

의사 샘플 신경망에서 학습 샘플 및 특징 선택 기법

허경용*, 박충식**, 이창우***

Training Sample and Feature Selection Methods for Pseudo Sample Neural Networks

Gyeongyong Heo*, Choong-Shik Park**, Chang-Woo Lee***

요약

의사 샘플 신경망은 학습 샘플의 수가 적은 경우 학습된 신경망이 국부 최적해에 빠져 성능이 저하되는 것을 보완하기 위해 기존 샘플들로부터 의사 샘플을 생성하고 이를 통해 해공간을 평탄화 시킴으로써 학습된 신경망의 성능을 향상시킬 수 있는 신경망의 변형이다. 이는 학습 샘플의 양에 관한 문제로 이 논문에서는 이에 더해 학습 샘플의 질을 향상시킴으로써 학습된 신경망의 성능을 더욱 높일 수 있는 방법을 제시하였다. 잡음이 적게 포함된 전형적인 학습 샘플들만이 주어지고 입력 특징 중 출력과 연관성이 높은 특징만을 사용함으로써 학습된 신경망의 성능을 높일 수 있음은 자명하다. 따라서 이 논문에서는 커널 밀도 추정을 통해 비전형적인 학습 샘플을 제거하고 입력값이 출력값에 미치는 영향을 나타내는 연관성 척도를 사용하여 연관성이 적은 특징을 제거함으로써 의사 샘플 신경망의 성능을 향상시킬 수 있음을 보였다. 제시한 방법의 유효성은 토석류 데이터를 이용한 실험을 통해 확인할 수 있다.

▶ Keywords : 의사 샘플 신경망, 샘플 선택, 특징 선택, 커널 밀도 추정, 연관성 척도

Abstract

Pseudo sample neural network (PSNN) is a variant of traditional neural network using pseudo samples to mitigate the local-optima-convergence problem when the size of training samples is small. PSNN can take advantage of the smoothed solution space through the use of pseudo samples. PSNN has a focus on the quantity problem in training, whereas, methods stressing the quality of training samples is presented in this paper to improve further the performance of PSNN.

•제1저자 : 허경용 •교신저자 : 이창우

•투고일 : 2013. 2. 21, 심사일 : 2013. 3. 29, 게재확정일 : 2013. 4. 10.

•이 논문은 2013년 한국컴퓨터정보학회 제47차 동계학술대회에서 발표한 논문("의사 샘플 신경망에서 특징 선택 기법")을 확장한 것임.

* 동대대학교 전자공학과(Dept. of Electronic Engineering, Dong-Eui University)

** 영동대학교 스마트IT학부(Dept. of Smart Information Technology, Youngdong University)

*** 국립산림과학원 산림방재연구과(Division of Forest Management, Korea Forest Research Institute)

It is evident that typical samples and highly correlated features help in training. In this paper, therefore, kernel density estimation is used to select typical samples and correlation factor is introduced to select features, which can improve the performance of PSNN. Debris flow data set is used to demonstrate the usefulness of the proposed methods.

▶ Keywords : Pseudo sample neural network, Sample selection, Feature selection, Kernel density estimation, Correlation factor

I. 서 론

신경망은 폰 노이만식 컴퓨터가 능동적 과정을 수행하도록 프로그램 하는 일이 매우 어려우며 뇌에서 수행되는 계산 과정은 컴퓨터에서 수행되는 계산 과정과 차이가 난다는 점을 인식하면서 생겨난 방법으로 고도의 병렬처리 방식과 정보의 분산 저장을 그 특징으로 한다. 신경망은 입력과 출력 사이의 임의의 관계를 학습할 수 있을 뿐만 아니라 체계적인 학습 방법이 정립되어 있어 분류기뿐만 아니라 회귀 분석의 도구로도 그 활용도가 높다[1].

신경망을 사용하는 과정은 크게 학습 과정과 사용 과정으로 나눌 수 있으며 학습 과정에서 학습 샘플의 중요성은 잘 알려져 있다. 학습 샘플의 개수가 적은 경우에는 학습된 모델이 극부 최적해에 빠져 성능이 저하되는 문제점이 있으며 이처럼 충분하지 못한 학습 샘플의 크기로 인해 학습된 시스템의 신뢰성이 떨어지는 문제를 SSS (Small Sample Size) 문제라고 한다[2-3]. SSS 문제의 해결을 위해서는 충분한 학습 샘플의 확보가 필요하지만 문제에 따라 충분한 학습 샘플을 확보하는 것이 어려울 수 있다. 이러한 SSS 문제를 완화하기 위해 기존 학습 샘플을 이용하여 의사 샘플을 생성하고 이를 학습에 사용하는 의사 샘플 신경망(Pseudo Sample Neural Network, PSNN)이 제안되었고 실험을 통해 학습 샘플의 크기 문제를 일부 해결할 수 있음이 입증되었다[4].

