

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6156071号
(P6156071)

(45) 発行日 平成29年7月5日(2017.7.5)

(24) 登録日 平成29年6月16日(2017.6.16)

(51) Int.Cl.

F I

G 0 6 F 21/62 (2013.01)

G 0 6 F 21/62 3 5 4

請求項の数 9 (全 27 頁)

(21) 出願番号 特願2013-230613 (P2013-230613)
 (22) 出願日 平成25年11月6日(2013.11.6)
 (65) 公開番号 特開2015-90617 (P2015-90617A)
 (43) 公開日 平成27年5月11日(2015.5.11)
 審査請求日 平成28年7月5日(2016.7.5)

(73) 特許権者 000005223
 富士通株式会社
 神奈川県川崎市中原区上小田中4丁目1番
 1号
 (74) 代理人 100103528
 弁理士 原田 一男
 (72) 発明者 山岡 裕司
 神奈川県川崎市中原区上小田中4丁目1番
 1号 富士通株式会社内
 審査官 宮司 卓佳

最終頁に続く

(54) 【発明の名称】 匿名化データ生成方法、装置及びプログラム

(57) 【特許請求の範囲】

【請求項 1】

データ格納部に格納されており且つ各々機密属性値と数値属性値とを含む複数のデータブロックにおける各数値属性で張られる空間における複数種類のメッシュの各々について、前記複数のデータブロックのうち、当該メッシュの1つのメッシュ要素に含まれ且つグループ化未了の第1のデータブロックを含み且つ機密属性値の度数分布が所定の条件を満たすデータブロックのグループを抽出し、

前記グループに属するデータブロックの数値属性値を、前記グループについての数値属性値で置換する

処理を含み、コンピュータにより実行される匿名化データ生成方法。

10

【請求項 2】

前記グループに属するデータブロックの機密属性値を削除する

処理をさらに含む請求項1記載の匿名化データ生成方法。

【請求項 3】

前記所定の条件が、機密属性値の種類数の下限値を含み、

前記抽出する処理は、

前記グループ化未了の第1のデータブロックを含み且つ前記1つのメッシュ要素に含まれるデータブロックの集合における機密属性値の種類が、前記機密属性値の種類数の下限値以上となっているか判断する

処理を含む請求項1又は2記載の匿名化データ生成方法。

20

【請求項 4】

前記抽出する処理は、

前記グループ化未了の第 1 のデータブロックを含み且つ前記 1 つのメッシュ要素に含まれるデータブロックの第 2 の集合についての機密属性値の度数分布が、前記所定の条件を満たすか判断し、

前記データブロックの第 2 の集合についての機密属性値の度数分布が前記所定の条件を満たさない場合には、前記所定の条件を満たすように前記データブロックの第 2 の集合から第 2 のデータブロックを除外することでデータブロックのグループを生成する

処理を含む請求項 1 乃至 3 のいずれか 1 つ記載の匿名化データ生成方法。

【請求項 5】

前記抽出する処理は、

前記データブロックのグループから、前記機密属性値の度数分布が前記所定の条件を満たす上で必須となるデータブロック以外の第 3 のデータブロックを抽出する処理

を含み、

データブロックの他のグループを抽出する処理において、前記第 3 のデータブロックが他の種類のメッシュにおけるメッシュ要素において抽出される

請求項 1 乃至 4 のいずれか 1 つ記載の匿名化データ生成方法。

【請求項 6】

メッシュ要素の境界を共有しないように、前記複数種類のメッシュを生成する

処理をさらに含む請求項 1 乃至 5 のいずれか 1 つ記載の匿名化データ生成方法。

【請求項 7】

これまでに生成されたメッシュの境界のうち最も近い境界との距離が最大となり、且つ前記これまでに生成されたメッシュの格子点との距離が最大となる基準点を選択し、

選択された前記基準点及び前記メッシュ要素のサイズに基づきメッシュを生成する

処理をさらに含む請求項 1 乃至 5 のいずれか 1 つ記載の匿名化データ生成方法。

【請求項 8】

データ格納部に格納されており且つ各々機密属性値と数値属性値とを含む複数のデータブロックにおける各数値属性で張られる空間における複数種類のメッシュの各々について、前記複数のデータブロックのうち、当該メッシュの 1 つのメッシュ要素に含まれ且つグループ化未了の第 1 のデータブロックを含み且つ機密属性値の度数分布が所定の条件を満たすデータブロックのグループを抽出し、

前記グループに属するデータブロックの数値属性値を、前記グループについての数値属性値で置換する

処理を、コンピュータに実行させるための匿名化データ生成プログラム。

【請求項 9】

データ格納部に格納されており且つ各々機密属性値と数値属性値とを含む複数のデータブロックにおける各数値属性で張られる空間における複数種類のメッシュの各々について、前記複数のデータブロックのうち、当該メッシュの 1 つのメッシュ要素に含まれ且つグループ化未了の第 1 のデータブロックを含み且つ機密属性値の度数分布が所定の条件を満たすデータブロックのグループを抽出するグループ化処理部と、

前記グループに属するデータブロックの数値属性値を、前記グループについての数値属性値で置換する匿名化処理部と、

を有する情報処理装置。

【発明の詳細な説明】**【技術分野】****【0001】**

本技術は、情報の匿名化技術に関する。

【背景技術】**【0002】**

複数の情報提供者から収集した、数値属性値を含むレコード群を、各レコードの情報提

10

20

30

40

50

供者識別子（以下、単にIDと略す）を秘密にして、他者に開示又は提供したい場合がある。このとき、IDを削除して開示又は提供しても、特徴ある数値属性値を有するレコードについては他者が情報提供者を推定できてしまう場合がある。

【0003】

例えば、個人の位置データの収集者が、情報提供者が分からない形で、分析者に位置データを提供することを考える。ここで、収集者としては位置データについてのサービス提供者、分析者としてはクラウドサービス提供者又はデータ二次利用者（例えば人口密度調査会社など）などが考えられる。

【0004】

ここで、収集者が収集した位置データが図1に示すものであるとする。図1の例では、各レコードには、行番号と、IDと、X（緯度）と、Y（経度）とが含まれる。ここでは、各レコードは、A、B及びCの3人のいずれかの位置データを表しており、全部で7レコードある。すなわち、同じIDのレコードが複数回出現する場合がある。なお、IDは、個人のユーザIDである場合もあれば、測定機器のIDである場合もある。また、所属する組織のIDである場合もある。

【0005】

図1に示されるデータを地図上にプロットすると、例えば図2に示すようになる。分析者は、図1及び図2のようなデータが得られれば、分析に役立てることができる。例えば、A宅及びB宅付近に人が集まっていることが分かる。

【0006】

しかしながら、例えば、収集者が情報提供者との間で、匿名化しない限り他者にデータを提供しないという契約を結んでいる状況が考えられる。情報提供者は、特定の時期にどこにいたかを収集者以外に知られたくないなどの理由で、匿名化を希望する場合がある。

【0007】

一方、分析者はID等の情報提供者の情報を利用しない場合もある。位置データの提供者が誰かを知らなくても人口密度調査のような分析はできるためである。

【0008】

このような場合、収集者は図1のデータに対して匿名化を行って、情報提供者の推定を困難にすれば良い。

【0009】

収集者による単純な匿名化方式として、IDを削除する方式がある。図1からIDを削除したデータを分析者が見ても、どのレコードが誰のデータなのかそのままでは分からない。しかし、位置データから情報提供者を推定可能なレコードがあるという問題がある。

【0010】

図1からIDを削除したデータを図2のように地図上にプロットすると、例えば第1レコードの位置データ(X, Y) = (6, 2)はA宅内であることが分かってしまう。すなわち、IDが削除されたデータしか見ることのできない分析者であっても、第1レコードの情報提供者がAであることが推定できてしまい、十分に匿名化されているとは言い難い。同様に、第7レコード以外は匿名化が不十分である。

【0011】

従来技術として、事前に定められた、重なりのない複数の数値範囲をグループとして把握し、各グループ内のレコード群をそれらの統計値に変換する方法がある。

【0012】

この従来技術では、緯度及び経度に基づいて地域をメッシングし、各メッシュ要素内のレコード群についての統計値を計算し、それを開示又は提供する。

【0013】

統計値としては、例えば「メッシュ要素M1には3レコード」というように、メッシュ要素毎のレコード数が用いられる。もしくは、各レコードのIDを削除し、位置をメッシュ要素の中心点に変換しても良い。

【0014】

10

20

30

40

50

例えば、図1の各レコードを一辺「5」のメッシュ要素によりグループ分けし、変換する場合を考える。その場合、例えば $(X, Y) = ([5, 10], [0, 5])$ などが1つのメッシュ要素、すなわちグループとなる。このメッシュ要素を仮にM10と名付けると、図1では第1レコードのみがM10に分類される。よって、「メッシュM10には1レコードあった」ことが開示される、あるいは第1レコードが $(X, Y) = (7.5, 2.5)$ (M10の中心点)に変換され開示されることになる。

【0015】

この従来技術では、メッシュサイズが十分大きければ匿名性に問題は生じないが、メッシュサイズを小さくすると匿名性が脅かされるという問題がある。例えば、メッシュ要素M10が、もしA宅の敷地内に包含された場合(例えば、A宅の敷地が $(X, Y) = ([2, 10], [0, 6])$ の場合など)、メッシュ要素M10に分類されたレコードの情報提供者はAだと推定できてしまう。メッシュサイズを小さくするほど、特定のIDしか存在し得ないような地域にメッシュ要素が包含される可能性が高くなる。

10

【0016】

一方、メッシュサイズを大きくするほど、位置の一般化度合いが大きくなり、分析者による分析の精度に大きな悪影響を与えるという問題がある。例えば、統計調査では一辺約1kmのメッシュ要素が使われることがあるが、その結果だけを使う限り1km単位より詳細な地域に関する分析結果を出すことは一般的にはできない。

【0017】

このように、この従来技術は、匿名性を担保するためにメッシュサイズを十分に大きくしなければならず、分析の精度に大きな悪影響を与えるという問題がある。

20

【0018】

また、グループを生成する別の従来技術として、事前に決めた値d及びkに対し、大きさd未満の範囲内に、k個以上のレコードが含まれるように、且つ別の範囲と重ならないように範囲の位置を調整し、その範囲に基づきグループ化する技術がある。

