数回の匿名化の各結果を含み、前記リスク評価条件データは、リスク評価指標の閾値を含み、前記リスク評価指標データ生成部は、あるデータ絞り込み対象の既知の個人情報データを使用した前記匿名化データセットのある回の匿名化の結果に対する前記リスク評価指標データのリスク評価指標が前記閾値以上である場合に、当該データ絞り込み対象の既知の個人情報データを使用して、前記匿名化データセットの次の回の匿名化の結果についての前記リスク評価指標データを生成する、リスク評価装置である。

【０００９】

（３）本発明の一態様は、リスク評価装置が、匿名化される前の複数の個人情報レコードのデータセットである匿名化前データセットと、前記匿名化前データセットが特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により匿名化された匿名化データセットと、前記匿名化データセットのリスク評価条件を示すリスク評価条件データとを入力する入力ステップと、前記リスク評価装置が、前記特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により入力データセットを匿名化する匿名化シミュレータとの間でデータを送受するシミュレータインタフェースステップと、前記リスク評価装置が、前記匿名化前データセットを前記シミュレータインタフェースステップにより前記匿名化シミュレータに入力して匿名化させる制御ステップと、前記リスク評価装置が、前記リスク評価条件データが示すデータ絞り込み対象の既知の個人情報データを使用して当該既知の個人情報データを有する個人情報レコードのみに前記匿名化データセットを絞り込み、当該絞り込み後の匿名化データセットと前記匿名化シミュレータにより匿名化されたデータセットであるシミュレーション結果データセットとの比較を行い、当該比較の結果を使用して、個人情報の漏洩のリスク評価指標を示すリスク評価指標データを生成するリスク評価指標データ生成ステップと、前記リスク評価装置が、前記リスク評価指標データを出力する出力ステップと、を含むリスク評価方法である。

【００１０】

（４）本発明の一態様は、コンピュータに、匿名化される前の複数の個人情報レコードのデータセットである匿名化前データセットと、前記匿名化前データセットが特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により匿名化された匿名化データセ

ットと、前記匿名化データセットのリスク評価条件を示すリスク評価条件データとを入力する入力機能と、前記特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により入力データセットを匿名化する匿名化シミュレータとの間でデータを送受するシミュレータインタフェース機能と、前記匿名化前データセットを前記シミュレータインタフェース機能により前記匿名化シミュレータに入力して匿名化させる制御機能と、前記リスク評価条件データが示すデータ絞り込み対象の既知の個人情報データを使用して当該既知の個人情報データを有する個人情報レコードのみに前記匿名化データセットを絞り込み、当該絞り込み後の匿名化データセットと前記匿名化シミュレータにより匿名化されたデータセットであるシミュレーション結果データセットとの比較を行い、当該比較の結果を使用して、個人情報の漏洩のリスク評価指標を示すリスク評価指標データを生成するリスク評価指標データ生成機能と、前記リスク評価指標データを出力する出力機能と、を実現させるためのコンピュータプログラムである。

【発明の効果】

【００１１】

本発明によれば、攻撃者のモデルが弱いことを前提にすることができ、リスク評価指標の有用性の向上を図ることができるという効果が得られる。

【図面の簡単な説明】

【００１２】

【図１】一実施形態に係るリスク評価システム１の構成例を示すブロック図である。

【図２】一実施形態に係る匿名化方法の一例の説明図である。

【図３】一実施形態に係るリスク評価方法の一例を示すフローチャートである。

【発明を実施するための形態】

【００１３】

以下、図面を参照し、本発明の実施形態について説明する。

本実施形態では、誰もが匿名化データセットにアクセスできるのではなく、特定の組織間のみで匿名化データセットが送受されることを前提にする。この前提により攻撃者の知識が限定されることになるので、誰もが匿名化データセットにアクセスできる場合に比して、本実施形態では攻撃者のモデルが弱くなる。以下、攻撃者の知識が限定されることを前提にして、本実施形態の説明を行う。攻撃者の知識が限定されることの一例として、ある個人の年齢は知っているが、他の個人情報、例えば住所や趣味は知らないことが挙げられる。

【００１４】

図１は、一実施形態に係るリスク評価システム１の構成例を示すブロック図である。図１において、リスク評価システム１は、リスク評価装置１０と、匿名化シミュレータ３０とを備える。リスク評価装置１０は、入力部１１と、制御部１２と、シミュレータインタフェース１３と、データ格納部１４と、リスク評価指標データ生成部１５と、出力部１６と、を備える。

