Utilizing Meta-Learning to Enhance Performance on Anomalous Spatio-Temporal Data

Wassay Qureshi
University of Connecticut
Storrs, Connecticut
abdul.qureshi@uconn.edu

Abstract

This study seeks to introduce a meta-learning framework to enhance spatio-temporal predictive models by enabling them to adapt to and predict anomalous traffic data, such as those observed during the COVID-19 pandemic. Traditional traffic prediction models are not built to specifically handle the obstacle of irregular or chaotic traffic data that has not been seen before. This necessitates more robust approaches that can handle these edge cases. Our proposed framework uses meta-learning to integrate a multi-input CNN model that can learn from historical traffic patterns to understand the underlying structures and utilize them to explain and predict anomalous data. Preliminary results highlight our framework's ability to handle this over traditional machine learning methods of training a model solely on said anomalous data. This work continues to extend past work dealing with spatio-temporal prediction models such as the STResNet model implemented in [1].

I. Introduction & Background

The motivation for our work stems from preliminary work completed over the last year in updating the STResNet model discussed in [5]. STResNet aimed to use biking data from New York City and taxi driver data from Beijing to teach a convolutional neural network how to understand and predict future traffic behavior. Due to the paper's age of almost 8 years, many of the packages and dependencies used to build STResNet had been deprecated entirely. This experience was a first step into seeing how traffic prediction models worked and how methodologies present in their papers were written and expressed in code.

Another more recent motivation was the CyberInfrastructure Hackathon held at UConn, where the STResNet model and data were used as inspiration for students to make traffic prediction models using New York City biking data. This event allowed for a deep dive into how raw open-source New York biking data could be processed in a manner that allowed for a machine-learning model to learn from.

These previous works allowed for the depth of understanding of creating spatio-temporal datasets and models that could learn from and generate predictions on this data. As the popularity of traffic prediction research grows, research into how to handle anomalies in traffic data also grows.

Works such as [1], [2], [4] have proposed solutions that get around this issue with meta-learning. In meta-learning a model is tasked with learning the underlying representation of the data to then make predictions on another data set that may have too few data samples to train a model directly. These works use meta-learning to understand the traffic patterns of a source city that lacks historical data, given a target city with a large historical traffic patterns dataset. Here, models are trained on the source cities' traffic patterns. These trained models are then used to gain an understanding of the traffic patterns for the target city, usually by introducing new data such as what the most popular areas are in the target city.

Without diving into the specifics of each paper, it should be noted that there is not as much work focused on using meta-learning specifically to increase the robustness of spatio-temporal models in the face of anomalous data. One of the most recent examples of anomalous data affecting traffic prediction models came with the many lockdowns and travel bans enacted during COVID-19. Here, much of the historical, pre-pandemic data was not able to be used in a manner where traffic prediction algorithms were as accurate as usual. This problem is one that I find incredibly appealing due to both my previous experience with STResNet and working with historical biking data from New York. Building models specifically made for situations where other models will struggle is this work's main purpose.

II. Related Work

This paper and the methodology being used draw inspiration from many previous works dealing with meta and transfer learning techniques applied to spatio-temporal data.

[1] Wan et al. utilize LSTM and transfer learning to offer a foundation for understanding spatio-temporal data within traffic systems. While their model is stable and performs well with stable traffic patterns, it's less equipped for anomalous and sudden disruptions in traffic behavior, like the pandemic. Our work draws inspiration from their work by developing a framework designed to handle those disruptions based on the meta-learning frameworks proposed in their paper.

[2] focuses more in-depth on the concept of meta-learning between cities. Our notion of using meta-learning to build more robust and adaptive models stems from their work showing how a predictive model could predict traffic patterns across different cities. This work's model learns the underlying

traffic patterns within a target city and using those patterns can predict the traffic patterns there, despite the different architecture and locations of essential services. [4] builds upon this idea by focusing on using the long-term traffic patterns from one city with a populous amount of spatio-temporal traffic data and using it to predict the traffic patterns in a city with little amounts of traffic data. These works were eye-opening as they allowed for an understanding of how much was currently possible with current meta-learning techniques. Our work deviated from theirs by focusing on a smaller scale. We are more focused on making the predictions in one city as robust as possible, even when previous historical traffic data fails to represent those patterns accurately.

