Stage d’été – Semaine 3

QGNN-TimeCausality
Avancement du projet

Université de Sherbrooke

UNIVERSITÉ TÉLUQ

01

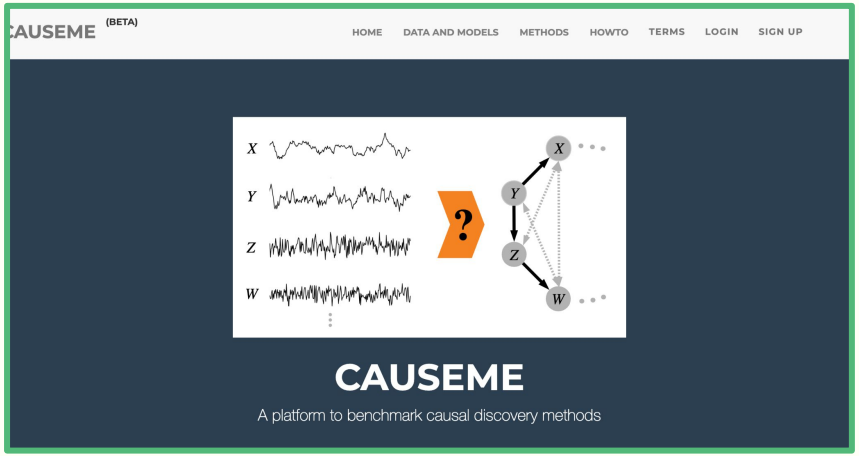
Avancement dans l’implémentation

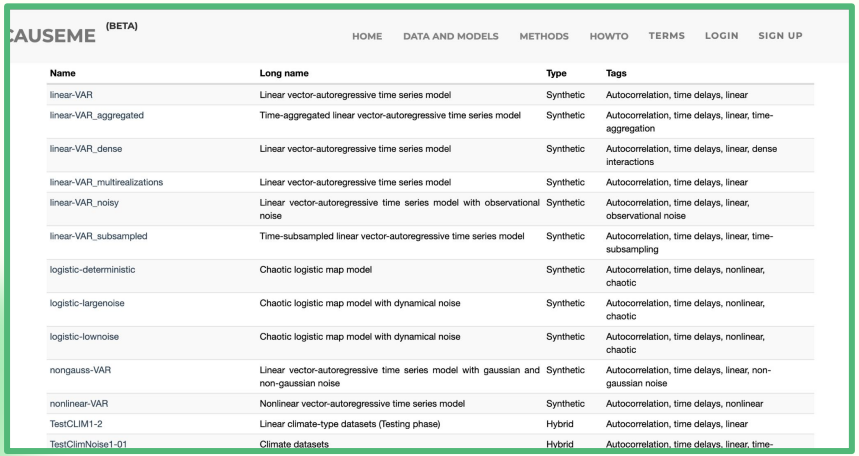
CauseMe

- J’ai exploré la plateforme CauseMe, qui est principalement un site de benchmarking où les utilisateurs partagent leurs méthodes de découverte causale.

- On y trouve quelques jeux de données, mais ils ne sont pas toujours bien documentés et ils ne sont pas associés à un domaine spécifique (données très générales, vecteurs linéaires, climat, etc.).

- À ce stade du projet, ces jeux de données ne semblent pas particulièrement pertinents pour l’entraînement initial du modèle. Cependant, une fois le modèle développé, ils pourraient être utiles pour des tests comparatifs ou pour évaluer la généricité des performances.

