

平成 29 年度
学士学位論文

部分復元可能な秘密分散バックアップした 医療データ検索方法

Retrieval method for Medical Data by confined
decodable Secret Sharing Scheme

1180362 沼 尚樹

指導教員 福本 昌弘

2018 年 3 月 16 日

高知工科大学 情報学群

要 旨

部分復元可能な秘密分散バックアップした医療データ検索方法

沼 尚樹

災害に備えて、高知県の病院では医療データを遠隔地にバックアップを行う取り組みがなされている。災害時に被災地域外からきた医師に対して、バックアップした医療データを提供することができれば、円滑に治療を行うことができる。しかし、医療データは個人情報であるため、治療に不必要な情報の提供を制限する必要がある。そのため、治療に必要な情報のみを提供すればよい。そこで、バックアップしたデータから必要な情報を提供できるように部分復元可能な秘密分散アルゴリズムが提案された。災害時、バックアップしたデータを利用する際、患者のデータを検索しなければならない。しかし、部分復元可能な秘密分散を行ったデータを検索する仕組みは考えられていない。そのため、検索を行うには名前などの情報を部分的に復元する必要がある。しかし、部分復元を行うには時間がかかってしまい、素早く検索することができない。

本論文では、部分復元可能な秘密分散を行った医療データを検索する仕組みを提案している。提案する方法は、バックアップを行うデータに検索用の情報を付与し検索を行えるようにする。そのため、付与する情報の設定を行っている。また、検索を素早く行うために、付与する情報のサイズの削減を提案している。60 個の姓のデータを用いて、姓の特定のために必要な文字数を調べた結果、先頭 3 文字を付与すればよいことを明かにしている。

キーワード 部分復元, データ検索

Abstract

Retrieval method for Medical Data by confined decodable Secret Sharing Scheme

NUMA Naoki

To provide against disaster situation, an effort toward making remote backup of medical data is implemented by hospitals in Kochi prefecture. Disaster Medical Assistance Team(DMAT) is a rescue team for disaster situations. In order to optimizing their rescue performance, an essential core of DMAT have to requests information of curing patients from data centers, medicines information. However, medical data is private information, so, it is not allowed to view all information. Therefore, only the necessary data, for curing victim, should be provided. In order to provide a selectively feature, the secret sharing scheme algorithm was proposed. The proposed algorithm can partly refine the specific information. But, there is no way to search for backed up data using this algorithm. Therefore, the proposed algorithm needs to restore a partial information to search for data. This takes time to do a search for data.

In this research, the computationally efficient mechanism to retrieve medical data with the remote backup of medical data and secret sharing scheme is proposed. The proposed method uses part of the medical data to give the backup data a share for the search. Therefore, the information to be granted is determined. Propose to reduce the amount of information to be granted to increase the search speed. The number of characters to be granted was examined using 60 family name data. The result showed that give the first three characters.

key words confined decoding, data retrieval

目次

第 1 章	序論	1
1.1	研究の背景と目的	1
1.2	本論文の構成	2
第 2 章	医療データのバックアップと災害時の活用	3
2.1	対象となる医療データ	3
2.1.1	電子カルテ	4
2.1.2	SS-MIX	4
2.2	医療データのバックアップ条件	4
2.3	秘密分散法を用いたバックアップ	6
2.4	部分復元可能な秘密分散法	6
2.5	部分復元可能な秘密分散バックアップした医療データの災害時での利用と 問題点	6
第 3 章	部分復元可能な秘密分散バックアップした医療データ検索方法	8
3.1	提案する検索方法	8
3.2	付与する検索情報	9
3.2.1	付与する検索情報の条件	9
3.2.2	医療データから得られる患者情報	10
3.2.3	付与する検索情報	10
3.3	付与した検索情報の優先度設定	11
3.3.1	検索情報の聞き出しやすさ	11
3.3.2	検索情報の優先度	12
3.4	付与する検索情報のサイズ削減と検索方法	12

目次

3.4.1	患者姓の付与	13
3.4.2	患者姓シェアの検索方法	13
3.4.3	西暦の付与	15
3.4.4	西暦のシェア作成と検索方法	15
	100 歳代の検索	16
	10 代未満, 10 代の検索	17
3.5	提案した検索方法の評価	18
3.6	考察	18
第 4 章	結論	20
4.1	まとめ	20
4.2	今後の課題	20
	謝辞	22
	参考文献	23

目次

2.1	標準化ストレージのディレクトリ構造	5
2.2	標準化された患者データ例	5
2.3	部分復元可能な秘密分散法の流れ	7
3.1	検索用情報の付与の流れ	9
3.2	姓を検索した場合	14
3.3	20 代の検索の流れ	16
3.4	100 歳代の検索の流れ	17

表目次

3.1	患者識別情報の種類	10
3.2	付与する検索情報	11
3.3	検索情報の聞き出しやすさ	12
3.4	検索情報の優先度	12
3.5	3 文字以上の姓を検索した場合の中央値と最大人数, 最小人数	13
3.6	1 文字ずつのシェアと 3 文字シェアの最大復元回数と演算量	14
3.7	1 桁ずつのシェアと 3 桁全てのシェアの最大復元回数と演算量	15
3.8	年代人口	15
3.9	2 桁目のみを復元した場合の該当する男女別の年代人口	16
3.10	1 桁目を復元した場合の該当する年代別の人数	16
3.11	1 桁目を復元した場合の該当する 100 歳代の人数	17
3.12	60 個の姓を検索した結果	18
3.13	年代を検索した場合の人数	18

