# シミュレーションデータの分析管理のためのデータウェアハウスについて

石川 佳治† 王 元元† 董 テイテイ† 杉浦 健人† 佐々木 勇和‡
†名古屋大学情報科学研究科 ‡名古屋大学未来社会創造機構

## 1 はじめに

科学の分野においては，情報技術の進展やコンピュータやストレージの高性能化・大容量化を受けてデータが急速に増大している．大規模データを扱うことを前提とした科学分野における取り組みは，しばしば**データサイエンス**（data science）と呼ばれ，新たな潮流として大いに着目を浴びている．シミュレーションにより生み出されるデータは膨大なものがあり，その利活用は重要なトピックとなっている [3]．本研究では，特に**時空間的なシミュレーション**を想定し，大量に発生するシミュレーションデータを管理し，分析処理を行うデータウェアハウスのシステム技術開発を目標にする．

本稿では，地震時の人の避難シミュレーションデータを題材として，避難状況分析のシナリオのもとでのデータウェアハウスを試作した事例について紹介する．システムのアーキテクチャおよび実装技術について述べ，分析のためのインタフェースや機能について説明する．

## 2 シミュレーションデータの分析

災害シミュレーションの領域では，地震や，それに伴い生じる津波や人の避難行動など，さまざまなシミュレーションが行われている．そのようなシミュレーションではしばしばある特定の領域が対象とされ，地震の発生後に，時間が経つにつれそれぞれの場所においてどのような変化が生じるかが計算される．たとえば津波の場合では，各地点の水の高さがどの程度になるかが各時点ごとに与えられ，人の避難の場合には，それぞれの避難者が各時点にどこに存在するかが求められる．得られるシミュレーションデータは，発生時刻などの時間と，座標値などの空間情報と結び付けられている．

災害に関する時空間的なシミュレーションが行われると，シミュレーション結果の時空間データが大量に発生するため，そのようなデータを管理し，その後の分析処理のために役立てることが求められる．たとえば，分析者が地震の規模や場所を指定して，複数のシミュレーションデータを統合して被害状況を分析することを考える．「4 月のウィークデイの正午に，千葉県沖の震源地 $X$ を中心とするマグニチュード 8 の地震が起きたとき，東京都心部における震度，津波の被害，人の避難の状況を統合的に分析したい」という要求があるとする．このような場合，まず，条件に合ったシミュレーションデータをデータベース内で特定する必要がある．一方，分析作業は一般的には探索的に行われる．詳細なシミュレーションデータをいきなり分析するのではなく，大ざっぱに粗くデータをさまざまな観点から眺めて注目すべき個所を発見し，その後より詳細な分析のために詳細度を上げてデータを調べていく．たとえば「地震発生後の 1km × 1km の粗いメッシュで，分析エリア内の人の避難状況を集計し，メッシュの各セル内に居る人の人数を求めよ」といった問合せが即座に実行でき，可視化などの手段で各時点における各セル中の人数の分布を大まかに捉える機能があれば，分析に大いに役立つものと考えられる．

このような要求を考えると，シミュレーション結果の事後的な分析のために**データウェアハウス**（data warehouse, DWH）[2, 6] の技術を活用することが考えられる．時空間シミュレーションでは時間情報と空間情報の双方が関わってくるので，それらに応じた対応が必要となる．また，科学分野ではビジネス分野と比べてより探索的な側面があり，対話的に試行錯誤しながらデータを分析することが求められる．

## 3 システムの設計

### 3.1 対象とするデータセット

試作システムで事例として取り上げるデータセットは，大規模地震の発生を想定した，高知市における避難時の人流のシミュレーションデータである*．今回実験で用いるサンプル人流は，9 時に地震が発生し，避難開始のピークが地震発生の 60 分後にあるという条件のもとで作成された 6 時間分のデータである．約 5 万人の人が避難する状況を，パーソントリップデータに基づいてシミュレーションしている．元データは CSV 形式で，各列は [パーソン ID，時刻，経度，緯度] という書式である．

### 3.2 システムのアーキテクチャ

システムのアーキテクチャを図 1 に示す．分析者はウェブブラウザを用いてシステムをによる分析を行う．

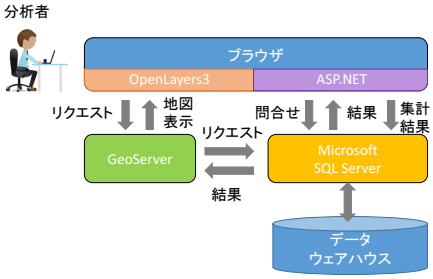


図 1 システムのアーキテクチャ

ブラウザ上で地図を表示するため，ブラウザ上で幾何データを扱うための JavaScript および CSS のライブラリである OpenLayers [5] を用いる．ブラウザ上の地図以外のビューには ASP.NET フレームワークをを利用し，バックエンドでデータベースへの問合せをしたり，データベースからの問合せ結果をもとに集計をし，GeoServer が参照する集計結果をデータベースに格納する．

GeoServer [1] は，地理情報の共有や編集を行うオープンソースのサーバソフトウェアである．要求に応じて空間データベースにアクセスし，データベースに格納された地理情報をベクタ形式やラスタ形式の図として配信できる．その他にも，さまざまな地理・空間データに対する機能を備えている．

Microsoft SQL Server では多次元データキューブを含むデータウェアハウスの基本機能が提供されているため，本研究で

* 東京大学 関本研究室からの提供による．

---

Data Warehousing for Analysis and Management of Simulation Data
Yoshiharu Ishikawa†, Yuanyuan Wang†, Tingting Dong†,
Kento Sugiura†, Yuya Sasaki‡
†Graduate School of Information Science, Nagoya University
‡Institute of Innovation for Future Society, Nagoya University

