

ĐẠI HỌC BÁCH KHOA HÀ NỘI
TRƯỜNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG



BÁO CÁO TUẦN 9

Môn học: Project II

*Chủ đề: Xây dựng mô hình dự đoán giá cổ phiếu bằng mạng
Transformer*

Giáo viên hướng dẫn:

Đỗ Tuấn Anh

Sinh viên thực hiện:

Lê Minh Triết

Mã số sinh viên:

20220045

Hà Nội - 2025

MỤC LỤC

Contents

MỤC LỤC.....	3
1. Công việc đã hoàn thành trong tuần	4
1.1. Triển khai Learnable Positional Encoding	4
1.2. Tối ưu tham số các mô hình sử dụng Learnable PE	4
2. Dự kiến các công việc tuần tới	4

1. Công việc đã hoàn thành trong tuần

1.1. Triển khai Learnable Positional Encoding

Trong các mô hình trước, em sử dụng Positional Encoding dạng tuần hoàn và **tĩnh**, nghĩa là nó không thay đổi trong quá trình huấn luyện.

Trong tuần này, em đã triển khai Learnable PE, nghĩa là encoding cũng được cập nhật trọng số trong quá trình huấn luyện.

Learnable PE giúp mô hình Transformer hội tụ nhanh hơn. Với các mô hình Transformer trước, thường mất khoảng 25-30 epoch để đạt loss trên tập valid dưới 0.1, trong khi với Learnable PE thì chỉ mất khoảng 15-20 epoch là bắt đầu hội tụ.

Trong quá trình triển khai, em lựa chọn 2 mô hình là Standard Transformer (theo chuẩn dùng Multi-head Attention) và Restricted Causal Transformer (dùng Restricted Causal Attention) và bước đầu đạt được kết quả khá tốt. Có thể xem ở version 12 của notebook:

<https://www.kaggle.com/code/trietp1253201581/stock-prediction-test-model>

Mô hình Standard Transformer thậm chí còn đạt được loss 3500 trên tập test, nghĩa là đã tiệm cận baseline Moving Average.

1.2. Tối ưu tham số các mô hình sử dụng Learnable PE

Em đã tối ưu trong version 13->16 của notebook

<https://www.kaggle.com/code/trietp1253201581/stock-prediction-test-model>

Do vấn đề giới hạn thời gian huấn luyện GPU miễn phí trên Kaggle, em mới chỉ thử tối ưu hai mô hình trên với khoảng 17-20 lần thử tham số mỗi mô hình.

Điều này làm cho mô hình Standard Transformer với tham số trả về từ optuna kém hơn khá nhiều so với tham số ban đầu. Vì vậy, từ giờ tới cuối kỳ, nếu còn thời gian em sẽ tối ưu thêm, nếu không thì sẽ lấy tham số cho kết quả tốt hơn cho 5 lần thử đầu tiên (quan sát mean và std).

Mô hình Restricted Causal + Learnable PE cho kết quả ổn định hơn, vì theo lý thuyết thì bản thân Attention này phù hợp với dữ liệu chuỗi thời gian

2. Dự kiến các công việc tuần tới

- Thử nghiệm Hybrid Model (LSTM + Transformer) cũng như tối ưu tham số cho mô hình này, chọn cách kết hợp tốt nhất giữa LSTM và Transformer. Đây là mô hình cuối cùng được thử nghiệm trước khi tập trung vào hiệu năng và tính giải thích cũng như tối ưu mô hình cuối cùng.