

# PaddleNLP Github 官方baseline

测评地址: <https://aistudio.baidu.com/aistudio/competition/detail/130/0/introduction>

## 各baseline地址

- AiStudio - PaddlePaddle: <https://aistudio.baidu.com/aistudio/projectdetail/2461530>
- Github - PaddleNLP: [https://github.com/PaddlePaddle/PaddleNLP/tree/develop/examples/text\\_matching/question\\_matching](https://github.com/PaddlePaddle/PaddleNLP/tree/develop/examples/text_matching/question_matching)
- PyTorch: <https://zhuanlan.zhihu.com/p/427995338>

## 目录结构

- `./data/`: 数据集
- `./model/`: 训练后的模型
- `./work/`:
  - `data.py`: 数据处理
  - `model.py`: 模型配置
  - `train.py`: 训练模块
  - `predict.py`: 预测模块
  - 修改/: 基于baseline修改的代码模块
    - `model - transformer` 系列: 预训练输出后接入transformer或LSTM (注释掉的部分)
    - `transformer - xxx` 系列: 手写的transformer的self-attention模块
  - `./baseline.ipynb`: 基线主程序
  - 注: 本文件夹下的 `.txt` 文件为数据集, 按照名称可予以区分 (afqmc 为蚂蚁金融语句对数据集)

## 运行方法

### 1. 环境配置

- 在本目录下开启 `cmd`
- 执行 `jupyter notebook --no-browser --notebook-dir=qianyan-question-matching`
- 打开 `baseline.ipynb`
- 注: `paddlepaddle >= 2.2.2`

### 2. 安装必需模块

- `baseline.ipynb` 中运行第1、2行代码, 若有需要的模块再单独安装

### 3. 结果文件

- 结果文件保存在 `./work/result/` 下

