

2022년 2학기 빅데이터기계학습 Final Exam

양소연

1. **Deep learning이 logistic regression, support vector machine, decision tree 등과 같은 전통적 machine learning 알고리즘들에 비해 성능이 좋은 이유를 간략히 설명하라. (6점)**

데이터가 부족했던 과거에는 전통적인 머신러닝 기법을 사용하였지만 오늘날에는 데이터의 양이 증가하고 있으며 이에 따라 딥러닝을 사용하는 것이 좋은 학습효과를 보여준다. 딥러닝의 경우 머신러닝과 달리 신경망(neural network)을 사용한다는 차이점이 있고 빅데이터 학습에 적합한 learning 기법이다. Neural network를 구성하는 neuron은 데이터의 feature값을 입력 받는다. 지도학습에서 딥러닝을 사용할 경우 특정 문제에 머신러닝 기법을 적용하여 가치를 향상시킨다. Neural network의 hidden layer에서는 고차원의 feature값을 찾아주며 뉴런의 수보다 layer의 수가 딥러닝의 성능을 결정한다.

2. **Gradient descent는 무슨 용도로 사용되는 알고리즘인지 간략히 쓰고, 주어진 neural network에 대해 대강 어떻게 동작하는 알고리즘인지 큰 그림만 간략히 설명하라. (6점)**

Gradient descent는 cost function을 최소화하는 파라미터 w (가중치)와 b (편향)값을 찾는 용도로 사용되는 알고리즘이다. Gradient descent는 경사가 가장 높은 곳에서부터 시작하여 global optima에 도달 할 때까지 반복적으로 실행된다. 해당 알고리즘을 반복적으로 거치면 파라미터 값인 w, b 가 변하면서 경사도가 가장 완만한 global optima를 찾는다.

3. **Backpropagation은 무슨 용도로 사용되는 알고리즘인지 간략히 쓰고, 주어진 neural network에 대해 대강 어떻게 동작하는 알고리즘인지 큰 그림만 간략히 설명하라. (6점)**

Backpropagation은 neural networks에서의 기울기나 미분값을 구하기 위하여 사용되는 알고리즘이다. Neural network에서의 output layer에서 input layer 방향으로 계산하며 weight 값을 업데이트 한다. Backpropagation에서는 연쇄법칙을 사용하는데 output layer와 hidden layer사이의 weight의 미분값을 받아서 앞에 있는 hidden layer에 배포하는 방식이다. 이것은 매개변수 가중치를 더 빠르게 계산하여 연산 처리 속도를 높이기 위해 사용된다.

4. **Neural network에서 layer가 많아질수록 발생하는 장점과 단점을 각각 간략히 쓰라. (6점)**

Neural network에서 layer가 많아지면 각 layer에서 feature 값을 추출할 수 있으므로 (부족한 것보다) 정교한 모델이 되어 성능이 향상된다는 장점이 있으나 layer가 지나치게 많아지면 연산이 느려지고 과적합이 발생할 확률이 높아진다. 이 뿐만 아니라, 특정 값 이상이 지나면 더 이상 학습되지 않는다는 단점이 있다. 따라서 적절한 layer 수를 구성하여야 모델의 성능이

향상된다.

5. **Regularization이 필요한 이유를 간략히 쓰고, regularization의 대표적인 기법 중 하나인 dropout의 동작과 원리를 각각 간략히 설명하라. (6점)**

딥러닝에서는 layer가 쌓임에 따라 강력한 학습 모델이 만들어지지만 이에 따라 과적합이 발생될 가능성 또한 높아진다. Regularization은 과적합을 줄이는 방법이며 dropout은 regularization의 대표적인 기법 중 하나이다. Dropout은 layer의 node 중 일부를 연산에서 제외시켜 correlation을 제거하는 동작이다. Dropout의 각 feature값에 의존하지 않고 weight를 분산시켜 과적합 문제를 해결하는 원리다.

