

# 地理データに対するデータマイニングの適用と応用に関する研究

株式会社 東芝

官公情報システム技術第2部 松本 雄一 土井 利次

研究開発センター 仲瀬 明彦 波田野 寿昭

## 1. はじめに

地理情報システム (GIS) は、地形情報だけでなく、天候、国勢調査結果など様々な情報を地図情報として統括し、表示することを可能にした。また、昨今の計算機性能の向上から、大量データの高速処理を可能にし、データから規則性を自動抽出するデータマイニング技術が注目されるようになった。本共同研究では、GIS とデータマイニングを融合した方式 (以下「空間データマイニング」と呼ぶ) を考案し、実験システムの開発を行った。空間データマイニングでは分析対象の観測データに GIS から得られる情報を附加して分析を行うため、観測データの不足を補完するとともに、分析結果を視覚的に表示できる。本稿では方式、システム概要、および適用実験結果について述べる。

## 2. 空間データマイニング方式

空間データマイニングの特長は、(1)収集されたデータの補完、(2)視覚的な表示による可読性の向上、(3)空間的な特徴の抽出にある。

### (1) 観測データの補完

GIS には地図に関する情報、すなわち、地形、河川などの自然情報に加え、市町村境界、道路、建造物などの人工的な情報、さらには、地域別人口密度、降水量などの動的な情報が埋めこまれたものがある。しかしながら、多目的に集められたデータが集約されているため、いわゆる関係データベースのような形式ではなく、できるだけコンパクトにデータを納めることを念頭に置いたデータ形式となっている。空間データマイニングでは、分析対象のデータに GIS から得られる情報を加え、情報の補完を行う。(図1)

補完情報は、観測地点の近傍を探索することで得られる。例えば、「50m 以内に工場がある場所は道路が傷みややすい」といった規則を発見するためには、各観測データに対して、「50m 以内に工場があるかどうか」を調べ、その結果を分析に加えるのである。探索はこのように

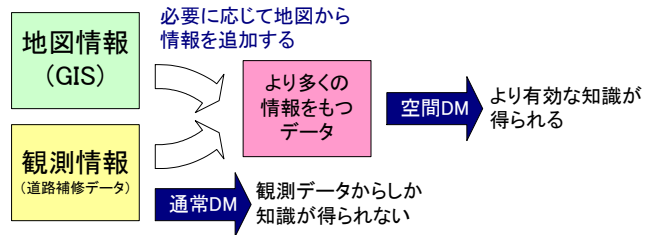


図1 空間データマイニング

Yes/No で答えられる項目以外にも、「50m 以内にある大型の建造物は何か」というような多値の離散情報や、「最寄りの工場までの距離」といった数値情報を集めることも考えられる。これらの情報と観測データを合わせて分析を行うことで、観測データだけでは得られない結論を得ることができる。

### (2) 空間的な情報把握

位置情報を含むデータは空間的に表示することで理解しやすくなる。図2は道路の損傷の仮想的な規則性を示したものである。図2(a)のように決定木上で「近くに河がある」と言葉で表現されるよりも、図2(b)のように、「近くに河がある」場所を地図上で示す方が情報の把握がしやすい。位置を伴う情報を地図上に表示させるシステムはこれまでに開発されているが、空間的な特性を持つデータを視覚的に表示することで人間の支援を行うことを目的としたものである。したがって、空間的な特性の発見そのものは人間自身が行わなければならない。

これに対して空間データマイニングでは、データの空

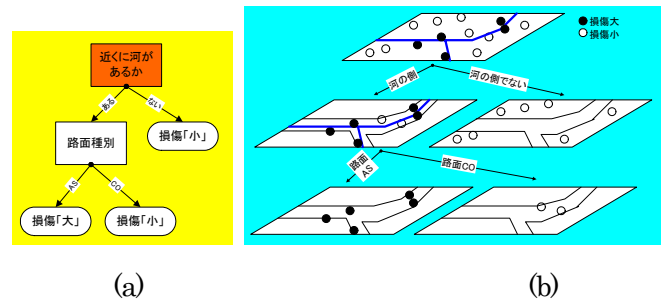


図2 データの視覚的表示

間的特性を自動的に獲得することを可能にしている。

### 3. システム概要

実世界の空間データを空間データマイニング方式を用いて分析し、その有効性を確認するため、空間データマイニングプロトタイプシステムを作成した。プロトタイプシステムは図3に示すように、決定木とGISシステムを統合することで実現している。