【００１５】

入力部１１は、匿名化前データセット１０１と、匿名化データセット１０２と、リスク評価条件データ１０３とを入力する。匿名化前データセット１０１は、匿名化される前の複数の個人情報レコードのデータセットである。匿名化データセット１０２は、匿名化前データセット１０１が特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により匿名化されたデータセットである。本実施形態に係る匿名化方法として、例えば、ｋ－匿名化、個人情報レコードの一部抽出、一般化、ミクロアグリゲーション（microaggregation）、ノイズの付加、属性削除などが挙げられる。リスク評価条件データ１０３は、匿名化データセット１０２のリスク評価条件を示すデータである。

【００１６】

ここで、図２を参照して、本実施形態に係る匿名化方法の一例を説明する。図２は、本実施形態に係る匿名化方法の一例の説明図である。図２には、匿名化方法がｋ－匿名化（ｋ＝２、２－匿名化）である場合の例が示されている。図２において、匿名化前データセ

ット１０１は、６個の個人情報レコードｒ＿０＿１，ｒ＿０＿２，・・・，ｒ＿０＿ｍから構成される（ｍ＝６）。各個人情報レコードｒ＿０＿１～ｒ＿０＿６は、５個の個人情報データ「名前」、「年齢」、「性別」、「住所」及び「口座残高」を有する。個人情報データ「名前」は、単体で個人を特定できる個人識別符号である。個人情報レコードｒ＿０＿１は、名前がＡである人の個人情報レコードである。個人情報レコードｒ＿０＿２は、名前がＢである人の個人情報レコードである。個人情報レコードｒ＿０＿３は、名前がＣである人の個人情報レコードである。個人情報レコードｒ＿０＿４は、名前がＤである人の個人情報レコードである。個人情報レコードｒ＿０＿５は、名前がＥである人の個人情報レコードである。個人情報レコードｒ＿０＿６は、名前がＦである人の個人情報レコードである。個人情報データ「年齢」、「性別」及び「住所」は、準識別子である。準識別子は、単体では個人を特定できないが、複数の組み合わせで個人を特定できる可能性がある情報である。

【００１７】

図２の例では、２－匿名化のために匿名化前データセット１０１に対して、個人識別符号が削除される。さらに、年齢に対してミクロアグリゲーションが実行される。さらに、住所に対して一般化が実行される。さらに、個人情報レコードｒ＿０＿６が削除される。これにより、図２に例示される２－匿名化された匿名化データセット１０２が生成される。当該２－匿名化された匿名化データセット１０２は、５個の個人情報レコードｒ＿１＿１，ｒ＿１＿２，・・・，ｒ＿１＿５から構成される。

【００１８】

本実施形態では、匿名化データセット１０２が攻撃者に取得された場合に、個人情報が当該匿名化データセット１０２に基づいて漏洩するリスク、を評価するための指標（リスク評価指標）を求める。

【００１９】

制御部１２は、リスク評価装置１０の制御を行う。シミュレータインタフェース１３は、匿名化シミュレータ３０との間でデータを送受する。匿名化シミュレータ３０は、シミュレータインタフェース１３から入力された入力データセットを匿名化し、当該匿名化の結果のデータセットであるシミュレーション結果データセットをシミュレータインタフェース１３に出力する。匿名化シミュレータ３０は、特定の複数の匿名化方法のうちいずれか一つ又は複数の匿名化方法により入力データセットを匿名化する。

【００２０】

シミュレータインタフェース１３は、入出力するデータセット形式が定められている。このデータセット形式に合えば、匿名化シミュレータ３０が実行する匿名化方法は任意であって限定されない。したがって、リスク評価装置１０は、匿名化シミュレータ３０として任意の匿名化方法を使用するものを利用できる。

【００２１】

なお、シミュレータインタフェース１３は、通信回線を介して匿名化シミュレータ３０とデータを送受してもよい。シミュレータインタフェース１３は、例えばインターネット等の通信ネットワークを介して、匿名化シミュレータ３０と通信を行ってもよい。

