Spherical convolutions on molecular graphs for protein model quality assessment

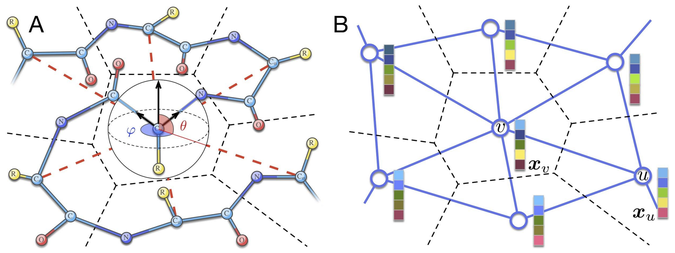
Ilia Igashov[1,2,*], Nikita Pavlichenko[2,*], Sergei Grudinin[1]

[1]Nano-D team, Inria & LJK CNRS, Grenoble, France [2]Moscow Institute of Physics and Technology, Dolgoprudniy, Russia [*]these authors contributed equally

Introduction

Processing information on 3D objects requires methods stable to rigid-body transformations, in particular rotations, of the input data. Irregular topology of graphs makes it challenging to define a rotation-equivariant convolution operation on these structures. In this work, we propose Spherical Graph Convolutional Network (S-GCN) that processes 3D models of proteins represented as molecular graphs. In a protein molecule, individual amino acids have common topological elements. This allows us to construct rotation-equivariant spherical filters that operate on angular information between graph nodes. S-GCN is comparable to state-of-the-art MQA methods, as we demonstrate on Critical Assessment of Structure Prediction (CASP) benchmarks. The proposed technique operates only on geometric features of protein 3D models. This makes it universal and applicable to any other geometric-learning task where the graph structure allows constructing local coordinate systems.

Protein graph



We construct a graph $\mathcal{G}$ of the protein molecule by splitting the surrounding space into cells using the Voronoi tessellation method Voronota [1]. Nodes of the resulting graph correspond to the protein residues and edges are associated with the pairs of residues whose Voronoi cells have a non-zero contact surface. Each node $\mathcal{V}$ of the graph $\mathcal{G}$ contains a feature vector associated with the corresponding protein residue. Feature vector consists of:

- one-hot vector representing the type of amino-acid (20 in total)
- surface of the contact area of the Voronoi voxel
- volume of the Voronoi voxel
- solvent-accessible surface area
- topological distance in the graph to the nearest solvent-accessible atom

Local coordinate system

For each residue we can define the normalized $C_\alpha - N$ vector as the *x*-axis, the unit vector lying in the $C - C_\alpha - N$ plane, orthogonal to *x*, and having positive dot product with $C_\alpha - C$ as the *y*-axis, and the vector product of *x* with *y* as the *z*-axis. Then, given a node $\mathcal{V}$, we can associate each neighbor $\mathcal{U}$ with a pair of spherical angles $\Omega_v^u = (\theta_v^u, \varphi_v^u)$. They specify the angular position of the projection of the node $\mathcal{U}$ onto a unit sphere in the local coordinate system of $\mathcal{V}$.

Spherical convolution

A matrix function $\boldsymbol{F} : S_1 \to \mathbb{R}^{d_1 \times d_2},\ d_1, d_2 \in \mathbb{N}$ that acts on a unit sphere S_1, can be expanded in a polynomial basis using spherical harmonics as the basis functions, and then approximated by cutting the series at the maximum expansion order L:

$$\boldsymbol{F}(\theta,\varphi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \boldsymbol{W}_l^m Y_l^m(\theta,\varphi)$$

$$\approx \hat{\boldsymbol{F}}(\theta,\varphi) = \sum_{l=0}^{L}\sum_{m=-l}^{l} \boldsymbol{W}_l^m Y_l^m(\theta,\varphi).$$

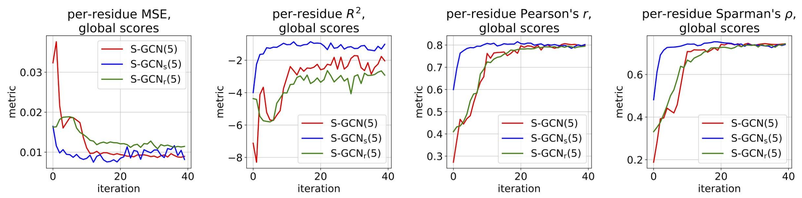
Finally, we can introduce the spherical convolution operation for the vertex $\mathcal{V}$ in the following way:

$$\boldsymbol{F} \circ v = \sum_{u \in \mathcal{N}(v)} \hat{\boldsymbol{F}}(\theta_v^u, \varphi_v^u)\boldsymbol{x}_v\,.$$