【0019】

この従来技術は対象データとして互いに異なるIDを有するレコード群を前提にしており、その場合は適切な匿名性が担保されるが、図1のように同じIDのレコードが複数存在し得るデータに対しては十分な匿名性を担保できないという問題がある。

【0020】

30

例えば、この従来技術の一部を適用し、図1の各レコードを一辺「5」未満の矩形($d = (5, 5)$)で、3個以上($k = 3$)のレコードが含まれるようグループ分けする場合を考える。この場合、例えばレコード{1, 2, 3}を含む矩形R43: $(X, Y) = ([2, 6], [2, 4])$ と、レコード{4, 5, 6}を含む矩形R49: $(X, Y) = ([2, 6], [8, 10])$ の2つのグループができる。しかし、上でも述べた例と同じように、特定のIDしか存在し得ないような地域に矩形が包含される可能性がある。例えば矩形R43がA宅の敷地内に包含された場合には、矩形R43に分類されたレコード{1, 2, 3}の情報提供者がAだと推定できてしまう。

【0021】

一般的に、図1のように同じIDのレコードが複数存在するレコード群をも取り扱うことができる手法の方が適用範囲が広くて良い。例えば、組織が情報提供者の場合は特に、複数の測定機器のデータに同じID(すなわち組織ID)が記録される場合もある。また、IDが同じ複数のレコードの存在を許すことで多くのレコードを一度に分析できるようになり、分析精度の向上が望める。しかしながら、この従来技術は、特殊な対象データでしか匿名性を担保できず、適用できる場面が少ないという問題がある。

40

【0022】

さらに、グループ化する別の従来技術として、IDのような機密属性値の種類が各グループ内で1種類以上ある(すなわち1-多様性を満たす)ようにする技術がある。この従来技術はグループを事前に決めた大きさ未満の範囲内にすることが難しいという問題がある。事前に決めた大きさ未満の範囲内にできないと、分析の精度に大きな悪影響を与える

50

という問題がある。

【先行技術文献】

【非特許文献】

【0023】

【非特許文献1】O. Abul, F. Bonchi, and M. Nanni. Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases. In Proceedings of the 24th International Conference on Data Engineering, ICDE 2008, pp.376-385 (2008).

【非特許文献2】A. Machanavajjhala, J. Gehrke, D. Kifer, M. Venkitasubramaniam. l-Diversity: Privacy Beyond k-Anonymity. ACM Transactions on Knowledge Discovery from Data, Vol. 1, Issue 1, Article No. 3, 2007.

10

【発明の概要】

【発明が解決しようとする課題】

【0024】

従って、本技術の目的は、一側面によれば、適切な分析精度を出すことができるように数値属性値を含むデータの匿名化を可能にするための技術を提供することである。

【課題を解決するための手段】

【0025】

本発明に係る匿名化データ生成方法は、(A)データ格納部に格納されており且つ各々機密属性値と数値属性値とを含む複数のデータブロックにおける各数値属性で張られる空間における複数種類のメッシュの各々について、複数のデータブロックのうち、当該メッシュの1つのメッシュ要素に含まれ且つグループ化未了の第1のデータブロックを含み且つ機密属性値の度数分布が所定の条件を満たすデータブロックのグループを抽出し、(B)上記グループに属するデータブロックの数値属性値を、上記グループについての数値属性値で置換する処理を含む。

20

【発明の効果】

【0026】

一側面によれば、適切な分析精度を出すことができるように数値属性値を含むデータを匿名化できるようになる。

【図面の簡単な説明】

【0027】

30

【図1】図1は、データの一例を示す図である。

【図2】図2は、データと他のデータとの重ね合わせの一例を示す図である。

【図3】図3は、本実施の形態に係る情報処理装置の構成例を示す図である。

【図4】図4は、第1データ格納部に格納されるデータの一例を示す図である。

【図5】図5は、本実施の形態における処理フローを示す図である。

【図6】図6は、メッシュ生成処理の処理フローを示す図である。

【図7】図7は、2種類のメッシュが生成された状態を示す図である。

【図8】図8は、メッシュの基準点のばらつきを表す図である。

【図9】図9は、複数種類のメッシュのメッシュ要素の重なりを示す図である。

【図10】図10は、グループ生成処理の処理フローを示す図である。

40

【図11】図11は、1番目のメッシュについての分類結果の例を示す図である。

【図12】図12は、レコード群Rs抽出処理の処理フローを示す図である。

【図13】図13は、除外処理の処理フローを示す図である。

【図14】図14は、除外処理において生成される度数分布表の一例を示す図である。

【図15】図15は、除外処理において生成される度数分布表の一例を示す図である。

【図16】図16は、除外処理において生成される度数分布表の一例を示す図である。

【図17】図17は、除外処理において生成される度数分布表の一例を示す図である。

【図18】図18は、除外処理において生成される度数分布表の一例を示す図である。

【図19】図19は、除外処理において生成される度数分布表の一例を示す図である。

【図20】図20は、レコード群Rs抽出処理の処理フローを示す図である。

50

【図 2 1】図 2 1 は、譲渡可能レコードの抽出処理の処理フローを示す図である。

【図 2 2】図 2 2 は、抽出処理において生成される度数分布表の一例を示す図である。

【図 2 3】図 2 3 は、抽出処理において生成される度数分布表の一例を示す図である。

【図 2 4】図 2 4 は、抽出処理において生成される度数分布表の一例を示す図である。

【図 2 5】図 2 5 は、抽出処理において生成される度数分布表の一例を示す図である。

【図 2 6】図 2 6 は、抽出処理において生成される度数分布表の一例を示す図である。

【図 2 7】図 2 7 は、メッシュとグループの対応付けテーブルの一例を示す図である。

【図 2 8】図 2 8 は、譲れるレコード及び譲れないレコードの区別を行うためのグループデータテーブルの一例を示す図である。

【図 2 9】図 2 9 は、行番号とグループ ID の対応付けテーブルの一例を示す図である。

【図 3 0】図 3 0 は、メッシュとグループの対応付けテーブルの一例を示す図である。

【図 3 1】図 3 1 は、譲れるレコード及び譲れないレコードの区別を行うためのグループデータテーブルの一例を示す図である。

【図 3 2】図 3 2 は、行番号とグループ ID の対応付けテーブルの一例を示す図である。

【図 3 3】図 3 3 は、2 番目のメッシュについての分類結果の例を示す図である。

【図 3 4】図 3 4 は、メッシュとグループの対応付けテーブルの一例を示す図である。

【図 3 5】図 3 5 は、譲れるレコード及び譲れないレコードの区別を行うためのグループデータテーブルの一例を示す図である。

【図 3 6】図 3 6 は、行番号とグループ ID の対応付けテーブルの一例を示す図である。

【図 3 7】図 3 7 は、3 番目のメッシュについての分類結果の例を示す図である。

【図 3 8】図 3 8 は、グループ化処理の結果の一例を示す図である。

【図 3 9】図 3 9 は、処理結果の一例を示す図である。

【図 4 0】図 4 0 は、コンピュータの機能ブロック図である。

【発明を実施するための形態】

【0028】

図 3 に、本発明の実施の形態に係る情報処理装置 100 の機能ブロック図を示す。情報処理装置 100 は、第 1 データ格納部 110 と、設定データ格納部 120 と、グループ化処理部 130 と、第 2 データ格納部 140 と、匿名化処理部 150 と、第 3 データ格納部 160 と、出力部 170 と、メッシュ生成部 180 と、第 4 データ格納部 190 とを有する。

【0029】

第 1 データ格納部 110 は、例えば図 4 に示すような匿名化前のデータを格納している。図 4 の例では、各レコード（データブロックとも呼ぶ）は、ID と、X（緯度）と、Y（経度）と、速さとを含む。行番号は、以下の説明のために付されている。

【0030】

また、設定データ格納部 120 は、生成すべきメッシュの種類数 x と、メッシュ要素のサイズ d と、度数分布についての条件（度数分布パターンとも呼ぶ）と、第 1 データ格納部 110 に格納されているデータのうちの機密属性（例えば ID 属性。機微属性とも呼ぶ。）及び数値属性（例えば緯度 X 及び経度 Y を含む位置データ）の指定とが格納される。度数分布パターンは、最小種類数 l と、減衰率 a とを含む。最小種類数 l は 2 以上の整数であり、減衰率 a は 1 以下の正の実数である。例えば、1 種類の ID について度数の多い順に n 番目の度数が $n - 1$ 番目の度数の a 倍以上であるという度数分布パターンが条件として設定される。

【0031】

グループ化処理部 130 は、設定データ格納部 120 及び第 4 データ格納部 190 に格納されているデータを用いて、第 1 データ格納部 110 に格納されているレコード群をグループ化する処理を行い、処理結果を第 2 データ格納部 140 に格納する。匿名化処理部 150 は、グループ化の結果に基づき、各グループに属するレコードの数値属性値を変換する処理を行い、処理結果を第 3 データ格納部 160 に格納する。出力部 170 は、第 3 データ格納部 160 に格納されているデータを、他のコンピュータ、表示装置や印刷装置

など出力する。

【0032】

メッシュ生成部180は、設定データ格納部120に格納されているメッシュの種類数 x 及びメッシュ要素のサイズ d に従って、メッシュのデータを生成し、第4データ格納部190に格納する。

【0033】

メッシュ要素のサイズ d は、各数値属性で張られる空間においてメッシュ要素の辺の長さを表す正の数値列である。たとえば、数値属性が (X, Y) の2次元なら、 $d = (6, 6)$ などと設定される。

【0034】

メッシュの種類数 x は、自然数である。例えば、異なるメッシュを6種類使うなら、 $x = 6$ となる。本実施の形態では、メッシュは n 次元矩形メッシュ(n は数値属性の数)とする。なお、メッシュは複数のメッシュ要素を含み、メッシュ要素間の境界(壁とも呼ぶ)は、2次元であれば辺であり、3次元であれば平面であり、4次元以上であれば超平面となる。また、メッシュは一般的には、三角形メッシュや六角形メッシュなど、数値属性の値がどのメッシュ要素に該当するかすぐに計算できるような規則に従っていれば、矩形に限らない。