はこれを活用する．システム実装においては，SQL Server で問合せ言語として提供されている **MDX** (MultiDimensional eXression) [4] を用いる．**多次元式**と呼ばれる言語であり，多次元データウェアハウスに特化した言語である．

## 4　システムの実装・評価

### 4.1　データウェアハウスのスキーマについて

構築したデータウェアハウスのスキーマを図 2 に示す．`EvacuationRecordTable` が**ファクトテーブル**（fact table）にあたる．`EvauationRecordNumber` は一意に付与される主キーの ID である．`PersonKey` は元のデータにおける人の ID であり，各避難者に対応している．`DateKey` と `PlaceKey` は，それぞれ時間の次元と場所の次元に対応する外部キーである．つまり，このテーブルでは，時間と場所の観点から人を捉えるということを表している．

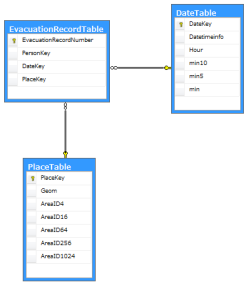


図 2　スキーマのグラフ

`DateTable` は**ディメンションテーブル**（dimension table）にあたる．`DateTimeInfo` は時刻の詳細情報を表す．`hour`, `10min`, `5min`, `min` は，時間単位，10 分単位，5 分単位，1 分単位で見たときに `DateTimeInfo` の値がいくつになるかを表しており，`hour-10min-5min-min` という概念階層に対応する．

`PlaceTable` もディメンションテーブルである．`Geom` は `geometry` 型をとるデータ（座標値）であり，人の位置を表す．ここでは，対象領域を 4 分割（2 × 2 分割），16 分割（4 × 4 分割），64 分割（8 × 8 分割），256 分割（16 × 16 分割），1024 分割（32 × 32 分割）に 5 パターンで均等分割することを考える．それぞれの場合に対し，分割されたセルに対しセル番号を割り当てている．このテーブルの `AreaID4`, `AreaID16`, `AreaID64`, `AreaID256`, `AreaID1024` には，そこにその人が存在しているセルの番号が保持される．

図 3 は，格納したデータを抜き出してプロットしたものである．ここでは空間を 16 分割している．高知市を中心として，高知県の湾岸に避難する人が点在しているのがわかる．

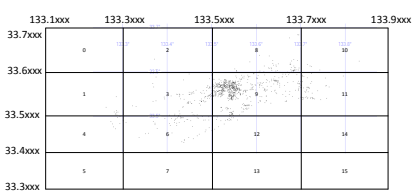


図 3　データの表示例

### 4.2　分析インタフェースの構築

前節で構築したユーザインタフェースを図 4 に示す．分析者はすべての操作をブラウザ上で行う．まず分析者が時間区間を指定すると，ブラウザ上には人流の時空間的な密度を表すグリッドが重ねられた地図が表示される．このグリッドはヒートマップになっており，密度が大きいセルは赤く，小さいセルは青く表示され，人流のないセルについては無色で表示される．初期状態では高知市を中心とした広い地域の地図が表示されており，更に詳細な情報を表示したい場合はセルを拡大表示する事ができる．ズームアップするとグリッドがより細かく分割されて表示される．拡大表示の例を図 5 に示す．分析者はこの操作を繰り返し，広域な地図から段階的に地域を絞りこみながら分析を行うことができる．

条件が指定されるとバックエンドでの問合せ処理が行われる．事前に用意してある MDX 問合せのテンプレートを用いて，SQL Server に対して問合せを行う．SQL Server では，指定された時間区間に対し，指定された分割数について各セル内を通過した人数を集計する．問合せにより得られた結果から，単位面積・単位時間あたりの人数を求めた後，それぞれ各分割数ごとに用意した集計結果テーブルに格納する．GeoServer は集計結果を参照し，データに応じたヒートマップレイヤを自動で生成し，OpenLayers を通してブラウザ上の地図に反映する．

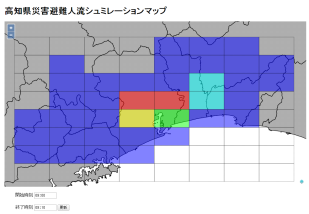


図 4　ユーザインタフェースにおける可視化

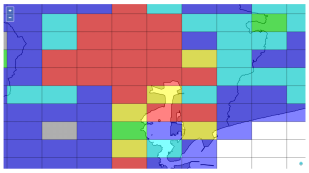

図 5　拡大表示した例

## 5　まとめと今後の課題

本稿では，試作した時空間災害シミュレーションデータに対するデータウェアハウスについて，アーキテクチャ，機能，インタフェースなどについて説明した．今後の課題として，津波予測などの他の災害シミュレーションデータの統合，可視化処理の高速化，インタフェース・分析機能の高度化などが挙げられる．


## 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」，文部科学省委託事業「DIAS-P」，および科研費（25280039）による．


## 参考文献


[1] GeoServer. `http://geoserver.org`.

[2] J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 3rd edition, 2011.

[3] T. Hey, S. Tansley, and K. Tolle eds. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, 2009.

[4] MultiDimensional eXpressions. `http://en.wikipedia.org/wiki/MultiDimensional_eXpressions`.

[5] OpenLayers 3. `http://openlayers.org`.

[6] A. Vaisman and E. Zimányi. *Data Warehouse Systems: Design and Implementation*. Springer, 2014.