

## 예측 불확실성과 이를 나타내는 새로운 지표

우제훈 김선규 강재우 (고려대학교)

gnswp21@korea.ac.kr

## Prediction uncertainty and new indicator of it

## 요 약

모델의 예측에 대한 신뢰도를 표현하는 것으로 우리는 reject 옵션을 효과적으로 고려할 수 있게 된다. 예측에 대한 신뢰도는 예측 불확실성으로 연구되고 있으며, 예측 불확실성은 데이터 불확실성과 모델 불확실성으로 나타낼 수 있다. 데이터의 불확실성은 모델이 logit값을 예측할 때, standard deviation을 같이 구하여 noised logit이 output이 되도록 모델을 설계하여 포착하였다. 모델 불확실성은 MC drop out 방식으로 logit의 standard deviation을 통해 포착하였고 이 둘을 동시에 포착하고 normalize term을 추가함으로써 예측 불확실성을 나타내는 새로운 지표를 소개하였다.

## 1. 서 론

기계학습 모델의 예측에 대한 신뢰도를 표현할 수 있다면 모델의 판단을 따를 지 아닐 지에 대해 고려할 수 있다. 암환자의 판정에 대해 모델의 신뢰도가 낮다면 모델의 판단을 reject 하여 전문가의 판단을 따르고, 반대로 높다면 모델의 판단을 따르는 것이 가능해져 안전하고 효과적으로 모델을 사용할 수 있게 된다. 따라서, 우리는 예측 불확실성(uncertainty)을 포착하는 모델을 설계해야 한다. 이번 논문에서 데이터 불확실성과 모델 불확실성을 포착하는 기계학습 모델을 구성하여 예측 불확실성을 포착하는 실험을 진행하였고 기존의 불확실성 계산식에 normalize term을 추가하는 것으로 새로운 방식으로 예측 불확실성을 포착하는 지표를 고안했다. Normalize term을 추가하였을 때, 기존의 방식과 유사하거나 어느 정도 향상된 성과를 보였다.

## 2. 불확실성

예측에 대한 신뢰도를 포착하는 방법은 예측 불확실성이라는 이름으로 연구되고 있다. 예측불확실성을 포착하기 위해 두가지 불확실성을 도입하여, 모델 불확실성 (epistemic uncertainty)과 데이터의 불확실성 (aleatoric uncertainty)으로 예측 불확실성을 설명하려는 연구들이 있었다. [1] 모델 불확실성은 모델이 얼마나 실제의 모델과 다른 지 나타낸다. 데이터의 불확실성은 예측 데이터가 얼마나 학습시 데이터와 다른 지 나타낸다. 모델의 불확실성은 쉽게 연상할 수 있다. 기계 학습의 근간을 이루는 방식이 데이터를 통해 우리가 알고 싶은 모델을

근사하는 것이기 때문에 실제 모델과 근사한 모델의 차이를 생각한다면 그것이 모델 불확실성이다. 데이터의 불확실성의 경우, 데이터를 얻는 과정을 고려해 보면 된다. 대부분 센서나 전문가의 조사를 통해 데이터를 얻는 경우가 빈번하다. 여기서 데이터의 불확실성이 발생한다. 데이터에는 언제나 노이즈가 발생한다. 간단하게  $y=f(x)$ 인 모델을 고려해 보자. 이 때,  $f$ 가 모델을  $x$ 가 데이터를 의미한다.  $f$ 와  $x$ 에 불확실성이 있다면 이는 식을 통해  $y$ 로 흘러간다. 이런 개념을 바탕으로 모델의 불확실성과 데이터의 불확실성을 포착하여 예측의 불확실성으로 최종적으로 나타낼 수 있다.

## 3. 불확실성 포착을 위한 모델

데이터의 불확실성은 예측 logit값의 분산으로 계산할 수 있다. 모델의 불확실성이 없다는 가정하에 모델이 deterministic하다면 데이터의 불확실성만이 예측의 불확실성으로 이어진다. 이렇게 간접적으로 데이터의 불확실성을 조사하는 이유는, 데이터 자체로는 이 데이터에 어느 정도의 노이즈가 있는지 알 수 없기 때문이다. NN의 마지막 fully connected layer에서 예측하는 output의 수만큼 그에 대한 standard deviation을 같이 모델이 학습하도록 한다.[1]

$$\hat{x}_{i,t} = f_i^W + \sigma_i^W \epsilon_t, \epsilon_t \sim N(0, I)$$

( $i$ 는 class,  $f$ 는 모델이 예측한 logit 값,  $\sigma$ 는 standard deviation 값,

$t$ 는 가우시안 분포를 따른 값을 만들기 위한 sampling에 따른 index)

Loss function 에 이러한 standard deviation 이 같이 적용됨으로써 logit 과 standard deviation 을 같이 구한다. Logit 값의 standard deviation 이 높은 값일수록 데이터의 노이즈가 많고 모델이 예측에 대해 더 적은 신뢰를 가지고 있다고 판단할 수 있다. 하지만 standard deviation 은 logit 값의 의존적이기 때문에 logit 값이 단순히 높은 것으로 standard deviation 도 따라 높아지게 된다. 따라서, logit 간의 normalized 된 standard deviation 을 원래의 값에 각각의 logit 값으로 나눠주는 것으로 구했다.