第 1 章

序論

1.1 研究の背景と目的

東日本大震災の津波により，病院に保管されてあった医療データが損失したこととで被災地での医療行為に支障がでた．このことから，高知県の病院では電子カルテを遠隔地にバックアップを行う取り組みが行われている [1]．バックアップを行った電子カルテを医療行為に活用することが想定される．災害時は，被災地域外から DMAT などの災害派遣医療チームも患者の治療に当たる．被災地域外から来た医師にも患者の医療データを閲覧することで，円滑な治療を行える．しかし，医療データは個人情報であるため，治療に必要な情報のみを提供できればよい．そこで，バックアップされたデータから治療に必要な情報のみを復元することのできる部分復元可能な秘密分散法アルゴリズムが提案された [2]．しかし，提案されたアルゴリズムを用いてバックアップを行った場合，患者のデータを検索する方法は考えられていない．そのため，検索を行うには，名前部分などを部分的に復元しなければならない．部分復元には復元に時間がかかるため患者データを災害時に検索方法としては適さない．

本研究では，部分復元可能な秘密分散法を用いてバックアップを行った医療データを少しでも，早く検索できるようにすることを目的とする．提案方式ではバックアップデータに検索情報を付与する．これによって検索時は付与した情報を検索し，該当する患者データを絞り込む．また，素早く検索できるようにするために，検索情報サイズの削減と検索方法を定める．

1.2 本論文の構成

本論文の構成について述べる．第 2 章では，医療データのバックアップと部分復元可能な秘密分散法について述べる．第 3 章では，本研究で提案した部分復元可能な秘密分散を用いてバックアップを行った医療データの検索方法について述べる．その後，提案方式の評価と考察を述べる．第 4 章では，本研究の結論と今後の課題について述べる．

第 2 章

医療データのバックアップと災害時の活用

東日本大震災では、津波の影響により病院に保管されてあった電子カルテなどの医療データが流失し被災地での医療行為に支障がでた。このことから、高知県の病院では医療データを遠隔地にバックアップする取り組みが行われている。また、バックアップを行った医療データを災害時に活用することが期待されている。災害時に患者の治療を行う医師の中には、被災地域外から来る医師もいる。そのため、被災地域外の医師にも医療データを提供することにより、円滑な治療が行える。しかし、医療データは個人情報のため治療に必要な情報のみを提供できればよい。そこで、バックアップしたデータから部分的な情報のみを提供可能とする部分復元可能な秘密分散法が提案された [2]。

本章では、本研究で対象となる医療データについて述べる。次に医療データのバックアップ条件について述べる。次に、部分復元可能な秘密分散法の流れについて述べる。そして、部分復元可能な秘密分散法を用いて医療データをバックアップした場合の問題点を述べる。

2.1 対象となる医療データ

本研究で対象とする医療データについて定める。医療データは様々な種類があるが、本研究で対象となるデータは電子カルテである。本節では、本研究で利用する医療データである電子カルテについて述べる。まず、電子カルテについて述べる。次に標準規格である SS-MIX について述べる。

2.2 医療データのバックアップ条件

2.1.1 電子カルテ

本研究で対象とする医療データは電子カルテである。電子カルテは、医師が患者を診察した際に診療結果や経過などを記述した診療録を電子化したものである。災害時にも電子カルテを閲覧できれば円滑な治療ができる。カルテを電子化することで、紛失のリスクやカルテの保存場所の削減になる。また、病院内のネットワークにつながり、閲覧権限を持っていればどこでも閲覧することが可能である。しかし、電子カルテは様々なベンダーによって実装されており、互換性や標準規約が定まっていない。そのため、病院間で電子カルテの情報を共有することが困難であった。

2.1.2 SS-MIX

電子カルテは様々なベンダーによって実装されているため、病院間で電子カルテを共有するのが困難である。そこで、医療情報の交換、共有を可能にするため厚生労働省電子的診療情報交換推進事業 (SS-MIX) が策定された。SS-MIX では各病院の電子カルテを HIS 情報ゲートウェイで変換を行い、標準化ストレージに標準化された医療データが格納される。SS-MIX はディレクトリ構造を用いている。SS-MIX のディレクトリ構造を図 2.1 に示す。ディレクトリ構造は各医療機関ごとにルートフォルダを置き患者 ID、診療日、データ種別順に分けられその配下に標準化された医療データが配置される。標準化されたデータの例をを図 2.2 に示す。標準化された患者データの中身は、関連する一連の情報項目のグループ (セグメント) ごとに記述されている。使用するセグメントはデータ種別ごとに異なる。

2.2 医療データのバックアップ条件

本節では、医療データのバックアップする上での条件を述べる。医療データのバックアップを行う上での条件を示す。1 つ目は、医療データは個人情報であるためバックアップしたデータから情報を読み取られないようにしなければならない。2 つ目は、バックアップデータから医療データを遠隔地にバックアップした際にバックアップ先のストレージが被災して