6. **Gradient descent, momentum, RMSprop, Adam 간의 관계를 간략히 쓰라. (6점)**

문제에서 언급된 알고리즘은 neural network가 빠르고 정확하게 학습하는 것을 목표로 하는 최적화 알고리즘이라는 공통점이 있다. Gradient descent에서는 mini-batch size값을 설정하여 mini-batch의 수만큼 반복하여 최적 값을 찾아간다. 다만 mini-batch size는 너무 크거나 작지 않아야 모델의 성능이 향상된다.

Momentum은 gradient descent의 지수 이동 가중평균 값을 사용하여 weight를 업데이트 시키고 이 기법은 Gradient descent보다 속도가 빠르다.

RMSprop은 과거 step에서의 gradient보다 최근 step에서의 gradient에 가중치를 부여하는 방법이며 w^2 미분값의 평균과 b^2 미분값의 평균을 나누는 방식이다.

Adam은 Momentum과 RMSprop을 합친 알고리즘이다.

7. **Batch normalization은 무슨 용도로 사용되는 알고리즘인지 간략히 쓰고, 주어진 neural network에 대해 대강 어떻게 동작하는 알고리즘인지 큰 그림만 간략히 설명하라. (6점)**

Batch normalization은 hyper parameter 탐색을 더 쉽게 만들어주고 neural network를 더욱 견고하게 하는 알고리즘이다. Batch norm은 neural network에 대해 normalization이 적용된 activation function을 통과하며 이 과정은 각 layer마다 적용된다. Batch normalization은 feature값을 normalize하여 학습속도를 증가시킨다는 장점이 있다.

8. **Standard neural network보다 CNN을 쓰는 것이 더 유리한 경우가 언제이고 그 이유는 무엇인지 간략히 설명하라. (6점)**

CNN은 computer vision 문제에 유리하다. 그 이유는 padding을 통해 데이터를 보존할 수 있음과 동시에 boundary data도 계산하기 때문이다. 또한 Standard neural network보다 적은 수의 파라미터 값으로 여러 개의 feature 값을 추출할 수 있어서 연산이 빠르며 위치에 관계없이 detect할 수 있다는 장점이 있다. 또한 각각의 layer와 sparse하게 연결되어 있어 과적합

이 될 가능성을 낮춰준다.

9. 다음 각 CNN 모델에 대해 기존 모델과 다른 해당 모델만의 핵심 아이디어를 간략히 쓰라. (6점)

- ① VGG-16 : 3*3 filter를 기반으로 하는 network를 가지고 max pooling한 network를 반복적으로 사용하여 많은 layer를 쌓는다.
- ② ResNet : residual block을 사용하여 network를 구성하며 skip connection을 이용하여 효과적으로 backpropagation을 한다.
- ③ Inception network : 서로 다른 크기의 filter의 feature값들을 max-pooling하여 가로, 세로를 줄이고 여러 개의 filter를 concat하여 내보낸다.
- ④ MobileNet : depth wise와 point wise를 사용하여 layer의 사이즈를 늘리고 그 안에서 feature값을 추출한다. 그 후 다시 압축하여 계산량을 줄인다.

10. YOLO가 CNN 기반의 다른 object detection model들에 비해 속도가 빠른 이유를 간략히 설명하라. (6 점)

YOLO는 이미지 전체를 단 한번(only once)만 보기 때문에 다른 object detection model들에 비해 속도가 빠르다. YOLO는 우선적으로 이미지의 input값을 resize하고 convolutional neural network에 이미지 데이터를 한 번만 통과시킨다. 해당 값을 기반으로 feature에 대한 분류 값을 추출하고 non-max suppression를 통해 output를 받는다.

11. Neural style transfer에서 content image C와 style image S가 주어졌을 때, 어떤 image G가 C의 content와 S의 style을 얼마나 따르고 있는지 측정하는 방법을 간략히 설명하라. (6점)

$$J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G)$$

Content image와 style image의 단위가 다르기 때문에 α 와 β 의 가중치를 주어 계산한다. $J_{content}(C, G)$ 는 content cost이며 C와 G의 값이 얼마나 일치하는지, 그리고 $J_{style}(S, G)$ 은 style cost이며 S와 G의 값이 얼마나 일치하는지를 나타낸다.

