DEIM Forum 2019 I7-6

災害を想定した被災情報のプライマリーとバックアップの地域内配置手法

田井 裕次郎† Le Hieu Hanh†† 横田 治夫††

† 東京工業大学情報工学科 〒 152–8550 東京都目黒区大岡山 2-12-1
†† 東京工業大学 情報理工学院 〒 152–8550 東京都目黒区大岡山 2-12-1
E-mail: †{tai,hanhlh}@de.cs.titech.ac.jp, ††yokota@cs.titech.ac.jp

あらまし 災害時には，避難に関する情報やデータの重要性が高い．しかし，避難に関する情報やデータを持つサーバーは災害時に失われてしまう可能性がある．バックアップデータを用意し，プライマリーデータが失われた際にもデータの復元が可能である状況を整えておくことが求められる．そこで，本研究ではクラスタリングや線形計画ソルバーを用いた，より良いデータ配置を検討する．具体的には，データセットから抜き出した避難所の位置情報を用いて，周囲の避難所のデータを集めるようにして幾つかの避難所にデータベースを設置することを想定する．データベースのバックアップは 1 つ取るものとして，データ損失を防ぐ為のデータ配置を検討する．

キーワード 避難場所情報, 災害, 位置情報, SIBM

1 はじめに

世界中には台風や地震などの災害による損失や人々の生活への影響が多く見られる．災害発生時，避難することは重要な作業である．その際，避難場所情報を初めとする避難に関する情報の管理もまた重要である．災害で及ぼされる影響は人々の生活に直接的なものに留まらず, 蓄えられたデータにも及ぶ. データが蓄えられたサーバーもまた被害を受け破壊される．

特に避難場所情報に関しては，避難場所情報管理システムを評価するための避難場所ベンチマーク SIBM(Shelter Information Benchmark) [1] の開発等からも重要視されていることがわかり，そのデータの保存性についてもまた重要視しなければならないものである．そこで，現実の問題としてバックアップデータをどこに置くのか，災害から守ることを考慮して外部のサーバー，それもかなり遠方に置く必要があると考えられる．

しかし，災害ではネットワーク障害等もあることから遠隔のバックアップデータだけでは，早急にデータが必要となる災害時に対応することができない場合がある．そこで求められるのは，同時に被災することのない安全な地域で且つバックアップデータが必要となる場所に近い位置にバックパップデータを置くということである．このバックアップデータにより，サーバーごと被災場所へと運び込むことで，ネットワークが使えない際に早急なデータの復元が可能となる．

実際のバックアップデータの管理例として，過去に大きな問題となった東日本大震災時 (平成 23 年 3 月) には,「地震発災時における地方公共団体の業務継続の手引きとその解説（第 1 版）」[2](平成 22 年 4 月) があり，バックアップの重要性も示していたものの，当時の地方公共団体での業務継続計画の策定率は低く，ネットワークが使用不可な時にデータの復元が出来ない地域があった．

しかし，内閣府はその後，各地方自治体に対して復元可能場所にバックアップデータを保管するよう「市町村のための業務継続計画作成ガイド」[3]（平成 27 年 5 月) を提示しており，管理の重要性を説いている．そして，最新版として「大規模災害発生時における地方公共団体の業務継続の手引き」(平成 28 年 2 月) [4] を提示しており，そこでは，全庁的なバックアップ体制の構築が重要とは書かれており都道府県内での市町村間での連携の重要性や，庁内でのバックアップシステムの整備を図る旨，バックアップ業者との契約の検討もかかれている．

しかし，これらの方針は手引きであり詳細は各地方自治体に委ねられている．実際のバックアップの取り方，効果的な取り方はここでは考えられていない．

そのようなことになっている問題として都道府県，市町村ごとにデータベースの配置が異なり正確な指示を出せないこと，地域によって注意するべき災害が異なってくることが原因に含まれると考えられる．

そこで，地方自治体に収まるような範囲においてネットワークが使用できない状況で早急なデータの復元を行うことを想定する．地方自治体に収まる範囲は，地理的に障害が少なくサーバーの運びこみが容易であると考えられる為，様々な自治体で同様の想定が出来ると考えられる．

その範囲において，様々な災害に対応できるデータベースの配置と，データベースにあるデータのバックアップデータの配置について最適なデータ配置を求めることを考える．プライマリーデータがあるデータベースの配置は，所属する避難所の数に大きな差がでないよう考慮したクラスタリングを用いて行う．バックアップデータの配置は，プライマリーデータとの距離が開くよう考慮した目的関数と制約条件を利用して行う．そして，データ配置の適性評価は，プライマリーデータ・バックアップデータ間の距離の平均と復元率を用いて行う．

本論文では，第 2 節では，本研究で用いられるクラスタリング，線形計画ソルバー，SIBM について説明し，第 3 節で， 関連研究を紹介する．第 4 節では，提案手法を提示し，第 5 節で，実験の手順，内容，結果とその考察をする．最後に，第 6 節でまとめと今後の課題を述べる．

2 背景知識

2.1 クラスタリング

クラスタリングとは，分け方を事前に指摘することなくデータを部分集合であるクラスタに切り分ける教師なし学習のことである．この時，それぞれの部分集合は，共通の特徴を持つようにすることが多い．本論文では，クラスタの数をデータ毎に一定に保った状態でクラスタリングのアルゴリズムを利用したいため，k means clustering を用いる．

2.1.1 k means clustering

k means clustering とは, 非階層型クラスタリングのアルゴリズムのことである．クラスタの平均を用い，与えられたクラスタ数 k 個に分類する．k-平均法（k-means），c-平均法（c-means）とも呼ばれる [5].

