Causal detector

Causal representation to build more robust models

Salim Cherkaoui

Work with: Frano, Ahmad, Po-Chien, Yuejiang

Data Science Project : 12 *credits*

EPFL VITA

Summary of the Midterm

- Goal : trying to classify an agent as causal or non-causal to the ego agent in a ORCA simulated trajectory
- Labels (deprecated) : generated from ORCA, a causal agent is an agent used by ORCA when generating the next step of an agent (neighbour at any given time step)

New labels

- We compute the ADE of the trajectory of the ego agent when removing an other agent
- Small (large) ADE : small (large) causal effect of the removed agent on the ego
- Use of predefined thresholds to label the data :
  - ADE < 0.02 : Non-Causal
  - ADE > 0.1 : Causal
  - For simplicity, we ignore what is between
  - (I’ll be using different threshold to evaluate the models)

Simple pair-comparison models

T0 : NC-threshold = 0.001 and C-threshold == 0.01
T1 : NC-threshold = 0.02 and C-threshold == 0.1

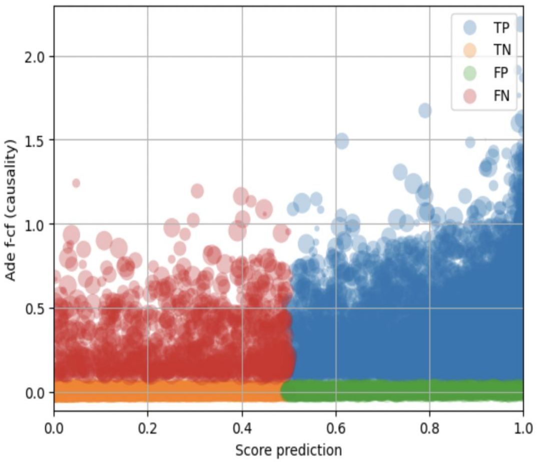
Here, the inputs of the models are the trajectories of the ego-agent and the agent that we want to classify (C or NC), we don't take into consideration the other agents (thus the interactions)

		Accuracy	F1	Specificity
MLP	T0	**0.838**	0.836	0.853
	T1	0.827	0.841	0.811
RN	T0	0.840	0.886	0.831
	T1	**0.851**	0.884	0.822
LSTM	T0	0.829	0.869	0.747
	T1	**0.833**	0.886	0.715

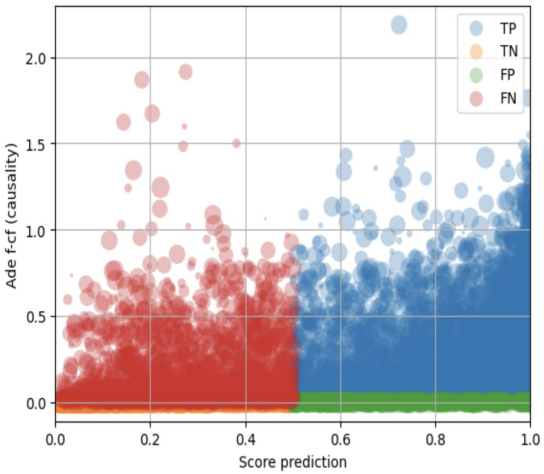
Simple pair-comparison models :

- This simple approaches gives better results that I was getting before, which means that most of cases seems trivials (ex : distance between two agents)
- The following graphs and the results above show that those models don't seem to be very receptive to the ADE

