AI による時系列データ等の解析事例

中日本建設コンサルタント（株）○山田 瑠莉子　中根 進

筆者の一人は、平成 20 年に AI の学習技術の１つであるニューラルネットワークを使って雨天時浸入水を含む流量の回帰を行い下水道協会誌に投稿している。回帰した流量から降雨時の晴天時流量を推定し、雨天時浸入水量を計算した。

その後、管路内水温の解析や脱水ケーキ含水率に及ぼす操作因子の推定など膨大な量の情報を AI の学習技術を使って解析してきた。

本稿では、AI の学習技術を回帰や分類に利用した事例を紹介するものである。

Key Words　：　人工知能、機械学習、ニューラルネットワーク、ランダムフォレスト

1. まえがき

AI(人工知能)は、「認識・判断」「認識・理解」「予測、推論」「計画・最適化」などを実行する人間の知能を人工的に模倣するための総合的な概念技術である。この総合技術は単一の要素技術でなく、多くの要素技術を含んでおり、技術を適切に選択し、目的に対して適用することにより、その能力を発揮させるものである。[1)]

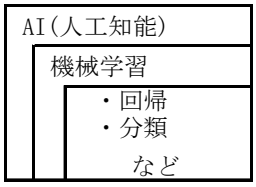

図 1.1 AI(人工知能)

人工的に知能を模倣するためには、法則やルール、事象を学習する必要があり、その学習技術を機械学習といい、**図 1.1** の回帰や分類などに用いる多くの学習方法がある。学習方法には、2 変量またはそれ以上の間の関係を回帰や分類する手法として、ニューラルネットワークやランダムフォレストという手法など多くの技術がある。機械学習の学習技術の種類を**図 1.2** に示す。[2)]

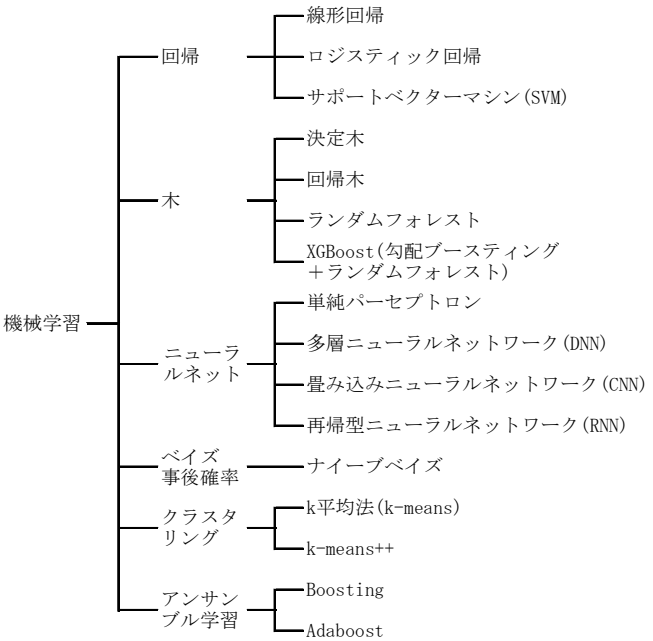

図 1.2 機械学習の学習技術[2)]

機械学習の手法として、「回帰(予測)」・「分類(判別)」のほかに「教師あり学習」・「教師なし学習」・「強化学習」という学習に分けることもある。教師あり学習は、目的変数を教師データとして説明変数との関係を回帰することや緊急度Ⅰ、Ⅱなどラベルがついた教師データに対して緊急度がついた要因を分類するなどの学習をいう。本稿では、筆者らが AI(人工知

能)のいくつかの学習技術を使い「教師あり学習」で「回帰」と「分類」を行った事例を紹介する。

2. 雨天時浸入水量解析の事例（ニューラルネット法）

2.1 ニューラルネット法の概要

AI の学習技術の１つであるニューラルネットワークは、人間の脳のしくみ（ニューロン間のあらゆる相互接続）から考えられたもので、脳機能の特性のいくつかをコンピュータ上で表現するために作られた数学モデルと言われている。