【0035】

次に、図5乃至図39を用いて、情報処理装置100の処理内容について説明する。

【0036】

メッシュ生成部180は、メッシュ生成処理を実行し、生成されたメッシュのデータを、第4データ格納部190に格納する(図5:ステップS1)。メッシュ生成処理については、図6乃至図9を用いて説明する。

【0037】

まず、メッシュ生成部180は、メッシュの空のリスト M を生成する(図6:ステップS11)。また、メッシュ生成部180は、原点 $(0, 0)$ をリスト M に追加する(ステップS13)。まず、メッシュ要素の1辺の長さが「1」であるメッシュを生成するものとする。従って、以降の処理では、基準の点についての各次元の値域が $[0, 1)$ に入るメッシュが、リスト M に追加される。

【0038】

そして、メッシュ生成部180は、リスト M に含まれるメッシュの種類数 $|M|$ は、指定された種類数 x に達したか否かを判断する(ステップS15)。ステップS13の直後であれば、 $|M| = 1$ であり、例えば $x = 6$ であるとする、この条件を満たしていない。

【0039】

ステップS15の条件を満たしていない場合には、メッシュ生成部180は、リスト M に属する各メッシュの境界のうち最も近い境界との距離が最大となる点 P' を選択する(ステップS19)。

【0040】

本実施の形態では、リスト M に含まれるメッシュの種類数 $|M|$ に応じた点(基準点とも呼ぶ)を発生させる。すなわち、 $c = \text{ceil}(\log_2(|M| + 1))$ (ceil は天井関数)と定義して、点 $(2^{-c}, 2^{-c})$ を基点として、 $[0, 1)$ 内を幅 2^{-c+1} で変動させた点を生成する。最初は、 $|M| = 1$ なので、 $c = 1$ となる。従って、 $(1/2, 1/2)$ のみが生成される。

【0041】

さらに、メッシュ生成部180は、選択された点 P' のうち、リスト M に属する各メッシュの格子点のうち近い格子点との距離が最大となる点のメッシュをリスト M に追加する(ステップS21)。最初は、基準点が $(1/2, 1/2)$ となるメッシュがリスト M に追加される。そして処理はステップS15に戻る。

【0042】

10

20

30

40

50

この状態では、図 7 に示すようなメッシュが生成されたことになる。なお、まだメッシュ要素の辺の長さは「1」のままである。

【0043】

ステップ S 15 において、リスト M に含まれるメッシュの種類数 $|M|$ が、指定された種類数 x に達した場合には、メッシュ生成部 180 は、メッシュ要素のサイズ d から、各メッシュを設定し、第 4 データ格納部 190 に格納する（ステップ S 23）。そして処理は呼び出し元の処理に戻る。

【0044】

上で述べた例では、 $|M| = 2$ で次にステップ S 19 に移行すると、 $c = 2$ となるので、 $P = \{(1/4, 1/4), (1/4, 3/4), (3/4, 1/4), (3/4, 3/4)\}$ が生成される。この段階では、どの点も、リスト M に含まれる各メッシュの格子点からの距離は同じなので、ステップ S 21 では、いずれかの点、例えば $p = (1/4, 1/4)$ を選び、リスト M に追加する。すなわち、 $M = \{(0, 0), (1/2, 1/2), (1/4, 1/4)\}$ となる。

【0045】

さらに、 $|M| = 3$ で次にステップ S 19 に移行すると、 $c = 2$ のままとなる。従って、前回と同じ P が生成されるが、 $(3/4, 1/4)$ 及び $(1/4, 3/4)$ は、 $(1/4, 1/4)$ のメッシュの辺上に載るので、それより距離が遠い $(3/4, 3/4)$ が採用される。ステップ S 21 でも、 $p = (3/4, 3/4)$ が選択されて、リスト M に追加される。すなわち、 $M = \{(0, 0), (1/2, 1/2), (1/4, 1/4), (3/4, 3/4)\}$ となる。

【0046】

さらに、 $|M| = 4$ で次にステップ S 19 に移行すると、 $c = 3$ となるので、 $P = \{(1/8, 1/8), (1/8, 3/8), \dots, (3/8, 1/8), (3/8, 3/8), \dots, (7/8, 7/8)\}$ が生成される。 P に含まれる点のうち、リスト M の各メッシュの格子点のうち、近い点と距離が最大となる点の 1 つである $p(1/8, 5/8)$ をリスト M に追加する。なお、近い格子点との距離が最大とは、たとえば、最も近い格子点とのユークリッド距離が最大で、それが同じ場合次に近い格子点とのユークリッド距離が最大で、といったように以降同様に各格子点との距離を比較して決める。例を挙げると、点 $(1/8, 5/8)$ の最も近い点の 1 つである $(1/4, 1/4)$ との距離は $10^{1/2}/8$ だが、 $(1/8, 3/8)$ の最も近い点 $(1/4, 1/4)$ との距離は $2^{1/2}/8$ であり、距離が短い後者の点は採用されない。そうすると、 $M = \{(0, 0), (1/2, 1/2), (1/4, 1/4), (3/4, 3/4), (1/8, 5/8)\}$ となる。

【0047】

さらに、 $|M| = 5$ で次にステップ S 19 に移行すると、 $c = 3$ のままとなる。そして同様に処理すると、 $(5/8, 1/8)$ が選択されて、 $M = \{(0, 0), (1/2, 1/2), (1/4, 1/4), (3/4, 3/4), (1/8, 5/8), (5/8, 1/8)\}$ となる。ここで、 $|M| = x = 6$ となり、 $d = (6, 6)$ から、以下のようなメッシュの基準点が生成される。 $M = \{(0, 0), (3, 3), (3/2, 3/2), (9/2, 9/2), (3/4, 15/4), (15/4, 3/4)\}$ 。そして、処理は呼び出し元の処理に戻る。

【0048】

図 6 の処理フローでは、メッシュの境界を共有しないように、ずらしてメッシュを生成するようにしている。このようにずらしてメッシュを生成することで、以下で述べるグループ化処理で、固定のメッシュ要素単位でグループ化を行うが、多様なグループ化が可能になる。

【0049】

なお、上で述べた基準点は、図 8 に示すようにばらつくようになる。2 次元の矩形格子を採用しているので、メッシュがそれぞれ少しずつずらして生成されていることがこの図

10

20

30

40

50

からも分かる。なお、図中の数字は生成順番を示している。また、図 9 に、実際のメッシュの重なり状態を示している。ここでは 6 番目に生成されるメッシュのメッシュ要素までを 1 つずつ示しているが、メッシュ要素の境界は共有されていない。

【 0 0 5 0 】

なお、図 6 の処理フローは、リスト M に含まれるメッシュから最も遠い位置のメッシュをリスト M に追加している処理とも言える。最も遠い位置の定義としては、以下のようなものとなる。

【 0 0 5 1 】

すなわち、リスト M と、2 つのメッシュ m_1 及び m_2 を入力として、M から m_2 の方が m_1 よりも遠い場合には - 1 を、 m_1 の方が m_2 より遠い場合には 1 を、 m_1 と m_2 が同じ遠さの場合には 0 を返す関数 $cmp(M, m_1, m_2) : \{-1, 0, 1\}$ を用意して、最も遠いメッシュの集合 F は、 $F = \{m \mid m', cmp(M, m, m') = 0\}$ で定義される。

【 0 0 5 2 】

例えば、図 6 の処理フローにおけるステップ S 1 9 及び S 2 1 の処理を $cmp(M, m_1, m_2)$ に当てはめると、以下ようになる。

1 . m_1 及び m_2 それぞれの任意の境界 (辺) と、リスト M の各メッシュの境界の中で最も近い境界との距離を d_1 及び d_2 としたとき、 $d_1 < d_2$ なら - 1 を返し、 $d_1 > d_2$ なら 1 を返す。

2 . m_1 及び m_2 それぞれの基準点と、リスト M に含まれる各メッシュの格子点の中で 1 番目に近い頂点との距離を d_1 及び d_2 としたとき、 $d_1 < d_2$ なら - 1 を返し、 $d_1 > d_2$ なら 1 を返す。

3 . そうでない場合、2 と同様に、2 番目に近い格子点との距離の比較を行って条件に合うか判断し、...、 $|M|$ 番目に近い格子点との距離の比較を行って条件に合うか判断する、ということを繰り返す、条件にあった時点で - 1 又は 1 を返す。

4 . そうでなければ 0 を返す。

【 0 0 5 3 】

このような関数 cmp によって F を求め、その中から基準点が最も原点に近いメッシュを選択すればよい。

【 0 0 5 4 】

図 5 の処理フローの説明に戻って、グループ化処理部 1 3 0 は、第 1 データ格納部 1 1 0 に格納されているレコードに対して、第 4 データ格納部 1 9 0 及び設定データ格納部 1 2 0 に格納されているデータに従って、グループ生成処理を実行する (ステップ S 3) 。グループ生成処理については、図 1 0 乃至図 3 8 を用いて説明する。

【 0 0 5 5 】

まず、グループ化処理部 1 3 0 は、第 4 データ格納部 1 9 0 におけるリスト M から、未処理のメッシュ m を 1 つ特定する (図 1 0 : ステップ S 3 1) 。そして、グループ化処理部 1 3 0 は、第 1 データ格納部 1 1 0 に格納されているレコード群を、特定されたメッシュ m のメッシュ要素で分類する (ステップ S 3 3) 。

【 0 0 5 6 】

具体的には、各レコードに対してメッシュ要素識別子 (ID) を算出し、ID と行番号との対応関係表を生成する。例えば、メッシュ要素 ID は、 $\text{floor}((\text{数値属性} - m) / d)$ として算出する。なお、 floor は床関数である。例えば、図 4 における行番号「8」のレコードは $(X, Y) = (6, 5)$ であり、 $m = (0, 0)$ 、 $d = (6, 6)$ であるから、 $(\text{floor}((6 - 0) / 6), \text{floor}((5 - 0) / 6)) = (1, 0)$ がメッシュ要素 ID となる。

