

인공신경망모형과 군집분석을 이용한 교각 세굴심 예측

Prediction of Scour Depth Using Incorporation of Cluster Analysis into Artificial Neural Networks

이창환* · 안재현** · 이주현*** · 김태웅****

Lee, Chang-Hwan · Ahn, Jae-Hyun · Lee, Joo Heon · Kim, Tea-Woong

Abstract

A local scour around a bridge pier is known as one of important factors of bridge collapse. Two approaches are usually used in estimating a scour depth in practice. One is to use empirical formulas, and the other is to use computational methods. But the use of empirical formulas is limited to predict a scour depth under similar conditions to which the formulas were derived. Computational methods are currently too expensive to be applied to practical engineering problems. This study presented the application of artificial neural networks (ANN) to the prediction of a scour depth around a bridge pier at an equilibrium state. This study also investigated various ANN algorithms for estimating a scour depth, such as Backpropagation Network, Radial Basis Function Network, and Generalized Regression Network. Preliminary study showed that ANN models resulted in very wide range of errors in predicting a scour depth. To solve this problem this study incorporated cluster analysis into ANN. The incorporation of cluster analysis provided better estimations of scour depth up to 42% compared with other approaches.

Keywords : *scour depth, artificial neural network, cluster analysis*

요 지

교각주위의 국부세굴은 교량붕괴의 주원인 중 하나로 알려져 있다. 세굴심을 산정하는 방법에는 경험식에 의한 방법과 수치모형을 이용한 시뮬레이션이 있다. 하지만 경험식에 의한 방법은 공식이 적용될 수 있는 유사한 상황에서만 제한적으로 사용가능하며, 수치모형을 이용한 방법은 비용이 많이 든다는 단점을 가지고 있다. 그러므로 본 연구에서는 세굴심 예측을 위한 CSU 공식, 다중회귀분석, 다양한 인공신경망 모형의 유용성을 비교분석하였다. 또한 세굴심을 산정하는데 있어 넓은 범위의 오차를 발생시키는 인공신경망 모형의 단점을 보완하기 위하여 본 연구에서는 인공신경망 모형에 군집분석을 결합하여 오차를 감소시키고자 하였다. 세굴심 예측을 위해 CSU 공식, 다중회귀분석, 다양한 인공신경망 모형을 적용해 본 결과 역전파알고리즘을 이용하는 인공신경망 모형이 가장 높은 정확성을 보였으며, 인공신경망 모형에 군집분석을 적용한 세굴심 예측에서는 군집수가 3일 때 가장 높은 정확도를 보였다. 군집분석을 적용한 인공신경망 모형의 정확도는 다른 모형과 비교할 때 최고 42.73%가 향상된 결과를 보여 인공신경망 모형내의 군집분석의 적용이 인공신경망의 오차를 줄이는데 큰 역할을 할 수 있음을 알 수 있었다.

핵심용어 : 세굴심, 인공신경망, 군집분석

1. 서 론

교량의 교각이나 교대와 같은 수공구조물의 설치에 유수단 면적을 축소시켜 유속의 증가와 수위의 상승을 유발하며 구조물 전후 하상에 와류를 발달시켜 국부세굴을 발생시킨다(박성식과 송재우, 1999). 이러한 국부세굴은 홍수 시 붕괴를 일으키는 주요 원인이 되고 있으며, 따라서 교각주위의 세굴은 교량의 안전성 문제와 관련성이 높다. 또한 홍수로 인한 교량의 교각 및 교대 주위의 세굴로 인한 교량의 붕괴나 구

조적인 피해는 운전자들의 안전에 심각한 영향을 미칠 뿐만 아니라 장기간에 걸쳐 사회와 경제에 막대한 손실을 끼칠 수가 있다.

국부세굴에 의한 교각주위의 세굴이 교량붕괴의 가장 중요한 원인으로 밝혀지면서 이에 대한 연구가 수십 년간 활발히 진행되어져 왔다. 윤태훈과 윤성범(1995)에 따르면, 우리나라 하천은 암반층이 10-20 m 이내에 얕게 존재하므로 대부분의 주요 교량의 교각기초가 암반층 위에 직접 닿아 있어 교각세굴에 의한 교량 붕괴가 드문 편이나, 1995년 발생

*정회원 · 한양대학교 대학원 건설환경공학과 석사과정 (E-mail : chang810906@hanmail.net)

**정회원 · 서경대학교 토목공학과 조교수 (E-mail : wrr@skuniv.ac.kr)

***정회원 · 중부대학교 토목공학과 부교수 (E-mail : leejh@joongbu.ac.kr)

****정회원 · 교신저자 · 한양대학교 건설환경시스템공학전공 조교수 (E-mail : twkim72@hanyang.ac.kr)

한 중복선 화성철교 붕괴사고와 같이 우리나라도 교량붕괴 사고에 대한 안전지대가 아니다. 윤용남 등(1995)은 2차원 수리모형을 이용하여 세굴심 추정 정확도를 높이고자 하였으며, 유량증가에 따라 세굴심이 대체로 선형적인 증가를 보이고 교각폭이 클수록 세굴심이 증가한다는 것을 입증하여 우리나라 실정에 맞는 경험식의 필요성을 강조하였다. 박성식과 송재우(1999)는 연속교각 설치 시 최대세굴심의 변화는 만곡부 보다 직유로에서 크게 발생하며 이는 만곡부 흐름특성으로 인한 교각의 접근유향 때문에 접근유속의 감소가 작고 최대 퇴적고가 외측으로 편기되었기 때문이라는 것을 증명하였다.

이러한 연구와 더불어 중요시 되는 것이 세굴심 산정에 대한 연구이다. 기존의 세굴심과 교각 세굴심 산정은 CSU 공식, SHE 공식, HAN 공식, BRE 공식, JAF 공식 등과 같은 경험식에 의해 산정하는 방법과 수치모형을 이용한 방법이 있다. 경험식에 의한 산정방법은 대부분 세굴심에 영향을 주는 유속, 수심, 교각폭 등을 고려하였으며, 하상재질, 교각형상, 교각의 접근각 등 모든 인자를 동시에 고려하지 못하고 있다. 세굴심 예측공식으로 많이 사용되는 CSU 공식의 경우 많은 인자를 포함하고 있지만, 불확실성을 내포하고 있는 하상재질을 하상형태에 따라 몇가지로 분류하여 계수를 곱해주는 형태로 사용되고 있을 뿐 중앙입경이나 기하학적 표준편차 등의 영향을 정확히 반영하지 못하고 있는 실정이다. 또한 경험공식은 실험실에서 산정되었기 때문에 현장에서의 실측 세굴심 자료와 비교하면 많은 오차를 보이고 있으며, 유사한 상황에서만 제한적으로 적용가능하다는 단점을 가지고 있다.