Let $\boldsymbol{A}_\Omega$ be a matrix of local angular coordinates for each node. We also denote $Y_l^m(\boldsymbol{A}_\Omega)$ as a result of the elementwise application of the spherical harmonics to the matrix of angular coordinates. Then, *k*th layer of of the S-GCN can be represented as follows,

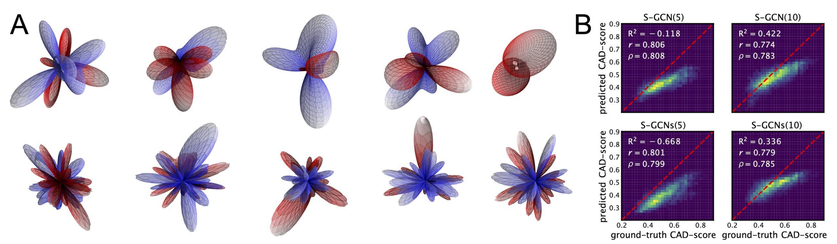
$$\boldsymbol{H}^k = \sigma\left(\sum_{l,m}^{L} Y_l^m(\boldsymbol{A}_\Omega)\boldsymbol{H}^{k-1}\boldsymbol{W}_l^m + \boldsymbol{H}^{k-1}\boldsymbol{W} + \boldsymbol{b}\right).$$

Training



Training was performed on CASP[8-11], the data was augmented with near-native conformations generated with NOLB [2]. Within each iteration, we trained each network in 4 parallel processes feeding 2048 models to each of them. We stored and processed the adjacency matrices in the sparse format.

Experiments



Examples of spherical filters learned by S-GCN of order 5 (top row) and S-GCN of order 10 (bottom row). The distance to the center is proportional to the absolute function value. The red color corresponds to the positive values, the blue color – to the negative ones. (B) Histograms comparing the ground-truth scores and the predictions of spherical graph convolutional networks on the CASP13 dataset.

Method	z-score	Global metrics				Per-target metrics			
		MSE	R^2	Pearson, *r*	Spearman, *ρ*	MSE	R^2	Pearson, *r*	Spearman, *ρ*
SBROD	1.453	0.050	−3.234	0.417	0.433	0.051	−22.455	0.805	0.761
VoroMQA	1.369	0.038	−2.197	0.659	0.688	0.038	−15.930	0.804	0.768
ProQ3	1.459	0.035	−1.969	0.726	0.728	0.035	−17.519	0.775	0.737
Ornate	1.403	0.009	0.193	0.786	0.799	0.009	−2.326	0.814	**0.786**
VoroCNN	1.516	**0.007**	0.368	0.764	0.767	**0.007**	−1.962	0.811	0.771
Baseline	0.865	0.017	−0.424	0.465	0.491	0.017	−6.375	0.648	0.619
S-GCN(5)	1.362	0.013	−0.118	**0.806**	**0.808**	0.013	−3.459	0.789	0.744
S-GCN(10)	1.247	**0.007**	**0.422**	0.774	0.783	**0.007**	**−1.348**	0.722	0.694
S-GCNs(5)	**1.582**	0.020	−0.668	0.801	0.799	0.020	−6.415	**0.820**	0.773
S-GCNs(10)	1.281	0.008	0.336	0.779	0.785	0.008	−1.760	0.742	0.702

Conclusion

In this work, we applied spherical convolutions to capture the 3D structure of a protein graph. The results demonstrate that our method gives a significant improvement in the quality of predictions compared to the baseline without orientational relations between the graph nodes. The spherical convolution method can also be combined with other approaches for the protein model quality assessment, and can also potentially use more input features. Thus, we believe it will be possible to achieve even higher prediction results adding biological and chemical information to the input graphs.

References

[1] K. Olechnovič & C. Venclovas. *J Comput Chem.*, **2014**, 30;35(8):672-8.
[2] A. Hoffmann & S. Grudinin. *J Chem Theory Comput.*, **2017**, 13 (5), pp.2123-2134.
[3] I. Igashov, N. Pavlichenko, and S. Grudinin. **2020**, arXiv:2011.07980.

CASP14 QA group: **graph-sh**