【 0 0 5 7 】

以上のような処理を行うと、図 1 1 に示すような分類がなされる。図 1 1 の例では、3 つのメッシュ要素に、各レコードが分類されている。

【 0 0 5 8 】

次に、グループ化処理部 130 は、レコードを含むメッシュ要素のうち、未処理のメッシュ要素 s を 1 つ特定する (ステップ S 35)。例えば、メッシュ要素 ID (0, 0) のメッシュ要素を特定する。

【0059】

そして、グループ化処理部 130 は、特定されたメッシュ要素 s 内において譲れないレコードとして第 2 データ格納部 140 に登録されているレコード以外のレコード R_a を抽出する (ステップ S 37)。最初の場合には、譲れないレコードとして登録されているレコードは存在しない。但し、以下で述べる処理では譲れないレコードとして登録されているレコードが存在する場合には、当該レコードを除外してレコード R_a を特定する。最初の例では、 $R_a = \{1, 2, 3, 4, 5, 6, 7\}$ となる。

10

【0060】

その後、グループ化処理部 130 は、レコード群 R_s 抽出処理を実行する (ステップ S 39)。このレコード群 R_s 抽出処理については、図 12 乃至図 19 を用いて説明する。なお、この処理が終了すると、処理は端子 A を介して図 20 の処理に移行する。

【0061】

まず、グループ化処理部 130 は、レコード R_a についての度数分布を生成し、レコード R_a には 1 種類以上の ID 属性値が含まれるか判断する (ステップ S 133)。レコード R_a に 1 種類以上の ID 属性値が含まれない場合には、グループ化処理部 130 は、レコード群 R_s を空に設定する (ステップ S 138)。そして、処理は呼び出し元に戻る。1 種類以上の ID 属性値が含まれていない場合には、グループ化を行うことができないので、呼び出し元の処理に戻る。

20

【0062】

一方、レコード R_a に 1 種類以上の ID 属性値が含まれる場合には、グループ化処理部 130 は、レコード R_a についての度数分布が、度数分布パターンにおける条件 a (= 減衰率) を満たしているか判断する (ステップ S 134)。1 = 3 及び $a = 0.5$ が設定されているものとする。度数が大きい順に並べて、 n 番目の度数が、 $n - 1$ 番目の度数の $a = 0.5$ 以上であることが条件となっている。例えば {B : 2, A : 1, C : 1} という度数分布が得られた場合には、この条件を満たしている。

【0063】

減衰率 a を用いるのは、度数分布の偏りが大きすぎるグループを作らない、すなわち安全性を高めるためである。減衰率 a は、1 種類以上あっても偏りが大きすぎる度数分布となるグループを許さないようにするための条件である。例えば、度数分布が {A : 100, C : 1} (A が 100 個、C が 1 個の意味) となるレコードを含む範囲を考える。この範囲は 2 種類の ID を含むので、この範囲が特定の 1 つの ID しか存在し得ないような地域に包含されることはない。しかし、度数の 99% 以上が A であり、この範囲のほとんどが A しか存在し得ないような地域である可能性がある。そうであった場合、この範囲を開示すると、この範囲のほとんどのレコードが A から提供されたことを推定しやすく問題となるからである。

30

【0064】

レコード R_a についての度数分布が、度数分布パターンにおける条件を満たしている場合には、グループ化処理部 130 は、レコード群 R_s に、レコード R_a を全て設定する (ステップ S 136)。そして、処理は呼出元の処理に戻る。

40

【0065】

一方、レコード R_a についての度数分布が、度数分布パターンにおける条件を満たしていない場合には、グループ化処理部 130 は、除外処理を実行する (ステップ S 135)。条件 1 を満たしていれば、条件 a を満たすようにレコードを除外することが可能である。従って、除外処理を実行する。除外処理については、図 13 乃至図 19 を用いて説明する。

【0066】

そして、グループ化処理部 130 は、予め定められているルールに従って、除外処理で

50

決定された数の除外すべきレコードを特定して、レコード R a から除外して、残余をレコード群 R s に設定する（ステップ S 1 3 7）。所定のルールは、例えばランダムといったような単純な方法でよい。

【 0 0 6 7 】

ここで、除外処理について説明する。まず、グループ化処理部 1 3 0 は、レコード R a について度数分布表 F を生成し、度数の昇順に整列させる（図 1 3：ステップ S 1 4 1）。上で述べた例では、除外処理は行われないので、ここでは図 1 4 に示すような度数分布表 F が生成されたものとする。また、 $l = 4$ 且つ $a = 0.5$ であるものとする。

【 0 0 6 8 】

そして、グループ化処理部 1 3 0 は、変数 p を初期化し（ステップ S 1 4 3）、変数 i を 0 に初期化する（ステップ S 1 4 5）。その後、グループ化処理部 1 3 0 は、i が度数分布表 F の行数 $|F|$ より小さいか判断する（ステップ S 1 4 7）。i が度数分布表 F の行数 $|F|$ より小さい場合には、グループ化処理部 1 3 0 は、 $i + l - 1$ が $|F|$ より小さいか判断する（ステップ S 1 4 9）。 $i + l - 1$ が $|F|$ より小さい場合には、グループ化処理部 1 3 0 は、変数 p に対して $F[i]$ を代入する（ステップ S 1 5 1）。 $F[i]$ は、F の $i + 1$ 行目の度数である。i = 0 であれば、変数 p には、F の 1 行目の度数「1」が代入される。

【 0 0 6 9 】

一方、 $i + l - 1$ が $|F|$ 以上であれば、グループ化処理部 1 3 0 は、変数 p に、 $\min(F[i], \text{floor}(p/a))$ を代入する（ステップ S 1 5 3）。 $\min(A, B)$ は、A と B のうち小さい方を出力する関数である。

【 0 0 7 0 】

ステップ S 1 5 1 又は S 1 5 3 の後に、グループ化処理部 1 3 0 は、 $F[i]$ に、 $F[i] - p$ を代入する（ステップ S 1 5 5）。i = 0 の時に、S 1 4 9 を実行すると、度数分布表 F は、図 1 5 のようになる。

【 0 0 7 1 】

その後、グループ化処理部 1 3 0 は、変数 i を 1 インクリメントし（ステップ S 1 5 7）、処理はステップ S 1 4 7 に戻る。

【 0 0 7 2 】

2 回目のステップ S 1 4 7 では、 $|F| = 5$ 、 $i = 1$ であるから、 $i < |F|$ となる。また、 $l = 4$ であるので、 $i + l - 1 < |F|$ となる。従って、ステップ S 1 5 1 で $p = 3$ であり、 $F[1] = 3 - 3 = 0$ となる。そうすると、度数分布表 F は、図 1 6 のようになる。その後 $i = 2$ となる。

【 0 0 7 3 】

3 回目のステップ S 1 4 7 では、 $|F| = 5$ 、 $i = 2$ であるから、 $i < |F|$ となる。また、 $l = 4$ であるので、 $i + l - 1 < |F|$ とはならず、ステップ S 1 5 3 に移行して、 $a = 0.5$ 及び $p = 3$ であるから、 $\min(F[i] = 4, \text{floor}(p/a) = 6) = 4$ となる。従って、 $F[2] = 4 - 4 = 0$ となる。そうすると、度数分布表 F は、図 1 7 のようになる。その後 $i = 3$ となる。

【 0 0 7 4 】

4 回目のステップ S 1 4 7 では、 $|F| = 5$ 、 $i = 3$ であるから、 $i < |F|$ となる。また、 $l = 4$ であるので、 $i + l - 1 < |F|$ とはならず、ステップ S 1 5 3 に移行して、 $a = 0.5$ 及び $p = 4$ であるから、 $\min(F[i] = 9, \text{floor}(p/a) = 8) = 8$ となる。従って、 $F[3] = 9 - 8 = 1$ となる。そうすると、度数分布表 F は、図 1 8 のようになる。その後 $i = 4$ となる。

【 0 0 7 5 】

5 回目のステップ S 1 4 7 では、 $|F| = 5$ 、 $i = 4$ であるから、 $i < |F|$ となる。また、 $l = 4$ であるので、 $i + l - 1 < |F|$ とはならず、ステップ S 1 5 3 に移行して、 $a = 0.5$ 及び $p = 8$ であるから、 $\min(F[i] = 10, \text{floor}(p/a) = 16) = 10$ となる。従って、 $F[4] = 10 - 10 = 0$ となる。そうすると、度数分布

10

20

30

40

50

表 F は、図 19 のようになる。その後 $i = 5$ となる。

【0076】

6 回目のステップ S 147 では、 $|F| = 5$ 、 $i = 5$ であるから、 $i < |F|$ が成り立たなくなる。そうすると、処理は呼出元の処理に戻る。すなわち、この時点における度数分布表 F (図 19) が、除外すべきレコードを示している。ここでは、ID が「E」のレコードを 1 つ除外することになる。除外するレコードについては、上で述べたように、ランダムに選択すればよい。

【0077】

図 10 の端子 A を介して図 20 の処理の説明に移行する。なお、上で述べていた例のように、 $Ra = \{1, 2, 3, 4, 5, 6, 7\}$ の場合、1 及び a の条件を満たすので、 $Ra = Rs$ となる。

【0078】

グループ化処理部 130 は、レコード群 Rs が空であるか判断する (ステップ S 41)。レコード群 Rs が空であると、グループ化は行われないので、処理はステップ S 55 に移行する。

【0079】

一方、レコード群 Rs が空でない場合には、グループ化処理部 130 は、レコード群 Rs に未グルーピングのレコードが存在するか否かを判断する (ステップ S 43)。未グルーピングのレコードが存在しない場合には、新たなグループの生成をわざわざ行うことはないので、処理はステップ S 55 に移行する。

【0080】

一方、レコード群 Rs に未グルーピングのレコードが存在する場合には、グループ化処理部 130 は、レコード群 Rs でグループを生成する (ステップ S 45)。 $Rs = \{1, 2, 3, 4, 5, 6, 7\}$ が 1 つのグループとなる。

【0081】

そして、グループ化処理部 130 は、生成されたグループに対して譲渡可能レコードの抽出処理を実行する (ステップ S 47)。すなわち、生成されたグループのうち、度数分布パターンを満たす上で必須のレコード以外のレコードを抽出する処理を実行する。より具体的には、図 21 乃至図 26 を用いて説明する。

