Martin Becerra (carlosb3@illinois.edu)
Brian Reinbold (brianjr3@illinois.edu)

STAN: Spatio-Temporal Attention Network for Pandemic Prediction Using Real-World Evidence

Citation to the original paper

Junyi Gao, Rakshith Sharma, Cheng Qian, Lucas M Glass, Jeffrey Spaeder, Justin Romberg, Jimeng Sun, Cao Xiao, STAN: spatio-temporal attention network for pandemic prediction using real-world evidence, *Journal of the American Medical Informatics Association*, Volume 28, Issue 4, April 2021, Pages 733–743, https://doi.org/10.1093/jamia/ocaa322**.**

What is the general problem this work is trying to do?

The paper proposes a model to make more accurate pandemic predictions.

What is the new specific approach being taken in this work, and what is interesting or innovative about it, in your opinion?

There are three primary innovations:

1. Utilizing demographic similarity and geographical proximity between locations.
2. Integrating pandemic transmission dynamics into a deep learning model.
3. Using patients' claim data that capture local disease status and medical resource utilization.

Many epidemiological models are fit to one location, failing to capture proximity and interaction with other locations. For example, cases are likely to rise in one city if they are rising in a neighboring city. What is interesting with this paper is that they combine a graph attention network (GAT) and a recurrent neural network (RNN) to simultaneously capture the spatial and temporal effects.

By incorporating the SIR equations into the model, epidemiological theory can act as a regularizer on STAN's predictions. Furthermore, the implications of these equations can be extended into the loss function to control for accurate prediction in the short term and long term. This is an intriguing way to incorporate "knowledge" into a model.

Finally, COVID case data is known to have quality issues. Data could be inaccurate, may be reported with a substantial lag, or reporting standards could change. By incorporating claims data related to COVID-19 symptoms, the authors can overcome these data quality issues.

What are the specific hypotheses from the paper that you plan to verify in your reproduction study?

The main hypothesis we want to test is that STAN can improve active case predictions for U.S. states for multiple prediction time horizons (5, 15, and 20 days) over the baselines. The paper compares STAN against the following baselines: SIR, SEIR, GRU, ColaGNN, and CovidGNN. Following the paper, we would evaluate using mean squared error, mean average error, and the average concordance correlation coefficient.

What are the additional ablations you plan to do, and why are they interesting?

We want to incorporate the Oxford COVID-19 Government Response Tracker (OxCGRT)[1] coding system as a data source to control the effect of public policy and additional demographic factors. It provides a systematic and objective account of the strength of Covid-19 response policies that have been instigated by U.S. state governments. It includes an overall government response index, a containment and health index, a stringency index, and an economic support index at daily frequency.

[1] https://github.com/OxCGRT/USA-covid-policy

These policies can encompass mask mandates, restrictions on movement, school/workplace closings, ect. Since there is ample evidence that these policies slow down infections, incorporating these features should improve the predictions of STAN.

We also want to see if these results hold when only using easily accessible, public data. By comparing the model performance with and without the proprietary EHR data, we can gauge how much EHR data improves predictions.

State how you are assured that you have access to the appropriate data.
Much of the data is publically available COVID-19 data like number of active cases, total cases, current number of hospitalizations; and easily accessible demographics factors like longitude, latitude, population size, and population density. Unfortunately, the paper utilizes proprietary data related to COVID-19 claims that we may or may not receive. Although we would not be able to replicate their exact results without this data, we would still be able to see if the proposed model improves predictions over the baseline only using publically available data.

Discuss the computational feasibility of your proposed work – make an argument that the reproduction will be feasible.
The reproduction will be feasible since the authors provide code to replicate STAN. We would need to implement the baselines, but this should be feasible. The SIR and SEIR models are just a system of differential equations, so it should be straightforward to implement. We should not have a problem implementing a standard gated recurrent neural network. However, implementing both ColaGNN and CovidGNN from scratch will be much more challenging. Also, the data is not large, so it should be quick to train.

State whether you will re-use existing code (and provide a link to that code base) or whether you will implement yourself.
We will use existing code to implement STAN. We will need to modify their code to incorporate additional features for our alblations, calculate performance metrics, and more systematically run and capture the results for each state. The model has to be trained and evaluated on each state individually. We will have to code the baselines ourselves.

SurvTRACE: Transformers for Survival Analysis with Competing Events

Citation to the original paper
Wang, Zifeng, and Jimeng Sun. "SurvTRACE: Transformers for Survival Analysis with Competing Events." *arXiv preprint arXiv:2110.00855* (2021).

What is the general problem this work is trying to do?
This work is trying to improve model predictions in survival analysis with competing events.

What is the new specific approach being taken in this work, and what is interesting or innovative about it, in your opinion?
The approach taken by the authors is to do survival analysis using transformers with competing events. The authors:

1. Use an inverse propensity score to deal with selection biases. Selection bias can be found in other classes of problems, so it will be useful to learn ways to deal with it.
2. Automatic feature engineering by learning interactions between covariates. By introducing attention, the model is more interpretable, which makes SurvTRACE more intriguing than other models.
3. Improve accuracy by doing a multi-task shared representation learning.

