

# 데이터 중요도의 사전 평가를 이용한 증가학습을 위한 데이터 선택 방법

이선영, 조성준, 방승양  
포항공과대학교 전자계산학과

## Data selection method for Incremental learning using prior evaluation of data importance

Sunyoung Lee, Sungzoon Cho, Sungyang Bang

Department of Computer Science and Engineering, Pohang University of Science and Technology

### 요 약

다층 퍼셉트론 학습은 학습 데이터의 능동적인 선택 여부에 따라 능동적 학습(Active learning)과 피동적 학습(Passive learning)으로 구분할 수 있다. 기존의 능동적 학습 방법들은 학습 데이터의 중요도를 측정할 수 있는 기준(measure)을 제시하고 이 기준에 따라 학습 데이터를 선택하는 방법을 취하고 있다. 이 방법들은 학습 데이터 선택을 위해 Hessian Approximation과 같은 복잡한 계산이나 학습 데이터를 선택하는 과정에 있어서 데이터의 중요도를 평가하기 위한 반복적인 계산을 필요로 한다. 본 논문에서는 학습 데이터 선택 시 반복적인 계산이 필요하지 않는 비교사 학습을 이용한 능동적 학습 데이터 선택 방법을 제안하고 그 수렴 특성과 일반화 성능을 분석한다. 또한 비교 실험을 통하여 제안된 방법이 기존의 능동적 학습방법보다 간단한 계산만으로 수렴 속도를 향상시키며 일반화에도 뒤떨어지지 않음을 보인다.

### 1. 서론

다층 퍼셉트론을 학습시키는데 있어서 가장 널리 사용되는 알고리즘인 오류 역전파 알고리즘(Error-backpropagation algorithm)은 gradient descent 방법으로 국부적 최적해(Local minima)에 도달하거나 수렴 속도(Convergence rate)가 느리다는 단점이 있다. 이러한 오류 역전파 알고리즘의 단점을 극복하기 위해 제안되어 온 방법은 더 나은 학습 알고리즘을 제안하는 방법과 다층 퍼셉트론의 수렴 속도 및 일반화 성능에 영향을 미치는 또 다른 중요한 요인인 학습 데이터를 조작하거나 능동적으로 학습에 도움이 되는 중요한 데이터를 선택하여 수렴속도를 향상시키는 방법이 있다. 이러한 능동적 학습은 초기에 학습 데이터를 임의로 선택하며, 학습 데이터 선택하는 과정에 있어서 Hessian Approximation과 같은 복잡한 계산이나 학습 데이터의 중요도를 평가하기 위한 반복적인 계산을 필요로 한다는 단점이 있다. 본 논문에서는 학습 데이터 선택 시에 비교사 학습을 이용하여 학습 데이터 선택을 위한 반복적인 계산이 아닌 단 한번의 계산만으로 학습데이터를 선택하는 보다 계산된 방법을 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 능동적 학습 방법에 대한 관련된 연구를 분석하고, 3장에서는 비교사 학습을 이용한 능동적 학습 알고리즘을 제시한다. 4장에서는 제시된 알고리즘을 인위적인 데이터 적용한 실험 결과와 기존의 능동적 학습방법과의 비교 실험 결과를 제시한다. 마지막으로 5장에서는 결론 및 향후 연구 방향을 기술한다.

### 2. 관련 연구

능동적 학습은 가능한 학습에 필요한 많은 정보를 가지고 있으면서 동시에 가장 적은 수의 데이터를 선택하여 학습하는 방법이다. 능동적 학습은 후보 데이터 집합(Candidate training set)에서 학습에 도움이 되는 유용한 정보를 가지고 있는 데이터를 등적으로 선택한다. 데이터 선택을 위해 학습 데이터의 중요도를 평가하는 기준이 제시되고 이 기준을 사용하여 데이터를 선택한다. 이러한 방법은 중요한 학습 데이터를 선택하여 학습함으로써 일반화 성능 향상과 학습 시간을 단축시키는 효과를 기대할 수 있다 [1, 2, 3].

