

# 베이지안 신경망을 활용하는 온라인 이미지 인식을 통한 순차 추정

손성호<sup>01</sup> 김지섭<sup>1</sup> 장병탁<sup>1</sup>

<sup>1</sup>서울대학교 컴퓨터공학부

shson@bi.snu.ac.kr, jkim@bi.snu.ac.kr, btzhang@bi.snu.ac.kr

## Online Image Recognition Using Bayesian Neural Network for Sequential Estimation

Seongho Son<sup>01</sup> Jiseob Kim<sup>1</sup> Byoung – Tak Zhang<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering, Seoul National University

### 요 약

온라인 학습 문제는 적은 데이터가 순차적으로 주어지며, 기존 데이터에 대한 학습을 마친 모델이 추가된 새 데이터를 학습하는 경우 학습 정보가 왜곡되어 파멸적 망각 현상이 발생하는 등 성능이 저하될 위험이 있다. 본 연구에서는 베이지안 이론 체계를 기반으로 하여 연결 가중치가 확률분포에 대응되는 베이지안 신경망에 재귀 베이지안 추정 기법을 결합한 온라인 학습 모델을 제시한다. 제시된 학습 모델은 베이지안 역전파 기법을 통해 각 확률분포의 매개변수를 학습하는 한편 데이터가 새로 주어질 때마다 신경망의 사전확률분포를 갱신하는 방법을 사용하며, 본 연구에서는 MNIST 데이터를 이용한 실험을 통해 베이지안 신경망을 이용한 온라인 학습이 가능함을 검증하였다.

### 1. 서론

온라인 학습 문제에서는 배치 학습에 비해 상대적으로 적은 양의 데이터를 순차적으로 학습한다. 이런 환경은 기존 데이터에 대한 학습 결과와 새로운 데이터에 대한 유연한 적응 사이의 균형을 맞춰야 하는 안정성/가소성 딜레마(Stability/Plasticity Dilemma)를 불러온다 [1]. 신경망을 그대로 온라인 학습에 적용하면 새로 학습하는 데이터가 이전의 데이터에 대한 학습 정보를 왜곡하는 파멸적 망각(Catastrophic Forget) 문제가 발생하여 이전에 학습한 데이터에 대한 성능이 저하된다. 기존의 연구는 파멸적 망각 문제를 해결하기 위해 신경망이 이미 학습한 데이터와 유사한 데이터를 만들어 새로운 데이터와 함께 학습하는 허위시연(Pseudorehearsal) 기법 등을 제시하였으나[2], 명확한 이론적 체계를 갖추지 못한 채 경험적인 기법에 의존했다는 한계를 지닌다.

베이지안 신경망(Bayesian Neural Network)에서는 각 가중치가 서로 다른 확률분포로 표현되어 학습 환경에

서의 과도적응 문제를 완화한다[3]. 또한 베이지안 신경망은 학습 및 추론 과정에서 각 매개변수의 사전 확률분포(Prior Probability Distribution)와 사후 확률분포(Posterior Probability Distribution)를 이용하는데, 이는 직전 상태의 사후확률분포를 다음 상태의 사전확률분포로 이용해 미지의 확률밀도함수를 추정하는 재귀 베이지안 추정(Recursive Bayesian Estimation) 기법을 결합했을 때 온라인 학습 모델로 응용될 수 있음을 시사한다.

본 연구에서는 온라인 이미지 인식을 효율적으로 수행하기 위해 베이지안 신경망에 재귀 베이지안 추정 기법을 결합하여 신경망의 사전확률분포가 주기적으로 갱신되는 모델을 제안한다. 그 후 제안된 모델이 온라인 이미지 인식 문제에 응용되었을 때의 성능을 MNIST 이미지 데이터를 활용한 실험으로 검증한다.

논문의 구성은 다음과 같다. 2장에서는 본 연구의 학습 모델을 이루는 재귀 베이지안 추정과 베이지안 신경망을 소개하며, 두 기법이 결합되어 온라인 학습을 수행

하는 원리를 설명한다. 3장에서는 MNIST 데이터를 이용한 실험 결과를 제시하며, 4장에서 결론을 내린다.