【0082】

まず、グループ化処理部 130 は、レコード Rs について度数分布表 F を生成し、度数の昇順に整列させる (図 21: ステップ S 171)。処理を分かり易くするために、図 22 に示すような度数分布表 F が生成されたものとする。

【0083】

そして、グループ化処理部 130 は、変数 ci に $|F| - 1$ を設定し、変数 min に、 $ceil(F[ci] * a)$ を設定する (ステップ S 173)。 $ceil(x)$ は天井関数であり、実数 x に対して x 以上の最小の整数を出力する関数である。 $F[i]$ は度数分布表 F の $i + 1$ 行目の度数を表し、 $|F|$ は度数分布表 F の行数を表す。 $ci = 5 - 4 = 1$ であり、 $min = ceil(2 * 0.5) = 1$ となる。

【0084】

また、グループ化処理部 130 は、変数 i を 0 に初期化し、変数 max を 0 に初期化する (ステップ S 175)。

【0085】

その後、グループ化処理部 130 は、 $i < |F|$ であるか判断する (ステップ S 177)。 $i < |F|$ であれば、グループ化処理部 130 は、変数 c を初期化する (ステップ S 179)。その後、グループ化処理部 130 は、 $i < ci$ であるか判断する (ステップ S 181)。 $i = 0$ であれば、 $ci = 1$ であるからこの条件は満たされている。

【0086】

$i < ci$ であれば、グループ化処理部 130 は、 c に 0 を設定する (ステップ S 183)。そうすると、グループ化処理部 130 は、 $F[i]$ に、 $F[i] - c$ を設定する (ス

10

20

30

40

50

テップS 1 8 5)。F [i] = 1であり、c = 0であるから、F [i] = 1となる。その後、グループ化処理部 1 3 0は、iを1インクリメントし(ステップS 1 8 7)、処理はステップS 1 7 7に戻る。

【0087】

i = 1になると、ステップS 1 8 1では、 $i < c$ は成り立たなくなるので、グループ化処理部 1 3 0は、 $i + 1 = |F|$ であるか判断する(ステップS 1 8 9)。i = 1であれば、 $i + 1 = 2$ であるから、この条件を満たさない。ステップS 1 8 9の条件を満たさない場合には、グループ化処理部 1 3 0は、cに、 $\text{ceil}(F[i+1] * a)$ を代入する(ステップS 1 9 1)。 $c = \text{ceil}(F[2] * 0.5) = 2$ となる。そして、グループ化処理部 1 3 0は、 $\max < c$ であるか判断する(ステップS 1 9 3)。 $\max = 0$ であるからこの条件を満たす。そうすると、グループ化処理部 1 3 0は、 \max にcを代入する(ステップS 1 9 7)。すなわち、 $\max = c = 2$ となる。その後、処理はステップS 1 8 5に移行する。従って、2回目のステップS 1 8 5では、 $F[1] = 2 - 2 = 0$ となる。従って、図23に示すような度数分布表Fになる。

10

【0088】

i = 2になると、ステップS 1 8 1では、 $i < c$ は成り立たなくなるので、ステップS 1 8 9に移行する。但し、 $i + 1 < |F|$ であるから、ステップS 1 9 1に処理は移行し、 $c = \text{ceil}(F[3] * 0.5) = 2$ となる。 $\max = 2$ であるから、 $\max < c$ の条件を満たしていない。そうすると、グループ化処理部 1 3 0は、cにminを代入する(ステップS 1 9 5)。min = 1であるから、 $c = 1$ となる。そして処理はステップS 1 8 5に移行して、3回目のステップS 1 8 5では、 $F[2] = 3 - 1 = 2$ となる。従って、図24に示すような度数分布表Fになる。

20

【0089】

i = 3になると、ステップS 1 8 1では、 $i < c$ は成り立たないので、ステップS 1 8 9に移行する。但し、 $i + 1 < |F|$ であるから、ステップS 1 9 1に処理は移行し、 $c = \text{ceil}(F[4] * 0.5) = 3$ となる。 $\max = 2$ で $c = 3$ であるから、 $\max < c$ の条件を満たしている。従って、 $\max = c = 3$ となる。そして、4回目のステップS 1 8 5では、 $F[3] = 4 - 3 = 1$ となる。従って、図25に示すような度数分布表Fとなる。

【0090】

30

i = 4になると、ステップS 1 8 1では、 $i < c$ は成り立たないので、ステップS 1 8 9に移行する。 $i + 1 = |F|$ を満たすので、ステップS 1 9 5に処理は移行し、 $c = \min = 1$ となる。そして処理はステップS 1 8 5に移行して、5回目のステップS 1 8 5では、 $F[4] = F[4] - c = 5 - 1 = 4$ となる。従って、図26に示すような度数分布表Fが得られる。

【0091】

その後i = 5になると、ステップS 1 7 7では $i < |F|$ の条件を満たさなくなるので、処理は呼出元の処理に戻る。従って図26に示すように、IDがAの1レコード、IDがCの2レコード、IDがDの1レコード、及びIDがEの4レコードが、譲渡可能なレコードとして特定されたことになる。

40

【0092】

図20の処理の説明に戻って、グループ化処理部 1 3 0は、ステップS 4 7の処理結果に基づき、予め定められているルールに従って、譲渡可能なレコードを具体的に特定する(ステップS 4 9)。所定のルールについては、例えばランダムであっても良いし、行番号が大きい順番であっても良い。

【0093】

上で述べた例では、 $Rs = \{1, 2, 3, 4, 5, 6, 7\}$ のうち、譲れるレコードは $\{2, 4, 6, 7\}$ であり、譲れないレコードは $\{1, 3, 5\}$ となる。

【0094】

ここまでの処理によって、第2データ格納部 1 4 0には、図27乃至図29のデータが

50

格納される。図 27 は、グループ化が行われたメッシュ及びメッシュ要素 ID とグループ ID (G_ID) との対応付けを行うテーブルを表している。グループ ID についてはユニークになるようにシリアルに番号付けする例を示している。また、図 28 では、グループ ID に対応付けて、当該グループに属するレコードのうち譲れないレコードの行番号集合と、譲れるレコードの行番号集合とが登録されるようになっている。さらに、図 29 は、行番号とグループ ID との対応付けを示している。

【0095】

そして、グループ化処理部 130 は、他のグループから譲ってもらったレコードが存在しているか否かを判断する (ステップ S51)。他のグループにグループ化されていたが、譲れるレコードに分類されていて今回生成されたグループに属するレコードが存在する場合には、所属グループを変更することになる。従って、他のグループから譲ってもらったレコードが存在している場合には、グループ化処理部 130 は、第 2 データ格納部 140 において、譲渡元グループから、譲ってもらったレコードの行番号を削除する (ステップ S53)。一方、他のグループから譲ってもらったレコードが存在していない場合には、処理はステップ S55 に移行する。

10

【0096】

そして、グループ化処理部 130 は、未処理のメッシュ要素が存在するか判断する (ステップ S55)。未処理のメッシュ要素が存在する場合には、処理は端子 B を介して図 10 のステップ S35 に戻る。一方、未処理のメッシュ要素が存在しない場合には、グループ化処理部 130 は、未処理のメッシュが存在するか否かを判断する (ステップ S57)。未処理のメッシュが存在する場合には、処理は端子 C を介して図 10 のステップ S31 に戻る。一方、未処理のメッシュが存在しない場合には、処理は呼び出し元の処理に戻る。

20

【0097】

なお、上で述べた例では、メッシュ (0, 0) についてメッシュ要素 ID (1, 0) に対する処理に移行する。但し、 $R_a = \{8\}$ となるので、1 及び a の条件を満たさないので、 R_s は空に設定されて、グループ化は行われない。

【0098】

また、メッシュ (0, 0) についてメッシュ要素 ID (1, 1) に対する処理に移行する。この場合、 $R_a = \{9, 10, 11, 12\}$ となって、1 及び a の条件を満たすので、 $R_a = R_s = \{9, 10, 11, 12\}$ が設定され、グループもそのまま生成される。なお、譲渡可能レコードの抽出処理 (図 21) を実行すると、 $\{12\}$ が譲れるレコードに特定されるので、譲れないレコードは $\{9, 10, 11\}$ となる。

30

【0099】

ここまで処理すると、第 2 データ格納部 140 には、図 30 乃至図 32 に示すようなデータが格納される。図 30 は、図 27 の後の状態を表しており、グループ ID 「2」が、メッシュ (0, 0) 及びメッシュ要素 ID (1, 1) に付与されている。また、図 31 は、図 28 の後の状態を表しており、グループ ID 「2」について、上で述べたように、譲れないレコードの行番号集合 $\{9, 10, 11\}$ 、及び譲れるレコードの行番号集合 $\{12\}$ が示されている。図 32 は、図 29 の後の状態を表しており、グループ ID 「2」についての行のデータが追加されている。

40

【0100】

次に、メッシュ (3, 3) の処理に移行して、レコードの分類を行うと、図 33 のような分類結果が得られる。2 つのメッシュ要素に分類されたことが分かる。

【0101】

まず、メッシュ要素 ID (-1, 0) のメッシュ要素に分類されたレコードを処理する。レコード $\{1, 2\}$ は、1 及び a の条件を満たさないので、メッシュ要素 ID (0, 0) の処理に移行する。

【0102】

そうすると、 $\{3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ のうち、これまでに

50

生成されたグループにおいて譲れないレコードは{ 1 , 3 , 5 , 9 , 10 , 11 }であるから、 $Ra = \{ 4 , 6 , 7 , 8 , 12 \}$ となる。この Ra は 1 及び a の条件を満たすので、 $Ra = Rs$ となる。

【 0 1 0 3 】

Rs の中にはレコード { 8 } が含まれているので、 Rs でグループを生成することになる。さらに、譲渡可能レコードの抽出処理 (図 2 1) を実行すると、譲れるレコードは { 7 , 8 } と特定され、譲れないレコードは { 4 , 6 , 12 } となる。なお、これらは図 3 1 から分かるように既にグループ化されているので、各グループから譲ってもらうことになる。