12. GRU와 LSTM은 이전 입력들에 대해 계산된 값 중 필요한 값을 계속 보존하기 위해 어떠한 기법을 쓰는지 각각 간략히 설명하라. (6점)

GRU에서는 memory cell을 사용하여 유지하고 싶은 값을 보존하고 LSTM에서는 gate를 통하여 새로운 것과 과거의 것을 어떻게 반영할지 비율을 정하여 보존한다. 이 둘의 다른 차이점은 gate 수다. GRU에서는 reset gate, update gate가 있지만 LSTM은 update gate, forget gate 그리고 output gate 3개로 구성되어 있다.

13. Word2Vec, negative sampling, GloVe가 생성한 word embedding이 서로 다른 이유를 각 방법이 사용 하는 모델 측면에서 간략히 설명하라. (6점)

Word2Vec에서는 단어를 벡터 값으로 바꾸어 주는 알고리즘이며 skip gram이 대표적이다. Skip gram은 특정 단어가 나타났을 때 주변 단어가 나타날 확률을 예측하는 방법이다.

Negative sampling은 단어의 위치에 기반하여 단어 간 거리의 확률을 나타낸다.

GloVe는 context와 target이 주어진 범위에서 몇 번 나타나는지 빈도를 계산하여 나타낸다.

14. Attention model이 전통적인 RNN 모델에 비해 좋은 성능을 내는 이유를 간략히 설명하라. (6점)

Attention model은 전통적인 RNN 모델에 비해 비교적 긴 문장에서도 성능이 좋다. 그 이유는 입력된 문장의 각 단어의 가중치를 부여하여 단어 token 중 어느 곳에 attention할 것인지를 보여주는 bi-directional한 모델이기 때문이다.

15. Transformer에서 self-attention과 multi-head attention이 하는 역할을 각각 간략히 설명하라. (6점)

Self-attention에서는 attention을 활용하여 각 단어의 의미를 잘 표현하는 벡터로 나타낸다. 단어 $x^{<n>}$ 에서는 query값인 $q^{<m>}$ 와 key값인 $K^{<n>}$ 을 내보내고 soft-max를 사용하여 $A(q, K, V)$ 산출한다. 모든 단어로부터 $Attention(q, K, V)$ 를 받으면 각 단어는 q 를 가지고 닮은 K 값을 찾는다. 따라서 해당 q, K, V 값은 학습으로 얻어진다.

Multi-head attention에서는 self-attention을 여러 번 반복하여 값을 산출한다. 단어 $x^{<n>}$ 에 대하여 여러 개의 $A(q, K, V)$ 값을 내보내고 CNN과 마찬가지로 해당 값들을 concat하여 사이즈를 줄여 feature값을 추출한다.

16. 수업에서 다루지 않은 문제 중 수업에서 설명한 deep learning model을 사용해서 풀어볼 만한 새로운 문제를 하나 제시하고(예: 주어진 텍스트를 이미지로 변환하는 문제, 주어진 동영상의 다음 장면을 예측하는 문제 등), 해당 문제를 풀기 위해 대강 어떤 deep learning model을 어떻게 적용할 수 있을지 간략히 설명하라. (10점)

CNN을 활용하여 결함이 있는 부품을 탐지할 수 있다. 카메라를 이용하여 detect하고자 하는 부품을 촬영한 뒤 A/B test를 진행하는 것이다. 테스트하려는 대상의 데이터가 표본 데이터의 일정 범위 이상에 오차가 있다면 그 부품은 결함이 있는 것으로 판단할 수 있다.

테스트하려는 이미지 데이터가 있다고 가정한다면 우선 이 데이터의 feature 값을 추출하여야 한다. 데이터의 feature 값을 추출하기 위해 합성곱 신경망을 거쳐 pooling layer를 쌓는다. 다음으로 fully connected layer를 추가하여 이미지의 특징을 분류한다.