MLP

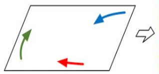


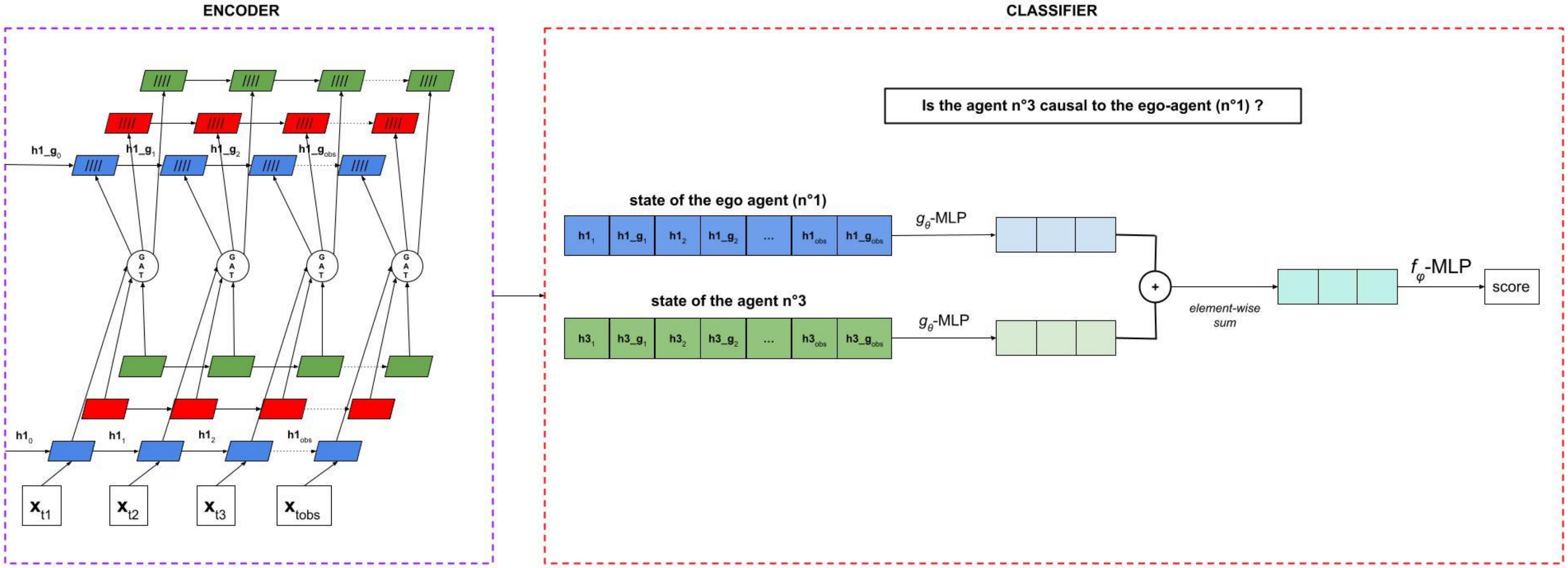
RN



STGAT-like encoder : model architecture

- Pretraining of the encoder : next step prediction
- State of an agent : every hidden states of the LSTM cells’ outputs
- Classifier : Relational Network to compare the state of both agents