2.2 医療データのバックアップ条件

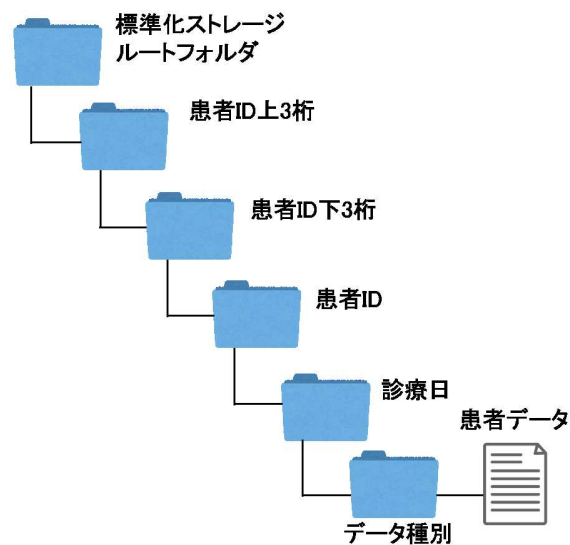


図 2.1 標準化ストレージのディレクトリ構造

```

MSH|^~\&|EGMAIN-GX|HIS|GW|GW|20130725191411.6701||OMP^009|20130725191411709|P|2.5|||||~ISO IR87||ISO 2022-199.
ZGW|9930000036|20130725191339|OMP-01^処方オーダー^L|137253020910400|INS|01^総合診療科^L←
SFT|Fujitsu Ltd.^D|V2|EGMAIN-GX←
NTE|0001|P|一包化する〈錠剤・カプセル剤〉←
PID|0001||9930000036||テスト^資源配布^~~~~^L^I^テスト^シゲンハイフ^~~~~^L^P||19750505|M|~~~~~|N||
PV1|0001|O|01^~~~~C|~~~~|01←
ZIN|01^政府管掌健康保険^HL7-0072←
ZIN|12^生活保護法による医療扶助^HL7-0072←
ORC|NW|137253020910400|137253020910400|1|~~~~|20130725191402|1000511^北村^和之^~~~~^L^~~~~^I^キタムラ^カズユキ^
TQ1|0001|1^日分|BID&1日2回&HL7-0335^PCM&朝食後&HL7-0335^PCD&昼食後&HL7-0335&&1日2回 朝・昼食後|1^日分|201307:
RXO|^ストラテラカプセル□ 4 0 m g^MDCHOT^I1021710^ストラテラカプセル□ 4 0 m g^L|0.5||CP^C P^L|CAP^カプセル剤
RXR|PO^□^HL7-0162←
  
```

図 2.2 標準化された患者データ例

も医療データを復元できなければならない。3 つ目は治療に必要な情報のみを提供できればよい。災害時、被災地域外から DMAT などの医療チームが患者の治療に当たる。被災地域外から来た医師にも患者の医療データを閲覧することで、円滑な治療を行えるが、医療データは個人情報であるため治療に必要な情報のみを提供できればよい。以上の条件をまとめる。

1. バックアップデータの秘匿化
2. バックアップデータの冗長化
3. 災害時の治療に必要な情報のみを提供可能

この 3 つの条件を満たしたバックアップ方法が部分復元可能な秘密分散である。

2.3 秘密分散法を用いたバックアップ

冗長化と秘匿化を行うバックアップ方法として、秘密分散を用いたバックアップがある。秘密分散法は分散と復元の工程がある。分散ではデータから複数の無意味な情報を生成する。これをシェアと呼ぶ。復元を行う際は、シェアを一定数集めることで復元することができる。シェアは一定数以上集めないとデータの情報をえることができないため、秘匿化が行える。また、全てのシェアを集めなくてもデータを復元できるので冗長化が行える。しかし、データを部分的に復元することができない。

秘密分散法では、部分的に復元することができないため、治療に必要な情報のみを提供することができない。そのため、部分復元可能な秘密分散法が提案された [2]。これにより、秘密分散を用いても部分復元を行うことが可能となり、冗長化、秘匿化、必要な情報のみを提供可能となる。

2.4 部分復元可能な秘密分散法

秘密分散法では分散したデータを部分的に復元することができないことから、部分復元可能な秘密分散法が提案された [2]。SS-MIX 形式の電子カルテをバックアップした場合を図 2.3 に示す。分散時には、データを項目ごとに分け、分けたデータをそれぞれを秘密分散しシェアを生成する。復元するには、復元したい項目のシェアを集め、シェアを結合し復元することで部分復元を行うことができる。

2.5 部分復元可能な秘密分散バックアップした医療データの 災害時での利用と問題点

本節では、バックアップしたデータの利用と問題点について述べる。

災害時、バックアップした医療データを提供することで、円滑に治療を行うことができるとされる。バックアップされた医療データから患者の情報を得るには、検索を行い該当する

2.5 部分復元可能な秘密分散バックアップした医療データの災害時での利用と問題点

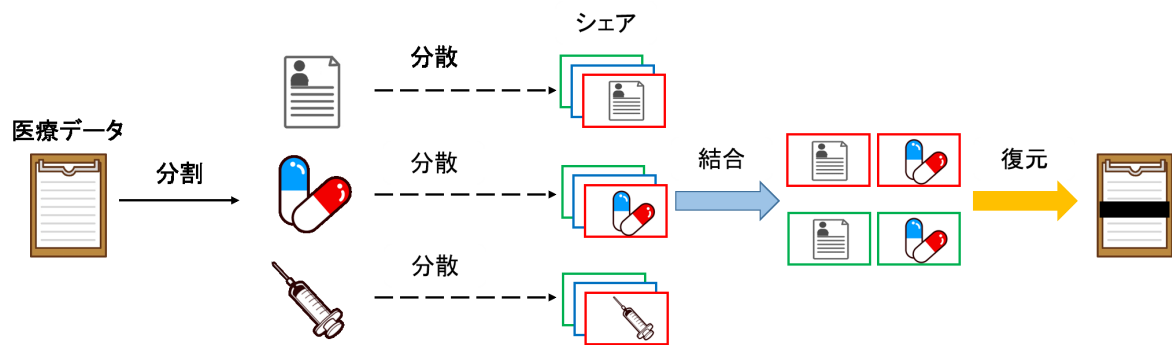


図 2.3 部分復元可能な秘密分散法の流れ

患者のデータを見つけなければならない。しかし、部分復元可能な秘密分散を行ったデータを検索する仕組みは考えられていない。そのため、部分復元可能な秘密分散を行ったデータの検索を行うには名前などの情報を部分的に復元を行い検索するしかない。しかし、部分復元可能なシェアのサイズは元データの大きさと等しいかそれ以上となってしまうので復元の処理時間がかかり、患者データを素早く検索することができない。

