Comparison between methods of explaining deep survival analysis models

Jakub Krajewski, Stanisław Frejlak, Maciej Wojtala

Plan for the presentation

1. Quick introduction to the problem
2. Methodology
3. Experiments on artificial data
4. Experiments on medical data
5. Computation time
6. Conclusions

Quick introduction

Report under the link

Methodology

- As the neural network model for survival analysis we have decided to use DeepHitSingle

Methodology

- As the neural network model for survival analysis we have decided to use DeepHitSingle
- We have compared explanations created using SurvSHAP(t) and Integrated Gradients

Methodology

- As the neural network model for survival analysis we have decided to use DeepHitSingle
- We have compared explanations created using SurvSHAP(t) and Integrated Gradients
- To highlight problem-specific effects we have used an artificial dataset with 5 variables, where only one had time-dependent effect and one was just a random noise

Methodology

- As the neural network model for survival analysis we have decided to use DeepHitSingle
- We have compared explanations created using SurvSHAP(t) and Integrated Gradients
- To highlight problem-specific effects we have used an artificial dataset with 5 variables, where only one had time-dependent effect and one was just a random noise
- We have also tested the methods on a medical dataset for heart failure

Methodology - cntd.

- To compare both methods, we have aggregated explanation attributions for every variable, dataset and moment in time over the whole dataset

Methodology - cntd.

- To compare both methods, we have aggregated explanation attributions for every variable, dataset and moment in time over the whole dataset
- The plots highlighted mean and standard deviation statistics

Methodology - cntd.

- To compare both methods, we have aggregated explanation attributions for every variable, dataset and moment in time over the whole dataset
- The plots highlighted mean and standard deviation statistics
- We have also used quantitative metrics: Faithfulness Correlation and Average Sensitivity

Methodology - cntd.

- To compare both methods, we have aggregated explanation attributions for every variable, dataset and moment in time over the whole dataset
- The plots highlighted mean and standard deviation statistics
- We have also used quantitative metrics: Faithfulness Correlation and Average Sensitivity
- Unfortunately, these metrics are only implemented for Integrated Gradients

Methodology - cntd.

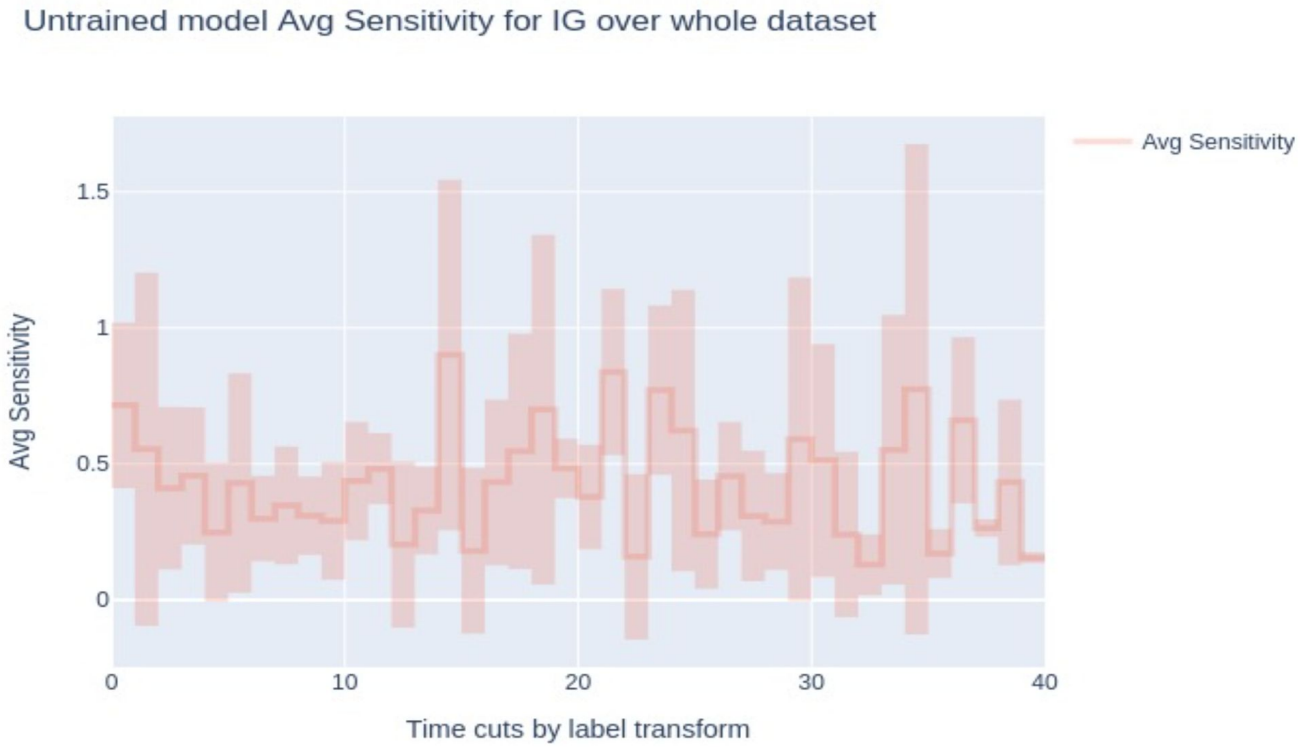
- To compare both methods, we have aggregated explanation attributions for every variable, dataset and moment in time over the whole dataset
- The plots highlighted mean and standard deviation statistics
- We have also used quantitative metrics: Faithfulness Correlation and Average Sensitivity
- Unfortunately, these metrics are only implemented for Integrated Gradients
- Therefore, as a baseline we have used explanations for an untrained model

