Explainable AI for attributing Molecules' Structure-Property relationship

SAPIENZA
UNIVERSITÀ DI ROMA

Faculty of Information Engineering, Informatics, and Statistics

Department of Computer, Control, and Management Engineering "Antonio Ruberti"

Master's Degree in Artificial Intelligence and Robotics

Alessio Ragno e Dylan Savoia

Introduction

- Medicinal Chemistry aims to find relationship between molecules' structure and their properties or activities with respect to targets

- AI brought big improvements with the application of ML models to this problem, but due to the complexity of Neural Networks it is difficult to explain which part of the molecule contributes to the final prediction

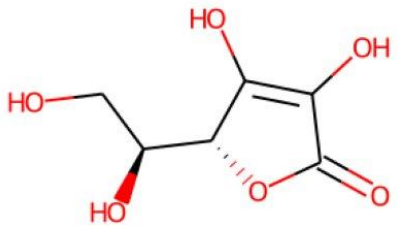
- Explainable AI (XAI) aims to generate attributions to the components of the input in order to visualize salient parts.

Molecules

- The structure can be represented as:
  - Descriptors (weight, number of bonds, etc.)
  - Fingerprints (Binary vectors indicating the presence of substructures)
  - Graphs (nodes are atoms and edges are bonds)
  - SMILES (Simplified Molecular-Input Line-Entry System)

OC[C@H](O)[C@H]1OC(=O)C(O)=C1O

SMILES to Molecule
(Vitamin C)



Our approach

- One recently used approach for predicting molecules activities and properties are Graph Convolutional Networks (GCN).
- Using graph representations allows to exploit both molecule's structure and atoms features.
- Once we have built a GCN model, we calculate input attributions according to the learned parameters.

Graph Convolutional Neural Networks[1]

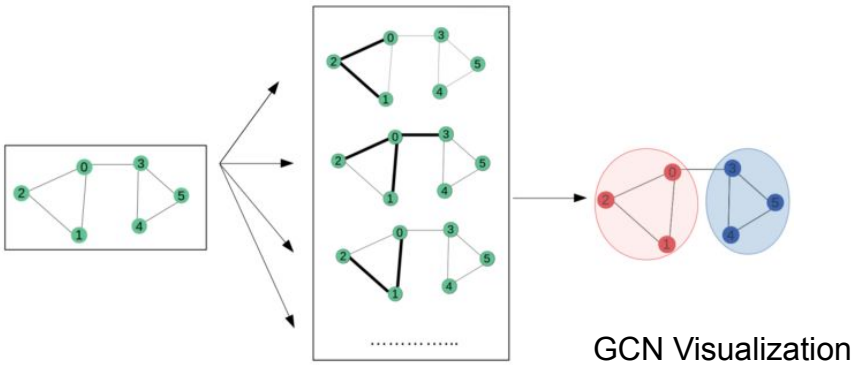
- Information is passed through the graph between adjacent nodes applying an actual convolution through the graph.
- With GCNs we can represent graph-data as an adjacency matrix and a vector of features for each node.
- Outputs a final vector which takes into account the structure of the molecule and the features of each node.

[1] Semi-Supervised Classification with Graph Convolutional Networks: https://arxiv.org/abs/1609.02907

Graph Convolutional Neural Networks (contd.)

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)$$

$\boldsymbol{A}$ being the Adjacency matrix (+ self-loops), $\boldsymbol{D}$ the Degree Matrix, $\boldsymbol{H}$ a features matrix containing the features of each node and $\boldsymbol{W}$ the matrix of learnable weights. $\boldsymbol{\sigma}$ represents an activation function (often a ReLU).



GCN Visualization

Problems

The project is structured in two main parts. We test several XAI methods on two datasets:

- Solubility Dataset: Composed of 1128 records containing molecule structures and their water solubility values from Kaggle[1]
- COVID19 Dataset: Composed of 1669 records containing molecule structures and their activity values against SARS-CoV2 from ChEMBL[2]

[1] https://www.kaggle.com/c/drug-solubility-challenge
[2] https://www.ebi.ac.uk/chembl/document_report_card/CHEMBL4303122/

Problems (contd.)

- First problem works like a benchmark. There are specific known relationships between atom presence and solubility (eg. Oxygen generally improves solubility)
- We can check whether our XAI approach is attributing sensible values to the substructures expected to be relevant
- COVID19 research has been highly active lately. We think XAI could contribute to faster development of effective drugs.

Methods: Simple Gradients Methods

- Working with gradients makes the method simple to compute and independent from the architecture: we can train any model and apply these techniques easily.

- One simple technique often used to give attributions is in fact simply computing gradients of the output w.r.t. the input (“Gradients Method”)

- A slightly improved technique is to multiply those gradients by the input (“InputXGradients”)

Methods: DeepLIFT and Guided-BP

- Guided-BP[1] was proposed to analyze CNNs and specifically tested on object recognition problems. The method is based on a variant of Deconvolutions. For these reasons it may not be suited for our molecules datasets.

- DeepLIFT[2] also computes its attributions w.r.t. a 'reference' (as Integrated Gradients) but it's not a gradients-based method. Scores are computed by a backpropagation-like algorithm and obtained by a single backward pass.

[1] Striving For Simplicity: The All Convolutional Net: https://arxiv.org/abs/1412.6806
[2] Learning Important Features Through Propagating Activation Differences: https://arxiv.org/abs/1704.02685

Methods: Integrated Gradients (IG)

- This method[1] exhibits the **Implementation Invariance** of gradients along with the property of **Sensitivity**.