本研究ではバックアップを行うデータに患者データを検索用として検索情報の付与を行い部分復元可能な秘密分散バックアップした医療データを検索する仕組みを提案する。

第 3 章

部分復元可能な秘密分散バックアップした医療データ検索方法

部分復元可能な秘密分散したデータを検索する仕組みは考えられていない。そのためバックアップされたデータを検索するには名前などの情報を部分復元を行い検索しなければならない。しかし、この方法は素早く患者データを検索できないという問題点がある。

本研究ではこれらの問題を解決する方法として、バックアップする医療データに検索用シェアの付与を行い検索する方法を提案する。また、検索を素早く行うために検索情報のサイズを削減と検索方法を提案する。

本章では、提案する検索方法について述べる。次に、どの情報を付与するのかを定め、その優先度を設定する。また、優先度の高い情報について、検索情報のサイズ削減と検索方法を述べる。最後に評価と考察を述べる。

3.1 提案する検索方法

本節では、提案する検索方法の流れについて述べる。

部分復元可能な秘密分散法でバックアップした医療データを検索する仕組みは考えられていない。そこでバックアップするデータに検索用の情報を付与し、検索を行う仕組みを提案する。図 3.1 に提案方法の流れを示す。患者の医療データをバックアップする際に医療データにある患者情報を検索用のデータとして秘密分散法で分散する。分散して作成された検索用シェアを分散した患者の医療データに付与しバックアップを行う。検索時には検索用シェ

3.2 付与する検索情報

アを復元し，該当する患者データを見つける．

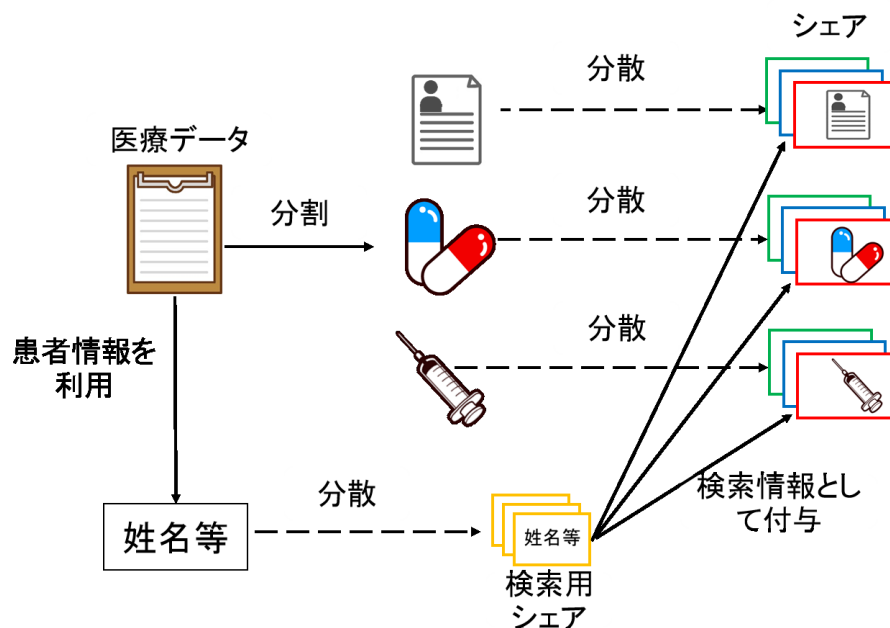


図 3.1 検索用情報の付与の流れ

3.2 付与する検索情報

提案方式では，バックアップを行うデータに検索情報を付与し検索する．そのため，検索用として付与する患者情報を定める．本節では，検索用として付与する情報の条件について述べる．次に条件を満たす患者情報について述べる．

3.2.1 付与する検索情報の条件

付与する患者情報の条件を示す以下に示す．

1. 医療データから得られる情報である事
2. 患者全員が持っている情報である事
3. 災害時に入手できる可能性のある情報である事

3.2 付与する検索情報

表 3.1 患者識別情報の種類

患者識別情報	患者 ID	患者姓	患者名	性別	西暦	月日	住所	電話番号
表記	数値	漢字とカタカナ表記	漢字とカタカナ表記	英 1 字	数値	数値	漢字など	数値

1 つ目については、本研究では医療データのバックアップを行うので付与する情報は医療データから得られる情報であり秘匿化しない部分であるとする。2 つ目は付与する情報の種類が異なると検索の際、患者ごとに検索する情報が異なってしまう。そのため、患者全員が持っている情報とする。3 つ目は災害時に入手できる情報である。災害時は患者自身の身分を証明できるものを持っていない可能性や患者から直接聞き出せない可能性がある。そのため、患者以外からでも聞き出せる可能性のある情報とする。以上が付与する情報の条件である。

3.2.2 医療データから得られる患者情報

付与する患者情報は医療データから得られる情報でなくてはならない。また、患者全員が持っており、災害時でも入手できる情報でなくてはならない。

標準化された患者医療データから得られる情報は様々である。しかし、条件を満たす情報が記載されているセグメントは患者識別セグメントである。患者識別セグメントに存在する情報とその表記を表 3.1 に示す。患者識別セグメントは SS-MIX 形式で標準化された患者医療データ全てにあるセグメントである。

