サポートベクトル回帰に基づく被害関数の機械学習

岐阜大学工学部　正会員　○能島暢呂
東京瓦斯（株）　正会員　猪股　渉

1．はじめに

地震被害想定では様々な被害予測が統計的方法によって行われている．地震動モニタリング情報を入力条件として膨大な地盤・施設データと重ね合わせ，被害関数を適用する被害予測は，多くの場合システム化されている．一方，オンサイトでのリアルタイム制御においては CPU やメモリ等の資源制約を受ける．本研究は，サポートベクトル回帰(SVR)[1)]を用いて地震動情報と被害予測との入出力関係を機械学習することによって，即時的被害予測を簡略化することを目指した基礎的研究である．なお SVR とは，カーネル関数を用いて非線形データを効率よく分析できるカーネル多変量解析の一手法であり，画像・音声・言語などのパターン認識分野における分類法として発展したサポートベクトルマシン(SVM)の原理を回帰分析に拡張した手法である[1)]．

2.　SI 値および被災パターンのデータ生成

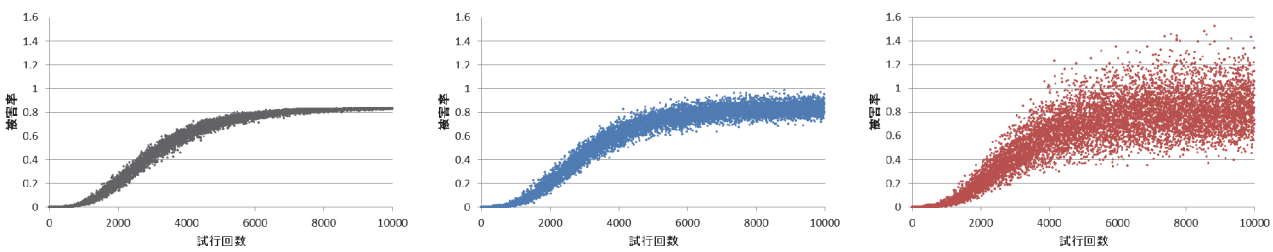
モンテカルロ・シミュレーション(MCS)に基づいて SI 値および被災パターンを生成する．その際，(1)表層地盤の SI 値増幅度の変動，(2)平均被害発生率に基づく被害箇所数の実現値の偶然的変動，(3)被害率関数自体の変動の 3 種類の不確定性を考慮する．まず基盤面 SI 値 SI_B は層別化して 5～100kine の範囲を試行回数 N(=10000)で等間隔に与える．SI_B は L ブロック内で均一とし，SI センサ位置の表層地盤の SI 値増幅特性を表す対数正規乱数を乗じて地表面 SI 値とする．L ブロック内のガス導管の本支管の被害箇所数 D_N は，標準被害率関数 ψ(SI)に管種補正係数 C_p と地盤補正係数 C_g を乗じた導管被害率 D_R に，導管延長 L を乗じて求める[2)]．

$$D_N = D_R \cdot L \qquad D_R = C_p \cdot C_g \cdot \varphi(SI) \qquad \varphi(SI) = R_0 \int_0^{SI} \frac{1}{\sqrt{2\pi}x\xi} \exp\left\{-\frac{(\ln x - \lambda)^2}{2\xi^2}\right\} dx \qquad (1)$$

ここに R_0=0.236, λ=4.298, ξ=0.387 である．L ブロック内には SI センサが複数設置されているため，M ブロック内の諸量を SI センサ個数で等分し，センサ位置→M ブロック→L ブロックの順に集計を進める．さらに，所与の被害発生率 λ (=D_N) のもとで被害がランダム発生すると仮定すると，確率変数としての被害箇所数 X は λ を母数としたポアソン分布に従う．また被害発生率のばらつきを考慮して，λ がガンマ分布 $f_\Lambda(\lambda;k,\theta)$に従うと仮定すると，被害箇所数の予測分布は負の二項分布 NB(x;k,θ)でモデル化できる．

$$\mathrm{NB}(x;k,\theta) = \int_0^\infty \mathrm{Po}(x;\lambda) f_\Lambda(\lambda;k,\theta) d\lambda = \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!} \cdot \frac{\lambda^{k-1}e^{-\lambda/\theta}}{\Gamma(k)\theta^k} d\lambda = \frac{\Gamma(x+k)}{x!\Gamma(k)} \cdot \left(\frac{1}{1+\theta}\right)^k \cdot \left(\frac{\theta}{1+\theta}\right)^x \qquad (2)$$