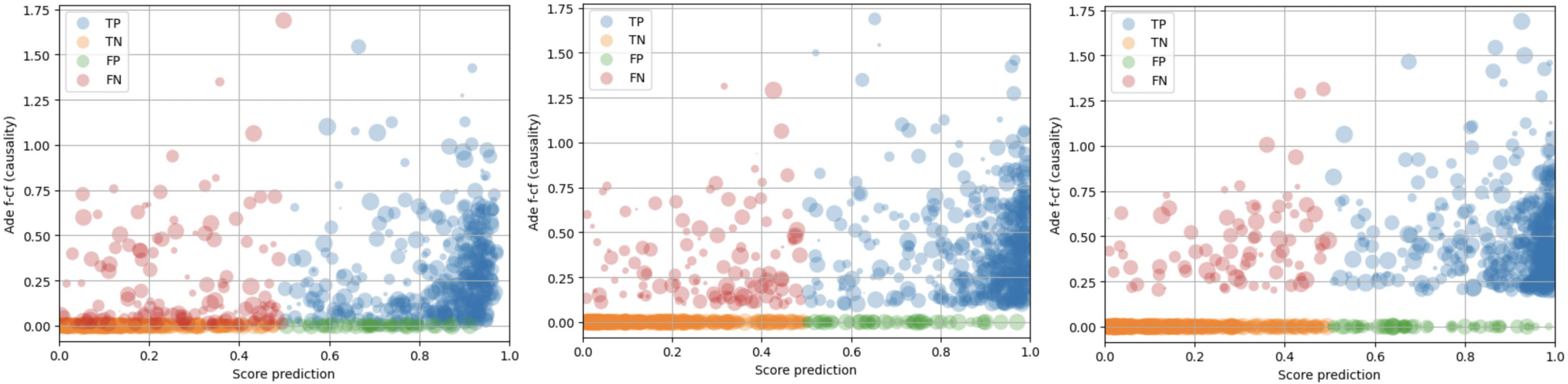




STGAT-like encoder : results

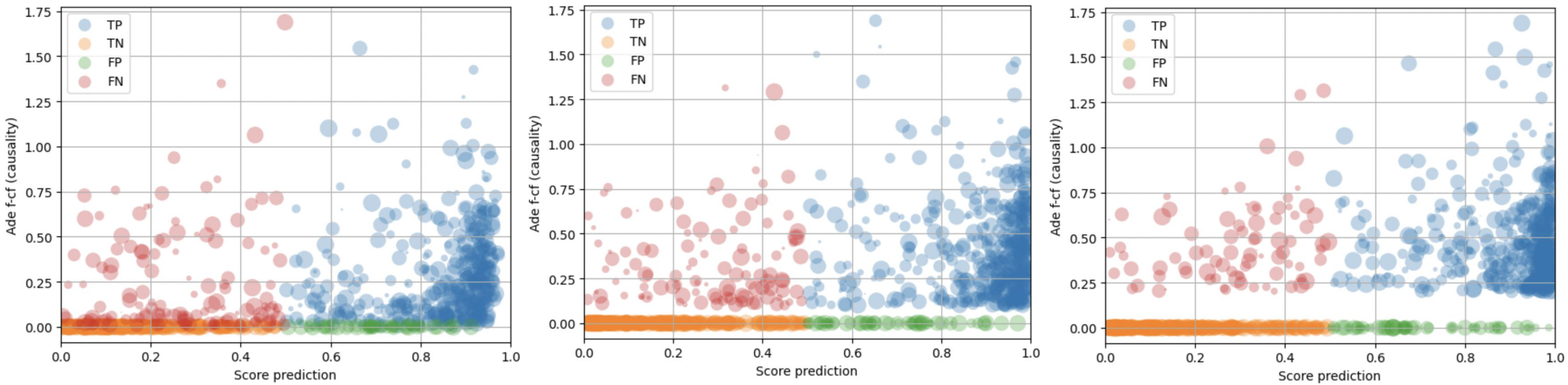
T0 : NC-threshold = 0.001 and C-threshold == 0.01
T3 : NC-threshold = 0.001 and C-threshold == 0.1
T4 : NC-threshold = 0.001 and C-threshold == 0.2

		Accuracy	F1	Specificity
STGAT-like encoder	T0	0.803	0.794	0.844
	T3	0.845	0.839	0.881
	T4	**0.875**	**0.872**	**0.895**



STGAT-like encoder : results

- Results seem more promising
- This model seems to be more “ADE dependant”:
  - The distribution of graphs are closer to the ideal one
  - The model is more sensitive to the threshold changes : the bigger gap between the thresholds and the higher the Causal threshold, the better the results :
    - Since the model focus more on the interactions, and the state of the agent in the all scene



Indirectly causal :

	Dataset	Accuracy	F1	Specificity
STGAT	Complete	**0.845**	0.839	0.881

	ADE label	Causal for n timesteps in ORCA	Accuracy
STGAT	Non-Causal	> 0	0.839
		> 10	0.758
	Causal	== 0	0.512
		< 5	0.552
		< 10	0.481
		< 15	0.829
		> 15	0.884

Indirectly causal :

	Dataset	Accuracy	F1	Specificity
MLP	Complete	**0.880**	0.882	0.861

	ADE label	Causal for n timesteps in ORCA	Accuracy
MLP	Non-Causal	> 0	0.862
		> 10	0.805
	Causal	== 0	0.743
		< 5	0.766
		< 10	0.713
		< 15	0.898

Causal approach :

- Considering our labels, this is a hard approach to implement :
  - Removing an agent from the scene doesn't really change the state of the ego agent since its trajectory remains the same (factual trajectories but without one agent)
  - Using the counterfactual trajectories won't be realistic

		Accuracy	F1	Specificity
C-RN	T0	0.765	0.858	0.409
	T1	0.735	0.799	0.613
C-LSTM2	T0	0.752	0.827	0.476
	T1	0.732	0.803	0.598

Discussion

- Removal of an agent can influence indirectly the ego agent, therefore, it is important to allow the model to reason on the interactions between agents
- Doubts :
  - The labels are computed using the counterfactual trajectory of the ego agent, it is hard to obtain good results using only the factual trajectory
  - A causal approach would make the most sense, but how can we get two comparable factual and counterfactual features without being unrealistic (impossible to get the required data in the real work
    - removing an agent and keeping the factual trajectories is not enough
  - The labels are very constraining
  - Are the models computing causality or are they just evaluating the interactions between agents ?