3.2.3 付与する検索情報

バックアップデータに付与する患者情報を述べる。

医療データから得られる情報として、患者識別セグメントの情報がある。患者識別セグメントは標準化された患者医療データすべてに共通して存在する項目である。また、患者識別セグメントの情報は災害時でも入手できる情報があるため付与する検索用の情報として利用できる。しかし、患者 ID は災害時に分からないため付与する情報に含まない。表 3.2 に付

3.3 付与した検索情報の優先度設定

表 3.2 付与する検索情報

患者姓	患者名	性別	西暦	月日	住所	電話番号
-----	-----	----	----	----	----	------

与する検索情報を示す。

3.3 付与した検索情報の優先度設定

患者データを検索するために、患者の情報を聞き出さなければならない。しかし、災害時のような急性期では、患者が自身を証明できるものを所持していない場合や患者自身が答えられない可能性もある。そのため、付与した情報の聞き出しやすさが異なる。また、患者データを特定するには、付与した情報を組み合わせて検索を行わなければならない。これらのことから、災害時にどのような順番で付与した情報を検索するのかを示す。本節では、情報の聞き出しやすさを基準に検索情報の優先度を設定し検索の順番を示す。

3.3.1 検索情報の聞き出しやすさ

災害時、患者本人から直接情報を聞き出せない場合がある。そのような場合に最も入手できる情報は、性別と西暦である。性別は身体から判断可能である。西暦は年代の検索に利用する。見た目から何歳代かを推測することが可能である。どちらも、外見から判断できるが、性別の方が正確なので、1番とする。次に入手できる情報は患者姓または名である。患者の姓または名が分からない場合、誕生日や住所、電話番号について知っている可能性は低いからである。患者姓名では、姓のみ知っている場合もあるため、姓を3番、名を4番とする。次の情報としては、電話番号である。電話番号はスマートフォンや携帯電話等を所持していれば分かる可能性があるため5番とする。最後は月日と住所である。どちらも、本人や家族以外に知らない可能性があるためである。聞き出しやすさについては、月日が数値の表記であり少ない情報量で済むため6番とした。表 3.3 に検索情報の聞き出しやすさを示す。

3.4 付与する検索情報のサイズ削減と検索方法

表 3.3 検索情報の聞き出しやすさ

検索情報	患者姓	患者名	性別	西暦	月日	住所	電話番号
聞き出しやすさ	3	4	1	2	6	7	5

表 3.4 検索情報の優先度

検索情報	患者姓	患者名	性別	西暦	月日	住所	電話番号
優先度	1	4	2	3	6	7	5

3.3.2 検索情報の優先度

聞き出しやすさだけの場合、外見だけで判断可能な性別が最も優先度が高くなる。また、年代も外見で判断できるので組み合わせで検索できることになる。しかし、性別と年代を組み合わせで検索しても患者データを特定することはできない。そのため、他の検索情報が必要となる。他の検索情報は、患者の姓名を知らない则他の情報も知らない可能性が高い。よって、患者姓名を知っている場合について述べる。

検索情報の優先度を表 3.4 に示す。姓が最も優先度が高い。これは、検索時に性別よりデータを絞れることができ、姓を先に検索することで、検索情報の復元するデータ数を減らせるためである。4 番以降は聞き出しやすさと変わらない。住所や電話番号は該当するデータを絞ることはできるが、情報を入手できない可能性があるためである。

3.4 付与する検索情報のサイズ削減と検索方法

バックアップデータから該当する患者の医療データを見つけるには、付与した検索情報を復元してデータを絞っていく。災害時のような急性期において患者データを素早く見つけ出さなければならない。しかし、付与する情報のサイズが大きければ復元の処理に時間がかかってしまう。そこで、付与する情報のサイズの削減を行い復元の処理にかかる演算量を減らす。

3.4 付与する検索情報のサイズ削減と検索方法

本節では、優先度の高い患者姓と年代の検索において、それぞれ何文字付与すればよいのか述べる。また、姓と年代それぞれの検索方法を述べる。

3.4.1 患者姓の付与

姓情報の削減について述べる。姓の読みから先頭何文字でどのぐらいの人数がヒットするのかを示す。日本人の姓と名の分布 [3] から東邦生命の被保険者と保険者の上位 60 位の姓の読みを利用し、先頭何文字でどのぐらいの人数 (744849 人) が当てはまるのかを調べた。表 3.5 に 3 文字以上の姓を検索した場合の中央値とヒットする最大の人数を示す。データでは姓の文字数の最大は 4 文字であるが、3 文字付与すれば 1 つの姓が特定できるため、4 文字付与した場合の中央値とヒットする最大人数に変化はない。最大で 3 文字付与すれば 4 文字の姓も特定でき、1 文字や 2 文字より検索したときの該当するデータを絞り込むことができる。よって、患者姓は先頭 3 文字付与する。最小ヒット人数が同数なのは、最小ヒットする姓の先頭の読みがデータに 1 つしかないからである。

3.4.2 患者姓シェアの検索方法

患者姓の検索用シェアの検索方法について述べる。災害時はバックアップされたデータから検索を行い素早く患者データを見つけなければならない。そのため、検索において演算回数の少ない検索方法を提案する。

姓の情報を付与するにあたって、最大で姓の先頭 3 文字を付与することを提案した。付与

表 3.5 3 文字以上の姓を検索した場合の中央値と最大人数, 最小人数

付与文字数	1 文字付与	2 文字付与	3 文字付与	4 文字付与
中央値 (人)	29517	12939	8068	8068
最大人数 (人)	117776	61969	54689	54689
最小人数 (人)	5701	5701	5701	5701

3.4 付与する検索情報のサイズ削減と検索方法

表 3.6 1 文字ずつのシェアと 3 文字シェアの最大復元回数と演算量

	1 文字ずつのシェア	3 文字シェア
最大復元回数	917314	744849
演算量	$805951 \times n^3$	$744849 \times (3n)^3$