강진구 등(2002)에서 경험식과 인공신경망(Artificial Neural Network, ANN)을 이용하여 국부세굴심의 정확도를 분석함과 동시에 교각 형상 및 접근각의 자료를 추가하여, CSU 공식과 Melville 공식에 적용한 결과 인공신경망의 정확도가 가장 높다는 것을 입증하였으며, Choi와 Cheng(2006)에서도 경험식과 인공신경망의 정확도를 비교하여 인공신경망의 정확도를 입증하였다. 이철웅(2000)은 임의의 분포함수를 따르는 확률변수간의 상관관계를 고려할 수 있는 직접계산법을 이용하여 교각 주위의 국부적인 세굴에 대한 신뢰성을 수행한 결과, Johnson(1992)이 CSU 공식을 개량한 경험식이 많은 불확실성을 내포하고 있다는 것을 입증하였다.

본 연구에서는 기존의 경험식 중 많이 사용되는 CSU 공식과 다중회귀분석 그리고 기존의 세굴심 산정 방법들의 단점을 보완하고 오차를 줄이고자 하였다. 이를 위해 변수들 간의 정확한 관계성 분석이 없이도 적용이 가능한 장점을 가지고 있는 역전파 알고리즘(Backpropagation Networks, BPN), 래디얼 기반 신경망(Radial Basis Function Network, RBFN), 일반화된 회귀신경망(Generalized Regression Networks, GRNN)와 같은 다양한 인공신경망을 이용하였다. 최성욱과 정상화(2003)는 인공신경망을 이용한 세굴심 예측시 실측자료에 대한 불확실성을 제거하기 위하여 난수를 발생시키는 방법을 이용하였다. 하지만 이 방법은 부족한 자료를 확률론적 개념으로 보강해야 한다는 번거로움과 실측자료가 아닌 임의의 자료를 사용해야 한다는 한계를 가지고 있다. 본 연구에서는 기존의 인공신경망에 관한 연구에서 나타난 오차

의 원인이 세굴심을 예측하는 인공신경망의 구조의 차이로 판단하고, 군집분석을 통하여 다수의 군집으로 그룹화한 후 인공신경망을 적용하여 그 결과를 기존의 세굴심 예측 모형과 비교해 보았다.

2. 연구방법

2.1 세굴심 자료

본 연구에서 이용된 세굴심 자료의 총수는 193개로 Jeng(2005)에서 사용되어진 자료를 사용하였다. 세굴심 예측을 위한 자료들이 국내에서는 충분히 축적되어져 있지 않고, 대부분이 수리모형을 통하여 획득된 자료이기 때문에, 기존에 사용된 세굴심 자료 중 세굴심을 산정하기 위한 많은 인자(그림 1)가 포함된 Jeng(2005)의 자료를 사용하였다. 또한 193개중 116개(60%)는 모형구축자료로 사용하였으며, 77개(40%)는 모형검증자료로 사용하였다. 표 1은 본 연구에 사용된 자료의 통계적 특성을 나타내고 있다. 평균입경(mean grain size, d_{50}), 평균유속(mean flow velocity, U), 한계유속(critical flow velocity, U_c)의 분산은 교각의 지름(pier diameter, D), 수심(flow depth, y), 세굴심(scour depth, d_{se} , d_{sc})에 비해 작은 값을 가지고 있어 자료 값들의 차이가 작음을 보여준다. 또한 6개의 입력 및 출력자료 모두 양의 왜도를 가지고 있음을 확인할 수 있다. 그림 2는 Jeng(2005)의 자료에 포함되어져 있는 세굴심과 예측인자들인 평균입경, 교각의 지름, 평균유속, 수심, 한계유속간의 상관관계(Correlation coefficient, CC)를 보여주며 5개의 인자들 중 교각의 지름이 가장 높은 상관관계를 보이고 있으나, 인자들 모두 세굴심과 비교적 높지 않은 상관관계를 가지고 있음을 확인할 수 있다.

2.2 CSU 공식

미국에서 가장 보편적으로 이용되고 있는 교각세굴방정식은 CSU 공식으로 미연방도로국(Federal Highway Administration, FHWA)에서 현장에서의 세굴조사, 세굴방지대책 수

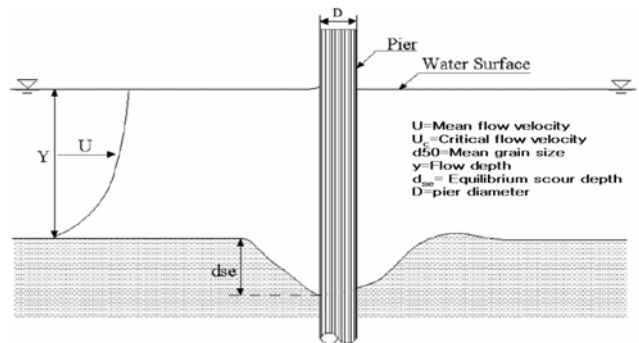


그림 1. 원형교각에서의 흐름과 국부세굴 (Jeng, 2005)

표 1. 자료의 통계적 특성

구분	입력 및 출력자료					
	d_{50} (mm)	D (mm)	U (m/s)	y (mm)	U_c (m/s)	d_{se} (mm)
평균	1.7733	156.0	0.4274	136.36	0.4692	114.55
분산	2.2447	31450	0.0734	12200	0.0346	4781.3
왜곡도	1.2127	2.0796	1.9917	2.3547	1.4345	1.0726

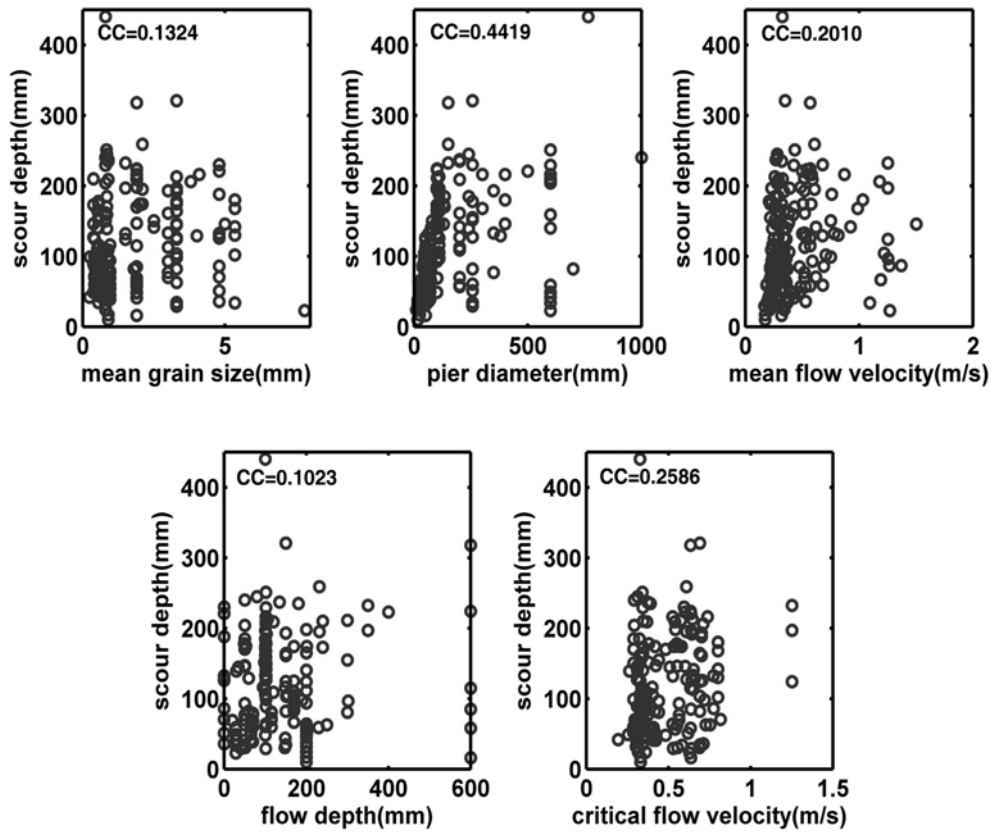


그림 2. 예측인자들과 세굴심의 상관분석

표 2. CSU 공식에 사용된 계수 값

계수	값
K_1 (교각형상에 따른 보정계수)	1.0
K_2 (흐름입사각에 대한 보정계수)	1.0
K_3 (하상조건에 대한 보정계수)	1.1
K_4 (하상재료의 크기에 대한 보정계수)	1.0