능동적 학습 방법은 크게 증가 학습(Incremental learning)과 선택 학습>Selective learning)으로 나눌 수 있다. 증가 학습은 초기에 후보 데이터 집합에서 임의로 선택된 데이터로 학습을 시작한다. 학습 동안 만족할 만한 학습 성능에 도달하면 학습을 끝내고 그렇지 못한 경우 heuristics을 사용하여 학습에 사용되지 않은 나머지 후보 데이터 집합에서 데이터를 선택한다. 선택된 데이터는 전 단계에서 학습에 사용된 데이터와 함께 학습을 위해 사용된다. 다층 퍼셉트론 학습을 위해 선택된 데이터들은 후보 데이터 집합에서 제거되므로 학습 과정이 진행됨에 따라 학습에 사용되는 데이터 집합은 증가하고 후보 데이터 집합은 감소하게 된다. 증가 학습 방법의 예로는 후크 데이터 집합에서 현재까지 학습된 신경망에서의 에러가 가장 큰 데이터들을 선택하는 방법 [2, 3] 및 Integrated squared bias을 최소화 하는 데이터를 선택하는 방법 [4], Sensitivity Analysis를 학습 네

이터 선택에 사용하는 방법[5] 등이 있다.

증가 학습과는 달리 선택 학습은 학습 데이터를 후보 집합에서 선택한 후 선택된 데이터를 후보 데이터 집합에서 제거하지 않는다. 따라서 각 데이터 선택 단계에서 모든 데이터들이 선택될 동등한 기회를 가지게 된다.

### 3. 비교사 학습을 이용한 능동적 학습

능동적 학습의 중요한 특징은 전체 데이터를 학습에 사용하는 것이 아니라, 초기에 전체 데이터 중 일부만을 선택하여 학습하고 학습된 신경망이 만족할 만한 성능에 도달하지 못한 경우 학습에 도움이 되는 중요한 데이터를 증가시켜 학습하는 것이다. 이 방법은 주어진 데이터가 잡음(Noise)이나 데이터의 중복(Redundancy)으로 인하여 각 데이터들이 학습에 기여하는 정도가 다를 경우 중요한 데이터만을 선택하여 학습함으로써 학습 속도 및 일반화 성능 향상에 기여한다는 장점이 있다. 능동적 학습에서 가장 중요한 부분은 어떻게 중요한 학습 데이터를 선택하느냐 하는 것이다. 이를 위해 학습 데이터의 중요도를 측정하는 기준이 필요하다.

#### 3.1 학습 데이터의 중요도 평가

$N$ 개의 주어진 데이터  $D_N$  으로 학습된 신경망은 데이터의 모델로 볼 수 있다. 우리의 목적은 현재  $N$ 개의 데이터로 학습한 신경망이 만족할 만한 성능에 도달하지 못한 경우  $N+1$ 번째 데이터를 선택하는 것이다. 이 때 데이터의 선택은 그 데이터가 선택됨으로써 신경망이 얻게 되는 정보의 양을 측정하여 수행될 수 있다.  $N+1$ 번째 데이터를 선택하기 전과 후의 파라미터의 확률분포를 각기  $P_N(w)$  와  $P_{N+1}(w)$  라고 할 때, Information Theory [6]에 의하면,  $P_N(w)$  와  $P_{N+1}(w)$  를 구분하기 위한 평균 정보의 거리(Mean Information Distance)는 다음과 같이 주어진다.

