

THIẾT KẾ NEURAL NETWORK DẠNG AUTOENCODER PHỤC VỤ VIỆC NÉN VÀ GIẢM NHIỀU CHO TÍN HIỆU CẢM BIẾN IOT

DESIGNING AN AUTOENCODER NEURAL NETWORK FOR COMPRESSION AND NOISE REDUCTION IN IOT SENSOR SIGNALS

Trịnh Tuấn Dương

Bộ môn Cơ điện tử, Khoa Cơ khí, Trường Đại học Giao thông vận tải

Email: trinhtuandung@utc.edu.vn

TÓM TẮT

Trong bối cảnh Internet vạn vật (IoT) ngày càng phát triển, việc triển khai hệ thống cảm biến IoT ở những địa điểm có dung lượng băng thông thấp và gặp nhiễu từ môi trường cũng phổ biến hơn. Nghiên cứu này đề xuất sử dụng mạng Autoencoder để nén dữ liệu cảm biến IoT, đồng thời giảm nhiễu, nhằm tối ưu hóa truyền dữ liệu. Mạng Autoencoder đã được triển khai mô phỏng trên Google Colab với bộ dữ liệu Air Quality UCI, đạt tỷ lệ nén 3x, MSE trung bình 0.0025, PSNR 26.1 dB, và SSIM 0.9802. Kết quả cho thấy Autoencoder không chỉ giảm kích thước dữ liệu mà còn giảm nhiễu, cải thiện chất lượng tái tạo, mở ra tiềm năng ứng dụng trong các hệ thống IoT cần quan tâm đến dung lượng truyền dữ liệu và xử lý nhiễu.

Từ khóa: Autoencoder; IoT; Cảm biến.

ABSTRACT

In the context of the rapidly evolving Internet of Things (IoT), the deployment of IoT sensor systems in locations with low bandwidth capacity and environmental noise interference has become increasingly common. This study proposes the use of an Autoencoder network to compress IoT sensor data while simultaneously reducing noise, aiming to optimize data transmission. The Autoencoder network was implemented and simulated on Google Colab using the Air Quality UCI dataset, achieving a compression ratio of 3x, an average MSE of 0.0025, a PSNR of 26.1 dB, and an SSIM of 0.9802. The results demonstrate that the Autoencoder not only reduces data size but also mitigates noise, improving reconstruction quality. This opens up potential applications in IoT systems where data transmission capacity and noise handling are critical concerns.

Keywords: Autoencoder; IoT; Sensors.

1. ĐẶT VẤN ĐỀ

Internet vạn vật (IoT) đã trở thành nền tảng quan trọng trong nhiều lĩnh vực như nông nghiệp thông minh, giám sát môi trường, và quản lý năng lượng. Do đó, các thiết bị IoT có thể cần được triển khai trong môi trường mạng hạn chế, nơi băng thông thấp và thường gặp nhiễu từ cảm biến. Việc nén dữ liệu cảm biến trước khi truyền và giảm nhiễu trong quá trình xử lý là cần thiết để đảm bảo hiệu quả và độ tin cậy của hệ thống [1].

Mạng Autoencoder, một loại mạng nơ-ron học không giám sát, có khả năng nén dữ liệu thành biểu diễn nhỏ gọn (latent space) và tái tạo dữ liệu gốc với sự giảm thiểu về nhiễu nhờ cơ chế của mạng là học các đặc trưng quan trọng. Do đó, mạng autoencoder có thể được ứng dụng để giảm thiểu dung lượng truyền dữ liệu, giảm nhiễu mà vẫn giữ nguyên được độ toàn vẹn của dữ liệu. Nghiên cứu này nhằm đánh giá hiệu quả của Autoencoder trong việc nén và tái tạo dữ liệu cảm biến IoT, sử dụng mô phỏng trên bộ dữ liệu mở. Mục tiêu là mạng autoencoder được thiết kế có thể đạt tỷ lệ nén cao, nhưng vẫn duy trì chất lượng tái tạo, và loại bỏ nhiễu hiệu quả.

2. THIẾT KẾ MẠNG AUTOENCODER

2.1. Bộ dữ liệu được sử dụng để huấn luyện và kiểm tra

Bộ dữ liệu được sử dụng là Air Quality Dataset từ UCI Machine Learning Repository. Đây là một bộ dữ liệu mở, bao gồm 9358 mẫu, được thu thập từ một mạng lưới cảm biến không khí đặt tại một thành phố ở Ý từ tháng 3 năm 2004 đến tháng 2 năm 2005. Trong bộ dữ liệu này có 12 cột ghi lại dữ liệu từ các cảm biến bao gồm: nồng độ khí CO, NO₂, NOx, benzen,

hiệt độ, độ ẩm tương đối, và các chỉ số khác liên quan đến chất lượng không khí [3]. Dữ liệu từ bộ dữ liệu trước khi đưa vào huấn luyện cho mạng autoencoder sẽ được chuẩn hóa về khoảng [0, 1] nhằm phù hợp với yêu cầu đầu vào của Autoencoder như sau:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Với: x là giá trị gốc, x_{min} , x_{max} là giá trị nhỏ nhất và lớn nhất tương ứng của cột dữ liệu, còn x_{norm} là giá trị đã được chuẩn hoá.