Transformers are one of the latest advances in deep learning, so it will be exciting to learn more about them. Also, we have not applied deep learning to survival analysis in previous homeworks and labs, so it will be an interesting application.

What are the specific hypotheses from the paper that you plan to verify in your reproduction study?
We want to test the authors' claim that their method is all-around superior to other baselines at different quantiles of event times, for single events on both the METABRIC and SUPPORT dataset, and for competing events on the SEER dataset. They use the time-dependent concordance index as the evaluation metric. For single events, they compare SurvTrace against these models: Cox Proportional Hazards, Random Survival Forests, DeepSurv, DeepHit, Piece-wise Constant Hazard, and Deep Survival Machines. For competing events, they compare SurvTrace against these modes: DeepHit, Deep Survival Machines, cause-specific Piece-wise Constant Hazard, and cause-specific Cox Proportional Hazards.

What are the additional ablations you plan to do, and why are they interesting?
This model uses the Scaled Exponential Linear Unit (SELU) activation function. Since this function was not introduced during the course, we want to try the more common ReLU. This will be interesting in the sense of measuring how important and impactful different activation functions can have on model performance and training time.

State how you are assured that you have access to the appropriate data.
We are confident that we have access to the appropriate data. Two of the datasets, SUPPORT and METABRIC, are publicly available and can be accessed through the library pycox[2]. We have requested and received access to the SEER database as well.

Discuss the computational feasibility of your proposed work – make an argument that the reproduction will be feasible.
Given that the author doesn't put much emphasis on the computational aspect of the experiments, and that the size of the relevant SEER data is only about 200MB, it gives us the impression that it should be doable with the resources available in Google Colab.

State whether you will re-use existing code (and provide a link to that code base) or whether you will implement yourself.
We will re-use existing code. The authors' have their model on GitHub. We will implement the baseline models ourselves.

Representation Learning for Robust Patient-Independent Epileptic Seizure Detection

Citation to the original paper
Zhang, Xiang et al. "Adversarial Representation Learning for Robust Patient-Independent Epileptic Seizure Detection." *IEEE journal of biomedical and health informatics* vol. 24,10 (2020): 2852-2859. doi:10.1109/JBHI.2020.2971610.

What is the general problem this work is trying to do?
This paper builds a model to detect seizures from patients' electroencephalography (EEG) signals.

[2] https://github.com/havakv/pycox

What is the new specific approach being taken in this work, and what is interesting or innovative about it, in your opinion?
Most previous research focuses on patient-dependent scenarios where a seizure is detected using a patient's own history. These methods achieve high accuracy since the training and testing sample are gathered from the same source and have similar distributions. Patient-independent methods, on the other hand, detect seizures by learning from other patients, which can be useful in alerting patients of potential seizure risk. However, patient-independent methods are easily corrupted by inter-patient noise. This paper proposes a model to better detect patient-independent seizures.

The interesting innovation in this paper is that it extracts a seizure embedding and a patient embedding to learn a robust set of general features useful for detecting seizures while simultaneously controlling the influence of inter-patient factors. The two embeddings run through parallel pipelines, and the model predicts whether a sample is having a seizure and the patient id. The loss function also incorporates the reconstruction error of the embedding and the L2 norm.

The other interesting aspect of the model is that the seizure detections incorporate an attention mechanism on the different EEG signal channels. These learned attention weights from a trained model give the importance of each channel in contributing to a seizure, which makes the model more interpretable.

What are the specific hypotheses from the paper that you plan to verify in your reproduction study?
We specifically want to test that the authors' proposed model outperforms the baselines in terms of accuracy for the majority of patients and averaging over accuracy. The paper compares their model against the following baselines: Support Vector Machine, Random Forest, K-Nearest Neighbors, Schirrmeister et al., Ansari et al, Lin et al, and Kiral et al.

What are the additional ablations you plan to do, and why are they interesting?
We want to try adopting ensembling strategies to improve the model's accuracy. For example, we could use voting by taking multiple EEG segments and then predicting the seizure state independently. The final decision is the majority of votes over all the prediction results. This is interesting because improving the diagnosis accuracy of the model (currently ~80%) could make it clinically deployable.

State how you are assured that you have access to the appropriate data.
We are confident that we have access to the appropriate data. We have requested and received access to the data from the Temple University Hospital EEG (TUH EEG) database.

Discuss the computational feasibility of your proposed work – make an argument that the reproduction will be feasible.
Based on their criteria for filtering and selecting, they have fourteen patients in their sample, which is around one gigabyte of data. They used a NVIDIA TITAN X GPU with 10 Gbps memory speed and 3584 cores, and it took 1417.4 seconds (about 24 minutes) to train. Since the training time is reasonable and the fact that the GPU specifications are similar to the Nvidia K80 GPUs one can access through Google Colab[3], the computations of this work should be feasible.

State whether you will re-use existing code (and provide a link to that code base) or whether you will implement yourself.
The code for the authors' model is on GitHub, and it is written in tensorflow. We will use their code, but if time permits we will rewrite their code using pytorch. However, we will have to write a script to process the data. We will also write the code for the baseline models ourselves.

[3] https://kazemnejad.com/blog/how_to_do_deep_learning_research_with_absolutely_no_gpus_part_2/