PSNN은 학습 샘플의 크기 문제를 완화시키는 방법이다. 하지만 학습 샘플의 크기와 더불어 학습 샘플의 품질 역시 학습에서 중요한 역할을 한다. 학습 샘플의 품질은 두 가지 측면에서 생각해볼 수 있으며 학습 샘플의 전형성(typicality)과 출력값 결정에 기여하는 정도가 그것이다. 전형적인 샘플은 잡음이 많이 포함되지 않은 학습 샘플을 말하며 비전형적인 샘플은 학습된 모델의 구조를 왜곡시켜 성능 저하를 유발하는 원인이 된다. 비전형적인 학습 샘플의 선택을 위한 기준

은 여러 가지가 있으며 분포 함수를 추정하고 주어진 샘플이 생성될 확률이 최소가 되는 샘플을 비전형적인 샘플로 선택하는 방법이 그 한가지이다. 하지만 실제 문제에서 샘플들이 알려진 분포 함수를 따르는 경우는 많지 않으며 특히 PSNN을 적용하는 문제의 경우 학습 샘플의 크기가 매우 작으므로 알려진 분포를 따른다고 가정하더라도 이를 추정하기에 어려움이 있다. 따라서 이 논문에서는 커널을 이용하는 비모수적인 분포 함수 추정을 사용하여 비전형적인 샘플을 선택하고 이를 학습에서 제외하는 방법을 제시하였다.

신경망은 입력으로 특징 벡터를 받아 출력값을 내지만 입력이 출력값을 결정하는데 기여하는 정도는 다르다. 출력값을 결정하는데 기여도가 높은 입력의 부분집합을 사용함으로써 학습된 모델의 성능을 높일 수 있는 것으로 알려져 있으며, 이처럼 특징의 부분집합을 선택하는 것을 특징 선택(feature selection)이라 한다. 특징 선택은 모든 가능한 부분집합을 대상으로 실험하여 그 중 최고의 성능을 보이는 부분집합을 선택하는 것이 가장 나은 방법이지만 이는 특징 개수에 따라 연산량과 시간이 지수적으로 증가하는 단점이 있다. 따라서 이 논문에서는 학습 샘플의 크기가 작은 경우 사용할 수 있는 특징 선택 기준으로 관계 지수(correlation index)를 제안하고 이를 특징 선택에 사용하였다. 제안한 샘플 및 특징 선택 방법을 PSNN에 적용하여 토석류 데이터에 적용한 결과 오류가 줄어들었음을 확인할 수 있었다.

이 논문의 구성은 다음과 같다. 2장에서는 PSNN에 대해 설명하며 3장과 4장에서는 PSNN의 성능을 높이기 위한 샘플 선택 및 특징 선택 방법을 제안한다. 5장에서는 실험 결과를 통해 제안한 방법이 효율적이며 PSNN의 성능이 향상됨을 보이고 향후 연구 방향은 6장에서 제시한다.

II. 의사 샘플 신경망

의사 샘플 신경망(PSNN)은 신경망의 변형으로, 학습 샘플

플의 크기가 작은 경우 학습된 신경망이 국부 최적해에 빠져 성능이 저하되는 문제를 보완하기 위해 제안된 방법이다. PSNN은 기존 학습 샘플을 이용하여 의사 샘플을 생성하고 이를 학습에 사용함으로써 해공간을 평탄화 시켜 국부 최적해에 빠질 확률을 줄임으로써 성능의 개선을 가져온다는 사실이 실험적으로 입증되었다[4].

