(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2011-123712
(P2011-123712A)
(43)公開日　平成23年6月23日(2011.6.23)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 21/24	(2006.01)	G06F	12/14	560Z	2C032
G06F 17/30	(2006.01)	G06F	17/30	120A	5B017
G09B 29/10	(2006.01)	G09B	29/10	A	5B075

審査請求 未請求　請求項の数 6　ＯＬ　（全 7 頁）

(21)出願番号　特願2009-281371 (P2009-281371)
(22)出願日　平成21年12月11日 (2009.12.11)

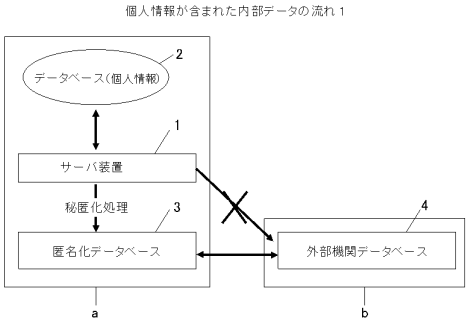
(71)出願人　309039255
田代　敦志
新潟県新潟市西区山田６４６－５
(72)発明者　田代　敦志
新潟県新潟市西区山田６４６－５

Ｆターム(参考)　2C032 HB08 HB22 HB31
5B017 AA04 BA08 BA09 CA06 CA16
5B075 KK54

(54)【発明の名称】個人情報の外部委託解析システム

(57)【要約】

【課題】　個人情報は、法に基づいた厳しい管理が求められることから、住所等の位置情報を含む顧客データを外部に提供して詳細な解析を行うことは事実上困難である。このような現状に対して本発明は、社内規則や自治体が持つ個人情報保護条例等に抵触することなく、外部機関で専門的なデータ解析が行えるようにする。
【解決手段】個人情報を保有する機関ａのサーバ装置１で対象となるデータベース２から氏名を削除し、システム上にロードされた秘匿化ソフトウエアにより解析に支障をきたさない程度まで位置情報を抽象化する情報処理を行った後に、秘匿化されたデータを元に匿名化データベース３を作成し、外部機関ｂと共通の管理用ＩＤを設定してデータを共有する。外部機関ｂは、外部機関データベース４を元に個人情報に付随する大量のデータに対してマーケティング等の空間解析を実施することを特徴とする。
【選択図】　図１



【特許請求の範囲】
【請求項１】
解析目的に応じて個人情報を秘匿化して外部提供するシステムであって、位置情報に関するアドレスマッチング処理について指示を受ける入力手段と、アドレスマッチングレベルに対応可能な位置座標データを記述した位置辞書と、処理レベルに応じて、位置辞書から位置座標データを特定する位置特定処理手段と、特定した位置座標データを出力する出力手段を備え、個人情報が含まれたデータベースを外部提供が可能な匿名化データベースに変換処理することを特徴としたシステム。
【請求項２】
アドレスマッチング処理において、どのレベルで秘匿化を行うか位置情報の抽象化処理を行うプロセスを加え、個人が特定されないよう位置座標データに変換し秘匿化するシステム。
【請求項３】
対象となる地域の人口密度や世帯密度等の特性に応じて、秘匿化のレベルを自動的に判断することを特徴としたアドレスマッチング処理システム。
【請求項４】
秘匿化処理後、個人情報に付随するデータを匿名化データベースの形で外部機関と共有し、マーケティング等の解析を外部機関において実施するシステム。
【請求項５】
抽象化処理を行うプロセスにおいて、住所等の位置情報に加えＧＰＳ等により得られた位置情報に対して格子等を用いた領域化を行い、格子の大きさや形を変化させることにより位置情報を抽象化して匿名化データベースを作成し、請求項４と同様にマーケティング等の解析を外部機関で実施するシステム。
【請求項６】
データ保有機関が位置情報のみ外部提供を認める場合、データを分離して位置情報の秘匿化を行う外部機関に提供し、抽象化処理した後にデータ保有機関において個人情報に付随するデータを再結合して匿名化データベースを作成し、請求項４と同様にマーケティング等の解析を外部機関で実施するシステム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、個人情報の秘匿化をコンピュータ上で行うことを目的としたソフトウエア、及びその秘匿化ソフトウエアを適用したアドレスマッチングによる情報処理を行うことにより、個人情報に付随するデータを外部機関へ提供し解析を可能にするシステムに関する。
【背景技術】
【０００２】
顧客データ等の個人情報は、個人情報保護法に基づいた厳しい管理を必要とすることから、これらの大量なデータを用いた解析（データマイニング）は、組織内で対応が可能な一部の企業を除くと手つかずの状態であり、多くのデータが解析されず眠ったままの状態にある。これまで個人情報に関連するデータを外部提供する場合、データを保有する機関において個人情報から個人データを分離して、残った個人識別情報を信託機関に提供し識別子により管理することで非個人化を行い、外部機関に提供する方法が知られていた。
【先行技術文献】
【特許文献】
【０００３】
【特許文献１】特開２０００-３２４０９４号公報
【発明の概要】
【発明が解決しようとする課題】
【０００４】
しかしながら、このような複雑なプロセスから成るシステム下で個人情報を外部提供する