[3] is one paper that is tied closely to ours. Yang and He focus on understanding traffic patterns within New York City's open-source biking dataset. Leveraging this information, Yang and He try to find the optimal location for an additional bike station to be placed. This paper's framework aligns very well with our work. We extend their work by using an additional model to understand traffic patterns over the previous months' biking data and determine their relevance to the current disruptive month, leading to a more robust model overall.

Lastly, Zhang et al.'s work [5] forms the backbone of our proposal. This paper goes in-depth about how to utilize a deep neural network to predict traffic patterns across a city like New York City or Beijing. This work was the focus of my research last semester. A lot of time and effort was spent updating their model to use modern machine learning focus packages like tensorflow. This paper was our first introduction to understanding traffic prediction by calculating the inflow and outflow of traffic movement of regions in New York City. We build upon their implementation and ideology by focusing on when traffic patterns do not match those seen in the historical data, rather than this work's focus on predicting overall movement. We borrow almost the same format of expressing New York City as a grid to predict inflow and outflow patterns as well, which is a vital aspect of our work that shall be mentioned in depth later.

Without the contributions of these previous works and the many that came before, it would not be possible for us to come up with our current work and progress. With this in mind, we would like to delve deeper into our work and our proof of concept.

III. Dataset Preparation

The dataset we would like to use for this work comes from New York City's open-source Citi Bike Dataset consisting of a vast amount of historical biking data from users driving their bikes from station to station. This data is a wonderful resource for understanding traffic patterns across New York City and is used in previous works [3] and [5].

This data comprises many different biking trips as stated before. The exact information stored from these trips is as follows: the starting time, the ending time, the name, ID, and coordinates of the station the trip begins at, the name, ID, and coordinates of the station the trip ends at, the bike ID, the type of user (whether they have a Citi bike subscription or not), the user's birth year, and the user's gender.

This wealth of data is very useful, but can't be directly used to train a model to form traffic predictions in its current format. Preprocessing is needed to get the data into a format that a machine learning algorithm can understand. Our current preprocessing techniques come directly from [5] and their use of inflow and outflow matrices.

This technique comprises of formatting New York City, specifically Manhattan in the form of a 16x8 grid. This grid can be thought of as an abstract map of Manhattan. From here the Citi Bike dataset can be used to calculate the number of biking trips going into and out of each station in each region of the 16x8 grid. Here, we store the information as 2 different versions of these 16x8 grids to preserve the inflow and outflow information. This leaves our final post-processed dataset consisting of a 2x16x8 tensor where the inflow matrix is first, followed by the outflow matrix.

A CNN model can now directly understand the post-processed tensor. This allows the model to learn from and predict the spatio-temporal patterns occurring in the underlying data. With an understanding of the data, let's discuss the technical components of our proposed framework.

IV. Proposed Directions of Technical Components

Figure 1 details the basis for our proposed framework. The goal of this framework is to build a model that can form accurate predictions given traffic data that does not occur in any of the previous historical data. For ease of readability, we will denote this anomalous month of data as A. This framework aims to understand the underlying structure of traffic patterns from these previous months, drawing relevance from past data to try and explain and generate accurate predictions on A.

We will assume that given our anomalous data A, while the traffic patterns differ, they will not be completely different from the patterns seen in the previous historical data. To explain this reasoning let's look back to the pandemic and how the ways that people moved changed, yet many aspects of this traffic also remained consistent. For example, while many people shifted to working from home or engaged in online schooling, essential activities like grocery shopping continued.

To generate predictions on A, we utilize four CNN models trained on their respective months of data. These models are not designed as deep neural networks but rather as simpler baseline models that aim to capture the trends and behaviors of traffic in their respective months. We leverage this data by adding the weights of their output layer and using this alongside A as inputs for the multi-input CNN model. This multi-input model can leverage the learned representations from the previous month's baseline models to form a more accurate prediction on A.