図 1 はある L ブロックを対象とした N=10000 回の被害率のサンプルである．SI 値増幅度のばらつきのみ考慮した場合，ポアソン分布に基づく場合，負の二項分布に基づく場合を比較した．なお(c)では δ_Λ=20%とした．



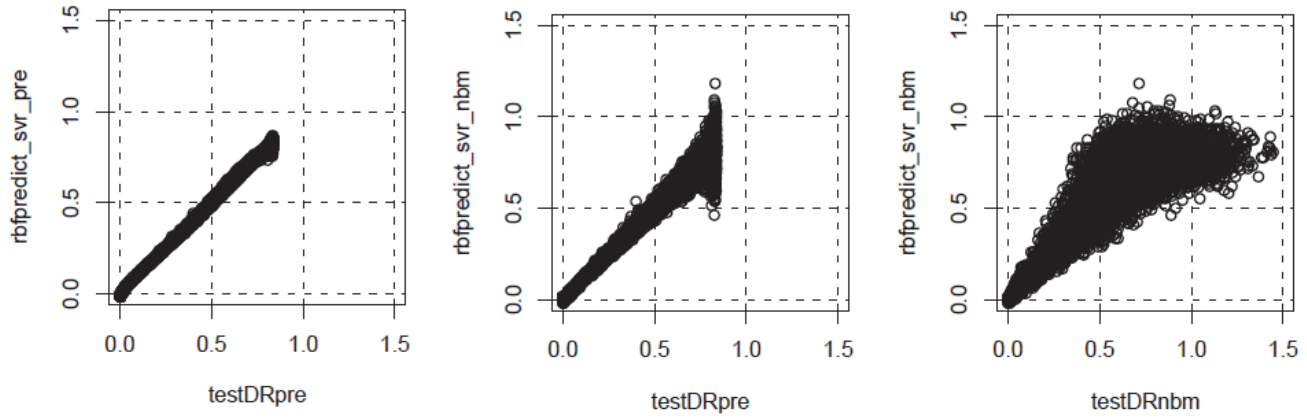
(a) SI 値増幅度のばらつきのみ考慮　　(b) ポアソン分布に基づく被害率　　(c) 負の二項分布に基づく被害率