PSNN에서 의사 샘플 X_{PS} 는 기존 샘플을 평균으로 가지는 가우스 분포에서 생성하며 각 가우스 분포의 분산은 학습 샘플의 분산을 샘플의 수로 나눈 값을 사용하였다. 가우스 분포 이외의 분포를 활용할 수도 있지만 샘플의 수가 적은 경우에는 분포를 추정하기가 어려우며, 모든 분포는 가우스 분포의 혼합으로 설명할 수 있고 샘플의 수가 많은 경우 모든 분포는 가우스 분포로 수렴한다는 중심 극한 정리(central limit theorem)[5]에 따라 가우스 분포를 사용하였다. N 개의 샘플 $X = \{x_1, x_2, \dots, x_N\}$ 이 주어진 경우 의사 샘플 X_{PS} 는 식 (1)과 같이 나타낼 수 있다.

$$X_{PS} = \bigcup_{i=1}^N \left\{ x_j' \mid x_j' \sim N\left(x_i, \frac{\Sigma}{N}\right), j=1, \dots, N_{PS} \right\} \quad (1)$$

이 때 Σ 는 샘플 X 의 분산-공분산 행렬을 나타내고, N_{PS} 는 각 분포에서 생성되는 샘플의 개수를 나타낸다. 생성된 의사 샘플 X_{PS} 는 PSNN의 학습에 사용한다. 학습 데이터에 따라서는 원 샘플의 중요성을 고려하여 원 샘플을 학습 데이터에 반복해서 첨가할 수 있으며 이 경우 PSNN의 학습 데이터는 식 (2)와 같이 주어진다.

$$X_{PS}' = X_{PS} \cup \left\{ \bigcup_{i=1}^{N_0} X \right\} \quad (2)$$

이 때 N_0 는 원 샘플이 학습 샘플에 반복해서 추가되는 횟수를 나타낸다. 그림 1은 2차 방정식 $y = \frac{1}{120}(x^2 + 2x - 1)$ 에서 11개의 샘플을 취하고 여기에 가우시안 노이즈를 추가한 예를 보여준다. 그림 1의 X 는 전통적인 신경망에서 학습 샘플로 사용되며 그림 2의 X_{PS} 는 PSNN의 학습에 사용된다. 두 학습 샘플 집합 X 와 X_{PS} 를 이용하여 신경망을 학습시키고 학습된 신경망에서의 결과를 비교한 예가 그림 3이다. 그림 3에서 보아 알 수 있듯이 의사 샘플을 이용하여 학습한 신경망의 경우 실제 함수에 보다 근접한 결과를 보여주고 있음을 알 수 있다.

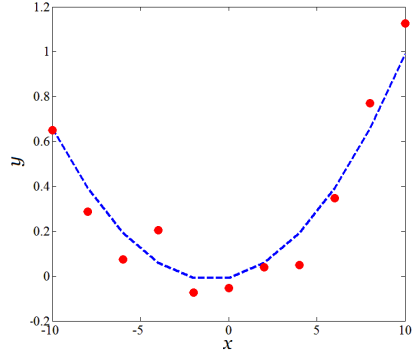


그림 1. 2차 방정식과 잡음이 섞인 샘플
Fig. 1. A quadratic equation and its noisy samples

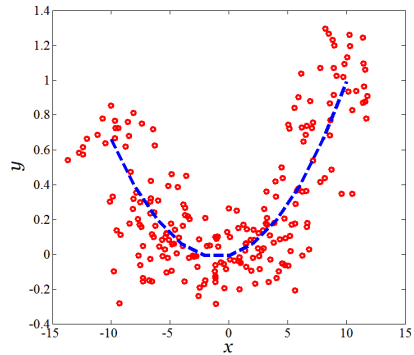


그림 2. 가우시안 분포를 이용하여 생성한 의사 샘플
Fig. 2. Pseudo samples using Gaussian distributions

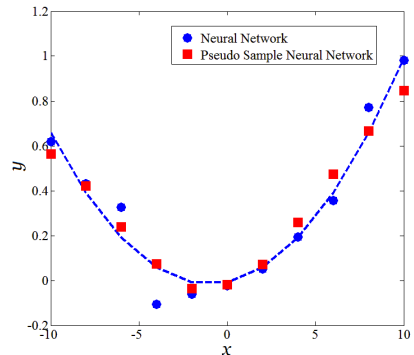


그림 3. 신경망과 의사 샘플 신경망의 학습 결과
Fig. 3. Training result of a traditional neural network and a pseudo sample neural network

III. 학습 샘플 선택

PSNN은 신경망의 학습을 위한 샘플이 충분히 확보되지 않은 경우 이를 보완하기 위한 방법으로 그 실효성이 입증되었다. 하지만 학습 샘플 자체의 품질 역시 학습된 모델의 품질에 영향을 미친다. 학습 샘플에 잡음이 많이 포함된 경우 잡음은 학습된 모델을 왜곡시키고 학습된 모델의 성능을 저하시키는 원인이 된다. 따라서 비전형적인, 잡음이 포함된 학습 샘플을 제거함으로써 성능 개선을 가져올 수 있다. 하지만 PSNN의 적용 대상이 되는 데이터는 그 크기가 작으므로 학습 샘플의 전형성을 판단하기가 쉽지 않다. 따라서 이 논문에서는 비모수적 밀도 추정을 통해 각 학습 샘플의 전형성을 판단하고 이를 통해 학습 샘플을 선택하는 방법을 제안하였다. 비모수적 밀도 추정 방법은 확률밀도 함수의 형태를 가정하지 않고 주어진 데이터로부터 함수 자체를 직접 추정한다.

