Week 11

Wentao Gao

Research questions

Propose a causality-driven predictive modeling approach to enhance the robustness and ensure the transferability of the model across different regions.

Idea:

Data → Latent variables

Original feature variables

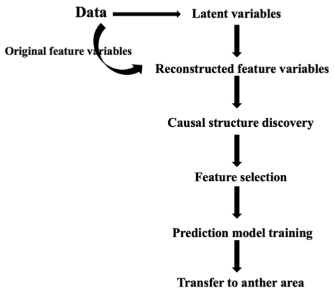
Reconstructed feature variables

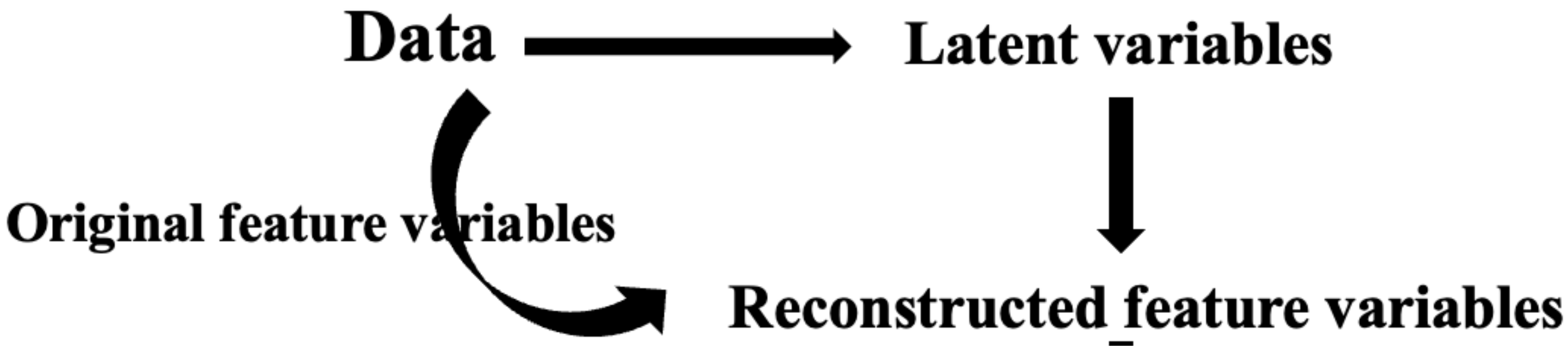
Causal structure discovery

Feature selection

Prediction model training

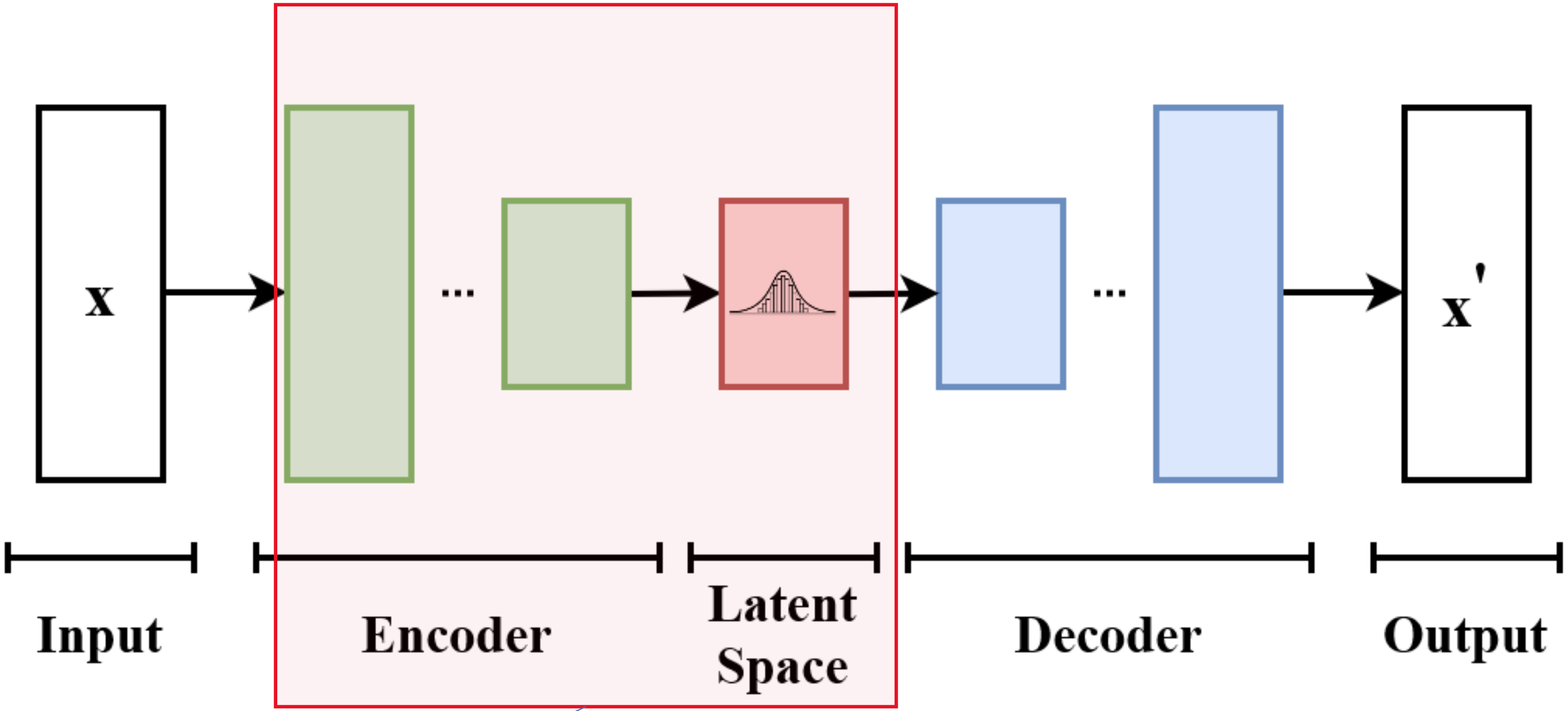
Transfer to anther area

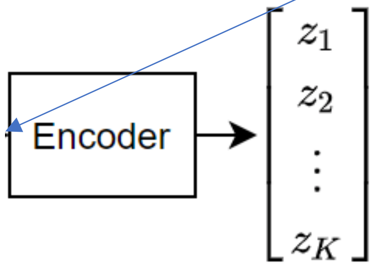




Using VAE to learn the latent feature, we suppose that the latent feature represent the feature that is unobservable but affect the precipitation and the feature variable form data.

VAE(Variational Auto-encoder)





The latent variable is what we need.

Architecture: VAE mainly comprises two parts: an encoder and a decoder.

Encoder: It maps the input data to a latent space (usually a space with lower dimensionality).

Decoder: It decodes from the latent space, attempting to reconstruct the original input.

For our purpose, we need to get latent variables learnt from original data.

So we only need encoder part, after we map the input data to a latent space, we sample a latent vector from latent space, the vector can be n dimensions depends on what we need.

