

Support Vector Machines를 이용한 교각주위 국부세굴 예측

Prediction of Local Scour around Bridge Piers using Support Vector Machines

최성욱*, 최성욱**

Seongwook Choi, Sung-Uk Choi

요 지

교각 주위에서의 국부세굴은 교각을 지나는 유체의 복잡한 흐름에 의해 발생한다. 이를 해석하기 위하여 많은 난류모형을 이용한 실내실험 및 수치실험을 수행하였으나 발생하는 와류를 하천 규모에서 전부 계산하기는 매우 어려운 문제다. 따라서 국부세굴 관련으로 최대 관심사인 최대 세굴심은 인공지능 기술에 근거한 다양한 기법을 적용해 계산하여 예측하기도 한다. 본 연구에서는 기계학습 분야 중 하나인 서포트 벡터 머신 (Support Vector Machines)을 이용하여 교각주위 국부세굴을 예측하였다. SVM은 본래 초평면을 이용하여 데이터를 분류시키는 기법이나 Vapnik (1995)이 제안한 ϵ 서포트 벡터 회귀 (ϵ -support vector regression)방법을 통해 회귀분석에도 활용할 수 있게 되었다. 학습을 위해 Charbert and Engeldinger (1956), Shen et al. (1969), Jain and Fischer (1979), 그리고 Dey et al. (1995)의 실험 자료를 이용하였고 검증에 위해 Yanmaz and Altinbilek (1991)의 실험 자료를 이용하였다. 커널함수로는 다항식 함수와 방사 기저 함수를 이용하였고 각 계수는 적합한 값을 찾기 위해 시행착오법을 사용하였다. 민감도 분석을 통해 각 계수들 중 ϵ 의 변화가 결과에 가장 민감하게 변화를 일으키는 것을 확인하였고 검증 결과 SVM가 충분히 국부세굴을 잘 예측하는 것을 확인하였다.

핵심용어: 국부세굴, 최대 세굴심, 인공지능 기술, 서포트 벡터 머신, 민감도 분석

1. 서론

우리나라 대부분의 지역에서는 도로 및 철도로 이동할 수 있도록 하천이나 내륙-섬 사이의 해안을 횡단하는 교량이 많이 준공되었다. 현재 우리나라는 준공 후 30년이 경과한 교량이 증가하고 있는 추세이며 이에 대한 유지 및 관리가 주요 이슈로 떠오르고 있다. 교량의 유지관리 대상 부재 중 하나인 기초는 국부 세굴에 의한 노출이 주요 관심사로 국부세굴은 하천이나 해안에서 교각 주위에서의 복잡한 흐름에 의해 발생하게 된다. 평형 세굴심이 기초보다 깊지 않도록 설계해야 하는데 평형 세굴심을 해석하기 위해서는 3차원 난류해석과 유사이동이 필수적이지만, 이는 실제 하천이나 해안에서와 같은 규모에 적용하기에는 아직은 어려운 문제이다. 이 문제를 해결하기 위해 기존의 많은 연구자들은 실내 수로나 실제 규모의 교량에서 세굴심을 측정하여 세굴심에 영향을 미치는 인자들을 연구하고 그 경험식을 제시하였다. 실제 설계에서는 계산의 용이성과 신뢰성에 의해 경험식을 활용하는 경우가 많다. 그러나 각 경험식은 적용 가능한 인자의 범위가 한정되므로

* 연세대학교 대학원 토목환경공학과 석·박사통합과정 · E-mail : csu1220@hotmail.com

** 정회원 · 연세대학교 공과대학 토목환경공학과 교수 · E-mail : schoi@yonsei.ac.kr

적용 대상에 대한 충분한 적용성 검토가 필요하다. 세굴심을 예측하는 다른 방법으로는 기계학습 기법을 이용하는 것이다. 기존 연구에서는 인공신경망, 퍼지 논리, 유전자 알고리즘 등을 다양한 기계학습 기법을 이용하여 교각주의 국부세굴을 예측하였다. 기계학습 기법 중 Vapnik (1995)에 의해 제안된 서포트 벡터 머신 (Support Vector Machines, SVM)은 뛰어난 일반화 능력에 의해 패턴인식과 분류문제에 있어서 좋은 성과를 나타내고 있기 때문에 적합한 계수를 찾아내도록 하면 다른 기법보다 우수한 결과를 보인다고 한다. 따라서 본 연구에서는 기계학습 기법인 SVM을 이용하여 교각주위 평형 세굴심을 예측하고 각 계수의 민감도를 분석하였다.