확률변수 X 가 iid(independent and identically distributed)일 때 샘플 $\{x_1, x_2, \dots, x_N\}$ 은 각각 $1/N$ 확률을 가지게 되므로 $X = x$ 에서의 경험적 누적분포함수 $\hat{F}(x)$ 는 샘플 중 x 보다 작거나 같은 값을 갖는 샘플의 비로 주어진다.

$$\hat{F}(x) = \frac{1}{N} \sum_{i=1}^N 1(x_i \leq x) \quad (3)$$

이 때 $1(A)$ 는 A 가 참일 때 1을 거짓일 때 영(zero)의 값을 갖는 지시 함수(indicator function)를 나타낸다. 경험적 확률밀도함수 \hat{f} 는 $\hat{F}(x)$ 를 미분하면 구할 수 있다. 즉, 구간 h 내 샘플의 비율을 계산함으로써 확률밀도함수를 구할 수 있다.

$$\begin{aligned} \hat{f}(x) &= \frac{\hat{F}(x+h) - \hat{F}(x)}{h} \\ &= \frac{1}{Nh} \sum_{i=1}^N 1\left(x - \frac{h}{2} < x_i \leq x + \frac{h}{2}\right) \end{aligned} \quad (4)$$

하지만 식 (4)는 불연속 함수라는 단점이 있다. 따라서 식 (4)에서 지시함수를 연속 함수인 $\kappa(\cdot)$ 으로 대체시킨 식 (5)를 사용한다.

$$\hat{f}(x) = \frac{A}{Nh} \sum_{i=1}^N \kappa\left(\frac{x_i - x}{h}\right) \quad (5)$$

이 때 A 는 식 (5)가 확률밀도함수이기 위해 필요한 상수이며, h 는 커널 폭으로 샘플 개수 N 의 함수로 $N \rightarrow \infty$ 에 따라 $h \rightarrow 0$ 이 된다. $\kappa(\cdot)$ 는 커널이라 부르며 일반적으로 식 (6)의 지수함수 형태가 많이 사용된다.

$$\kappa(x) = \exp\left(-\frac{\|x_i - x\|^2}{h}\right) \quad (6)$$

식 (6)은 확률밀도를 추정할 샘플 x 주변에 많은 샘플이 있는 경우 국부적으로 높은 밀도를 부여하는 방식으로 주어진 샘플을 기초로 확률밀도함수를 추정하는 방식이다[5-7]. 식 (6)을 이용하여 학습 샘플 내에서의 비전형적인 샘플의 번호는 식 (7)로 계산할 수 있다.

$$i_{atypical} = \arg \min_{i=1, \dots, N} \sum_{j=1}^N \exp\left(-\frac{\|x_i - x_j\|^2}{h}\right) \quad (7)$$

식 (7)은 학습 샘플 내 비전형적인 샘플 하나를 선택하는 방법이며 식 (7)을 반복 적용함으로써 임의의 개수의 비전형적인 샘플을 선택할 수 있다. 이 논문에서는 가우시안 분포에서 전형적인 샘플들이 2σ 내에 분포한다고 가정하고 95%의 학습 샘플만을 학습에 사용하였다.

IV. 특징 선택

학습 샘플 선택과 더불어 학습에서 중요한 요인 중 하나는 입력값이 출력값 결정에 영향을 미치는 정도를 파악하는 것이다. 학습 샘플 선택은 잡음이 많이 포함된 샘플을 선택하고 제거함으로써 직접적으로 학습 샘플의 품질과 연관을 가진다. 반면 특징 선택은 출력값 결정에 영향을 미치지 않는 입력을 제거함으로써 모델 내 학습해야 할 파라미터의 개수를 줄여 잡음 민감성을 감소시키는 간접적인 연관성이 있다. 입력값이 출력값 결정에 미치는 영향을 정량화할 수 있는 일반적인 방법은 존재하지 않으며 입력 중 일부를 선택하고 이를 이용한 반복 실험을 통해 최선의 결과를 내는 입력의 부분집합을 선택하는 방법이 주로 사용된다. 하지만 특징의 개수가 증가함에 따라 지수적으로 연산량이 증가하는 단점이 있다. 따라서 이 논문에서는 입력값과 출력값의 scatter plot 분석을 통해 입력값이 출력값 결정에 미치는 영향을 정량화할 수 있는 관계 지수(correlation index)를 제안하고 이를 특징 선택에

사용하였다.

