

配列指向 DBMS を用いた 避難シミュレーションデータの格納と分析

河井 悠佑[†] 杉浦 健人[§] 趙 セイ[§] 石川 佳治[§] 脇田 佑希子[§]

[†] 名古屋大学工学部電気電子・情報工学科 [§] 名古屋大学大学院情報科学研究科

1 はじめに

大規模なシミュレーションデータに対する効果的な解析ツールが求められている。災害対策を例に挙げると、予測や被害の想定のためにシミュレーションが用いられている。計算機の性能向上及びストレージの容量拡大により、大規模なシミュレーションデータが生成され、その活用が課題になっている。大規模なシミュレーションデータが多数生み出されている今日では、高速な応答によるインタラクティブな分析が重要である。

最近では大規模データのインタラクティブな解析を実現する配列指向 DBMS (array DBMS) が登場している [3]。本稿ではこの配列指向 DBMS に着目し、大規模な時空間シミュレーションデータの格納及び分析について述べる。時空間情報を配列の次元と結び付けることで多次元配列として扱うことができ、配列指向 DBMS で効率的に扱うことができる。対象となるデータは大規模地震発生時の避難シミュレーションである [1, 4]。このシミュレーションデータは東京工業大学の東大研究室から提供を受けたものである。

2 シミュレーションデータ

提供を受けたシミュレーションモデル [1, 4] の概要を説明する。首都直下型地震を想定した避難シミュレーションであり、対象となる場所は東京都足立区北千住である。この地域は荒川と隅田川に囲まれており、地域外への広域避難が困難である。さらに木造住宅密集地域が多いため、建物の倒壊及び火災の危険性が高い。このシミュレーションでは、18 時に東京湾北部で震度 6 強の地震が発生し、北千住地域で火災が発生した状況を想定している。この状況で発災から 24 時間経過までの人・建物・道路の様子シミュレーションが行われている。提供を受けたデータは、多数のシナリオによるシミュレーションログデータのうちの 1 つである。

人・建物・道路のシミュレーションログデータのうち、本研究では人の避難の様子に関して格納及び分析を行う。元データは CSV 形式で、様々な情報が記録

されているが、今回用いる属性は [経過時間, 人の固有 ID, x 座標, y 座標, 人の状態] である。なお、配列指向 DBMS で扱うデータはグリッド構造のデータを想定しているため、準備段階としてシミュレーションデータを時空間グリッドごとで集計する。詳細は後述する。

3 配列指向 DBMS

本研究で使用する DBMS は SciDB [2, 3] である。SciDB は科学的分析向けに開発された多次元配列形式の大規模並列 DBMS である。リレーショナル DBMS ではリレーション (テーブル) 形式のデータを扱うのに対し、配列指向 DBMS は大規模配列の蓄積と問合せに特化したシステム構成がなされており、対応するインタフェースが提供されている。

多次元配列では各次元に対して順序付けが行われている。SciDB では意味的に近いセル同士は参照されやすいと想定し、配列データ上で意味的に近いセルを物理的にも近い位置で扱うように設計されている。この設計に用いられるのがチャンクである。チャンクは SciDB におけるデータ処理の基本的な物理単位である。SciDB に大量のデータが格納されたとき、意味的に近いデータ同士が同じチャンクに集められる。SciDB 上の演算や問合せもチャンクごとに参照される。したがって、SciDB ではチャンクに基づいて空間的に集まっているデータを効率的に処理できる。RDBMS における索引を、SciDB では次元とチャンクを組み合わせることで実現する。また、SciDB には複数のインスタンスが用意されている。これらにチャンクを均等に振り分けることにより、効率的な並列処理が実現する。チャンク分割及びインスタンスによる並列処理の様子を図 1 に示す。

チャンクの大きさは設定可能である。チャンク内の総セル数が等しくても、各次元の長さの取り方によって処理の効率性は変わる。参照されるチャンク数が少なくなるように、データの性質や利用方法に応じて適切なチャンクサイズを定める必要がある。

4 ユースケース

上記に示した避難シミュレーションログデータの分析には様々なものが考えられるが、例として避難状況に基づく人数の集計、及び避難時における移動の可視化について考える。

Storing and Analyzing Evacuation Simulation Data Using an Array-oriented DBMS

Yusuke Kawai[†], Kento Sugiura[§], Jing Zhao[§], Yoshiharu Ishikawa[§], Yukiko Wakita[§]

[†] Department of Information Engineering, School of Engineering, Nagoya University

