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

2018年2月20日

1180362　沼　尚樹

情報学群

ネットワーク信号処理研究室

背景

東日本大震災の津波により医療データが流出し、被災地での医療行為に支障

↓

遠隔地にバックアップを行う取り組み

- 災害時にバックアップした医療データを活用

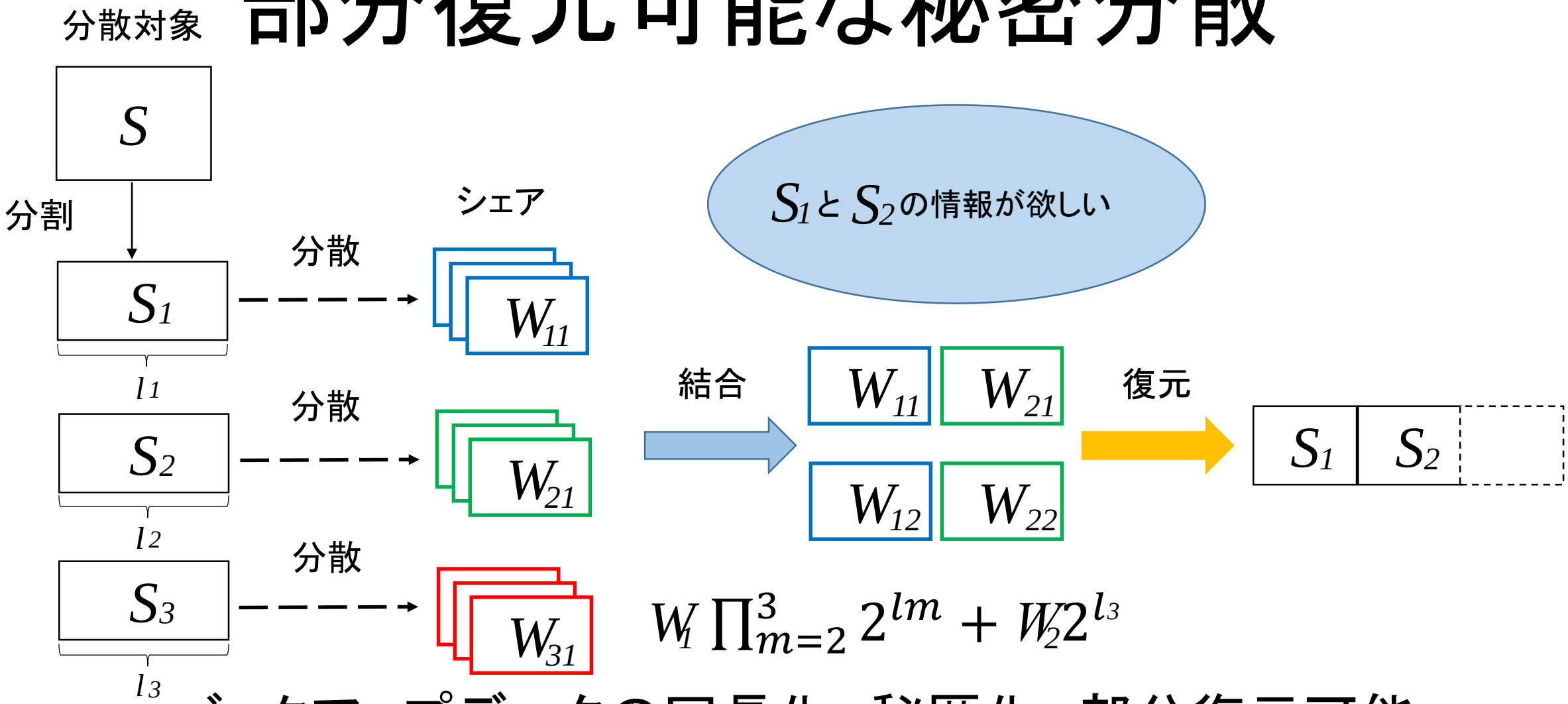
↓

秘密分散法を利用したバックアップ

秘密分散バックアップした医療データの部分復元(田中2015)