립 등을 위한 지침서로 발간된 HEC-18(Hydraulic Engineering Circular No.18)에서 추천되고 있으며, 우리나라의 하천설계 기준(한국수자원학회, 2005)에서도 국부세굴공식으로 명시되고 있다. CSU 공식의 산정식은 다음과 같다.

$$\frac{y_s}{y_1} = 2.0K_1K_2K_3K_4\left(\frac{a}{y_1}\right)^{0.65} Fr_1^{0.43} \quad (1)$$

여기서 y_s 는 세굴심도, y_1 은 구조물 설치 상류부 수심, K_1 은 교각형상에 대한 보정계수, K_2 는 흐름 입사각에 대한 보정계수, K_3 은 하상조건에 대한 보정계수, K_4 는 하상재료의 크기에 대한 보정계수, a 는 교각의 폭, Fr_1 은 구조물 설치 직상류부의 Froude 수이다. 식 (1)에 주어진 관련계수(K_n)는 하천설계기준(한국수자원학회, 2005)에 따라 표 2의 값을 적용하였다.

2.3 다중회귀모형

다중회귀모형은 k 개의 독립변수를 가지고 1개의 종속변수(Y)를 예측하는데 주로 적용되며, 일반적인 다중회귀모형은 식 (2)와 같다.

$$E(y_i | x_{1i}, \dots, x_{ki}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \quad (2)$$

종속변수(y_i)와 독립변수의 관계(x_i)는 부분회귀계수(Partial

regression coefficient, β_k)에 의해서 결정된다. 단순회귀모형의 회귀계수를 추정하는 방법과 동일하게 다중회귀분석에서도 최소제곱법과 최우도법을 사용하여 회귀계수를 추정할 수 있다. 최소제곱법에 의해 회귀계수를 추정하는 방법은 식 (3)의 잔차(e_i)들의 제곱합이 최소가 되도록 추정량을 구하는 방법을 취한다.

$$e_i = y_i - \hat{y}_i = y_i - (b_0 + b_1 x_{1i} + \dots + b_k x_{ki}) \quad (3)$$

그러므로 최소제곱추정값은 식 (4)의 조건을 만족시키는 값 b_0, b_1, \dots, b_k 이다.

$$\text{minimize} \sum (y_i - b_0 - b_1 x_{1i} - \dots - b_k x_{ki})^2 \quad (4)$$

2.4 인공신경망 이론

Haykin(1999)은 인공신경망이란 “자연적으로 경험적인 지식을 저장하고 유용하게 만드는 경향을 지닌 간단한 처리 단위로 구성된 거대하며 병렬적으로 분포된 연산 처리 장치”라고 정의하였다. 인공신경망의 여러 장점들 중에서 본 연구에 부합하는 세 가지는 다음과 같다. 첫번째로 인공신경망은 선형 또는 비선형적일 수 있으나, 뉴런들간의 상호 연결에 의한 인공신경망은 비선형적이며 이 비선형성은 인공신경망 전체에 걸쳐 분포되어진다는 것이다. 두번째로 인공신경망은 입력과 출력간의 사상에 학습을 한다는 것이며, 세번째로 인공신경망은 주위 환경에 적응되어지도록 연결강도를 조절할 수 있는 능력을 가지고 있다는 것이다.

인공신경망 모형의 가장 큰 특징인 학습이란 입력층(Input layer), 은닉층(Hidden layer), 출력층(Output layer)으로 구성된 다층신경망의 각 층 뉴런들의 연결강도를 최적의 상태

로 적응시키는 과정을 말한다. 인공신경망은 수학적 연산능력을 가진 뉴런들이 상호 연결된 것으로 적절한 학습규칙에 의하여 작동한다. 즉 각 뉴런들은 결합함수와 전달함수(활성화 함수)에 의해 수학적 연산을 수행한다. 본 연구에서 사용된 인공신경망의 종류는 역전파 알고리즘(BPN), 래디얼 기반 인공신경망(RBFN), 일반화된 회귀 인공신경망(GRNN)이며, 역전파 알고리즘(BPN)에 포함된 활성화 함수는 은닉층에서 탄젠트 시그모이드(tangent sigmoid)함수와 출력층에서의 선형(linear)함수를 사용하였다.

2.4.1 역전파 알고리즘 인공신경망(BPN)

역전파 알고리즘(BPN)은 다층(Multilayer)으로 구성되어 있으며 다층인식자(Multilayer Perceptron, MLP)모델 중 하나이다. 학습의 방법은 지도 학습(supervised learning)으로, 원하는 값(target data)을 얻기 위하여 입력자료(input data)에 적용되어지는 적절한 가중치를 찾는 과정이다. 인공신경망의 출력값(output data)는 오차($e = \text{output data} - \text{target data}$)를 가지며, 오차에 비례하여 출력층의 가중치를 갱신하고, 그 다음 은닉층의 가중치를 갱신한다. 은닉층의 출력은 식 (5)와 같다(김만식과 이영신, 2007).

$$t_j = \phi(r_j) = \phi\left(\sum W_{ij}x_i\right) \quad (5)$$

여기서 W_{ij} 는 입력층과 은닉층 사이의 연결강도, x_i 는 입력자료, t_j 는 출력뉴런에 의해 전달되는 은닉층의 출력이다. 출력층에서 조합함수(combine function)는 주어진 실제출력 $y^{(k)}$ 에 대한 접근 $Y^{(k)}$ 에 따라 가중되어 합해지며 식 (6)와 같다(안상진과 전계원, 2001).

$$y^{(k)} \approx Y^{(k)} = \sum_{j=1}^n c_j^{(k)} t_j = \sum_{j=1}^n c_j^{(k)} \phi\left(\sum W_{ij}x_i\right) \quad (6)$$

여기서 c_j 는 은닉층과 출력층 사이의 연결강도, W_{ij} 는 입력층과 은닉층 사이의 연결강도이다. 본 연구에 사용된 BPN의 구조는 그림 3과 같으며 세굴심을 예측하기 위한 5개의 입력데이터 값이 입력되면 가중치를 곱하고 더하는 과정을 거쳐 출력 값이 발생하면 다시 오차에 비례하여 출력층과 은닉층의 가중치를 갱신하여 오차를 줄여나감으로써 실측치와 근접한 세굴심 값을 얻게 된다.