2.1.2 k means clustering のアルゴリズム

（1） データの数を n，クラスタの数を k とする

（2） 各データ $x_i(i = 1, ..., n)$ に対してランダムにクラスタを割り振る

（3） 割り振ったデータをもとに各クラスタの中心 $V_j(j = 1, ..., k)$ を計算する

（4） 各 x_i と各 V_j との距離を求め，x_i を最も近い中心のクラスタに割り当て直す

（5） 上記の処理で全ての x_i のクラスタの割り当てが変化しなかった場合，あるいは変化量が事前に設定した一定の閾値を下回った場合に，収束したと判断して処理を終了する．そうでない場合は新しく割り振られたクラスタから V_j を再計算して上記の処理を繰り返す.

2.1.3 k means clustering の利用

結果は，最初のクラスタのランダムな割り振りに大きく依存するため，1 回の結果で最良のものが得られるとは限らない．そのため，何度か繰り返して行って最良の結果を選択する手法を利用する．

実験では，データを避難所の位置情報，クラスタの数を設置するデータベースの数として計算を行う．

2.2 線形計画ソルバー

線形計画法ソルバーとは，複数の変数を含む幾つかの 1 次不等式および 1 次等式を制約条件とし，制約条件の範囲内である 1 次式を最大化または最小化する解の組み合わせを求めるものである．本論文では，プログラムの内部で線形計画法ソルバーを用いるため，API 経由で呼び出して利用することが出来る lp_solve を利用する．

2.2.1 lp_solve

lp_solve とは，線形計画問題や整数計画問題，混合整数計画問題を解くことができるフリーのソフトウエアである．lp_solve では，複数の変数の値を変化させながら変数の相互関係を判断し，最適な値を算出することができる．ある変数に特定の制約条件をつけたり，あるいは特定の変数が最大値・最小値を得るために他の変数の値を変化させたりすることもできる．lp_solve を用いれば，連立方程式の解や，複数の項目が連動する事業計画の試算といった複雑な演算ができる [6] [7].

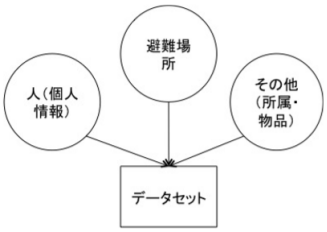

図 1 データセットの構成

2.2.2 lp_solve の特徴

- 改訂単体法と分枝限定法を実装している
- 基本的に非線形計画問題を解くことはできない
- 解ける問題サイズに制限はない

2.2.3 lp_solve の利用

実験では，各データベースについてすべてのパターンのバックアップデータの置き方を想定するため，各データベース間の距離を係数としプライマリーデータとバックアップデータのつながりがあるかどうかの存在を表す変数と併せた，項数が (避難所の数)^2 となる目的関数と，それに対する制約条件が lp_solve に使われる．

2.3 SIBM

SIBM は避難場所情報と避難作業の関係者情報を生成可能である [1]．本実験は，この SIBM を用いて情報を生成し，そのデータから得られる避難所の位置情報をもとに進められる．

SIBM は以下の 2 つの目標に基づいて設計されている：

（1） 避難する際，実際に発生する事情を再現できるデータを生成すること．

（2） 生成した情報を問い合わせするクエリセットを用意すること．

それ以外にも，研究分野や使用目的に応じて，様々な場面で活用できることが考えられる． SIBM では，避難場所情報を現実に近い情報源を生成する目的を持ち，以下の 2 つの特徴をもつ：

- SIBM では，指定された避難場所数によるデータを生成することが可能であり，生成した情報が実際の避難場所情報や被災者間の関係などを再現できる．
- SIBM が生成した情報に対する問い合わせクエリセットを用意し，RDF データ構造に対するクエリや， 実際の情報を問い合わせするケースに対するクエリを作成する [8]

SIBM が用意したクエリセットでは， 3 つの情報：避難場所情報， 関係者の情報とその関連情報（関係者の所属情報や避難場所の 物品情報など）を対象とする．その 3 つの要素からなるデータ セットは図 1 のように構成される

避難場所情報が避難場所の詳細情報を表すものである．その構造は以下のようになる:

- 地上の座標，住所
- 収容人数

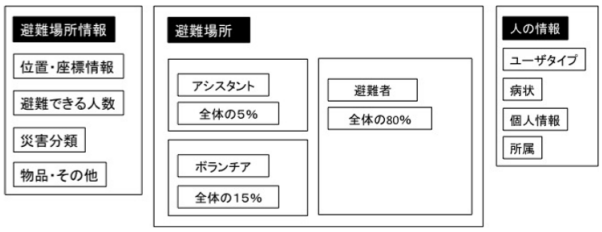

図 2　避難場所全体の構成

図 3　都道府県単位の避難場所構成

- 災害分類（地震・津波・火山・洪水・その他・未確定）
- 施設の種類

関係者には 3 つのタイプがある：アシスタント，避難者，ボランティア．関係者全体に対する基本個人情報が生成され，タイプ別に対する所属情報や怪我状態などの情報が追加される．例えば，一般個人情報には，ユーザの氏名・年齢・住所・性別などの情報が持つ．ユーザタイプは避難者であるならば，怪我情報が追加され，また，ユーザの年齢によって会社や学校などに所属する情報が追加される．

避難場所情報全体には，避難場所の詳細情報と，その避難所で避難する・作業する人たちの情報が含まれる．関係者の分配割合が適切に生成される（図 2）．

また，生成した避難場所情報を都道府県単位でまとめる（図 3）．本稿では，平成 24 年度国土交通省による公開された全国の避難場所情報（役 12.5 万箇所）を使用している．

3　関 連 研 究

2011 年に発生した東日本大震災では，多くの被害がもたらされた．その際には，ネットワークが接続できなくなり，情報提供システムも機能を止め，安否の確認や負傷者の管理もできなくなった．T.Nakamura らは，Japan National Project に参加し迅速で確実なデータ復旧が行える技術の提案をした [9]．

復旧の方法として，2 つの状況を想定し，提案された．1 つは，ネットワークが使用できる状況でデータ復旧する方法である．もう 1 つは，東日本大震災において広範囲でネットワークの復旧までに 1 か月がかかったために想定された，ネットワークを使わない復旧方法である．

ネットワークを使用する方法では，近隣の地域のデータと併せて遠隔に置いたバックアップから復元する．

ネットワークを使用しない方法では，近くのサーバーにバックアップデータを置き，サーバーが失われた場所にサーバーごと運び込むことでデータの復旧を図る．そのため，バックアップデータが遠すぎず，プライマリーデータと同時に失われないような安全な位置に置くことを考える．ハザードマップやデータ量を参考にし，各地域において被災確率と被災した際の損失の大きさから，その地点にバックアップを置く危険度を示すリスク関数を用意する．このリスク関数を用いることで，安全なデータ配置ができるようにしている．

この研究では，1 つの災害に対して対応する際には早い時間での復旧を可能にし，高いリカバリー率を得る．その一方で，ハザードマップが対応することができる状況のみを想定しており，ハザードマップに存在しない災害が起きた場合には対応できない．例として，千葉県千葉市や北海道札幌市などでは，地震の際の建物崩壊の危険性を纏めたハザードマップが存在することに対して，千葉県長生郡長生村や北海道登別市などでは，存在しない．

ハザードマップは自治体ごとに作られている種類や数が異なるため，この研究の手法を用いることが出来ない状況があるという欠点もある．また，同時に複数の災害が起きた場合を想定すると，リスク関数の値の設定が非常に難しいという問題点も挙げられる．

4　提 案 手 法

そこで本論文では，ネットワークを使用しない方法での災害の種類や複合災害等に左右されないデータ復元について考える．データの損失を防ぐためバックアップデータを安全な位置に置く方法を考え，使用不能のデータベースサーバーが現れてしまった際に，バックアップデータを用いることで，データを補うことが出来る方法を検討する．

4.1　プライマリーデータの配置

準備で用意した避難所の座標を基に，データベースの配置を決める．この時，データベース毎にデータ量に差がつかないように考える．本論文では，避難所毎にデータ数を一律としているため，クラスタ毎のデータ量を平等に保つには避難所数が偏らないようにする必要がある．k means clustering algorithm を利用して行うが，単純な k means clustering algorithm では，クラスタに属する避難所数に偏りが出てしまう為，配置の決定は以下のような手順を取った．

Algorithm1

Input: クラスタ数　Output: データベースの座標

step1 各避難所をランダムでクラスタに所属させる

step2 クラスタ毎に，クラスタに属する避難所の座標の平均を取る

step3 平均座標に最も近い避難所にデータベースを設定する

step4 各避難所を，データベースが最も近い位置にあるクラスタに改めて所属させる (図 4)

step5 クラスタ毎に，クラスタに属する避難所の座標の平均を取る

step6 平均座標に最も近い避難所にデータベースを設定する (図 5)

step7 step4，step5，step6 の動作をランダムに数回行う

step8 クラスタ毎に属する避難所の数に偏りが出ないよう，

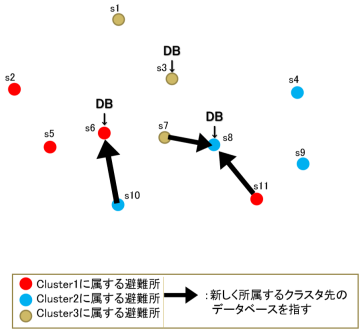

図 4　クラスタ再設定の様子 (step4)

