

汎用統計ソフトを用いたデータマイニングによる上水道の地震被害形態の解析

(株)地崎工業 正会員 須藤 敦史
 武蔵工業大学 正会員 星谷 勝
 (株)地崎工業 正会員 渋谷 卓

1. はじめに

近年、コンピュータのハードウェア性能の向上やネットワーク技術の発達により、様々なデータを大量に蓄積することが可能となっている。しかし、その反面データ量の増加に伴い、本来要求されている情報・知識の検索が難しくなっているのも事実である。そのため、蓄積された膨大なデータを有効に活用する解析技術の必要性が高まってきており、簡単にデータ分析を行うための汎用ツールの開発が進められている。特に、大量なデータの中に隠れた知識（規則性や相関関係など）の発見を目的としたデータマイニングに関する研究が盛んに行われている。

本研究では阪神大震災の地震による神戸市・芦屋市・西宮市の上水道配水管の被害データをもとに汎用的な統計解析ツールである Excel を用いたデータマイニングより、上水道配水管の被害状況や影響を与える様々な要因についての解析を行っている。

2. データマイニング

データマイニング（Data Mining）とは、直訳すれば「データの発掘」であり膨大なデータの中に隠れている知識や規則を客観的に発見することである。またデータマイニングは、データベースからの知識発見（KDD：Knowledge Discovery in Databases）とも呼ばれており、図-1 に示すように6段階のプロセス（手順）が基本となっている。

データマイニングは、様々な仮説を立てて、これをデータにより検証する「仮説検証型」と、データを単純な表現形式に変換して実際のデータから隠れた知識を発見する「仮説生成型」に分類される。本研究では前者の「仮説検証型」を用いてデータマイニングを行っている。

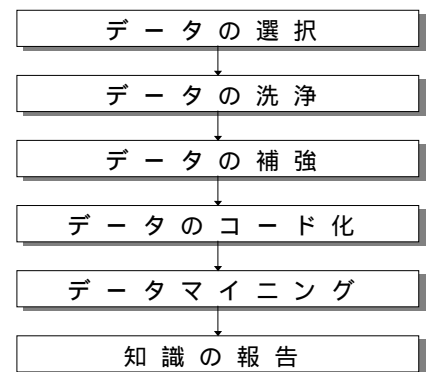


図-1 KDDの過程

3. 統計解析ソフト

最近、パソコンの急激な普及とネットワーク技術の整備に伴い、統計ソフトが手軽に使用できるようになってきており、代表的な統計ソフトは STATISTICA, SPSS, SAS, S-PLUS などがあげられる。ここで汎用的かつ簡易なデータマイニングを行うにはパソコンで稼動する統計ソフトをいかに活用するかが重要なことであり、データマイニングを実現するパソコンの汎用的な統計ソフトが有する要件は以下となる。

- a)多くの統計手法がサポートされている b)多量なデータを高速で処理できる c)データ数・変数等の制約が少ない d)安価で手軽に使用できる e)必要に応じて統計プログラム（マクロ等）の作成ができる f)他の表集計ソフトとの連携がよい g)マニュアル・ヘルプ等が完備・充実している h)グラフ機能が充実している

以上より、マイクロソフト Excel は現在の環境では汎用的な統計解析ツールであるため、本研究では Excel を用いてデータマイニングを行っている

4. 地震被害データへの適応

データマイニングに用いたデータベースは、神戸市・芦屋市・西宮市における上水道配水管の被害データを用いており、観測項目には震度・地形・管径・管種などがある。

表-1 都市別仮説検証

	被害件数			被害率		
	神戸市	芦屋市	西宮市	神戸市	芦屋市	西宮市
改變山地	No	No	No	No	-	No
段丘	No	No	No	No	No	No
谷・旧水部	No	No	No	No	No	No
沖積平野(液状化無し)	Yes	Yes	No	No	Yes	Yes
沖積平野(液状化有り)	No	-	No	No	No	No
良質地盤	-	-	-	Yes	Yes	No

これらの観測項目の中で、配水管の被害への相関について仮説「震度が大きくなるに従って被害も拡大する」が成立するかどうかを検証した。ここで観測項目ごとに被害件数、被害率において仮説と観測項目との関係を3段階（Yesは71～100%で仮説成立、Noは0～30%で仮説成立、-は31～70%で仮説成立か比較不可能）で表した。ここで解析結果を都市別にまとめたもの表-1、管径別を表2、管種別を表3に示す。さらに詳しく検証するために被害率を用いて図-2を作成した。

