

Support vector regression과 최적화 알고리즘을 이용한 하천수위 예측모델

River stage forecasting models using support vector regression and optimization algorithms

서영민* · 김성원**

Youngmin Seo · Sungwon Kim

요 지

본 연구에서는 support vector regression (SVR) 및 매개변수 최적화 알고리즘을 이용한 하천수위 예측모델을 구축하고 이를 실제 유역에 적용하여 모델 효율성을 평가하였다. 여기서, SVR은 하천수위를 예측하기 위한 예측모델로서 채택되었으며, 커널함수 (Kernel function)로서는 radial basis function (RBF)을 선택하였다. 최적화 알고리즘은 SVR의 최적 매개변수 (C , cost parameter or regularization parameter; γ , RBF parameter; ϵ , insensitive loss function parameter)를 탐색하기 위하여 적용되었다. 매개변수 최적화 알고리즘으로는 grid search (GS), genetic algorithm (GA), particle swarm optimization (PSO), artificial bee colony (ABC) 알고리즘을 채택하였으며, 비교분석을 통해 최적화 알고리즘의 적용성을 평가하였다. 또한 SVR과 최적화 알고리즘을 결합한 모델 (SVR-GS, SVR-GA, SVR-PSO, SVR-ABC)은 기존에 수자원 분야에서 널리 적용되어온 신경망 (Artificial neural network, ANN) 및 뉴로퍼지 (Adaptive neuro-fuzzy inference system, ANFIS) 모델과 비교하였다. 그 결과, 모델 효율성 측면에서 SVR-GS, SVR-GA, SVR-PSO 및 SVR-ABC는 ANN보다 우수한 결과를 나타내었으며, ANFIS와는 비슷한 결과를 나타내었다. 또한 SVR-GA, SVR-PSO 및 SVR-ABC는 SVR-GS보다 상대적으로 우수한 결과를 나타내었으며, 모델 효율성 측면에서 SVR-PSO 및 SVR-ABC는 가장 우수한 모델 성능을 나타내었다. 따라서 본 연구에서 적용한 매개변수 최적화 알고리즘은 SVR의 매개변수를 최적화하는데 효과적임을 확인할 수 있었다. SVR과 최적화 알고리즘을 이용한 하천수위 예측모델은 기존의 ANN 및 ANFIS 모델과 더불어 하천수위 예측을 위한 효과적인 도구로 사용될 수 있을 것으로 판단된다.

핵심용어 : River stage forecasting, Support vector regression, Grid search, Genetic algorithm, Particle swarm optimization, Artificial bee colony

1. 서 론

인공지능 (Artificial intelligence) 또는 기계학습 (Machine learning) 분야에서 분류, 회귀 및 이상치 탐색기법으로서 알려져 있는 support vector machines (SVMs)은 최근 토양수분, 강우-유출, 하천유량, 저수지 수위 예측 등과 같은 수문학적 분야에서 그 관심이 증가하고 있다. SVMs은 신경망 (artificial neural network, ANN) 등과 같은 기존의 기법들에 비해 과대적합, 국소 최적해, 낮은 수렴률 등과 같은 문제 측면에서 상대적으로 우수한 것으로 알려져 있지만, SVMs의 학습 및 일반화 능력은 매개변수의 선택에 따라 매우 민감한 문제점을 가지고 있다. 따라서 SVMs의 최적 매개변수를 효과적으로 결정하기 위하여 매개변수 최적화 알고리즘 (Optimization algorithm)을 적용하는 것이 필수적이다. 따라서 본 연구에서는 효과적인 하천수위 예측을 위하여 SVR 및 매개변수 최적화 알고리즘을 이용한 하천수위 예측모델을 구축하고, 이를 실제 유역 (안동댐 유역, 도산 수위관측소 지점, 일수위 자료 적용)에 적용하여 모델 효율성을 평가하였다.

