

PHÂN TÍCH DỮ LIỆU CHUỖI THỜI GIAN TRONG BÀI TOÁN ĐÁNH GIÁ VÀ DỰ ĐOÁN KHẢ NĂNG HỒI PHỤC CỦA BỆNH NHÂN ĐỘT QUỴ

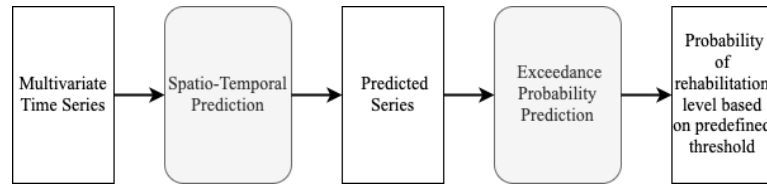
Trần Lê Phương Thảo
thao.tlp200604@sis.hust.edu.vn

Lê Văn Bằng
banglv1@viettel.com.vn

INTRODUCTION

Cải thiện khả năng đi lại là mục tiêu chung của những người sau đột quỵ và là trọng tâm chính của việc hồi phục chức năng. Với dữ liệu chuỗi thời gian của các đặc trưng là số bước đi và tốc độ đi bộ, mục tiêu của nghiên cứu là:

- (1) dự báo số bước và tốc độ đi bộ dựa vào chuỗi thời gian trong quá khứ;
- (2) ước lượng khả năng hồi phục của bệnh nhân trong khoảng thời gian được xác định trong tương lai.

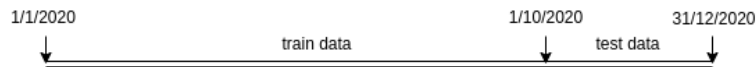


Hình 1. Mô tả về bài toán

DATA ANALYSIS

Dataset

Tập dữ liệu RSPDC bao gồm chuỗi thời gian về số bước đi và tốc độ đi bộ của 31 đối tượng bệnh nhân đột quỵ.



Gap Filling Techniques

Phương pháp nội suy tuyến tính (linear interpolation)

Kiểm định nghiệm đơn vị

Kiểm định Augmented Dickey Fuller (ADF) cho ra kết quả các chuỗi đều cần sai phân bậc 1 để có tính dừng.

Kiểm định độ trễ (lag)

Tối ưu giá trị của tiêu chuẩn AIC thu được độ trễ tối ưu lag=3

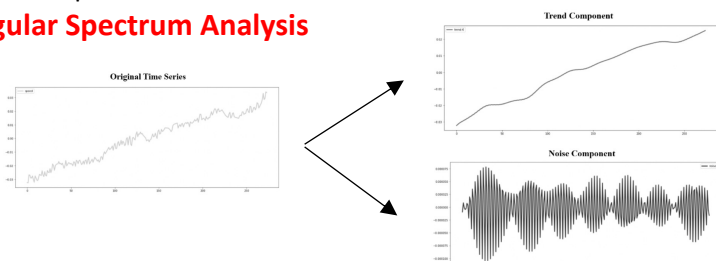
Kiểm định nhân quả Granger

Granger Causality Test cho kết quả cả 2 biến counts và speed có mối tương quan với nhau dựa vào p-value < 0.05

	counts-X	speed-X
counts-Y	1.000000	0.000000
speed-Y	0.000904	1.000000

Bảng 1. Granger Causation Matrix

Singular Spectrum Analysis



Hình 2. Phân tách chuỗi dữ liệu ban đầu thành 2 yếu tố thành phần là trend và noise

METHODOLOGY

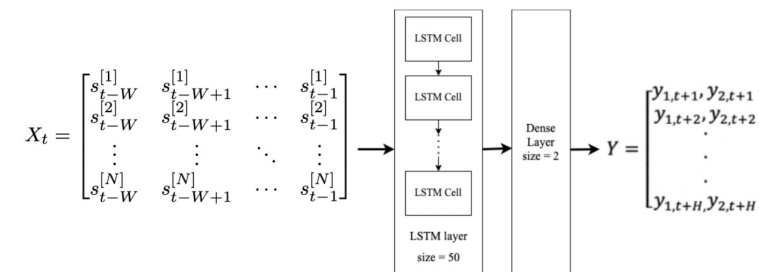
Vector Autoregressive Model (VAR)

$$\begin{bmatrix} 1 & b_{1,2} \\ b_{2,1} & 1 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} b_{1,0} \\ b_{2,0} \end{bmatrix} + \underbrace{\begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} \\ \varphi_{2,1} & \varphi_{2,2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix}}_{\text{Auto-Regressive (AR)}} + \underbrace{\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}}_{\text{Moving Average (MA)}}$$

LightGBM

- Supervised Learning Setting
- One-step Forecasting

Long short-term memory (LSTM)