数の多いクラスタに属する避難所を，データベースから遠い避難所から順に空きのあるクラスタの中で最も近いクラスタに所属させる (図 6)

step9 クラスタ毎に，クラスタに属する避難所の座標の平均を取る

step10 平均に最も近い避難所にデータベースがあると設定する (図 7)

step11 step4 に戻る (データベースの座標が変わらなかったら正常終了する．また，クラスタの変更がなくなるか、同じ変更が繰り返されるようになったら終了とする．

図 4，図 5 では，クラスタリングのアルゴリズムの中で避難所がクラスタを変える際の様子を表している．各点は，避難所を表しており，赤，青，茶の色はそれぞれのクラスタであることを示しており，それぞれ Cluster 1，Cluster 2，Cluster 3 としている．DB と示されているものは，その位置にデータベースが設置されていることを表している．

図 4 では，自身の所属するクラスタのデータベースよりも近くに別クラスタのデータベースが存在する避難所のデータを，近くにあるデータベースへと移す様子を表しており，s7，s11 は s8 にあるデータベースに，s10 は s6 にあるデータベースにデータを移す．

図 5 では，それに伴い避難所のクラスタを変更し，各クラスタでデータベースを再設定した後の様子を表している．s7，s11 は Cluster2 に所属を変更し，s10 は Cluster1 に所属を変更している．データベースの再設定では，Cluster 1 のデータベースの位置が s6 から s5 へと変わっている．

図 6 では，所属避難所が多い Cluster2 の中で最もデータベースから遠い避難所を別のクラスタに移す様子を表しており，Cluster2 のデータベースがある s8 から最も遠い s4 が，Cluster2 以外で最も近いデータベースがある s3 にデータを移す．

図 7 では，それに伴い s4 の避難所のクラスタを変更し，データベースの再設定をした後の様子を表している．

4.2　バックアップデータの配置

データベースを設置した後，データベースの座標を使ってプライマリーデータとバックアップデータの理想的な配置を考えていく．プライマリーデータとバックアップデータの間の距離が遠いほど同時に失われることがないという考えのもと，残りのバックアップデータをより効果的な場所に置くために次のよ

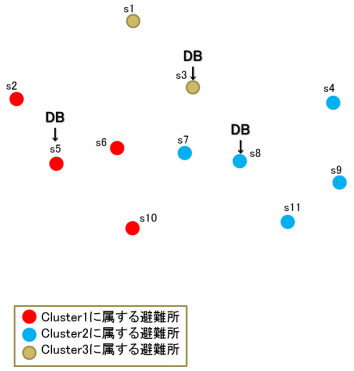

図 5　クラスタ再設定後，DB 再設置後の様子 (step6)

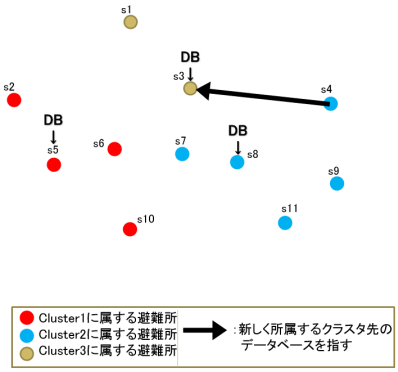

図 6　避難所の数の調整後，DB 再設置後の様子 (step8)

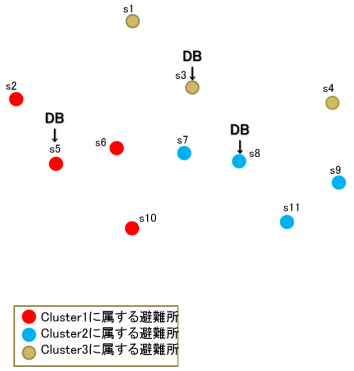

図 7　避難所の数の調整 (step10)

うに考える．

Algorithm2

Input: データベースの座標と数　Output: バックアップの配置

step1 データベースの数を M とする

step2 データベース i とデータベース j の間の距離を $l_{i,j}$ とする

step3 データベース i にあるプライマリーデータと，データベース j にあるそのデータのバックアップとの繋がりを表す存在関数を $x_{i,j}(=0,1)$ とする

step4 プライマリーデータとそのバックアップデータを同じ場所に置かないようにするため $x_{u,v}=0$ $(u=v)$

step5 バックアップデータは既に最も近い位置にあるデータベースには置いてあると考えるので $l_{u,v}=0$ $(v|\min l_{u,v})$

step6 1 つのデータについて幾つバックアップデータを取るかを決める (ここでは P とする)．データベース i にあるデータに関して $\sum_{j=1}^{M} x_{i,j} = P$

step7 1 つのデータベースに置くことの出来るバックアップの制限を決める (ここでは Q とする)．制限を表す関数を $\sum_{i=1}^{M} x_{i,j} \leqq Q$ とする

step8 これらの条件のもと，プライマリーデータとバックアップデータの間の距離の総和 $\sum_{i=1}^{i=M} \sum_{j=1}^{j=M} x_{i,j} l_{i,j}$ が最大となるようなバックアップの配置を考える

