Journal Club: Convolutional Networks on Graphs for Learning Molecular Fingerprints (Duvenaud *et al.*)

Stefan O. Gugler

Massachusetts Institute of Technology
Department of Chemical Engineering

May 8, 2018

Abstract

- Convolutional Neural Network on graph
- end-to-end learning of prediction pipelines
- input: arbitrary graphs
- generalization of circular fingerprints
- interpretability & performance

State of the art

- Currently: Fixed size and off-the-shelf fingerprints
- New: Replace fingerprint with a differentiable neural network whose input is a molecular graph
- First layers convolutional, later layers pooling
- benefits: optimized fingerprints outperform off-the-shelf fingerprints
- benefits: don't need 43,000 sized vector because we only use relevant features
- benefits: notion of similarity enhances interpretability

Circular Fingerprints

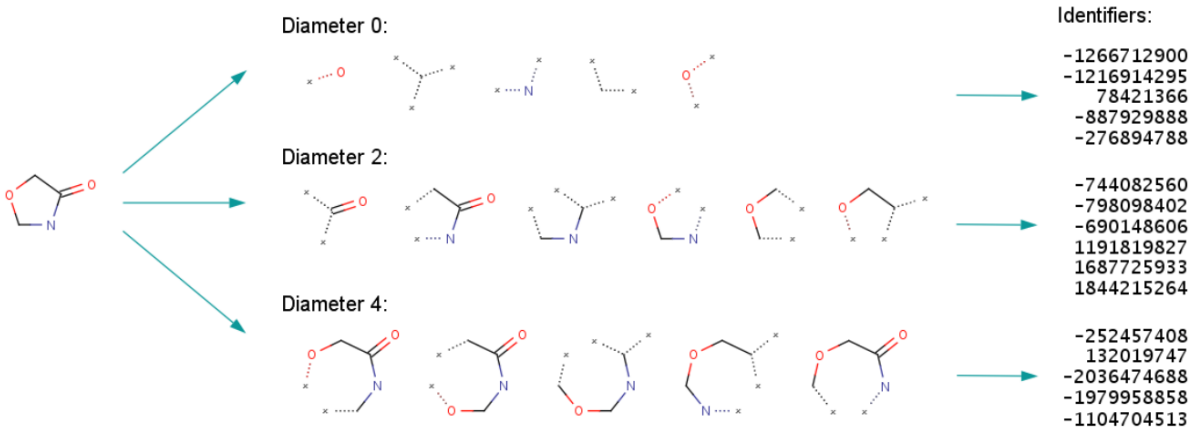
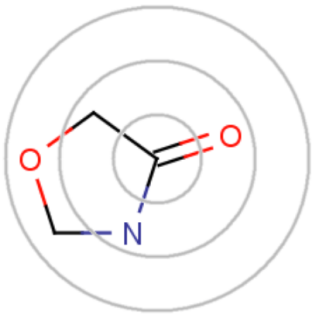