筆者らの一人が文献[3)]で利用したニューラルネットワークのアルゴリズムを**図 2.1** に示す。このニューラルネットワークは入力層－中間層－出力層の３層と中間層を複数持つ深層からなるフィードフォワード型ニューラルネットワークで、学習方法には誤差逆伝搬法（バックプロパゲーション法）を用いている。**図 2.2** は中間層を 2 層設定した図となっている。それぞれユニット(ニューロン)は複数設定できる。今までの解析した結果では、中間層が１層でも複数層でも解析結果が大きく変わるものではない。

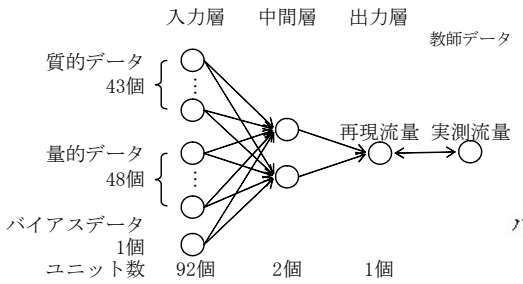

図2.1 中間層1層のニューラルネットワーク

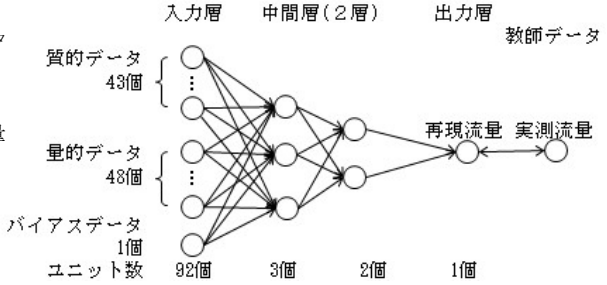

図 2.2 中間層２層のニューラルネットワーク

管きょ流量調査や処理場流入流量の時系列データを解析する場合のモデルを説明するが、詳細は文献[3)]を参照されたい。

入力層には、カレンダー情報などの説明変数を割り当てるため、雨天時浸入水量の推定では 92 個のユニット(ニューロン)を設けた。出力層は、ある時刻の計測データのみに対応するため、１個のユニット(ニューロン)を設ける。また、中間層にはｎ個のユニット(ニューロン)を設ける。ユニット数により出力値と教師データ間の相関が変わるため、相関が高くなるユニット数を設定する。今までの解析事例からは中間層のユニット数ｎは３～５個で十分であった。

2.2 ニューラルネット法による雨天時浸入水量解析

ニューラルネット法により流量の時系列データを設定した説明変数を使って回帰し、計測流量を再現する。得られた説明変数に対する重みを用いて降雨をなし(0)として再計算することにより雨天時の晴天時水量を推定する。

解析結果の一部を**図 2.3** に示す。流入水量の計測値を打点(●印)で示し、ニューラルネット法で回帰した流入水量を青色実線で示す。打点した計測流量と回帰した推定線はよく一致している。降雨時における推定晴天時流量(降雨のない流量)をピンク色実線で示す。

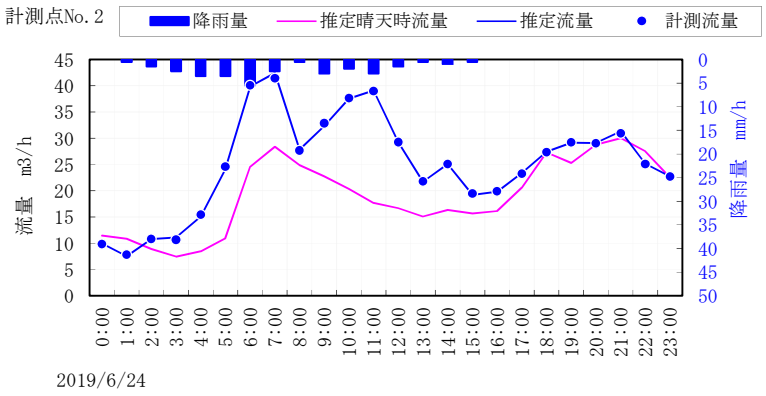

図 2.3 計測流量・降雨量と推定流量と推定晴天時流量の推移

計測流量と推定流量の当てはまり具合を**図 2.4** に示す。両流量は 45 度線上に乗っている。また、推定晴天時流量と推定流量の関係は**図 2.5** となり、45 度線より上が雨天時浸入水量となる。