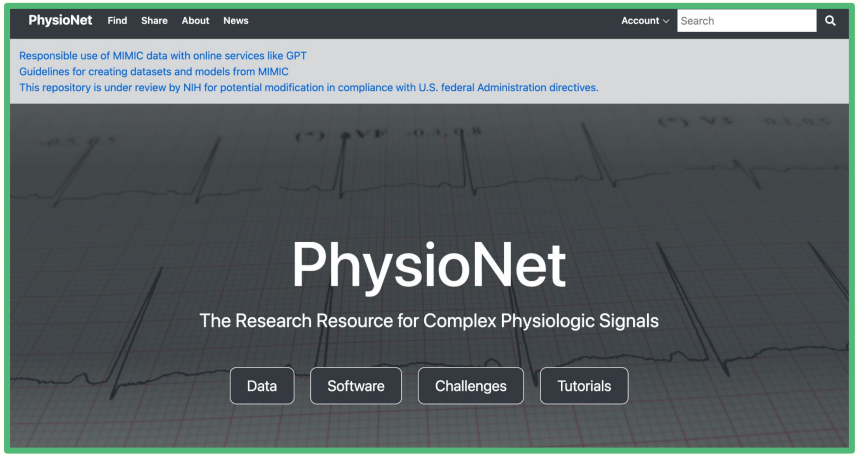




Name	Long name	Type	Tags
linear-VAR	Linear vector-autoregressive time series model	Synthetic	Autocorrelation, time delays, linear
linear-VAR_aggregated	Time-aggregated linear vector-autoregressive time series model	Synthetic	Autocorrelation, time delays, linear, time-aggregation
linear-VAR_dense	Linear vector-autoregressive time series model	Synthetic	Autocorrelation, time delays, linear, dense interactions
linear-VAR_multirealizations	Linear vector-autoregressive time series model	Synthetic	Autocorrelation, time delays, linear
linear-VAR_noisy	Linear vector-autoregressive time series model with observational noise	Synthetic	Autocorrelation, time delays, linear, observational noise
linear-VAR_subsampled	Time-subsampled linear vector-autoregressive time series model	Synthetic	Autocorrelation, time delays, linear, time-subsampling
logistic-deterministic	Chaotic logistic map model	Synthetic	Autocorrelation, time delays, nonlinear, chaotic
logistic-largenoise	Chaotic logistic map model with dynamical noise	Synthetic	Autocorrelation, time delays, nonlinear, chaotic
logistic-lownoise	Chaotic logistic map model with dynamical noise	Synthetic	Autocorrelation, time delays, nonlinear, chaotic
nongauss-VAR	Linear vector-autoregressive time series model with gaussian and non-gaussian noise	Synthetic	Autocorrelation, time delays, linear, non-gaussian noise
nonlinear-VAR	Nonlinear vector-autoregressive time series model	Synthetic	Autocorrelation, time delays, nonlinear
TestCLIM1-2	Linear climate-type datasets (Testing phase)	Hybrid	Autocorrelation, time delays, linear
TestClimNoise1-01	Climate datasets	Hybrid	Autocorrelation, time delays, linear, time-

PhysioNet

- J’ai exploré la plateforme PhysioNet, qui propose un large éventail de jeux de données médicaux.
- Ces jeux de données sont bien documentés et couvrent des thématiques cliniques diverses, ce qui les rend prometteurs et potentiellement très utiles pour le projet.
- Toutefois, tous les ensembles ne sont pas directement exploitables dans le cadre d’une tâche de découverte causale. (Pour que le dataset soit pertinent, il est essentiel qu’il contienne des variables avec des relations causales claires ou exploitables, ou qu’il permette d’en inférer.)
- Je vais donc approfondir mes recherches sur PhysioNet afin d’identifier des jeux de données qui respectent ces critères.



Open databases

- Abdominal and Direct Fetal ECG Database: Multichannel fetal electrocardiogram recordings obtained from 5 different women in labor, between 38 and 41 weeks of gestation.
- A Comprehensive Dataset of Pattern Electroretinograms for Ocular Electrophysiology Research: The PERG-IOBA Dataset: 336 CSV records with 1354 PERG responses (microvolts) from 304 subjects at IOBA. Includes age (years), gender, diagnoses, and visual acuity in logMar scale.
- AF Termination Challenge Database: ECG recordings created for the Computers in Cardiology Challenge 2004, which focused on predicting spontaneous termination of atrial fibrillation.
- AHA Database Sample Excluded Record: Two ECG signals that were excluded from the 1980 American Heart Association database.
- A large scale 12-lead electrocardiogram database for arrhythmia study: A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients
- A multi-camera and multimodal dataset for posture and gait analysis: Multimodal dataset with 166k samples for vision-based applications with a smart walker used in gait and posture rehabilitation. It is equipped with a pair of Depth cameras with data synchronized with an inertial MoCap system worn by the participant.
- A Multimodal Dataset for Investigating Working Memory in Presence of Music: A multimodal dataset containing fNIRS data along with a wide range of physiological signals like EDA, HR, PPG, etc over the course of n-back experiments in presence of music.
- A Multi-Modal Satellite Imagery Dataset for Public Health Analysis in Colombia: Multi-Modal Satellite imagery

Implementation

02

Questions & Prochaines étapes

Questions

- La structure des datasets en general pour ce projet..
- Update concernant l’accès a l’ordinateur quantique de ibm

Prochaines etapes

- Vérifier un peu plus mon raisonnement et mon code
- Rendre l’encodage plus prêt pour les QGNNs (avec PyTorch ou d’autres outils)

Merci de votre attention!

Université de Sherbrooke

UNIVERSITÉ TÉLUQ