以上を纏めると，目的関数と制約条件は以下のようになる．

目的関数：

$$Maximize\ L = \sum_{i=1}^{i=M} \sum_{j=1}^{j=M} x_{i,j} l_{i,j}$$

制約条件：

$$x_{u,v} = \{0,1\} (1 \leqq u, v \leqq M)$$

$$x_{u,v} = 0\ (u = v)$$

$$l_{u,v} = 0\ (v | \min l_{u,v})$$

$$\sum_{j=1}^{M} x_{i,j} = P\ (i = 1, \ldots, M)$$

$$\sum_{i=1}^{M} x_{i,j} \leqq Q$$

以上の式から，線形計画ソルバーよりデータの復元率が向上するバックアップの配置を行う．このバックアップは，4.2 データベースの配置で設定したデータベースに置く．

4.3 データ配置の適性評価

以上のアプローチによるデータ配置の適性について以下の 2 つで評価する．

- プライマリーデータの配置された避難所とバックアップデータ間の距離の配置された避難所間の距離の平均
- 復元率

プライマリーデータの配置された避難所とバックアップデータ間の距離の配置された避難所間の距離の平均は，大きい程良い評価とする．これは，距離が離れている方が同時に災害を受けることが少なくなると考えられる為である．復元率は，$\frac{\text{復元可能なデータベース}}{\text{使用不可になったデータベース}} \times 100(\%)$ で表し数値が高い方が良い結果となる．図 8 では設定した被害範囲内のデータベースは使用不可となり，範囲外のデータベースは使用可能であることを示している．使用不可となったデータの中で範囲外にバックアップデータがあるデータが復元可能となっている．

5 評価実験

提案手法で求めたデータ配置と，ランダムでバックアップデータの配置場所を決めるデータ配置と，ランダムでプライマリーデータの配置場所を決めるデータ配置の性能を比較する実験を行う．それぞれ，データベース・バックアップデータ間の距離の平均と復元率を計算し，比較した．バックアップの配置を決める式の計算は lp_solve を利用して行う．

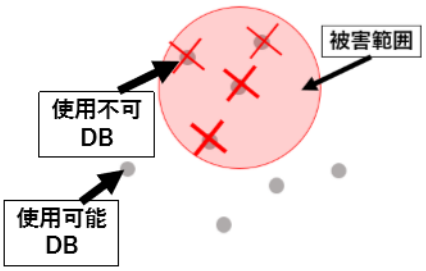

図 8　被害範囲とデータベース

```
Kanagawa_14205_0
35.3981029991,139.460302
Kanagawa_14205_1
35.401436,139.467523
Kanagawa_14205_2
35.396076,139.472273
Kanagawa_14205_3
35.389383,139.46544
Kanagawa_14205_4
35.386856,139.461468
Kanagawa_14205_5
35.38441T,139.466524
Kanagawa_14205_6
35.3900209991,139.47794
```

図 9　各避難所毎の座標

5.1 実験手順

5.1.1 座標表示

既存の SIBM では避難所の位置座標情報は，座標コード (P20_100473 など) で表記されており，緯度経度の表記ではない．本手法では，避難所の位置座標から避難所間の距離を利用するために緯度経度の表記が必要となる．そのため，以下のように SIBM のプログラムを変更した．

- 緯度経度の出力を可能にした
- 各避難所毎の座標をまとめて出力できるようにした

図 9 は SIBM で作られるデータから，本実験で必要な情報 (データセットの都道府県名，行政区画番号，避難所番号とその避難所の座標) を纏めたものを出力したデータの一部である．

避難所の位置情報を実験として正確に利用するために，計算のための準備として以下を行った．

- 緯度経度表記では，球状のマップを想定しなければならず実験時の距離計算をする際に不便であるため，メートル法に則り平面化を行う
- 緯度経度を元に全避難所の中心を原点と取り，各避難所の位置を距離 (単位:km) で計算し，x y 座標を用いて表した
- このとき，x 軸 y 軸は緯線経線とそれぞれ平行且つ，上記の中心を通るように定めた

5.1.2 プライマリーデータの配置

（1） クラスタの数を，⌊(避難所の数) ÷ 15⌋ とする

（2） 求めたクラスタ数で Algorithm1 を行う

（3） step7 での動作は 1-3 回とする

表 1　提案手法と比較する手法の差異

	提案手法	比較手法 1	比較手法 2
プライマリーの配置	Algorithm 1	Random (Algorithm 1 step4-11 を省いたもの)	Algorithm 1
バックアップの配置	Algorithm 2	Algorithm 2	Random (Algorithm 2 を省きランダムにバックアップを振り分けたもの [1])

Algorithm1 step1，step8 では，クラスタに属する避難所の数は，$\lfloor \frac{(\text{避難所の数})}{(\text{クラスタの数})} \rfloor + 1$ 以下になるように調整する．

5.1.3　バックアップデータの配置

バックアップデータは，1 つのプライマリーデータから 1 つ，バックアップデータは，1 つのデータベースに 3 つまでとする．求められたプライマリーデータの配置を用いて，Algorithm2 を行う．