には、時間とコストがかかることが予想される。また、組織内で高度な解析を行う場合においても専門スタッフの確保を含めコストがかかり、一部の企業を除くとマーケティング等に関するデータマイニングを組織の内部で実施する事は難しい。さらに、データを外部提供する際の秘匿化に関しても、個人情報保護を担保する定められたルールが存在しないのが現状である。

【０００５】

一方で、大量のデータからマーケティング等に関する空間解析を実施する場合には、個人が持つ住所等の位置情報を位置座標に変える事で、カーネル密度関数等を利用して個人に付随した情報を元にＧＩＳ（地理情報システム）により特定集団の分布地図を作成する方法が実用化されており、この場合はピンポイントで個人の位置情報を特定する必要がない。したがって、このような条件下で個人情報の秘匿化を行う場合、氏名を削除した後に個人が特定されないよう位置情報を位置座標に変換することが可能と考えられる。

【０００６】

変換によって得られた位置座標から個人が特定されないためには、個人が存在する地域の特性が大きく関わる。例えば、過疎地域における個人の位置情報は過密地域における位置情報と比較してより厳密な抽象化が求められる。住所を例に取ると　○町△丁目Ｘ番地□号において、○町△丁目Ｘ番地を代表する位置座標で十分秘匿化される場合と、○町△丁目を代表する位置座標でないと十分秘匿化されない場合など、地域の特性や解析目的により要求されるアドレスマッチングの処理レベルも種々のケースが想定される。

【０００７】

そこで、本発明は、大量の個人情報をコンピュータ上で外部提供できる状態まで秘匿化する際に、地域の特性を考慮した最適の位置情報の抽象化プロセスを加えることで、外部機関において解析に支障をきたさない程度まで個人情報が含まれるデータベースを秘匿化し、匿名化されたデータベースを共有することにより個人情報に付随するデータについて、外部機関でマーケティング等の解析を可能にする事を目的とする。

【課題を解決するための手段】

【０００８】

以上の課題を解決するため、国勢調査等で公表されている人口密度や世帯密度等を元にした丁目単位等の密度分布地図をベースに位置情報の抽象化レベルを複数定め、アドレスマッチング処理をレベル別に実施可能とする。これらの基礎データを元に、解析目的に応じて地域の特性に合ったアドレスマッチングの処理レベルを自動的に判断し、コンピュータ上で種々の抽象化手法を用いたデータ処理を行うソフトウエアをシステム上で実行し、個人情報が含まれたデータベースから匿名化データベースを作成し、これを外部機関と共有するシステムである。

【発明の効果】

【０００９】

本発明に基づくシステムを導入することで、企業のみならず自治体が保有する個人情報についても、法を遵守した上で外部委託機関において個人情報に付随するデータの空間解析が低いコストで可能となり、現状把握と解決すべき問題への介入等、多くの部署において保有データの有効活用が期待できる。