2. 기본이론

2.1 교각주위 국부 세굴

기계학습을 수행하기 위해서는 가장 먼저 평형 세굴심에 영향을 미치는 변수를 결정할 필요가 있다. 유체의 흐름특성, 하상입자의 특성, 교각의 형태 특성에 대한 다양한 변수가 있으나 Sheppard et al. (2014)이 다양한 연구자들의 연구 결과를 분석한 결과 원통형 교각이 있는 모래하상의 하천일 경우 평형 세굴심에 대하여 다음과 같이 표현하였다.

$$d_{se} = f(V, V_c, b, y, d_{50}) \quad (1)$$

여기서 d_{se} 는 평형 세굴심, V 는 교각에 접근하는 유체의 평균유속, V_c 는 하상입자가 부유하는 임계속도, b 는 교각 폭, y 는 수심, d_{50} 은 하상입자의 대표입경이다. 각 인자들의 관계에 대하여 무차원화하여 관계식을 만드는 것이 일반적이지만, Choi et al. (2015)에 의하면 무차원변수를 이용하여 세굴심을 예측하는 것 보다 차원을 가지는 변수를 이용할 때 예측능력이 더 높은 것을 확인하였으므로 본 연구에서는 차원을 가지는 변수를 이용하여 평형 세굴심을 예측하였다.

2.2 SVM

SVM은 Vapnik (1995)이 제안한 기계학습 기반 분류기법이다. 소프트 마진, 커널 함수를 통한 비선형 분류 등으로 최적화된 분류기법으로 널리 사용되었다. 주어진 학습용 데이터 집합 안에서 ($D = [(x_1, y_1), \dots, (x_n, y_n)]$) 분류가 가능한 두 클래스 ($y_i = -1$ or 1)가 있을 때 두 클래스를 구분 가능한 초평면이 있다고 하면 다음과 같이 표현할 수 있다.

$$w \cdot x - b = 0 \quad (2)$$

여기서 w 는 초평면의 법선벡터이다. 식 (2)의 초평면과 가장 가까운 점에서 생성되는 초평면 두 개는 각각 $w \cdot x - b = 1$, $w \cdot x - b = -1$ 이 되므로 두 초평면의 거리는 기하학적으로 $2/\|w\|$ 가 된다. 서포트 벡터를 지나는 초평면 사이에는 데이터가 존재하지 않아야 하므로 다음과 같은 식이 성립된다.

$$\begin{aligned} w \cdot x_i - b &\geq +1, & \text{if } y_i = +1 \\ w \cdot x_i - b &\leq -1, & \text{if } y_i = -1 \end{aligned} \quad (3)$$

따라서 SVM 문제는 다음과 같은 최적화 문제로 표현할 수 있다.

$$\min \|w\| \quad \text{subject to } y_i(w \cdot x - b) \geq 1, \text{ for all } 1 \leq i \leq n \quad (4)$$

그 외에도 자료를 완벽히 분류할 수 없는 경우와 특정한 형태의 분류법이 필요한 경우를 고려하여 여유변수와 커널함수를 이용한다. 또한 Smola and Vapnik (1997)이 제안한 ϵ -support vector regression에 의해 회귀분석을 위한 최적화 문제를 적용하면 다음과 같이 표현할 수 있다.

$$\begin{aligned} \min & \left\{ \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right\} \\ \text{subject to} & \begin{cases} u_i - w \cdot x_i - b \leq \epsilon + \xi_i \\ w \cdot x_i + b - u_i \leq \epsilon + \xi_i^* \\ \xi_i \geq 0, \quad \xi_i^* \geq 0 \end{cases} \quad \text{for all } 1 \leq i \leq n \end{aligned} \quad (5)$$

여기서 u_i 는 학습된 값, ξ_i 는 여유변수, ϵ 은 u_i 와 초평면과의 허용오차, C 는 여유계수이다.

