

# Deep\_LSTM\_Manual

☰ 멘토

## ▼ 실행 방법

- C 코드 추론 진행 방법
  1. /Deep\_LSTM/main.cpp 빌드 후 실행
  2. /Deep\_LSTM/output/main.exe 실행
- Python 코드 추론 진행 방법
  1. /Deep\_LSTM/Tools/LSTM\_Infer\_Real.py 실행

## ▼ 디렉토리 구조

```
Deep_LSTM/ # LSTM C코드 프로젝트
|
|-- Data/ # 정상 및 비정상 데이터 디렉토리
| |
| | -- anormal.txt # 비정상 데이터 (128 시퀀스)
| |
| |
| | -- normal.txt # 정상 데이터 (128 시퀀스)
| |
|-- output/ # 빌드 파일 디렉토리
| |
| | -- main.exe # main.cpp 빌드 실행 파일
| |
|-- Tools/ # Python 스크립트 파일 디렉토리
| |
| | -- LSTM_Gen.py # LSTM(16,16,16,1) 테스트 모델 생성 코드
| |
| |
| | -- LSTM_Infer_Real.py # 이이랑 생성 모델 + 실제 정상 데이터 추론 코드
```

```

| |
| | -- LSTM_Infer_Test.py # 테스트 모델 + 테스트 데이터 추론 코드
| |
| |
| | -- LSTM_Weight.py # 이이랑 생성 모델 가중치 추출 코드(..Weight/ 에 저장됨)
| |
| |
| | -- test_lstm.onnx # LSTM(16,16,16,1) 테스트 모델
| |
| |
| | -- yirang.onnx # 이이랑 생성 모델
| |
|-- Weight/ # 가중치 디렉토리
| |
| | -- tensor_B_1_16.bin # 가중치 바이너리 파일
| |
| |
| | -- tensor_B_1_16.txt # 가중치 텍스트 파일
| |
| | -- ...
| |
|-- LSTM_Ops.cpp # IP로 생성하기 위한 LSTM 연산 모듈 함수가 존재하는 파일
| |
|-- main.cpp # 연산 모듈 및 Weight 파일을 바탕으로 LSTM(16,16,16,1) 구현 파일
| |
|-- WeightResize.cpp # 텍스트 형식(txt) 가중치를 배열에 저장 함수 존재
| |
|-- WeightResizeBin.cpp # 바이너리 형식(Bin) 가중치를 배열에 저장 함수 존재
| |
|-- WeighSplit.cpp # 가중치 분리(input, output, forget, update) 함수 존재

```

#### ▼ C코드 가중치 파일 이름 규칙

ex) tensor\_B\_1\_16.txt

- tensor\_{W,R,B}\_{1,2,3,4}\_{16,1}.{txt,bin}

- {W,R,B}
  - W: Weights
  - R: Recurrence Weights
  - B: Bias
- {1,2,3,4}
  - Layer 계층 순서
- {16, 1}
  - LSTM 출력 크기
- {txt, bin}
  - 텍스트, 바이너리

#### ▼ Pytorch LSTM 모델 가중치 크기 원리

- LSTM 모델의 출력의 크기(Hidden)에 따라 가중치 크기가 달라짐
- Hidden의 개수가 num일 경우
  - $W[4*num][1]$ ,  $R[4*num][num]$ ,  $B[8*num]$  정의됨
- W, R, B에 대한 정의
  - W는 입력 데이터 연산에 사용되는 가중치
  - R은 이전 데이터 연산에 사용되는 가중치
  - B는 바이어스
- $W[4*num]$  및  $R[4*num][num]$ 에서 num에 4를 곱하는 이유
  - Input, Output, Forget, Update가 하나의 가중치에 저장되어 있어서
  - $[0: num-1]$  : Input Gate 가중치
  - $[num : 2*num - 1]$  : Output Gate 가중치
  - $[num : 3*num - 1]$  : Forget Gate 가중치
  - $[num : 4*num - 1]$  : Update Gate 가중치
- $B[8*num]$ 에서 num에 8을 곱하는 이유

- {Input, Output, Forget, Update} → 4개
- {입력 데이터, 이전 데이터} → 2개
- $4 * 2 = 8$ 의 Bias가 필요함