$$I(P_{N+1}, P_N) = \int P_{N+1}(w) \ln \frac{P_{N+1}(w)}{P_N(w)} dw$$

$I(P_{N+1}, P_N)$  값이 클수록 두 확률 분포의 차이가 크게 된다. 즉, 파라미터  $w$ 에 대하여 더 많은 정보를 가지고 있다는 의미가 된다. 즉, 현재의 분포  $P_N(w)$  이 주어진 경우,  $P_N(w)$  와  $P_{N+1}(w)$ 의 차이를 크게 함으로써 최대의 정보를 얻을 수 있다.  $P_N(w)$  를 다음과 같이 정의한 경우 현재의 파라미터에 대해  $E(D_{N+1}|w)$  (Error function)를 가장 크게 하는 데이터를 선택하여  $P_N(w)$  와  $P_{N+1}(w)$  차이를 최대화 할 수 있다[7]

$$R_N(w) = R(D_N|w) = \prod_{p=1}^N P(y_p | x_p, w)$$

where

$$P(y_p | x_p, w) = \frac{\exp(-\beta E(y_p | x_p, w))}{Z(\beta)}$$

이때  $Z(\beta)$  는 normalization constant이다.

따라서,

$$\Delta E_{N+1} = E(D_{N+1}|w) - E(D_N|w)$$

을 최대화 하는 데이터가 학습에 도움이 되는 중요한 데이터가 된다

[3]. 이것은 현재까지 선택된 데이터들과 행동 방식이 다른 데이터일 수록 중요한 정보를 가지고 있는 데이터일 가능성이 높다는 의미이다. 이러한 이론에 근거하여 주어진 학습 데이터를 Class 별로 비교사 학습을 이용하여 행동 방식이 비슷한 데이터들끼리 Clustering 한 후 각 Cluster의 Mean에 가장 가까운 데이터들을 초기에 학습한다. 학습한 결과가 만족할 만한 성능에 도달하지 못한 경우 그 다음 데이터로 자신이 속한 Cluster의 Mean과의 거리가 가장 멀고 다른 Class에 속한 Cluster들의 Mean과는 거리가 가장 가까운 데이터를 차례로 선택하여 학습한다. 이 데이터들은 각 Cluster들 간의 경계에 존재하는 데이터들로써 모호성(Ambiguity)이 큰 데이터들이다. 이러한 데이터들은 Decision boundary에 존재하여 분류하기 모호한 데이터에 해당된다. 이제까지의 많은 연구들이 데이터를 임의로 선택해서 학습하는 것보다 Decision boundary에 존재하는 데이터를 학습하는 것이 일반화 성능이 향상된다는 것을 보여주고 있다[1]

본 논문에서 제시한 방법은 초기에 주어진 학습 데이터를 대표하는 데이터들을 학습하게 된다. 따라서, 초기에 학습된 신경망이 어느 정도의 학습 성능을 나타낼 수 있어서 기존의 방법보다 데이터 선택 단계를 감소시킬 수 있다. 또한 데이터의 중요도를 매 데이터 선택 단계에서 계산하지 않고, 데이터를 Clustering한 후 데이터의 중요도를 미리 결정할 수 있으므로 학습동안 데이터 선택을 위한 반복적인 계산과정을 필요로 하지 않는다.

#### 3.2 비교사 학습을 이용한 능동 학습 알고리즘

3.1에서 제시한 방법의 알고리즘은 다음과 같다

- [단계1] 주어진 학습 데이터를 Target 밀로 비교사 학습(Unsupervised learning)을 사용하여 Clustering 한다.
- [단계2] 각 데이터에 대해 다음 식에서 제시한 중요도를 평가하는 기준을 이용하여 데이터의 중요도를 계산한다. 중요도 값이 클수록 학습에 도움이 되는 데이터가 된다.

데이터  $x_j$  의 중요도 평가하는 기준(Criticality Measure)

$$C(x_j) = \frac{|x_j - mean_j|}{\sum_{i=1}^t \sum_{k=1}^c |x_j - mean_k|}$$

$x_j$  - Class 1 의  $j$  번째 Cluster에 속하는 학습 데이터  
 $t$  - Class의 개수  
 $c_k$  - Class  $k$  의 Cluster 개수  
 $mean_k$  - Class 1 의  $j$  번째 Cluster의 Mean

[단계3] 초기에 각 Cluster의 평균에 가장 가까운 데이터들을 선택하여 학습 데이터 집합  $D$ 에 포함시키고 나머지 데이터들을 후보 데이터 집합  $C$ 에 포함시킨다.