3. 모형의 적용

3.1 적용 대상

본 연구에서는 선행 실험 자료를 조합하여 학습 및 검증자료로 사용하였으며 네 개의 실험에서 측정된 99개의 자료에 대한 정보는 Table 1, 2와 같다. V_C 가 측정되지 않은 실험의 경우 Sheppard et al. (2014)이 제시한 식을 이용하여 산정하였다.

3.2 민감도 분석

SVM을 이용한 회귀분석의 성능은 계수 선택에 큰 영향을 받는다. 각 계수는 slack parameter (C), kernel parameter (d), 그리고 threshold parameter (ϵ)이다. kernel function은 가장 대표적인 radial basis function (RBF)을 사용하였다. 각 계수가 일정한 범위를 가지고 변동될 때, 평형 세굴심의 예측오류를 얼마나 크게 변동시키는지 검토하였다. C 는 너무 큰 값을 사용시 과적합을 발생시키므로 0.01에서 100 사이의 값을 변동시켰고 ϵ 은 상대오차 개념으로 하여 1부터 10^{-5} 까지 변동시켰다. 또한 d 역시 과적합을 발생시킬 수 있으므로 1에서 5 사이로 결정하였다.

3.3 적용 결과

Figure 1은 SVM을 이용하여 64가지 실험자료를 학습시킨 후 각각 학습자료를 적용한 예측결과와 검증자료를 적용한 예측결과를 도시한 것이다. MAPE는 각각 4%와 6%로 상당히 우수한 예측 성능을 보이고 있다. 검증자료로 사용된 결과의 MAPE가 작은 이유는 검증자료의 변수 범위가 학습자료의 범위 내에 있기 때문에 SVM에 의해 일반화되면서 오차가 적은 값을 가지게 된 것으로 판단된다. Figure 2는 SVM의 각 계수들이 변화할 때 검증자료의 예측 오차에 대한 민감도 분석을 수행한 결과이다. Figure 2 (a)를 보면 C 는 0.1부터 20사이가 적합한 것으로 나오는데, 이는 C 가 너무 작으면 최적의 초평면을 찾지 못하며 너무 크면 과적합에 의하여 다른 데이터에 적용 시 예측이 불가능하게 되는 문제를 발생시키기 때문으로 보인다. Figure 2 (b)를 보면 ϵ 은 0.01 이하부터가 적합한 것으로 예측하는데, 10^{-5} 보다 10^{-4} 에서 10^{-3} 사이가 적합한 양상이 보인다. 이것 역시 과적합과 어느 정도 관계가 있다. Figure 2 (c)를 보면 d 의 경우 특정한 규칙이 없이 진동하는 양상을 보인다. 이는 RBF 사용시 d 의 크기가 예측 정확도에 심각한 영향을 주지는 않는 것을 의미하는 것으로 판단된다. 따라서 C 와 ϵ 의 적절한 값 선정이 예측성능을 크게 결정하는 것을 알 수 있다.

4. 결론

본 연구에서는 SVM을 이용해 모래 하상에서 평형 세굴심을 예측하였다. 예측 결과 검증자료를 통한 예측시 MAPE가 6%로 우수한 성능을 보였으며, 또한 예측 성능은 SVM의 계수 C 와 ϵ 에 의해 민감하게 변동할 수 있음을 확인하였다.

감사의 글

이 논문은 2014년도 정부 (미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2014R1A2A1A11054236). 연구비 지원에 감사드립니다.