Experiments on artificial dataset

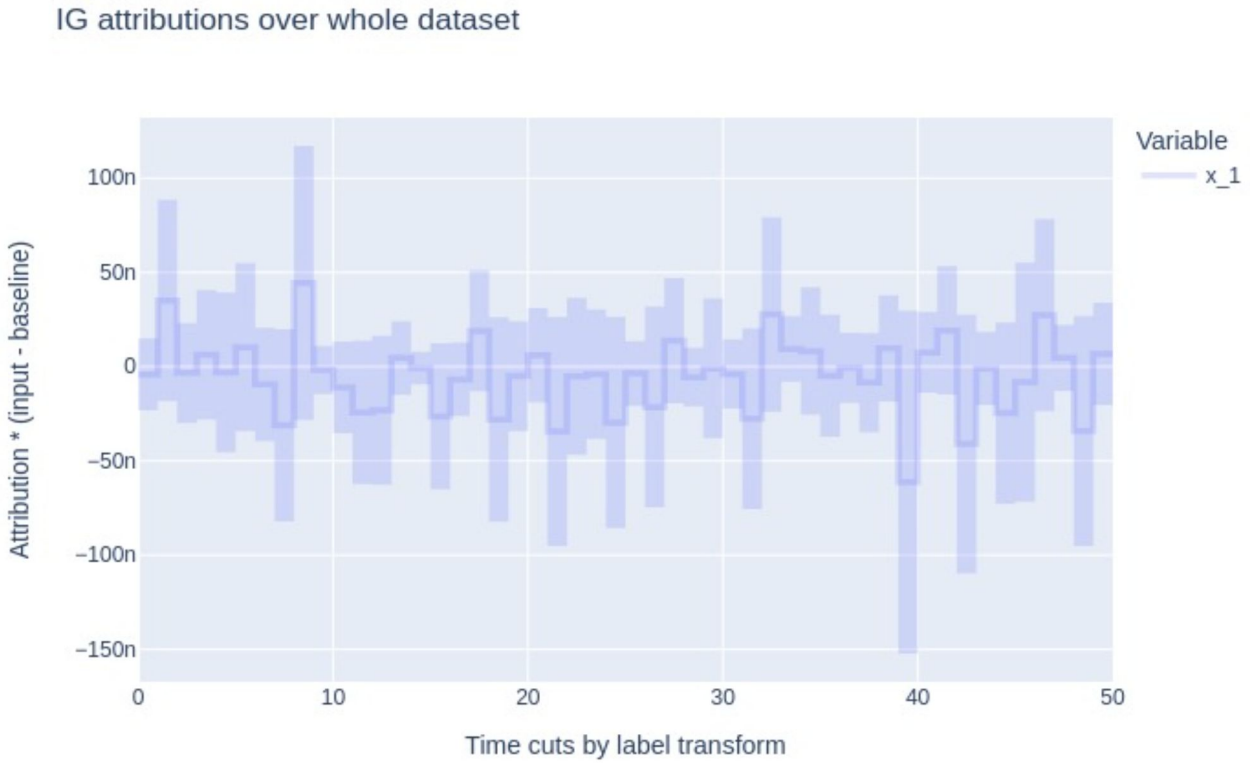
Experiments on artificial dataset

Experiments on artificial dataset

Experiments on artificial dataset



Experiments on medical dataset



Experiments on medical dataset

Experiments on medical dataset

Experiments on medical dataset

Time measurement

Comparison of computation time

Table 1: Time of computation for compared methods on both datasets

	artificial data	medical data
SurvSHAP(t)	102 ms ± 7.82 ms	2.71 s ± 13.6 ms
Integrated Gradients	670 ns ± 16.2 ns	680 ns ± 15.6 ns

Conclusions

- SurvSHAP(t) is a good, problem-specific, but model-agnostic method of explanations

Conclusions

- SurvSHAP(t) is a good, problem-specific, but model-agnostic method of explanations
- The computation time is the main drawback, however it is not a limiting factor in most cases

Conclusions

- SurvSHAP(t) is a good, problem-specific, but model-agnostic method of explanations
- The computation time is the main drawback, however it is not a limiting factor in most cases
- We can largely reduce the computation time

Conclusions

- SurvSHAP(t) is a good, problem-specific, but model-agnostic method of explanations
- The computation time is the main drawback, however it is not a limiting factor in most cases
- We can largely reduce the computation time
- However, this will also come at cost of losing accuracy