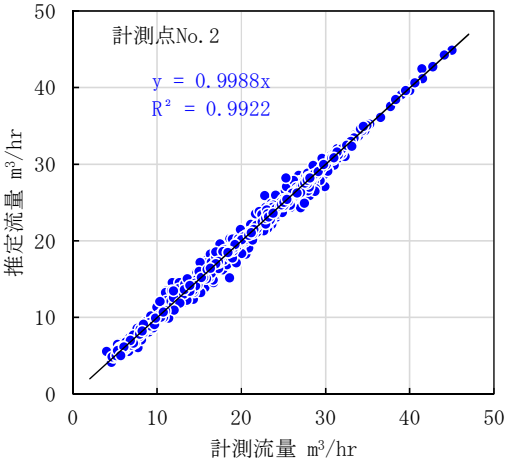

図 2.4 計測流量と実測流量の関係

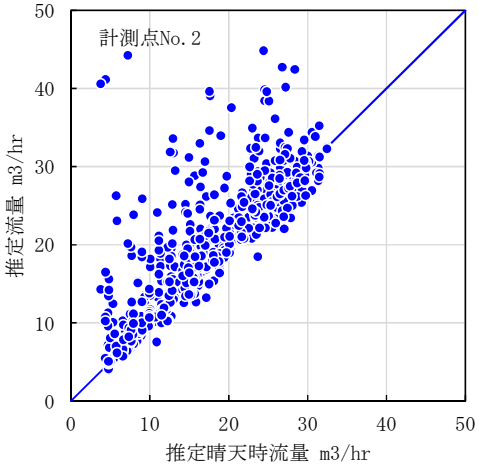

図 2.5 推定晴天時流量と推定流量の関係

2.3 雨天時浸入水量解析結果

計測期間中の雨天時浸入水量を積算し、総計測流量に対する雨天時浸入水量の割合が得られる(**表 2.1**)。この割合を計測点ごとで比較することにより雨天時浸入水の影響の大きい区域を特定できる。

表 2.1 ニューラルネット解析による雨天時浸入水量の割合

項目	単位	諸値	備考
①総降雨量	mm	212	
②総計測流量	m^3	14, 855	
③総推定流量	m^3	14, 852	
④総推定晴天時流量	m^3	13, 346	
⑤総雨天時浸入水量	m^3	1, 506	
雨天時浸入水量割合	―	0. 101	⑤／④