図 1　被害率関数に不確定性を見込んだ場合の被害率（横軸は N=10000 の試行回数番号 i で SI_B=5～100kine に相当）

キーワード　被害関数，機械学習，サポートベクトル回帰
連絡先　　〒501-1193　岐阜市柳戸 1-1　岐阜大学工学部　Tel 058-293-2416

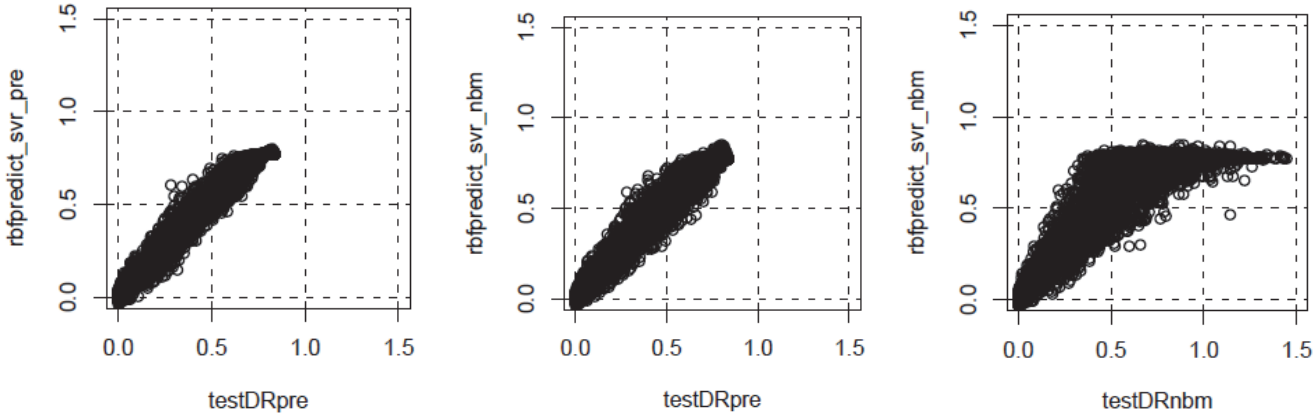
3. サポートベクトル回帰による学習と予測

分析手順は以下に示す通りである．まず MCS による 10000 回の試行による地表面 SI 値と導管被害率を，訓練データとテストデータに振り分ける．試行回数番号 i の剰余系を考え，$i \pmod m$=0, (m=20, 10, 5, 2) を満たすものを訓練データとする（各 500，1000，2000，5000 個）．訓練データの地表面 SI 値と導管被害率の関係をサポートベクトル回帰で分析し，残りのテストデータの地表面 SI 値から導管被害率を予測し，正解（テストデータの導管被害率）と比較する．カーネル関数としては RBF カーネル $K(\mathbf{x},\mathbf{x}')=\exp\{-\gamma\|\mathbf{x}-\mathbf{x}'\|^2\}$ を用いて，パラメータ γ=0.01～1.0 の範囲を対象とした．マージン幅を制御するための正則化係数については C= 1.0～50 の範囲を対象とした．分析には被害率のばらつきを考慮しない図 1(a)のデータセット A，負の二項分布に基づくばらつきを考慮した図 1(c)のデータセット C を用いる．予測パターンとしては，A を学習して A を予測（パターン AA），C を学習して A を予測（パターン CA），C を学習して C を予測（パターン CC）の 3 種類とした．検討の結果，1000～2000 回程度で高度な学習効果があること，正則化係数については対象範囲内で差がほとんどないことが確認された，ここでは学習回数 2000 回，C= 5 とする．図 2 に γ=0.02 とした場合の例を示す．パターン AA では予測誤差は小さく，式(1)の非線形な被害率関数を含む SI 値と被害率との入出力関係が適切に学習されている．パターン CA ではばらつきを含む訓練データのため，被害率の上限値付近の予測誤差が大きくなっているが，低い被害率ではさほど大きくない．ばらつきが二重化するパターン CC では予測誤差がさらに拡大している．図 3 は γ=0.5 としてカーネル関数の幅を狭めた場合である．全体的にばらつきは拡大するがパターン CA では一部縮小している．今後，さらに多様な条件で検討を進める予定である．



(a)パターン AA　(b)パターン CA　(c)パターン CC

図 2 サポートベクトル回帰による被害率の予測結果（γ=0.02，横軸：テストデータの実際の値，縦軸：予測値）



(a)パターン AA　(b)パターン CA　(c)パターン CC

図 3 サポートベクトル回帰による被害率の予測結果（γ=0.5，横軸：テストデータの実際の値，縦軸：予測値）

謝辞：本研究では東京瓦斯(株)の SUPREME に関するデータを使用させていただきました．関係各位に記して謝意を表します．

参考文献　1) 赤穂昭太郎：カーネル多変量解析―非線形データ解析の新しい展開（シリーズ確率と情報の科学），岩波書店，2008.
2) 猪股 渉・乗藤雄基・石田栄介・塚本博之・山崎文雄：東日本大震災における東京ガスの設備被害の概況と超高密度地震観測情報に基づく低圧ガス導管被害推定の精度検証，日本地震工学会論文集 第 13 巻，第 2 号（特集号），pp.2_37-2_44，2013.3.