입력값과 출력값이 주어진 경우 이를 scatter plot으로 나타낸 것이 그림 4이다. 신경망은 입력과 출력의 비선형적인 임의의 관계도 학습할 수 있지만 scatter plot 상에서 특정 지역에 데이터가 밀집되어 나타나는 경우 이를 분별하기는 쉽지 않다. 따라서 scatter plot 상에서 데이터 포인트들 사이의 거리가 가까운 경우 학습에 부정적인 역할을 미치는 것으로 생각할 수 있다. 그림 4에서 볼 수 있듯이 4-(a)는 전체적으로 넓게 분포하는 반면 4-(b)의 경우는 특정 지역에 밀집된 양상을 보인다. 즉, 그림 4-(b)에 사용된 입력이 성능 저하의 원인이 되는 것으로 판단할 수 있다.

두 벡터의 상관관계 파악을 위해서는 피어슨의 상관 분석이 많이 이용되지만[8] 선형 관계만을 파악할 수 있고 실제 데이터가 선형 관계를 가지는 경우는 드물기 때문에 사용할 수 없다. 따라서 scatter plot 상에서 데이터의 밀집 정도를 나타내는 관계 지수를 제안하였다.

N 개의 스칼라 특징값 $X_{*a} = \{x_{ia} | i = 1, \dots, N\}$ 와 스칼라 출력값 $Y_{*b} = \{y_{ib} | i = 1, \dots, N\}$ 가 주어졌다고 가정하자. 일반적으로 신경망에서 입력값 및 출력값은 벡터로 주어지므로 a 와 b 는 벡터 중 하나의 원소를 나타낸다. X_{*a} 와 Y_{*b} 로 이루어지는 scatter plot 상에서 i 번째 포인트 (x_{ia}, y_{ib}) 와 j 번째 포인트 (x_{ja}, y_{jb}) 사이의 거리는 식 (8)과 같이 계산된다.

$$d_{ij} = \sqrt{(x_{ia} - x_{ja})^2 + (y_{ib} - y_{jb})^2} \quad (8)$$

입력 X_{*a} 와 출력 Y_{*b} 의 관계 지수(correlation index) ρ_{ab} 는 거리 $d_{ij} (i, j = 1, \dots, N)$ 의 분산으로 식 (9)와 같이 정의한다.

$$\rho_{ab} = \text{Var}(d_{ij}), \quad 1 \leq i, j \leq N \quad (9)$$

입력 특징의 개수를 A , 출력값의 개수를 B 라 하면 a 번째 특징이 전체 출력과 가지는 관계 지수는 식 (10)과 같이 정의할 수 있다.

$$\rho_a = \sum_{b=1}^B \rho_{ab} \quad (10)$$

이 논문에서는 식 (10)의 관계 지수를 이용하여 입력값이 출력값 결정에 미치는 영향을 정량화하고 최소의 관계 지수를 갖는 특징을 제거하였다. 출력값과 무관한(irrelevant) 특징 번호는 식 (11)과 같이 계산된다.

$$a_{irrelevant} = \arg \min_{a=1, \dots, A} \rho_a \quad (11)$$

V. 실험 결과

제한한 학습 샘플 선택 및 특징 선택은 의사 샘플 신경망에서와 마찬가지로 신경망 학습 이전의 전처리 단계로 사용되었으며 전체적인 흐름도는 그림 5와 같다.