$$Pred.Uncertainty = normal(std)$$

모델 불확실성을 표현하는 한 방식으로 베이지안 딥러닝 기법이 있다. 기존의 딥러닝 기법이 deterministic 하게 파라미터를 정하던 것에 반해, 베이지안 딥러닝 기법은 파라미터의 distribution 을 추정한다. Maximum Likelihood Estimation(MLE) 방식과 파라미터의 사전분포를 함께 고려하는 Maximum A Posterior 방식(MAP)으로 파라미터를 학습한다는 의미이다. 파라미터의 분포로 예측의 분포를 구할 수 있고 이 분포의 standard deviation 으로 불확실성을 표현할 수 있다. 그러나 MAP 방식은 MLE 방식에 비해 자원이 너무 많이 사용되어 일반적으로 MAP 방식이 아닌 MLE 방식을 주로 사용한다. 이에 대한 해결책으로 Monte Carlo (MC) drop out 방식을 사용하여 MAP 방식을 근사하는 방식이 있다.[2] MC drop out 은 각각의 레이어에 걸쳐 있는 dropout layer 를 예측에서도 그대로 사용하여 여러 개의 모델을 통해 예측하는 ensemble 의 효과를 내어 예측에 대한 sampling 이 가능해진다. 이를 normal 분포로 추정하여 예측에 대한 분포를 나타내는 것이다. MC drop out 방식은 sampling 된 logit 값의 standard deviation 을 예측 불확실성으로 사용한다.

$$Pred.Uncertainty = std(logit)$$

이번 실험에서는 추가로 앞서 다룬 데이터의 불확실성, 즉 각각 logit 값의 standard deviation 을 같이 구함으로써 MC drop out 한번 한번의 sampling 결과를 mean 과 variance 를 갖는 gaussian 분포로 취급할 수 있었고 최종 예측에 사용하는 값을 Mixture Of Gaussian(MOG)으로 취급할 수 있다. 이에 따라, MOG의 standard deviation 인 sampling logit 의 standard deviation 과 sampling standard deviation 의 mean 의 합을 구하고 이를 예측의 불확실성을 나타낼 수 있었다.

$$\begin{aligned} Pred.Uncertainty \\ &= std(logit) + mean(normal(std)) \\ &= Model Uncertainty + Data Uncertainty \end{aligned}$$

앞서 데이터의 불확실성을 구한 경우에 standard deviation 을 logit 값에 대해 normalize 하는 과정을 거쳤는데, 이번에도 이를 사용하여 sampling logit 값의 standard deviation 이 모델 불확실성을 normalized sampling standard deviation 의 평균이 데이터 불확실성을 포착해주는 지표가 되게 해주었다

#### 4. 실험

ACCURACY Table

	CNN	Alexnet	VGG11
Base	0.92	0.92	0.92
Data	0.91	0.92	0.92
Data, Reject(20%)	0.92	0.94	0.92
Data + Model	0.91	0.92	0.92
Data + Model, Reject(20%)	0.94	0.96	0.96

실험에 사용된 데이터는 FashionMNIST 데이터셋이다. FashionMNIST 데이터셋은 운동화, 셔츠, 샌들과 같은 작은 이미지들의 모음이며, 열 가지로 분류될 수 있는 28×28 픽셀의 이미지 7 만개로 이루어져 있다. 이 중에서 5 만개의 데이터로 학습하고 1 만개로 validation 나머지 1 만개 데이터를 통해 test 하였다.

실험의 모델은 Convolution Neural Net (CNN), Alexnet, Vgg11 모델이 사용되었으며, 기본적인 모델에서 데이터 불확실성을 추가한 모델, 데이터 불확실성과 모델 불확실성을 추가한 모델에 대해 10 가지 물체에 대한 분류 작업을 진행하였다. 추가로 예측 불확실성을 계산하고 불확실성이 높은 순으로 reject 함으로써 모델의 성능이 변화하는 정도를 파악했다.

#### 5. 실험 결과

모델에 대한 실험의 결과는 표 1과 같다. 불확실성을 포착하는 모델들의 성능은 기존 모델에 비해 변화가 없거나 오히려 낮아졌다. 이는 모델이 10개의 class의 logit값을 추론하던 것에서 추가로 10개의 standard deviation을 추론하는 것으로 더욱 복잡해지기 때문이다. 그러나, 20%의 reject 옵션을 고려한 경우 성능이 향상된 모습을 볼 수 있다. 그림 1의 1행의 그래프들을 통해 데이터 불확실성의 경우 standard deviation이 normalized 된 값이 예측 불확실성을 더욱 잘 포착하는 것을 볼 수 있다. 심지어 단순히 standard deviation을