## 2. 학습 모델

### 2.1. 재귀 베이지안 추정

재귀 베이지안 추정은 상태(state)  $x_0 \sim P(x_0)$ 들이 마르코프 과정에 따라 변화할 때 특정 시점의 상태  $x_t (t = 0, 1, \dots)$  를 추정하는 것이다[4]. 사전확률  $P(x_0)$ , 전이 확률(Transition Probability)  $P(x_{t+1}|x_t)$ , 우도(Likelihood)  $P(y_t|x_t)$  가 주어졌을 때 시점  $t$  까지의 측정값  $y_{1,2,\dots,t}$ 를 모은 벡터를  $Y_t$  라 하면

$$P(x_t|Y_t) = \frac{P(y_t|x_t)P(x_t|Y_{t-1})}{P(y_t|Y_{t-1})},$$

$$P(x_{t+1}|Y_t) = \int_{R^n} P(x_{t+1}|x_t)P(x_t|Y_t) dx_t$$

가 성립한다. 이는 재귀 베이지안 추정 문제에서  $Y_{t-1}$ 가 주어졌을 때  $x_t$ 를 얻을 확률  $P(x_t|Y_{t-1})$  이 다음 상태  $x_t$  에서의 사전확률로서 기능함을 보이고 있다.

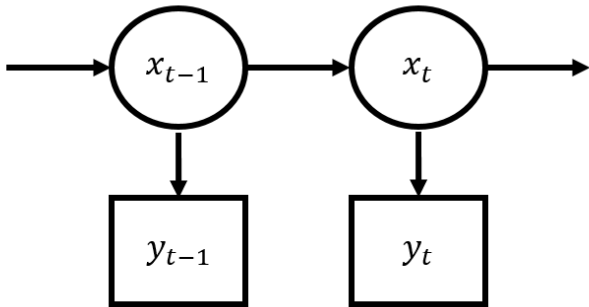


그림 1. 재귀 베이지안 추정의 시각적 모델

### 2.2. 베이지안 역전파 (Bayes by Backprop)

신경망에서의 베이지안 추론은 학습 데이터가 주어졌을 때 신경망의 가중치에 대한 사후 확률  $P(w|D)$ 를 계산하여 수행한다. 이 사후확률분포는 주어진 테스트 데이터  $\hat{x}$  가  $\hat{y}$  로 분류될 확률  $P(\hat{y}|\hat{x}) = E_{P(w|D)}[P(\hat{y}|\hat{x}, w)]$  을 구하는 데 이용된다. 이때 변분법적 학습(Variational Learning)을 이용하여  $P(w|D)$ 와의 쿨백-라이블러 괴리도(Kullback-Leibler Divergence)를 최소화하는 분포  $q(w|\theta)$  의 매개변수  $\theta$  를 찾는다[3].

$$\begin{aligned} \theta^* &= \operatorname{argmin}_{\theta} \operatorname{KL}[q(w|\theta)||P(w|D)] \\ &= \operatorname{argmin}_{\theta} \int q(w|\theta) \log \frac{q(w|\theta)}{P(w)P(D|w)} dw \end{aligned}$$

$$= \operatorname{argmin}_{\theta} \operatorname{KL}[q(w|\theta)||P(w)] - E_{q(w|\theta)}[\log P(D|w)] \quad (1)$$

식 (1)의 괴리도를  $F(D, \theta)$ 라 하고,  $w^{(i)}$ 를 변분법적 사후확률분포(Variational Posterior)  $q(w|\theta)$ 에서 추출한  $i$  번째 몬테 카를로 표본이라고 할 때  $F(D, \theta)$ 는 아래와 같이 근사할 수 있다.

$$\begin{aligned} F(D, \theta) &= \operatorname{KL}[q(w|\theta)||P(w)] \\ &\quad - E_{q(w|\theta)}[\log P(D|w)] \\ &\approx \sum_{i=1}^n \log q(w^{(i)}|\theta) - \log P(w^{(i)}) - \log P(D|w^{(i)}) \end{aligned} \quad (2)$$