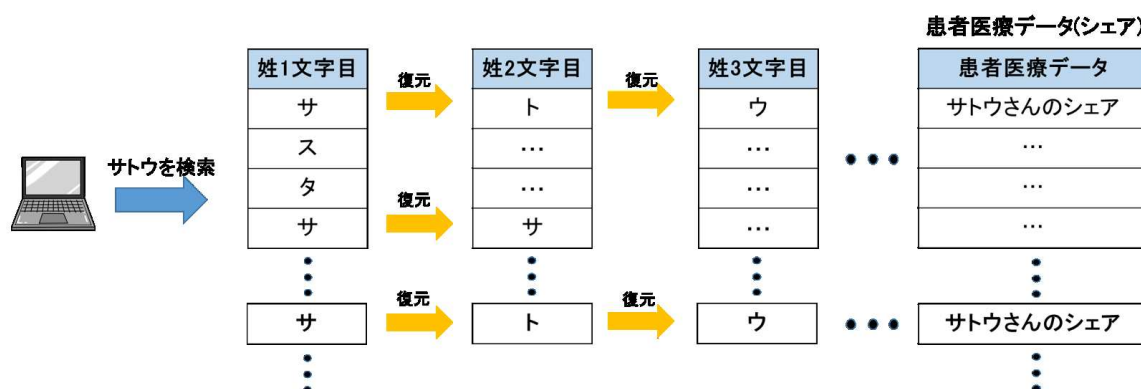


図 3.2 姓を検索した場合

した姓を分散し検索を行う方法として2通りの方法がある。1つ目は1文字ずつ分散し検索には1文字ずつ復元を行い復元するデータの数減らしていく方法。2つ目は3文字全てを分散し検索には全てのデータを検索する方法である。この2つの方法でどちらが演算回数が少ないかを60個の姓の種類を利用し示す。表3.6に60個の姓全て検索した場合の最大の復元回数とその演算回数を示す。演算量は1文字のシェアの復元演算量を n^3 とし、3文字のシェアは $(3n)^3$ とする。表3.6では復元回数は3文字シェアが少ない。しかし、3文字シェアは復元にかかる演算量が1文字ずつと比べ27倍であるため、演算量は1文字ずつの方が少ない。よって姓のシェアは1文字ずつ分散を行い検索では先頭から1文字ずつ復元していき、復元するデータ数を減らしていく方法をとる。1文字ずつ復元をを行い検索する流れを図3.2に示す。

3.4 付与する検索情報のサイズ削減と検索方法

表 3.7 1 桁ずつのシェアと 3 桁全てのシェアの最大復元回数と演算量

	1 桁ずつのシェア	3 桁全てのシェア
最大復元回数	849581	744849
演算量	$849581 \times n^3$	$744849 \times (3n)^3$

表 3.8 年代人口

年齢	10 代未満	10 代	20 代	30 代	40 代	50 代	60 代	70 代	80 代	90 代	100 代
人数	60430.36	68038.52	71654.93	89153.1	111286.5	90827.25	109141.6	82641.18	50236.5	11069.66	387.25
男性人数	30939.06	34895.06	36664.54	45351.07	56474.35	45541.72	53233.29	37701.2	19100.83	2645.28	53.62055
女性人数	29479.39	33119.63	34996.34	43802.03	54800.2	45279.57	55878.57	44945.94	31135.67	8424.384	339.59

3.4.3 西暦の付与

西暦の付与する桁数について述べる．現在の日本人の男女別の平均寿命は男 80.79 歳，女 87.05 歳である [4]．そのため，だいたいの年代は西暦の下 2 桁までで判断できる．しかし，100 歳を越えた人もいるため 100 歳代と 10 代，10 歳未満の世代は下 2 桁まででは判断できない．そのため，西暦は下 3 桁を付与する．

3.4.4 西暦のシェア作成と検索方法

西暦も一桁ずつ処理した方が演算量が少ないため，1 桁ずつ分散復元を行う．

検索では，2 桁目から復元を行い検索する．2 桁目を復元した場合に一番データを絞り込めるからである．本研究では，患者姓と 2016 年の日本人の年齢，男女別人口の統計データ [5] を利用し，744849 人を年齢人口の割合に当てはめた．3 桁分散し復元した場合と 1 桁ずつ分散し復元した場合の最大の復元回数と演算量を表 3.7 に示す．表 3.8 に当てはめた年代別の人口を示す．また，年齢，男女別の人口のデータがあるため男女の割合も当てはめた．

2 桁目のみだと正しく世代ごとに分けきれない．次に 2 桁目を復元後どの桁を復するかを決める．2 桁目を復元後，次に復元する桁で人数を絞れるのは 1 桁目である．2 桁目を復元後，1 桁目を復元し検索した場合の人数を表 3.10 に示す．図 3.3 に患者が 20 代だった

3.4 付与する検索情報のサイズ削減と検索方法

表 3.9 2 桁目のみを復元した場合の該当する男女別の年代人口

年齢	10 代未満	10 代	20 代	30 代	40 代	50 代	60 代	70 代	80 代	90 代	100 代
男性人数	55020.64	69815.71	78119.18	97088.94	103666.4	96713.6	94753.47	71714.5	31040.34	27267.28	55020.64
女性人数	53263.08	66528.38	74812.58	93812.13	101116.4	97339.17	102099.5	88736.05	50909.73	35785.58	53263.08