2.4.2 래디얼 기반 인공신경망 (RBFN)

안상진과 전계원(2001)에 따르면 래디얼 기반 인공신경망(RBFN)은 Moody와 Darken(1989)에 의해 래디얼 함수(radial function)에 기반을 둔 네트워크를 제시함으로써 이

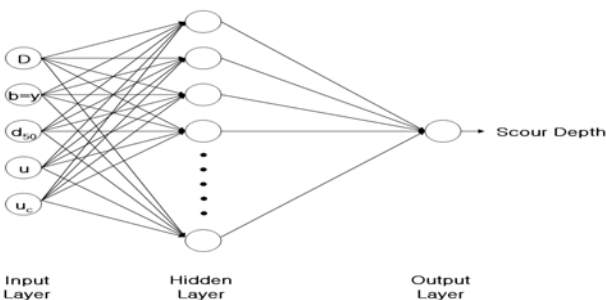


그림 3. 역전파 알고리즘 인공신경망(BPN)의 구조

에 대한 연구가 활발히 진행되어 왔다. RBFN은 학습속도가 빠르고 구성이 간단하며 분류능력이 우수한 장점이 있다. 그리고 은닉층에서 MLP가 시그모이드 함수(sigmoid function)를 사용하는 것에 반해 RBFN에서는 가우시안 함수(gaussian function)를 사용한다. RBFN의 학습 알고리즘은 대부분이 두 단계로 나누어 학습을 한다. 은닉층에서의 학습은 클러스터링 알고리즘을 이용한 자율학습(unsupervised learning)을 하고, 출력층에서의 학습은 은닉층의 학습이 끝나고 난 뒤 하게 되며 이때 은닉층 사이에서는 연결된 연결강도만 변하게 되고, 일반적으로 최소평균자승(Least Mean Squares, LMS) 알고리즘을 이용하여 학습을 하게 된다.

2.4.3 일반화된 회귀 인공신경망(GRNN)

안상진과 전계원(2001)에 따르면 일반화된 회귀 인공신경망(GRNN)은 입력층(input layer), 패턴층(pattern layer), 합산층(summation layer), 출력층(output layer)의 4개층으로 구성되어진 비선형 회귀이론을 기초로 한 순방향(feed-forward) 신경망이다. 대부분의 신경망처럼 반복적으로 학습하지 않는 대신에 패턴 층에서 각 입력패턴을 저장하고, 합산층에서 가중치들을 계산하는 방식으로 학습된다. GRNN의 각층별 특징은 다음과 같다. 첫번째 층은 입력층으로 패턴층과 완전 연결되어 있다. 두번째 층은 패턴층으로 각 입력패턴에 관해 하나의 뉴런을 가진다. 세번째 층은 합산에 관한 층으로 S-합산 뉴런과 단일 D-합산 뉴런의 두 가지 유형의 뉴런의 출력으로 나뉜다. GRNN의 마지막 층은 출력층이고, 각 S-합산 뉴런의 출력을 D-합산 뉴런의 출력으로 나뉜다. 또한 일반화된 회귀신경망의 출력은 식 (7)과 같은 조건부 평균이 된다.

$$\hat{Y} = \frac{\sum_{i=1}^N W_i \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^N \exp\left(-\frac{D_i^2}{2\sigma^2}\right)} \quad D_i = |T_i - X_n| \quad (7)$$

여기서 W_i 는 i 번째 training pattern의 결과, T_i 는 i 번째 training pattern, X_n 는 n 번째 test pattern, σ 는 너비상수를 나타낸다.

3. 연구결과

3.1 CSU 공식을 이용한 세굴심 산정

본 연구에서 사용된 CSU 공식의 계수의 값들은 표 2와 같다. 본 연구에서는 77개(40%)의 모형검증자료를 CSU 공식을 이용하여 예측값을 산정하고 정확도를 확인하기 위하여 최소자승평균오차(Root Mean Square Error, RMSE), 상관계수(Correlation Coefficient, CC), 평균절대백분비오차(Mean Absolute Percent Error, MAPE), 상대평균절대오차(Relative Mean Absolute Error, RMAE), 분산(Varinace) 등을 산정하였다. RMSE는 실측치와 예측치가 산포도 상에서 확산된 정도를 나타내는 측정치이며, CC는 실측치와 예측치의 상관관계를 나타낸다. MAPE는 예측오차와 실측치간의 상대적 비율을 퍼센트로 표시한 것이다. RMAE는 예측

오차의 평균을 실측치의 평균으로 나눔으로써 예측오차와 실측치간의 상대적 비율을 평균적인 개념에서 나타낸 것이다. VARIANCE는 실측치와 예측치 간의 분포양상을 나타내어주는 값이다.

$$RMSE = \frac{1}{n} \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (8)$$

$$CC = \frac{\sum_{k=1}^n (y_k - \bar{y}_k)(x_k - \bar{x}_k)}{\sqrt{\sum_{k=1}^n (y_k - \bar{y}_k)^2} \sqrt{\sum_{k=1}^n (x_k - \bar{x}_k)^2}} \quad (9)$$

$$MAPE = \frac{1}{n} \sum_{k=1}^n \frac{|x_k - y_k|}{x_k} \times 100 \quad (10)$$

$$RMAE = \frac{\frac{1}{n} \sum_{k=1}^n |y_k - x_k|}{\frac{1}{n} \sum_{k=1}^n x_k} \quad (11)$$

$$VARIANCE = \frac{1}{n} \sum_{k=1}^n (y_k - x_k)^2 - \left(\frac{1}{n} \sum_{k=1}^n y_k - x_k \right)^2 \quad (12)$$

여기서, x_k 는 세굴심의 실측치, y_k 는 모형의 세굴심 예측치이다.

3.2 다중회귀모형을 이용한 세굴심 산정

모형구축자료(116개)를 이용하여 다중회귀분석을 실시하여 회귀직선을 산정한 후 모형검증자료(77개)를 이용하여 회귀직선의 정확도를 검증하였다. CSU 공식의 정확도를 산정할 때와 동일하게 RMSE, CC, MAPE, RMAE, VARIANCE를 산정하였다. 표 3은 CSU 공식과 다중회귀분석의 정확도를 나타낸다. CSU 공식과 다중회귀분석의 정확도를 산정해본 결과 정확도를 판단하는 5가지 척도 모두에서 다중회귀분석이 CSU 공식보다 정확함을 알 수 있다.

