DSC180A: Spatiotemporal Q2 Project Proposal

Kailing Ding k5ding@ucsd.edu
Judy Jin z3jin@ucsd.edu
Miles Labrador mlabrado@ucsd.edu
Derek Leung djleung@ucsd.edu

1 Introduction

During Q1, our group's major focus was to enable uncertainty quantification feature in torchTS, a PyTorch-based time-series forecasting library. This open-sourced library aims to provide an efficient and user-friendly time-series deep learning library that would benefit the entire research and industry community.

2 Problem Statement

However, during the development of this library and the replication of the UQ research paper, we found that there are 2 major problems to be potentially solved which provide a lot of value:

High Learning Curve When we started developing torchTS, it took us 2 or 3 weeks to get familiarized with the library. Although on the official website, there are mathematical explanations on most models that the library provides, there weren't many code examples for users to look at, and thus they might have to look into source code in order to use the library, which increases user friction.

Limited Number of Spatiotemporal Features We understand that torchTS is currently under active development and young in development, but we currently only have graph models (DCRNN) and Sesq2Seq implemented as well as an uncertainty quantification feature. If we can add more spatiotemporal-related features into the library, torchTS would likely become more valuable and stand out from other similar libraries.

3 Proposed Solution

3.1 Idea Summary

The proposed solution aims to reduce the learning curve of torchTS and increase the number of spatiotemperal-related features within the library. Thus, the solution consists of 3 parts: TorchTS research, user-facing contributions, and end-to-end data analysis using TorchTS. We believe that these 3 parts will make torchTS more valuable and user-friendly.

3.1.1 TorchTS Research

Given that the goal of this project aims to make accessible and useful predictions, ensuring the usefulness of these predictions requires selecting optimal methods and models. To this end, our team intends to explore a conformal quantile regression model and several other models as an extension of previous exploration into the accuracy of quantile regression. The conformal quantile regression model seeks to combine the strengths of conformal prediction, which produces a confidence interval using finite samples, without making distributional assumptions, and quantile regression, which is adaptive to heteroscedasticity. While before testing, we are not certain this will be a final model, this research phase into conformal quantile regression and other potential models will provide several models and results for comparison and evaluation.

Preprint. Under review.

Our implementation of the models we determine to be the most accurate through testing will be evaluated on mean absolute error, mean interval score, and interval size relative to our pre-existing models.

3.1.2 User-facing Contributions

TorchTS is one library part of a movement towards more open and efficient tools that help enable state of the art data analysis in a convenient manner. One of the best and most important methods to ensure that our library is usable for the common user is to keep update documentation on how our library works and the best practices involved in interacting with it.

We propose a quarter-long commitment to extending the current documentation with more in-depth and explanatory pages documenting torchTS. This can involve, but is not limited to, setting up pages for individual models that can help visualize the input and output parameters of the code, walk-through guides on producing viable implementations of the code, and potentially general housekeeping notes as the library evolves.

With the current state of libraries such as torchTS being naturally open and collaborative, we acknowledge that the documentation will greatly benefit users of the library while also providing an easier experience for anyone that is looking to push code developments to the library in the future.

3.1.3 End-to-end Data Analysis Using TorchTS

Besides user friendly documentations, we also plan to work on creating real life examples on spatiotemporal analysis to first give an easy to follow case study for the users, and second to give insights to the developers about what to improve on the model. Ideally, we propose to perform data analysis on a real life open resource dataset, for example the METR-LA traffic dataset. Data analysis will consist of a well documented data preparation pipeline, a training run file, final forecasting results, and some visualization on the forecasting results. We plan to also do qualitative and quantitative analysis on the prediction and the final results to get a comprehensive view on our model's ability and limitations.

This data analysis should be a quarter long task which aims at attaining results from four different models in torchTS with different datasets. During the analysis, we expect to encounter bugs/inconsistency or inconvenience from the base code, which could further motivate developers to do certain modifications for the library.

3.2 Expected Outcomes

The outcome of this project will consist of few notable parts:

- There will be documentation on the official website for each method and model we test in order to understand the mathematical reasoning behind the method or model. This documentation should be a condensed, higher level explanation with any essential mathematical formulas explained
- We will be implementing the methods and models we use into the torchTS library for anyone who would like to use it.
- There will be an example of each model implemented in torchTS in order to demonstrate how to use each method and see what kind of outputs can be expected.