

Vysvětlitelnost neuronových sítí

Diplomová práce

Vojtěch Drahý

ČVUT FEL

Otevřená informatika: Kybernetická bezpečnost

Prezentace diplomové práce

20. 6. 2023



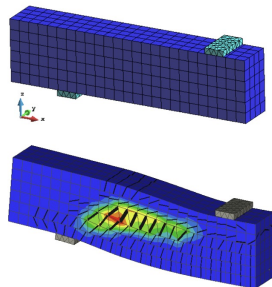
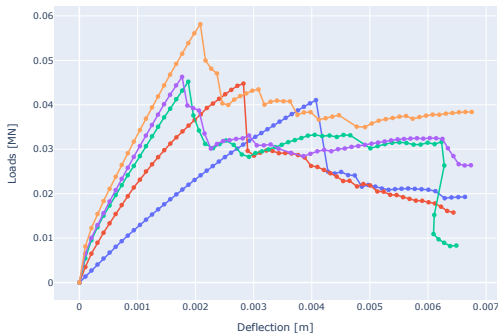
- Neuronové sítě vnímány jako univerzální blackbox.
- Bez porozumění fungování nelze aplikovat na problémy kritické infrastruktury a zdravotnictví.
- Rozsáhlé modely zabírají příliš místa v paměti.

Model	Velikost (MB)	Parametry
VGG16	528	138.4M
ResNet152	232	60.4M
ConvNeXtXLarge	1310	350.1M

- Activation mapping (Grad-CAM).
- Integrované gradienty.
- Ablace.
- Teorie her (Shapleyho čísla).
- Sledování gradientu.
- Korelace mezi výstupy neuronů a labely.

Problematika dat

Určení materiálových parametrů betonové konstrukce dle křivky závislosti zátěže na deflekcii.



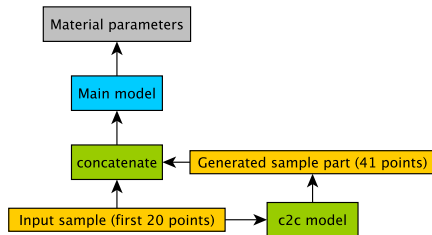
- Nedostatek vzorků.
- Heteroskedasticita.
- Outlieři.
- Nelinearita chování materiálů.

- Úlohy

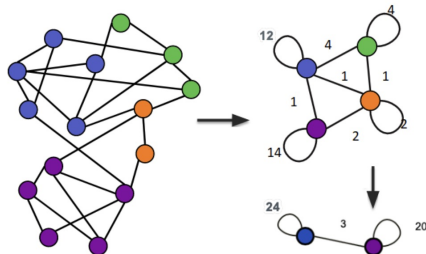
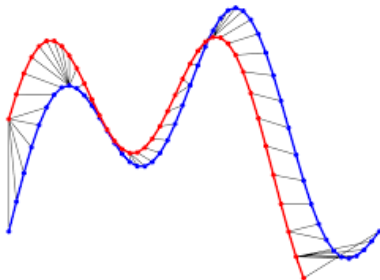
- Zpracování kompletní křivky.
- Zpracování částečné křivky.

- Modely

- CNN, Dense NN, B-LSTM NN.
- Atypické architektury využívající komunity na křivkách vzorků.



- Zpracování vlastností dat dle komunit - snížení výpočetní náročnosti.
- Navržení podobnostního grafu - Dynamic time warping.



- Hierarchická detekce komunit - Louvainská metoda.

	CNN	Dense NN	LSTM NN
Parametry	23 904	71 204	6 762

	L-Maen	L-A-Maen	F-Maen
Parametry	12 308	2 600	33 028

- Pro malý počet vzorků lepší výsledky na atypických architekturách.
- S rostoucím počtem dat Dense NN lépe pojala úlohu.
- Shapley: důležitá pružná, částečně nepružná deformace + chování po rozlomení.
- Sledováním korelací mezi neurony a labely Dense NN zredukována přibližně o 10%.

Vysvětlitelnost a stavba modelů:

- V případě malých datasetů má smysl se zabývat předzpracováním dat.
- Atypické modely dokáží lépe pojmut problém i na malých datech a umožňují snažší vysvětlitelnost.
- Má smysl se zabývat redukcí počtu parametrů u rozsáhlých modelů.

Úloha materiálového inženýrství:

- Nutnost rozšířit dataset.
- Zvýšit frekvenci vzorkování pro pružnou deformaci.