표 3. CSU 공식과 다중회귀모형의 정확도

구분	Validation				
	RMSE	CC	MAPE	RMAE	VARIANCE
CSU 공식	125.05	-0.0074	116.43%	0.7575	14837
다중회귀모형	73.409	0.4004	72.56%	0.4722	5040

3.3 인공신경망을 이용한 세굴심 산정

인공신경망 모형 중 BPN, RBFN, GRNN을 사용하여 세굴심을 예측하였다. Jeng(2005)에서는 세굴에 가장 많은 영향을 주는 인자인 평균입경, 교각의 지름, 평균유속, 수심, 한계유속을 기본인자로 선정하여 입력변수의 수를 변화시켜본 결과 입력층의 노드(neuron)의 수가 5개일 때 가장 정확한 예측값을 보여주었다. 따라서 본 연구에서는 위의 연구결과를 토대로 입력층의 노드 수는 세굴심에 영향을 주는 5개의 인자로, 출력층의 노드수는 1개 인자를 사용하였으며 은닉층의 노드수는 3, 5, 8, 10, 15, 20개로 변화시켜가며 Case 1~6으로 분류하여 결과를 정리하였다. 또한 전처리 과정으로 인공신경망의 입력 및 출력변수들을 표준화시켜 적용하였다. 각 인공신경망 모형에 동일하게 반복횟수는 10000번 반복하고 학습률은 0.6, 시스템오차는 0.01로 고정하였다.

각각의 Case 마다 모형구축자료 116개(60%)를 이용한 모형구축을 통해 인공신경망을 구축하고 RMSE, CC를 산정한 후 구축된 인공신경망을 이용하여 모형검증자료 77개(40%)의 자료를 이용한 모형검증을 실시하였다. 모형검증 또한 RMSE, CC를 산정하여 Case별 정확도를 판단하였다. 표 4은 각 Case별 모형구축과 모형검증자료와 예측자료간의 RMSE, CC값을 나타낸다. 또한 표 5는 모형검증에 가장 정확한 정확도를 나타낸 Case에 대하여 모형검증에 대한 RMSE, CC, MAPE, RMAE, VARIANCE 등을 산정한 표이다. 인공신경망의 정확도를 산정한 결과 BPN의 은닉층의 수가 3일 때 가장 높은 정확도를 보였다. 하지만 BPN의 은닉층의 수가 3일 때 모형을 구축하고 모형검증 자료를 이용하여 정확도를 산정해본 결과, 모형구축자료를 이용하여 정확도를 산정한 결과보다는 정확성이 떨어졌다. 이는 실측

표 4. 은닉층 뉴런수에 따른 예측 결과

구분	BPN 은닉층 뉴런수	Training		Validation	
		RMSE	CC	RMSE	CC
Case1	3	43.605	0.7091	67.209	0.4992
Case2	5	40.136	0.7691	76.121	0.4295
Case3	8	43.8301	0.7165	75.2271	0.3624
Case4	10	31.5595	0.8824	93.9340	0.4080
Case5	15	32.6004	0.8720	104.0642	0.2998
Case6	20	29.3504	0.8888	106.2466	0.3924
	GRNN	60.7984	0.4662	78.7537	0.5275
	RBFN	1.9027	0.9995	39516	0.1095

표 5. 검증자료에 대한 정확도 (BPN(Case 1))

구분	BPN 은닉층 노드수	Validation				
		RMSE	CC	MAPE	RMAE	VARIANCE
Case1	3	67.209	0.4992	66.82%	0.4107	4505.8

자료의 부족과 본 연구에서 사용된 자료들과 세굴심과의 상관성 부족이 원인인 것으로 판단된다. GRNN은 모형구축자료를 이용하여 정확도를 산정한 결과와 모형검증 자료를 이용하여 정확도를 산정한 결과가 비슷한 값을 보였으며, RBFN은 모형구축자료를 이용한 정확도 산정에서는 가장 정확한 결과를 보이고 있으나 과적합(overfitting)되어 모형검증 자료를 이용한 정확도 산정에서는 가장 낮은 정확도를 보였다.

3.4 군집분석을 적용한 인공신경망을 이용한 세굴심 산정

세굴 영향인자에 대한 자료들은 확률론적 특징을 가지고 있기 때문에 수리수문학적 관측오차를 포함할 수 밖에 없으며 세굴에 관한 실측자료는 대부분 충분하지 못한 경우가 많다(최성욱과 정상화, 2002). 기존의 연구결과를 세굴 영향인자 및 실측자료가 부족할 시 난수를 발생시켜 자료를 생성하는 방법을 적용하였다(최성욱과 정상화, 2002). 하지만 본 연구에서는 실측자료를 임의로 새롭게 생성하지 않고 문제를 해결하기 위하여 인공신경망을 적용하기 전에 K-means

방법을 이용한 군집화를 실시하는 방법을 적용하였다. K-means 방법은 군집분석(Cluster analysis)에서 널리 사용되는 방법으로 K는 군집의 수 또는 중심점의 수를 의미한다. 사용자의 의해 K개의 중심이 정해지면 각 점들은 가장 가까운 중심점에 할당되어 K개의 군집을 형성한다. 각 군집안의 점들은 군집의 중심점으로부터 최대한 가깝게 형성하며 각각의 군집사이에는 최대한 멀게 형성한다. 각 군집의 중심점들은 매번 실행 시 마다 변화하며 군집화된 점들의 평균값을 의미한다. 이 과정을 중심점이 변하지 않을 때 까지 반복하여 최종 군집을 형성한다.

모형구축자료(116개)를 K-means 군집분석을 이용하여 5개 이상으로 군집화할 경우는 각 군집의 데이터의 수가 과도하게 적어지는 경향을 보임으로 K=2, 3, 4로 군집화를 실시하였다. 군집화한 자료들은 군집별로 인공신경망을 통하여 은닉층의 노드수를 3, 5, 8, 10, 15, 20으로 변화시켜가며 학습과정을 거쳐 각각의 모형을 형성한다. 위의 인공신경망 모형과 동일하게 반복횟수는 10000번, 학습률은 0.6, 시스템