2.2. Thiết kế cấu trúc mạng autoencoder

Mạng Autoencoder được khởi tạo thiết kế với kiến trúc fully-connected như sau [2]:

- Input Layer: 12 nút, tương ứng với 12 cột dữ liệu cảm biến.

- Encoder: Hai tầng ẩn với lần lượt 8 và 4 nút, sử dụng hàm kích hoạt ReLU: $f(x) = \max(0, x)$.

- Bottleneck: 4 nút, là kích thước của latent space.

- Decoder: Hai tầng ẩn với lần lượt 8 và 12 nút, kết thúc bằng hàm sigmoid để đảm bảo đầu ra trong khoảng [0, 1]: $\sigma(x) = \frac{1}{1 + e^{-x}}$

- Loss Function: Mean Squared Error (MSE) được dùng để tính sai số giữa đầu ra của mạng và dữ liệu gốc: $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ với y_i là giá trị thực tế của mẫu thứ i và $\hat{y}_i$ là giá trị dự đoán từ autoencoder.

- Optimizer: Adam, với learning rate 0.001.



Cấu trúc mạng trên được lựa chọn để cân bằng giữa việc nén dữ liệu giúp giảm băng thông khi truyền dữ liệu của sensor đồng thời không làm mất đi tính chi tiết của dữ liệu. Dựa vào kết quả của quá trình thử nghiệm, cấu trúc của mạng sẽ được thay đổi để thu được đáp ứng tốt hơn.

2.3. Huấn luyện mạng autoencoder

Để đảm bảo mạng autoencoder học một cách hiệu quả về thời gian, và tránh hiện tượng overfitting vì dữ liệu 9358 mẫu là tương đối nhỏ, mô hình được huấn luyện trên Google Colab trong 20 epochs, với batch size 32. Ngoài ra, để mô phỏng việc các cảm biến chịu ảnh hưởng bởi nhiễu, dữ liệu gốc được thêm nhiễu Gaussian như sau:

$$x_{\text{noise}} = \text{clamp}(x + N(0, \sigma^2), 0, 1)$$

Trong đó:

- x : Dữ liệu gốc.
- $N(0, \sigma^2)$: Nhiễu Gaussian.
- clamp : Hàm giúp giới hạn giá trị trong khoảng $[0, 1]$ để tránh vượt ngưỡng yêu cầu của mạng.

3. KẾT QUẢ VÀ PHÂN TÍCH

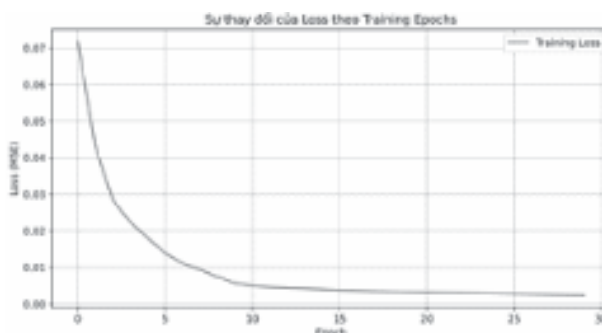
Sau quá trình thử nghiệm, cấu trúc mạng được thiết kế lại với việc tăng một tầng ẩn và số nút để có đáp ứng tốt hơn với một số cảm biến có MAE cao như NOx. Qua việc thử nghiệm với một số kiểu cấu trúc khác nhau, mạng autoencoder được thiết kế lại với cấu trúc là $(12 \rightarrow 20 \rightarrow 10 \rightarrow 4)$, so với cấu trúc mạng cũ là $(12 \rightarrow 8 \rightarrow 4)$.

