Executive Summary: Physics-Informed and Graph Neural Networks for Modeling Spatiotemporal COVID-19 Dynamics

Navid Bahadoran

Overview

This project develops a hybrid machine learning framework to model the spatiotemporal evolution of COVID-19 case counts across U.S. counties using two complementary approaches:

- **Physics-Informed Neural Networks (PINNs)** for solving partial differential equations (PDEs) governing disease dynamics in a continuous domain.
- **Graph Neural Networks (GNNs)** for learning discrete spatiotemporal patterns on a static county-level graph using structured COVID-19 and demographic data.

The objective is to accurately predict infection spread patterns by incorporating physical principles and spatial dependencies, while benchmarking performance under real-world noisy and sparse data.

Data and Preprocessing

The dataset combines:

- Time-series COVID-19 case counts per U.S. county.
- County-level geospatial data (latitude, longitude, shapefiles).
- Population data to normalize case counts.

We preprocess the data by aggregating values, normalizing by population (to obtain infection rates), and scaling features with `MinMaxScaler`. For PINNs, we define a continuous spatial-temporal domain using collocation and boundary condition points. For GNNs, we build a spatial graph where nodes represent counties and edges are defined based on geographic adjacency.

Modeling Approaches

1. PINN

We solve a reaction-diffusion PDE of the form:

$$\frac{\partial u}{\partial t} = D\Delta u + f(u, x, t)$$

using a fully connected neural network trained to minimize a physics-based loss combining PDE residuals, initial conditions, and boundary constraints. Adversarial training is optionally employed to improve robustness under noisy observations.

2. GNN

We train a 3-layer Graph Convolutional Network (GCN) to predict normalized infection rates per county. The model uses node features such as historical case values and population, and edge connectivity based on spatial proximity. We scale both input and target variables and apply `Softplus` activation to ensure non-negative predictions. Early stopping and learning rate scheduling improve convergence stability.

Results and Evaluation

Evaluation metrics include:

- **Root Mean Square Error (RMSE)** and **Mean Absolute Error (MAE)** on original and scaled values.
- **Heatmaps and scatter plots** comparing predicted vs actual values spatially and temporally.

Key results:

- GNN achieved RMSE $\approx$ 6,100 and MAE $\approx$ 3,100 on original case scale.
- PINN reconstructed spatial diffusion patterns effectively using PDE structure, though sensitive to data sparsity.
- Both models were interpretable: PINN through learned PDE coefficients; GNN through learned node embeddings.

Conclusion

This project demonstrates the viability of integrating domain knowledge via PINNs and structural information via GNNs to model complex epidemic dynamics. While PINNs capture physics-based mechanisms, GNNs adapt flexibly to real data structures. Future work may combine the two into a unified spatiotemporal framework with uncertainty quantification and transfer learning.