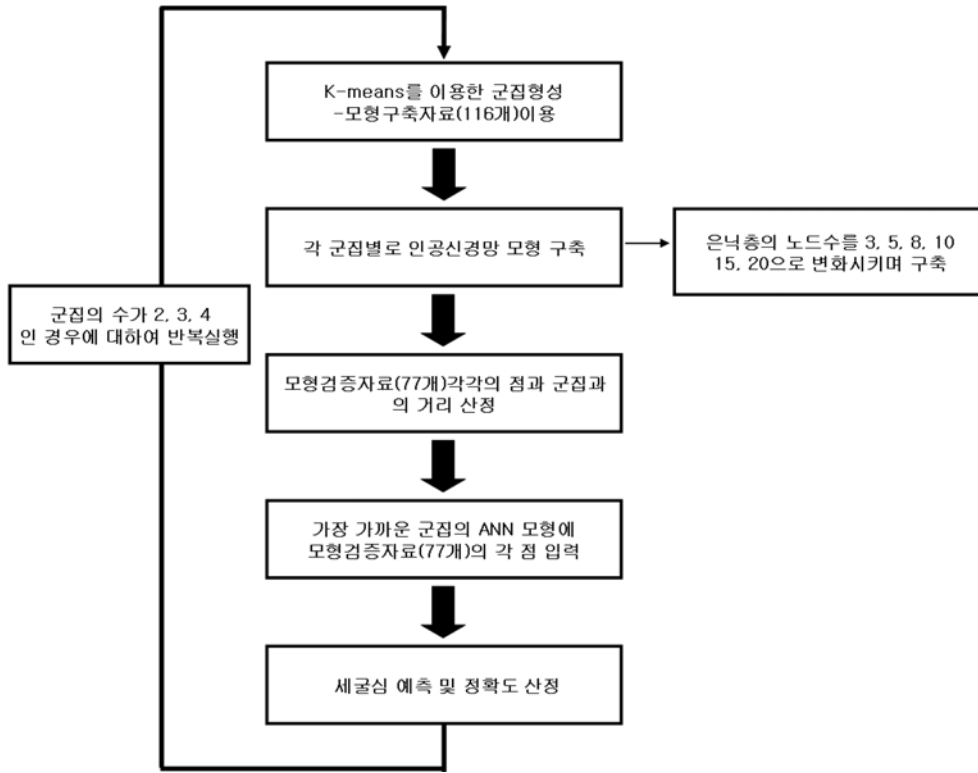


그림 4. 군집분석을 적용한 인공신경망을 이용한 세굴심 예측 과정

표 6. 군집화한 자료의 통계적 특성

구분		입력 및 출력자료					
		d ₅₀	D	U	y	U _c	d _{sc}
Cluster 1	평균	1.381	105.98	0.3687	87.371	0.3964	98.305
	분산	1.8632	5944.3	0.0308	917.43	0.0228	3083.4
	왜곡도	1.9635	1.0311	1.9382	-0.346	1.7337	1.0941
Cluster 2	평균	0.8661	60.54	0.4906	219.09	0.4336	102.48
	분산	0.2122	1842	0.1471	10026	0.0588	3542.2
	왜곡도	1.3399	1.777	1.4087	2.8521	2.762	0.8335
Cluster 3	평균	2.4448	530	0.6103	86.595	0.4935	142.05
	분산	4.1956	13150	0.0922	2480.7	0.0325	5370.1
	왜곡도	0.9345	-0.615	0.9631	1.1753	0.0642	-0.260

표 7. 군집분석을 적용한 인공신경망의 정확도

구분	은닉층 뉴런수	군집수					
		2		3		4	
		RMSE	CC	RMSE	CC	RMSE	CC
Case1	3	69.5426	0.5533	38.4862	0.8798	84.4921	0.4775
Case2	5	77.8549	0.5169	41.6776	0.8628	89.8079	0.4374
Case3	8	70.9898	0.6306	62.2505	0.7710	138.4459	0.4491
Case4	10	74.4680	0.6226	65.6058	0.7660	125.1256	0.5150
Case5	15	85.7535	0.6168	51.8854	0.8396	129.253	0.3041
Case6	20	149.5809	0.4244	104.0501	0.6564	249.0988	0.4178

표 8. 검증자료에 대한 정확도 (Case 1-2)

구분	은닉층 뉴런수	군집수: 3				
		RMSE	CC	MAPE	RMAE	VARIANCE
Case1-2	3	38.4862	0.8798	30.082%	0.1455	1412.2

오차는 0.01로 고정하고 모형을 형성하였다. 모형을 완성한 후 모형검증자료(77개)의 각점과 모형구축자료(116개)를 군집화 한 군집의 중심점과의 거리를 산정하여, 가장 가까운 군집의 ANN모형의 입력 층 자료로 모형검증자료(77개)의 각점을 입력한다. 결과 값으로 나온 예측치와 실측치를 이용하여 모형검증의 Case별 정확도를 판단한다.

그림 4은 군집분석을 적용한 인공신경망을 이용한 세굴심 예측과정을 나타내 주고 있으며, 표 6은 가장 정확한 결과를 보인 군집수가 3일 때의 군집별 데이터의 통계적 특성을 보여준다. 입력 및 출력자료의 군집별 분산값들을 표 1의 분산값들과 비교해 볼 때 분산값들이 감소함을 확인할 수 있다. 이러한 현상과 군집들의 중심과 가까운 검증 데이터들을 이용하여 검증을 실시하였기 때문에 군집분석을 적용하여 인공신경망을 구축하고 검증하는데 있어 정확성을 높인 것으로 판단된다. 표 7은 군집분석을 적용한 인공신경망의 정확도를 보여주며 Case1에서 군집수가 3일 때(Case 1-2)의 정확도가 가장 높음을 확인할 수 있다. 표 8은 가장 정확도가 높은 Case인 은닉층의 뉴런수가 3이고 군집수가 3일 경우(Case 1-2)에 대해서 RMSE, CC, MAPE, RMAE, VARIANCE를 산정한 결과이다. 표 5의 값들과 비교해 볼 때 군집분석을 적용한 인공신경망의 정확도가 증가하였음을 알 수 있다.

4. 결과 분석 및 비교

모형검증자료(77개)를 이용하여 CSU 공식과 다중회귀분석의 RMSE, CC, MAPE, RMAE, VARIACE 값을 산정한 결과 다중회귀분석을 이용한 방법이 CSU 공식을 이용한 방법보다 5가지 지표 모두에서 더 나은 결론을 얻을 수 있었다. BPN를 이용한 Case별 RMSE값을 비교분석한 결과 은닉층의 노드수가 3인 Case에서 가장 낮은 RMSE값을 나타내었다. 은닉층의 노드수가 3일 때 BPN를 이용하여 산정한 RMSE값이 CSU 공식을 이용한 RMSE값보다 46.25% 감소된 결과를 보였으며, 다중회귀분석의 RMSE값보다 8.45% 감소된 결과를 보였다. 하지만 다중회귀분석은 BPN의 은닉층의 노드수가 3일 때(Case2)를 제외한 나머지 경우보다는 정확한 결과를 나타내었다. 그림 5은 CSU 공식과 다중회귀분석을 이용한 산포도를 나타내고, 그림 6은 BPN의 각 Case별 산포도를 나타낸다. 그래프의 X축은 세굴심의 실측치(Observed values)를 나타내며 Y축은 세굴심의 예측치(Predicted values)를 나타낸다.

GRNN의 RMSE와 CC는 BPN의 Case4, Case5, Case6 보다는 정확성이 있지만 Case1, Case2, Case3 보다는 정확성이 낮다. RBFN는 훈련과정에서는 다른 모형들과 비교하여 월등한 결과를 보이고 있으나 검증과정에서 산정한