참고 문헌

1. Chabert, J. and Engeldinger, P. (1956). Etude des Affouillements Autour des Piles des Ponts. *Laboratoire Nationale d'Hydraulique*, Chatou, France.
2. Choi, S-U, Choi, B., and Lee, S. (2015). Prediction of Local Scour around Bridge Piers using the ANFIS Method. *Neural Computing & Applications*, 26(8), 1-10.
3. Dey, S., Bose, S.K., and Sastry, G.L.N. (1995). Clear Water Scour at Circular Piers: A Model. *IAHR, Journal of Hydraulic Research*, 15(4), pp.869-876.
4. Jain, S.C. and Fischer, E.E. (1979). Scour around Bridge Piers at High Froude Numbers. *Report FHWA-RD-79-104*, Federal Highway Administration, Washington D.C.
5. Shen H.W., Schneider V.R., and Karaki S. (1969). Local scour around bridge piers. *Journal of Hydraulic Division*, 95(HY5), 1919 - 1940.
8. Sheppard, D.M., Melville, B., and Demir, H. (2014). Evaluation of Existing Equations for Local Scour at Bridge Piers. *Journal of Hydraulic Engineering*, ASCE, 140(1), 14-23.

7. Smola, A. and Vapnik, V.N. (1997). Support vector regression machines. *Advances in neural information processing systems*, 9, 155–161.

6. Vapnik, V.N. (1995). *The nature of statistical learning theory*, Springer, New York.

Table 9. datasets for training and validation

Data source	used	number of data	variables
Chabert and Engeldinger (1956)	training	9	$y, V, V_C, b, d_{50}, d_{se}$
Dey et al. (1995)	training	18	$y, V, V_C, b, d_{50}, d_{se}$
Jain and Fischer (1979)	training	14	$y, V, V_C, b, d_{50}, d_{se}$
Shen et al. (1969)	training	23	$y, V, V_C, b, d_{50}, d_{se}$
Yanmaz and Altinbilek (1991)	validation	33	$y, V, V_C, b, d_{50}, d_{se}$

Table 10. ranges of variables used in the prediction

variable	training		validation	
	min	max	min	max
y (m)	0.035	0.671	0.045	0.165
V (m/s)	0.152	1.41	0.167	0.362
V_C (m/s)	0.213	0.64	0.282	0.371
b (m)	0.05	0.914	0.047	0.067
d_{50} (mm)	0.24	3.0	0.84	1.1
d_{se} (m)	0.026	0.671	0.032	0.107

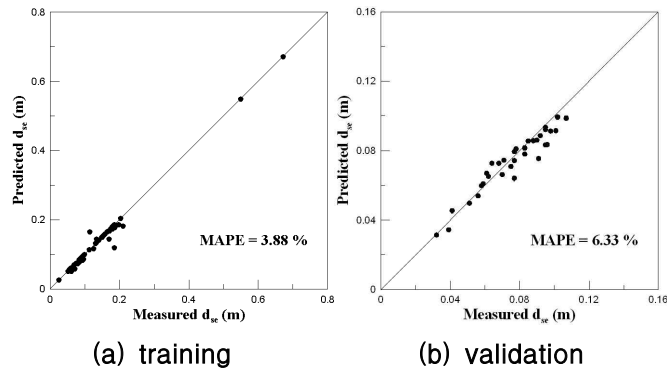


Figure 1. result of prediction using SVM

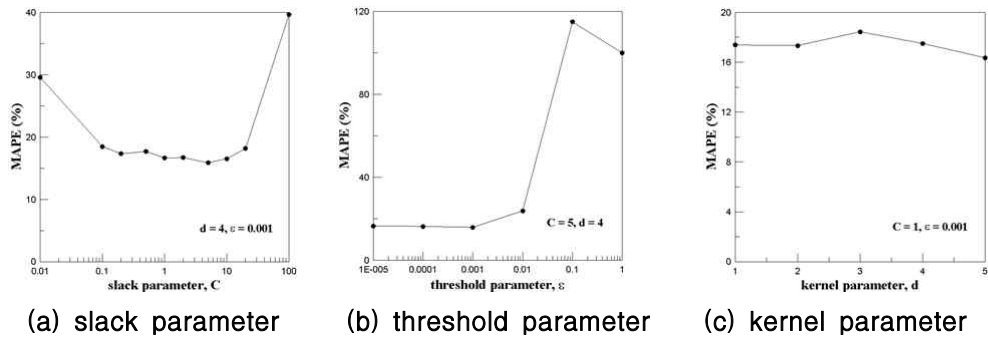


Figure 2. Sensitivity analysis