5.1.4　データ配置の適性評価

与えられたデータ配置から，プライマリーデータの配置された避難所とバックアップデータ間の距離の配置された避難所の距離の平均を求める．

復元率は，被害範囲の半径を 0.5km, 1.0km, 1.5km, ...,10km とし，算出する．1 つの半径に対して，被害範囲の中心を全てのデータベースに順に設定していき，それぞれで算出した復元率の平均をとる．その値を，その半径の被害範囲での復元率として扱う．

また，本実験では，上記の手順を 1000 回行い，その平均を取っている．その結果を，実験結果として利用する．

5.2　比 較 対 象

Algorithm 1 と，Algorithm 2 を変えたものを用意する．それぞれの手法を比較手法 1，比較手法 2 として提案手法との違いを纏めたものが下の表になる (表 1)．

SIBM で生成した，実験で使用するデータを表 2 に纏めた．data1,data2,data3 の避難所の配置は以下のようになっている．(図 10，図 11，図 12)

data1，data2，data3 は SIBM で生成されるデータであり，群馬県，神奈川県，和歌山県の避難所情報を半径 10km の円で出したものである．人口が普通の県，多い県，少ない県の例としてそれぞれの県を選んでいる．

data1_5km，data2_5km，data3_5km とは，生成したデータを中心から約半径 5km の円で区切ったものであり，data1_10km，data2_10km，data3_10km は図 10，図 11，図 12 の範囲内の全ての避難所を含むものである．また，避難所の持つデータのデータ量は一律で考える．

これを用いて提案手法の評価をした．

5.3　結　　果

5.3.1　プライマリーデータの配置

それぞれの手法で実際に配置されたプライマリーデータの

1：プライマリーデータを持つデータベースと最近傍のデータベースを除いたデータベースのいずれかに振り分けられる

2：クラスタの数が少ない場合は実験は行わない

表 2　実験で使用するデータセット

	データセットの名前					
データセット	data1_5km	data1_10km	data2_5km	data2_10km	data3_5km	data3_10km
データ	data1 の中心から約 5km 以内のデータを利用	data1 の中心から約 10km 以内のデータを利用	data2 の中心から約 5km 以内のデータを利用	data2 の中心から約 10km 以内のデータを利用	data3 の中心から約 5km 以内のデータを利用	data3 の中心から約 10km 以内のデータを利用
避難所の数	75	216	81	418	19	65
クラスタの数	5	14	5	27	1[2]	4

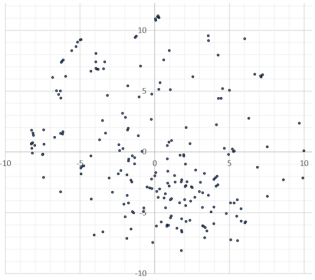

図 10　data1(群馬県) の避難所の配置

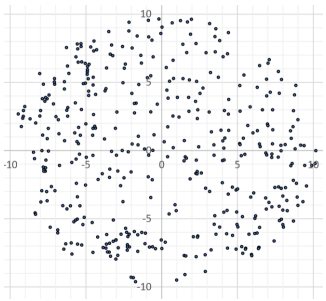

図 11　data2(神奈川県) の避難所の配置

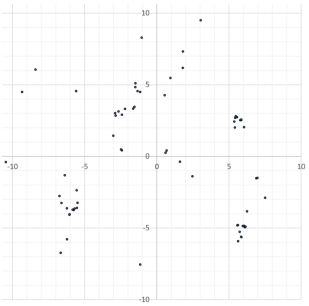

図 12　data3(和歌山県) の避難所の配置

データ配置の様子を以下に示す．データセットは data1_10km を用いた (図 13，図 14)．黒点が避難所の位置を示しており，色のついた点が配置されたデータベース (プライマリーデータ) を示している．それぞれ 3 度の実験で配置されたデータベースの配置図を重ねたものであり，赤，青，緑の大きな点は，それぞれ 1 度目，2 度目，3 度目の実験で配置されたデータベースとなる．図 13 は Algorithm1 を 3 度実行して得られた結果を重ねたもの，図 14 は Random を 3 度実行して得られた結果を重ねたものになっている．

この結果から，提案手法と比較手法 2 で用いている Algorithm 1 によって，比較手法 1 と比べてデータベースが分散して配置されることがわかる．

5.3.2　バックアップデータの配置

それぞれの手法について，プライマリーデータ・バックアップデータ間の距離の平均を表に纏めた (表 3)．値は全て 1000

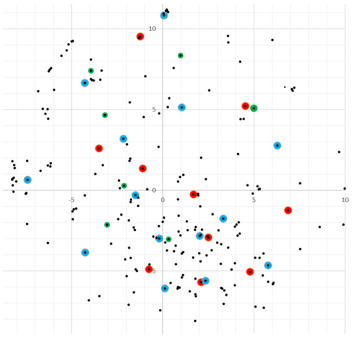

図 13　Algorithm1 によるデータベースの配置

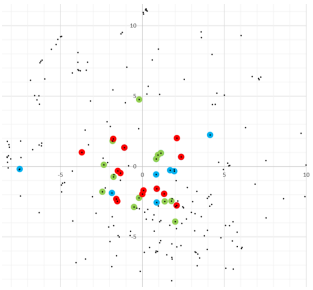

図 14　Random のデータベースの配置

表 3　プライマリー・バックアップ間の距離の平均

	データセットの名前					
データセット	data1_5km	data1_10km	data2_5km	data2_10km	data3_5km	data3_10km
クラスタの数	5	14	5	27	1	4
提案手法 (km)	6.36	13.26	6.02	15.40	-	11.90
比較手法 1 (km)	2.88	5.74	2.60	7.53	-	4.95
比較手法 2 (km)	4.61	8.11	4.47	9.49	-	9.27