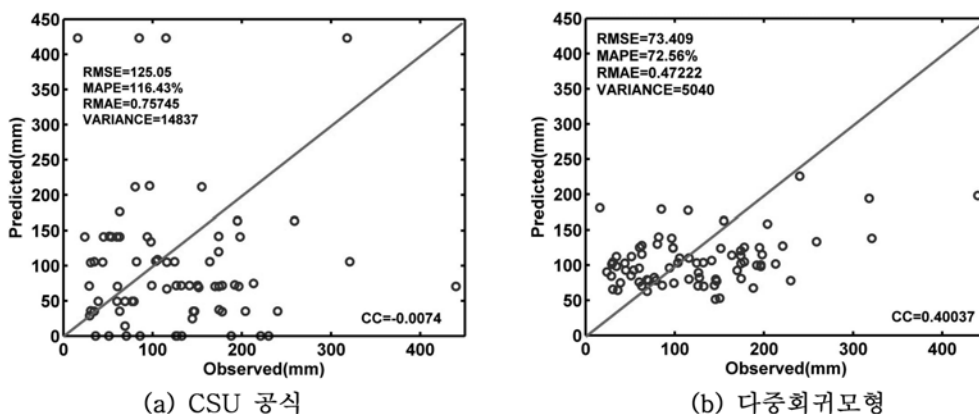


그림 5. CSU 공식 및 다중회귀모형에 대한 산포도

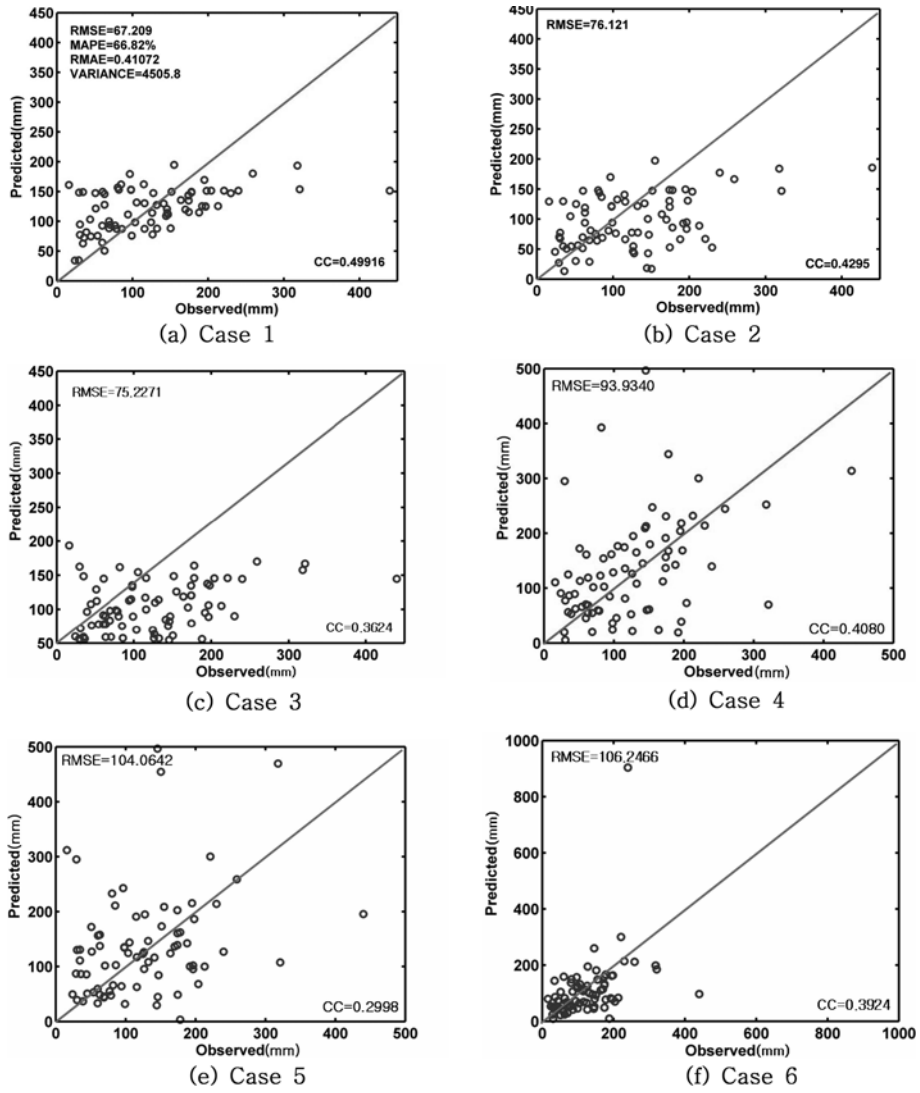


그림 6. 은닉층 뉴런수에 따른 산포도

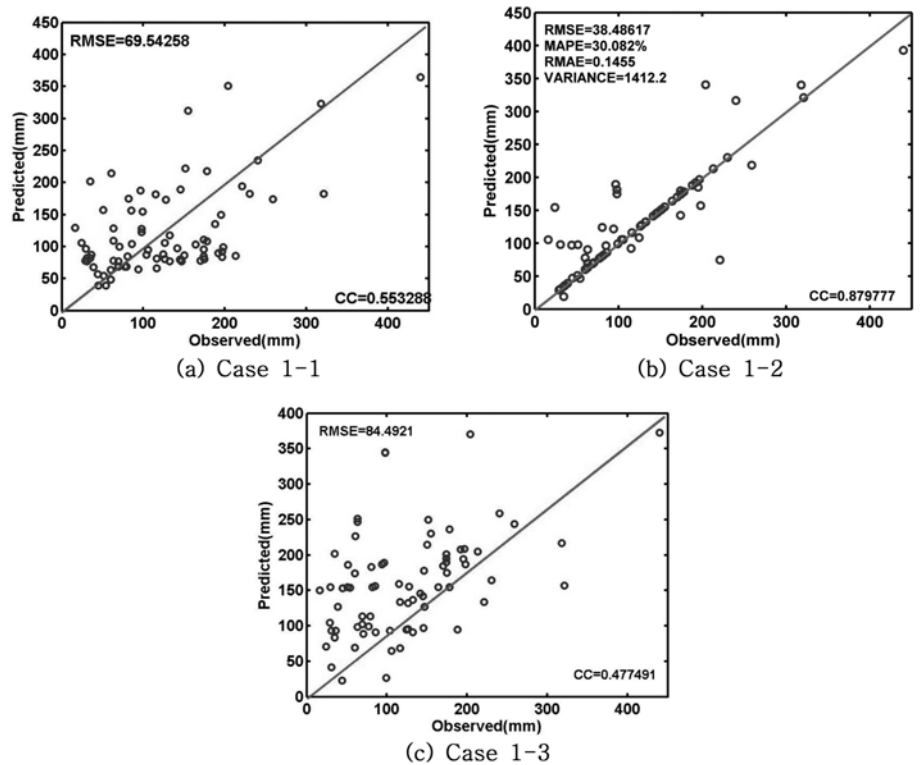


그림 7. K-means 군집분석을 적용한 신경망 모형의 산포도 (a) Case 1-1(은닉층 뉴런수: 3, 군집수: 2) (b) Case 1-2(은닉층 뉴런수: 3, 군집수: 3) (c) Case 1-3(은닉층 뉴런수: 3, 군집수: 4)

표 9. 세굴심 예측 방법별 정확도 비교

구분	정확도				
	RMSE	CC	MAPE	RMAE	VARIANCE
C.S.U	125.05	-0.00744	116.43%	0.75745	14837
다중회귀모형	73.409	0.40037	72.56%	0.47222	5040
역전파알고리즘 (Case1)	67.209	0.49916	66.82%	0.41072	4505.8
군집분석을 적용한 인공신경망 (Case1-2)	38.486	0.879777	30.082%	0.1455	1412.2

RMSE와 CC는 본 연구에서 실시한 방법 중 가장 낮은 정확도를 보이고 있다. 이러한 원인은 Case4, Case5, Case6에서도 확인할 수 있듯이 훈련과정에서의 과적합(Overfitting)이 검증에서의 부정확성을 증가시키는 것으로 판단된다. 과적합된 인공신경망은 훈련에 사용된 자료에 대한 오차는 매우 작은 값을 얻도록 해주지만, 그 인공신경망에 제공되는 새로운 자료에는 오차가 더 커질 수 있다(이성진 등, 2002). 그러므로 인공신경망의 종류 중 BPN이 가장 높은 정확도를 보이고 있음을 알 수 있다.