[§] Graduate School of Information Science, Nagoya University

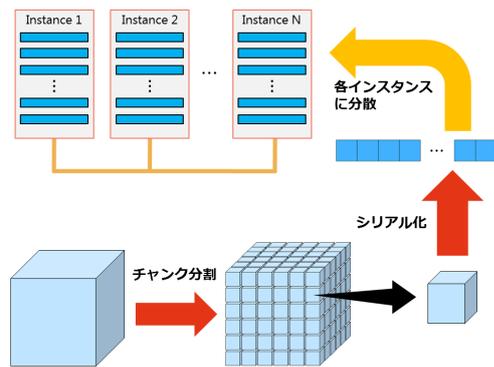


図 1: チャンク分割及び並列処理

4.1 避難状況に基づく人数の集計

経過時間と平面グリッドに加えて状況別で人数を集計することを考える。経過時間別である区画の人数や避難状況を把握することができる。具体的には、避難が %完了したときの時間を求めたり、死者や避難困難者がどの地域に多いかを把握することができる。この集計データの応用として、人数の変化を見るためにある時間単位での差分を取るなどの分析が考えられる。人数が変化する区画は、人々が動いているまたは状況が変化している箇所に対応しており、情報として重要である。また、集計データを時間単位や xy グリッドの粒度を様々な度合いに変えて用意することで、段階的な分析を行うことができる。粒度が低いデータを参照することで、大まかな情報を把握し、必要な部分に関しては粒度を高くして詳細な分析を行うことが考えられる。

4.2 避難時における移動の可視化

移動の様子を集計することで、移動方向や避難経路を把握することができる。移動距離と移動方向という極座標形式で格納することを想定する。この場合、移動方向を一目で把握することができる。移動距離の大小で、急いで避難を行っていたり、避難が滞留気味であるなどの情報を把握することができる。移動の様子に関しても、人数の集計と同様に様々な粒度のデータを用意することで、段階的な分析を行うことができる。

集計のイメージを図 2 に示す。発災から 1 時間経過したときの人数の集計及び移動の様子を $50\text{m} \times 50\text{m}$ グリッド単位で示す。ただし、この図では人の置かれた状況は考慮していない。人数の大小は赤の濃淡で表している。矢印は 1 分後の移動の様子を表している。矢印が長いほど移動距離が長い。

5 実装

まず前処理としてシミュレーションデータからユーザーケースに応じたグリッド構造のデータを生成する。この処理は R 言語で行う。今回は xy 平面に関して $5\text{m} \times 5\text{m}$ を一つのグリッドとして人数の集計を行う。

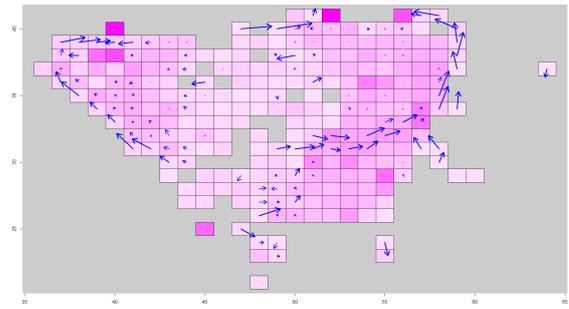


図 2: 集計のイメージ

経過時間は 0.1 分刻みである。約 1,000 万のグリッドデータが生成され、これを SciDB で次元 {経過時間, x 座標, y 座標, 人の状態}, 属性値 [人数] の 4 次元配列として格納する。

これに対し範囲指定の問合せを行ったところ、ある 1 つの経過時間を指定するだけであれば 0.2 秒ほどで応答することを確認した。時空間的な範囲指定では 10 万程度の大きさであっても 1 秒以内に応答できることを確認した。

粒度の変更に関して、SciDB 上でグリッドの統合が行えるので、最初に用意するグリッドデータを高い粒度で格納しておけば、SciDB 上で粒度を低くして、よりサマリな分析が可能である。今回の例では xy 平面上の 2×2 のグリッドを 1 つのグリッドに統合することで、 $10\text{m} \times 10\text{m}$ のグリッドとして集約できる。

6 まとめ

本研究では配列指向 DBMS である SciDB を用いて避難シミュレーションの格納を行い、簡単な問合せを実行した。今後はユーザーケースに応じて具体的な実装を進めていく。SciDB ではチャンクがデータ処理の基本的な単位なので、状況に応じたチャンクの設定を考える必要がある。

謝辞

本研究の一部は、CREST「大規模・高分解能数値シミュレーションの連携とデータ同化による革新的地震・津波減災ビッグデータ解析基盤の創出」および科研費 (16H01722) による。

参考文献

- [1] 沖 拓弥, 大佛 俊泰. 東京都足立区北千住における大地震時の徒歩帰宅・広域避難リスク評価. 日本建築学会学術講演梗概集, pp. 1055–1056, 2014.
- [2] Paradigm4: Creators of SciDB a computational DBMS. <http://www.paradigm4.com/>.
- [3] M. Stonebraker, P. Brown, J. Becla, and D. Zhang. SciDB: A database management system for applications with complex analytics. *IEEE Computational Science & Engineering*, 15(3):54–62, 2013.
- [4] 沖 拓弥, 大佛 俊泰. 大地震時の地域住民による救助活動と逃げ遅れについて (都市計画). 日本建築学会関東支部研究報告集, 85:425–428, 2015.