Latent variables

+

Original variables

=

Reconstructed feature variables

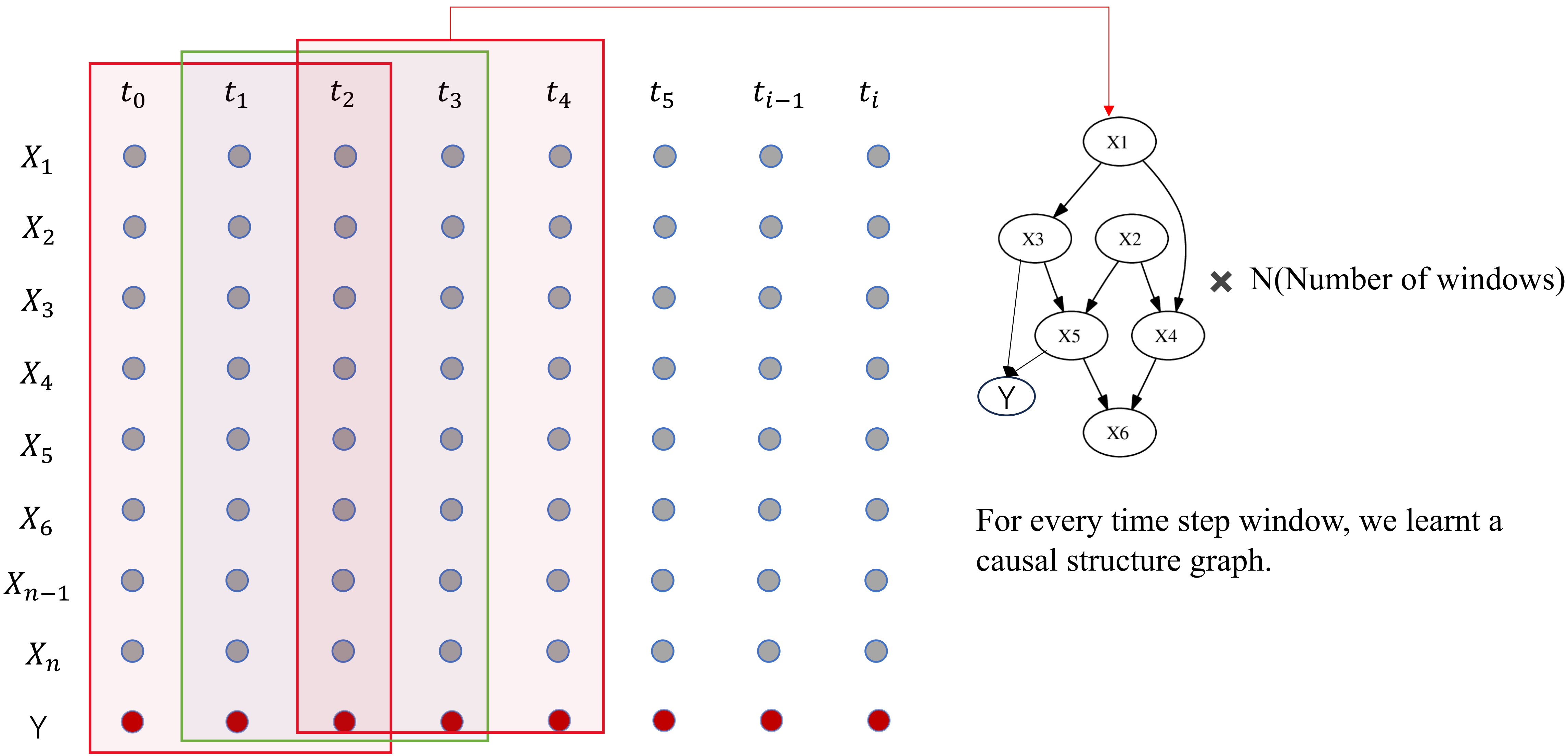
Causal structure discovery

↓

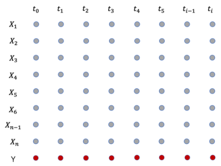
Feature selection

The causal structure we learned is limited in a time window because we suppose the most important feature is based on the recent information.

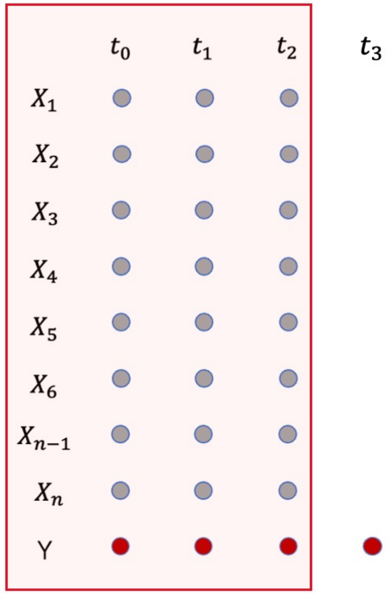
In this case, we use VAE algorithm to learn the latent feature combined with feature variable to learn the causal structure every time window.



For every time step window, we learnt a causal structure graph.



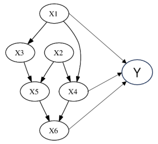
Causal structure learning algorithms



Method	Essence/Principle	Advantages	Disadvantages
Granger Causality	If predicting variable A using its own past values improves by including past values of variable B, then B Granger-causes A.	Simple; well-established.	Linear assumptions; doesn't prove true causality.
Vector Autoregression (VAR)	Multivariate linear model that predicts a variable from its own lagged values and the lagged values of all other variables in the system.	Can model multiple interdependent time series.	Assumes linearity and equal effects over time.
Structural VAR (SVAR)	Extension of VAR with structural (economic) constraints to identify causal impact.	Can model exogenous shocks.	Needs a-priori information.
Transfer Entropy	Measures the amount of uncertainty reduced in predicting future values of variable A by knowing past values of variable B, without assuming linearity.	Captures non-linear dependencies.	Computationally intensive; requires larger samples for accuracy.
Dynamic Bayesian Networks (DBNs)	Graphical models representing probabilistic relationships among a set of variables that change over time.	Can capture complex dependencies and non-linearities.	Requires a good understanding of structure and can be computationally intensive.