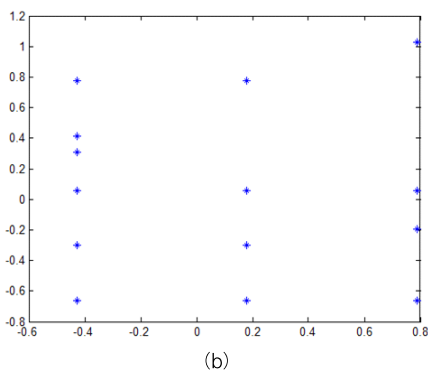
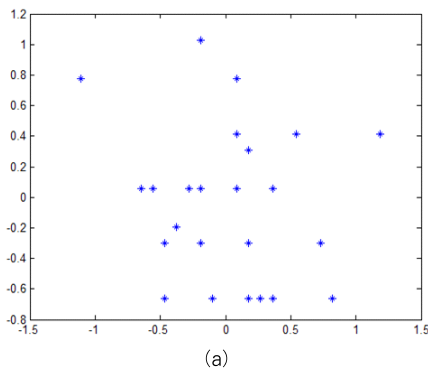


그림 4. 특징값과 출력값의 scatter plot
Fig. 4. Scatter plots between feature and output values

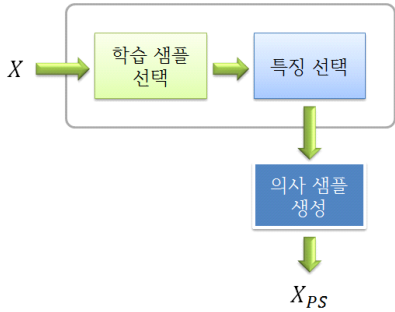


그림 5. 개선된 의사 샘플 신경망 학습 과정
Fig. 5. Training steps for modified pseudo sample neural networks

실험에 사용한 데이터는 토석류 데이터로 산사태에 의해 발생하는 피해지역 예측을 위해 사용된다. 입력은 산사태 발생 지역의 지형 및 지질 특징을 나타내는 5가지로, 출력은 산사태 피해 예측을 위한 Random Walk Model의 3가지 파라미터로 구성되며 5개 지역에서 수집된 27개 데이터가 사용되었다[9].

신경망의 사용을 위해서는 먼저 신경망의 구조를 결정하여야 한다. 신경망의 구조는 은닉층(hidden layer)의 개수와 은닉층의 노드 개수를 조절하면서 오류가 최소가 되는 구조를 선택하였다. 은닉층의 개수는 1~3개, 은닉층의 전체 노드 개수는 4~22개 범위에서 실험하였으며 200회 반복 실험을 수행하였다. 모델 선택을 위한 실험에서는 전체 데이터를 모두 학습에 사용하는 test-on-train 방법을 사용하였으며 오류는 신경망의 출력값 $O = \{o_{ib} | 1 \leq i \leq N, 1 \leq b \leq B\}$ 와 기대 출력 $Y = \{y_{ib} | 1 \leq i \leq N, 1 \leq b \leq B\}$ 의 자승 오류 합으로 계산하고 비교하였다.

$$E = \sum_{i=1}^N \sum_{b=1}^B |y_{ib} - o_{ib}|^2 \quad (12)$$

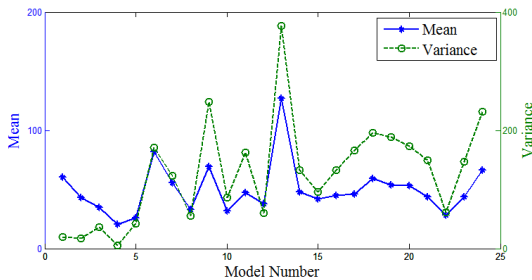


그림 6. 신경망 구조에 따른 오류
Fig. 6. Errors with respect to network structure

그림 6은 신경망의 구조에 따른 오류를 나타낸 것으로 4번째 구조(1개의 은닉층과 7개의 은닉 노드)에서 최소의 평균 오류 및 분산을 나타내고 있음을 알 수 있다. 이후 실험에서는 이 실험에서 선택된 구조를 사용하여 실험하였으며 이후의 오류들은 10-fold crossvalidation을 적용하고 30회 반복 실험한 결과를 평균한 값을 제시하였다.

의사 샘플 신경망에 학습 샘플 선택만을 적용한 결과가 그림 7과 표 1에 나타나 있다. 그림 7은 제거되는 학습 샘플의 개수에 따른 오류의 평균 및 분산을 나타낸 그래프로 1개의 샘플을 제거한 경우 최소의 오류를 보이고 있다. 이는 전체 샘플의 약 4.1%에 해당하는 값으로 가우시안 분포에서 2σ 내에 전형적인 샘플들이 위치한다는 가정과 일치한다.