* 정회원 · 발표자 · 공학박사 · 경북대학교 과학기술대학 건설방재공학부 겸임교수 · E-mail : ymse0@knu.ac.kr

** 정회원 · 교신저자 · 공학박사 · 수자원개발기술사 · 동양대학교 철도토목학과 부교수 · E-mail : swkim1968@dyu.ac.kr

2. Support vector regression 및 최적화 알고리즘

2.1 Support vector regression

SVMs은 통계학적 학습이론 (Statistical learning theory)에 기초한 기계학습기법 (Machine learning method)으로서 주어진 자료에 대해 그 자료를 분리하는 초평면 (Hyperplane) 중에서 자료들과 가장 먼 초평면 (Hyperplane)을 찾는 개념에 기초하며, 회귀문제에 적용하기 위한 SVMs을 support vector regression (SVR)이라고 한다. SVR의 기본개념은 식 (1)과 같은 최적화 문제를 해석하는 것이다 (Sudheer et al., 2014).

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi + \xi^*) \\ & \text{subject to } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon_i + \xi \\ \langle w, x_i \rangle + b - y_i \leq \epsilon_i + \xi^* \\ \xi, \xi^* \geq 0 \quad i = 1, \dots, l \end{cases} \end{aligned} \quad (1)$$

2.2 최적화 알고리즘

SVR의 매개변수 최적화를 위한 알고리즘으로는 grid search (GS), genetic algorithm (GA), particle swarm optimization (PSO), artificial bee colony (ABC) 등이 적용될 수 있다 (Yoon et al., 2011; Yan et al., 2012; Yuan, 2012; Sudheer et al., 2014). GS는 매개변수 공간에서 기 설정된 부분집합을 통한 완전탐색을 수행하는 단순한 방법으로서 매개변수들에 대한 여러 조합을 생성한 후 교차타당성 분석을 통하여 정확도가 가장 높게 나타나는 매개변수를 최종 선택하는 방법이다. GA는 자연세계의 진화과정에 기초한 전역 최적화 기법 (Global optimization method)으로서 선택 (Selection), 교차 (Crossover), 변이 (Mutation), 대체 (Replacement) 등과 같은 주요 연산을 통해 최적 매개변수를 탐색하게 된다. PSO는 새 또는 물고기 무리와 같은 생체군집의 사회적 행동양식을 토대로 한 최적화 기법으로서 개별 해와 해집단을 각각 입자 (Particle)와 무리 (Swarm)로 표현하며, 각 입자는 최적 해를 얻기 위해 해공간을 이동하면서 입자 자신과 주변 정보를 이용하여 최적 해의 위치로 이동하게 된다. ABC는 꿀벌 무리의 수렵채집 거동에 근거한 최적화 알고리즘으로서 꿀벌의 군집을 고용벌 (Employed bees), 관망벌 (Onlookers), 정찰벌 (Scouts)로 구성하며, 각 꿀벌들의 역할을 통해 가장 많은 식량원, 즉 전역해 (Global solution)를 탐색하게 된다.

3. 적용 및 결과

하천수위 예측을 위한 SVR 및 매개변수 최적화 알고리즘의 적용성 평가를 위해 안동댐 유역의 도산 수위관측소에 대한 일수위 자료 (Training data, 2002-2010; testing data, 2011-2013)를 적용하였다. SVR의 커널함수 (Kernel function)로서는 radial basis function (RBF)을 선택하였다. 최적화 알고리즘은 SVR의 최적 매개변수 (C , cost parameter or regularization parameter; γ , RBF parameter; ϵ , insensitive loss function parameter)를 탐색하기 위하여 적용되었다. SVR과 최적화 알고리즘을 결합한 모델 (SVR-GS, SVR-GA, SVR-PSO, SVR-ABC)의 적용성을 평가하기 위하여 기존에 널리 적용되어온 신경망 (ANN) 및 뉴로퍼지 (ANFIS) 모델과 모델 효율성을 비교하였다. Table 1은 모델 효율성 평가지표를 비교한 것이고, Figure 2는 관측치와 예측치에 대한 산점도를 나타내고 있다. Table 1과 Figure 2로부터 모델 효율성 측면에서 SVR-GS, SVR-GA, SVR-PSO 및 SVR-ABC는 ANN에 비해 우수한 결과를 나타내었으며, ANFIS와는 비슷한 결과를 나타내었다. SVR-GA, SVR-PSO 및 SVR-ABC는 SVR-GS보다 상대적으로 우수한 결과를 나타내었으며, 모델 효율성 측면에서 SVR-PSO 및 SVR-ABC는 비교모델 중 가장 우수한 모델 성능을 나타내었다.