[단계4] 학습된 신경망이 만족할만한 성능에 도달할 때까지 다음 과정을 반복한다

[단계4.1] 신경망 학습 단계  
 학습 데이터 집합  $D$ 의 데이터들로 오류 역전파 알고리즘을 사용하여 신경망을 학습시킨다.

[단계4.2] 학습 데이터 선택 단계

후보 데이터 집합  $C$ 에 있는 데이터 중, 중요도 값이 큰 데이터들을 차례로 선택하여 학습 데이터 집합  $D$ 에 포함시킨다.

[표1] 기존 방법과의 비교 실험 결과

Data Selection Method	Validation error(# of used data)
Passive learning	4.41 (300)
Active learning (select data with max error)	5.54 (156)
Active learning using unsupervised learning	4.70 (120)

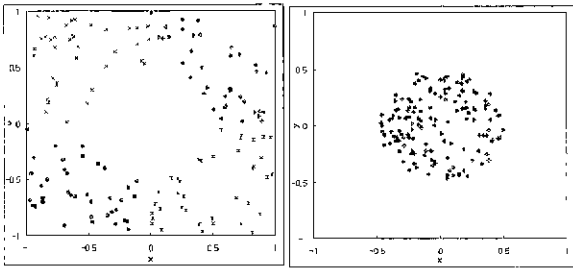
4. 실험 결과

다음은 인위적으로 만들어 낸 분류문제에 적용한 실험 결과이다. 동일한 조건에서 기존의 능동 학습 방법과 본 논문에서 제시한 방법으로 실험하였다. 사용된 신경망 구조는 2-3-1이며 *Learning rate*는 0.1을 사용하였다.

실험한 문제는 원점을 중심으로 반경이 0.5인 원 영역과 원 밖의 사각 영역을 분류하는 문제이다

$$class = \begin{cases} 0 & \text{if } \sqrt{x^2 + y^2} \leq 0.5 \\ 1 & \text{otherwise} \end{cases}$$

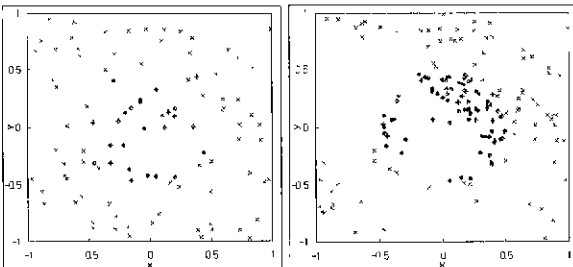
300개의 학습 데이터를 [-1,1] 범위에서 *Uniform*한 분포를 갖도록 생성시켰으며, *Validation*을 위해 [-1,1] 범위를 0.1 간격으로 나눈 *Grid* 데이터를 361개 생성시켜 사용하였다. [그림 1], [그림 2]은 각 *Class*별로 *Clustering*한 결과이다.



[그림 1] Class 1의 Clustering결과      [그림 2] Class 0의 Clustering결과

*Class 1*은 4개의 *Cluster*로 *Clustering*하였으나 *Class 0*은 1개의 *Cluster*로 *Clustering*하였다. 학습 데이터 선택 단계에서는 *Class 1*의 각 *Cluster*에서 2개씩을 *Class 0*의 *Cluster*에서 4를 선택하여 총 12개의 데이터를 선택하였다 [그림 3]는 능동적 학습방법을 사용하지 않은 경우와 비슷한 *Validation error*를 갖는 10번째 데이터 선택 단계에서 학습에 사용된 120개의 데이터를 나타낸다.

