

# TCT-기술인증테스트

## AI

## 서술형 문제지

[ 2020년 #차 ]

사번	성명
유의 사항	<ul style="list-style-type: none"><li>▪ 본 서술형 문제지에 제시된 시나리오와 문제를 읽고, 별도 배포되는 답안지에 응시자의 답안을 작성해 주시기 바랍니다.</li><li>▪ 배포된 문제지는 시험 종료 후 답안지와 함께 감독관에게 제출해 주시기 바랍니다.</li><li>▪ 공정한 평가를 위하여 답안 작성 시 동료를 도와 주는 행위, 보여주는 행위를 금지하고 있으며, 평가에 대한 부정행위 적발 시, 응시한 평가는 0점 처리됩니다.</li></ul>

[서술형 1번] 딥러닝을 활용한 이진(Binary) 분류 문제 해결 과정 (5점)

심층 신경망 모델(Deep Neural Networks Model)을 활용하여 지도학습(supervised learning)으로 2 클래스 분류(classification) 문제를 해결하고자 한다. 해당 태스크(Task) 예시는 텍스트 데이터인 경우 긍정, 부정 분류의 자연어 감성분석(sentiment analysis) 프로젝트, 이미지 데이터인 경우 양호, 불량 분류의 비전검사(vision inspection) 프로젝트에 해당한다. 해당 프로젝트에 투입된 철수는 판정 모델의 개발을 담당하게 되었으며, 최저 비용으로 최고의 성능(Performance<sup>주1</sup>)을 내는 모델을 개발하고자 한다.

- 철수는 고객에게 데이터를 요청하여 긍정(양호), 부정(불량)으로 라벨링(labeling)된 텍스트(이미지) 데이터 3,500개를 수령하고, 라벨(label) 품질을 확인하였다.
- 철수는 수령한 데이터를 학습(training) 데이터셋, 검증(validation) 데이터셋, 테스트(test) 데이터셋으로 나누고, 각 데이터셋의 긍정(양호)과 부정(불량) 비율이 운영 시스템에서 실제 발생하는 비율과 동일함을 확인하였다.
- 철수는 이진 분류 문제를 풀기 위해 텍스트 또는 이미지 데이터에 대해 일반적으로 사용되는 여러 개의 최고 성능(SOTA, State-Of-The-Art) 심층 신경망 모델들(Deep Neural Networks Models) 중에서 가장 큰 수용력(Capacity)을 가진 모델로 학습을 진행하였다.
- 철수는 심층 신경망 모델의 구조는 유지한 채 하이퍼파라미터(Hyperparameter)를 변경하면서 여러 번 모델을 학습시킨 후, 그 중 테스트 데이터셋을 기준으로 가장 높은 Accuracy를 보인 학습 완료된 모델을 운영 시스템에 배치(deployment)하였다.

다음의 3가지 관점에서 철수의 개발 과정에 문제가 없는지 판단하고, 문제가 있을 경우 바람직한 해결책을 서술하시오.

- 1) 데이터 준비하기
- 2) 성능 지표(Performance metrics) 정의 하기
- 3) 학습할 심층 신경망 모델 선정하기

<sup>주1</sup> 성능(Performance): 모델의 판정 수준을 통칭함(예: 정확도 등).

[서술형 2번] 딥러닝의 실용적 방법론(Practical Methods) 점검 (4점)

다음에 제시된 딥러닝의 실용적 방법론 4개에 대해 각각 맞음(○)과 틀림(X)을 판단하고(○/X로 표시), 그 이유를 간략하게 서술하시오.

- 1) 심층 신경망 모델(Deep Neural Networks Model)을 일반화(Generalization)할 수 있는 방법 중에 가장 좋은 방법은 실제 데이터를 더 많이 모아서 학습에 사용하는 것이다.
- 2) 많은 양의 데이터로 학습하였을 때 좋은 성능(Performance <sup>주)1</sup>)을 보인 심층 신경망 모델 (Deep Neural Networks Model)을 새로운 도메인의 데이터에서 학습 시키고자 한다면, 기존 학습에서 사용한 학습률 감소(Learning rate decay)를 사용하는 것이 좋다.
- 3) 여러 개의 하이퍼파라미터(Hyper-parameter)를 동시에 튜닝(tuning)하면서 최적의 조합을 찾고자 할 때는 탐색 구간 내의 후보 하이퍼파라미터 값들을 일정한 간격으로 확인하는 그리드 탐색(Grid search)보다 무작위 탐색(Random search)을 쓰는 게 더 효율적이다.
- 4) 주어진 데이터에 비해 수용력(Capacity)이 큰 모델은 과대적합(Overfitting)의 가능성이 높지만 Early Stopping을 사용해 이를 완화할 수 있다.

<sup>주)1</sup> 성능(Performance): 모델의 판정 수준을 통칭함(예: 정확도 등).