변분법적 사후확률분포와 사전확률분포를 대각 가우시안분포라고 가정하고, 사후 확률과 사전 확률의 매개변수를 각각  $\theta = (\mu, \rho)$ 와  $\theta^* = (\mu^*, \rho^*)$ 로 두며, 정규분포의 표본을  $\epsilon \sim N(0, I)$  라 할 때 식 (2)의 확률분포들을 아래와 같이 모델링하였다.  $P(D|w)$ 를 정의하는 식에서  $x$ 는 주어진 데이터,  $t$ 는  $x$ 에 대한 올바른 분류값,  $y$ 는 신경망이  $x$ 를 입력받아 내놓은 분류 결과를 뜻한다.

$$q(w|\theta) = N(\mu, \log(1 + \exp(\rho)))$$

$$P(w) = N(\mu^*, \log(1 + \exp(\rho^*)))$$

$$P(D|w) = P(y = t|x; w)$$

(3)

이 때 표본추출이 개입된 가중치 계산을 편리하게 하기 위해 가중치를 아래와 같이 모델링한다.

$$w = \mu + \log(1 + \exp(\rho)) \circ \epsilon \quad (4)$$

베이지안 신경망으로 매개변수  $\theta$  를 학습할 때는 식 (4)의 가중치를 이용해 식 (2)의 비용함수에 경사 하강 기법(Gradient Descent)을 적용하여 수행한다.

### 2.3. 베이지안 신경망을 이용한 온라인 학습

선행 연구에서는 베이지안 역전파를 통해 신경망의 매개변수를 조정할 때 사전확률  $P(w)$ 의 분포를 고정하는 것이 사후확률  $q(w|\theta)$ 의 분포와 함께 조정되는 것보다 더 높은 인식 성능을 보이는 것을 보고한 바 있다 [3]. 이는 사전확률분포가 조정되도록 허용하는 경우 사전확률분포의 매개변수들이 학습 초기의 추정된 분포를 빠르게 학습하여 성능을 떨어뜨리기 때문이다.

본 연구에서는 효율적인 온라인 학습을 위해 베이지안 신경망에 재귀 베이지안 추정기법을 결합한다. 신경

망이 주어진 미니배치(Minibatch)의 데이터를 기반으로 사후확률의 매개변수  $\theta$ 에 대한 경사 하강을 수행하는 동안 사전확률분포  $P(w)$ 의 매개변수  $\theta^*$ 는 고정된다. 신경망이 사후확률분포의 매개변수 학습을 마치면 기존의 사전확률 매개변수는 폐기되고 사후확률의 매개변수가 복사된다. 즉, 신경망이 새로운 미니배치 데이터를 이용한 학습을 수행하기 전  $\theta^* \leftarrow \theta$ 가 일어난다.

### 3. 실험 결과 분석 및 논의

#### 3.1. 신경망 구성

본 연구에서 실험한 신경망은 학습하는 이미지마다 가중치에 대응되는 확률분포에서 표본을 10개씩 추출하여 비용함수를 계산한 값의 평균을 학습에 활용하였다. 각 층의 활성화함수(Activation Function)로는 은닉층에서 시그모이드(Sigmoid) 함수를, 출력층에서 출력값이 10개인 소프트맥스(Softmax) 함수를 사용하였다. 신경망이 한 번에 학습하는 미니배치의 크기는 250이며, 최초의 두 미니배치에서는 일반 신경망을 활용하여 가중치를 학습한 뒤 베이지안 신경망에서의  $\mu$ 값으로 이용하였다.

#### 3.2. MNIST 데이터 분류

본 연구에서는 MNIST 숫자 필기체 데이터[6]를 활용하여 베이지안 신경망의 감독학습을 수행하였다. MNIST 데이터셋은 각 변의 길이가 28픽셀인 정사각형 이미지로 이루어져 있으며, 60,000개의 학습용 이미지와 10,000개의 테스트용 이미지로 이루어져 있다.