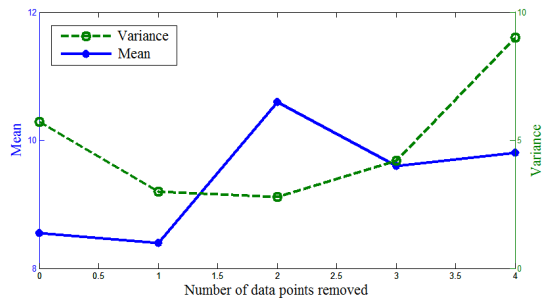


그림 7. 학습 샘플 제거 개수에 따른 오류
Fig. 7. Errors w.r.t. the number of samples removed

표 1. 학습 샘플 제거 개수에 따른 오류
Table 1. Errors w.r.t. the number of samples removed

제거되는 샘플 수	오류 평균	오류 분산
0	8.55	5.71
1	8.41	3.02
2	10.62	2.83
3	9.63	4.21
4	9.88	9.04

의사 샘플 신경망에 특징 선택만을 적용하여 식 (10)의 관계 지수를 계산한 결과는 표 2에 나타나 있다. 표 2에서 볼 수 있듯이 4번째 특징의 관계 지수가 가장 작다. 즉, scatter plot 상에서 데이터가 특정 부분에 밀집되어 나타난다.

표 2. 입력과 출력의 연관성 ρ_a

Table 2. ρ_1 between inputs and outputs

특징 번호 a	관계 지수 ρ_a
1	4.302
2	4.528
3	4.201
4	4.072
5	4.179

그림 8과 표 3은 각 특징을 제거한 경우의 실험 결과를 나타낸 것으로 4번째 특징을 제거한 경우 최소의 오류를 보임을 알 수 있다. 이는 표 2에서 4번째 특징의 관계 지수가 가장 작다는 사실과 일치한다.

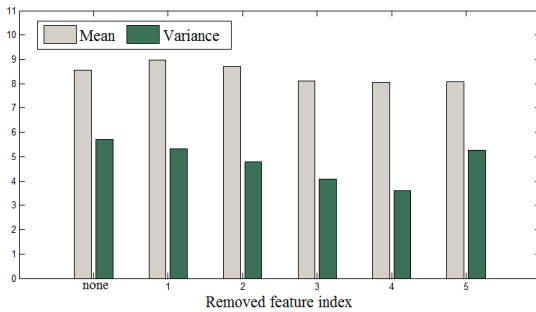


그림 8. 특징 제거에 따른 오류

Fig. 8. Error w.r.t. the index of removed feature

표 3. 특징 제거에 따른 오류

Table 3. Error w.r.t. the index of removed feature

제거한 특징 번호	오류 평균	오류 분산
none	8.55	5.71
1	8.98	5.33
2	8.69	4.77
3	8.11	4.09
4	8.07	3.61
5	8.08	5.27

토석류 데이터에 그림 5의 과정을 적용하여 위의 실험들과 비교한 결과는 표 4에 나타나 있다. 전통적인 신경망(NN)에 비해 의사 샘플 신경망(PSNN)은 그 성능이 확연히 좋아짐을 알 수 있으며 이는 국부 최적해에 수렴함으로써 학습된 모델의 성능이 아주 나쁜 경우가 많이 발생하였기 때문이다. PSNN에 학습 샘플 선택(Training Sample Selection,

TSS)이나 특징 선택(Feature Selection, FS)만을 적용한 경우에도 PSNN에 비해 나은 결과를 얻을 수 있었지만 두 방법을 모두 적용하였을 때 최선의 결과를 얻을 수 있었다. 학습 샘플 선택 및 특징 선택은 그림 5의 순서로 적용되었으며 특징 선택을 먼저 시행한 경우에도 토석류 데이터의 경우 거의 동일한 결과를 얻을 수 있었다.

표 4. 실험 결과 요약

Table 4. Summary of the experimental results

사용 방법	오류 평균	오류 분산
NN	50.14	38.42
PSNN	8.55	5.71
PSNN+TSS	8.41	3.02
PSNN+FS	8.07	3.61
PSNN+TSS+FS	7.95	2.99

VI. 결론 및 향후 연구 방향

의사 샘플 신경망(PSNN)은 학습 샘플이 적은 경우 학습된 신경망의 성능이 저하되는 문제를 완화하기 위해 의사 샘플을 이용하여 학습을 수행함으로써 성능 저하를 줄이는 신경망의 변형이다. 하지만 학습 샘플의 수가 적은 경우 샘플에 포함된 잡음과 출력값을 결정하는데 관계가 적은 입력값 역시 학습된 신경망의 품질에 영향을 미친다. 전자가 학습 샘플의 양(quantity)에 주안점을 두고 있다면 후자는 질(quality)에 주안점을 두고 있다고 하겠다. 이 논문에서는 학습 샘플의 질 개선을 통해 학습된 신경망의 성능을 높일 수 있는 방법을 제안하였다. 이를 위해 이 논문에서는 커널 밀도 추정을 통해 잡음이 많이 포함된 샘플을 선택하고 제거할 수 있는 학습 샘플 선택 기법과 입력과 출력의 상관관계를 나타내는 관계 지수를 정의하고 이를 이용하여 관련성이 높은 특징들만을 학습에 사용하도록 하는 특징 선택 기법을 제안하였다. 제안한 방법들은 토석류 데이터를 이용한 실험을 통해 그 유효성이 입증되었다.