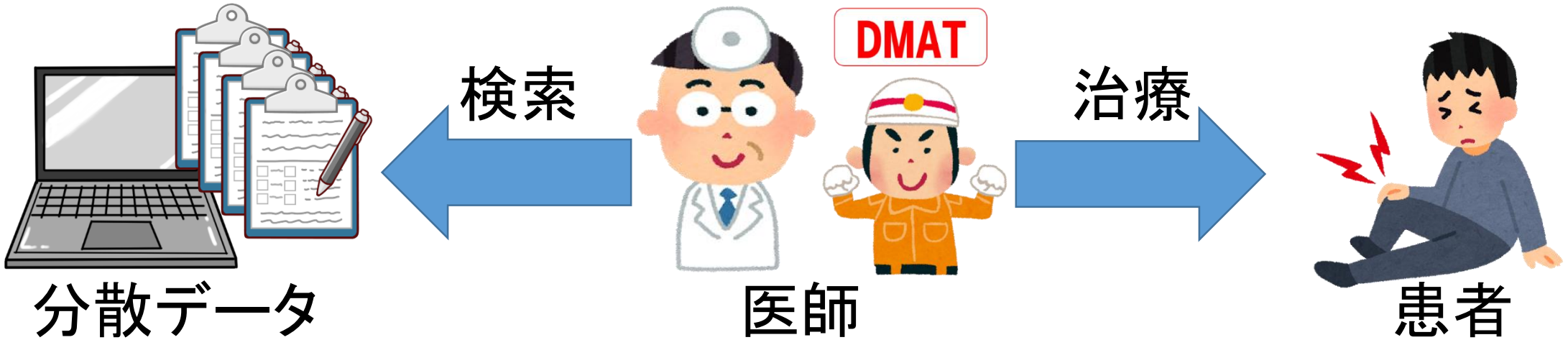
- バックアップデータの冗長化、秘匿化、部分復元可能

部分復元可能な秘密分散



バックアップデータの冗長化、秘匿化、部分復元可能

災害時のバックアップデータの活用



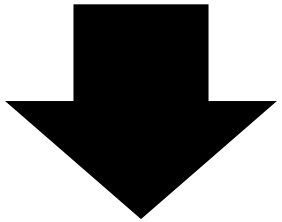
部分復元可能な秘密分散には検索の仕組みがない

名前部分などを部分復元して検索

- 部分復元を行うには時間がかかる

研究の目的

災害時、検索のために部分復元を行うには復元処理の演算に時間がかかる



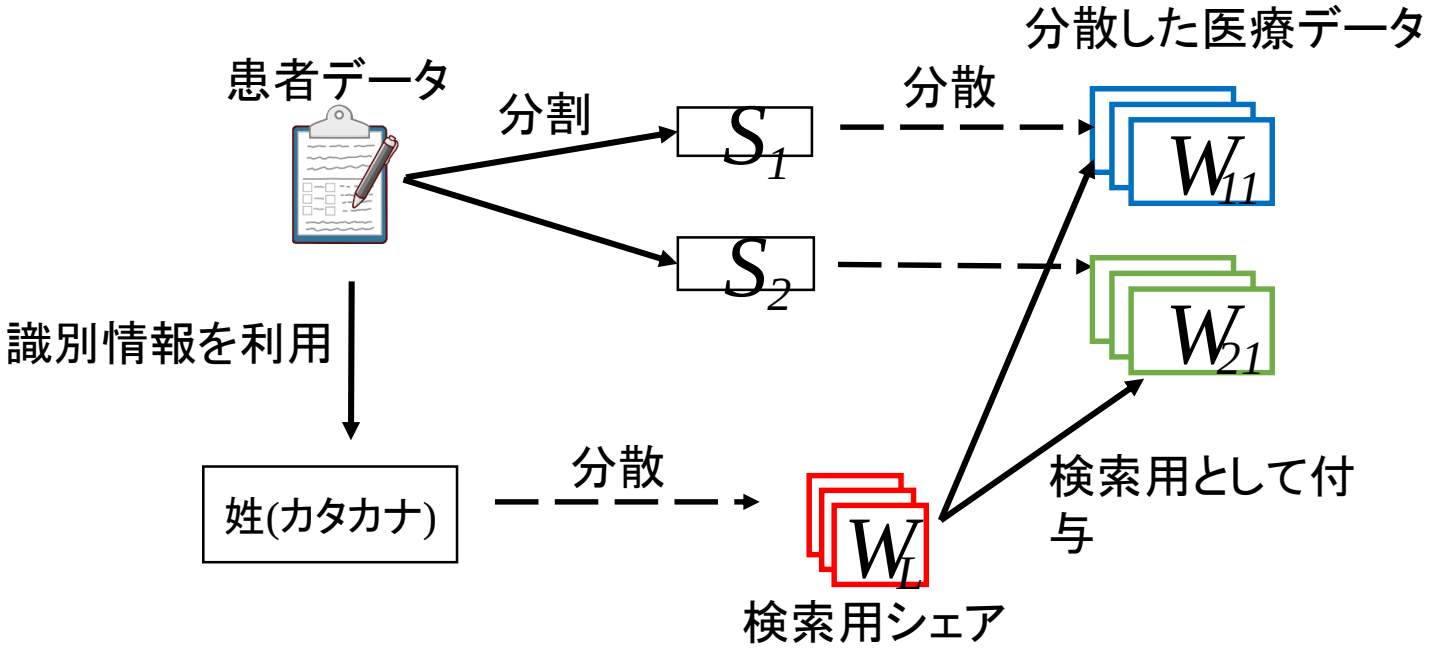
バックアップを行うデータに検索用の患者情報を付与

検索のための部分復元を行うデータ数を減らす

検索用の患者情報付与

医療データから患者識別情報を利用

医療データのシェアに検索用シェアを付与



検索では検索用シェアだけを復元

医療データの情報

標準化された医療データを利用

標準化されたデータからの患者識別できる情報

識別情報	患者姓名	性別	住所	生年月日
表記	漢字とカタカナ	全角英1文字	漢字	数値

4つの情報を付与して患者を特定

付与する患者情報

災害時における優先度の設定

	患者姓名	性別	住所	生年月日
入力条件	カタカナ	全角英1文字	漢字	数値
検索の手間	簡単 身内以外にも知っている可能性が高い	簡単 身体の情報でわかるデータが絞れない	困難 他人が知らない可能性が高い	簡単 他人が知らない可能性がある