Fig. 2. ECFP generation process



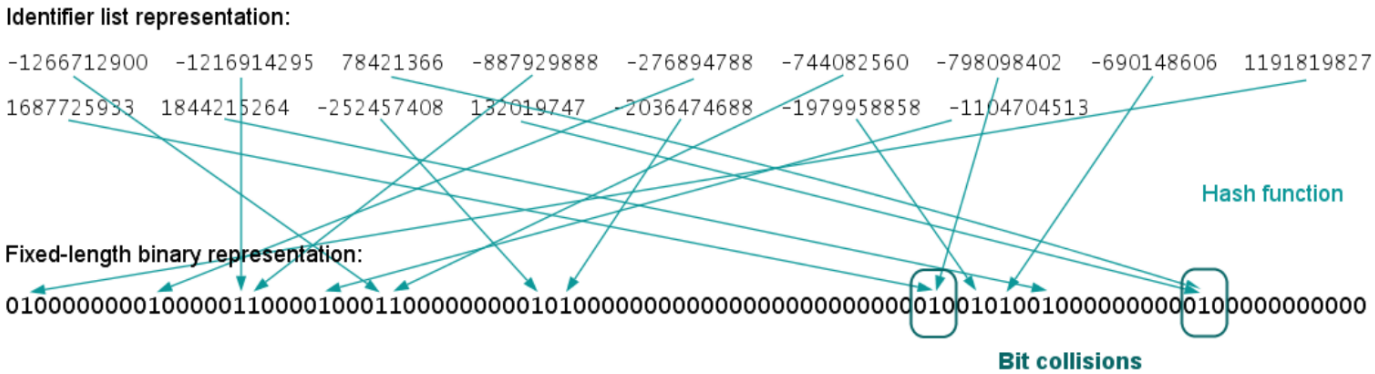


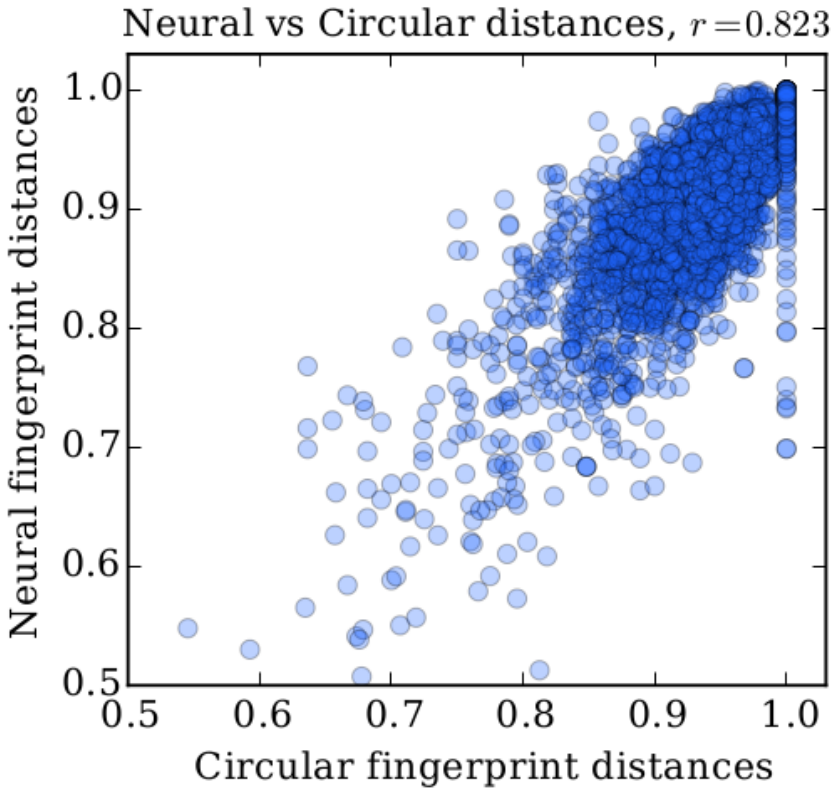
Fig. 3. Generation of the fixed-length bit string ("folding")

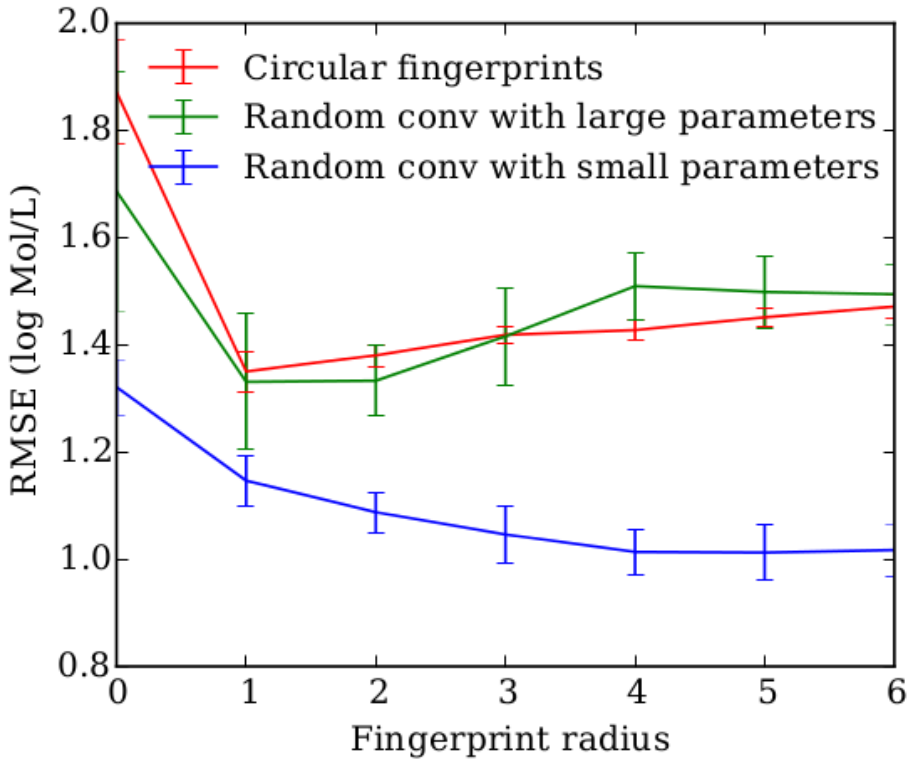
Differentiable Fingerprints

- discrete hash function $\rightarrow$ smooth NN layer
- indexing $\rightarrow$ softmax
- canonicalization: sorting $\rightarrow$ permutation-invariant function like summing

Large random weights in NN fingerprints are the limit of discrete fingerprints.