제한한 학습 샘플 및 특징 선택 방법들은 학습 데이터가 극히 적어 학습 샘플의 분포를 추정하거나 신경망의 입출력 관계를 정확히 파악할 수 없는 경우 유용하다. 하지만 학습 샘플 및 특징 선택을 위해 다양한 방법들이 제시되고 있으므로 이 논문에서 제안한 방법들 역시 학습 데이터의 크기와 상관없이 사용할 수 있는 방법으로서 일반화가 가능할 것으로 판단되며 이는 향후 연구 과제로 남아있다. 또한 학습 샘플의

선택에서는 95%의 학습 샘플을, 특징 선택에서는 1개의 입력값을 제거하였지만 데이터에 따라 그 비율은 달라질 수 있으므로 데이터에 따른 비율 결정 방법 역시 중요한 향후 연구 과제라 하겠다.

the Optimal Parameters Extraction of Random Walk Model – Case on Debris Flow Area of Bonghwa County in Gyeongbuk Province.” Journal of Korean Forest Society, Vol. 100, No. 4, pp. 664-671, Apr. 2011.

참고문헌

- [1] C. M. Bishop, "Pattern Recognition and Machine Learning," 2nd ed. Springer, 2007.
- [2] R. P. W. Duin, "Small sample size generalization," Proceedings of the 9th Scandinavian Conference on Image Analysis, pp. 957-964, Oct. 1995.
- [3] T. G. van Niel, T. R. McVicar, and B. Datt, "On the relationship between training sample size and data dimensionality: Monte Carlo analysis of broadband multi-temporal classification," Remote Sensing of Environment, Vol. 98, No. 4, pp. 468-480, 2005.
- [4] Gyeongyong Heo, Chang-Woo Lee, and Choong-Shik Park, "Parameter Estimation in Debris Flow Deposition Model Using Pseudo Sample Neural Network," Journal of the Korea Society of Computer and Information, Vol. 17, No. 11, pp. 11-18, Nov. 2012.
- [5] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," Annals of Mathematical Statistics, Vol. 27, No. 3, pp. 832-837, Sep. 1956.
- [6] E. Parzen, "On estimation of a probability density function and mode," Annals of Mathematical Statistics, Vol. 33, No. 3, pp. 1065-1076, Sep. 1962.
- [7] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," Annals of Statistics, Vol. 38, No. 5, pp. 2916-2967, Oct. 2010.
- [8] J. Aldrich, "Correlations Genuine and Spurious in Pearson and Yule," Statistical Science, Vol. 10, No. 4, pp. 364-376, Nov. 1995.
- [9] Chang-Woo Lee, Choongshik Woo, and Ho-Joong Youn, "Analysis of Debris Flow Hazard Zone by

저자소개



허 경 용
 1994: 연세대학교
 전자공학과 공학사.
 1996: 연세대학교
 전자공학과 공학석사.
 2009: University of Florida
 컴퓨터공학과 공학박사
 현 재: 동의대학교 전자공학과 교수
 관심분야: 인공지능, 패턴인식, 로봇공학
 Email : hgycap@deu.ac.kr



박 충 식
 1985: 한양대학교
 전자공학과 공학사.
 1987: 연세대학교
 전자공학과 공학석사.
 1994: 연세대학교
 전자공학과 공학박사
 현 재: 영동대학교
 스마트IT학부 교수
 관심분야: 인공지능, 패턴인식,
 인지과학
 Email : leciel@yd.ac.kr



이 창 우
 1992: 영남대학교
 산림자원학과 농학사
 1994: 영남대학교
 산림자원학과 농학석사
 1998: 일본 동경대학
 농학생명과학연구과
 산림과학전공 농학박사
 현 재: 국립산림과학원
 관심분야: 토석류, 사방, 산사태,
 토질역학
 Email : leecw@forest.go.kr