4. 결론

본 연구에서는 하천수위 예측을 위하여 SVR 및 매개변수 최적화 알고리즘 (GS, GA, PSO, ABC)을 이용한 하천수위 예측모델을 구축하고 이를 실제 유역 (안동댐 유역, 도산 수위관측소 지점, 일수위 자료 적용)에 적용하여 모델 효율성을 평가하였다. 그 결과 SVR-PSO 및 SVR-ABC는 비교모델 중 가장 우수한 모델 성능

을 나타내었으며, 이로부터 본 연구에서 적용한 매개변수 최적화 알고리즘은 SVR의 매개변수를 최적화하는데 효과적임을 확인할 수 있었다. SVR과 최적화 알고리즘을 이용한 하천수위 예측모델은 기존의 ANN 및 ANFIS 모델과 더불어 하천수위 예측을 위한 효과적인 도구로 사용될 수 있을 것으로 판단된다.

참 고 문 헌

1. Sudheer, C., Maheswaran, R., Panigrahi, B.K., Mathur, S. (2014). "A hybrid SVM-PSO model for forecasting monthly streamflow." *Neural Computing and Applications*, 24, 1381-1389.
2. Yan, Y., Zhang, Y., Gao, F. (2012). "Dynamic artificial bee colony algorithm for multi-parameters optimization of support vector machine-based soft-margin classifier." *EURASIP Journal on Advances in Signal Processing*, 2012(1), 1-13.
3. Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O., Lee, K.K. (2011). "A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer." *Journal of Hydrology*, 396, 128-138.
4. Yuan, F.C. (2012). "Parameters optimization using genetic algorithms in support vector regression for sales volume forecasting." *Applied Mathematics*, 3, 1480-1486.

Table 1. Performance evaluation

Models	CE	d	r^2	RMSE (m)	MAE (m)	MSE ($10^{-3}m^2$)	MSRE (10^{-7})	MS4E ($10^{-4}m^4$)
ANN	0.965	0.991	0.967	0.079	0.053	6.195	2.154	11.909
ANFIS	0.980	0.995	0.981	0.060	0.036	3.600	1.262	2.729
SVR-GS	0.980	0.995	0.981	0.060	0.036	3.569	1.240	3.632
SVR-GA	0.981	0.995	0.975	0.058	0.034	3.373	1.171	3.947
SVR-PSO	0.981	0.995	0.982	0.058	0.034	3.352	1.164	3.833
SVR-ABC	0.981	0.995	0.982	0.058	0.034	3.352	1.164	3.787