3.1. Chất lượng tái tạo tổng thể

Chất lượng tái tạo được đánh giá qua các chỉ số MSE, PSNR và SSIM trên 100 mẫu dữ liệu cho thấy cả hai mạng autoencoder đều cho kết quả tốt, và cấu trúc mạng được thiết kế lại cho đáp ứng tốt hơn so với mạng khởi tạo. Với cấu trúc Autoencoder mới $(12 \rightarrow 20 \rightarrow 10 \rightarrow 4)$, mô hình đạt $\text{MSE} = 0.0025$ (giảm từ 0.0188 so với cấu trúc cũ $12 \rightarrow 8 \rightarrow 4$), $\text{PSNR} = 26.1 \text{ dB}$ (tăng từ 17.26 dB), và SSIM trung bình $= 0.9802$ (tăng từ 0.8542). Các chỉ số này cho thấy dữ liệu tái tạo rất gần với dữ liệu gốc, với độ méo thấp và cấu trúc được bảo toàn tốt (SSIM gần 1).

Bảng 1. So sánh chất lượng tái tạo dữ liệu giữa các cấu trúc mạng Autoencoder

Cấu trúc mạng	MSE	PSNR(dB)	SSIM
$12 \rightarrow 8 \rightarrow 4$	0.0188	17.26	0.8542
$12 \rightarrow 20 \rightarrow 10 \rightarrow 4$	0.0025	26.10	0.9802

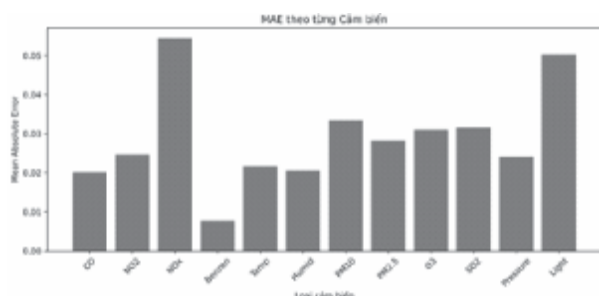


Hình 1. Sự thay đổi MSE trong quá trình huấn luyện mạng $(12 \rightarrow 20 \rightarrow 10 \rightarrow 4)$.

3.2. Sai số từng cảm biến

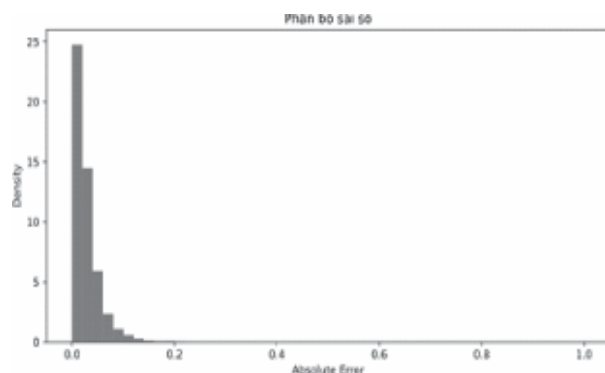
Hình 2 cho thấy MAE trung bình tính theo từng cảm biến thu được từ mạng $(12 \rightarrow 20 \rightarrow 10 \rightarrow 4)$. Cảm biến NOx ở mạng khởi tạo có MAE cao (0.17-0.2), nhưng với mạng autoencoder cấu trúc mới, MAE NOx giảm xuống dưới 0.06. Gần như tất cả các cảm biến

khác đều có MAE rất thấp (<0.04), cho thấy tái tạo đồng đều trên hầu hết cảm biến.



Hình 2. Kết quả MAE theo từng cảm biến thu được từ mạng ($12 \rightarrow 20 \rightarrow 10 \rightarrow 4$).

3.3. Phân bố sai số



Hình 3. Phân bố sai số tuyệt đối.

Sai số chủ yếu nằm trong khoảng 0-0.04, chứng minh chất lượng tái tạo cao.

4. KẾT LUẬN

Bài báo đã trình bày việc thiết kế mạng autoencoder với mục đích nén và giảm nhiễu dữ liệu trong quá trình truyền nhận các cảm biến IOT. Kết quả thu được cho thấy, mạng autoencoder nếu được thiết kế đúng cách hoàn toàn có thể vừa giảm được dung lượng dữ liệu truyền nhận, vừa có thể giảm nhiễu do môi trường tác động. Điều này sẽ giúp nâng cao khả năng ứng dụng, triển khai các dự án IOT tại các môi trường gặp khó khăn trong việc truyền nhận dữ liệu và ảnh hưởng nhiễu bởi nhiễu. ❖

Ngày nhận bài: 12/3/2025

Ngày phản biện: 25/3/2025

Tài liệu tham khảo:

- [1]. G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks". Science, vol. 313, no. 5786, pp. 504-507, 2006.
- [2]. I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning". Cambridge, MA, USA: MIT Press, 2016.
- [3]. S. De Vito et al., "On-field Calibration of an Electronic Nose for Benzene Estimation in an Urban Pollution Monitoring Scenario". Sensors and Actuators B: Chemical, vol. 129, no. 2, pp. 750-757, 2008.