プロトタイプシステムの基本機能として、(1)観測データの緯度経度情報をもとに、ポイントを地図上に表示、(2)観測データの緯度経度情報をもとに、各ポイントの指定された属性値を地図上に表示、(3)観測データの各ポイントの指定された複数の属性値の時系列表示、(4)地図上の領域ごとに、観測データの指定された属性値の平均値を求め、指定された色で領域を塗り分け、(5)観測データ内の1つの属性をクラス属性とし、他の指定された属性を説明属性とした決定木を構築、(6)作成された決定木の部分木を指定し、その部分木内に含まれる観測データのクラス属性値を地図上に表示、(7)作成された決定木を新たな領域のデータに適用し、クラス属性の予測結果を地図上に表示、(8)地理情報システムの持つデータ（ランドマーク、地形情報）などを観測データの1属性として追加、

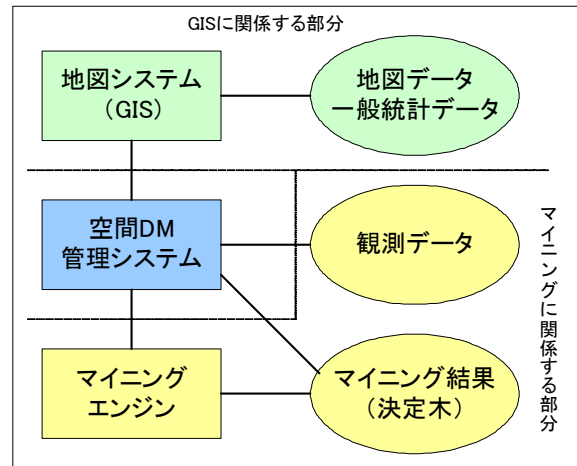


図3 システム構成

などが実装されている。システムはWindows2000上で動作し、GIS部はインフォマティク社SISシステムを採用している。

### 4. 適用実験

数値地図2500（岐阜県）、および岐阜県内一部の道路舗装点検データを題材に空間データマイニングの適用実験を行った。分析中のシステム画面を図4に示す。図4

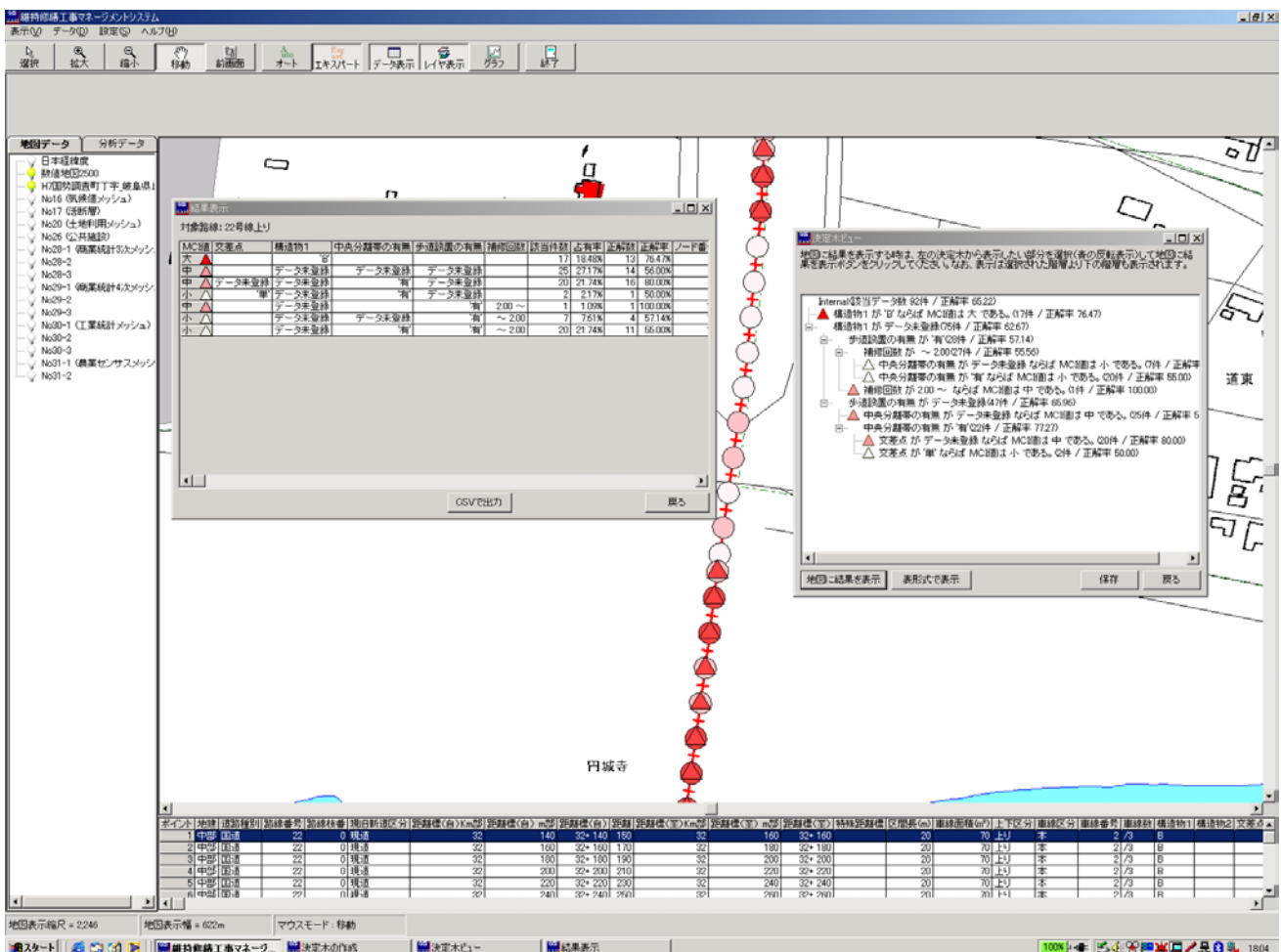


図4 システム画面

の中央を縦にのびる道路を一定区間ごとに点検したデータから、損傷度合を大中小の3段階で表し、これらの損傷度合を特徴付ける規則抽出を行うことを分析の目標とした。損傷度合はMC I 値に基づき、損傷小 (MCI 値8以上)、損傷中 (6以上8未満)、損傷大 (6未満) と区分した。

画面内部に開いている2つのウィンドが分析結果である。右側が分析結果を決定木で表示したものであり、左側は表形式で表示したものである。表形式表示の項目は、表内の該当件数：表の各レコードの条件にマッチするデータ件数、占有率：全データ件数中で該当件数の占める割合、正解数：条件通り正しく分類されたデータの数、正解率：正しく分類されたデータの割合を意味している。ノード番号は、決定木との対応を取るためのデータで分析結果とは関係ない。