K-means 군집분석을 적용한 ANN을 이용하여 모형검증을 실시한 결과 BPN과 동일하게 은닉층의 노드수가 3인 Case에서 가장 낮은 RMSE 값을 얻을 수 있었다. 또한 군집의 수를 3으로 했을 때 가장 정확도가 높음을 확인할 수 있으며, 그 값은 BPN의 Case1보다 42.73% 상승한 결과를 나타내었다. 그림 7는 은닉층의 뉴런수가 3일 때 군집의 수에 따른 산포도를 나타낸다. 군집분석을 실시한 인공신경망의 적용이 정확도를 측정하는 모든 값에서 가장 좋은 결과를 나타내는 것을 알 수 있다. 또한 실측치와 예측치를 이용하여 나타낸 산포도의 그림에서도 GRNN과 RBFN 방법은 세굴심 예측에 적합하지 않으며 K-means 군집분석을 적용한 ANN 모형 Case 1-2는 BPN의 Case1 보다 42.73%, GRNN 보다 51.13%, RBFN 보다 99% 증가한 결과 값을 보이고 있다.

5. 결 론

본 연구에서는 기존 연구방법과는 달리 관측자료를 이용하여 군집분석을 이용한 인공신경망의 적용을 통하여 그 정확도를 높이는 방법을 제시하였다. 본 연구에서 제시한 세굴심 예측방법들의 정확도를 비교한 결과 결론은 다음과 같다.

1. 다중회귀분석을 적용한 결과 CSU 공식 보다는 정확도가 높았으나, ANN 모형 중 BPN의 모형보다는 정확성이 높지 않다는 결론을 얻었다.
2. ANN 모형 중 BPN를 은닉층의 뉴런수를 여러 가지 경우로 바꾸어가며 적용한 경우 CSU 공식보다는 높은 정확성을 확인할 수 있었으며, 은닉층의 뉴런수가 3인 경우에는 다중회귀모형보다 정확성이 높은 것이 검증되었다. 또한 은닉층의 뉴런수가 3인 경우는 CSU 공식보다 46.25%, 다중회귀분석보다 8.45% 높은 정확도를 나타내었다.
3. ANN 모형 중 GRNN을 적용한 결과 BPN의 일부 Case 보다는 높은 정확성을 나타내었지만, 일부의 Case 보다는 낮은 정확성을 나타내어 세굴심 산정에서 BPN 보다 낮은 정확성을 나타낸다는 것을 확인할 수 있었다.

4. ANN 모형 중 RBFN은 가장 낮은 정확성을 나타내어 세굴심 산정에 가장 적합하지 않다는 것이 검증되었다.
5. Case4, Case5, Case6, RBFN의 결과에서 보여 지듯이 훈련과정에서의 과적합(Overfitting)은 검증과정에서 부정확성을 증가시키는 것을 확인할 수 있었다.
6. K-meas 군집분석을 적용한 인공신경망은 군집의 수를 변화시키며 적용한 결과 은닉층의 뉴런수가 3, 군집의 수가 3일 경우, GRNN, RBFN보다 정확성이 높은 BPN의 Case1(은닉층의 뉴런수가 3인 경우)보다 42.73% 높은 정확성을 보였다. 또한 K-means 방법을 사용하여 군집을 나누었을 때 군집별 분산값이 전체자료의 분산값보다 작아지는 현상과 군집의 중심과 가까운 검증 데이터들을 이용하여 검증을 실시하였기 때문에 인공신경망의 훈련 및 학습의 정확성을 증가시킨 것으로 판단된다. 따라서 군집분석을 적용한 인공신경망의 적용이 본 연구에서 사용된 방법 중 세굴심 산정에 가장 적용성이 높은 방법임과 동시에 정확도가 높은 방법임이 확인되었다.

참고문헌

강진구, 심우배, 송재우(2002) 신경망 이론을 이용한 교각 주위 국부세굴심 예측. **대한토목학회논문집**, 대한토목학회, 제22권, 제2-B호, pp. 125-133.

김만식, 이영신(2007) 하천수질예측을 위한 역전파알고리즘(BP) 적용. **대한토목학회논문집**, 대한토목학회, 제22권, 제2-B호, pp. 125-133.

박성식, 송재우(1999) 하천 민곡부의 연속교각 세굴심 변화 특성. **2007년도 정기학술대회논문집**, 대한토목학회, pp. 4185-4192.

안상진, 전계원(2001) RBF를 이용한 홍수유출량 예측. **대한토목학회논문집**, 대한토목학회, 제21권, 제6-B호, pp. 599-607.

이성진, 이승래, 장범수(2002) 인공신경망 모델을 이용한 불포화도 결빙기점착력 추정에 관한 연구. **대한토목학회논문집**, 대한토목학회, 제22권, 제3-C호, pp. 331-344.

윤용남, 윤재영, 이재수(1995) 2차원 수리모형을 이용한 홍수시 교량에서의 최대세굴심도 예측. **대한토목학회논문집**, 대한토목학회, 제15권, 제6호, pp. 1689-1696.

윤태훈, 윤성범(1995) 국내의 세굴에 의한 교량붕괴 사례 및 감시대책. **한국수자원학회지**, 한국수자원학회, 제28권, 제6호, pp. 28-34.

이철웅(2000) 교각주위의 국부 세굴에 대한 신뢰성 해석. **대한토목학회논문집**, 대한토목학회, 제20권, 제4-B호, pp. 491-501.

최성욱, 정상화(2002) Monte-Carlo 시뮬레이션이 포함된 인공신경망을 통한 세굴심 예측. **2002년도 학술발표대회 논문집**, 대한토목학회, pp. 596-599.

한국수자원학회(2005) **하천설계기준**. (주)건설교통저널.

Choi, S.-U. and Cheng, S. (2006) Prediction of local scour around bridge piers using artificial neural networks. *Journal of the American Water Resources Association*, Vol. 43, pp. 487-493.

Haykin, S. (1999) *Neural Networks A Comprehensive Foundation*.

- Macmillian College Publishing Company, New York, N.Y.
- Jeng, D.-S. (2005) Neural Network Assessment for Scour Depth Around Bridge Piers. Research Report, NO. R855, University of Sydney, Australia.
- Jonhson, P.A. (1992) Reliability-based pier scour engineering. *Journal of Hydraulic Engineering, ASCE*, Vol. 118, No. 10, pp. 1344-1358.
- Moody, J. and Darken, C.J. (1989) Fast learning in networks of locally-tuned processing units. *Neural Computation*, Vol. 1, pp. 281-294.
- (접수일: 2008.7.22/심사일: 2008.9.2/심사완료일: 2009.1.16)