Interview-talk

for Student Assistant position
at XAI Group at Fraunhofer HHI

Ivan Timofeev

Berlin, 11.09.2024

About me

- 4th semester MSc Data Science (Freie Universität Berlin)
- BSc + MSc Physics
- 6 years in experimental research

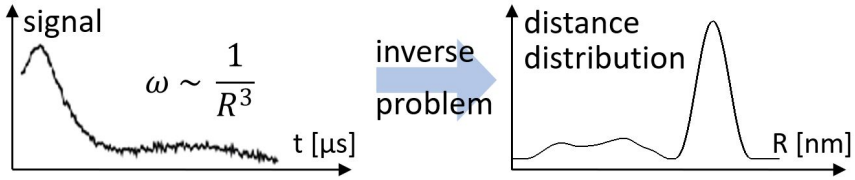
Plan for the talk

- Previous research in bio-chemical physics
- Data Science experience
- Recent project on XAI

Research in bio-chemical physics

Studied properties of biomolecules

using electron magnetic resonance spectroscopy



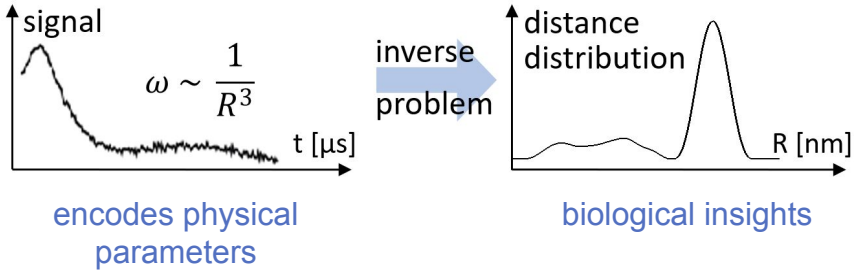
encodes physical parameters

biological insights

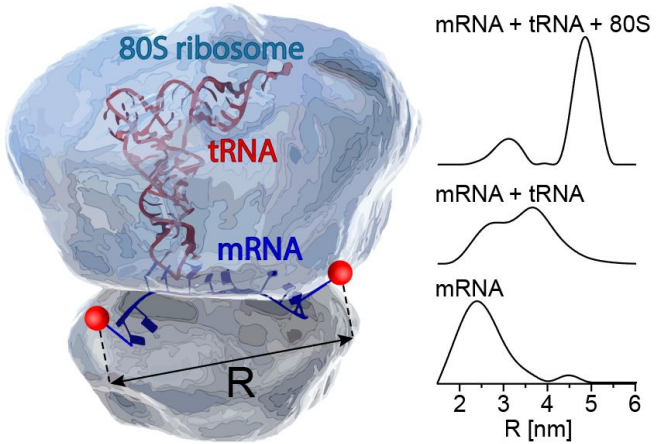
Research in bio-chemical physics

Studied properties of biomolecules

using electron magnetic resonance spectroscopy



- Research directions:
  - Ribosomal complexes
  - Boosting reliability (regression, Monte Carlo simulations, neural network)
  - Method development
- 11 papers (4 as primary contributor)



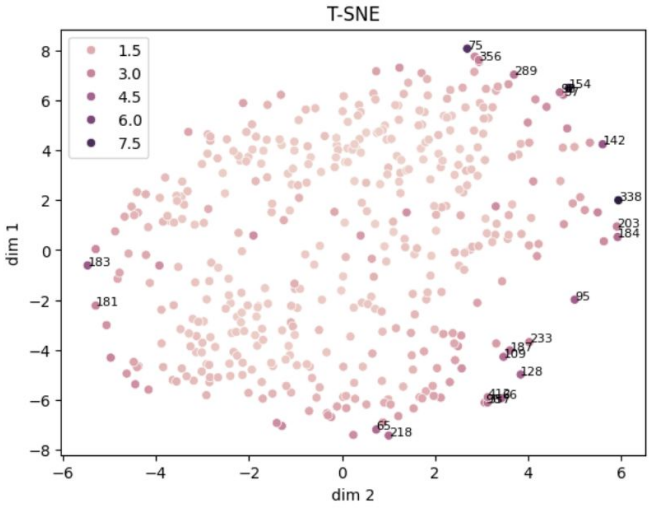
Experience in data science

- 3 years of studying (classic ML, DL, tool proficiency)
- Implemented architectures:
  - Developed NN framework from scratch using numpy
  - Implemented ResNet for defect detection in solar cell images
  - Implemented self-supervised CV models for representation learning (BYOL, contrastive learning)
  - Implemented Transformer for text generation (Attention Is All You Need)
- Tool proficiency:
  - Tools: Unix/Linux, VCS Git, GitHub CI/CD, Docker
  - Languages: Python (PyTorch, TensorFlow), C++, MATLAB, Mathematica, SQL, bash