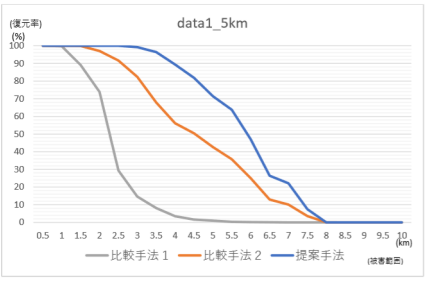

図 15　data1_5km を使用した際の被害範囲毎の復元率

回ずつ行った実験から得られた数値の平均値を取っている．

全てのデータセットで，数値が提案手法＞比較手法 2 ＞比較手法 1 となっている．プライマリーデータ配置の方法が異なる提案手法と比較手法 1 で大きな差がつき，バックアップデータの配置方法の異なる提案手法と比較手法 2 の間の差よりも明らかに大きくなった．

5. 3. 3　復元率の算出

復元率を求めたものをグラフにした．(図 15-図 19)

グラフの横軸は被害半径を 0.5km 刻みで表記しており，縦軸は復元率を 10%刻みで表記している．

5. 4　考　　察

グラフ (図 15-図 19) から，全ての点で提案手法が最も優れていることがわかる．また，比較手法 2 と比較手法 1 では比較手法 2 の方が優れており，この結果からバックアップの配置を工夫して復元率を高めようとしても元となるデータベースの配置が工夫されていないと効果が薄いことが確かめられた．このことは，表 3 に示されているプライマリーデータ・バックアップデータ間の距離の差からも確かめることができる．

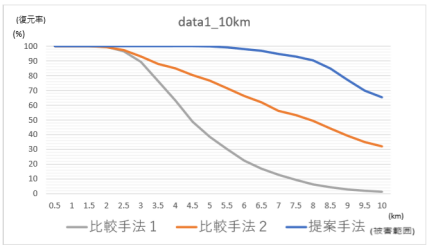

図 16　data1_10km を使用した際の被害範囲毎の復元率

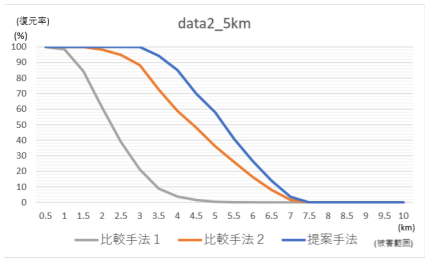

図 17　data2_5km を使用した際の被害範囲毎の復元率

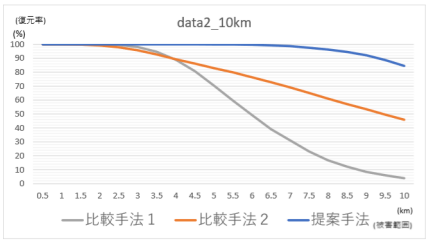

図 18　data2_10km を使用した際の被害範囲毎の復元率

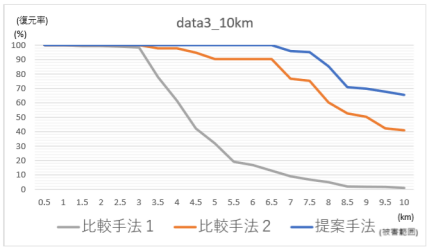

図 19　data3_10km を使用した際の被害範囲毎の復元率

それぞれのクラスタ数と復元率を対応させてみると，クラスタ数が多い方が提案手法と比較手法 2 で差ができ、逆に、比較手法 1 とは，クラスタ数が少ない方が差が出来ることがわかります．このことからクラスタ数が多い時は、バックアップデータの配置の影響が大きく，クラスタ数が少ない時は，プライマリーデータの配置の影響が大きいと考えられます．

data2_5km と data3_10km のようなクラスタ数が少ないデータセットでは，他のデータセットを用いた実験と比べて提案手法と比較手法 2 の復元率の差が少ないが，要因はバックアップの配置の際に提案手法で選択するバックアップ先と同じデータベースをバックアップ先として選択する可能性が高いためと考

えられる．一方で同じくクラスタ数の少ない data1_5km のような，避難所の位置の偏りが顕著でありデータベースの配置場所にも偏りが生じたものは，その２つのグラフと比べて提案手法と比較手法２の復元率の差が生まれてしまう．その要因として，先の２つと同じく提案手法と同じバックアップ先を選ぶ可能性が高い状況ではあるが，配置の偏りにより１つのバックアップの設置場所の違いが復元率に大きく影響を与えてしまっているのだと考えられる．