【 0 1 0 4 】

そうすると、第 2 データ格納部 1 4 0 に格納されるデータは、図 3 4 乃至図 3 6 に示すようなデータに更新される。図 3 4 は、図 3 0 の後の状態を表しており、グループ ID に対応付けてメッシュ (3 , 3) のメッシュ要素 ID (0 , 0) が追加されている。さらに、図 3 5 は、図 3 1 の後の状態を表しており、グループ ID 「 3 」に対応付けて、譲れないレコード { 4 , 6 , 12 }、及び譲れるレコード { 7 , 8 } が登録されている。レコード { 8 } 以外は、他のグループから譲ってもらったので、グループ「 1 」及び「 2 」から、それらのグループの行番号が削除されている。図 3 6 は、図 3 2 の後の状態を表しており、レコード { 8 } が新たに追加され、譲ってもらったレコードについてのグループ ID が変更になっている。

【 0 1 0 5 】

さらに、メッシュ (3 / 2 , 3 / 2) についての処理に移行する。そうすると、図 3 7 に示すような分類結果が得られる。図 3 7 から分かるように、メッシュ要素 ID (- 1 , 0) (0 , 1) (1 , 1) は 1 及び a の条件を満たさない。メッシュ要素 ID (0 , 0) の場合には、これまでに生成されたグループにおいて譲れないレコードは { 1 , 3 , 4 , 5 , 6 , 9 , 10 , 11 , 12 } であるから、 $Ra = \{ 2 , 7 , 8 \}$ となる。これは、1 及び a の条件を満たすので、 $Ra = Rs$ となる。但し、{ 2 , 7 , 8 } は皆グループビンギ済みであるので、新たなグループが生成されないことになる。

【 0 1 0 6 】

以下、他のメッシュについて処理を行っても、新たにグループが生成されることはない。

【 0 1 0 7 】

以上の処理結果を、平面上に表したのが図 3 8 である。図 3 8 において、三角は ID 「 A 」のレコードを表し、×は ID 「 B 」のレコードを表し、丸は ID 「 C 」のレコードを表し、このような記号の脇の番号は行番号を表す。

【 0 1 0 8 】

グループ「 1 」は矩形 1 0 0 1 に対応し、レコード { 1 , 2 , 3 , 5 } が所属する。また、グループ「 2 」は矩形 1 0 0 2 に対応し、レコード { 9 , 10 , 11 } が所属する。グループ「 3 」は矩形 1 0 0 3 に対応し、レコード { 4 , 6 , 7 , 8 , 12 } が所属する。このように、重なっている領域部分に含まれるレコードを分け合っており、各矩形において特定の ID のみが所属することはない。

【 0 1 0 9 】

図 5 の処理の説明に戻って、匿名化処理部 1 5 0 は、第 2 データ格納部 1 4 0 に格納されているデータから、グループ毎に匿名化を実行し、処理結果を第 3 データ格納部 1 6 0 に格納する (ステップ S 5)。例えば、グループに対応するメッシュ要素の中心点に、そのグループに所属するレコードが存在するようにデータを変換すると共に、機密属性である ID を削除する。例えば、図 4 に示したレコード群の場合には、図 3 9 に示すようなデータが得られる。図 4 の (X , Y) が、メッシュ要素の中心座標で置換されていることが分かる。このように匿名化されて、個々のレコードが特定されることはない。

【 0 1 1 0 】

そして、出力部 1 7 0 は、第 3 データ格納部 1 6 0 に格納されているデータを、出力装

10

20

30

40

50

置（例えば表示装置、印刷装置やネットワークで接続された他のコンピュータ）に出力する（ステップS7）。

【0111】

以上のように処理を行えば、匿名性と分析精度を両立したデータを開示することができるようになる。また、メッシュ生成は高速に行われるので、全体の匿名化処理をも高速に行える。

【0112】

多様性の条件1を満たすようにグルーピングされたレコードだけが開示されるが、各グループには機密属性値について1種類以上あるので、各グループのレコードが特定のIDしか存在し得ないようなメッシュ要素に包含されることはなく、匿名性が担保される。

10

【0113】

また、開示される各メッシュ要素は大きさd未満となるため、小さいdを指定することで高精度な分析を期待される。但し、dが小さいほどいずれのメッシュ要素にも分類されないレコードが増え、そのようなレコードは開示されないため、dを小さくし過ぎても良くない。

【0114】

さらに、複数種類のメッシュで多様性の条件1を満たすか否かを判定するため、多様性1の条件を満たさないレコード数を減らすことができる。これは、開示されるレコード数を増やす、つまり分析に使えるデータ量が増えることになるので、高精度な分析が期待される。

20

【0115】

以上本発明の実施の形態を説明したが、本発明はこの実施の形態に限定されるものではない。例えば、図3に示した機能ブロック図は一例であって、プログラムモジュール構成やファイル構成とは一致しない場合もある。処理フローについても、処理結果が変わらない限り、ステップの順番を入れ替えたり、複数ステップを並列に実行するようにしても良い。

【0116】

なお、図6のメッシュ生成処理は一例であって、他の手順で同様のメッシュを生成しても良いし、さらに異なる態様のメッシュを生成するようにしても良い。予め単位サイズのメッシュを所定種類だけ生成しておき、サイズdに応じてサイズ変換を行うようにしても良い。

30

【0117】

なお、上で述べた情報処理装置100は、例えばコンピュータ装置であって、図40に示すように、メモリ2501とCPU2503とハードディスク・ドライブ（HDD）2505と表示装置2509に接続される表示制御部2507とリムーバブル・ディスク2511用のドライブ装置2513と入力装置2515とネットワークに接続するための通信制御部2517とがバス2519で接続されている。オペレーティング・システム（OS：Operating System）及び本実施例における処理を実施するためのアプリケーション・プログラムは、HDD2505に格納されており、CPU2503により実行される際にはHDD2505からメモリ2501に読み出される。CPU2503は、アプリケーション・プログラムの処理内容に応じて表示制御部2507、通信制御部2517、ドライブ装置2513を制御して、所定の動作を行わせる。また、処理途中のデータについては、主としてメモリ2501に格納されるが、HDD2505に格納されるようにしてもよい。本技術の実施例では、上で述べた処理を実施するためのアプリケーション・プログラムはコンピュータ読み取り可能なリムーバブル・ディスク2511に格納されて頒布され、ドライブ装置2513からHDD2505にインストールされる。インターネットなどのネットワーク及び通信制御部2517を経由して、HDD2505にインストールされる場合もある。このようなコンピュータ装置は、上で述べたCPU2503、メモリ2501などのハードウェアとOS及びアプリケーション・プログラムなどのプログラムとが有機的に協働することにより、上で述べたような各種機能を実現する。

40

50

【 0 1 1 8 】

以上述べた本実施の形態をまとめると、以下ようになる。

【 0 1 1 9 】

本実施の形態に係る匿名化データ生成方法は、(A) データ格納部に格納されており且つ各々機密属性値と数値属性値とを含む複数のデータブロックにおける各数値属性で張られる空間における複数種類のメッシュの各々について、複数のデータブロックのうち、当該メッシュの1つのメッシュ要素に含まれ且つグループ化未了の第1のデータブロックを含み且つ機密属性値の度数分布が所定の条件を満たすデータブロックのグループを抽出し、(B) 上記グループに属するデータブロックの数値属性値を、上記グループについての数値属性値で置換する処理を含む。

10

【 0 1 2 0 】

機密属性値の度数分布が所定の条件を満たすデータブロックのグループが生成されるので、確実な匿名化がなされる。また、複数種類のメッシュを用いることで、いずれかのメッシュのメッシュ要素でグループ化が行われる可能性があるため、開示されないデータブロックの数を減らすことができる。すなわち、分析に用いることができるデータ量が増加し、分析精度が向上する。また、メッシュ要素のサイズについても調整可能であり、この点についても分析精度向上の要因となる。メッシュ要素のサイズは、メッシュの各種類について同じにする場合もある。さらに、任意の位置にグループ化のための範囲を規定することによりグループ化を行う場合に比して処理が高速化される。

【 0 1 2 1 】

さらに、本匿名化データ生成方法は、上で述べたグループに属するデータブロックの機密属性値を削除する処理をさらに含むようにしても良い。これによって安全なデータを開示できるようになる。

20

【 0 1 2 2 】

さらに、上で述べた所定の条件が、機密属性値の種類数の下限値を含む場合もある。この場合、上で述べた抽出する処理は、(a 1) グループ化未了の第1のデータブロックを含み且つ上記1つのメッシュ要素に含まれるデータブロックの集合における機密属性値の種類が、機密属性値の種類数の下限値以上となっているか判断する処理を含むようにしても良い。このようにすれば、秘匿化の安全性が高まる。

【 0 1 2 3 】

さらに、上で述べた抽出する処理は、(a 2) グループ化未了の第1のデータブロックを含み且つ上記1つのメッシュ要素に含まれるデータブロックの第2の集合についての機密属性値の度数分布が、所定の条件を満たすか判断し、(a 3) データブロックの第2の集合についての機密属性値の度数分布が所定の条件を満たさない場合には、所定の条件を満たすようにデータブロックの第2の集合から第2のデータブロックを除外することでデータブロックのグループを生成する処理を含むようにしても良い。例えば、データブロックの第2の集合についての機密属性値の種類数が、その下限値以上となっていれば、適切なデータブロックを除外することで度数分布の条件を満たすようになる場合もある。

30

【 0 1 2 4 】

また、上で述べた抽出する処理は、(a 4) データブロックのグループから、機密属性値の度数分布が所定の条件を満たす上で必須となるデータブロック以外の第3のデータブロックを抽出する処理を含むようにしても良い。この場合、データブロックの他のグループを抽出する処理において、第3のデータブロックが他の種類のメッシュにおけるメッシュ要素において抽出される場合もある。このようにすれば、第3のデータブロックが他のグループ生成に用いられるので、削除されてしまうデータブロックを削減することができるようになる。

40

【 0 1 2 5 】

なお、本実施の形態に係る匿名化データ生成方法は、(C) メッシュ要素の境界を共有しないように、複数種類のメッシュを生成する処理をさらに含む場合もある。このようにすれば、匿名化されないデータブロックをより効率的に削減できるようになる。

50

【 0 1 2 6 】

また、本実施の形態に係る匿名化データ生成方法は、(D) これまでに生成されたメッシュの境界のうち最も近い境界との距離が最大となり、且つこれまでに生成されたメッシュの格子点との距離が最大となる基準点を選択し、(E) 選択された基準点及びメッシュ要素のサイズに基づきメッシュを生成する処理をさらに含むようにしても良い。このようにすれば、匿名化されないデータブロックをより効率的に削減できるようになる。