【００２２】

データ格納部１４はデータを格納する。リスク評価指標データ生成部１５は、リスク評価指標データ２００を生成する。リスク評価指標データ２００は、匿名化データセット１０２が攻撃者に取得された場合に、当該匿名化データセット１０２に基づいて個人情報が漏洩するリスクの評価指標（リスク評価指標）を示すデータである。リスク評価指標として、例えば、個人識別確率の最大値Ｐ、個人識別確率がＰである個人の数Ｎ＿Ｐなどが挙げられる。出力部１６は、リスク評価指標データ２００を出力する。

【００２３】

リスク評価装置１０の機能は、リスク評価装置１０が備えるＣＰＵ（Central Processing Unit：中央演算処理装置）がコンピュータプログラムを実行することにより実現される。なお、リスク評価装置１０として、汎用のコンピュータ装置を使用して構成してもよ

く、又は、専用のハードウェア装置として構成してもよい。また、リスク評価装置１０と匿名化シミュレータ３０とは、各々単独の装置として構成されてもよく、又は、同じ一つの装置として構成されてもよい。例えば、一つのコンピュータ装置が、リスク評価装置１０の機能を実現させるためのコンピュータプログラムと、匿名化シミュレータ３０の機能を実現させるためのコンピュータプログラムとを実行してもよい。

また、入力部１１は、リスク評価装置１０にデータを入力するための入力デバイスを備える。入力デバイスは、例えば、利用者が実際に操作するデバイス（例えば、キーボード、テンキー、マウス等）であってもよく、又は、データが印刷された印刷物から当該データを読み取るデバイスであってもよく、又は、データが記録された記録媒体から当該データを読み出すデバイスであってもよく、又は、通信によりデータを受信するデバイスであってもよい。入力部は、入力デバイスにより、匿名化前データセット１０１と、匿名化データセット１０２と、リスク評価条件データ１０３とをリスク評価装置１０に入力する。

なお、入力部１１は、リスク評価装置１０の外部の装置からリスク評価条件データ１０３を取得してもよい。例えば、入力部１１は、リスク評価条件データ１０３の所在を示す所在情報（例えば、ＵＲＬ（Uniform Resource Locator））の指定を受け付け、当該指定された所在情報で示される場所からリスク評価条件データ１０３を通信により受信する。

また、入力部１１は、リスク評価条件データ１０３を決定してもよい。例えば、入力部１１は、リスク評価装置１０の外部の装置から複数のリスク評価条件の候補を入力し、当該複数のリスク評価条件の候補の中から、使用するリスク評価条件を選択して入力してもよい。

また、出力部１６は、リスク評価装置１０からデータを出力するための出力デバイスを備える。出力デバイスは、例えば、データを表示画面に表示するデバイスであってもよく、又は、紙等の印刷媒体にデータを印刷するデバイスであってもよく、又は、記録媒体にデータを書き込むデバイスであってもよく、又は、通信によりデータを送信するデバイスであってもよい。出力部１６は、出力デバイスにより、リスク評価指標データ２００を出力する。

【００２４】

次に図３を参照して本実施形態に係るリスク評価方法を説明する。図３は、本実施形態に係るリスク評価方法の一例を示すフローチャートである。

【００２５】

（ステップＳ１）リスク評価装置１０の入力部１１は、匿名化前データセット１０１と、匿名化データセット１０２と、リスク評価条件データ１０３とを入力する。データ格納部１４は、入力部１１により入力された匿名化前データセット１０１、匿名化データセット１０２及びリスク評価条件データ１０３を格納する。

【００２６】

本実施形態の一例として、匿名化前データセット１０１は、「Ｄ＿０＝｛ｒ＿０＿１，ｒ＿０＿２，・・・，ｒ＿０＿ｍ｝」である。匿名化前データセット１０１「Ｄ＿０」は、ｍ個の個人情報レコードｒ＿０＿１，ｒ＿０＿２，・・・，ｒ＿０＿ｍから構成される。本実施形態の一例として、匿名化データセット１０２は、ｎ回の匿名化の各結果Ｄ＿１，Ｄ＿２，・・・，Ｄ＿ｎから構成される。１回匿名化データセットＤ＿１は、１回の匿名化の結果である。２回匿名化データセットＤ＿２は、２回の匿名化の結果である。ｎ回匿名化データセットＤ＿ｎは、ｎ回の匿名化の結果である。