表 3.10 1 桁目を復元した場合の該当する年代別の人数

年齢	10 代未満	10 代	20 代	30 代	40 代	50 代	60 代	70 代	80 代	90 代	100 代
男性人数	30970.6	34917.14	36664.54	45351.07	56474.35	45541.72	53233.29	37701.2	19100.83	2645.28	54877.65
女性人数	29679.15	33259.46	34996.34	43802.03	54800.2	45279.57	55878.57	44945.94	31135.67	8424.384	52524

場合の年代の検索の流れを示す。しかし、100 歳代は違う桁を復元する方が絞れる。2 桁目の次に 1 桁目を復元すると 10 代未満、10 代、100 歳代以外は分けることが可能である。

100 歳代の検索

分けられない世代を特定するには 3 桁目を復元しなければならない。

2 桁目を復元し検索を行ったとき 100 歳代を検索すれば 2000 年以降の人のデータも該当する。そのため、2 桁目を復元しても 2000 年以降の人のデータと区別がつかない。100 歳代と比べ 2000 年以降の生まれの人口が多いので 100 歳代のデータを絞ることができない。

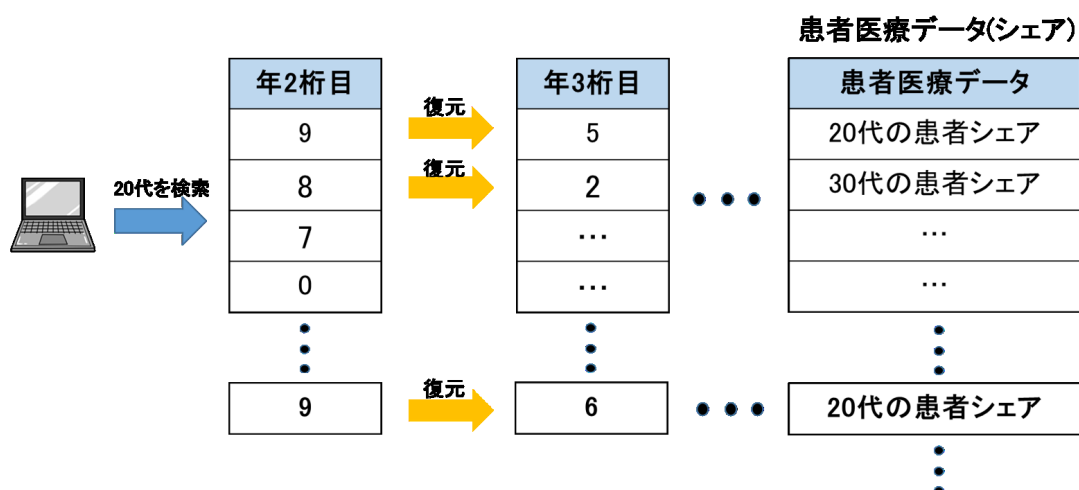


図 3.3 20 代の検索の流れ

3.4 付与する検索情報のサイズ削減と検索方法

表 3.11 1 桁目を復元した場合の該当する 100 歳代の人数

年齢	100 歳代
男性人数	1269.02
女性人数	1078.36

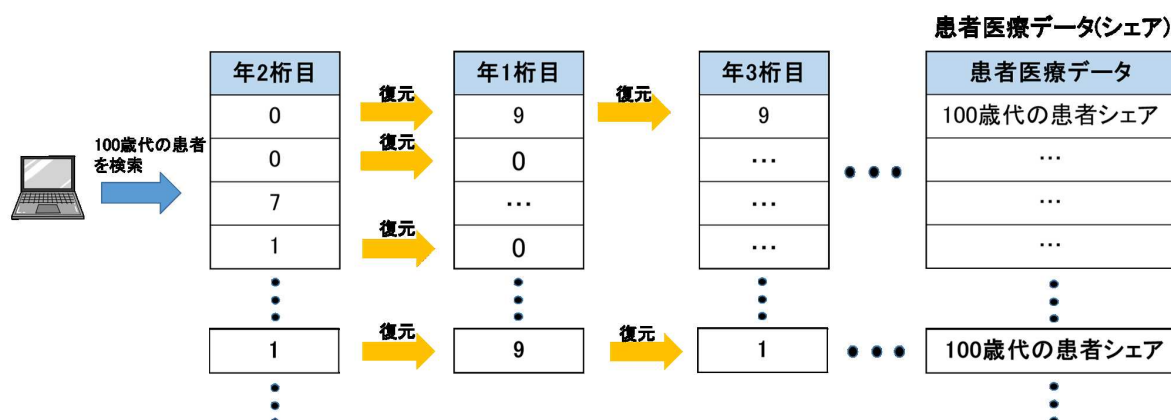


図 3.4 100 歳代の検索の流れ

そこで、100 歳代の検索では 2 桁目を復元した後、3 桁目を復元し検索する方がデータを絞ることができる。そうすることにより、3 桁目を復元した時点で 2000 年以降以降生まれのデータは含まれていない。よって 100 歳代を検索する場合は 2 桁目を復元した後、3 桁目を復元する。その後、1 桁目を復元することにより 100 歳代のデータを求めることができる。表 3.11 に 2 桁目を復元後、3 桁目を復元した場合の 100 歳代の該当する人数を示す。図 3.4 に患者が 100 歳代だった場合の年代の検索の流れを示す。

10 代未満、10 代の検索

10 代未満、10 代の検索は 2 桁目を復元後 1 桁目、3 桁目を復元する。10 代未満と 10 代は 2 桁目までだと 100 歳代の患者データが含まれるが、人口的に 100 歳代の人口が少ないためこの順番となる。