【図面の簡単な説明】

【００１０】

【図１】本システムにおける個人情報を含むデータの流れである。
【図２】位置情報から位置座標データへの抽象化処理方法である。
【図３】地域密度レベルに基づいた位置情報の抽象化処理方法の一例である。
【図４】位置情報を処理し、秘匿化された位置座標の一例である。
【図５】位置特定処理手段における抽象化処理方法１である。
【図６】位置特定処理手段における抽象化処理方法２である。
【図７】抽象化処理を外部機関で行う場合の個人情報を含むデータの流れである。

【発明を実施するための形態】

【００１１】
本発明の実施方法について図を用いて説明する。データ保有機関ａが管理する個人情報が含まれた内部データを、解析を行う外部機関ｂに提供するまでの流れを図１に示す。個人情報を保有する機関ａは、購入もしくはレンタルで本システムにおいて使用するソフトウエアをあらかじめ組織内部のサーバ装置１にロードし、個人情報の秘匿化処理についてプライベートＬＡＮの端末より指示が受けられる状態とする。
【００１２】
ＬＡＮに接続された端末からの指示で、データベース２より対象となるファイルのデータから氏名を削除し、代わりに管理用ＩＤが付与されたデータファイルを作成する。この操作の後に、あらかじめロードされた秘匿化ソフトウエアを用いて位置情報の抽象化処理をサーバ装置１で実施する。
【００１３】
秘匿化ソフトウエアによる位置データの処理過程を図２に示す。対象となるデータは、アドレスマッチング処理について指示を受ける入力手段２０より位置座標を特定する位置特定処理手段２１に送られ、位置辞書２２を用いて位置特定処理手段２１に設けた住所参照テーブル上で位置座標データを特定し、必要な抽象化処理を行った後に位置座標データとして出力手段２３より出力する。
【００１４】
個人の位置情報を地域密度に対応した位置座標に抽象化して変換するアドレスマッチングについて説明する。位置辞書２２内に住所に対応した位置座標データと国勢調査に基づいた人口密度や世帯密度等より得られた地域の密度レベルデータを格納し、位置特定処理手段２１に設けた、図３に示す住所欄と地域密度欄、座標欄からなる住所参照テーブルにおいて、個人の位置情報（住所等）と地域密度別の位置座標（緯度経度等）の対応づけを行う。
【００１５】
位置座標の抽象化について図３で一例を示す。あらかじめ町字、丁目、番地、号まで可能な限り詳細な位置座標を設定し、地域密度が標準的な地域においては、詳細なアドレスマッチングを行った後、緯度経度の小数点Ｎ（図３ではＮ＝５）桁以下（Ａ，Ｂ）を乱数変換（Ｃ，Ｄ）する。人口密度の高い地域においては個人が特定され難いことから、詳細なアドレスマッチング処理を行った後に、必要に応じて緯度経度の小数点Ｎ＋１桁以下をランダムに座標変換する。一方人口密度の低い地域においては、個人が特定され易い事を考慮して、同様のアドレスマッチング処理した後に、緯度経度の小数点以下Ｎ－１桁以下をランダムに座標変換する。このようにコンピュータ上で国勢調査に基づいた人口密度や世帯密度等より得られた地域の密度レベルデータを元にアドレスマッチング精度を自動的に判断し、個人が特定されず解析に支障をきたさない必要十分なレベルで位置情報（住所等）を抽象化処理する。
【００１６】
実際の住所に相当する位置座標　（ｘ座標、ｙ座標）を　（x^1、y^1）とすると、抽象化処理をした後の座標　（x^2、y^2）との距離のずれＳは、以下の計算式より近似値が求められる。R_0：平均曲率半径　m_0：座標系の原点における縮尺係数
【００１７】
【数１】

$$S=\frac{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}}{\frac{s}{S}}$$

【００１８】
【数２】

$$\frac{s}{S}=m_0\left\{1+\frac{1}{6R_0^2m_0^2}\left(y_1^2+y_1y_2+y_2^2\right)\right\}$$