【００２７】

本実施形態の一例として、リスク評価条件データ１０３は、データ絞り込み対象の既知の個人情報データｒ＿ｉ（例えば年齢「３２」）と、リスク評価指標の閾値（以下、リスク閾値と称する）Ｒｉｓｋと、攻撃者モデル「Ａ＝｛Ａ＿１，Ａ＿２，・・・，Ａ＿ｑ｝」とを示すデータである。攻撃者モデルＡ＿ｑは、ｑ番目の攻撃者についての匿名化シミュレーション条件である。

【００２８】

（ステップＳ２）リスク評価装置１０の制御部１２は、匿名化データセット１０２をシミ

ュレータインタフェース１３により匿名化シミュレータ３０に入力して、当該匿名化データセット１０２の匿名化を実行させる。本実施形態の一例として、制御部１２は、匿名化データセット１０２に対してｎ回の匿名化を実行させる。これにより、匿名化シミュレータ３０は、匿名化データセット１０２のｎ回の匿名化の各シミュレーション結果ｓｉｍＤ＿１，ｓｉｍＤ＿２，・・・，ｓｉｍＤ＿ｎから構成されるシミュレーション結果データセットｓｉｍＤを、シミュレータインタフェース１３に出力する。匿名化シミュレーション結果ｓｉｍＤ＿１は、１回の匿名化の実行の結果である。匿名化シミュレーション結果ｓｉｍＤ＿２は、２回の匿名化の実行の結果である。匿名化シミュレーション結果ｓｉｍＤ＿ｎは、ｎ回の匿名化の実行の結果である。

【００２９】

また、本実施形態の一例として、制御部１２は、ｑ個の攻撃者モデルＡ＿１，Ａ＿２，・・・，Ａ＿ｑのそれぞれに対して、匿名化シミュレータ３０により匿名化前データセット１０１の匿名化を実行させる。これにより、ｑ個のシミュレーション結果データセットｓｉｍＤ（Ａ１），ｓｉｍＤ（Ａ２），・・・，ｓｉｍＤ（Ａｑ）が、匿名化シミュレータ３０により生成されて、シミュレータインタフェース１３に出力される。データ格納部１４は、匿名化シミュレータ３０からシミュレータインタフェース１３により入力されたｑ個のシミュレーション結果データセットｓｉｍＤ（Ａ１），ｓｉｍＤ（Ａ２），・・・，ｓｉｍＤ（Ａｑ）を格納する。

【００３０】

なお、各攻撃者モデルＡ＿１，Ａ＿２，・・・，Ａ＿ｑは、例えば、それぞれの攻撃者のノイズ値である。制御部１２は、例えば攻撃者モデルＡ＿１についての匿名化には、攻撃者モデルＡ＿１のノイズ値を匿名化前データセット１０１に加えた結果のノイズ付加匿名化前データセットを、匿名化シミュレータ３０の入力データセットに使用する。

【００３１】

（ステップＳ３）リスク評価装置１０のリスク評価指標データ生成部１５は、リスク評価条件データ１０３が示すデータ絞り込み対象の既知の個人情報データｒ＿ｉ（例えば年齢「３２」）を使用して、当該既知の個人情報データｒ＿ｉを有する個人情報レコードのみに、リスク評価指標データ生成対象の匿名化データセットを絞り込む。これにより、リスク評価指標データ生成対象の匿名化データセットの絞り込み結果が有する個人情報レコードは、個人情報データｒ＿ｉ（例えば年齢「３２」）を有する個人情報レコードのみになる。なお、匿名化データセット１０２のｎ個のデータセットＤ＿１，Ｄ＿２，・・・，Ｄ＿ｎのうち、最初のリスク評価指標データ生成対象の匿名化データセットは１回匿名化データセットＤ＿１である。

【００３２】

（ステップＳ４）リスク評価指標データ生成部１５は、ｑ個のシミュレーション結果データセットｓｉｍＤ（Ａ１），ｓｉｍＤ（Ａ２），・・・，ｓｉｍＤ（Ａｑ）と、リスク評価指標データ生成対象の匿名化データセットの絞り込み結果とを比較する。この比較の対象は、同じ回数の匿名化のデータセット同士である。例えば、リスク評価指標データ生成対象の匿名化データセットが１回匿名化データセットＤ＿１である場合には、シミュレーション結果データセットｓｉｍＤ（Ａ１），ｓｉｍＤ（Ａ２），・・・，ｓｉｍＤ（Ａｑ）のうち１回の匿名化の実行の結果である匿名化シミュレーション結果ｓｉｍＤ＿１（Ａ１），ｓｉｍＤ＿１（Ａ２），・・・，ｓｉｍＤ＿１（Ａｑ）のそれぞれと、１回匿名化データセットＤ＿１とを比較する。また、比較の方法として、比較対象のデータ間の一致か不一致かを判断してもよく、又は、比較対象のデータ間の類似度を算出してもよい。