分析結果の解釈方法は、例えば表中の第1行目の場合、「構造物がB」（橋の上の道路であることを意味している）である道路は、MCI 値が大、すなわち損傷が小である。この条件に該当する箇所は17箇所あり全体の18.48%である。実際にこの条件でMC I 値が大であった箇所は13箇所であり、分類の正解率は76.47%である」と解釈される。分析の結果得られる規則は、正解率だけでなく、該当件数も大きなものが有用であると考えられている。それは、数少ない例外的な事例にのみ有効な規則より、多くの事例に有効な規則の方が現実に利用される場面が多いからである。

実際の分析は広範囲の舗装点検データを用い、すべての観測項目、GIS から得られる空間情報を使って行い、多くの有効な結果を得ることができた。

## 5. おわりに

本研究による空間データマイニングは、GIS システム、GIS の地図上の地物データ、データマイニングエンジンを組み合わせることで、地図上に表されるデータに関する様々な知識を発見することが可能となる。

これまでの単なる背景としての地図利用から、地図データそのものを分析に用いるなど、地理情報を用いたデータ分析への要求は高まっている。地図データを多角的に処理できるGIS技術は、地図データの高度利用には不可欠であり、今後、GIS技術の発展と共に、より高度なデータ処理への応用が可能となる。

本研究により、GISの新しい活用方式の実用化が見えてきた事は、非常に有意義であった。今後は、本システムを道路保守、交通シミュレーション、行動ナビゲーションなどに応用してゆきたいと考えている。

## 問い合わせ先

株式会社東芝 e-ソリューション社

官公情報システム事業部 官公情報システム技術第2部

松本 雄一、土井 利次

〒105-8001 東京都港区芝浦 1-1-1

Tel 03-3457-2597 FAX 03-5444-9273

株式会社東芝 研究開発センター

コンピュータ・ネットワークラボラトリ

仲瀬 明彦、波田野 寿昭

〒212-8582 神奈川県川崎市幸区小向東芝町1

Tel 044-549-2235 FAX 044-520-1841