3. 脱水ケーキ含水率の推定事例（ニューラルネット法）

次を目的として脱水機の操作因子を説明変数として脱水ケーキ含水率(目的変数)をニューラルネットワークで解析する。本稿では、紙面の都合で以下の結果は示していない。

ⅰ. 脱水ケーキ含水率に影響を与える最適な操作因子を明らかにする。

ⅱ. 運転中の脱水機投入汚泥の性状や操作因子(ろ過速度、薬注率)から脱水ケーキ含水率を推定する。

下水処理場において約４年間、稼働日１日１回測定した脱水ケーキ含水率の推移を**図 3.1**(ピンク色)に示す。

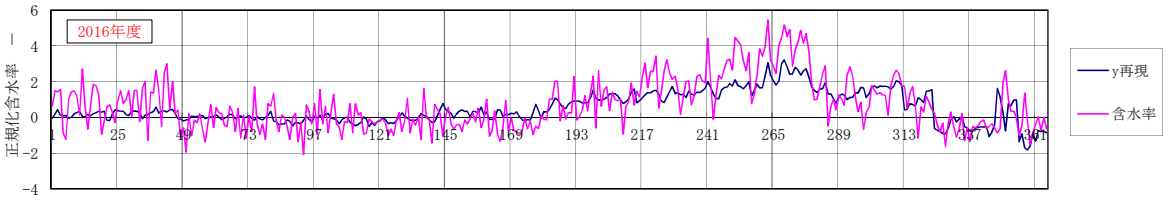

図 3.1 脱水ケーキ含水率と回帰脱水ケーキ含水率の推移

脱水ケーキ含水率(目的変数)に及ぼすと思われる以下に示す脱水機投入汚泥の性状や操作因子(ろ過速度、薬注率)などの管理情報(説明変数)を正規化してニューラルネットワークで解析する。管理情報から①気温、雨量、②PAC 注入量、③濃縮槽 pH、④生汚泥固形物割合、消化汚泥固形物割合、余剰汚泥固形物割合、⑤高分子凝集剤 13 種類、⑥高分子凝集剤注入率、⑦脱水機ろ過速度、⑧カレンダー情報(月、曜日)を抽出し説明変数する。

解析した結果は前述図 3.1 に併せて示すが、推定した脱水ケーキ含水率(青色実線)は、ばらつきのある計測した脱水ケーキ含水率(ピンク色)を再現できておらず、平均的な値を示している。ばらつきのある脱水ケーキ含水率と回帰値を一致させるためには上記①～⑧の以外の説明変数を探して加える必要がある。現状、新たな説明変数は見つかっていない。

4. 管路の劣化要因分析事例（ニューラルネットワーク法、ランダムフォレスト、SVM）

4.1 ニューラルネットワークによる劣化要因の推定

コンクリート管のＴＶカメラ調査診断結果である劣化程度を表す緊急度(Ⅰ,Ⅱ,Ⅲ,Ⅳ)に及ぼす影響のある要因(説明変数)をニューラルネットワーク法で推定する。

推定した要因(説明変数)を使い、自治体の持つ管きょの緊急度を推定し、緊急度の高い管きょから点検・調査を実施するなどの机上のスクリーニングとして利用する。

国総研管渠劣化データベース Ver.2 にある情報を使い、コンクリート管の緊急度(目的変数)と説明変数にする下記項目間の関係を解析する。

管渠劣化データベースにあるコンクリート管に関する説明変数は、次に示す管径から土被りの７項目がある。

管径　　：250 mm～1,000 mm
路線延長　　：5.36 m～64.03 m
管本数　　：3 本～30 本
取付け管本数：1 本～8 本
経過年数　　：7 年～35 年
歩車道区分　：歩道、車道
土被り　　：0.8 m～63.3 m

データベース中に説明変数と緊急度の記載があるスパン数を 415 スパン抽出し、このデータを解析する。隠れ層

表 4.1 調査診断緊急度と推定緊急度

スパン番号	調査診断緊急度	推定緊急度の確率			推定緊急度	推定判定
		Ⅱ	Ⅲ	Ⅳ		
1	Ⅲ	0.01157	0.82014	1.68E-01	Ⅲ	1
2	Ⅳ	0.04116	0.39599	5.63E-01	Ⅳ	1
3	Ⅳ	0.025315	0.333857	6.41E-01	Ⅳ	1
4	Ⅳ	0.015027	0.345599	6.39E-01	Ⅳ	1
5	Ⅳ	0.000876	0.446202	5.53E-01	Ⅳ	1
6	Ⅳ	5.35E-09	0.340125	6.60E-01	Ⅳ	1
7	Ⅳ	0.014906	0.44914	5.36E-01	Ⅳ	1
8	Ⅲ	0.022364	0.310713	6.67E-01	Ⅳ	0
9	Ⅳ	1.57E-07	0.79117	2.09E-01	Ⅲ	0
409	Ⅲ	0.109479	0.47431	4.16E-01	Ⅲ	1
410	Ⅲ	0.149534	0.68735	1.63E-01	Ⅲ	1
411	Ⅲ	5.85E-08	0.498566	5.01E-01	Ⅳ	0
412	Ⅲ	3.68E-06	0.714269	2.86E-01	Ⅲ	1
413	Ⅲ	2.7E-07	0.720947	2.79E-01	Ⅲ	1
414	Ⅲ	3.14E-08	0.606716	3.93E-01	Ⅲ	1
415	Ⅳ	8.79E-06	0.641503	3.58E-01	Ⅲ	0