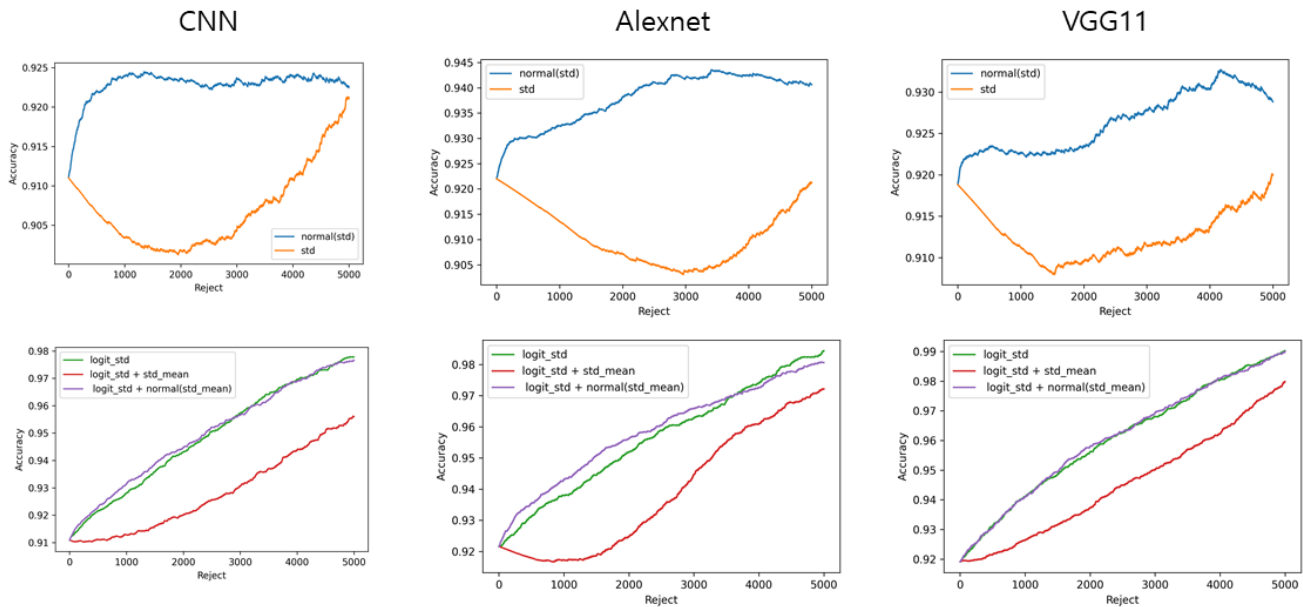


그림 1

예측 불확실성으로 취급한다면 logit값이 높은 값을 reject하는 경우가 많아지므로 오히려 accuracy가 떨어지는 모습을 볼 수 있다. 이를 통해 standard deviation을 normalize하는 방법으로 logit값으로 나눠준 방식이 효과가 있음을 확인하였다.

이와 같은 경향은 MC drop out 방식에도 이어진다. 우선 예측 불확실성으로 sampling logit의 standard deviation만 사용한 경우 (그림1의 초록선)가 sampling standard deviation의 평균도 같이 사용한 경우(그림1의 빨간선)보다 더 좋은 지표로 작용했음을 알 수 있다. 이는 예측의 불확실성으로 예측 logit값의 standard deviation을 그대로 사용하는 것은 좋은 성능을 내지 못한다는 것을 의미한다. 따라서, sampling standard deviation 값에 normalize를 해준 결과 (그림1의 보라선) sampling logit의 standard deviation만 사용한 경우와 비슷하거나 약간 더 나은 성능을 보여주었다. 이처럼 적절한 normalize term을 통해 데이터 불확실성과 모델 불확실성을 모두 포착하는 더 좋은 성능으로 예측 불확실성을 나타내는 것이 가능했다.

## 6. 결론

예측 불확실성과 예측 불확실성을 포착할 수 있는 모델을 소개하고 그에 따른 reject 옵션에 대해 알아보았다. 예측 불확실성을 데이터 불확실성과 모델 불확실성으로 나누어 포착할 수 있었고, 데이터 불확실성을 표현할 때는 standard deviation에 함께 구한 logit 값으로 normalize 해준 것이 더 정확히 불확실성을 포착하는 지표가 되는 것을 실험을 통해

확인하였다. 이러한 경향성은 데이터 불확실성과 모델 불확실성을 모두 포착하는 방식에도 이어졌다. 이번 실험을 통해 예측 불확실성을 포착하는 대한 새로운 지표에 대해 알아볼 수 있었다.

## 7. 사사의 글

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업이며 (NRF-2020R1A2C3010638) 국가초고성능컴퓨팅센터로부터 초고성능컴퓨팅 자원과 기술지원을 받아 수행된 연구성과임 (KSC-2019-CRE-0186). 또한, 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT 명품인재양성사업의 연구결과로 수행되었음 (IITP-2021-2020-0-01819).

## 8. 참고 문헌

- [1] Kendall, A., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision?. *arXiv preprint arXiv:1703.04977*.
- [2] Gal, Y., & Ghahramani, Z. (2016, June). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (pp. 1050-1059). PMLR.