【００１９】

１０進法表示された緯度経度において、仮に小数点以下５桁目を乱数表示すると、図４に示すように実際の座標と最大１００ｍ程度のずれが生じる。個人が特定されないよう住所を座標変換して大量のデータから空間解析を行う場合、解析目的にもよるがこの値を超えないずれは、都道府県単位の解析を行う場合において結果に大きな影響を与えないと考えられ、過疎地を除けば、位置情報の抽象化レベルとして適当な情報処理と考えられる。さらに、人口密集地域を含む市区町村単位の解析を行う場合であれば、小数点以下６桁目の乱数表示により実際とのずれを最小にして精度の高い空間解析も可能であり、本発明におけるアドレスマッチングレベルの設定は、解析目的と人口密度、世帯密度等の地域特性に応じて変更することが可能である。

【００２０】

位置特定処理手段２１において、地域の密度レベルデータを元に位置情報（住所等）を抽象化処理する他の方法を、図５に示す。Ｃに示すように、実際の座標Ａ１を一定の範囲ｒにあるＡの位置にランダムに座標変換する以外に、Ｄに示すように実際の座標Ｂ１を番地や号レベルのエリアを代表とする座標Ｂ（重心座標等）に変換処理する方法等がある。

【００２１】

位置情報が住所以外の郵便番号等の文字情報で提供される場合など、必要に応じて文字情報を代表するポリゴン（領域）の重心座標等（図５のＢ）を基準にして、図５のＣに示した一定の範囲にランダムに座標変換する等の方法を組み合わせて変換処理を実施することも可能である。

【００２２】

最初のアドレスマッチングで、番地、号までマッチングできない場合は、最も近いレベルのデータを暫定的に出力し、空間解析への影響の有無を判断できるよう非マッチングデータとし、必要に応じて再処理できるようグループ化する

【００２３】

秘匿化処理が終わったデータは、アドレスマッチング処理のエラー等について検証した後に、共有ＩＤにより匿名化データベース３で管理し、インターネットもしくはフラッシュメモリー等の記録媒体により専門的な解析を行う外部機関ｂのデータベース４に提供を行う。

【００２４】

外部機関ｂにおいて空間解析を実施し、はずれ値や非マッチングデータを含め不適切な秘匿化処理が疑われるデータについては、共有ＩＤにより照会し必要に応じて個人情報を保有する機関ａのシステム上でデータの再処理を行い匿名化データベース３の修正を行う。外部機関ｂは、完成した匿名化データベース３より作成されたデータベース４を用いて、依頼を受けたマーケティング等の解析を実施する。

【００２５】

さらに、アドレスマッチング処理を必要としないＧＰＳによる位置情報サービス等を利用する場合や、より詳細に住所等の位置情報を位置特定処理手段２１おいて抽象化処理する際には、図６に示すように密集地域Ｅにおいては、実際の位置情報Ａ１～Ａ３を格子等を用いて小さく領域化して格子を代表する重心座標Ａ等の異なった座標に変換する。さらに、変換後の重心座標を基準にして、図５のＣに示すように一定の範囲にランダムに座標変換する方法を組み合わせることも可能である。

【００２６】

格子の大きさや形は対象地域の密度レベルに応じて変化させ、過疎地域Ｆの位置情報Ｂ１～Ｂ３はより大きな領域の重心座標Ｂ等に変換することで厳密に個人の位置情報の抽象化処理が可能である。これらの方法で作成された匿名化データベース３を元に、外部機関ｂにおいてマーケティング等の解析を同様に実施する。

【００２７】

以上は、位置情報（住所等）の外部漏出を防ぐために、個人情報の秘匿化処理をデータ保有機関ａの内部で行うシステムであるが、データ保有機関ａが位置情報のみ外部提供を認める場合は、図７に示すように秘匿化処理を行う外部機関５にＩＤ管理された位置情報の