推定判定
0：不一致
1：一　致

(中間層)のニューロン数を 4 とし、出力層に緊急度Ⅱ、Ⅲ、Ⅳ(健全)を配置して解析した。

解析した結果を**表 4.1** に示す。各スパンの緊急度に対して、出力は、各緊急度の確率値として示される。スパン番号 1 の出力結果は、緊急度Ⅱ＝0.012、緊急度Ⅲ＝0.820、緊急度Ⅳ(健全)＝0.168 であり、確率の最も高い緊急度Ⅲを推定緊急度とする。表 4.1 をとりまとめて**表 4.2** に示す。

調査診断の緊急度と推定した緊急度は、全 415 スパンに対して 60.0%しか一致していない。調査時における緊急度Ⅱのスパン数は 28 あるが、Ⅲ、Ⅳと推定してしまったスパン数はそれぞれ 14 あり、まったく推定できていない。

表 4.2 ニューラルネットワークによる緊急度推定結果

	推定結果		
	Ⅱ	Ⅲ	Ⅳ
Ⅱ	0	0	0
Ⅲ	14	167	54
Ⅳ	14	84	82
診断結果	28	251	136
的中率	0.000	0.665	0.603

総合的中率 0.600

表 4.3 ランダムフォレストによる緊急度推定結果

	推定結果		
	Ⅱ	Ⅲ	Ⅳ
Ⅱ	26	1	0
Ⅲ	2	250	3
Ⅳ	0	0	133
診断結果	28	251	136
的中率	0.929	0.996	0.978

総合的中率 0.986

4.2 ランダムフォレストによる劣化要因の推定

学習用のデータをランダムにサンプリングして多数の決定木を作成し、作成した決定木をもとに多数決で結果を決める方法である。精度、汎用性が高く扱いやすい分析手法と言われている。

4.1 と同じデータを用いコンクリート管のＴＶカメラ調査診断結果である緊急度(Ⅰ, Ⅱ, Ⅲ, Ⅳ)に及ぼす影響のある要因(説明変数)をランダムフォレストで推定する。

ランダムフォレストによる解析結果は**表 4.3** となる。調査時における緊急度Ⅱのスパン数は 28 であり、ランダムフォレストによる推定では、緊急度Ⅱと推定したスパン数が 26 で、誤って緊急度Ⅲとしたのが 2 スパンあり、緊急度Ⅱの的中率は 92.9%であった。同様に緊急度Ⅲは、ランダムフォレストで 251 スパン中 250 スパンを推定しており、的中率は 99.6%になった。ランダムフォレストでは緊急度Ⅱ、Ⅲ、Ⅳの総合的中率は 98.6%となり、ニューラルネットワーク法の 60.0%より高い的中率となった。