- Attributions are given by an approximation of a path **integral between a "baseline" input and the actual input** we feed into the network of which we want to compute the attributions.

- A baseline depends on the type of data. For images a good baseline would be a black image (all zeros). Here for molecules we also use a graph made of zero-valued nodes.

[1] Axiomatic Attribution for Deep Networks: https://arxiv.org/abs/1703.01365

Methods: Comparison

- The methods have been compared by Taly et al. using Sensitivity and Implementation Invariance. They then developed IG with the purpose of satisfying both.

- The problem with **gradients** is that they **break Sensitivity**, a property that all attribution methods should satisfy.

- **DeepLIFT breaks** the **Implementation Invariance** property while Deconvolution and **Guided-BP** methods **do not satisfy Sensitivity**.

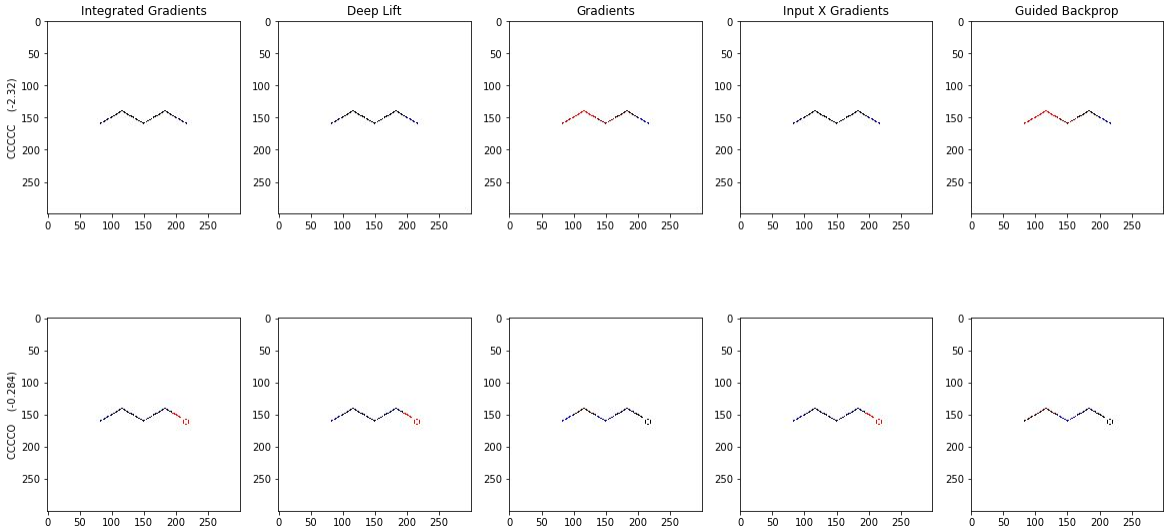
Our Architecture

- The architecture we use is made of **3 GCN** layers alternated with **2 Edge Pooling** layers.
- Edge Pooling tries to merge blocks of connected nodes into one single node giving more importance to chemical groups rather than single atoms.
- The final output is given by the sum of the two pooling layers fed to a fully-connected layer.

Solubility Results

- We know from Chemistry that there are certain rules that influence solubility. Here we expect a high score for the hydroxyl group **(-OH)**

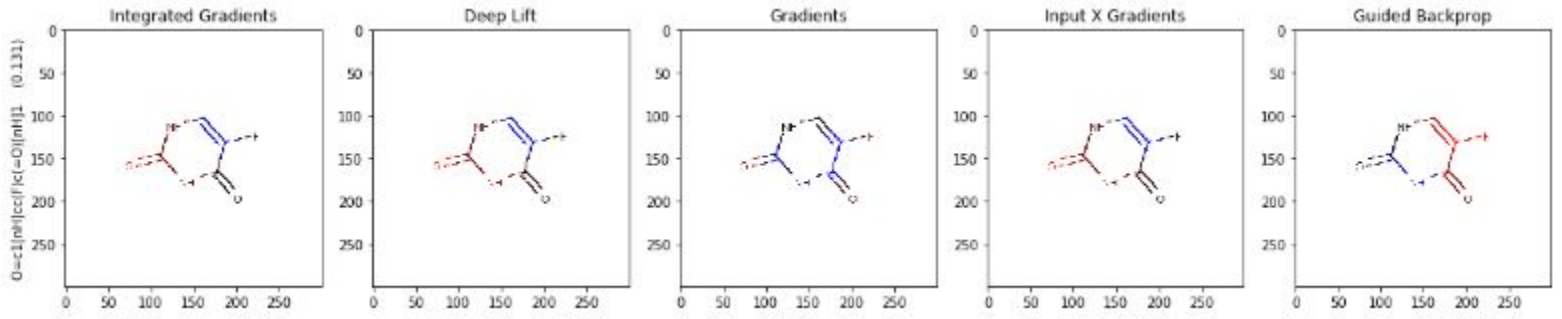
- Not all methods highlight it correctly. IG and DeepLIFT both give sensible attributions. Guided-BP and Gradients are less accurate. InputXGrad despite being simple, performs well.



COVID19 Results

- We do not have the means to validate the attributions for COVID activity but we can check consistency between the different methods

- Similarly to previous results IG, DeepLift and InputXGradients seem coherent while Gradients and Guided Backprop don't support previous methods' attributions.



Conclusions

- With a high-quality dataset it is possible to develop models that are able to predict interesting properties of a molecule from their structure.
- Graph Convolutional Networks are especially suited to work on molecules given their natural representation in graph-form.
- Using XAI methods on these models can give valuable information for faster development of drugs that have a higher activity against a specific target.