

Personal Note 06 - Deep Learning For NLP

1. How to build a deep learning model for NLP

- Xử lý data:
 - Việc xử lý data cần thực hiện tối ưu và tiết kiệm chi phí nhất có thể do nó không thể đẩy lên GPU để thực hiện.
 - Cần preprocess data trước tiên sau khi giải nén.
 - Thực hiện chia tokens.
- Text Embedding:
 - Chuyển đổi dữ liệu sang dạng số (như vector, hoặc ma trận tần suất)
- Chọn model
 - Một số model có thể cân nhắc:
 - Basic nhất vẫn là MLP, sau đó đến các model đặc thù khác như LSTM
- Training model
 - Lựa chọn bộ tham số phù hợp
- Predict & Evaluation

2. Vấn đề với project

- Dữ liệu lớn khiến cho quá trình thử nghiệm tốn nhiều thời gian và tài nguyên.
 - Việc xây dựng mô hình LSTM với data được tối ưu gặp khó khăn.
- Project chưa được hoàn thiện, vẫn còn gặp lỗi và chưa được đánh giá