また，提案手法同士を比べると，data1_5km と data2_5km のグラフで，data1_5km の方が提案手法の結果が良くなっている．これは，data1_5km では避難所が中心から離れた位置に多くあり，広い範囲にデータベースが配置されているため，どの避難所を被害範囲の中心にとっても被害を小さく抑えられたため復元率が高くなったのだと考えられる。

6 おわりに

6.1 まとめ

ネットワークが使用不可能の状況で，災害によりサーバーに被害を受け失われたデータを素早く復元するための効果的なプライマリーデータとバックアップデータの配置を検討した．

避難所の位置座標から，クラスタリングを用いて避難所を分け，分けたそれぞれの場所にプライマリーデータを持つデータベースを設置した．

バックアップデータは，プライマリーデータとバックアップデータの間の距離が遠いほど同時に失われることがないという考えのもと，プライマリーデータから離れた位置に置くこととした．各データベースについてすべてのパターンのバックアップデータの置き方を想定するため，各データベース間の距離を係数としプライマリーデータとバックアップデータのつながりがあるかどうかの存在を表す変数と併せた，項数が (避難所の数)^2 となる目的関数が最大値をとるようなデータ配置を求めた．

その後の，復元率を求める実験では，復元率は避難所の位置の偏りによって生じる，データベースの配置場所の偏りの影響を受けてしまうことがわかった．また，データベースは分散している方がプライマリーデータとバックアップデータの同時損失の可能性を減らすことができ，データ管理の点で安全である．また，バックアップデータの配置を工夫する前にプライマリーデータの配置を工夫することで大幅に復元率が向上させることが可能であり，一方でプライマリーデータの配置を考慮せず，バックアップデータの配置のみを考慮することは効果が薄いことがわかった．

6.2 今後の課題

復元率を向上させるため，プライマリーデータを離して配置することが出来るよう工夫していくことや，１つのプライマリーデータから作成するバックアップデータを増やすことが考えられる．プライマリーデータを離すために，座標に応じた重み付けをするといった方法が考えられる．

また，１つのプライマリーデータから作成するバックアップデータを増やすことで復元の際に近くのバックアップデータを選ぶといった選択が可能になり，復元にかかる時間の短縮が見込める．これにより災害時に素早い復元が必要になるという背景に沿った評価指標を立てることも可能になると考えられる．

更に，課題として現実性を向上させていく必要があると考える．本手法ではデータ量の差を考慮していないため，１つのプライマリーに対するバックアップの数の最適化，データ量を考慮したクラスタの数の最適化といったものが今後の改善策として考えられる．また，バックアップデータ配置において海抜などの地理情報を考慮することによる現実性の向上の必要もあると考える．

謝　辞

本研究の一部は，厚生労働行政推進調査補助金政策科学推進研究事業「ナショナルデータベース (NDB) データ分析における病名決定ロジック作成のための研究」の助成により行われた．

文　献

[1] NGUYEN Hoai Nam, 荒堀 喜貴, 横田 治夫, “SIBM - 避難場所情報に対する RDF データセット ベンチマークツール”，第 7 回データ工学と情報マネジメントに関するフォーラム，2015.

[2] 地震発災時における地方公共団体の業務継続の手引きとその解説 第 1 版, “http://www.bousai.go.jp/taisaku/chuogyoumukeizoku/chiou/pdf/h22kaisetu.pdf” ,2019 年 01 月 10 日閲覧.

[3] 市町村のための業務継続計画作成ガイド, “http://www.bousai.go.jp/taisaku/chihogyoumukeizoku/pdf/H27bcpguide.pdf” , 2019 年 01 月 10 日閲覧.

[4] 大規模災害発生時における地方公共団体の業務継続の手引き, “http://www.bousai.go.jp/taisaku/chihogyoumukeizoku/pdf/H28tebiki.pdf”, 2019 年 01 月 10 日閲覧.

[5] MacQueen, J. B. “Some Methods for classification and Analysis of Multivariate Observations” Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. pp. 281–297.

[6] Mixed Integer Linear Programming (MILP) solver, “https://sourceforge.net/projects/lpsolve”, 2019 年 01 月 10 日閲覧.

[7] 最適化問題の数理・計算的解法, “http://www.edu.kobe-u.ac.jp/fmsc-hrymlab/Lecture/lp_solve/lpsolve.pdf”, 2019 年 01 月 10 日閲覧.

[8] E. F. Codd, “A Relational Model of Data for Large Shared Data Banks”, Communications of the ACM (CACM), Vol. 13, No. 6, pp. 377–387, 1970.

[9] Takaki Nakamura, Shinya Matsumoto, Hiroaki Muraoka “Discreet Method to Match Safe Site-Pairs in Short Computation Time for Risk-Aware Data Replication”, IEICE TRANSACTIONS on Information and Systems Vol.E98-D No.8 pp.1493-1502