XAI project: part 1/3

Task: identify anomalies in tabular dataset in unsupervised manner and explain using LRP

- Kernel density estimation $\tilde{p}(\boldsymbol{x}) = \frac{1}{N}\sum_{k=1}^{N}\exp(-\gamma\,\|\boldsymbol{x}-\boldsymbol{u}_k\|^2)$
- *x'* – anomaly threshold $\tilde{p}(\boldsymbol{x}) < \tilde{p}(\boldsymbol{x}')$



[ref] G. Montavon, J. R. Kauffmann, W. Samek, and K.-R. Müller. Explaining the Predictions of Unsupervised Learning Models, volume 13200 of Lecture Notes in Computer Science, pages 117–138. Springer, 2020

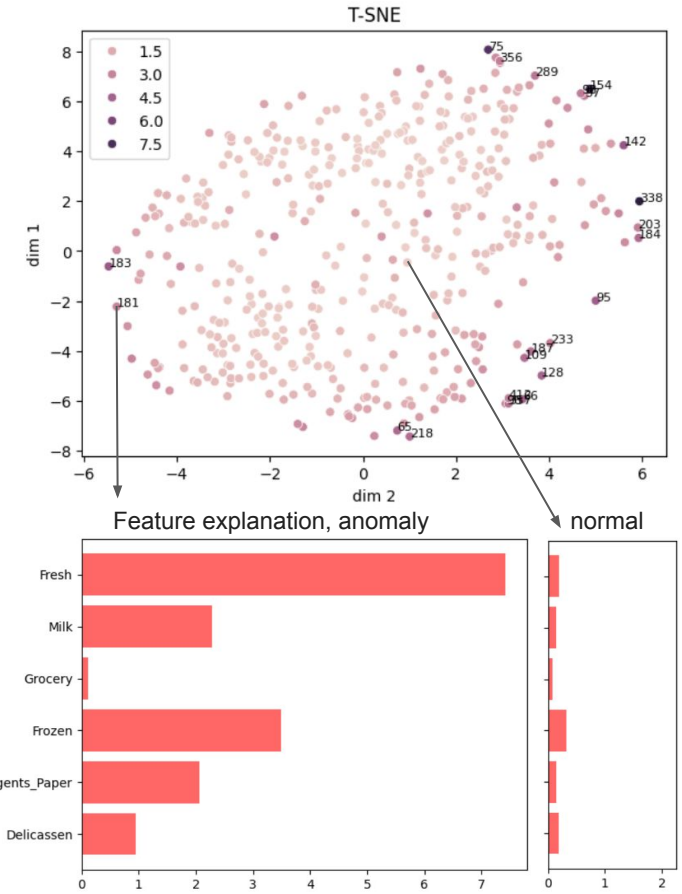
XAI project: part 1/3

Task: identify anomalies in tabular dataset in unsupervised manner and explain using LRP

- Kernel density estimation $\tilde{p}(\boldsymbol{x}) = \frac{1}{N}\sum_{k=1}^{N} \exp(-\gamma \left\|\boldsymbol{x} - \boldsymbol{u}_k\right\|^2)$
- *x'* – anomaly threshold $\tilde{p}(\boldsymbol{x}) < \tilde{p}(\boldsymbol{x}')$
- Layer-wise relevances

$$R_k^{(j)} = \frac{\exp(-\gamma z_{jk})}{\sum_{k \neq j} \exp(-\gamma z_{jk})} \cdot y_j$$

$$R_i^{(j)} = \sum_{k \neq j} \frac{[\boldsymbol{x}_k - \boldsymbol{x}_j]_i^2}{\left\|\boldsymbol{x}_k - \boldsymbol{x}_j\right\|^2} \cdot R_k^{(j)}$$



[ref] G. Montavon, J. R. Kauffmann, W. Samek, and K.-R. Müller. Explaining the Predictions of Unsupervised Learning Models, volume 13200 of Lecture Notes in Computer Science, pages 117–138. Springer, 2020

XAI project: part 2/3

Task: predict atomization energies of molecules based on

- atomic numbers
- 3D coordinates
- atomization energies

- Linear model with pair-encoded representation
  - Explanation: energy(molecule) = Σ energy(atomic pair)
  - Good metrics, but physically incorrect pair energies