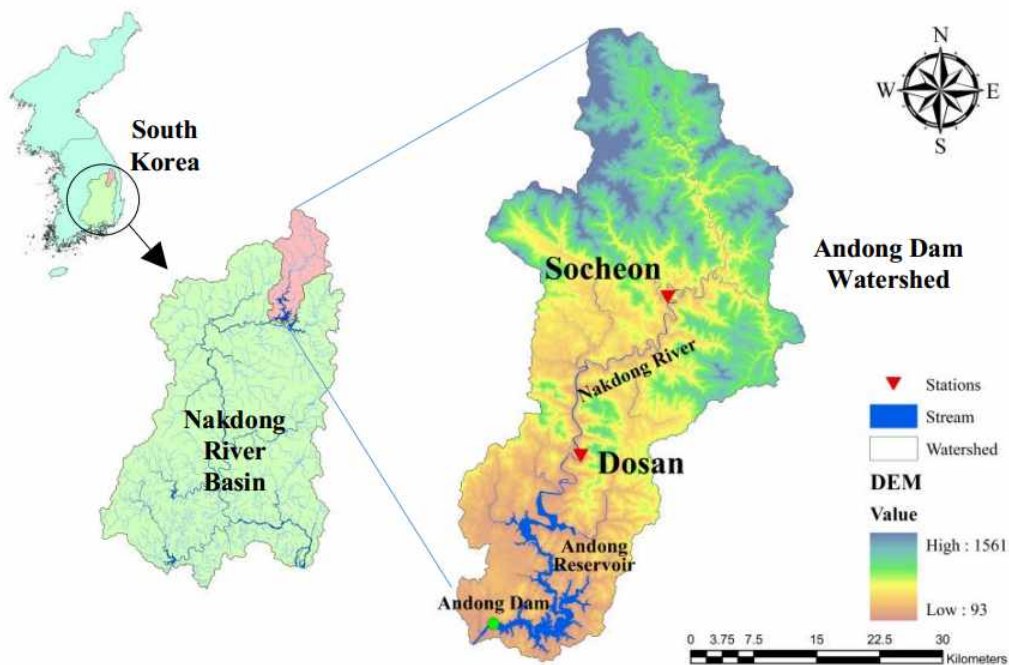


Figure 1. Study region and locations of observations

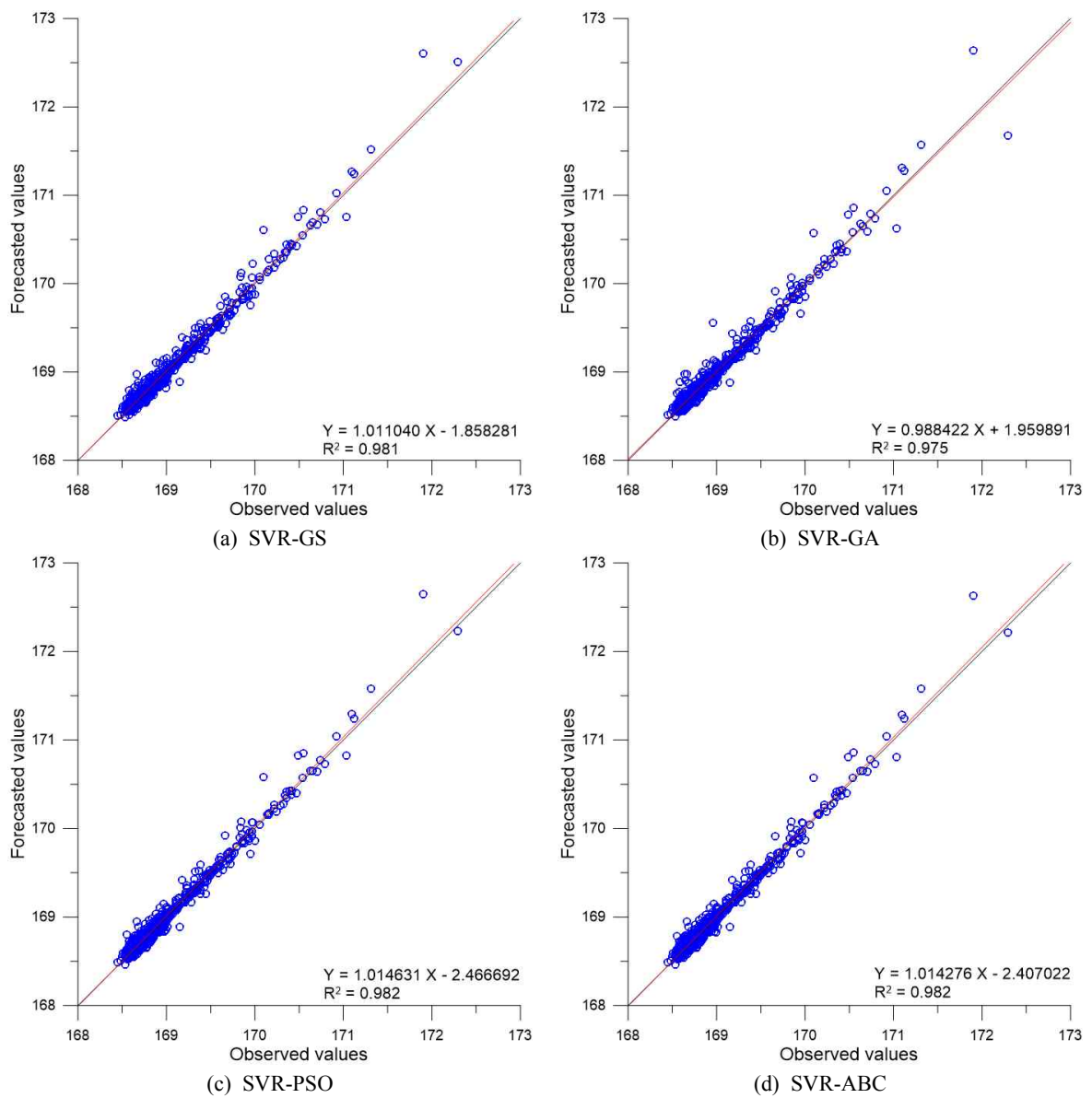


Figure 2. Scatter plots