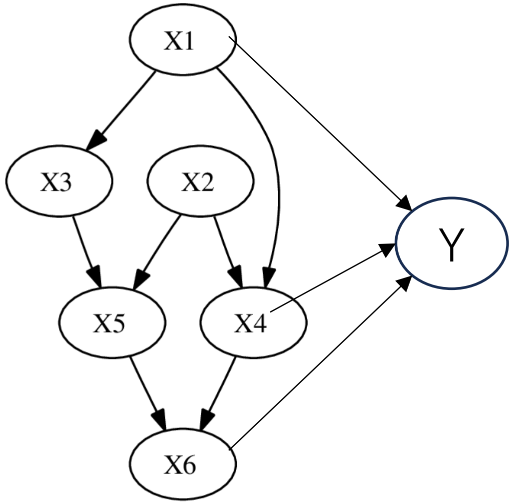
For every time window, we want to know the causal graph from the data in this window(X and Y) and next time step (Y)

Repeat it in every time window. Get *j* causal graphs, combine these causal graphs, count the causal relation lines in causal graph. Delete some low counts relations.



Feature selection

Ideally, finally we can learn a causal graph like this,



One solution is that we choose the variables that have a direct cause to Y.
In this causal graph, we choose X1, X4 and X6 to be our feature.

Another solution might be that using the measure of causal effect. If X have a relatively high causal effect on Y, we say have a relatively high importance