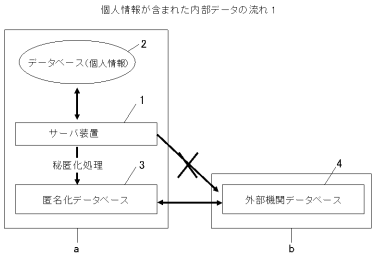
みを提供し、前述したデータの抽象化処理を実施する。この後データ保有機関ａにおいて、個人情報に付随するデータを再結合して匿名化データベース３を作成し、これを元に解析を行う外部機関ｂにおいてマーケティング等のデータマイニングを実施する。

【符号の説明】

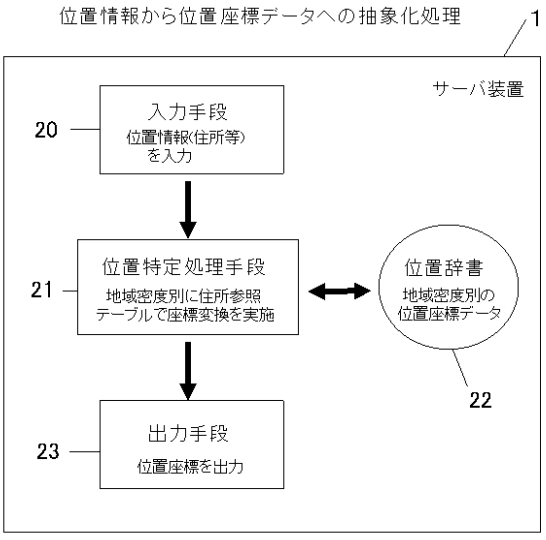
【００２８】

１　サーバ装置
２　個人情報が含まれたデータベース
３　匿名化データベース
４　外部機関データベース
５　秘匿化を行う外部機関
ａ　データ保有機関
ｂ　解析を行う外部機関
２０　入力手段
２１　位置特定処理手段
２２　位置辞書
２３　出力手段
ｒ　一定の範囲
Ａ１～３　実際の座標
Ａ　変換後の座標
Ｂ１～３　実際の座標
Ｂ　変換後の座標
Ｃ　一定の範囲にランダムに配置する抽象化
Ｄ　エリアを代表する座標に配置する抽象化
Ｅ　密集地域
Ｆ　過疎地域

【図１】



【図２】



【図３】

位置情報(住所等)と位置座標(緯度経度)の対応づけ

住所					地域密度	実際の座標	
市区町村	町字	丁目	番地	号	(n段階)	緯度	経度
中央区	○△	X	Y	Z	n	37.××××A	139.××××B

地域密度別に抽象化した座標	
緯度	経度
37.××××C	139.××××D

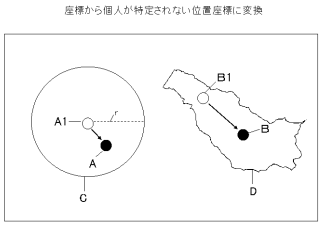
【図４】

秘匿化した位置座標における地域密度別のずれ

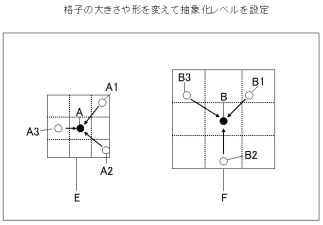
住所					地域密度	実際の座標	
市区町村	町字	丁目	番地	号	（3段階）	緯度	経度
中央区	○×	X	Y	Z	3(密集)	37.901562	139.070292
北区	△×	X	Y	Z	2(標準)	37.901562	139.070292
南区	××	X	Y	Z	1(過疎)	37.901562	139.070292

住所					抽象化座標		位置のずれ
市区町村	町字	丁目	番地	号	緯度	経度	m
中央区	○×	X	Y	Z	37.901569	139.070299	9.8
北区	△×	X	Y	Z	37.901522	139.070211	83.8
南区	××	X	Y	Z	37.901134	139.070626	558.6

【図５】



【図６】



【図７】