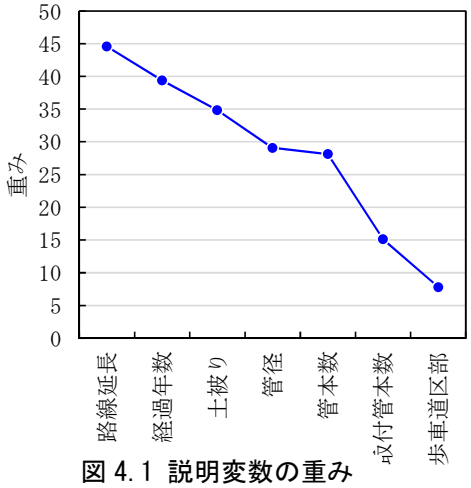

図 4.1 説明変数の重み

ランダムフォレストによる緊急度に及ぼす説明変数の重みは、**図 4.1** となり、スパン長の重みが大となり、つづいて経過年数、土被の重みが大となった。

表 4.4 原因施設または陥没位置集計結果

原因施設または陥没位置	件数	割合
本管	2,659	15%
取付管	9,109	53%
人孔	818	5%
桝	1,212	7%
本管と人孔の接続部	601	3%
本管と取付管の接続部	1,407	8%
取付管と人孔の接続部	130	1%
取付管と桝の接続部	897	5%
圧送管下流の本管	6	0%
圧送管下流の人孔	1	0%
圧送管本体	15	0%
不明	323	2%
合計	17,178	100%

文献[4]の**表 4.4** には取付け管は、「道路陥没の原因施設や陥没位置」に 53%も挙がっており、緊急度に及ぼす要因と考えられたが、取付け管本数の重

みが小さいのは意外な結果であった。

4.3 サポートベクターマシン(SVM)による劣化要因の推定

サポートベクターマシン(SVM)は、Support Vector Machine の略で教師あり学習に分類される。線形、非線形の識別関数があり現在知られている多くの学習モデルの中では最も優れた識別能力があるとされている[5)]。サポートベクターマシン(SVM)を使って 4.1 と同じデータを用い、解析する。解析結果は**表 4.5** となる。緊急度Ⅱ、Ⅲ、Ⅳを含めた全体の的中率は、78.1%であり、緊急度Ⅱの的中率は小さく、緊急度Ⅲの的中率は、91.6%であった。劣化程度を表す緊急度Ⅲ、Ⅳの的中率が高くても、緊急度の高い(Ⅱ)スパンを見つけるというような机上のスクリーニング技術としては、学習結果の利用方法はあまりない。

表 4.5 SVM による緊急度推定結果

	推定結果		
	Ⅱ	Ⅲ	Ⅳ
Ⅱ	3	0	0
Ⅲ	21	230	45
Ⅳ	4	21	91
診断結果	28	251	136
的中率	0.107	0.916	0.669

総合的中率　0.781

4.4 機械学習３手法のまとめ

管路の劣化要因の分析にはニューラルネットワークや SVM よりランダムフォレストの方が緊急度の高いスパンを的中させるという点で優位な手法であった。

5. まとめ

筆者らは人工知能といわれる機械学習を回帰や分類に使っている。雨天時浸入水量、脱水ケーキ含水率は、回帰の事例であり、管路の劣化要因の分析は分類に属するものである。学習技術は回帰、分類の結果に応じて使い分けて利用している。

機械学習で得られた説明変数の重みを使って自治体の持つ管きょのスパンの緊急度を推定し、推定緊急度の高い管きょから点検や調査を始めるというような机上のスクリーニングに使うことや所要の脱水ケーキ含水率を得るような操作因子を見出すことなどの利用を想定している。

機械学習は、用意したデータを入力すれば自動的に学習し、モデル化できるものではない。今後活用を勧めるためには、精度が高くなるような学習技術を選択するとともに入力データを取捨選択して、学習させる必要がある。

＜参考文献＞

１）人工知能学会 HP　学会概要　https://www.ai-gakkai.or.jp/
２）株式会社 SPJ：決定木の２つの種類とランダムフォレストによる機械学習アルゴリズム入門　https://spjai.com/regression-tree/
３）中根 進：ニューラルネットワークを用いた雨天時浸入水を含む流量の再現 下水道協会誌論文集 平成 20 年 8 月号
４）横田敏宏,深谷渉,宮本豊尚：下水道管路施設に起因する道路陥没の現状(2006-2009 年度)国土交通省 国土技術政策総合研究所 p.16
５）R 言語で SVM(Support Vector Machine)による分類学習
http://yut.hatenablog.com/entry/20120827/1346024147