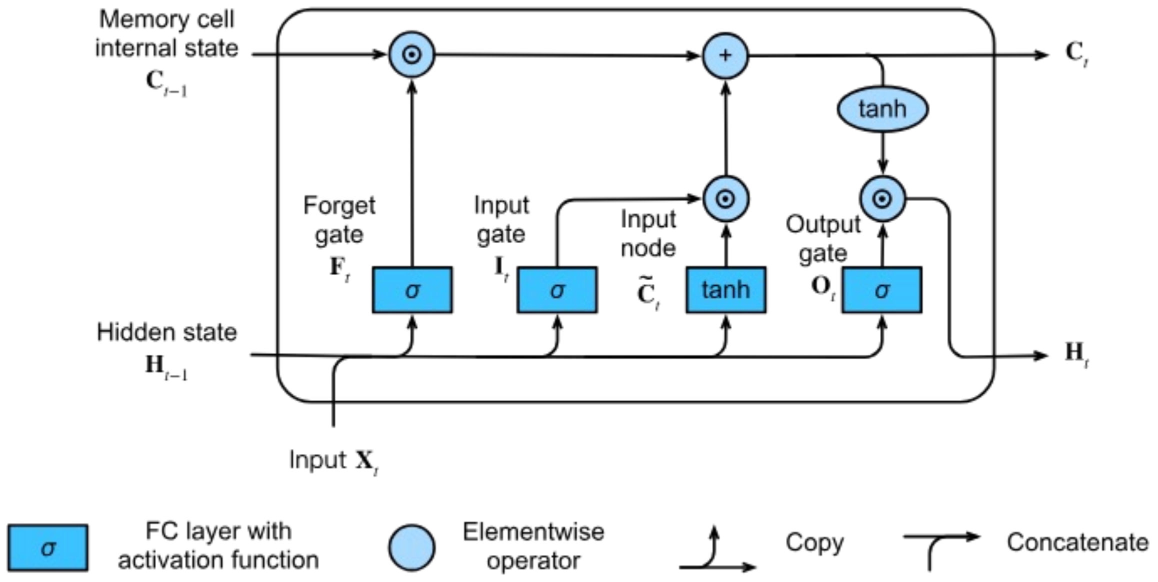
Prediction model training

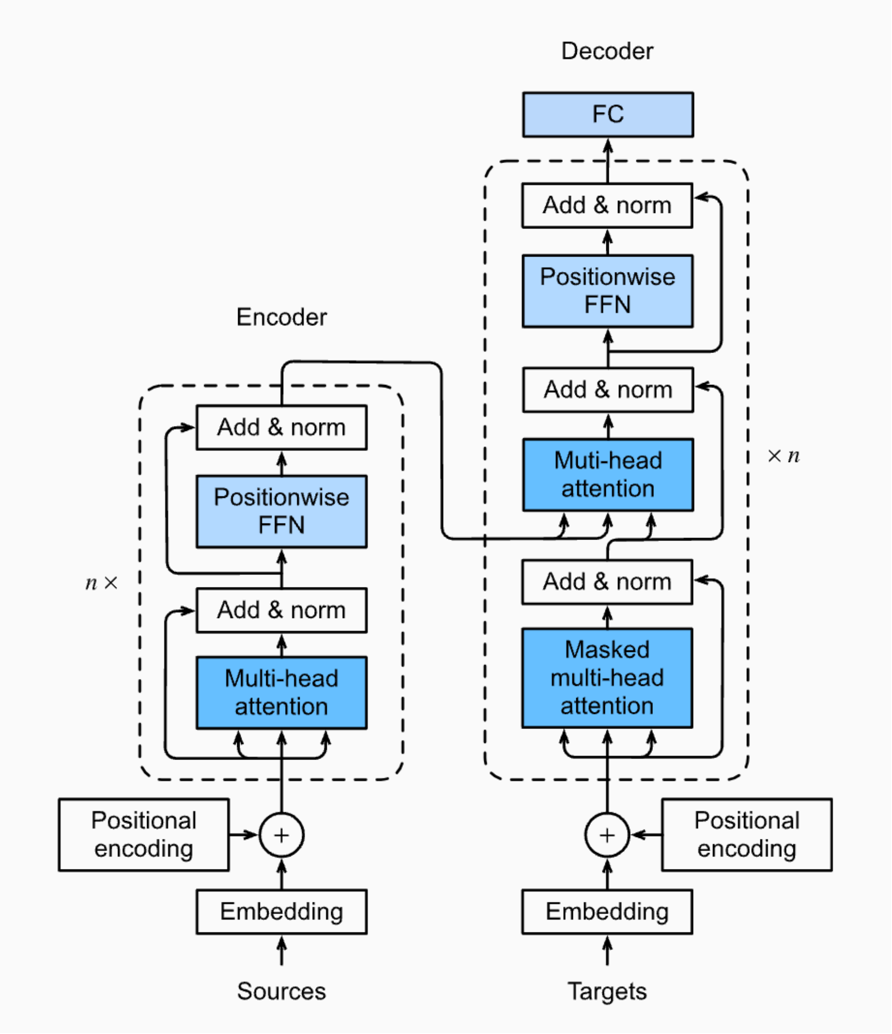
Transfer to anther area

When we select the feature from causal graph, put it in to the prediction model for training. The model we train is a small model which is based on the selected core feature. We hope to make it useful for different area.

And retrain the model based on the small model we have for now.(transfer learning)

Prediction Model





LSTM or Transformer depends on the data scale.

More specifically, in our data, if we use monthly data, LSTM will be better; if we use daily data, Transformer will be better.

After we have got the model, We want it have a good robustness performance.

There are two understandings in this robustness performance.

One is we train our model in an area, and we put it to a very similar area to test its performance. Kind of like open-set prediction because the data in this similar area an brand new data.

The other is that We employ a two-phase approach to model climate change dynamics. Initially, we train a baseline model on region-specific data to capture localized climate patterns. Subsequently, using this as a pretrained foundation, we fine-tune the model on new regional data. This method combines general climate understanding with regional adaptations, striving for both overarching insights and localized precision in predictions.

For research proposal

Title: Develop human-AI interactive models for accountable drought predictions

Introduction: Introduce what method have been used in drought prediction for now, what is the problem of that.
We define this problem as time series prediction problem.
What we want to do is using causality to improve the performance of time series prediction.
List three research questions on this problem.

Literature review: Drought prediction(base climate change bg), Time series prediction(base method), Causal structure learning(from classical to TS), Granger causality.

Research questions and expected outcomes: Describe the questions in details.

Method: Briefly introduce the general solution of the questions, including pipeline and experiment settings.