【００３３】

（ステップＳ５）リスク評価指標データ生成部１５は、シミュレーション結果データセットｓｉｍＤ（Ａ１），ｓｉｍＤ（Ａ２），・・・，ｓｉｍＤ（Ａｑ）と、リスク評価指標データ生成対象の匿名化データセットの絞り込み結果との比較の結果を使用して、リスク評価指標データを生成する。本実施形態の一例として、リスク評価指標データ生成部１５は、リスク評価指標として、個人識別確率の最大値Ｐを算出する。

【００３４】
（ステップＳ６）リスク評価装置１０の出力部１６は、リスク評価指標データ２００を出力する。本実施形態の一例として、リスク評価指標データ２００は、リスク評価指標「個人識別確率の最大値Ｐ」を示すデータである。
【００３５】
（ステップＳ７）リスク評価装置１０の制御部１２は、リスク評価指標「個人識別確率の最大値Ｐ」がリスク閾値Ｒｉｓｋ以上であるか否かを判断する。この判断の結果、リスク閾値Ｒｉｓｋ以上である場合にはステップＳ８に進み、そうではない場合には図３の処理を終了する。
【００３６】
（ステップＳ８）制御部１２は、匿名化データセット１０２のうち次の回の匿名化の結果を、新しいリスク評価指標データ生成対象に設定する。この後、ステップＳ３に戻る。例えば、最初のリスク評価指標データ生成対象の匿名化データセット「１回匿名化データセットＤ＿１」について、ステップＳ７の結果がリスク閾値Ｒｉｓｋ以上である場合には、次の回の匿名化の結果「２回匿名化データセットＤ＿２」を、新しいリスク評価指標データ生成対象に設定する。この後、ステップＳ３に戻る。これにより、リスク評価指標データ生成対象の２回匿名化データセットＤ＿２に対して、ステップＳ３以降の処理が実行される。
【００３７】
上述した実施形態によれば、ある個人情報データのみを持つ攻撃者（つまり、保有する知識が限定された攻撃者）に対するリスク評価指標を算出することができる。これは、攻撃者のモデルが弱いことが前提にして、匿名化データセットが十分な安全性を担保するように強固に匿名化されている場合を除いたときのリスク評価指標を算出することができることである。これにより、リスク評価指標の有用性の向上を図る効果が得られる。
【００３８】
以上、本発明の実施形態について図面を参照して詳述してきたが、具体的な構成はこの実施形態に限られるものではなく、本発明の要旨を逸脱しない範囲の設計変更等も含まれる。
【００３９】
上述した実施形態では、リスク評価条件データ１０３として、データ絞り込み対象の既知の個人情報データ（例えば年齢）を使用したが、これに限定されない。例えば、リスク評価条件データ１０３として、匿名化前データセット１０１「Ｄ＿０＝｛ｒ＿０＿１，ｒ＿０＿２，・・・，ｒ＿０＿ｍ｝」のうちの一部の個人情報レコードと、当該個人情報レコードのみを持つ攻撃者の攻撃者モデルとを使用してもよい。この場合、当該個人情報レコードのみを持つ攻撃者の攻撃者モデルのみに対して、匿名化シミュレータ３０により匿名化前データセット１０１の匿名化を実行させる。
【００４０】
また、上述した各装置の機能を実現するためのコンピュータプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行するようにしてもよい。なお、ここでいう「コンピュータシステム」とは、ＯＳや周辺機器等のハードウェアを含むものであってもよい。
また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ＲＯＭ、フラッシュメモリ等の書き込み可能な不揮発性メモリ、ＤＶＤ（Digital Versatile Disc）等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。
【００４１】
さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムが送信された場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリ（例えばＤＲＡＭ（Dynamic Random Access Memory））のように、一定時間プログラムを保持しているものも含むものとする。

また、上記プログラムは、このプログラムを記憶装置等に格納したコンピュータシステムから、伝送媒体を介して、あるいは、伝送媒体中の伝送波により他のコンピュータシステムに伝送されてもよい。ここで、プログラムを伝送する「伝送媒体」は、インターネット等のネットワーク（通信網）や電話回線等の通信回線（通信線）のように情報を伝送する機能を有する媒体のことをいう。