表-2 管径別仮説検証

	被害件数				被害率			
	75-75	100-150	200-450	500-800	75-75	100-150	200-450	500-800
改变山地	No	No	No	-	-	-	-	-
段丘	No	No	-	-	-	No	No	-
谷・旧水部	No	No	No	Yes	No	No	No	-
沖積平野(液状化無し)	-	-	Yes	No	Yes	Yes	Yes	-
沖積平野(液状化有り)	No	No	No	No	No	No	Yes	-
良質地盤	No	-	-	Yes	-	-	-	-

表-3 管種別仮説検証

	被害件数		被害率	
	DIP-A.K.T	CIP	DIP-A.K.T	CIP
改变山地	No	No	No	-
段丘	No	No	No	No
谷・旧水部	No	No	No	No
沖積平野(液状化無し)	-	-	No	Yes
沖積平野(液状化有り)	No	No	No	No
良質地盤	-	-	-	Yes

5. 解析結果

表-1の被害率表に示すように、沖積平野（液状化無し）と良質地盤には仮説が成立し、改变山地、段丘、谷・旧水部、沖積平野（液状化有り）には仮説が成立しないことが分かった。つまり前者は震度が大きくなるに従って被害も拡大する（相関性を有する）ため、被害予想がしやすい地形であるが、後者は震度が大きくなるに従って被害が拡大するとは限らないため、震度の大きさによる被害を予測しにくい地形であることを示唆している。また表-2からは管径が200～450mmの管のほうが、他の管径の管よりも被害予測しやすいことがわかり、表-3からは管種がCIPの管のほうがDIPの管よりも被害予測しやすいことが分かった。

図-2より、改变山地・段丘・谷・給水部では震度5程度の震度でも地盤の被害が大きいため、震度5からの対策を考えていくことを示唆していると言える。また同じ沖積

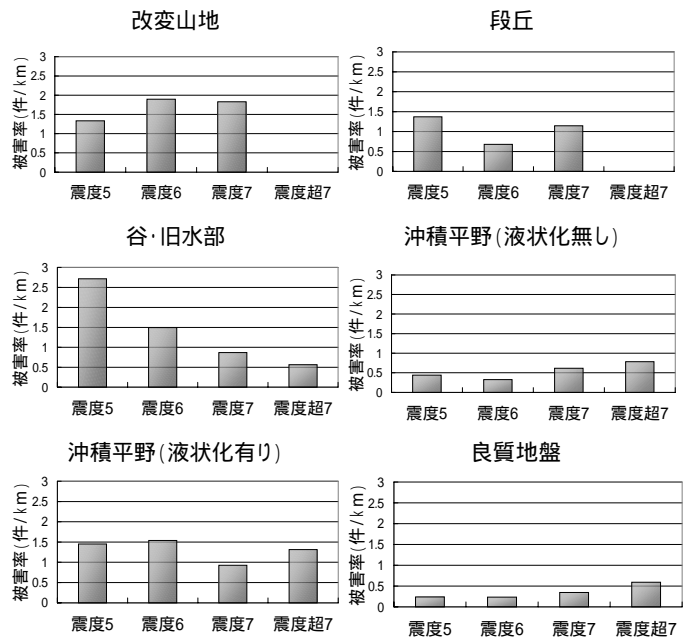


図-2 震度別被害率

平野でも液状化している地形としていない地形ではまったく異なる結果が得られた。沖積平野(液状化無し)では、比較的仮説が成立するが沖積平野(液状化有り)では震度の大きさに関わらず全体的に大きな被害が出ているため、被害予測しにくい地形でもあり、防災も困難な地形であると言える。加えて良質地盤・沖積平野(液状化無し)では、被害率が小さいが仮説が成立しており、被害予測は比較的しやすい地形である。

6. まとめ

Excelを用いて上水道配水管の被害に影響を与える様々な要因についてデータマイニングを行った結果、仮説「震度が大きくなるに従って被害も拡大する」は管種や管径よりも、地形の種類に大きくかかわってくることが導き出された。よって地形の種類により地震の被害予測・対策を考えていくのが、効率もよく適切ではないかと考えられる。

また、今年度はニューラルネットワークより汎用的な Excelを用いて解析を行ったが、これによりデータマイニングの本来の長所である、専門（地震工学）知識がなくてもでも簡単に解析が行えることの一部が証明されたのではないかと考える。

[参考文献]

- 1) Pieter Adriaans・Dolf Zantinge 著 山本英子・梅村恭司訳：データマイニング，共立出版，1998.
- 2) 地震による水道管路の被害予測，社団法人日本水道協会，1998.