그림 2는 2장에서 제시한 비용함수(f)와 비용함수를 구성하는  $\log q(w|\theta)$  (q),  $-\log P(w)$  (-p),  $-\log P(D|w)$  (-l)가 학습이 진행됨에 따라 값이 변화하는 양상을 그래프로 나타낸 것이다. 그림 2에서는 최초 미니배치에서 비용함수의 그래프가 데이터 우도의 그래프와 유사한 모양을 보이고 있는데, 이는 비용함수를 미분하여 구하는 그래디언트(Gradient)에서 데이터 우도가 차지하는 비중이 높았음을 보여준다. 반면 사전확률  $P(w)$ 는 첫 번째 미니배치가 끝난 뒤 크게 증가하여 비슷한 값을 유지하는데, 이는 최초의 미니배치에서 사후확률분포가 학습되는 과정에서 매개변수의 값이 비용함수에 미치는 영향이 크게 줄어들었음을 시사한다.

그림 3은 데이터를 학습한 베이지안 단층 신경망의 입력층 가중치 평균값을 그림으로 나타낸 것이며, 학습

이 어느 정도 진행됨에 따라 신경망의 확률분포가 데이터의 특성을 반영하게 됨을 보이고 있다.

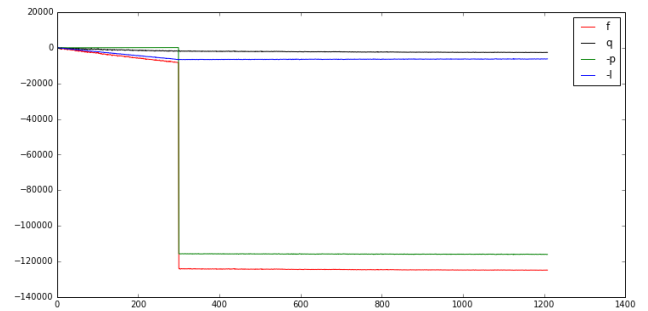


그림 2. MNIST 데이터 실험의 비용함수 그래프

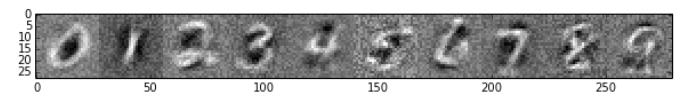


그림 3. 단층 신경망의 가중치 평균값 그래프

### 4. 결론

본 연구에서는 온라인 이미지 인식을 위해 변분법적 베이지안 학습을 기반으로 한 베이지안 역전파 기법과 재귀 베이지안 추정 기법을 결합한 신경망 모델을 제시하였으며, MNIST 데이터를 대상으로 한 실험을 수행하여 그 성능을 검증하였다.

### 5. 감사의 글

이 연구는 삼성전자의 지원을 받았으며, 2016년도 정부(미래창조과학부, 국방부)의 재원으로 정보통신기술진흥센터(R0126-16-1072-SW스타랩), 한국산업기술평가관리원(10044009-HRI.MESSI, 10060086-RISF), 국방과학연구소(UD130070ID-BMRR)의 지원을 일부 받았음.

### 참고문헌

[1] Robins, A., "Sequential learning in neural networks: A review and a discussion of pseudorehearsal based methods", *Intelligent Data Analysis* 8.3, 301-322., 2004.  
 [2] Ans, B. & Rousset, S., "Neural networks with a self-refreshing memory: knowledge transfer in sequential learning tasks without catastrophic forgetting", *Connection science* 12.1: 1-19, 2000.  
 [3] Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D., "Weight uncertainty in neural networks", *arXiv preprint arXiv:1505.05424*, 2015.  
 [4] Bergman, N., "Recursive Bayesian Estimation", *Department of Electrical Engineering, Linköping University, Linköping Studies in Science and Technology. Doctoral dissertation*, 579, 1999.  
 [5] Graves, A., "Practical variational inference for neural networks", In *Advances in Neural Information Processing Systems* (pp. 2348-2356), 2011.  
 [6] Yann, L., Corinna C., "The MNIST database of handwritten digits", URL <http://yann.lecun.com/exdb/mnist/>, 1998.