Comparison circular and neural fingerprints

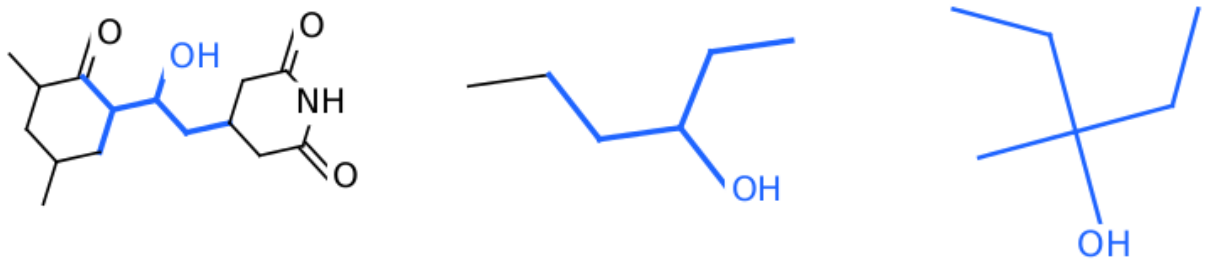




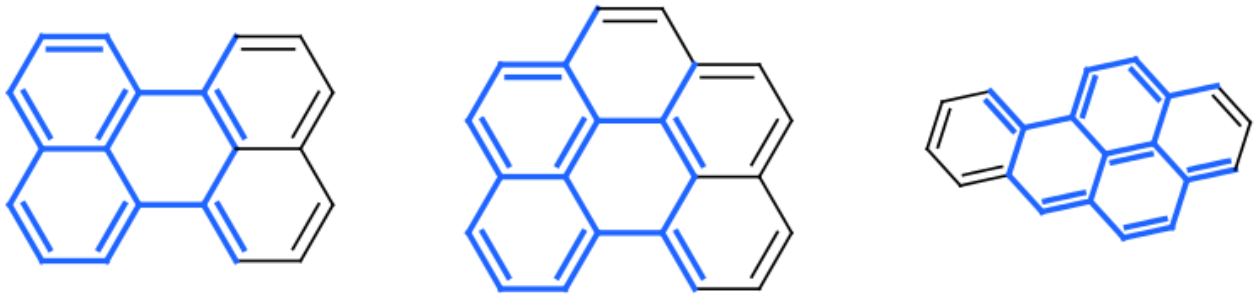
Interpretability

Each feature of the circular fingerprint can only be activated once, hence the long feature vector. Neural features can be activated by similar substructures, resultin gin shorter more interpretable vectors.

Fragments most activated by pro-solubility feature



Fragments most activated by anti-solubility feature



Predictive performance

- Solubility (1,144)
- Druf efficacy (10,000)
- Photovoltaic efficiency (20,000)

Dataset Units	Solubility [4] log Mol/L	Drug efficacy [5] EC_{50} in nM	Photovoltaic efficiency [8] percent
Predict mean	4.29 ± 0.40	1.47 ± 0.07	6.40 ± 0.09
Circular FPs + linear layer	1.71 ± 0.13	$\mathbf{1.13 \pm 0.03}$	2.63 ± 0.09
Circular FPs + neural net	1.40 ± 0.13	1.36 ± 0.10	2.00 ± 0.09
Neural FPs + linear layer	0.77 ± 0.11	$\mathbf{1.15 \pm 0.02}$	2.58 ± 0.18
Neural FPs + neural net	$\mathbf{0.52 \pm 0.07}$	$\mathbf{1.16 \pm 0.03}$	$\mathbf{1.43 \pm 0.09}$

Drawbacks

- Computational cost due to matrix multiplications
  $\mathcal{O}(\text{depth} \cdot \text{fp-length} \cdot \text{features} \cdot \text{atoms} + \text{depth} \cdot \text{atoms} \cdot \text{features}^2)$
- Information Propagation: Worst case $\frac{N}{2}$ layers to distinguish N atoms.
- Enantiomers and *cis/trans* isomers not handled