Rather than just adding the weights of the baseline models' output layers, we propose a technique utilized in [5], where we assign each of these months a weighting based on how well

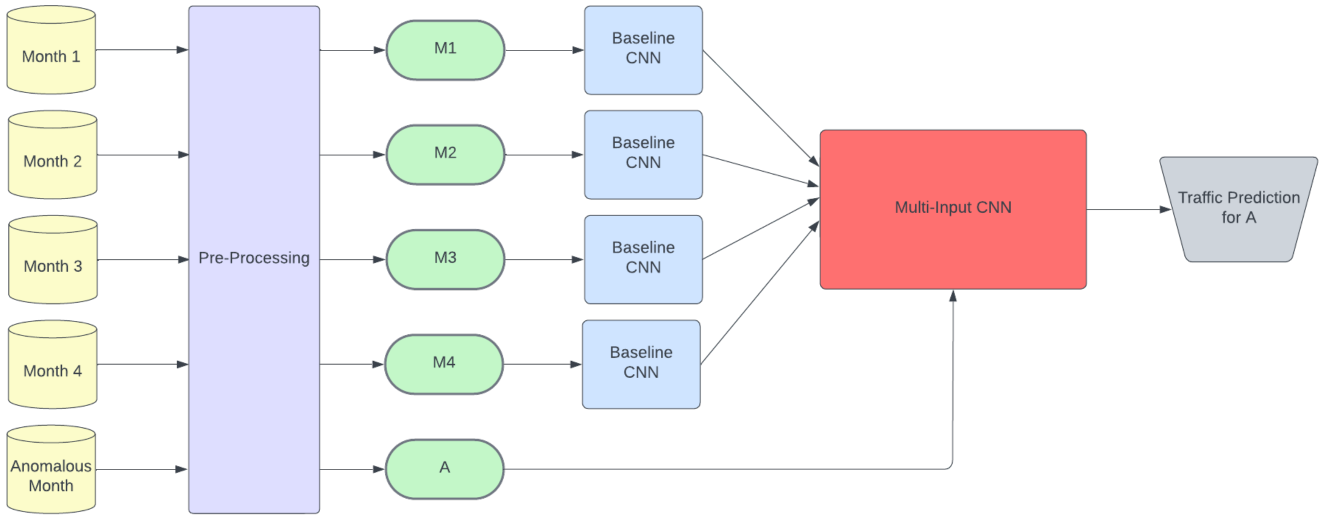

Fig. 1. Proof of Concept Framework for Predicting Anomalous Data

they align with the patterns in the anomalous data. Months that better match these trends are assigned higher importance, and vice versa. The multi-input CNN model will use this weighted sum alongside A to form a more robust understanding of A.

V. Experiment

To demonstrate our results, we highlight an experiment using the above framework. In this experiment, we try to use the process highlighted in Figure 1 to predict traffic patterns during the first month of the pandemic, April 2020 in Manhattan. To train respective baseline CNN models, we use data from November 2019 to February 2020. March 2020 will not be used for training as many of the lockdown policies for COVID-19 were put in place late in March, and this could lead to the baseline models potentially learning some of these post-pandemic traffic patterns, which would inflate the results covered in the next section.

The baseline CNN models used are a 3 convolutional layers CNN model consisting of 128 neurons. Each layer consists of 32 then 64 then 128 filters. As mentioned, the baseline models are designed not to be deep neural networks and to find a good latent representation for each month.

The multi-input CNN model takes in the weights from each of the previous branch models, Despite its multiple inputs, this CNN model is not classified as a deep neural network, as it comprises only 4 layers, excluding the layers from the baseline branch models

To ensure our framework is indeed an improvement over traditional methods, we will also do another experiment where a baseline model is trained on only the April 2020 data. In both experiments, the models will be trained on the first half of April 2020 data and forming predictions on the second half of April 2020.

VI. Results & Discussion

This section details the results of both of the above methods. Comparison of how well the multi-input CNN and the baseline-only methods do to predict the traffic patterns for April 2020.