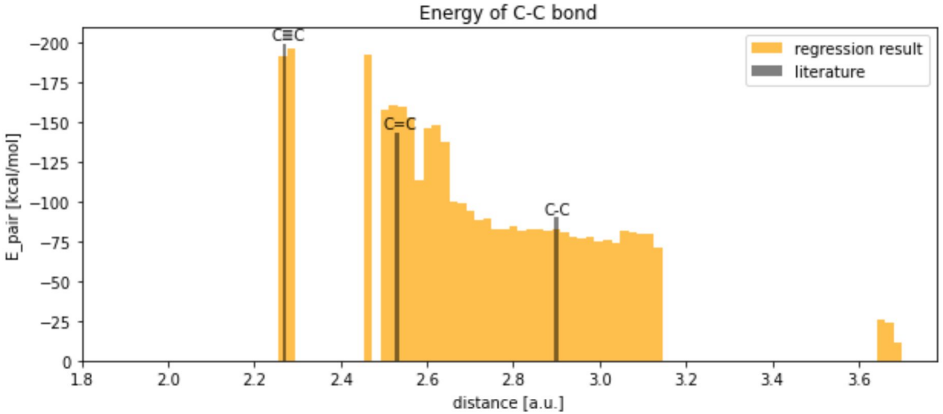
XAI project: part 2/3

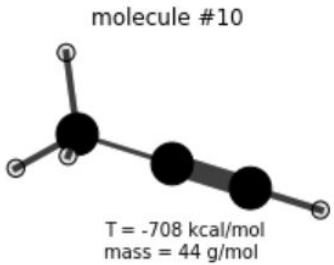
Task: predict atomization energies of molecules based on

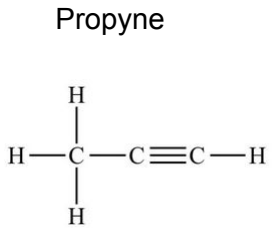
- atomic numbers
- 3D coordinates
- atomization energies

- Linear model with pair-encoded representation
  - Explanation: energy(molecule) = Σ energy(atomic pair)
  - Good metrics, but physically incorrect pair energies

- Physical constraints – meaningful pair energies







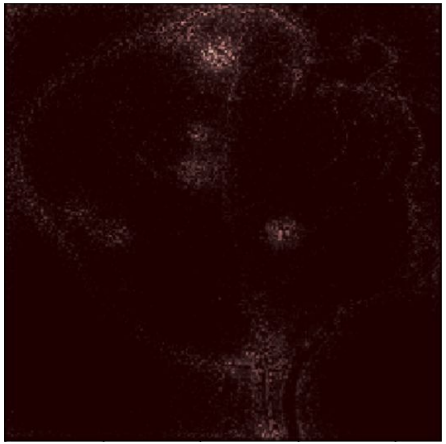
XAI project: part 3/3

Task: gain insights into healthy/diseased leaf dataset

- Pretrained VGG16 + simple binary classifier
- Robust gradient-based explanation – prioritizes excitatory effect over inhibitory effects in the network

label: diseased

gradient score



Final remarks

How I can contribute to the work of the XAI Group at Fraunhofer HHI:

- DL proficiency
- Expertise in multiple scientific domains
- Physics+math background
- Academic background

\+ Overall highly motivated to grow in data science
\+ Interested in applying and developing XAI methods
\+ Open to discussing any available research topic in your group

Thank you for your attention!

XAI project: part 1/3

$$\tilde{p}(\boldsymbol{x}) = \frac{1}{N}\sum_{k=1}^{N}\exp(-\gamma\,\|\boldsymbol{x}-\boldsymbol{u}_k\|^2) \qquad \tilde{p}(\boldsymbol{x}) < \tilde{p}(\boldsymbol{x}') \qquad o(\boldsymbol{x}) \triangleq -\frac{1}{\gamma}\log\tilde{p}(\boldsymbol{x})$$

$$\begin{aligned} h_k &= \|\boldsymbol{x}-\boldsymbol{u}_k\|^2 && \text{(layer 1)} \\ o(\boldsymbol{x}) &= \mathrm{LME}_k^{-\gamma}\{h_k\} && \text{(layer 2)} \end{aligned}$$

$$\mathrm{LME}_k^{\alpha}\{h_k\} \;=\; \tfrac{1}{\alpha}\log\left(\tfrac{1}{N}\textstyle\sum_{k=1}^{N}\exp(\alpha\,h_k)\right)$$

$$R_i = \sum_k \frac{[\boldsymbol{x}-\boldsymbol{u}_k]_i^2}{\epsilon + \|\boldsymbol{x}-\boldsymbol{u}_k\|^2} R_k$$

$$R_k = \frac{\exp(-\beta h_k)}{\sum_k \exp(-\beta h_k)} \cdot o(\boldsymbol{x})$$

[1] G. Montavon, J. R. Kauffmann, W. Samek, and K.-R. Müller. Explaining the Predictions of Unsupervised Learning Models, volume 13200 of Lecture Notes in Computer Science, pages 117–138. Springer, 2020