【 0 1 2 7 】

なお、上で述べたような処理をコンピュータに実施させるためのプログラムを作成することができ、当該プログラムは、例えばフレキシブル・ディスク、CD-ROMなどの光ディスク、光磁気ディスク、半導体メモリ(例えばROM)、ハードディスク等のコンピュータ読み取り可能な記憶媒体又は記憶装置に格納される。なお、処理途中のデータについては、RAM等の記憶装置に一時保管される。

10

【 0 1 2 8 】

以下、本実施の形態に係る付記を記す。

【 0 1 2 9 】

(付記 1)

データ格納部に格納されており且つ各々機密属性値と数値属性値とを含む複数のデータブロックにおける各数値属性で張られる空間における複数種類のメッシュの各々について、前記複数のデータブロックのうち、当該メッシュの1つのメッシュ要素に含まれ且つグループ化未了の第1のデータブロックを含み且つ機密属性値の度数分布が所定の条件を満

20

すデータブロックのグループを抽出し、
前記グループに属するデータブロックの数値属性値を、前記グループについての数値属性値で置換する

処理を含み、コンピュータにより実行される匿名化データ生成方法。

【 0 1 3 0 】

(付記 2)

前記グループに属するデータブロックの機密属性値を削除する

処理をさらに含む付記1記載の匿名化データ生成方法。

【 0 1 3 1 】

(付記 3)

前記所定の条件が、機密属性値の種類数の下限値を含み、

前記抽出する処理は、

前記グループ化未了の第1のデータブロックを含み且つ前記1つのメッシュ要素に含まれるデータブロックの集合における機密属性値の種類が、前記機密属性値の種類の下限值以上となっているか判断する

30

処理を含む付記1又は2記載の匿名化データ生成方法。

【 0 1 3 2 】

(付記 4)

前記抽出する処理は、

前記グループ化未了の第1のデータブロックを含み且つ前記1つのメッシュ要素に含まれるデータブロックの第2の集合についての機密属性値の度数分布が、前記所定の条件を満たすか判断し、

40

前記データブロックの第2の集合についての機密属性値の度数分布が前記所定の条件を満たさない場合には、前記所定の条件を満たすように前記データブロックの第2の集合から第2のデータブロックを除外することでデータブロックのグループを生成する

処理を含む付記1乃至3のいずれか1つ記載の匿名化データ生成方法。

【 0 1 3 3 】

(付記 5)

前記抽出する処理は、

前記データブロックのグループから、前記機密属性値の度数分布が前記所定の条件を満

50

たす上で必須となるデータブロック以外の第3のデータブロックを抽出する処理を含み、

データブロックの他のグループを抽出する処理において、前記第3のデータブロックが他の種類のメッシュにおけるメッシュ要素において抽出される

付記1乃至4のいずれか1つ記載の匿名化データ生成方法。

【0134】

(付記6)

メッシュ要素の境界を共有しないように、前記複数種類のメッシュを生成する処理をさらに含む付記1乃至5のいずれか1つ記載の匿名化データ生成方法。

【0135】

(付記7)

これまでに生成されたメッシュの境界のうち最も近い境界との距離が最大となり、且つ前記これまでに生成されたメッシュの格子点との距離が最大となる基準点を選択し、選択された前記基準点及び前記メッシュ要素のサイズに基づきメッシュを生成する処理をさらに含む付記1乃至5のいずれか1つ記載の匿名化データ生成方法。

【0136】

(付記8)

データ格納部に格納されており且つ各々機密属性値と数値属性値とを含む複数のデータブロックにおける各数値属性で張られる空間における複数種類のメッシュの各々について、前記複数のデータブロックのうち、当該メッシュの1つのメッシュ要素に含まれ且つグループ化未了の第1のデータブロックを含み且つ機密属性値の度数分布が所定の条件を満たすデータブロックのグループを抽出し、

前記グループに属するデータブロックの数値属性値を、前記グループについての数値属性値で置換する

処理を、コンピュータに実行させるための匿名化データ生成プログラム。

【0137】

(付記9)

データ格納部に格納されており且つ各々機密属性値と数値属性値とを含む複数のデータブロックにおける各数値属性で張られる空間における複数種類のメッシュの各々について、前記複数のデータブロックのうち、当該メッシュの1つのメッシュ要素に含まれ且つグループ化未了の第1のデータブロックを含み且つ機密属性値の度数分布が所定の条件を満たすデータブロックのグループを抽出するグループ化処理部と、

前記グループに属するデータブロックの数値属性値を、前記グループについての数値属性値で置換する匿名化処理部と、

を有する情報処理装置。

【符号の説明】

【0138】

- 110 第1データ格納部
- 120 設定データ格納部
- 130 グループ化処理部
- 140 第2データ格納部
- 150 匿名化処理部
- 160 第3データ格納部
- 170 出力部
- 180 メッシュ生成部
- 190 第4データ格納部

10

20

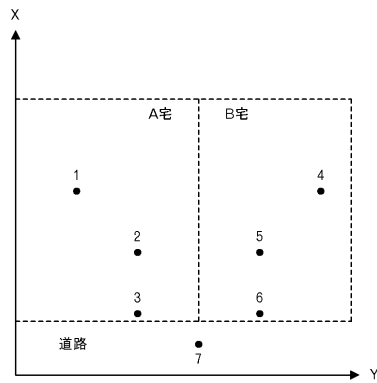
30

40

【図 1】

行番号	ID	X(緯度)	Y(経度)
1	A	6	2
2	A	4	4
3	A	2	4
4	B	6	10
5	B	4	8
6	B	2	8
7	C	1	6

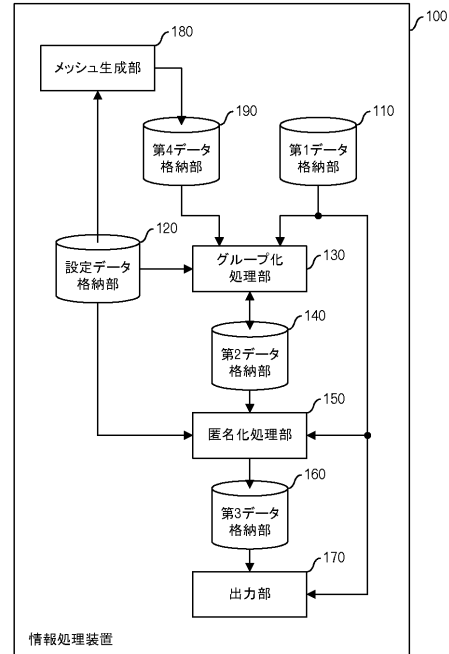
【図 2】



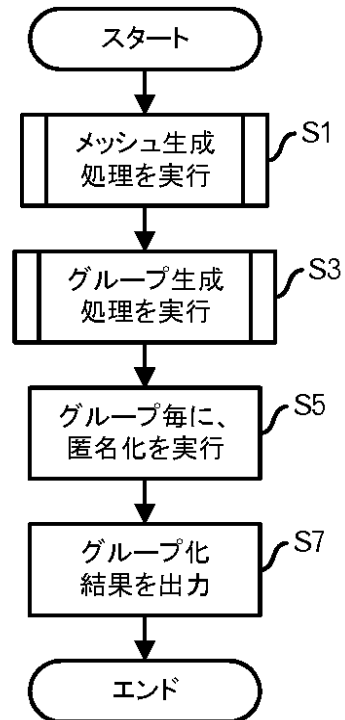
【図 4】

行番号	ID	X	Y	速さ
1	A	1	4	0
2	A	2	5	2
3	B	3	3	1
4	B	4	5	1
5	C	5	3	2
6	C	5	4	3
7	B	5	5	1
8	C	6	5	1
9	B	6	6	2
10	C	7	6	3
11	A	7	8	1
12	A	8	8	4