3.5 提案した検索方法の評価

表 3.12 60 個の姓を検索した結果

中央値 (人)	8068
最大人数 (人)	54689
最小人数 (人)	5701

表 3.13 年代を検索した場合の人数

年齢	10 代未満	10 代	20 代	30 代	40 代	50 代	60 代	70 代	80 代	90 代	100 代
男性人数	30939.6	34895.06	36664.54	45351.07	56474.35	45541.72	53233.29	37701.2	19100.83	2645.28	53.62
女性人数	29479.39	33119.63	34996.34	43802.03	54800.2	45279.57	55878.57	44945.94	31135.67	8424.384	339.59

3.5 提案した検索方法の評価

本研究では、部分復元可能な秘密分散によってバックアップしたデータの検索方法を提案した。今回は姓と西暦 (年代) のそれぞれで、どのぐらいの人数が絞れるのかを検証した。

60 種の姓データ [3] を用いて姓を検索した場合の中央値と最大数, 最小数を表 3.12 に示す。次に世代を検索したときの年代, 男女別の人数を 2016 年の日本の割合 [5] を利用し, 744849 人の人数に当てはめてた場合を表 3.13 に示す。姓, 年代だけでは, 該当する患者データを見つけることができないため, 他の情報と組み合わせて検索する必要がある。

本来ならば姓で絞ったデータから性別, 年代を検索した場合のデータ数を示さなければならないが, データが揃わなかったため別々で示した。また, 別の付与情報と合わせて検索した場合にどのぐらい絞れるのかを検証しなければならない。今回は姓と西暦の文字数の削減しか示していない。そのため, 他の付与情報の削減と検索方法を検討する必要がある。

3.6 考察

本研究では, バックアップを行うデータに検索用の情報を付与し検索を行う仕組みを提案した。しかし, 姓と西暦の文字数の削減, 検索方法しか述べれていない。そのため, 他の検索情報についても削減を行い演算量を減らす必要がある。性別では, 英 1 文字で 2 択であるため 0 か 1 で 1bit で表現できるようにすれば性別のサイズを削減できると考えられる。他

3.6 考察

の検索情報についても，データを集め削減する方法を検討しなけえばならない．

第 4 章

結論

4.1 まとめ

災害時の医療行為に必要な情報のみを提供できるようにするために，部分復元可能な秘密分散法が提案された [2]．災害時はバックアップされたデータから該当する患者の医療データを検索しなければならない．しかし，部分復元可能な秘密分散法を用いてバックアップを行った医療データを検索する仕組みはなく，名前などを部分復元を行い検索しなければならない．しかし，部分復元を行うには時間がかかってしまう．そこで，本研究では，部分復元可能な秘密分散を行った医療データを検索する仕組みの提案を行った．検索を行えるようにするためにバックアップを行うデータに検索情報を付与することを提案した．また，付与する検索情報を定め，姓と年代のサイズの削減方法と検索方法について述べた．研究では他の検索情報について示すことができなかった．また，患者データを特定するために必要な検索情報を調べ検索にかかる時間を示す必要がある．

4.2 今後の課題

部分復元可能な秘密分散法を用いてバックアップを行ったデータの検索方法を提案した．本研究では，優先度の高い姓と西暦のみのサイズの削減と検索方法しか述べることができなかった．姓については 60 種類の場合を示したため，姓についてのデータを集める必要がある，また，他の検索用情報についても検討する必要がある．患者データの特定を行うには付与した情報を組み合わせて検索する必要がある．そのため，付与した情報を組み合わせてど

4.2 今後の課題

のぐらい人数を絞れるのかを検証し，検索にかかる時間を調べる必要がある．他に，災害時に患者の情報を聞き出せない可能性もあるため他に付与する情報を検討する必要がある．

謝辞

本研究を行うにあたり、ご指導いただきました福本昌弘教授には、大変感謝致します。何度も期限を守らず申し訳ございませんでした。また、最後まで見捨てずご指導して頂きとても感謝致します。本研究の副査をしていただいた敷田幹文教授、鵜川始陽准教授にも大変感謝いたします。また、本研究で用いたデータの提供をいただいた高知県医療センター情報システム室北村和之氏にも感謝致します。

NOCの福富英次氏にも感謝致します。何度もお食事に連れて行って貰ったり、研究のアドバイスやミキサーの作業を手伝っていただいたりして本当にありがとうございます。

研究室の同期の佐藤文也氏、竹中壮磨氏にも感謝します。いつも文句ばかり言って申し訳ございませんでした。研究室のイベントがあるごとに徹夜で作業したのはいい思い出です。また、辞めそうになったときに止めていただきありがとうございます。

最後に4年間支えてくださった方々に感謝致します。

参考文献

- [1] 福本昌弘, “南海トラフ巨大地震に対する医療情報の保全のための高知県での取り組み”,
Mercato, 東北情報通信懇談会, Vol.89, pp.18-20, 2014.
- [2] 田中麻実, 福富英次, 福本昌弘, “秘密分散バックアップした医療データの部分復元”,
信学技報 IA2015-74, p.31-36, Dec. 2015.
- [3] 田中康仁, “日本人の姓と名の分布”, http://www.orsj.or.jp/~archive/pdf/bul/Vol.23_06_357.pdf, 日本オペレーションズ・リサーチ学会, pp.381, Jun. 1978.
- [4] 厚生労働省, “主な年齢の平均寿命”, <http://www.mhlw.go.jp/toukei/saikin/hw/life/life15/dl/life15-02.pdf>, 2018 年 2 月 27 日閲覧.
- [5] 政府統計の総合窓口 (e-Stat), “年齢 (各歳), 男女別人口及び人口性比-総人口, 日本人人口 (平成 28 年 10 月 1 日現在)”, 2018 年 2 月 27 日閲覧.