優先度	1	2	4	3

患者姓を高速に検索する方法を提案

姓情報の削減

東邦生命の被保険者と契約者の頻度上位60位までの姓を利用(データ数744849件)

文字数によって特定できる姓の数

1文字目	2文字目	3文字目
12個	41個	60個

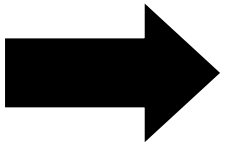
姓の先頭3文字で1つの姓を特定可能

姓の先頭3文字を検索用のシェアとして付与

検索用シェアの作成

3文字を付与する場合の検索方法

- 1文字づつ分散、復元
- 3文字全てを分散、復元



演算量の見積もりを比較

演算量　1文字づつのシェア n^3　　3文字づつのシェア $(3n)^3$

サトウ(54689件)を検索した場合

	1文字づつのシェア	**3文字シェア**
復元回数	917314回	744849回
演算量	$917314 \times n^3$	$744849 \times (3n)^3$

1文字づつ処理する方が高速

検索用シェアを用いた検索方法

- 患者姓を付与する場合
  - 姓の先頭3文字を付与
  - 1文字づつ分散、復元を行い検索

- 他の識別情報について
  - 高速に検索する方法を考える

まとめ

- 分部復元可能な秘密分散バックアップした医療データの検索方法を提案
  - 患者識別情報を検索用シェアとしバックアップデータに付与
  - 患者姓の検索方法
- 今後の課題
  - 実際のデータを使って検証
  - 他の識別情報の検索方法
  - 他にどのような情報を付与すべきかの検討