また、上記プログラムは、前述した機能の一部を実現するためのものであっても良い。さらに、前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるもの、いわゆる差分ファイル（差分プログラム）であっても良い。

【符号の説明】

【００４２】

１…リスク評価システム、１０…リスク評価装置、１１…入力部、１２…制御部、１３…シミュレータインタフェース、１４…データ格納部、１５…リスク評価指標データ生成部、１６…出力部、３０…匿名化シミュレータ

【図１】

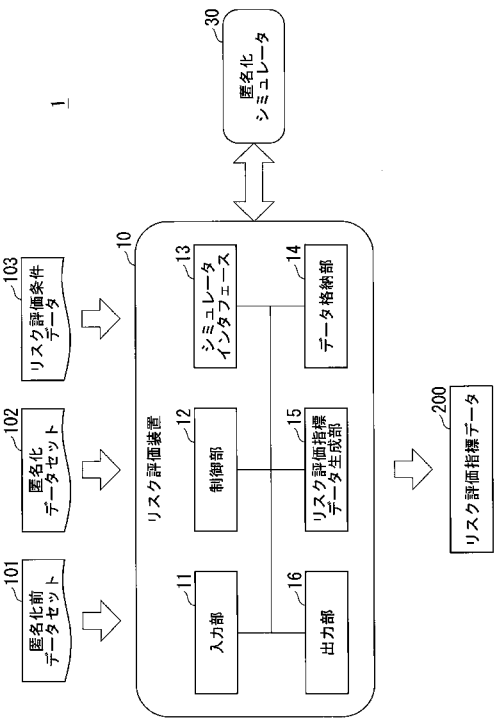


図1

【図２】

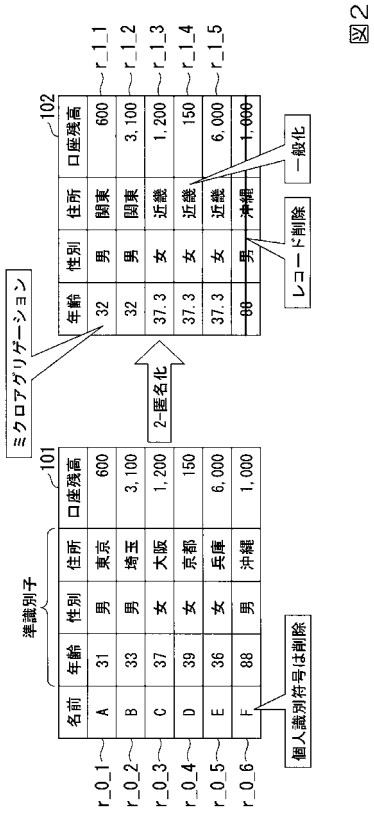


図2

【図３】

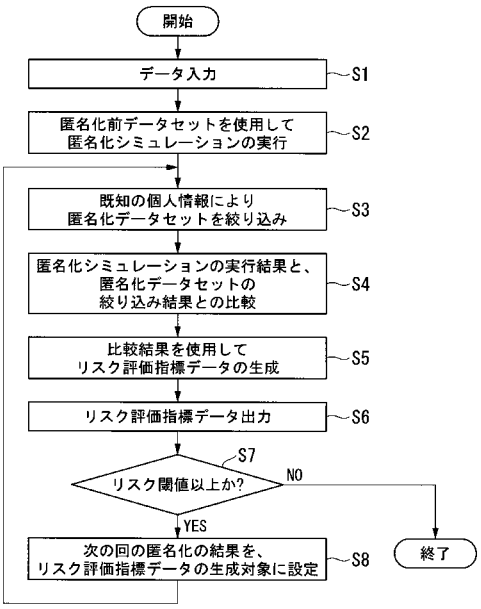

開始
データ入力
S1
匿名化前データセットを使用して
匿名化シミュレーションの実行
S2
既知の個人情報により
匿名化データセットを絞り込み
S3
匿名化シミュレーションの実行結果と、
匿名化データセットの
絞り込み結果との比較
S4
比較結果を使用して
リスク評価指標データの生成
S5
リスク評価指標データ出力
S6
S7
リスク閾値以上か?
NO
YES
終了
次の回の匿名化の結果を、
リスク評価指標データの生成対象に設定
S8

図３