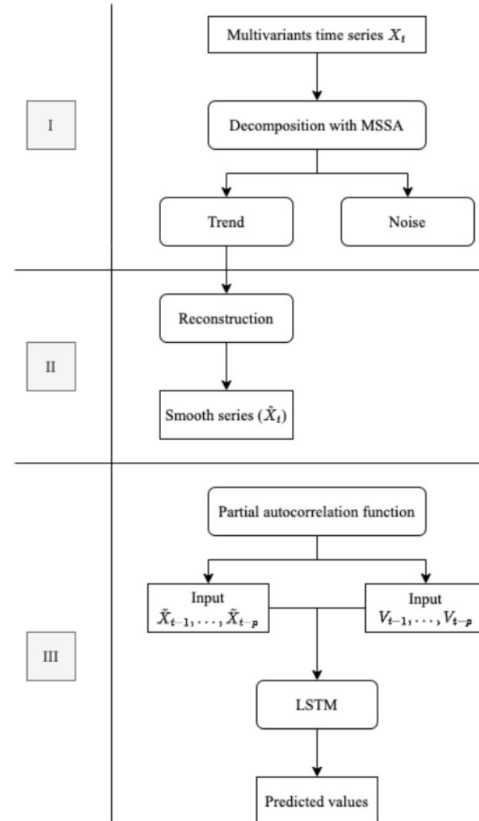
Hình 3. Sơ đồ của mô hình dự báo LSTM cho nhiều biến

Combine LSTM with SSA

Stage 1. Phân tách theo phương pháp đơn phổ (SSA) từ chuỗi dữ liệu thời gian ban đầu thành 2 yếu tố thành phần là xu hướng (trend) và phần nhiễu (noise).

Stage 2. Tái cấu tạo chuỗi thời gian sau khi loại bỏ phần nhiễu (noise).

Stage 3. Sử dụng mô hình dự đoán LSTM, dựa vào hàm tự tương quan một phần PACF (Partial autocorrelation function) để xác định độ trễ.



Hình 4. Sơ đồ của mô hình kết hợp SSA và LSTM

EXPERIMENTS

Performance Metrics

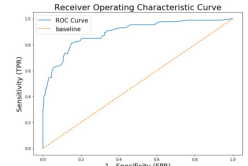
Spatio-Temporal Prediction.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

Exceedance Probability Prediction



$$TPR = \frac{TP}{TP + FN}$$

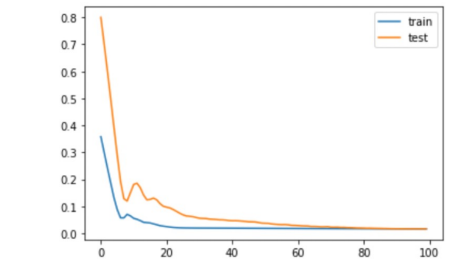
$$FPR = \frac{FP}{TN + FP}$$

$$BrierScore = \frac{1}{N} \sum_{t=1}^N (f_t - o_t)^2$$

Results

Baseline	RMSE		MAE		MAPE	
	counts	speed	counts	speed	counts	speed
VAR	3.800168	0.008569	2.902174	0.007102	0.006333	0.005751
LightGBM	4.376173	0.007773	3.277381	0.006616	0.007158	0.005363
LSTM	2.316863	0.004731	1.928027	0.004129	0.004218	0.003015
SSA-LSTM	1.442477	0.003198	1.172822	0.003017	0.002598	0.002541

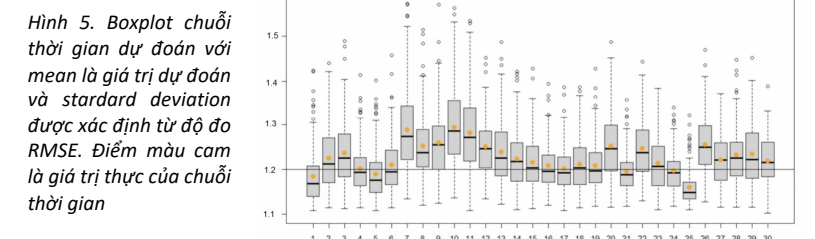
Bảng 2. Kết quả độ đo của mô hình dự báo chuỗi thời gian.



Hình 5. Giá trị hàm loss trong huấn luyện của model SSA-LSTM

Baseline	ROC	Brier Score
Binary Classification	0.918803	0.144711
Forecast Ensemble	0.254273	0.964696

Bảng 3. Kết quả độ đo của mô hình ước lượng xác suất hồi phục dựa trên ngưỡng xác định trước.



Hình 5. Boxplot chuỗi thời gian dự đoán với mean là giá trị dự đoán và standard deviation được xác định từ độ đo RMSE. Điểm màu cam là giá trị thực của chuỗi thời gian

FUTURE WORK

- Thêm biến chuỗi thời gian phù hợp với các đặc trưng có thể
- Xử lý chuỗi thời gian đếm bằng phương pháp Croston
- Tiền xử lý dữ liệu bằng biến đổi Fourier, Wavelet
- Phương pháp lấy mẫu Bootstrap cho bài toán ước lượng xác suất khả năng hồi phục