[서술형 3번] 딥러닝 응용(Application) 시스템 구축 (5점)

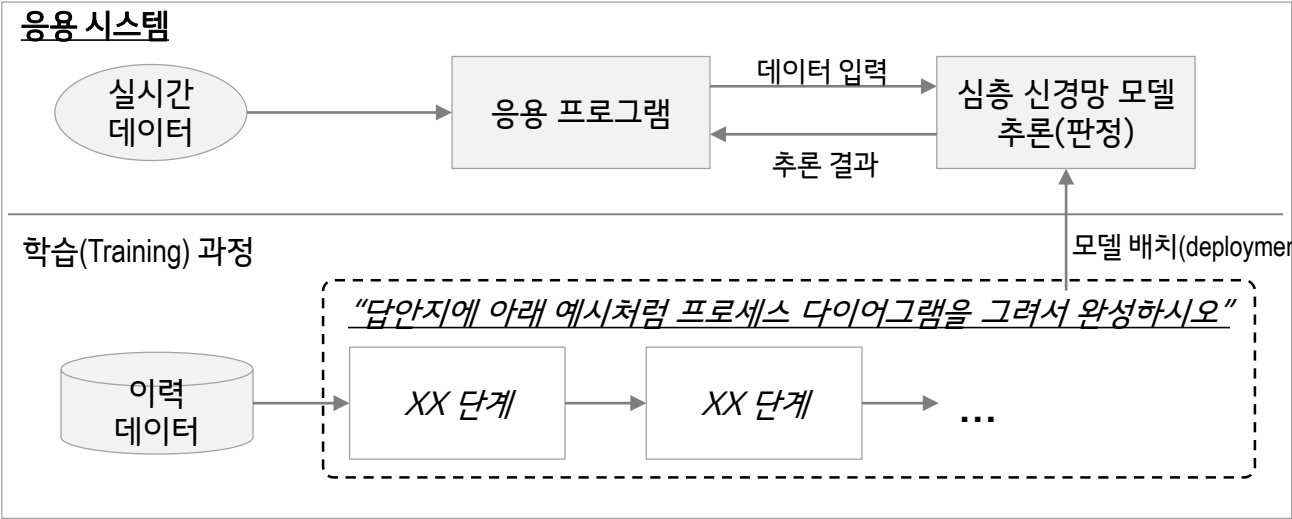
딥러닝(Deep Learning)으로 자동 판정하는 응용 시스템을 구축하고자 한다. 주어진 [고객 요구사항]을 만족하도록 아래 [그림]에서 딥러닝의 주요 학습 과정 등을 표현할 수 있도록 점선 표시 영역 안에 들어갈 1)프로세스 다이어그램을 그리고 2)각 단계별 주요 기능을 서술하시오.

※ 주의: 답안지에 프로세스 다이어그램을 그릴 것.

[고객 요구사항]

1. 심층 신경망 모델(Deep Neural Networks Model)의 성능(Performance<sup>주1</sup>)이 지속 유지되기 위해서 학습(Training) 가능한 시스템이 구축되어야 함.
2. 때로는 과거에 판정 모델로 사용되었던 심층 신경망 모델을 복구할 수 있어야 함.

[그림]



<sup>주1</sup> 성능(Performance): 모델의 판정 수준을 통칭함(예: 정확도 등).

[서술형 4번] 딥러닝 적용 여부 판단 (6점)

시각 지능(이미지 데이터) 및 언어 지능(자연어 데이터) 외에 정형 데이터(structured data, tabular data) 분석 및 예측(prediction)에 딥러닝을 적용하고자 한다.

다음 두 가지 사례를 참조하여 전통적 머신러닝 보다 딥러닝이 높은 성능(Performance<sup>주)1</sup>)을 낼 수 있는 1)데이터 차원(dimension) 또는 특징(feature)의 조건과 2)해당 조건에서 딥러닝이 높은 성능을 보이는 이유를 설명하시오.

사례1)

S은행은 『고객 신용도 예측』을 위해 5가지 컬럼(column)을 가진 100만건의 정형 데이터(tabular data)를 다양한 머신러닝, 딥러닝 모델로 분석을 시도했다. 그 결과 가장 높은 성능의 모델은 심층 신경망 모델이 아닌 XGBoost이다.

사례2)

D회사는 『단백질 구조 예측』을 위해 155가지 컬럼을 가진 50만건의 정형 데이터를 다양한 머신러닝, 딥러닝 모델로 분석을 시도했으며, 특히 컬럼간의 상호관계가 중요하여 155 X 155 가지 정보 활용이 필요했다. 그 결과 가장 높은 성능의 모델은 심층 신경망 모델 중 하나인 CNN(Convolutional Neural Networks)이다.

<sup>주)1</sup> 성능(Performance): 모델의 판정 수준을 통칭함(예: 정확도 등).