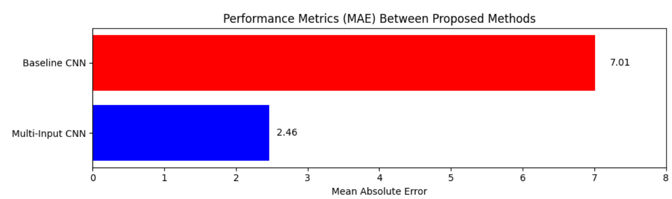

Fig. 2. Mean Absolute Error Metrics Between 2 Methods

The results of the experimental evaluation highlight the performance of our multi-input CNN approach as well as a comparison of its performance to the traditional baseline CNN direct testing/training method when predicting the anomalous conditions seen in April 2020 from the COVID-19 pandemic. Examining the results as seen in Figure 2 the multi-input CNN model had a Mean Absolute Error (MAE) of 2.46 as opposed to the baseline model's MAE of 7.01. This serves as the primary indicator of each method's prediction accuracy.

This discrepancy in MAE values between the multi-input model and the baseline experiment showcases our framework's ability to capture broader trends in the biking data and to integrate it into its predictions successfully. While the baseline model was designed to be a simple model, its performance compared to our proposed method is very promising.

Figure 3 takes a further examination into how the multi-input model came to its predictions in this experiment. The figure visualizes the importance assigned to each month within the multi-input model indicating how much each month's data influenced the multi-input model's predictions for April.

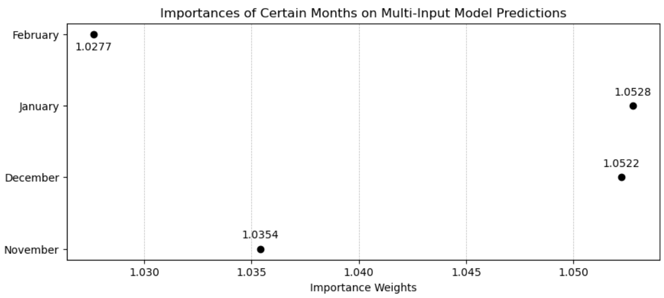

Fig. 3. Importances found for each branch model's weights

Months with a higher weight suggest a higher similarity observed between the month and our anomalous data from April. As seen, February and January may have been given a higher importance as they are closer to April than November and December. These results suggest that the multi-input model gained an understanding of the temporal relevance of the historical data and how not all months contribute equally to the trends seen in April.

These figures showcase the efficacy of our framework in handling and accurately predicting anomalous data it was never trained on. Future work can explore the use of more sophisticated models in place of the baseline CNN models used in both our proposed framework and also in the performance comparison. This could lead to a potentially improved ability of the multi-input model to gain a better understanding of the trends occurring in its branching CNN baseline models and assign them a more accurate importance weighting. This may lead to an improved ability of the multi-input model to explain anomalous data.

Acknowledgments

Finally, I would like to give thanks to all the authors and contributors to the papers used in our work. Each work played a vital role in the development of this paper. Without their work, we never would have been able to create ours.

Another thanks to Professors Bing Wang and Suining He who led the CyberInfrastructure Hackathon and allowed me to help as an administrator. Learning how to pre-process this data efficiently was instrumental to the development of this work.

Again, special thanks to Professor Suining He for his support and guidance throughout the last year for both my independent research and thesis. Before working with Professor He on this work, I had no prior experience building machine learning models from scratch. I genuinely appreciate all the patience, kindness, and guidance you have given me throughout this last year.

References

[1] Wan, X., Liu, H., Xu, H., & Zhang, X. (2022). Network traffic prediction based on LSTM and Transfer Learning. IEEE Access, 10, 86181–86190. https://doi.org/10.1109/access.2022.3199372

[2] Wei, Y., Zheng, Y., & Yang, Q. (2016). Transfer knowledge between cities. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. https://doi.org/10.1145/2939672.2939830

[3] Yang, X., & He, S. (2020). Towards Dynamic Urban Bike Usage Prediction for Station Network Reconfiguration. https://doi.org/10.48550/arXiv.2008.07318

[4] Yao, H., Liu, Y., Wei, Y., Tang, X., & Li, Z. (2019). Learning from multiple cities: A meta-learning approach for spatial-temporal prediction. The World Wide Web Conference. https://doi.org/10.1145/3308558.3313577

[5] Zhang, J., Zheng, Y., & Qi, D. (2017). Deep spatio-temporal residual networks for citywide crowd flows prediction. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1). https://doi.org/10.1609/aaai.v31i1.10735