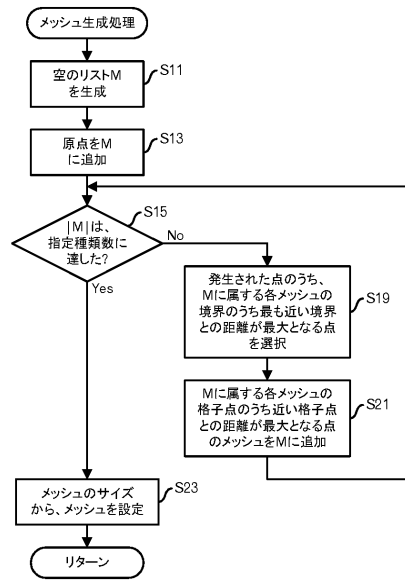
【図 3】



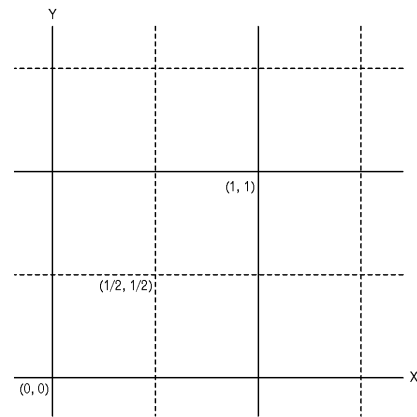
【図 5】



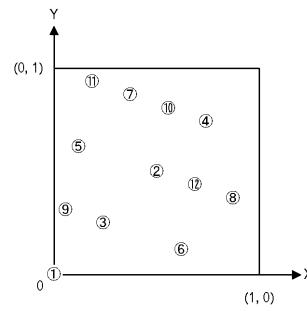
【図 6】



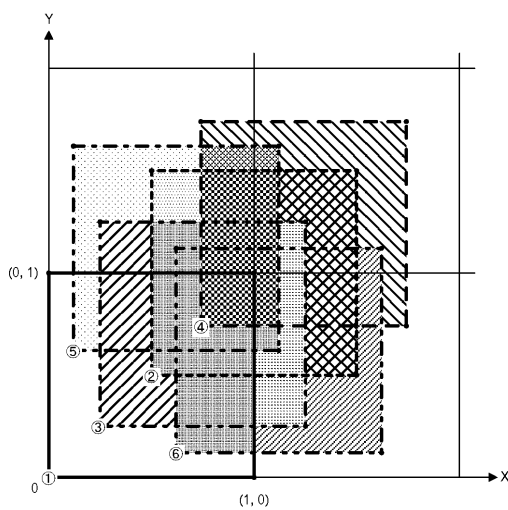
【図 7】



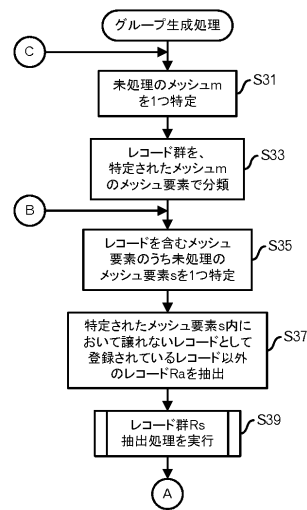
【図 8】



【図 9】



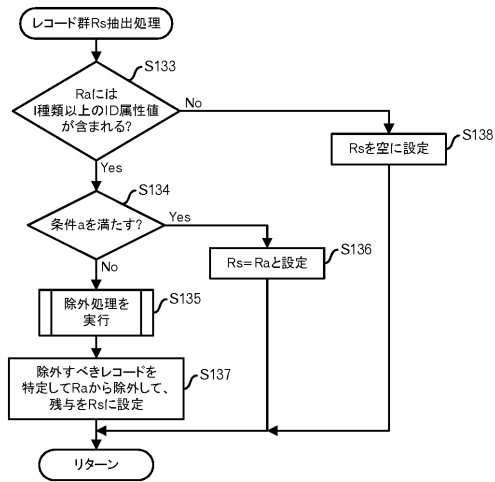
【図 10】



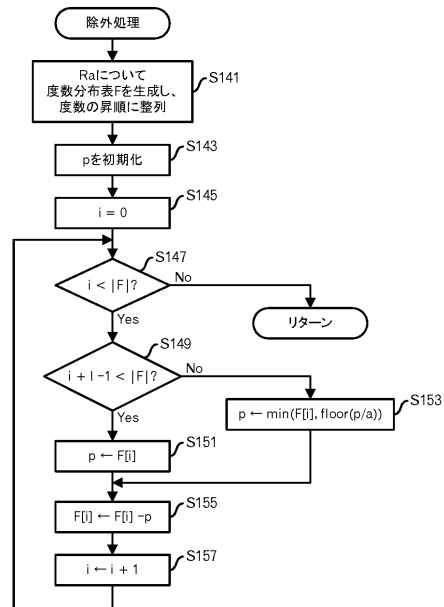
【図 11】

メッシュ要素ID	行番号
(0, 0)	1, 2, 3, 4, 5, 6, 7
(1, 0)	8
(1, 1)	9, 10, 11, 12

【図 1 2】



【図 1 3】



【図 1 4】

ID	度数
A	1
C	3
D	4
E	9
B	10

【図 1 6】

ID	度数
A	0
C	0
D	4
E	9
B	10

【図 1 5】

ID	度数
A	0
C	3
D	4
E	9
B	10

【図 1 7】

ID	度数
A	0
C	0
D	0
E	9
B	10

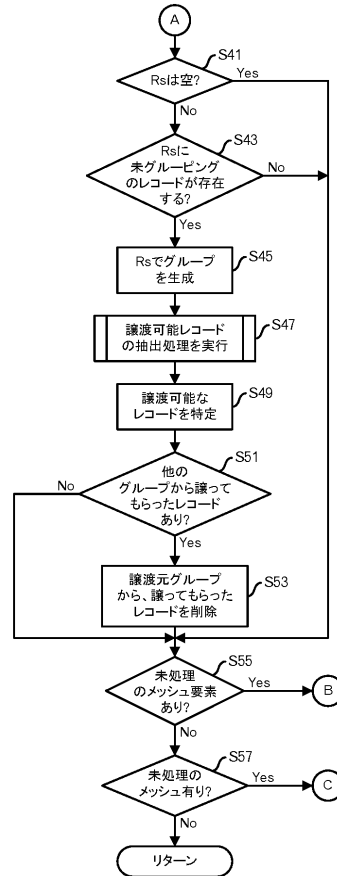
【図 18】

ID	度数
A	0
C	0
D	0
E	1
B	10

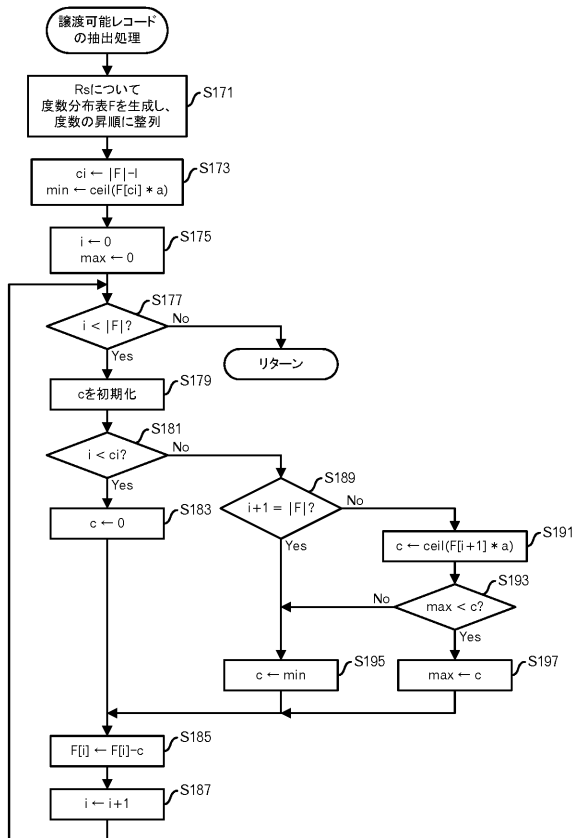
【図 19】

ID	度数
A	0
C	0
D	0
E	1
B	0

【図 20】



【図 21】



【図 22】

ID	度数
A	1
B	2
C	3
D	4
E	5

【図 23】

ID	度数
A	1
B	0
C	3
D	4
E	5

【図 2 4】

ID	度数
A	1
B	0
C	2
D	4
E	5

【図 2 5】

ID	度数
A	1
B	0
C	2
D	1
E	5

【図 2 6】

ID	度数
A	1
B	0
C	2
D	1
E	4

【図 2 7】

G_ID	メッシュ	メッシュ要素ID
1	(0, 0)	(0, 0)

【図 2 8】

G_ID	譲れない 行番号	譲れる 行番号
1	1, 3, 5	2, 4, 6, 7

【図 2 9】

行番号	G_ID
1	1
2	1
3	1
4	1
5	1
6	1
7	1

【図 3 1】

G_ID	譲れない 行番号	譲れる 行番号
1	1, 3, 5	2, 4, 6, 7
2	9, 10, 11	12

【図 3 2】

行番号	G_ID
1	1
2	1
3	1
4	1
5	1
6	1
7	1
9	2
10	2
11	2
12	2

【図 3 0】

G_ID	メッシュ	メッシュ要素ID
1	(0, 0)	(0, 0)
2	(0, 0)	(1, 1)

【図 3 3】

メッシュ要素ID	行番号
(-1, 0)	1, 2
(0, 0)	3, 4, 5, 6, 7, 8, 9, 10, 11, 12

【図 3 4】

G_ID	メッシュ	メッシュ要素ID
1	(0, 0)	(0, 0)
2	(0, 0)	(1, 1)
3	(3, 3)	(0, 0)

【図 3 5】

G_ID	譲れない 行番号	譲れる 行番号
1	1, 3, 5	2
2	9, 10, 11	
3	4, 6, 12	7, 8

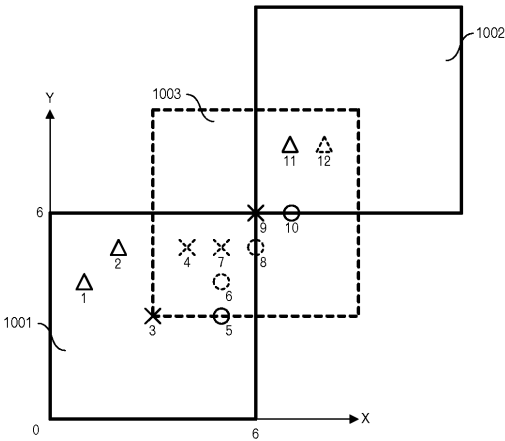
【図 3 6】

行番号	G_ID
1	1
2	1
3	1
4	3
5	1
6	3
7	3
8	3
9	2
10	2
11	2
12	3

【図 3 7】

メッシュ要素ID	行番号
(-1, 0)	1
(0, 0)	2, 3, 4, 5, 6, 7, 8, 9, 10
(0, 1)	11
(1, 1)	12

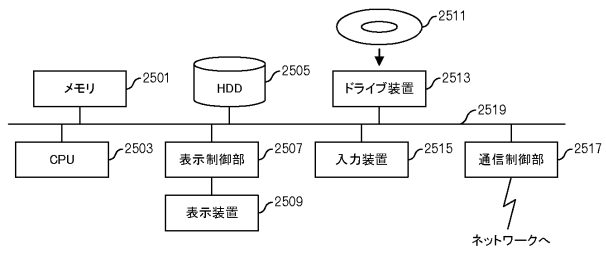
【図 3 8】



【図 3 9】

行番号	X	Y	速さ
1	3	3	0
2	3	3	2
3	3	3	1
4	6	6	1
5	3	3	2
6	6	6	3
7	6	6	1
8	6	6	1
9	9	9	2
10	9	9	3
11	9	9	1
12	6	6	4

【図 40】



フロントページの続き

(56)参考文献 特開2012-159982(JP, A)

国際公開第2012/090628(WO, A1)

国際公開第2011/142327(WO, A1)

山岡裕司他, 1 - 多様性を達成するプライバシー保護位置データ開示, 2013年 暗号と情報
セキュリティシンポジウム SCIS2013 [CD-ROM] 2013年 暗号と情報セ
キュリティシンポジウム概要集, 電子情報通信学会, 2013年 1月, p.1-p.8

(58)調査した分野(Int.Cl., DB名)

G06F 21/62