[그림 4]는 기존의 능동적 학습방법 중 현재까지 학습된 신경망에서



[그림 3] 10번째 학습 데이터 선택 단계에서 선택된 학습 데이터 (Unsupervised learning을 이용한 방법)      [그림 4] 13번째 학습 데이터 선택 단계에서 선택된 학습 데이터 (기존의 능동적 학습방법)

의 *Error*가 가장 큰 데이터를 선택하는 방법으로 실험하였을 경우 능동적 학습방법을 사용하지 않는 경우와 비슷한 *Validation error*를 갖는 13번째 데이터 선택 단계에서 학습에 사용된 156개의 데이터를 나타낸다 [표1]은 주어진 모든 데이터를 학습에 사용하는 피동적 방법과 비교사 학습을 이용한 능동적 학습, 그리고 기존의 능동적 학습 방법인 현재의 신경망에서의 *Error*가 가장 큰 데이터를 선택하는 방법을 제시된 분류문제에 대해 실험한 결과이다. 괄호 안의 값은 학습에 사용된 데이터의 개수를 나타낸다 [표1]에서 알 수 있듯이 능동적 학습은 주어진 300개의 데이터 중 약 1/2에 해당하는 150개의 데이터 만으로도 주어진 모든 데이터를 사용하여 학습하는 경우와 비슷한 성능을 낼 수 있다. 또한 본 논문에서 제시한 비교사 학습을 이용한 능동적 학습 방법은 120개의 데이터 만을 사용하여 피동적 학습 방법과 비슷한 성능을 낼 수 있다. [그림 4]에서 보듯이 기존의 능동적 학습은 *Decision boundary* 근처에 있는 데이터들을 선택한다. 비교사 학습을 이용한 방법은 *Decision boundary* 근처의 데이터를 선택할 것이라는 기존의 예상과는 달리 주어진 데이터들 가운데 중복되지 않으면서 전체 데이터 분포를 대표하는 것들을 선택하여 학습하는 것을 볼 수 있다([그림 3] 참조).

5. 결론 및 향후 연구 방향

본 논문에서는 비교사 학습을 이용한 능동적 학습 방법을 제안하였다 제시된 방법의 장점은 기존의 방법이 복잡한 계산이나 반복적인 계산이 필요한데 반해 간단한 계산만으로 기존의 방법보다 학습 속도를 향상시킬 수 있다는 것이다. 본 논문에서는 인위적으로 생성된 데이터에 대한 실험만을 하였다 앞으로의 과제는 첫째, 실제 문제에도 제시된 방법을 적용하여 다각적인 분석을 하며 둘째, 제시한 방법을 좀 더 개선시켜 *Decision boundary*에 가까운 데이터를 선택하는 방법을 모색하며 마지막으로 능동적 학습방법을 발전시켜 데이터의 질을 평가하는 방법을 찾아보는 것이다.

참고 문헌

[1] D. Cohen, L. Atlas, R. Ladner, Improving Generalization with Active Learning, *Machine Learning*, Vol 15, pp 201-221, 1991  
 [2] A. Robel, The dynamic pattern selection algorithm: Effective training and controlled generalization of backpropagation neural networks, *Technische Universität Berlin, Germany, Tech. Rep.*, 1994  
 [3] B-T. Zhang, Accelerated Learning by Active Example Selection *International Journal of Neural Systems*, Vol 5, No 1, pp 67-75, 1994  
 [4] M. Pfitrowski, H. Witte, Selecting Concise Training Sets from Clean Data, *IEEE Transactions on Neural Networks*, Vol. 4 No 2, pp 305-318, 1993  
 [5] A. Engelbrecht, I. Cloete, Selective Learning using Sensitivity Analysis, *IJCNN98*, Vol2, pp 1150-1155, 1998  
 [6] S. Kullback, *Information Theory and Statistics*, New York Wiley, 1959